



Bases génétiques du déterminisme du sexe chez le tilapia du Nil en populations naturelles

Cecile Triay

► To cite this version:

Cecile Triay. Bases génétiques du déterminisme du sexe chez le tilapia du Nil en populations naturelles. Sciences agricoles. Université Montpellier, 2021. Français. NNT : 2021MONTG025 . tel-03351827

HAL Id: tel-03351827

<https://theses.hal.science/tel-03351827>

Submitted on 22 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Génétique et Génomique

École doctorale GAIA

Unité de recherche UMR 5554 ISEM

Bases Génétiques du Déterminisme du Sexe chez le Tilapia du Nil en Populations Naturelles

Présentée par Cécile TRIAY
Le 30 Juin 2021

Sous la direction de Hélène D'COTTA

Devant le jury composé de

Yann GUIGUEN, Directeur de Recherche, LPGP Rennes, INRAE

Nicolas PERRIN, Professeur Émérite, DEE Université de Lausanne

Philippe JARNE, Directeur de Recherche, CEFE Montpellier, CNRS

Gabriel MARAIS, Directeur de Recherche, LBBE UBC Lyon 1, CNRS

Astrid Böhne, Group Leader, Zoological Research Museum Alexander Koenig, Bonn

Hélène D'COTTA, Cadre Scientifique des EPIC, ISEM Montpellier, CIRAD

Thomas D. KOCHER, Professor, University of Maryland

Rapporteur

Rapporteur

Président du Jury

Examineur

Invitée

Directrice de Thèse

Co-encadrant de Thèse



UNIVERSITÉ
DE MONTPELLIER

À ma famille,

Papa, Maman, Lise et Mathieu, Mes grands-mères et Jinger.

J'ai une pensée émue pour mon grand-père, amoureux de la nature, qui n'aurait certainement eu d'autres mots que « Bo bo bo... Rien que ca y nous manque » s'il avait su que je ferai une thèse sur le déterminisme du sexe.

**Je ne dédicace pas ces pages à Maxime... Il a tellement travaillé dessus... Elles sont les siennes !
Merci.**

Remerciements

Je souhaiterais tout d'abord remercier les membres du jury qui ont accepté de relire ce travail. J'espère que ce manuscrit donnera lieu à des discussions scientifiques enrichissantes pour nous tous. Par avance, merci.

Evidemment, cette thèse n'aurait pas pu avoir lieu sans Helena et Jean-François. Merci pour votre confiance ! Je pense qu'ensemble nous avons ajouté une petite pierre à l'édifice de la compréhension du déterminisme du sexe chez le tilapia du Nil. Nous avons peut-être soulevé encore plus de questions que nous n'avons réussi à fermer de portes... Mais n'est-ce pas de bonne augure en sciences ? Je vous laisse avec pas mal de boulot sur les bras, alors bon courage !

Je voudrais remercier tout le monde à l'ISEM et plus particulièrement l'équipe EPOISS pour m'avoir si bien accueillie et soutenue dans mes travaux : Bruno, Christelle, Eric, Frédérique, Nicolas, Jean-François A., Antoine, Erwan, Hala, Sélim, Madoka. Bonne continuation !

Il y a de toute évidence les partenaires du projet qui ont permis de le rendre possible et pour en parler, je vais passer à l'anglais pour quelques lignes.

David, Kerry and John, thank you for welcoming me so nicely at the University of Stirling. We eventually managed to produce ddRAD libraries and it's definitely thanks to you! All the best for your futur project and/or retirement!

Thank you also to Astrid, Athimed and Nico for taking great care of me at the University of Basel. I enjoyed a lot my time there and you were all of a precious help. Astrid, thank you so much for the time you invested in this PhD.

I would like to thank the whole team at UMD I had the pleasure to join for a couple of months! I cherish the time spent with you all (especially at Crazy Kebob ;)... what an experience!). It was really awesome to be part of a great team with so much enthusiasm, unity and kindness. I have a very emotional thoughts for Betsy and Matt: of course we accomplished a lot of work together (with intense emotional rollercoasters), but more importantly you welcomed me as if I've always been part of the team, thank you so much, and thank you for your help all along. I wish you all the best for your forthcoming projects.

Karen, I thank you so much for welcoming me not only in your lab, but also in your home. You are forever welcome in South of France!

Tom, you joined this thesis halfway to the end when accepting to endorse the role of Co-supervisor. I'm really glad you did as you gave a whole new turn to this PhD project. I'm sorry you had to listen to my French accent for so many hours on Zoom! Be sure that you gave me two goals in life: First obtain my PhD; and then equip my basement as well as yours for woodwork and metalwork (Though I'll probably won't aim for building a plane and rather keep on with bench and shelves design!).

J'ai aussi eu l'opportunité de côtoyer des partenaires du Burkina Faso. Aboubacar et Rokyatou, merci pour votre implication ! C'était un plaisir de travailler avec vous et j'espère que nous parviendrons à terminer le travail que nous avons commencé. Rokyatou, je te remercie tout particulièrement pour le soutien que tu as m'a apporté pendant ces quelques années. Tu es une femme forte et pleine de sagesse !

Rencontrer les bonnes personnes au bon moment, c'est ce que m'a dit Pascal il y a près de 9 ans lorsque j'ai commencé l'Université. Alors merci Pascal, Pierrick, Mylène, Marion, Christophe, Agnès et Céline d'avoir été des enseignants inspirants et de supers encadrants, au bon moment.

D'aucun sait qu'une thèse est un parcours d'endurance et c'est plus agréable quand on est bien entouré. Antoine, Célia, Yannis, vous avez été là depuis le jour 1 de cette thèse et jusqu'à la fin. On a pleuré et on s'est pas mal énervé mais heureusement on a aussi su rire de la situation parfois. Nous n'aurions peut-être pas su relativiser aussi bien sans Antoni, Maxime, Roger le Bien Né, Etienne... Merci pour tout. On a pour sûr laissé des marques de notre passage et pas uniquement à cause des traces de café sur le canapé.

J'ai la chance d'avoir pu compter sur les copains également. Merci aux Flo², Alex, Loïs et merci à tous les Copiloutres (Vincent, Morgane, Randy, Sarah, Elsa, Sebastien, Florian) d'avoir été présents lorsque j'avais besoin de me vider la tête. Merci Florent de m'avoir épaulé pendant un bon bout de ce parcours universitaire : depuis le BAC il y a eu un sacré chemin. Il y a quelques années, qui aurait cru que nous parviendrions tous les deux à boucler des thèses et que tu aurais un poste avant même que j'ai le temps de soutenir.

Enfin, je veux remercier ma famille. Elle déjà grande et belle et s'est agrandi encore avec vous les Pouget & Courcelle. Je suis consciente de la chance que j'ai de tous vous avoir. Papa, Maman, Lise, Mathieu, mes Maminous et Maxime, vous remercier juste pour la thèse n'aurait pas de sens... Je ne serai pas la personne que je suis sans vous. Merci pour ça mais surtout tout le reste, je vous aime.

TABLE DES MATIÈRES

Introduction générale : Déterminisme du sexe chez les vertébrés	15
I. Historique et mythes des chromosomes sexuels	17
II. Diversité des déterminismes du sexe	19
III. Des transitions entre systèmes qui ne sont pas rares	32
IV. Les téléostéens, champions de la diversité	34
V. Le Tilapia du Nil, un modèle de choix	37
Contexte et objectifs de la thèse	55
Contexte	55
Objectifs	55
Context and objectives of the thesis project	59
Context	59
Objectives	59
Chapitre 1 : Déterminisme génétique du sexe de deux populations naturelles	63
Objectifs	65
Introduction et contexte	65
Matériels et méthodes	66
Principaux résultats et discussion	67
Conclusion	72
Publication associée	77
Analyses complémentaires	119
Chapitre 2 : Diversité et perte de l'haplotype Y ancestral en populations naturelles	127
Objectifs	129
Introduction et contexte	129
Matériels et méthodes	130
Principaux résultats et discussion	132
Conclusion	134
Publication associée	139
Le challenge de l'assemblage des chromosomes sexuels	179
The challenge of sex chromosome assembly	188

Chapitre 3 : Différenciation du sexe chez une population qui ne présente pas le chromosome Y ancestral	197
Objectifs	199
Introduction et contexte	199
Matériels et méthodes	201
Principaux résultats et discussion	202
Conclusion	205
Publication associée	213
Discussion générale et Perspectives	261
Un système ancestral qui garde des zones d'ombres	263
La variabilité au niveau intra-spécifique, un atout pour comprendre l'évolution des chromosomes sexuels	265
Identifier le système de déterminisme du sexe de Hora	266
Les limites de nos études	268
Conclusion	276
Contributions scientifiques et enseignements	285

TABLE DES FIGURES

Introduction générale : Déterminisme du sexe chez les vertébrés _____15

Figure 1.1. Plaques équatoriales observées par Nettie Stevens sur le <i>Tenebrio molitor</i> permettant de mettre en évidence un chromosome plus petit uniquement chez les mâles et démontrant le lien entre chromosome et déterminisme du sexe.	17
Figure 1.2. Déterminisme et changement de sexe chez les vertébrés. GSD et ESD coexistent dans plusieurs clades de vertébrés pour différentes espèces.....	20
Figure 1.3. Possible patrons d'épistasie entre deux locus de déterminisme du sexe.....	24
Figure 1.4. Turnovers de chromosomes sexuels parmi 28 espèces de grenouilles.	25
Figure 1.5. Illustration schématisée des trois patrons de TSD et d'espèces présentant ce type de déterminisme.	29
Figure 1.6. Méthodologie pour déterminer le type de déterminisme d'une espèce (ou une population) présentant ou non des chromosomes sexuels hétéromorphiques.....	31
Figure 1.7. Évolution des systèmes de déterminisme du sexe chez les geckos.	32
Table 1.1. Gènes du déterminisme du sexe formellement identifiés ou fortement suspectés chez les poissons téléostéens.	36
Figure 1.8. Évolution des chromosomes sexuels chez les cichlidés africains.	37
Figure 1.9. Représentation triangulaire schématisée de la complexité du déterminisme du sexe chez le tilapia du Nil.	38
Table 1.2. Bilan des origines, dénominations et déterminants génétiques du sexe chez les souches domestiques et commerciales de tilapia du Nil.	39
Figure 1.10. Localisation géographique des fourchettes de température des environnements d'où proviennent les populations sauvages étudiées dans ce projet de thèse.....	42

Contexte et objectifs de la thèse _____55

Context and objectives of the thesis project _____59

Chapitre 1 : Déterminisme génétique du sexe de deux populations naturelles _____63

Figure 2.1. Illustration de la détection d'une duplication par l'étude de lectures courtes sur un génome complet, visionné dans IGV (Integrative Genome Viewer).....	70
Figure 2.2. Relations entre les régions du chromosome B d' <i>Astatotilapia latifasciata</i> et les LGs de <i>Metriaclicha zebra</i>	71
Figure 2.3. Manhattan plots des $-\log_{10}(\text{p-value})$ issues des Test Exact de Fisher pour l'association des polymorphisme nucléotidique et sexe phénotypique.....	121
Figure 2.4. Emplacement des polymorphismes nucléotidiques simples (SNPs) significativement dimorphiques entre mâles et femelles sur LG19 d'après l'alignement sur le génome de référence d' <i>O. niloticus</i>	121
Figure 2.5. Génotypes de la descendance par rapport à leurs parents inférés à partir des sites significatifs du FET du ddRAD.	122

Figure 2.6. Diagramme en barres du nombre de segments et du nombre de locus identifiés comme appartenant au chromosome B dans le pool de femelles Koka et le pool de mâles Kpandu utilisés dans Triay et al., 2020.	124
--	-----

Chapitre 2 : Diversité et perte de l'haplotype Y ancestral en populations naturelles _____ 127

Figure 3.1. Intérêt du séquençage lectures longues (long reads) pour correctement assembler une région dupliquée.	131
Figure 3.2. Problèmes rencontrés lors de la cartographie d'un individu hétérogamétique sur un génome de référence homogamétique.	181
Figure 3.3. Représentation schématique de graphe et de « bulles » de long reads.....	184
Figure 3.2. Problems encountered when mapping a heterogametic individual on a homogametic reference genome.....	190
Figure 3.3. Schematic representation of graph and « bubbles » of long reads and hybrid assembly...	193

Chapitre 3 : Différenciation du sexe chez une population qui ne présente pas le chromosome Y ancestral _____ 197

Figure 4.1. Modèle de réseaux de gènes antagonistes spécifiques à chaque sexe contrôlant la différenciation sexuelle chez les poissons.....	200
---	-----

Discussion générale et Perspectives _____ 261

Figure 5.1. Aire de répartition géographique se chevauchant pour deux espèces d'Oreochromis.....	274
Figure 5.2. Proportion du génome introgressé en fonction du temps passé depuis l'hybridation.....	274

Contributions scientifiques et enseignements _____ 285

INTRODUCTION GÉNÉRALE : DÉTERMINISME DU SEXE CHEZ LES VERTÉBRÉS

I. Historique et mythes des chromosomes sexuels

On parle d'un déterminisme génétique du sexe lorsque le phénotype sexuel et donc le type de gonades développées par un individu est défini par la présence d'un (ou plusieurs) locus spécifique(s) dans son génome. Ces locus sont appelés déterminants du sexe (SD) et les chromosomes qui les portent sont généralement qualifiés de chromosomes sexuels (en opposition aux autosomes). L'ensemble des gènes portés par la paire de chromosomes sexuels n'influe pas forcément sur le développement des gonades, mais peut parfois affecter la valeur sélective des individus en étant plus favorable pour un sexe que pour l'autre. Différents systèmes de déterminismes génétiques ont été décrits depuis les années 1900 et le début de l'étude des chromosomes sexuels.

I.1. Les débuts de l'analyse des chromosomes sexuels

La première description empirique de chromosomes sexuels remonte à 1905 par Nettie Steven lors de l'observation de caryotypes de ténébrion meunier (*Tenebrio molitor*), petits coléoptères dont les larves sont communément appelées « vers de farine ». Elle constate que les individus possédant 19 gros chromosomes et un petit se développent en mâles alors que les individus possédant 20 gros chromosomes se développent en femelles (Figure 1.1).

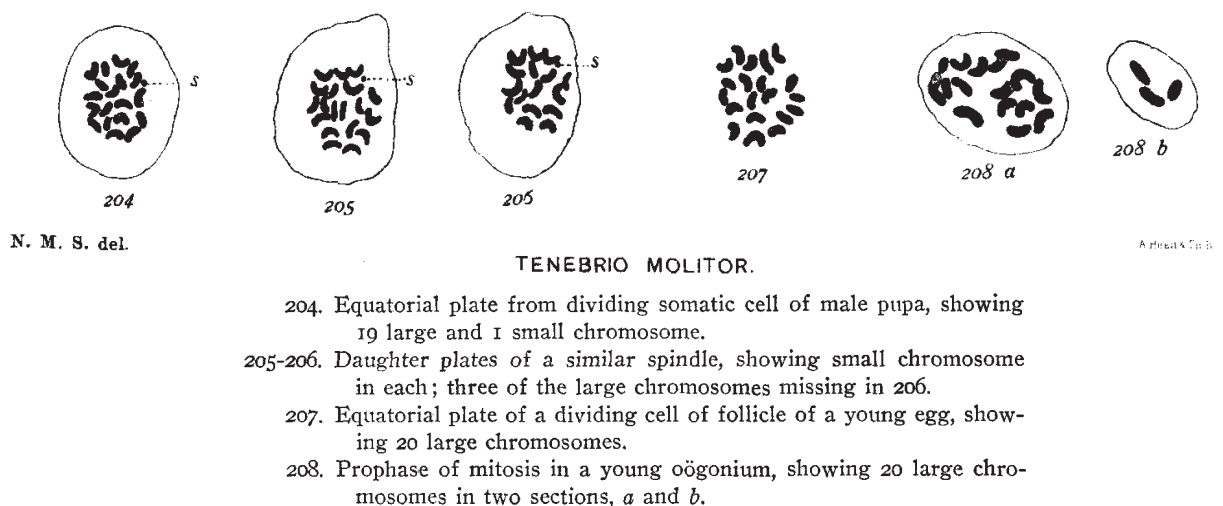


Figure 1.1. Plaques équatoriales observées par Nettie Stevens sur le *Tenebrio molitor* permettant de mettre en évidence un chromosome plus petit uniquement chez les mâles et démontrant le lien entre chromosome et déterminisme du sexe.

Les figures 204, 205 et 206 montrent des plaques équatoriales de cellules somatiques mâles où un chromosome « accessoire » plus petit que les autres est identifié par un S. Les figures 207, 208 (a et b) correspondent des plaques équatoriales d'une cellule en division d'un follicule d'un jeune ovule. Dans les cellules femelles on observe 20 grands chromosomes et le petit chromosome (S) n'est pas présent. Issue de Stevens, 1905.

Le petit chromosome est appelé « chromosome accessoire » et correspond à ce qu'on appelle aujourd'hui un chromosome Y. En effet, le ténébrier meunier est une espèce dont le déterminisme génétique du sexe est de type XX/XY. Les individus peuvent soit porter une paire de chromosomes identiques (deux chromosomes X), soit une paire de chromosomes différents l'un de l'autre (chromosome X et chromosome Y). Les individus portant les chromosomes différents représentent le sexe hétérogamétique, en opposition au sexe homogamétique. Dans le cas du système XX/XY, les mâles sont les individus hétérogamétiques et les femelles sont homogamétiques.

1.2. Des espèces modèles particulièrement étudiées

Certaines espèces ont particulièrement attiré l'attention des biologistes, soit parce qu'elles ont un intérêt pour la société (médical ou économique par exemple), ou parce qu'elles sont pratiques à étudier pour comprendre et illustrer des concepts théoriques. Ces espèces couramment utilisées en laboratoire, appelées espèces « modèles », sont choisies pour leur caractéristiques biologiques commodités pour le maintien d'élevages en conditions contrôlées. Elles ont souvent des cycles de vie courts, des grandes tailles de descendance ou encore des coûts faibles d'entretien (infrastructure ou soin nécessaire). C'est typiquement le cas des drosophiles (mouches à fruits), des petits rongeurs (souris et rats) ou des nématodes.

L'homme fait de toute évidence partie des espèces qui ont souvent été placées au centre des réflexions et notamment dans le cadre de l'étude du déterminisme du sexe. Chez l'humain, comme chez le ténébrier meunier, le sexe est déterminé par des chromosomes sexuels de type XX/XY. Le chromosome X porte de nombreux gènes qui ont été perdus sur le chromosome Y. De fait, on observe une taille réduite du Y, qui est dit « dégénéré ». Parmi les gènes restant sur le Y, seule l'expression dans l'embryon du gène *SRY* (de l'anglais « Sex-determining Region of Y ») suffit à initier la cascade de différenciation des gonades en testicules (Sinclair et al., 1990). Les petits rongeurs étudiés en laboratoire, comme la souris, le rat ou le cochon d'Inde voient également leur sexe déterminé par le gène *Sry*, comme la vaste majorité des mammifères, même si toutes les espèces ne présentent pas systématiquement un chromosome Y aussi dégénéré que celui de l'humain (Koopman et al., 1995). Dans le cas des espèces modèles non-mammifères, telles que les mouches à fruits (*Drosophila melanogaster*) ou les petits nématodes *Caenorhabditis elegans*, les déterminismes du sexe sont comparables à ceux des mammifères (XX/XY) même si *Sry* n'est pas le gène déterminant.

Cependant, l'étude des systèmes très stables de ces espèces de laboratoire et des mammifères a mené à une représentation faussée de l'omniprésence des systèmes de type XX/XY et de leur stabilité dans le règne animal. En réalité, d'autres types de systèmes de déterminisme génétique ont été décrits peu de temps seulement après la première preuve empirique de système XX/XY. Morgan montre en 1909 l'existence de système du type ZZ/ZW chez des insectes. A l'inverse des chromosomes XX/XY, dans un système ZZ/ZW ce sont les femelles qui représentent le sexe hétérogamétique (ZW) alors que les mâles sont homogamétiques (ZZ). Ce système ZZ/ZW est aussi mis en évidence plus tard chez les oiseaux (Ohno et al., 1966) et semble très stable au sein de ce groupe où le W n'est pas toujours fortement dégénéré en comparaison de Z (Ellegren et al., 2010).

Dans les sections suivantes, nous nous intéresserons à la diversité des mécanismes du déterminisme du sexe rencontrés dans une catégorie particulière du règne animal : les **vertébrés gonochoriques** (espèces de vertébrés dont les individus ne portent qu'une seule fonction reproductrice à la fois, mâle ou femelle). En effet, la plupart des vertébrés sont gonochoriques (Devlin & Nagahama, 2001; Herpin & Scharl, 2015; Mei & Gui, 2015), même si l'on retrouve dans de nombreux embranchements de l'arbre du vivant des espèces hermaphrodites (dont chaque individu dispose des deux fonctions mâles et femelles simultanément ou successivement) (Bachtrog et al., 2014; Shärer et al., 2017). Les mécanismes intervenant dans l'hermaphrodisme ainsi que les pressions évolutives qui s'appliquent sur ces systèmes peuvent être très différents de ceux des espèces avec un gonochorisme. Comme dans ce projet de thèse, nous avons concentré notre attention sur le tilapia du Nil (un poisson téléostéen gonochorique), nous ne nous attarderons pas ici sur les systèmes hermaphrodites. Pour une revue très complète des déterminismes du sexe chez le vivant, incluant les modèles hermaphrodites, la revue de Adolphi et al., (2019) est certainement un très bon point de départ.

II. Diversité des déterminismes du sexe

Bien que l'on connaisse des systèmes de déterminismes différents de celui des mammifères ou des espèces modèles, ces mécanismes sont restés pendant longtemps dans les consciences collectives comme des exceptions à la règle. Le développement du séquençage et de la biologie moléculaire a permis au fil du temps de révéler la grande diversité des systèmes déterminismes du sexe chez les vertébrés (voir revues de la littérature

de Valenzuela & Lance, 2004; Bachtrog et al., 2014; Capel et al., 2017; Smirnov & Trukhina, 2019). En effet, des espèces proches peuvent présenter des mécanismes de déterminisme du sexe différents (Figure 1.2). C'est le cas par exemple chez les amphibiens où des systèmes XX/XY et ZZ/ZW peuvent être observés dans différentes populations d'une même espèce (Miura et al., 2017). De même, si les oiseaux présentent un système stable, les reptiles dans leur ensemble présentent une grande variabilité de systèmes incluant des déterminismes génétiques (GSD) et des déterminismes environnementaux du sexe (ESD) ou bien encore une combinaison des deux (Sarre et al., 2011). Dans le cas d'une espèce à ESD, un même génome peut développer deux phénotypes sexuels différents en fonction des conditions environnementales dans lesquelles un individu se développe. De fait, les plus proches parents des oiseaux, les crocodiliens, possèdent eux un déterminisme ESD où la température du milieu contrôle le devenir du sexe des individus (Lang & Andrews 1994; Deeming, 2004). Enfin, les tortues présentent des déterminismes du sexe de type XX/XY, ZZ/ZW et ESD (Ewert et al., 2004). Les sections suivantes s'attèlent donc à donner une idée de la diversité à la fois des mécanismes génétiques et environnementaux chez les vertébrés gonochoriques.

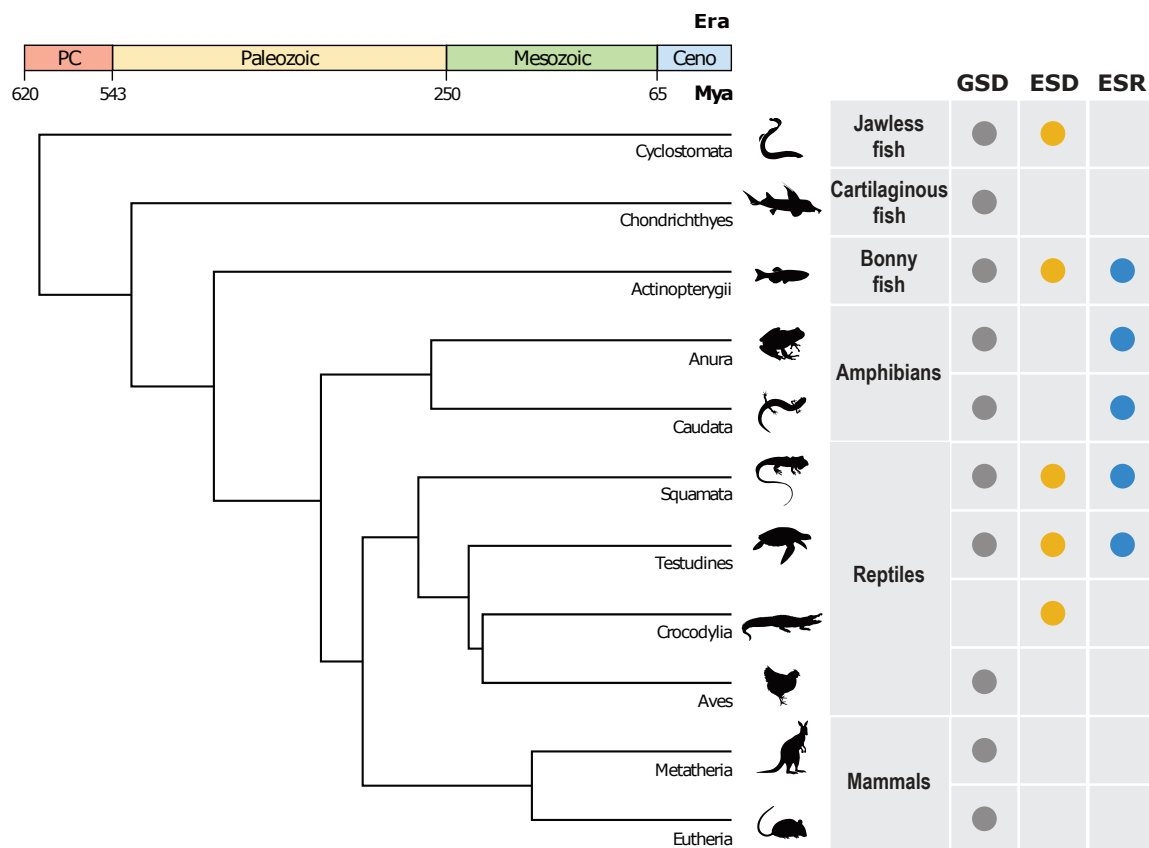


Figure 1.2. Déterminisme et changement de sexe chez les vertébrés. GSD et ESD coexistent dans plusieurs clades de vertébrés pour différentes espèces.

Suite de la légende : L'ESR pendant des fenêtres temporelles critiques affecte la détermination du sexe chez la plupart des clades de vertébrés, à l'exception des euthériens, et il y a un manque de preuves chez les chondrichthyens. Les poissons font preuve d'une plasticité sexuelle remarquable et peuvent subir un ESR suite à des changements dans les conditions extérieures ou même un changement complet de sexe à l'âge adulte. Les temps de divergence utilisés pour construire l'arbre ont été obtenus à partir de la base de données TimeTree. Abréviations : Ceno, Cénozoïque ; ESD, déterminisme environnemental du sexe par l'environnement ; ESR, inversion du sexe par l'environnement ; GSD, déterminisme génétique du sexe ; PC, Précambrien. Issu, traduit et modifié de Ortega-Recalde et al., (2020).

II.1. Les déterminismes génétiques du sexe

II.1.1. Une diversité de formes

Contrairement à l'image reçue du Y dégénéré, dont le contenu en gènes est drastiquement réduit en comparaison du X, les chromosomes sexuels ne sont pas toujours fortement différenciés. On retrouve en effet tous les cas possibles le long d'un gradient allant de la stricte homomorphie à une complète hétéromorphie (Bachtrog et al., 2014). Dans le cas des mammifères, un simple caryotype peut souvent permettre d'identifier la paire de chromosomes sexuels du fait de la différence de taille importante entre le X et le Y et de leur manque d'homologie moléculaire, en lien avec le manque de recombinaison notamment (Skaletsky et al., 2003). Cependant, cette situation est loin d'être la règle et l'on retrouve des poissons téléostéens, des amphibiens ou encore des reptiles dont les chromosomes sexuels sont fortement homomorphes (Miura et al., 2017; Gammerdinger & Kocher, 2018). De plus, dans un même clade on peut également trouver des chromosomes sexuels avec différents degrés de différenciation. Ainsi, les Boidae et Pythonidae présentent des chromosomes sexuels homomorphiques alors que les Viperidea portent des chromosomes sexuels fortement hétéromorphiques et enfin les Colubridae eux possèdent des chromosomes sexuels modérément différenciés (Gamble et al., 2017). C'est aussi le cas chez les oiseaux chez qui les ratites, contrairement à l'ensemble de leurs congénères, présentent des chromosomes sexuels grandement homomorphiques (Yazdi & Ellegren, 2014). Il existe également des cas extrêmes de chromosomes sexuels homomorphiques comme par exemple chez le *Takifugu rubripes*, un poisson chez qui la région de déterminisme du sexe se résume au polymorphisme d'un nucléotide entre le X et le Y (Kamiya et al., 2012).

Il existe donc une grande diversité de formes des paires de chromosomes sexuels.

II.1.2. Une diversité de systèmes

Si les deux types de déterminismes décrits précédemment (XX/XY et ZZ/ZW) sont les plus connus, il en existe de nombreuses variantes.

On connaît par exemple des cas où le chromosome Y n'est plus seulement fortement dégénéré mais a complètement disparu. C'est le cas par exemple chez plusieurs rongeurs fouisseurs du genre *Ellobius* (*E. lutescens*, *E. tancrei* et *E. talpinus*) qui ont perdu le chromosome Y ainsi que le gène *Sry* (Mulugeta et al., 2016; Matveevsky et al., 2017). Chez *E. tancrei* et *E. talpinus* la perte du Y a donné lieu à la duplication du chromosome X : mâles et femelles sont donc XX. En revanche, chez *E. lutescens* tous les individus sont XO. On retrouve également une perte du chromosome Y chez une espèce de rats épineux du genre *Tokudaia* (*T. osimensis*), chez qui des fractions du Y ont été transloqué sur le X ou sur d'autres autosomes. Ainsi, mâles et femelles sont également XO (Murata et al. 2016). Cependant, si la cascade de différenciation du sexe est conservée, le ou les gènes remplaçant le *Sry* ne sont pas identifiés (Otake & Kuroiwa, 2016).

A l'inverse, le Campagnol de l'Oregon (*Microtus oregoni*) a non pas subi une perte du Y mais une perte d'un X chez les femelles. Ainsi, plusieurs génotypes ségrègent mais les femelles de cette espèce peuvent notamment être XO alors que les mâles sont XY (Charlesworth & Dempsey, 2001).

Il existe aussi des espèces pour lesquelles plusieurs paires de chromosomes peuvent interagir pour déterminer le sexe d'un individu. Le *Strabomantis biporcatus* (un amphibien endémique du Vénézuéla, aussi nommé *Eleutherodactylus maussi*) possède deux paires de chromosomes X distinctes. Un facteur masculinisant s'est transposé sur l'un des X d'une des paires, agissant alors comme un Y et résultant en un déterminisme du type X1X1X2X2 chez les femelles et X1Y1X2X2 ou X1X2Y2 chez les mâles (Schmid et al., 2003). Sur un autre modèle, le Poisson-tigre (*Hoplias malabaricus*) possède lui deux chromosomes Y, ce qui induit un déterminisme du type X1X1 (femelles) et X1Y1Y2 (mâles) (Cioffi and Bertollo, 2010). Un exemple impliquant plus encore de paires de chromosomes sexuels est illustré par l'ornithorynque (*Ornithorhynchus anatinus*) qui possède 5 paires de chromosomes sexuels de type XX/XY (Grützner et al., 2004).

Le même type de variant existe pour les espèces dont les chromosomes sexuels suivent un patron ZZ/ZW. Un système avec deux paires de chromosomes sexuels ZW détermine par

exemple le sexe chez un poisson du genre *Ancistrus* (*Ancistrus* sp.2 “*Barcelos*”) où les femelles sont Z1Z2W1W2 tandis que les mâles sont Z1Z1Z2Z2 (De Oliveira et al., 2008). De même une variante du système ZZ/ZW contrôle le sexe chez le poisson-lézard (*Trachinocephalus myops*) et les femelles sont alors ZW1W2 tandis que les mâles sont ZZ (Ueno et al., 2001).

Des systèmes classiques peuvent aussi se cumuler même s'ils font intervenir plusieurs locus qui ne suivent pas forcément le même patron. On retrouve ainsi des espèces pour lesquelles le sexe dépend d'une combinaison de déterminants de type XX/XY et ZZ/ZW. L'épistasie entre les locus définit alors le développement sexuel des individus (Figure 1.3). De fait, on retrouve des espèces qui présentent un déterminisme XX/XY/ZZ/ZW avec $Y > W > Z > X$ ou encore $W > Y > Z > X$. Le sexe du platy (*Xiphophorus maculatus*) est ainsi déterminé par la combinaison de trois chromosomes sexuels : W, Y, et X, avec $W > Y > X$. Les femelles de cette espèce peuvent alors être XX, XW, ou YW, et les mâles peuvent être YY ou XY (Schultheis et al., 2009). *Xenopus tropicalis*, un crapaud d'Afrique de l'ouest, possède également trois chromosomes sexuels Y, W et Z avec les relations suivantes : $Y > W > Z$. Les femelles sont alors WW ou ZW et les mâles sont YZ, YW ou ZZ (Roco et al., 2015).

Finalement on peut également retrouver des systèmes de déterminisme génétique où le sexe est défini par un chromosome surnuméraire et non essentiel dans le caryotype et pour le développement des individus. Ce chromosome est généralement appelé chromosome « B », en opposition au jeu de chromosomes originels de l'espèce (« A »). Les chromosomes B peuvent endosser le rôle de chromosomes sexuels s'ils portent un locus qui agit comme un déterminant du sexe. Ainsi, le sexe des individus est déterminé en fonction de l'interaction du chromosome B avec les systèmes existant avant son apparition dans le génome. C'est le cas chez *Metriaclimma lombardii*, un cichlidé du lac Tanganyika chez lequel le chromosome B agit comme un chromosome W, en plus de la présence d'un système XX/XY ancestral (Clark et al., 2017). Dans cette espèce, les relations de dominances entre les chromosomes sont les suivantes : $W > Y > X$ et les individus WXY se développent en femelles tout comme les individus XX, alors que les individus XY se développent classiquement en mâles.

$Y > X > W > Z$	ZZ	ZW	WW
XX			
XY			
YY			

$W > Z > Y > X$	ZZ	ZW	WW
XX			
XY			
YY			

$Y > W > X > Z$	ZZ	ZW	WW
XX			
XY			
YY			
$W > Y > Z > X$	ZZ	ZW	WW
XX			
XY			
YY			
$Y > W > Z > X$	ZZ	ZW	WW
XX	*		
XY			
YY			
$W > Y > X > Z$	ZZ	ZW	WW
XX	*		
XY			
YY			

Figure 1.3. Possible patrons d'épistasie entre deux locus de déterminisme du sexe.

Modèles possibles d'épistasie simple dans le cas de deux loci déterminant le sexe. Les relations épistatiques entre les allèles sont indiquées par les systèmes d'inégalités à côté de chaque tableau. Le nombre de systèmes possibles est limité par les modèles de dominance à chaque locus ($Y > X$ et $W > Z$). Le sexe attendu de chaque génotype est indiqué par une couleur (bleu = mâles, rose = femelles). Les trois systèmes de gauche représentent un Y épistatiquement dominant, tandis que les trois systèmes de droite représentent un W épistatiquement dominant. Des différences dans les relations épistatiques des allèles récessifs (X et Z) peuvent inverser le phénotype attendu du locus épistatiquement dominant (*). Issu et traduit de Tao et al., (2021).

II.1.3. Les évènements entraînant des turnovers de chromosomes sexuels

La grande diversité des systèmes décrits précédemment montre une plasticité des déterminismes génétiques du sexe chez les vertébrés et reflète les nombreux évènements de changement de chromosomes sexuels (turnover) qui ont eu lieu dans l'histoire évolutive de ces taxons. Les mécanismes qui permettent ces turnovers et les évènements qui les favorisent sont introduits ci-après.

a) Les mécanismes de turnovers

Les turnovers de chromosomes sexuels sont classés en deux catégories : les turnovers homologues et les turnovers non-homologues. Le premier cas correspond au déplacement du déterminant sexuel sur une autre paire d'autosomes. Le déterminant reste donc le même mais la paire de chromosomes sexuels change. C'est ce phénomène qui a été identifié chez le saumon Atlantique où le gène du déterminisme du sexe (*sdY*) peut être retrouvé sur différentes paires de chromosomes selon les souches (Lubieniecki et al., 2015).

Dans le cas d'un turnover non-homologue en revanche, il s'agit d'un nouveau locus sur une autre paire de chromosomes qui prend la tête de la cascade du déterminisme du sexe. Ainsi, chez les grenouilles, 13 turnovers de chromosomes sexuels ont été identifiés sur 28 espèces (en un laps de temps de 55 millions d'années) avec certains chromosomes recrutés plus fréquemment que d'autres (Figure 1.4) (Jeffries et al., 2018).

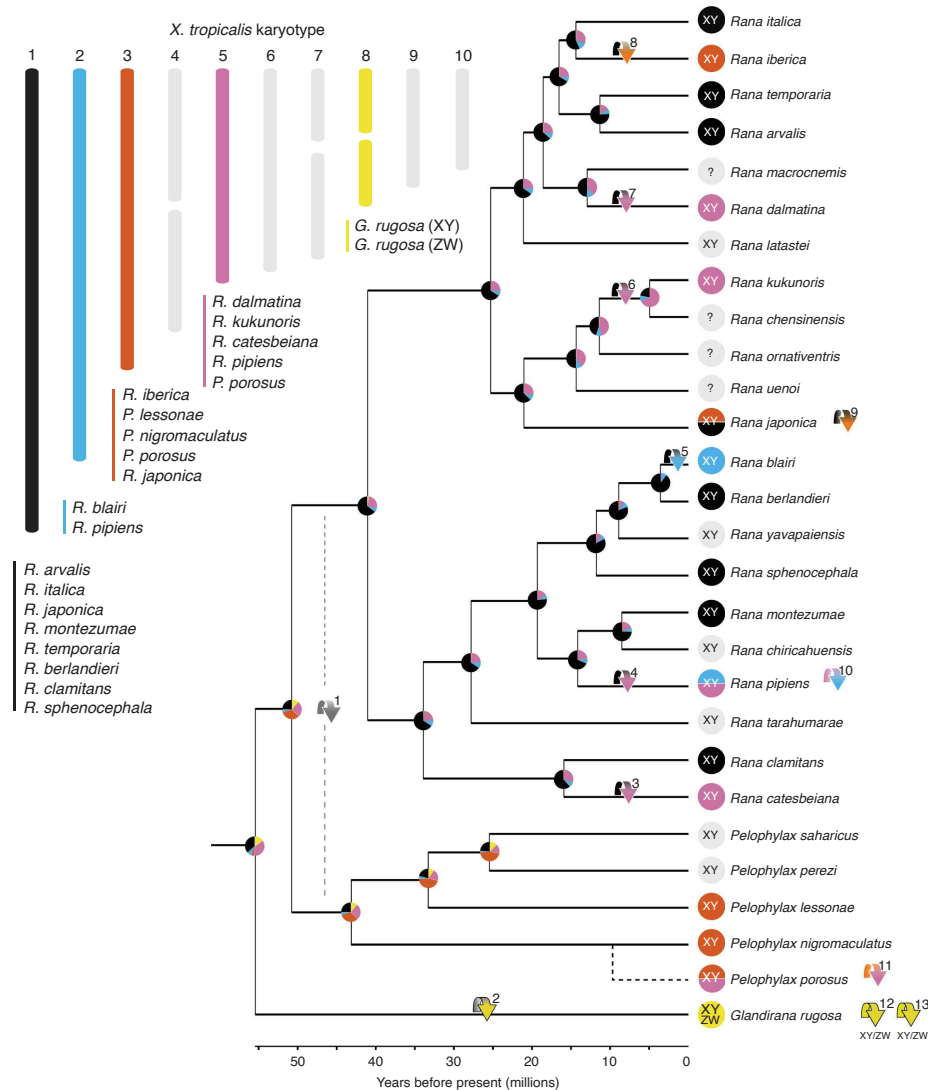


Figure 1.4. Turnovers de chromosomes sexuels parmi 28 espèces de grenouilles.

Le caryotype (en haut à gauche, chromosomes non à l'échelle) indique le nombre d'espèces utilisant chaque chromosome pour le déterminisme du sexe et les couleurs correspondent aux flèches, aux diagrammes circulaires des nœuds et aux extrémités. Les flèches de couleur montrent la branche sur laquelle les changements présumés se produisent sur la base des analyses de cartographie stochastique (figure supplémentaire 8) et les diagrammes circulaires aux nœuds représentent la proportion d'arbres simulés dans la cartographie stochastique avec chacun des états à ce nœud. Les extrémités bicolores représentent des événements de renouvellement intraspécifiques, la transition étant décrite par la flèche colorée après le nom de l'espèce. Le gris représente les identités inconnues des chromosomes sexuels dans les deux pointes et les flèches de rotation 1 et 2, et les points d'interrogation aux extrémités représentent un système d'hétérogamie inconnu. Issu et traduit de Jeffries et al., (2018), avec une simplification méthodologique de la légende.

Ce scénario est également décrit chez les boas et les pythons qui possèdent un déterminisme XX/XY mais qui ne font pas intervenir les mêmes chromosomes (Gamble et al., 2017). Le passage d'un déterminisme ancestral certainement du type ZZ/ZW vers XX/XY résulte d'une convergence évolutive entre ces deux lignées où de nouveaux déterminants ont été fixés.

Les turnovers ne sont pas toujours instantanés et il peut y avoir une coexistence des deux systèmes en parallèle notamment si le nouveau déterminant suit un patron différent de celui déjà existant. Comme vu précédemment, le déterminisme du sexe du platy (*Xiphophorus maculatus*) et du crapaud d'Afrique de l'ouest *Xenopus tropicalis* dépend de trois chromosomes sexuels respectivement W, Y, et X, avec $W > Y > X$ et Y, W et Z avec $Y > W > Z$ (Schultheis et al., 2009; Roco et al., 2015).

Ces turnovers de chromosomes sexuels peuvent avoir lieu si le déplacement ou l'apparition *de novo* d'un nouveau déterminant est fixé par dérive, ou encore s'ils augmentent la valeur sélective des individus.

Enfin, comme illustré précédemment avec le chromosome B chez le cichlidé *Metriaclimbra lombardii*, un turnover non homologue de chromosomes sexuels peut avoir lieu si un élément génétique égoïste vient endosser le rôle de déterminant du sexe (Clark et al., 2017).

b) Les conditions favorisant un turnover de chromosomes sexuels

Pour commencer peut-être la situation la plus évidente : un nouveau déterminant du sexe peut être sélectionné s'il améliore la valeur sélective des individus.

Comme décrit chez les mammifères, les chromosomes sexuels peuvent parfois être fortement hétéromorphes. Les causes pour ces différences de tailles sont souvent associées à une perte de recombinaison entre les deux chromosomes sexuels (Wright et al., 2016; Furman et al., 2020; Tao et al., 2021). On peut par exemple aisément imaginer qu'une diminution de la recombinaison entre les chromosomes sexuels peut permettre l'accumulation de locus bénéfiques aux mâles sur un chromosome Y, améliorant la valeur sélective du sexe mâle. Dans le cas où il existe une sélection sexuelle forte et des conflits génétiques, la pression de sélection peut-être encore plus forte pour le maintien des mutations sexuellement antagonistes sur des chromosomes différents (X vs Y par exemple), et donc la diminution, voire l'arrêt de la recombinaison (Rice, 1987). Cependant, un arrêt de la recombinaison

implique également un arrêt des balayages sélectifs dans ces régions chromosomiques, laissant la possibilité à des allèles faiblement délétères de s'accumuler sur le chromosome (Bachtrog, 2013). Ainsi, une translocation du déterminant du sexe sur une autre paire de chromosome, ou bien l'apparition d'un nouveau déterminant sur une paire de chromosome différente peut être sélectionnée si le Y devient particulièrement délétère chez les mâles (revue dans Tao et al., 2021). Ce processus peut être imaginé exactement à l'inverse pour un déterminisme de type ZZ/ZW. Cependant, comme nous l'avons déjà explicité précédemment, tous les chromosomes sexuels ne sont pas hétéromorphes et tous ne s'arrêtent pas systématiquement de recombinaison. La dégénération du Y n'est donc pas le seul moteur de turnovers de chromosomes sexuels.

Si une mutation confère à un gène existant une nouvelle fonction, (Kamiya et al., 2012; Myosho et al., 2012) ou entraîne un changement de la régulation de la cascade du sexe (Herpin et al., 2010), ces locus peuvent prendre le rôle de déterminant majeur.

Les duplications de gènes sont également un fort promoteur de l'apparition de nouveaux déterminants du sexe car l'une des copies paralogues peut acquérir une nouvelle fonction et prendre la tête de la cascade du déterminisme du sexe, et cela d'autant plus si la duplication a eu lieu sur un gène faisant déjà partie de cette cascade (Mank & Avise, 2009; Ortega-Recalde et al., 2020).

La dérive peut aussi faire disparaître un déterminant ou au contraire aider la fixation d'un déterminant qui ségrégerait par exemple en faible fréquence dans une population, particulièrement si celle-ci subit un goulot d'étranglement, comme cela peut être le cas lors de domestications (Wilson et al., 2014; Taslima et al., 2021).

Enfin, les éléments transposables ou les chromosomes B peuvent accélérer les turnovers de chromosomes sexuels non pas parce qu'ils sont sélectionnés pour rétablir des sex-ratios équilibrés, mais parce qu'ils agissent sous le modèle d'éléments égoïstes (Herpin et al., 2010; Clark and Kocher, 2019). En contrepartie, un élément égoïste qui entrainerait de la dérive méiotique pourrait promouvoir la sélection d'un déterminant génétique qui permette de compenser les biais de l'élément génétique égoïste (Cocquet et al., 2012).

Les possibilités pour expliquer les turnovers de chromosomes sexuels sont donc nombreuses et donnent une idée de la plasticité des systèmes de déterminisme du sexe (Furman et al., 2020; Tao et al., 2021).

II.2. Quand l'environnement rentre en jeu

Les mécanismes de déterminismes environnementaux du sexe sont retrouvés dans différents groupes taxonomiques (crocodiles, tortues, poissons...) mais les processus précis par lesquels ils agissent sont encore mal compris (Shen & Wang, 2018). Les changements environnementaux se traduisent par des modifications moléculaires qui peuvent impacter des voies métaboliques. Ces variations peuvent alors altérer l'expression des gènes de la différenciation du sexe. Les gènes intervenant dans les systèmes ESD sont donc certainement ceux qui s'expriment aux stades embryonnaires précoces où les gonades sont indifférenciées ou lorsqu'elles sont encore labiles. L'épigénétique, c'est à dire l'étude des mécanismes réversibles et héréditaires modifiant l'expression et/ou la fonction des gènes sans modifier la séquence ADN, est vue aujourd'hui comme l'une des clés pour comprendre les régulations d'expressions induites par les changements environnementaux. De récentes avancées en épigénétiques chez les reptiles et les poissons notamment permettent de mieux comprendre l'interaction entre l'environnement et la différenciation du sexe (Navarro-Martín et al., 2011; Shao et al., 2014; Deveson et al., 2017; Ge et al., 2018; Georges & Holleley, 2018; Weber et al., 2020).

Différents types de déterminismes environnementaux ont été observés dans la nature et parmi eux, le mieux décrit est certainement le déterminisme du sexe par la température (TSD). Ce type de déterminisme est retrouvé chez des espèces qui peuvent être sensibles aux conditions environnementales, et dans le cas de la température ce sont typiquement des espèces ectothermes. La température a un impact sur la différenciation du sexe pendant une période de temps définie (période thermosensible), durant laquelle les gonades sont encore labiles et indifférenciées. Les mécanismes de TSD sont répartis en trois classes chez les reptiles (Figure 1.5) (Valenzuela & Lance 2004). On qualifie de *TSD Ia*, (ou patron mâle-femelle) les espèces pour lesquelles à basse température les individus se développent en mâles, et à hautes températures se développent en femelles. Il existe donc un seul seuil de température pivot. C'est le mécanisme de déterminisme retrouvé notamment chez la tortue à nez de cochon (*Carettochelys insculpta*) où l'incubation des œufs en deçà de 32°C donne lieu à la différenciation en mâles, et au-delà de 32°C donne lieu au développement des voies femelles (Georges, 1992; Young et al., 2004). Le *TSD Ib* (aussi appelé patron femelle-mâle) est similaire au *TSD Ia* en ce sens qu'il n'y a qu'une seule température pivot, mais cette fois les températures basses induisent la différenciation des femelles, alors que des températures

élevées donnent lieu à la différenciation des mâles. Le Sphénodon ponctué (*Sphenodon punctatus*) présente ce type de déterminisme avec le développement majoritairement de mâles au-dessus de la température pivot de 22°C (Mitchell et al., 2006). Enfin, le troisième type est appelé *TSD II* ou encore patron femelle-mâle-femelle. Dans ce dernier cas, il existe deux températures pivots. Les individus soumis à des températures extrêmes (hautes et basses) se développent en femelles alors que les températures intermédiaires donnent lieu au développement de mâles. Cette situation est retrouvée par exemple chez l'alligator d'Amérique (*Alligator mississippiensis*) où 100% des œufs incubés entre 32,5 et 33,0°C se développent en mâles, alors que les femelles se développent au-dessous et au-dessus de ces températures pivots, avec 100% de femelles entre 29 et 31,5°C et au-delà de 34,5°C (Lang & Andrews, 1994). Le gecko léopard (*Eublepharis macularius*) possède une fourchette de température plus large qui conduit au développement des mâles (Viets et al., 1993). En effet, le développement des mâles prédomine aux températures intermédiaires de 31 à 33°C. Puis, entre 26 et 28°C et au-delà de 34°C, la différenciation du sexe aboutit à la formation de femelles (90 à 100% de femelles).

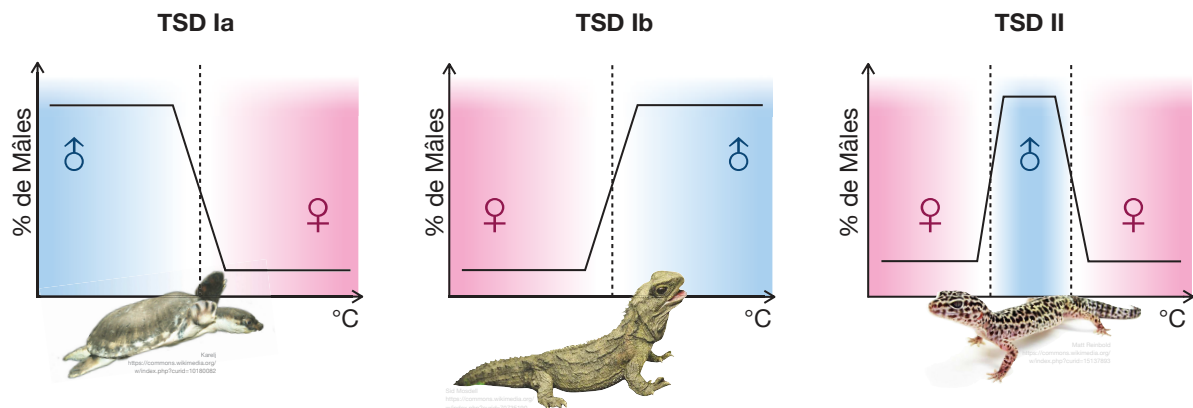


Figure 1.5. Illustration schématique des trois patrons de TSD et d'espèces présentant ce type de déterminisme.

Les traits pointillés représentent les « températures pivots » à partir desquelles les tendances du pourcentage de mâles dans les descendance s'inversent. La tortue nez de cochon (*Carettochelys insculpta*) présente un déterminisme TSD Ia, où les mâles sont produits après incubation des œufs à température basse, alors que les femelles sont produites à température élevée. Le déterminisme du Sphénodon ponctué (*Sphenodon punctatus*) est de type TSD Ib, ce qui induit la formation de femelles à basses températures et de mâles à hautes températures. Enfin, le Gecko léopard (*Eublepharis macularius*) est une espèce arborant un déterminisme de type TSD II : les températures intermédiaires donnent lieu au développement de mâles, et de part et d'autre des températures pivots, des femelles se développent.

II.3. La mince frontière entre GSD et facteurs environnementaux

Facteurs génétiques et environnementaux ne sont cependant pas toujours exclusifs et la frontière entre GSD et ESD est parfois difficile à cerner, et ce particulièrement chez les poissons. Ainsi, certaines espèces présentent un déterminisme du sexe génétique, avec des chromosomes sexuels différenciés ou non, qui peut être altéré ou outrepassé par des effets environnementaux. C'est le cas par exemple des fortes températures chez de nombreuses espèces de poissons, comme entre-autres les Cichlidés, certains Carassins ou encore le Pejerrey (Ospina-Álvarez & Piferrer, 2008; Baroiller et al., 2009). En dehors de la température, d'autres facteurs environnementaux peuvent influencer la différenciation du sexe chez les poissons. La photopériode peut avoir un effet sur les sexe-ratios de certaines espèces (Brown et al., 2014) et une exposition continue à la lumière induit par exemple une masculinisation chez les femelles de *Chirostoma estor* (Corona-Herrera et al., 2018). La couleur de l'aquarium joue sur les sex-ratios chez le Cardeau de Floride (*Paralichthys lethostigma*) en augmentant la proportion de mâles lorsque les individus sont élevés dans des aquariums bleus, en comparaison d'aquariums noirs et gris (Mankiewicz et al., 2013). Les pH acides (5.5) entraînent également des sex-ratios biaisés vers les mâles chez le cichlidé africain *Pelvicachromis pulcher* (Reddon & Hurd, 2013). Enfin, une forte densité d'individus peut entraîner une augmentation du développement des mâles chez le poisson-zèbre (*Danio rerio*), chez différentes espèces d'anguilles (*Anguilla*), ainsi que chez le Pejerrey (*Odontesthes bonariensis*) - également thermosensibles - malgré la présence de locus sexuels (Geffroy and Bardonnnet, 2016; Ribas et al., 2017; García-Cruz et al., 2020).

Même si les mécanismes précis par lesquels ces changements environnementaux impactent la différenciation du sexe restent flous, ces variations peuvent générer du stress. De fait, pour beaucoup d'espèces sensibles à l'environnement, des changements de conditions sont associées à une augmentation du cortisol (l'hormone du stress chez les vertébrés) (Fernandino et al., 2012; Geffroy & Bardonnnet, 2016; Ribas et al., 2017; Goikoetxea et al., 2017). De fortes températures ou de grandes densités d'individus chez le Pejerrey sont associés à une augmentation des niveaux de cortisol qui favorise la production de testostérone, 11-Ketotestosterone (11-KT) qui sont des signes moléculaires de la masculinisation (Hattori et al., 2009 ; García-Cruz et al., 2020). Egalement d'une manière dose-dépendante, le cortisol induit la masculinisation des femelles chez le *Paralichthys lethostigma* chez qui des températures intermédiaires donnent des sex-ratios équilibrés, alors

qu'un peu à l'image d'un TSD II inversée, les températures élevées et basses donnent des proportions plus importantes de mâles (Mankiewicz et al., 2013). Il en est de même chez le *Paralichthys olivaceus* qui possèdent un déterminisme de type XX/XY mais dont l'exposition des juvéniles à de fortes températures peut induire la masculinisation des individus XX en lien avec une l'augmentation du niveau de cortisol dans les gonades (Yamaguchi et al., 2010). Chez certaines espèces, le cortisol semble donc être un élément clé pour comprendre le lien entre masculinisation et effets environnementaux (et particulièrement des fortes températures) au moins chez les téléostéens. Il est cependant important de noter que chez les reptiles à ESD, la voie métabolique du stress ne semble pas être systématiquement impactée par les fortes températures (Castelli et al., 2021). D'autres acteurs tels que la régulation cellulaire du calcium et de l'oxydo-réduction ou l'action de facteurs d'épissage pourraient être impliqués, et traduire un stress cellulaire (et non pas hormonal comme chez les poissons) (Castelli et al., 2020, 2021; Weber et al., 2020).

Dans ces derniers cas le déterminisme n'est pas de type TSD mais plutôt GSD influencé par des Effets Environnementaux (EE). La limite entre ces deux mécanismes est parfois floue, et une confusion existe dans la littérature sur ces dénominations. Certains auteurs ont donc suggérés des protocoles afin de déterminer si l'on est en présence d'un mécanisme GSD, ESD ou GSD + EE (Valenzuela et al., 2003; Ospina-Álvarez & Piferrer, 2008; Shen & Wang, 2018). Ainsi, sur un spectre des mécanismes de déterminismes du sexe, il y aurait à chacun des extrêmes GSD et ESD, et une multitude de possibilités intermédiaires de systèmes GSD sur lesquels les effets environnementaux ont plus ou moins d'impact (Figure 1.6).

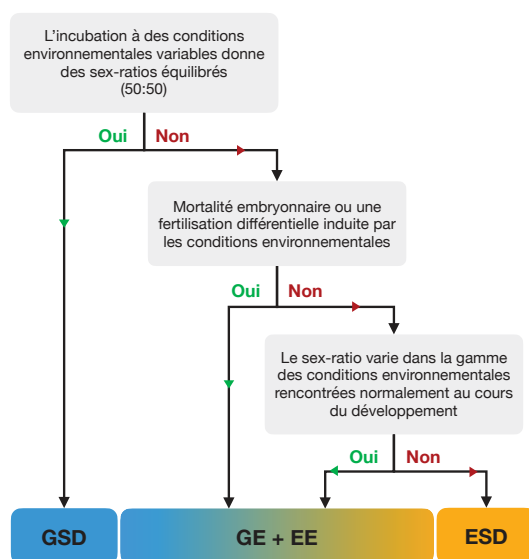


Figure 1.6. Méthodologie pour déterminer le type de déterminisme d'une espèce (ou une population) présentant ou non des chromosomes sexuels hétéromorphiques.

L'objectif principal est de différencier un système ESD strict, d'un système faisant interagir des effets génétiques et environnementaux pour déterminer le sexe. GSD : Déterminisme Génétique du sexe ; ESD : Déterminisme environnemental du sexe ; GE : Effets Génétiques ; EE : Effets environnementaux. Cette figure est une synthèse des figures proposées dans Valenzuela et al., (2003); Ospina-Álvarez & Piferrer, (2008); Shen & Wang, (2018) dans l'objectif de généraliser le plus possible les protocoles proposés.

Malgré des définitions rigides, les déterminismes en milieu naturel ne le sont pas toujours et les transitions d'un système à l'autre sont possibles même entre espèces proches comme chez les Geckos (Figure 1.7) (Gamble et al., 2015). En effet, les turnovers entre GSD et ESD ne sont pas rares au sein des vertébrés (Pennel et al., 2018). Les différentes causes favorisant ces transitions sont décrites plus en détails dans les sections suivantes.

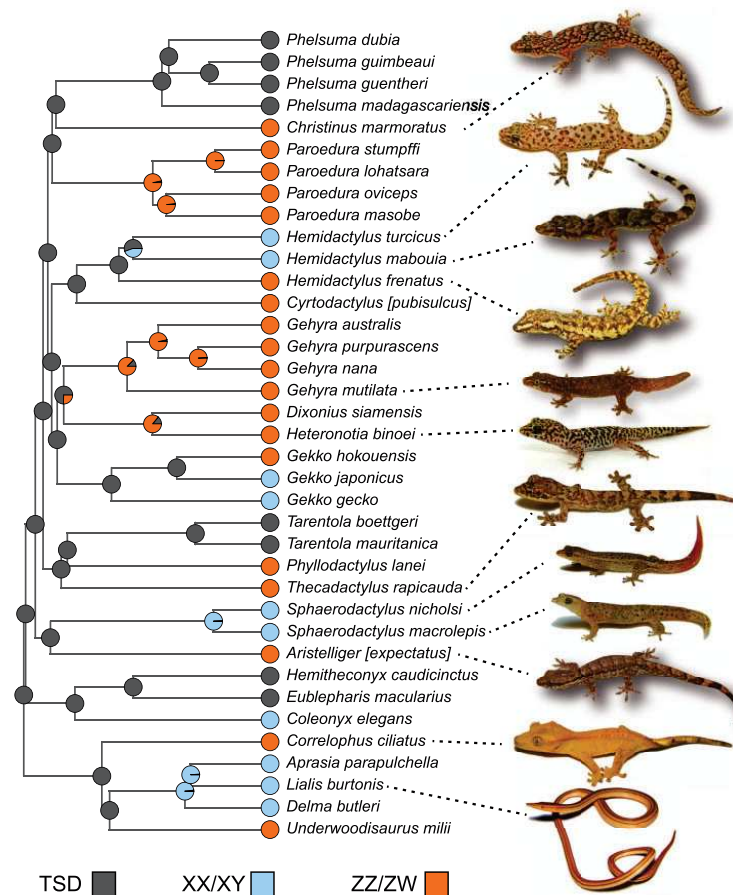


Figure 1.7. Évolution des systèmes de déterminisme du sexe chez les geckos.

Les cercles colorés aux extrémités de l'arbre phylogénétique indiquent les systèmes de déterminisme du sexe des espèces échantillonnées. Les cercles aux nœuds internes indiquent les reconstructions issues de 1 000 simulations de cartographie stochastique utilisant le modèle TRAP. Les noms entre crochets [] indiquent les espèces dont le mode de déterminisme du sexe est connu et qui étaient représentées sur la phylogénie originale par un taxon différent, étroitement apparenté. Issu et traduit de Gamble et al., (2015).

III. Des transitions entre systèmes qui ne sont pas rares

L'ensemble des exemples cités ci-dessous montre que contrairement à l'image dessinée par l'étude des mammifères, il existe une grande flexibilité du déterminisme du sexe et de

l'évolution des chromosomes sexuels chez les vertébrés. Certaines conditions peuvent favoriser ces transitions entre GSD et ESD et ces turnovers entre chromosomes sexuels.

Dans certains cas, un déterminisme sexuel peut être contre-sélectionné si celui-ci conduit à la production de descendance dont les sex-ratios ne sont plus favorables. Par exemple, en présence de conditions environnementales imprévisibles, un mécanisme ESD pourrait entraîner des sex-ratios extrêmement biaisés. De même, si l'environnement n'est pas (ou trop peu) variable, les biais de sex-ratios induits par un déterminisme ESD pourraient être important si les conditions ne sont pas régulièrement favorables au développement des deux sexes. Dans ce genre de situations, l'apparition d'un déterminisme de type GSD pourrait être sélectionné (Fisher, 1930). Cependant, dans le cas où un sex-ratio biaisé peut améliorer la valeur sélective d'un des sexes (*e.g.* s'il existe une compétition pour l'accès à un partenaire, ou si la température est favorable au développement d'un des deux sexes) alors l'apparition ou le maintien d'un système ESD peut être favorisé (Bull, 1983; Kocher, 2004; Valenzuela & Lance 2004). Une illustration de ces phénomènes existe chez les tortues. En effet, les espèces avec un GSD ont des espérances de vie plus courtes (et donc moins d'opportunités dans le temps pour produire un nombre équilibré d'individus de chaque sexe) que celles qui ont un patron ESD (Sabath et al., 2016). Sur le long terme, des biais de sexe-ratios peuvent être compensés chez les espèces à grande longévité, mais, sur des cycles de vies courts, un système stable de GSD est certainement plus fortement sélectionné.

La domestication est également un moteur pour le changement d'un type de déterminisme à un autre. Lorsqu'une espèce est prélevée du milieu naturel pour être maintenue en élevage, le nombre d'individus fondateurs des lignées domestiquées est souvent faible et un fort goulot d'étranglement s'exerce donc sur la population. La taille efficace plus faible des stocks commerciaux ou de laboratoires rend les espèces domestiquées plus sensibles à la dérive génétique et celle-ci peut par exemple entraîner la perte d'un déterminant du sexe trouvé dans le milieu naturel, particulièrement suite à des croisements successifs. Les populations sauvages de poisson-zèbres (*Danio rerio*) possèdent un déterminisme du sexe de type ZZ/ZW avec un déterminant majeur sur le chromosome 4 que ne présentent pas les souches de laboratoires, et notamment celle utilisée pour construire le génome de référence de cette espèce modèle (Wilson et al., 2014). Pour renforcer encore ce

propos, le déterminisme sexuel de la souche domestiquée est complexe, polyfactoriel et influencé par des effets environnementaux (Wilson et al., 2014; Ribas et al., 2017).

La thermosensibilité et son implication dans le déterminisme du sexe est certainement l'une des clés pour comprendre les transitions entre différents modes de déterminismes qui ont eu lieu à plusieurs reprises chez les amphibiens, les reptiles et les poissons (Sarre et al., 2011). Enfin, il apparaît que les transitions de déterminismes ESD vers GSD sont plus courantes que la réciproque chez les poissons (et chez les squamates également) ce qui suggère qu'un déterminisme GSD est plus stable (Gamble et al., 2015; Pokorna & Kratochvíl, 2009; Sabath, Itescu, et al., 2016; Pennel et al., 2018). Dans le contexte de changement climatique actuel, les espèces à ESD (et particulièrement TSD) vont devoir rapidement s'adapter pour limiter les biais de sex-ratios (qui sont déjà recensés), impliquant que les espèces à ESD seront particulièrement plus enclines aux transitions vers un GSD sur le long terme (Mitchell & Janzen, 2010; Jensen et al., 2018; Pennel et al., 2018). En revanche, les transitions d'un système GSD vers ESD ne sont pas impossibles et peuvent se faire rapidement en réponse à des conditions environnementales extrêmes particulièrement dans une petite population et s'il existe une thermosensibilité, comme chez l'Agame barbu (*Pogona vitticeps*) (Holleley et al., 2015).

IV. Les téléostéens, champions de la diversité

Les poissons téléostéens correspondent au groupe de vertébrés avec la plus grande richesse spécifique (plus de 30,000 espèces) et à eux seuls représentent la moitié des espèces de vertébrés (Helfman et al., 2009; Fricke et al., 2021). Ils sont présents dans tous les types d'eaux du globe, adaptés à des conditions parfois extrêmes de températures et de salinités notamment et arborent une grande diversité morphologique (Wootton et al., 2012). En plus de cette richesse spécifique et de cette diversité écologique, les poissons sont également des exemples de diversité et de plasticité du déterminisme et de la différenciation du sexe (Wang et al., 2018). Pendant longtemps, le gène *Sry*, déterminant du sexe chez les mammifères thériens, était le seul déterminant identifié chez les vertébrés. En dehors des poissons, seulement deux autres déterminants du sexe ont été identifiés chez l'ensemble des vertébrés : le gène *Dmrt1* (*Doublesex and mab-3 related transcription factor 1*) qui se trouve sur le chromosome Z chez les oiseaux (Smith et al., 2009); et le gène *Dm-w* chez le Xénope lisse (ou

Xénope du Cap, *Xenopus laevis*) qui est un paralogue du gène *dmrt1* et se situe sur le chromosome W (Yoshimoto et al., 2008). *A contrario*, chez les poissons, 8 déterminants du sexes ont été identifiés : le gène *dmy* (aussi appelé *dmrt1bY*) chez le médaka (*Oryzias latipes*) (Matsuda et al., 2002; Nanda et al., 2002) ; *dmrt1* chez le poisson plat *Cynoglossus semilaevis* (Cui et al., 2017) ; l'*amhY* chez le Pejerrey (*Odontesthes hatchery*) (Hattori et al., 2012) et également chez le brochet (*Esox lucius*) (avec la dénomination *amhbY*) (Pan et al., 2019) ; l'*amhr2* chez le fugu (*Takifugu rubripes*) (Kamiya et al., 2012) ; *gsdfY* chez *Oryzias luzonensis*, un proche parent du médaka (Myosho et al., 2012) ; *sdY* chez la truite arc-en-ciel (*Oncorhynchus mykiss*) (Yano et al., 2012) ainsi que chez d'autres salmonidés (Guiguen et al., 2018); *gdf6Y* chez *Nothobranchius furzeri* (Reichwald et al., 2015) ; et enfin *sox3Y* qui est un excellent candidat pour le déterminisme du sexe chez une troisième espèce d'*Oryzias* (*Oryzias dancena*) (Takehana et al., 2014) (Table 1). On peut noter que parmi les déterminants déjà identifiés, les membres de la superfamille des facteurs de croissance transformant, ou TGF- β (de l'anglais transforming growth factor), sont particulièrement représentés puisque les gènes de l'*amh*, *gsdf*, *gdf6* et de l'*amhr2* en font partie. Les TGF- β pour beaucoup jouent un rôle de régulation dans le développement embryonnaire et notamment dans le contrôle de la prolifération des cellules germinales et des spermatogonies chez les poissons osseux comme chez le poisson-zèbre ou la truite arc-en-ciel (Sawatari et al., 2007; Wang & Orban, 2007).

Sur l'ensemble de ces huit déterminants sexuels chez les poissons, cinq sont issus d'évènement de duplications où l'un des paralogues a pris le rôle de la tête leader de la cascade du déterminisme du sexe, et témoignent de la plasticité des génomes chez ce groupe taxonomique (Table 1.1). En effet, les téléostéens ont globalement des taux de mutations et de duplications supérieurs à ceux des autres vertébrés (Mank & Avise, 2009). Or, ces duplications, lorsqu'elles interviennent sur gènes déjà impliqués dans la cascade de déterminisme ou de la différenciation du sexe sont un phénomène efficace pour créer de nouveaux déterminants car ceux-ci sont déjà intégrés à la cascade.

Table 1.1. Gènes du déterminisme du sexe formellement identifiés ou fortement suspectés chez les poissons téléostéens.

Les déterminants qui ne sont pas issus d'évènement de duplication sont devenus déterminants majeurs suite à une diversification allélique. D'après Guiguen et al., (2018) et Li & Gui, (2018).

Déterminant du sexe	Taxon	Système	Gène ancestral	Issue d'une Duplication	Famille génique	Références
<i>dmy</i> (<i>dmrt1bY</i>)	<i>Oryzias latipes</i>	XY	<i>dmrt1</i>	✓	DM-domain	Matsuda et al., 2002
<i>dmrt1</i>	<i>Cynoglossus semilaevis</i>	ZW	<i>dmrt1</i>	✗	DM-domain	Cui et al., 2017
<i>amhY</i>	<i>Odontesthes hatcheri</i>	XY	<i>amh</i>	✓	TGF- β	Hattori et al., 2012
<i>amhbY</i>	<i>Esox lucius</i>	XY	<i>amh</i>	✓	TGF- β	Pan et al., 2019
<i>amhr2Y</i>	<i>Takifugu rubripes</i>	XY	<i>amhr2</i>	✗	TGF- β	Kamiya et al., 2012
<i>gsdfY</i>	<i>Oryzias luzonensis</i>	XY	<i>gsdf</i>	✗	TGF- β	Myosho et al., 2012
<i>gdf6Y</i>	<i>Nothobranchius furzeri</i>	XY	<i>gdf6</i>	✗	TGF- β	Reichwald et al., 2015
<i>sdY</i>	<i>Oncorhynchus mykiss</i>	XY	<i>irf9</i>	✓	Irf	Yano et al., 2012
<i>sox3Y</i>	<i>Oryzias dancena</i>	XY	<i>sox3</i>	✗	Sox	Takehana et al., 2014

Les Cichlidés représentent peut-être la famille dont les radiations évolutives sont les plus impressionnantes, et particulièrement dans les grands lacs de l'Est Africain, et l'on recense plus de 3000 espèces de cichlidés entre l'Amérique du Sud, l'Afrique et l'Inde (Kocher, 2004; Matschiner et al., 2020).

La vaste majorité des cichlidés est gonochorique et il existe une grande variabilité des locus déterminants du sexe et des chromosomes/groupes de liaisons qui les portent. Ainsi, on connaît plus de 20 régions différentes associées au sexe sur 17 chromosomes/groupes de liaisons, incluant les chromosomes B, chez les cichlidés (Figure 1.8) (Gammerdinger & Kocher, 2018; Böhne et al., 2019; El Taher et al., 2020).

Les chromosomes sexuels identifiés jusqu'à présent sont systématiquement homomorphiques (revue dans Gammerdinger & Kocher, 2018) mais présentent différents degrés de différenciation génétique avec certains pour lesquels la région est étroite, et d'autres pour lesquels elle est étendue sur la presque totalité d'un groupe de liaison (El Taher et al., 2020). Enfin, certains chromosomes ont été plusieurs fois recrutés chez les cichlidés pour endosser le rôle de chromosomes sexuels, comme les LGs 5, 7 et 19. Enfin, on retrouve chez les cichlidés des patrons de GSD XX/XY et ZZ/ZW de déterminismes du sexe, mais aussi des systèmes plus complexes comme XX/XY + ZZ/ZW ou encore XX/XY + Chromosome B (spécifique aux femelles et agissant comme un W) (Clark & Kocher, 2018; Gammerdinger & Kocher 2018; El Taher et al., 2020) et aussi une importance des facteurs environnementaux et notamment de la température (Baroiller et al., 2009; Baroiller & D'Cotta, 2016).

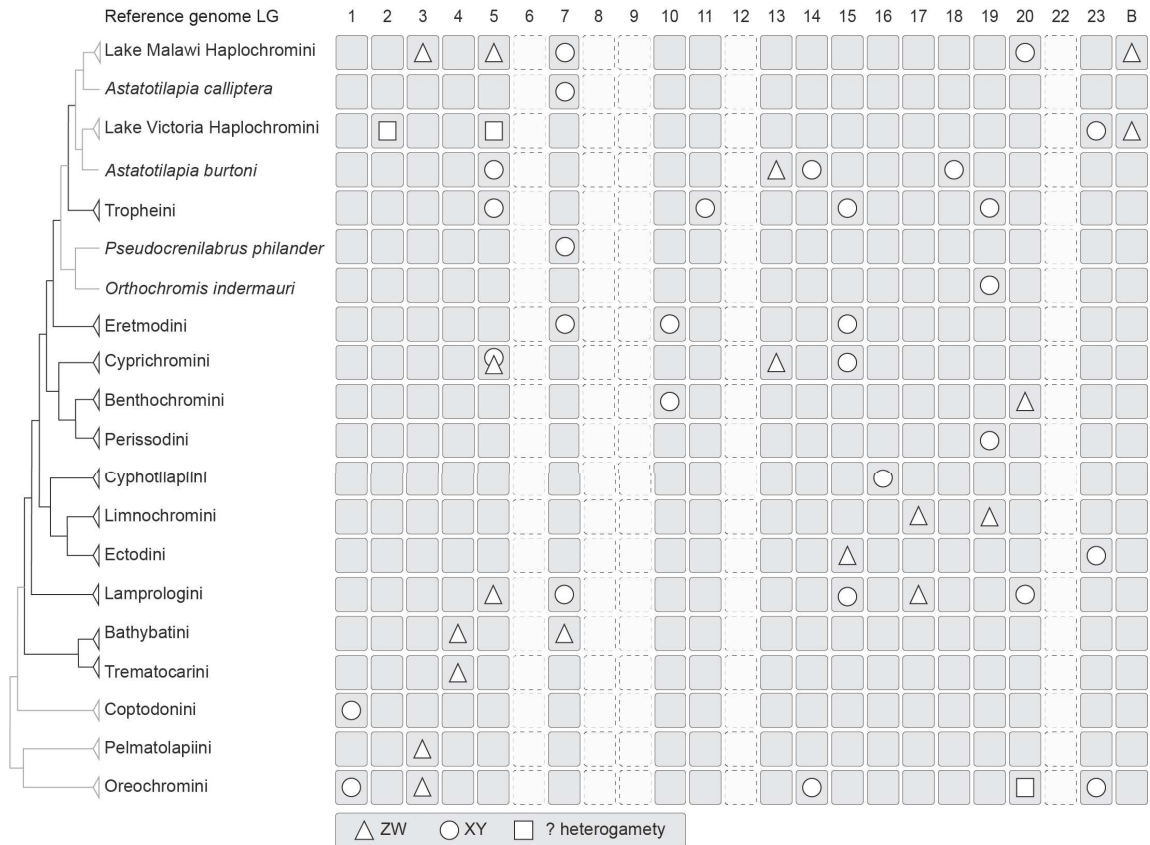


Figure 1.8. Évolution des chromosomes sexuels chez les cichlidés africains.

Relations phylogénétiques et apparition des chromosomes sexuels par rapport au génome de référence du tilapia du Nil (*O. niloticus*) chez les cichlidés africains. Les lignées de cichlidés présentes dans le lac Tanganyika sont indiquées en noir, les cichlidés provenant d'autres lacs ou rivières en gris. Issue de El Taher et al., 2020 (version de bioRxiv).

V. Le Tilapia du Nil, un modèle de choix

V.1. Diversité du déterminisme du sexe en populations domestiques

Le tilapia du Nil (*Oreochromis niloticus*) est un cichlidé africain d'importance économique majeure pour l'aquaculture puisqu'il représente l'une des espèces de poissons les plus produites et consommées au monde avec 5,5 millions de tonnes produites en 2019 (FAO, 2021). Les mâles de cette espèce ont une croissance plus rapide, et disposent d'une masse musculaire plus développée que les femelles. Malgré un fort intérêt économique pour la production de mâles, le déterminisme du sexe de cette espèce est complexe et reste mal compris. Depuis les années 1970, il a été établi que le déterminisme du sexe chez le tilapia du Nil suit un système GSD XX/XY (Jalabert et al., 1974) mais des facteurs génétiques mineurs, hérités des parents, peuvent impacter les sex-ratios des descendance (Baroiller & D'Cotta, 2001). De plus, ce déterminisme génétique peut également être supplanté par des effets

environnementaux puisqu'une température élevée pendant les phases précoces du développement peut affecter la différenciation du sexe et générer alors des mâles XX viables et fertiles (Baroiller et al., 1995; Baroiller et al., 2009) (Figure 1.9). Cette sensibilité à la température (thermosensibilité), définie comme la proportion de mâles XX obtenue, est un trait variable et héritable qui peut être sélectionné en seulement quelques générations pour fixer des souches très ou très peu thermosensibles (Baroiller and D'Cotta, 2001; Wessels et Hörstgen-Schwark 2007).

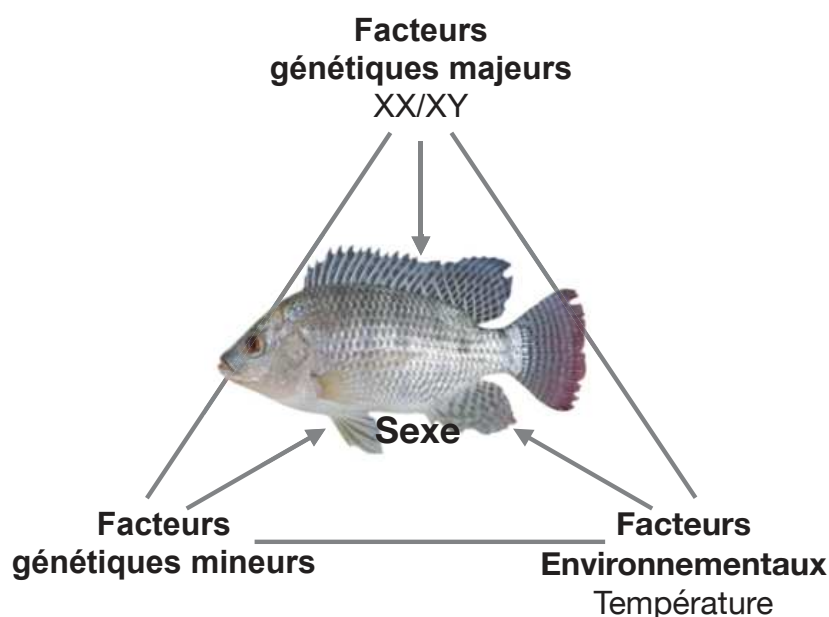


Figure 1.9. Représentation triangulaire schématisant la complexité du déterminisme du sexe chez le tilapia du Nil.

Modifié d'après Baroiller et al., (2009).

Jusqu'à présent, le déterminisme du sexe chez cette espèce a principalement été étudié sur des souches domestiques et commerciales. Si la présence d'un déterminisme XX/XY est acquise, la paire de chromosomes sexuels portant le déterminant majeur est encore débattue. Des premières études, basées sur de la cytogénétique et l'observation de complexes synaptonémaux chez *O. niloticus* ont montré un non-appariement des extrémités de la plus grande paire de chromosome (LG3) chez les mâles XY mais pas chez les mâles XX, suggérant que cette paire endosse le rôle de chromosomes sexuels (Carrasco et al., 1999). Des analyses génétiques ont indiqué une association des marqueurs sexuels sur une paire de chromosomes plus petite (LG1) (Lee et al., 2003; Cnaani et al., 2008; Lee et al., 2011). Finalement, en utilisant des technologies plus récentes et des analyses génomiques, la région

du déterminisme du sexe a été identifiée sur LG1 (Lee et al. 2003; Cnaani et al., 2008; Palaikostas et al. 2013; Gammerdinger et al. 2014; Palaikostas et al. 2015) ou sur LG23 (Eshel et al. 2012, 2014; Li et al. 2015; Wessels et al., 2017; Cáceres et al. 2019; Taslima et al. 2020) ou bien sur les deux en même temps (Taslima et al., 2021) selon les souches étudiées. Aucun gène précis n'a été proposé sur LG1. Sur LG23 en revanche, en étudiant des souches domestiques Japonaises, la région du sexe a été identifiée : les mâles sont porteurs sur le Y d'une duplication en tandem comprenant deux copies du gène de l'Hormone Anti-Müllérienne (*amh*) dont une est intacte (*amhY*) et l'autre tronquée (*amhΔY*) (Eshel et al., 2014; Li et al., 2015). Une étude sur une population naturelle a également montré que le sexe était associé au gène de l'*amh* sur LG23 chez des individus du lac Kou au Burkina-Faso (Sissao et al., 2019).

Table 1.2. Bilan des origines, dénominations et déterminants génétiques du sexe chez les souches domestiques et commerciales de tilapia du Nil.

Laboratoire/Commercial	Souche	Origine du stock	Dénomination	LG	Locus du sexe	Reference LG	Reference Origine
Université de Stirling (Royaume-Uni)	Manzala	Manzala (Egypte)	Manzala-Stirling	LG1 + LG23	? + <i>amh</i>	Taslima et al., 2021	Taslima et al., 2021
Université de Liège (Tihange, Belgique)	Manzala	Manzala-Stirling	Manzala-Thiange	LG23	<i>amh</i>	Sissao et al., 2019	Taslima et al., 2021
Université de Göttingen (Allemagne)	Manzala	Manzala-Stirling	Manzala-Göttingen	LG23	<i>amh</i>	Wessels et al., 2017	Kronert et al., 1989
Agricultural Research Organization (Israël)	Manzala	Manzala-Stirling-Swansea	Manzala-ARO	LG23	<i>amh</i>	Eshel et al., 2014	Taslima et al., 2021
Agricultural Research Organization (Israël)	Manzala (+ Ghana)	Manzala-Stirling-Swansea	Manzala-ARO	LG23	<i>amh</i>	Curzon et al., 2019	Curzon et al., 2019
Laboratory of Reproductive Biology (Okazaki, Japon)	Egypte	Egypte (possible introgression par Manzala-Stirling par l'utilisation de mâles YY)	Japenese	LG23	<i>amh</i>	Li et al., 2015	Li et al., 2015
Souches commerciales d'Amérique du Sud	Malaysie (plus ou moins lié au GIFT)	AquaAmerica (Brésil)		LG23	<i>amh</i>		
	Croisement d'individus d'Israël, de Singapour, de Taiwan et de Thaïlande présents aux Philippines (Carmen Aquafarm) dans la fin des années 1980	Acuacorporación Internacional (ACI, Costa Rica) (Obtenu du Brésil)		LG23	<i>amh</i>	Cáceres et al., 2019	Cáceres et al., 2019
	GIFT + 2 souches originaires d'Afriques ayant servie à fonder la souche GIFT	Acuacorporación Internacional (ACI, Costa Rica)		LG23	<i>amh</i>		
World Fish	GIFT (*)	Hybrid of several strains	GIFT	LG23	<i>amh</i>	Taslima et al., 2020	Eknath et al., 1998; Ponzoni et al., 2011

Laboratoire/Commercial	Souche	Origine du stock	Dénomination	LG	Locus du sex	Reference LG	Reference Origine
Guangxi Fisheries Institute (Nanning, Guangxi, Chine)	Ancien GIFT (**)	2 programmes successifs de sélection mis en place par la Chine sur l'ancien GIFT (ICLARM non associé)	Chinese GIFT	LG23			Zhao et al., 2019
USA Souche commerciale	Amherst, Massachusetts USA	?	USA Commercial Stock	LG1	?	Gammerdinger et al., 2014	Gammerdinger et al., 2014
Université N. Boni, (Burkina Faso)	Kou	Lac Kou	Kou-BF	LG23 + ?	<i>amh</i> + ?	Sissao et al., 2019	Sissao et al., 2019

(**) Hybride entre 4 populations de *O. niloticus* (Egypte, Ghana, Kenya et Sénégal), et 4 souches d'aquaculture (Israël, Singapour, Taïwan et Thaïlande).

(**) 3ème Génération du GIFT transmis en 1994 par l'ICLARM (International Center for Living Aquatic Resources Management (ancêtre du Worldfish).

Des marqueurs PCR permettant de génotyper le sexe des individus ont été développés pour les souches présentant un déterminisme du sexe par le gène de l'*amh* (Eshel et al., 2014; Li et al., 2015; Sissao et al., 2019). Ces marqueurs ne permettent cependant pas de génotyper efficacement toutes les populations sauvages (Baroiller & D'Cotta, 2016), suggérant un polymorphisme du chromosome Y dans les populations naturelles. En effet, ces marqueurs PCR sont inutilisables sur les populations éthiopiennes (aucune amplifications de l'*amhΔY*), et des incongruences existent entre les marqueurs de l'*amhY* et de l'*amhΔY* dans la population sauvage de Kou (Burkina-Faso), alors que ces deux copies sont issues d'une duplication en tandem et donc que la présence de l'un devrait systématiquement indiquer la présence de l'autre (Sissao et al., 2019).

V.2. Une espèce à large aire de répartition et diversité écologique

L'influence de l'environnement et particulièrement de la température sur le déterminisme du sexe des espèces GSD + EE n'a pas toujours été étudiée en tenant compte des conditions pouvant réellement être rencontrés dans le milieu naturel. En l'occurrence, le tilapia du Nil est une espèce à l'aire de répartition très large à travers l'Afrique, des bassins du Nil au Sénégal, aux plateaux des vallées du Grand Rift en Ethiopie, en passant par les provinces de Kivu et du Nord du Tanganyika (Trewavas 1983; Bezault et al., 2011). Du fait de ces nombreuses colonisations de différents cours d'eau et lac, cette espèce s'est adapté à une large gamme de températures dans le milieu naturel et on la retrouve dans des eaux « froides » (17-26°C), tempérées (28-34°C) et même des sources chaudes (36-40°C) (Bezault et al., 2007; Ndiwa et al. 2014). Or, une exposition à une température supérieure à 32°C pendant les phases précoces

du développement des gonades (10 à 20 jours après fécondation des œufs) peut induire une réversion du sexe et donner lieu à la différenciation en mâles d'individus pourtant XX (Baroiller et al., 1995; Baroiller et al., 2009). Cette thermosensibilité (plus ou moins forte selon les familles ou les souches) chez le tilapia du Nil a aussi été démontrée pour l'ensemble des populations étudiées issues du milieu naturel (Bezault et al. 2007; Baroiller et al., 2009).

En milieu naturel, après une période d'incubation buccale puis de garde, les femelles relâchent les alevins pendant leur période thermosensible et ceux-ci rejoignent les zones d'eau peu profondes où les températures sont aussi plus élevées (Philippart & Ruwet, 1982). Les alevins peuvent donc être exposés à des températures qui dépassent la température de masculinisation de 32°C. Une étude en conditions contrôlées sur les préférences en *continuum* thermique a d'ailleurs montré que 20% des alevins avaient une préférences pour les eaux plus chaudes (36°C) pendant les premiers jours puis dans l'ensemble montraient une préférence pour les bassins à 32°C malgré la présence d'un bassin à 28°C accessible (Nivelle et al., 2019).

Le tilapia du Nil est donc une espèce présentant un GSD+EE dont les effets environnementaux pouvant influencer la différenciation du sexe sont rencontrés en milieu naturel. Du fait de la diversité des mécanismes rencontrés dans les souches de laboratoire, ainsi que de son déterminisme du sexe complexe dans lequel facteurs génétiques et environnementaux interagissent, le tilapia du Nil représente un modèle idéal pour l'étude de l'évolution du déterminisme du sexe à l'échelle intraspécifique. De plus, l'existence de populations sauvages adaptées à une large gamme de température fait également du tilapia du Nil un bon modèle pour étudier l'impact du réchauffement des eaux sur la dynamique des sex-ratios des espèces de poissons thermosensibles dans un contexte de changement climatique. Toutefois, il n'est pas évident de savoir en quelle mesure les différents mécanismes de déterminisme du sexe observés chez le tilapia du Nil sont dues à une diversité propre à l'espèce ou si elles sont dues aux événements de domestications (avec des introgressions et hybridations, accidentelles ou non).

Ainsi, dans ce projet de thèse l'objectif était, dans un premier temps, de comparer les bases génétiques du déterminisme du sexe de populations d'*O. niloticus* du milieu naturel dont les écologies thermiques sont variées (Figure 1.10) pour comprendre si la diversité observée dans les stocks de laboratoires est le reflet d'une variabilité dans les populations sauvages ; dans un

second temps de pouvoir caractériser en détails la région du sexe et identifier des régions (si possibles universelles) qui permettent le développement de marqueurs pour génotyper sexuellement les individus sauvages ; enfin dans un troisième temps, il était souhaité d'explorer les gènes précocement impliqués dans la différenciation du sexe et l'effet de la température sur cette cascade.

Ce manuscrit se compose donc de trois chapitres au format article qui tentent de répondre à ces problématiques ou du moins de démêler l'histoire complexe du déterminisme du sexe chez le tilapia du Nil.

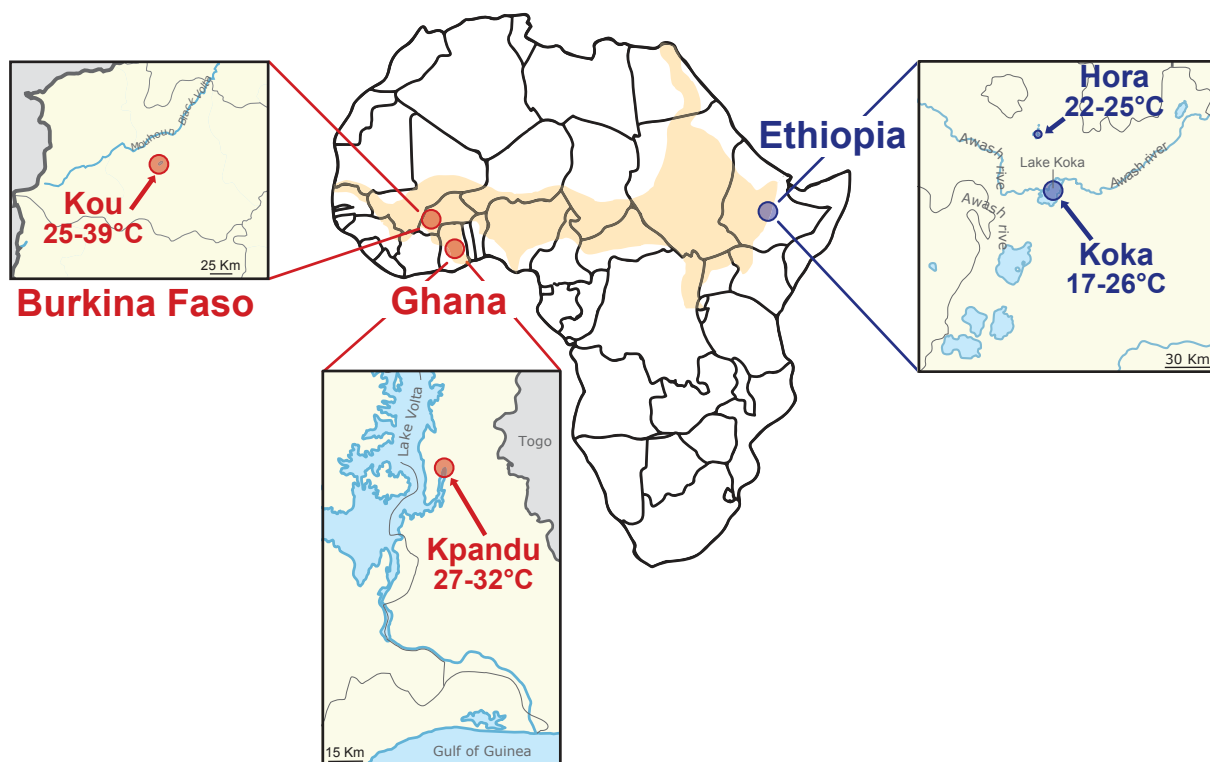


Figure 1.10. Localisation géographique des fourchettes de température des environnements d'où proviennent les populations sauvages étudiées dans ce projet de thèse.

Références

- Adolfi, M.C., Nakajima, R.T., Nóbrega, R.H., and Scharf, M. (2019). Intersex, Hermaphroditism, and Gonadal Plasticity in Vertebrates: Evolution of the Müllerian Duct and Amh/Amhr2 Signaling. *Annu. Rev. Anim. Biosci.* 7, 149–172.
- Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics* 14, 113–124.
- Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., et al. (2014). Sex determination: why so many ways of doing it? *PLOS Biology* 12, e1001899.
- Baroiller, J.F., and D’Cotta, H. (2001). Environment and sex determination in farmed fish. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 130, 399–409.
- Baroiller, J.-F., and D’Cotta, H. (2016). The Reversible Sex of Gonochoristic Fish: Insights and Consequences. *Sexual Development* 10, 242–266.
- Baroiller, J.F., Chourrout, D., Fostier, A., and Jalabert, B. (1995). Temperature and sex chromosomes govern sex ratios of the mouthbrooding Cichlid fish *Oreochromis niloticus*. *J. Exp. Zool.* 273, 216–223.
- Baroiller, J.F., D’Cotta, H., and Saillant, E. (2009a). Environmental effects on fish sex determination and differentiation. *SXD* 3, 118–135.
- Baroiller, J.F., D’Cotta, H., Bezault, E., Wessels, S., and Hoerstgen-Schwark, G. (2009b). Tilapia sex determination: where temperature and genetics meet. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 153, 30–38.
- Bezault, E., Clota, F., Derivaz, M., Chevassus, B., and Baroiller, J.-F. (2007). Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture* 272, S3–S16.
- Bezault, E., Balaesque, P., Toguyeni, A., Fermon, Y., Araki, H., Baroiller, J.-F., and Rognon, X. (2011). Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genetics* 12, 102.
- Böhne, A., Weber, A.A.-T., Rajkov, J., Rechsteiner, M., Riss, A., Egger, B., and Salzburger, W. (2019). Repeated Evolution Versus Common Ancestry: Sex Chromosome Evolution in the Haplochromine Cichlid *Pseudocrenilabrus philander*. *Genome Biology and Evolution* 11, 439–458.
- Brown, E.E., Baumann, H., and Conover, D.O. (2014). Temperature and photoperiod effects on sex determination in a fish. *Journal of Experimental Marine Biology and Ecology* 461, 39–43.
- Bull, J.J. (1983). Evolution of sex determining mechanisms. *Evolution of Sex Determining Mechanisms*.
- Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Travisany, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., et al. (2019). Fine mapping

- using whole-genome sequencing confirms anti-müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3 (Bethesda)* 9, 3213–3223.
- Capel, B. (2017). Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat. Rev. Genet.* 18, 675–689.
- Carrasco, L.A.P., Penman, D.J., and Bromage, N. (1999). Evidence for the presence of sex chromosomes in the Nile tilapia (*Oreochromis niloticus*) from synaptonemal complex analysis of XX, XY and YY genotypes. *Aquaculture* 173, 207–218.
- Castelli, M., Georges, A., and Holleley, C.E. (2021). Corticosterone does not have a role in temperature sex reversal in the central bearded dragon (*Pogona vitticeps*). *J Exp Zool A Ecol Integr Physiol* 335, 301–310.
- Castelli, M.A., Whiteley, S.L., Georges, A., and Holleley, C.E. (2020). Cellular calcium and redox regulation: the mediator of vertebrate environmental sex determination? *Biological Reviews* 95, 680–695.
- Charlesworth, B., and Dempsey, N.D. (2001). A model of the evolution of the unusual sex chromosome system of *Microtus oregoni*. *Heredity* 86, 387–394.
- Cioffi, M.B., and Bertollo, L. a. C. (2010). Initial steps in XY chromosome differentiation in *Hoplias malabaricus* and the origin of an X 1 X 2 Y sex chromosome system in this fish group. *Heredity* 105, 554–561.
- Clark, F.E., and Kocher, T.D. (2019). Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. *Scientific Reports* 9, 1–10.
- Clark, F.E., Conte, M.A., Ferreira-Bravo, I.A., Poletto, A.B., Martins, C., and Kocher, T.D. (2017). Dynamic sequence evolution of a sex-associated B chromosome in Lake Malawi cichlid fish. *JHERED* 108, 53–62.
- Clark, F.E., Conte, M.A., and Kocher, T.D. (2018). Genomic characterization of a B chromosome in Lake Malawi cichlid fishes. *Genes* 9, 610.
- Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., D'Hont, A., Baroiller, J.-F., D'Cotta, H., Penman, D.J., et al. (2008). Genetics of sex determination in tilapiine species. *Sexual Development* 2, 43–54.
- Cocquet, J., Ellis, P.J.I., Mahadevaiah, S.K., Affara, N.A., Vaiman, D., and Burgoyne, P.S. (2012). A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. *PLOS Genetics* 8, e1002900.
- Corona-Herrera, G.A., Arranz, S.E., Martínez-Palacios, C.A., Navarrete-Ramírez, P., Toledo-Cuevas, E.M., Valdez-Alarcón, J.J., and Martínez-Chávez, C.C. (2018). Experimental evidence of masculinization by continuous illumination in a temperature sex determination teleost (Atherinopsidae) model: is oxidative stress involved? *Journal of Fish Biology* 93, 229–237.
- Cui, Z., Liu, Y., Wang, W., Wang, Q., Zhang, N., Lin, F., Wang, N., Shao, C., Dong, Z., Li, Y., et al. (2017). Genome editing reveals *dmrt1* as an essential male sex-determining gene in Chinese tongue sole (*Cynoglossus semilaevis*). *Scientific Reports* 7, 42213.

- Curzon, A.Y., Seroussi, E., Benet-Perelberg, A., Zak, T., Dor, L., Shirak, A., and Ron, M. (2019). Hybrid Origin of the Thai-Chitralada Tilapia Strain Using DNA Barcoding and Microsatellite Analysis. *Israeli Journal of Aquaculture - Bamidgeh*.
- Deeming, D.C. (2004). Chapter 4 - Prevalence of TSD in crocodilians. In *Temperature-Dependent Sex Determination in Vertebrates*, (Smithsonian), pp. 33–41.
- Deveson, I.W., Holleley, C.E., Blackburn, J., Marshall Graves, J.A., Mattick, J.S., Waters, P.D., and Georges, A. (2017). Differential intron retention in *Jumonji* chromatin modifier genes is implicated in reptile temperature-dependent sex determination. *Sci. Adv.* 3, e1700731.
- Devlin, R.H., and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208, 191–364.
- Eknath, A., Dey, M., Rye, M., Gjerde, B., Abella, T.A., Sevilleja, R., Tayamen, M.M., Reyes, R.A., and Bentsen, H. (1998). Selective breeding of Nile tilapia for Asia.
- El Taher, A., Ronco, F., Matschiner, M., Salzburger, W., and Böhne, A. (2020). Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes (*Evolutionary Biology*).
- Ellegren, H. (2010). Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution* 25, 283–291.
- Eshel, O., Shirak, A., Weller, J.I., Hulata, G., and Ron, M. (2012). Linkage and physical mapping of sex Rregion on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3: Genes, Genomes, Genetics* 2, 35–42.
- Eshel, O., Shirak, A., Dor, L., Band, M., Zak, T., Markovich-Gordon, M., Chalifa-Caspi, V., Feldmesser, E., Weller, J.I., Seroussi, E., et al. (2014). Identification of male-specific amh duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 15, 774.
- Ewert, M.A., Etchberger, C.R., and Nelson, C.E. (2004). Chapter 3 - Turtle Sex-Determining Modes and TSD Patterns, and Some TSD Pattern Correlates. In *Temperature-Dependent Sex Determination in Vertebrates*, (Smithsonian), pp. 21–32.
- FAO (2021). *The State of World Fisheries and Aquaculture 2021* (FAO).
- Fernandino, J.I., Hattori, R.S., Kishii, A., Strüssmann, C.A., and Somoza, G.M. (2012). The Cortisol and Androgen Pathways Cross Talk in High Temperature-Induced Masculinization: The 11 β -Hydroxysteroid Dehydrogenase as a Key Enzyme. *Endocrinology* 153, 6003–6011.
- Fisher, R.A. (1930). *The genetical theory of natural selection*. (Oxford: Clarendon Press).
- Fricke, R., Eschmeyer, W.N., and Van der Laan, R. (2021). *Eschmeyer's Catalog of Fishes: Genera, Species, References*.
- Furman, B.L.S., Metzger, D.C.H., Darolti, I., Wright, A.E., Sandkam, B.A., Almeida, P., Shu, J.J., and Mank, J.E. (2020). Sex chromosome evolution: so many exceptions to the rules. *Genome Biol Evol* 12, 750–763.
- Gamble, T., Coryell, J., Ezaz, T., Lynch, J., Scantlebury, D.P., and Zarkower, D. (2015). Restriction Site-Associated DNA Sequencing (RAD-seq) Reveals an Extraordinary

- Number of Transitions among Gecko Sex-Determining Systems. *Molecular Biology and Evolution* 32, 1296–1309.
- Gamble, T., Castoe, T.A., Nielsen, S.V., Banks, J.L., Card, D.C., Schield, D.R., Schuett, G.W., and Booth, W. (2017). The Discovery of XY Sex Chromosomes in a Boa and Python. *Current Biology* 27, 2148–2153.e4.
- Gammerdinger, W.J., and Kocher, T.D. (2018). Unusual diversity of sex chromosomes in African cichlid fishes. *Genes* 9, 480.
- Gammerdinger, W.J., Conte, M.A., Acquah, E.A., Roberts, R.B., and Kocher, T.D. (2014). Structure and decay of a proto-Y region in *Tilapia*, *Oreochromis niloticus*. *BMC Genomics* 15, 975.
- García-Cruz, E.L., Yamamoto, Y., Hattori, R.S., de Vasconcelos, L.M., Yokota, M., and Strüssmann, C.A. (2020). Crowding stress during the period of sex determination causes masculinization in pejerrey *Odontesthes bonariensis*, a fish with temperature-dependent sex determination. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 110701.
- Ge, C., Ye, J., Weber, C., Sun, W., Zhang, H., Zhou, Y., Cai, C., Qian, G., and Capel, B. (2018). The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science*.
- Geffroy, B., and Bardonnet, A. (2016). Sex differentiation and sex determination in eels: consequences for management. *Fish and Fisheries* 17, 375–398.
- Georges, A. (1992). Thermal-Characteristics and Sex Determination in Field Nests of the Pig-Nosed Turtle, *Carettochelys-Insculpta* (Chelonia, Carettochelydidae), From Northern Australia. *Aust. J. Zool.* 40, 511–521.
- Georges, A., and Holleley, C.E. (2018). How does temperature determine sex? *Science* 360, 601–602.
- Goikoetxea, A., Todd, E.V., and Gemmell, N.J. (2017). Stress and sex: does cortisol mediate sex change in fish? *Reproduction* 154, R149–R160.
- Grützner, F., Rens, W., Tsend-Ayush, E., El-Mogharbel, N., O'Brien, P.C.M., Jones, R.C., Ferguson-Smith, M.A., and Marshall Graves, J.A. (2004). In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature* 432, 913–917.
- Guiguen, Y., Fostier, A., and Herpin, A. (2018). Sex Determination and Differentiation in Fish: Genetic, Genomic, and Endocrine Aspects. In *Sex Control in Aquaculture*, (John Wiley & Sons), pp. 35–64.
- Hattori, R.S., Fernandino, J.I., Kishii, A., Kimura, H., Kinno, T., Oura, M., Somoza, G.M., Yokota, M., Strüssmann, C.A., and Watanabe, S. (2009). Cortisol-Induced Masculinization: Does Thermal Stress Affect Gonadal Fate in Pejerrey, a Teleost Fish with Temperature-Dependent Sex Determination? *PLOS ONE* 4, e6548.
- Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., and Strüssmann, C.A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *PNAS* 109, 2955–2959.

- Helfman, G.S., Colette, B.B., Facey, D.E., and Bowen, B.W. (2009). The diversity of fishes: biology, evolution, and ecology (Chichester, UK ; Hoboken, NJ: Blackwell).
- Herpin, A., and Schartl, M. (2015). Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Reports* 16, 1260–1274.
- Herpin, A., Braasch, I., Kraeussling, M., Schmidt, C., Thoma, E.C., Nakamura, S., Tanaka, M., and Schartl, M. (2010). Transcriptional Rewiring of the Sex Determining *dmrt1* Gene Duplicate by Transposable Elements. *PLOS Genetics* 6, e1000844.
- Holleley, C.E., O'Meally, D., Sarre, S.D., Marshall Graves, J.A., Ezaz, T., Matsubara, K., Azad, B., Zhang, X., and Georges, A. (2015). Sex reversal triggers the rapid transition from genetic to temperature-dependent sex. *Nature* 523, 79–82.
- Jalabert, B., Moreau, J., Planquette, P., and Billard, R. (1974). Déterminisme du sexe chez *tilapia macrochir* et *tilapia nilotica* : action de la méthyltestostérone dans l'alimentation des alevins sur la différenciation sexuelle; proportion des sexes dans la descendance des mâles «inversés». *Annales de Biologie Animale Biochimie Biophysique Vol. 14*, 729–739.
- Jeffries, D.L., Lavanchy, G., Sermier, R., Sredl, M.J., Miura, I., Borzée, A., Barrow, L.N., Canestrelli, D., Crochet, P.-A., Dufresnes, C., et al. (2018). A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nature Communications* 9.
- Jensen, M.P., Allen, C.D., Eguchi, T., Bell, I.P., LaCasella, E.L., Hilton, W.A., Hof, C.A.M., and Dutton, P.H. (2018). Environmental Warming and Feminization of One of the Largest Sea Turtle Populations in the World. *Current Biology* 28, 154-159.e4.
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., et al. (2012). A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genet* 8.
- Kocher, T.D. (2004). Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet* 5, 288–298.
- Koopman, P. (1995). The molecular biology of SRY and its role in sex determination in mammals. *Reprod. Fertil. Dev.* 7, 713–722.
- Kronert, U., Hörstgen-Schwark, G., and Langholz, H.-J. (1989). Prospects of selecting for late maturity in tilapia (*Oreochromis niloticus*): I. Family studies under laboratory conditions. *Aquaculture* 77, 113–121.
- Lang, J.W., and Andrews, H.V. (1994). Temperature-dependent sex determination in crocodilians. *Journal of Experimental Zoology* 270, 28–44.
- Lee, B.-Y., Penman, D.J., and Kocher, T.D. (2003). Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics* 34, 379–383.
- Lee, B.-Y., Coutanceau, J.-P., Ozouf-Costaz, C., D'Cotta, H., Baroiller, J.-F., and Kocher, T.D. (2011). Genetic and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus*). *Mar Biotechnol (NY)* 13, 557–562.
- Li, X.-Y., and Gui, J.-F. (2018). Diverse and variable sex determination mechanisms in vertebrates. *Sci. China Life Sci.* 61, 1503–1514.

- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genetics* 11, e1005678.
- Lubieniecki, K.P., Lin, S., Cabana, E.I., Li, J., Lai, Y.Y.Y., and Davidson, W.S. (2015). Genomic instability of the sex-determining locus in Atlantic salmon (*Salmo salar*). *G3 Genes | Genomes | Genetics* 5, 2513–2522.
- Mank, J.E., and Avise, J.C. (2009). Evolutionary Diversity and Turn-Over of Sex Determination in Teleost Fishes. *SXD* 3, 60–67.
- Mankiewicz, J.L., Godwin, J., Holler, B.L., Turner, P.M., Murashige, R., Shamey, R., Daniels, H.V., and Borski, R.J. (2013). Masculinizing effect of background color and cortisol in a flatfish with environmental sex-determination. *Integr Comp Biol* 53, 755–765.
- Matschiner, M., Böhne, A., Ronco, F., and Salzburger, W. (2020). The genomic timeline of cichlid fish diversification across continents. *Nat Commun* 11, 5895.
- Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., et al. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417, 559–563.
- Matveevsky, S., Kolomiets, O., Bogdanov, A., Hakhverdyan, M., and Bakloushinskaya, I. (2017). Chromosomal Evolution in Mole Voles *Ellobius* (Cricetidae, Rodentia): Bizarre Sex Chromosomes, Variable Autosomes and Meiosis. *Genes* 8, 306.
- Mei, J., and Gui, J.-F. (2015). Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish. *Sci. China Life Sci.* 58, 124–136.
- Mitchell, N.J., and Janzen, F.J. (2010). Temperature-dependent sex determination and contemporary climate change. *Sex Dev* 4, 129–140.
- Mitchell, N.J., Nelson, N.J., Cree, A., Pledger, S., Keall, S.N., and Daugherty, C.H. (2006). Support for a rare pattern of temperature-dependent sex determination in archaic reptiles: evidence from two species of tuatara (*Sphenodon*). *Front Zool* 3, 9.
- Miura, I. (2017). Sex Determination and Sex Chromosomes in Amphibia. *SXD* 11, 298–306.
- Morgan, T.H. (1909). Sex determination and parthenogenesis in phylloxerans and aphids. *Science* 29, 234–237.
- Mulugeta, E., Wassenaar, E., Sleddens-Linkels, E., van IJcken, W.F.J., Heard, E., Grootegoed, J.A., Just, W., Gribnau, J., and Baarends, W.M. (2016). Genomes of *Ellobius* species provide insight into the evolutionary dynamics of mammalian sex chromosomes. *Genome Res.* 26, 1202–1210.
- Murata, C., Kuroki, Y., Imoto, I., and Kuroiwa, A. (2016). Ancestral Y-linked genes were maintained by translocation to the X and Y chromosomes fused to an autosomal pair in the Okinawa spiny rat *Tokudaia muenninki*. *Chromosome Res* 24, 407–419.
- Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., Naruse, K., Hamaguchi, S., and Sakaizumi, M. (2012). Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics* 191, 163–170.

- Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A., et al. (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *PNAS* 99, 11778–11783.
- Navarro-Martín, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Croce, L.D., and Piferrer, F. (2011). DNA Methylation of the Gonadal Aromatase (*cyp19a*) Promoter Is Involved in Temperature-Dependent Sex Ratio Shifts in the European Sea Bass. *PLOS Genetics* 7, e1002447.
- Ndiwa, T.C., Nyingi, D.W., and Agnese, J.-F. (2014). An Important Natural Genetic Resource of *Oreochromis niloticus* (Linnaeus, 1758) Threatened by Aquaculture Activities in Lobo Drainage, Kenya. *PLOS ONE* 9, e106972.
- Nivelle, R., Gennotte, V., Kalala, E.J.K., Ngoc, N.B., Muller, M., Mélard, C., and Rougeot, C. (2019). Temperature preference of Nile tilapia (*Oreochromis niloticus*) juveniles induces spontaneous sex reversal. *PLOS ONE* 14, e0212504.
- Ohno, S. (1966). Conservation of the Original Z-Chromosome by Diverse Avian Species and Homology of the Z-linked Genes. In *Sex Chromosomes and Sex-Linked Genes*, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 73–81.
- De Oliveira, R.R., Feldberg, E., dos Anjos, M.B., and Zuanon, J. (2008). Occurrence of multiple sexual chromosomes (XX/XY1Y2 and Z1Z1Z2Z2/Z1Z2W1W2) in catfishes of the genus *Ancistrus* (Siluriformes: Loricariidae) from the Amazon basin. *Genetica* 134, 243–249.
- Ortega-Recalde, O., Goikoetxea, A., Hore, T.A., Todd, E.V., and Gemmell, N.J. (2020). The Genetics and Epigenetics of Sex Change in Fish. *Annual Review of Animal Biosciences* 8, null.
- Ospina-Álvarez, N., and Piferrer, F. (2008). Temperature-Dependent Sex Determination in Fish Revisited: Prevalence, a Single Sex Ratio Response Pattern, and Possible Effects of Climate Change. *PLOS ONE* 3, e2837.
- Otake, T., and Kuroiwa, A. (2016). Molecular mechanism of male differentiation is conserved in the SRY -absent mammal, *Tokudaia osimensis*. *Scientific Reports* 6, 32874.
- Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2013). Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS ONE* 8, e68389.
- Palaiokostas, C., Bekaert, M., Khan, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2015). A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 16, 171.
- Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.-M., Bobe, J., Concordet, J.-P., et al. (2019). Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genetics* 15, e1008013.
- Pennell, M.W., Mank, J.E., and Peichel, C.L. (2018). Transitions in sex determination and sex chromosomes across vertebrate species. *Mol Ecol* 27, 3950–3963.
- Philippart, J.-C., and Ruwet, J.-C. (1982). Ecology and distribution of tilapias (ICLARM-International Center for Living Aquatic Resources Management).

- Pokorná, M., and Kratochvíl, L. (2009). Phylogeny of sex-determining mechanisms in squamate reptiles: are sex chromosomes an evolutionary trap? *Zoological Journal of the Linnean Society* 156, 168–183.
- Ponzoni, R.W., Nguyen, N.H., Khaw, H.L., Hamzah, A., Bakar, K.R.A., and Yee, H.Y. (2011). Genetic improvement of Nile tilapia (*Oreochromis niloticus*) with special reference to the work conducted by the WorldFish Center with the GIFT strain. *Reviews in Aquaculture* 3, 27–41.
- Reddon, A.R., and Hurd, P.L. (2013). Water pH during early development influences sex ratio and male morph in a West African cichlid fish, *Pelvicachromis pulcher*. *Zoology* 116, 139–143.
- Reichwald, K., Petzold, A., Koch, P., Downie, B.R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., et al. (2015). Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell* 163, 1527–1538.
- Ribas, L., Valdivieso, A., Díaz, N., and Piferrer, F. (2017). Appropriate rearing density in domesticated zebrafish to avoid masculinization: links with the stress response. *J Exp Biol* 220, 1056–1064.
- Rice, W.R. (1987). The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. *Evolution* 41, 911–914.
- Roco, Á.S., Olmstead, A.W., Degitz, S.J., Amano, T., Zimmerman, L.B., and Bullejos, M. (2015). Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis*. *PNAS* 112, E4752–E4761.
- Sabath, N., Itescu, Y., Feldman, A., Meiri, S., Mayrose, I., and Valenzuela, N. (2016). Sex determination, longevity, and the birth and death of reptilian species. *Ecology and Evolution* 6, 5207–5220.
- Sarre, S.D., Ezaz, T., and Georges, A. (2011). Transitions Between Sex-Determining Systems in Reptiles and Amphibians. *Annu. Rev. Genom. Hum. Genet.* 12, 391–406.
- Sawatari, E., Shikina, S., Takeuchi, T., and Yoshizaki, G. (2007). A novel transforming growth factor- β superfamily member expressed in gonadal somatic cells enhances primordial germ cell and spermatogonial proliferation in rainbow trout (*Oncorhynchus mykiss*). *Developmental Biology* 301, 266–275.
- Schärer, L. (2017). The varied ways of being male and female: THE VARIED WAYS OF BEING MALE AND FEMALE. *Mol. Reprod. Dev.* 84, 94–104.
- Schmid, M., Feichtinger, W., Steinlein, C., Visbal García, R., and Fernández Badillo, A. (2003). Chromosome banding in Amphibia. *Cytogenet Genome Res* 101, 62–73.
- Schultheis, C., Böhne, A., Scharl, M., Volff, J.N., and Galiana-Arnoux, D. (2009). Sex Determination Diversity and Sex Chromosome Evolution in Poeciliid Fish. *SXD* 3, 68–77.
- Shao, C., Li, Q., Chen, S., Zhang, P., Lian, J., Hu, Q., Sun, B., Jin, L., Liu, S., Wang, Z., et al. (2014). Epigenetic modification and inheritance in sexual reversal of fish. *Genome Res* 24, 604–615.

- Shen, Z., and Wang, H. (2018). Environmental Sex Determination and Sex Differentiation in Teleosts – How Sex Is Established. In *Sex Control in Aquaculture*, (John Wiley & Sons), pp. 85–116.
- Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.-M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346, 240–244.
- Sissao, R., D’Cotta, H., Baroiller, J.-F., and Toguyeni, A. (2019). Mismatches between the genetic and phenotypic sex in the wild Kou population of Nile tilapia *Oreochromis niloticus*. *PeerJ* 7.
- Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423, 825–837.
- Smirnov, A.F., and Trukhina, A.V. (2019). Comparison of Sex Determination in Vertebrates (Nonmammals). *Sex-Determination System*.
- Smith, C.A., Roeszler, K.N., Ohnesorg, T., Cummins, D.M., Farlie, P.G., Doran, T.J., and Sinclair, A.H. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature* 461, 267–271.
- Stevens, N.M. (1905). *Studies in Spermatogenesis*. Carnegie Institution of Washington Vol. 36.
- Takehana, Y., Matsuda, M., Myosho, T., Suster, M.L., Kawakami, K., Shin-I, T., Kohara, Y., Kuroki, Y., Toyoda, A., Fujiyama, A., et al. (2014). Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nat Commun* 5, 4157.
- Tao, W., Conte, M.A., Wang, D., and Kocher, T.D. (2021). Network architecture and sex chromosome turnovers: Do epistatic interactions shape patterns of sex chromosome replacement? *BioEssays* 43, 2000161.
- Taslima, K., Wehner, S., Taggart, J.B., de Verdal, H., Benzie, J.A.H., Bekaert, M., McAndrew, B.J., and Penman, D.J. (2020). Sex determination in the GIFT strain of tilapia is controlled by a locus in linkage group 23. *BMC Genet* 21, 49.
- Taslima, K., Khan, M.G.Q., McAndrew, B.J., and Penman, D.J. (2021). Evidence of two XX/XY sex-determining loci in the Stirling stock of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 532, 735995.
- Trewavas, E. (1983). Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*.
- Ueno, K., Ota, K., and Kobayashi, T. (2001). Heteromorphic sex chromosomes of lizardfish (Synodontidae): focus on the ZZ-ZW1W2 system in *Trachinocephalus myops*. *Genetica* 111, 133–142.
- Valenzuela, N., and Lance, V. (2004). Temperature-Dependent Sex Determination in Vertebrates (Smithsonian).
- Valenzuela, N., Adams, D.C., and Janzen, F.J. (2003). Pattern Does Not Equal Process: Exactly When Is Sex Environmentally Determined? *The American Naturalist* 161, 676–683.

- Viets, B.E., Tousignant, A., Ewert, M.A., Nelson, C.E., and Crews, D. (1993). Temperature-dependent sex determination in the leopard gecko, *Eublepharis macularius*. *Journal of Experimental Zoology* 265, 679–683.
- Wang, X.G., and Orban, L. (2007). Anti-Müllerian hormone and 11 β -hydroxylase show reciprocal expression to that of aromatase in the transforming gonad of zebrafish males. *Developmental Dynamics* 236, 1329–1338.
- Wang, H., Piferrer, F., and Chen, S. (2018). *Sex Control in Aquaculture* (John Wiley & Sons).
- Weber, C., Zhou, Y., Lee, J.G., Looger, L.L., Qian, G., Ge, C., and Capel, B. (2020). Temperature-dependent sex determination is mediated by pSTAT3 repression of Kdm6b. *Science* 368, 303–306.
- Wessels, S., and Hörstgen-Schwark, G. (2007). Selection experiments to increase the proportion of males in Nile tilapia (*Oreochromis niloticus*) by means of temperature treatment. *Aquaculture* 272, S80–S87.
- Wessels, S., Krause, I., Floren, C., Schütz, E., Beck, J., and Knorr, C. (2017). ddRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC Genomics* 18, 531.
- Wilson, C.A., High, S.K., McCluskey, B.M., Amores, A., Yan, Y., Titus, T.A., Anderson, J.L., Batzel, P., Carvan, M.J., Scharl, M., et al. (2014). Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. *Genetics* 198, 1291–1308.
- Wootton, R.J. (2012). *Ecology of Teleost Fishes* (Springer Science & Business Media).
- Wright, A.E., Dean, R., Zimmer, F., and Mank, J.E. (2016). How to make a sex chromosome. *Nat Commun* 7, 12087.
- Yamaguchi, T., Yoshinaga, N., Yazawa, T., Gen, K., and Kitano, T. (2010). Cortisol Is Involved in Temperature-Dependent Sex Determination in the Japanese Flounder. *Endocrinology* 151, 3900–3908.
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., Cabau, C., Bouchez, O., Fostier, A., and Guiguen, Y. (2012). An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Current Biology* 22, 1423–1428.
- Yazdi, H.P., and Ellegren, H. (2014). Old but Not (So) Degenerated—Slow Evolution of Largely Homomorphic Sex Chromosomes in Ratites. *Molecular Biology and Evolution* 31, 1444–1453.
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., Matsuda, Y., Takamatsu, N., Shiba, T., and Ito, M. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proceedings of the National Academy of Sciences* 105, 2469–2474.
- Young, J.E., Georges, A., Doody, J.S., West, P.B., and Alderman, R.L. (2004). Pivotal range and thermosensitive period of the pig-nosed turtle, *Carettochelys insculpta* (Testudines: Carettochelydidae), from northern Australia. *Can. J. Zool.* 82, 1251–1257.
- Zhao, Y., Chen, H.J., Wang, Y.Y., Mei, Y., Huang, L.B., Wang, H., and Ji, X.S. (2019). Gonad development examination of high-temperature-treated genetically female Nile tilapia. *Aquaculture* 515, 734535.

CONTEXTE ET OBJECTIFS DE LA THÈSE

Contexte

Nous avons vu dans l'introduction générale que le sexe chez le tilapia du Nil est un processus complexe avec un déterminisme génétique du sexe de type XX/XY qui peut être influencé par la température du milieu et par des facteurs génétiques mineurs (parentaux). Les températures au delà de 32°C peuvent ainsi conduire des individus XX à se développer en mâles fonctionnels et biaiser les sex-ratios. Ce phénomène a été observé dans les populations sauvages (Bezault et al., 2007). Cependant, à part des analyses microsatellites sur des individus sauvages qui ont donné des résultats peu clairs (Bezault, 2005), les études précédentes sur le déterminisme génétique du sexe n'ont été réalisées que sur des souches domestiquées (commerciales ou de laboratoire). Du fait de l'importance du tilapia du Nil en aquaculture, ces souches ont fait l'objet de nombreux transferts de stocks, de consanguinité et de goulots d'étranglement génétiques. Par conséquent, au début de cette thèse, nous ne savions pas dans quelle mesure la complexité du déterminisme génétique du sexe, avec une liaison sexuelle sur LG1 ou LG23, était due aux processus de domestication ou si elle reflétait les différences dans le déterminisme du sexe en milieu naturel. En outre, en raison de l'adaptation des populations de tilapia du Nil à différents régimes thermiques, parfois extrêmes, nous nous sommes demandé quel était le système dominant de déterminisme génétique du sexe, et de quelle manière il pouvait coexister avec l'influence de la température au sein de ces populations sauvages. Des études préliminaires, avant mon arrivée dans le groupe, réalisées sur des populations sauvages en utilisant des marqueurs PCR de l'*amh*, ont suggéré que le sexe phénotypique n'était pas lié au LG23 dans les populations éthiopiennes.

Objectifs

L'objectif de ma thèse était d'étudier les bases génétiques du déterminisme du sexe dans les populations sauvages.

La première partie de la thèse visait à étudier si les populations sauvages suivaient un système XY hétérogamétique mâle similaire à celui détecté dans les souches domestiques, et de trouver quel(s) groupe(s) de liaison ségrégeait(ent) dans différentes populations sauvages pour définir le sexe phénotypique. Pour cela, nous avons choisi la population Kpandu du Lac Volta en Afrique de l'Ouest appartenant à la vaste région soudano-sahélienne et la population Koka, une population éthiopienne du Lac Koka en Afrique de l'Est comprenant le bassin de l'Awash qui a été isolé pendant la formation du rift éthiopien. Ces populations vivent dans des habitats thermiques différents et avec de fortes variations saisonnières : entre 28 et 34°C dans le lac Volta et entre 22 et 26°C dans le lac Koka. Nous avons réalisé deux approches génomiques complémentaires en utilisant des échantillons déjà en notre possession. La première était une approche de séquençages ddRAD (double digestion enzymatique pour le Séquençage d'ADN associé à un site de restriction) sur des familles élevées dans des conditions de température contrôlée, engendrées par des individus capturés dans la nature. La seconde approche consistait en un séquençage du génome complet (WGS) d'individus capturés dans la nature et groupés par sexe phénotypique.

Après avoir trouvé que le sexe était associé à une duplication de *l'amh* sur LG23 dans les populations Kpandu et Koka, **le second objectif** était d'étudier plus finement cette région du sexe et de détailler l'ordre des gènes de cette structure. Pour cela, nous avons utilisé le séquençage MinION long reads (séquençage de lectures longues) sur des individus mâles. Un autre des objectifs de cette étude était d'identifier des régions (si possible universelles) qui nous permettraient de génotyper le sexe des poissons éthiopiens ainsi que dans d'autres populations non étudiées, afin de pouvoir analyser les sex-ratios génétiques vs phénotypiques d'individus sauvages vivant dans différents régimes thermiques.

Comme les études préliminaires suggéraient que les populations éthiopienne ne possédaient pas *l'amh* comme déterminant du sexe, **le troisième objectif** de ce projet était de réaliser une étude transcriptomique sur la population éthiopienne Hora (du lac Hora) afin d'étudier s'il pourrait y avoir des modifications dans la cascade de la différenciation sexuelle. Nous avons donc voulu comparer les gènes impliqués dans la différenciation du sexe chez cette population à ce qui était déjà connu dans la littérature sur les souches domestiques, et analyser les effets précoces des hautes températures sur la voie de la différenciation sexuelle. Nous nous attendions également à observer, au stade précoce de la différenciation sexuelle

une expression dimorphique d'un gène qui agirait potentiellement comme déterminant du sexe. Pour ce troisième objectif, nous disposons dans nos infrastructures du CIRAD de descendants de poissons sauvages appartenant à la population éthiopienne du lac Hora. Ces derniers ont été croisés et la descendance traitée dans deux conditions de température : une température contrôle de 27°C et une température masculinisante de 36°C.

Références

- Bezault, E. (2005). Etude du Système de déterminisme du sexe au sein de populations naturelles de Tilapia du Nil, *Oreochromis niloticus* (linnaeus, 1758) : importance des composantes génétiques et environnementales. Université de Paris-Sud.
- Bezault, E., Clota, F., Derivaz, M., Chevassus, B., and Baroiller, J.-F. (2007). Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture* 272, S3–S16.

CONTEXT AND OBJECTIVES OF THE THESIS PROJECT

Context

We have seen in the general introduction that sex in Nile tilapia is complex with a XX/XY genetic sex determination that can be influenced by temperature as well as minor genetic (parental) factors. Temperatures above 32°C can thus lead to XX individuals developing into functional males and bias the sex-ratios. This phenomenon has been observed in wild populations (Bezault et al., 2007). However, apart from microsatellite sex linkage studies on wild individuals which gave unclear results (Bezault, 2005), previous studies on the genetic sex determination had been performed only on domesticated strains (commercial or laboratory). Due to the importance of Nile tilapia in aquaculture, these strains have been subject to numerous stock transfers, inbreeding and genetic bottlenecks. Therefore, at the beginning of this thesis, we did not know to what extent the complexity of genetic sex determination, with sex-linkage on LG1 or LG23, was due to domestication processes or whether they stemmed from population differences in sex determination. Furthermore, because of the adaptation of Nile tilapia populations to different, sometimes extreme, thermal regimes, we questioned what the prevailing sex determinant system, and how it could coexist with the influence of temperature in these wild populations. Preliminary studies, prior to my arrival in the group, using *amh* PCR markers suggested that the phenotypic sex was not linked to LG23 in Ethiopian populations.

Objectives

The objective of my thesis was to investigate the genetic basis of sex determinism in wild populations.

The first part of the thesis aimed to investigate whether wild populations follow a male heterogamous XY system as detected in domestic strains, and to determine which linkage group(s) segregate(s) in the different wild populations to determine phenotypic sex. For this, we chose the Kpandu population from Lake Volta in West Africa belonging to the vast

Sudano-Sahelian region and the Koka population, an Ethiopian population from Lake Koka in East Africa comprising the Awash Basin that was isolated during the formation of the Ethiopian Rift. These populations live in different thermal habitats and with strong seasonal variations: between 28 and 34°C in Lake Volta and between 22 and 26°C in Lake Koka. We performed two complementary genomic approaches using samples already in our possession. The first was a double digest restriction-site associated DNA sequencing approach (ddRAD) on families reared under controlled temperature conditions, sired from wild caught individuals. The second approach consisted of whole genome sequencing (WGS) of wild-caught individuals pooled by phenotypic sex.

Having found that sex was associated with a duplication of the *amh* gene on LG23 in both the Kpandu and Koka populations, **the second objective** was to study more finely this sex region and to determine the gene order within this structure. For this, we used MinION long reads sequencing on male individuals. Another objective of this study was to identify regions (putatively universal) which would allow us to genotype the sex of the Ethiopian fish as well as in other unstudied populations, with the aim for future studies on the genetic vs phenotypic sex-ratios of wild individuals living in different temperature regimes.

As preliminary studies suggested that the Ethiopian populations did not have the *amh* gene as a sex determinant, **the third focus** of this project was to carry out a transcriptomic study on the Ethiopian Hora population (from Lake Hora) to investigate whether there might be changes in the sex differentiation pathway. We therefore wanted to compare the genes involved in sex differentiation in this population with what was already known in the literature on domestic strains, and to analyse the early effects of high temperatures on the sex differentiation pathway. We also expected to observe the dimorphic expression of a putative master sex gene at early stages of sex-differentiation. For this third objective, we had in our CIRAD infrastructures the descendants of wild-caught fish belonging to the Ethiopian Lake Hora population, which were reproduced, and the progeny treated under control conditions of 27°C and high masculinizing temperatures of 36°C

References

Bezault, E. (2005). Etude du Système de déterminisme du sexe au sein de populations naturelles de Tilapia du Nil, *Oreochromis niloticus* (linnaeus, 1758) : importance des composantes génétiques et environnementales. Université de Paris-Sud.

Bezault, E., Clota, F., Derivaz, M., Chevassus, B., and Baroiller, J.-F. (2007). Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture* 272, S3–S16.

CHAPITRE 1 : DÉTERMINISME GÉNÉTIQUE DU SEXE DE DEUX POPULATIONS NATURELLES

Objectifs

Au vu de résultats longtemps apparus comme contradictoires dans la littérature des souches domestiques sur le déterminisme du sexe du tilapia du Nil, le premier objectif était d'explorer les bases génétiques du déterminisme du sexe en populations naturelles pour cette espèce afin de comprendre si la diversité observée dans les souches domestiquées est liée à une diversité existant également dans le milieu naturel.

Ce chapitre a donné lieu à une publication dans *Genes* (MDPI) en août 2020.

Introduction et contexte

Le déterminisme du sexe chez le tilapia du Nil a été largement étudié sur des stocks domestiques (commerciaux ou de laboratoires). Malgré un intérêt économique évident pour comprendre le déterminisme du sexe de cette espèce en vue de produire plus de mâles, celui-ci n'est toujours pas clair. Des analyses de sex-ratios ont confirmé que le déterminisme suit un patron de type XX/XY (Jalabert et al., 1974). Cependant, la paire de chromosomes sexuels est homomorphique et il n'est donc pas possible de les identifier par une simple observation du caryotype. Des observations cytogénétiques de complexes synaptonémaux ont montré qu'il n'y avait pas d'appariement près des extrémités du groupe de liaison 3 (LG3) – la plus grande paire de chromosomes chez le tilapia – chez les mâles hétérogamétiques XY contrairement à ce qui est observé chez les mâles homogamétiques XX ou YY, laissant penser qu'il s'agissait de la paire de chromosomes sexuels (Carrasco et al., 1999). Cependant, des analyses génétiques et génomiques ont montré par la suite des associations du sexe avec LG1 (Lee et al., 2003; Cnaani et al., 2008; Lee et al., 2011; Palaiokostas et al., 2013; Gammerdinger et al., 2014; Palaiokostas et al., 2015; Taslima et al., 2020) et avec LG23 selon les souches étudiées (Eshel et al., 2012, 2014; Li et al., 2015; Wessels et al., 2017; Cáceres et al., 2019; Taslima et al., 2021). Si aucun gène précis n'a été proposé sur LG1, sur LG23 en revanche, une région comprenant une duplication en tandem du gène de l'Hormone Anti-Müllérienne (*amh*) est proposée comme déterminant du sexe sur le chromosome Y (Eshel et al., 2014; Li et al., 2015). Dans cette duplication, une copie, l'*amhY*, est conservée, alors que la seconde, l'*amhΔY* présente des indels dont une insertion de 5bp qui entraîne un codon stop prématuré (Li et al., 2015). Des marqueurs PCR ont été développés pour sexer les individus sur la base de caractéristiques de la duplication en tandem de l'*amh*. Cependant, si ces marqueurs

fonctionnent sur certaines souches domestiques, ils ne permettent pas d'expliquer la totalité des phénotypes observés dans la population naturelle Kou du Burkina-Faso (Sissao et al., 2019). Par exemple, 2% des mâles et 2% des femelles ont des génotypes inattendus d'après les marqueurs PCR à l'*amh* puisqu'ils amplifient l'*amhX* et l'*amhY* mais pas l'*amhΔY*, suggérant la possibilité d'un polymorphisme de cette région dans les populations sauvages.

Malgré cela, le gène de l'*amh* demeure un candidat robuste pour le déterminisme du sexe puisqu'il intervient dans le développement des gonades et la régulation des cellules germinales de certains téléostéens et s'exprime même dans le cerveau chez certaines espèces (Poonlaphdecha et al., 2011; Pfenning et al., 2015). De plus, des duplications du gène de l'*amh* ont déjà été réquisitionnées au sommet de la cascade du déterminisme par exemple chez le Brochet (*Esox lucius*) et le Pejerrey (*Odontesthes hatcheri*) par exemple (Hattori et al., 2012; Pan et al., 2019).

Comme les études sur les bases génétiques du sexe de cette espèce ont été conduites principalement sur des souches de laboratoires ou des stocks commerciaux, les individus sont susceptibles d'avoir subi de forts goulots d'étranglements et de forts taux de consanguinité du fait d'effectifs réduits des fondateurs ainsi que de nombreuses sélections consécutives pour les croisements voire de potentielles hybridations (volontaires ou non) avec des espèces proches ou d'autres souches plus performantes. Un déterminisme du sexe différent de celui des populations naturelles suite à une domestication en laboratoire a déjà été observé chez le poisson-zèbre (*Danio rerio*) par exemple (Wilson et al., 2014). De fait, il est difficile de savoir si la diversité observée du déterminisme génétique du sexe chez le tilapia du Nil est le reflet d'un polymorphisme du milieu naturel ou s'il s'agit en réalité des conséquences de la domestication de ces souches.

Matériels et méthodes

Pour répondre à ce questionnement, nous avons choisi d'étudier deux populations naturelles d'*O. niloticus* de l'Est (Ethiopie, Lac Koka) et de l'Ouest (Ghana, Lac Volta, « Kpandu ») de l'Afrique. Ces deux populations vivent dans des conditions écologiques différentes, notamment en termes de régimes thermiques du milieu oscillant entre 17 et 26°C pendant l'année pour le lac Koka et de 27 à 32°C pour le lac Volta.

Deux approches différentes ont été utilisées : dans un premier temps une approche exploratoire de ddRAD (double digestion enzymatique) concentrée sur une famille de chaque population (descendance + géniteurs) afin d'analyser un grand nombre d'individus et explorer l'ensemble du génome à moindre coût ; et dans un second temps une approche plus précise de Séquençage de Génome Complet (WGS) sur un pool de mâles et un pool de femelles pour chacune des populations afin de caractériser au mieux les régions identifiées par ddRAD en utilisant des individus de F0 directement issus du milieu naturel.

La mise au point du ddRAD a nécessité un séjour d'apprentissage à l'Université de Stirling (Écosse) dans l'équipe de David Penman, puis une adaptation et une optimisation de la méthode aux dispositifs présents sur les plateformes de biologie moléculaire de Montpellier. De plus, en vue de l'utilisation d'un grand nombre d'individus pour l'ensemble des volets du projet de thèse, nous avons automatisé et optimisé l'extraction et la quantification de l'ADN.

Un séjour d'apprentissage pour l'analyse des données de WGS a également été effectué à l'Université du Maryland (USA) au sein de l'équipe de Thomas D. Kocher.

Pour les deux types de données, les analyses ont suivies un même fil conducteur avec un alignement sur le dernier génome de référence disponible (génome d'une femelle XX de la souche Manzala-Stirling), puis les sites polymorphes pour les deux sexes ont été recensés le long du génome et des mesures statistiques de type *Fst* et Test Exact de Fisher ont été effectuées entre mâles et femelles afin de déterminer les sites présentant un fort dimorphisme entre les deux sexes le long du génome.

Pour les données de génomes complets, l'utilisation de lectures paires (paired-end reads) a également permis de se pencher sur la structure de la région du déterminant du sexe. En effet, les lectures courtes de type Illumina peuvent être limitantes dans l'analyse de régions répétées. Nous avons donc utilisé des outils analysant l'orientation des lectures ainsi que la taille de l'insert (l'espace entre deux lectures paires) afin de déterminer la structure chromosomique de cette région.

Principaux résultats et discussion

Pour résumer les principaux résultats, nous avons mis en évidence que :

i. Il existe une forte association avec le sexe sur LG23 dans les deux populations

Les analyses de ddRAD ont montré une forte association avec le sexe phénotypique sur LG23 dans les deux populations sauvages. Sur ce groupe de liaison, la majorité du signal sexe-spécifique est concentrée sur la région de 24 Mb à 40Mb, c'est à dire centré autour du locus de l'*amh*. Ces résultats vont dans le sens de ceux observés sur les souches commerciales Sud Américaines, la souche Japonaise ainsi que la population sauvage du Burkina-Faso notamment (Li et al., 2015; Cáceres et al., 2019; Sissao et al., 2019).

Dans la population de l'Ouest (Ghana), un signal dimorphique entre mâles et femelles est également observé sur LG3. Cependant, ces signaux semblent être liés à des problèmes d'assemblages du génome sur cette région, possiblement dûs à la présence de séquences répétées et de familles multigéniques (Conte et al., 2019).

En revanche, aucun signal sexuel n'a été détecté sur LG1 comme cela a pu être observé dans le stock commercial Américain ou encore la souche Manzala-Stirling écossaise (Gammerdinger et al., 2014; Palaiokostas et al., 2015).

ii. Les patrons hérités dans la descendance confirment un système hétérogamétique mâle

La comparaison des génotypes (homozygotes ou hétérozygotes) sur LG23 entre les géniteurs et leur descendance confirme l'existence d'un système hétérogamétique mâle. Les génotypes des descendants sur les sites les plus proches de l'*amh* permettent de prédire le sexe phénotypique des individus et montrent également l'existence de crossing-overs proches de l'*amh* suggérant que les chromosomes X et Y recombinent toujours près de la région du déterminisme du sexe. L'utilisation du ddRAD est cependant limitante pour analyser en détails cette région puisque seules les régions présentes entre les sites des deux enzymes de restrictions utilisées pour construire la banque, et ayant passé les filtres sur la taille, sont séquencées et donc analysées par la suite. De fait, cette méthode donne naissance à un ensemble que l'on pourrait assimiler à un texte à trous, ce qui empêche parfois d'explorer une région particulière si elle ne fait pas partie des sites retenus lors de la préparation de la banque. Ainsi, il était impossible d'observer les génotypes à l'*amh* exactement par exemple, ni même effectuer une comparaison exacte sites à sites pour l'ensemble des séquences des différents individus séquencés.

iii. Des variants mâles spécifiques sur le gène de l'Hormone Anti-Müllerienne

iii.a. Confirmation de la duplication de l'*amh* chez les mâles

Une analyse attentive de l'appariement des lectures du WGS a permis de confirmer l'existence d'une duplication du gène de l'*amh* uniquement chez les mâles malgré l'utilisation de lectures courtes (Figure 2.1). Ainsi, nous avons pu confirmer l'existence de deux copies de l'*amh* chez la population de l'Ouest (déjà suggéré par l'analyse de marqueurs PCR), mais également chez la population Ethiopienne (chez qui les marqueurs PCR ne permettaient pas de mettre en évidence la présence de duplication chez les mâles). De plus, même s'il est difficile de distinguer les deux copies avec les méthodes utilisées, dans la population de l'Ouest (Kpandu) l'une des copie est conservée tandis que l'autre présente les indels déjà décrits dans la souche japonaise par Li et al., (2015) entraînant notamment un codon stop prématuré dans la séquence codante et caractérisant la copie de l'*amh* Δ Y. A contrario, aucun des indels spécifiques de l'*amh* Δ Y n'a été mis en évidence dans la population de l'Est (Koka), laissant sans réponse la question du rôle fonctionnel de la seconde copie sur le Y.

iii.b. Première suggestion précise de la région du sexe sur l'haplotype Y

De plus, l'utilisation d'un séquençage de génome complet a permis de reconstruire précisément la région de la duplication en identifiant des variants structuraux spécifiques aux mâles. Ainsi, nous avons pu définir pour la première fois les limites de cette duplication et décrire un haplotype Y attendu. Il s'agit de la première caractérisation de la duplication complète de la région du sexe chez le tilapia du Nil et elle confirme les fortes similitudes de cette structure entre les deux populations, indiquant certainement une origine ancestrale de la duplication chez les populations naturelles. Certains SNPs mâles-spécifiques ont également pu être mis en évidence et parmi eux se trouvent des variants alléliques déjà décrits dans des études antérieures, qui constituent des candidats intéressants pour le déterminisme du sexe. En revanche, l'utilisation de lectures courtes ne nous a pas permis d'assembler la séquence précise de l'haplotype Y et donc de préciser sur quelles copies des gènes se situaient ces SNPs au sein de la duplication.

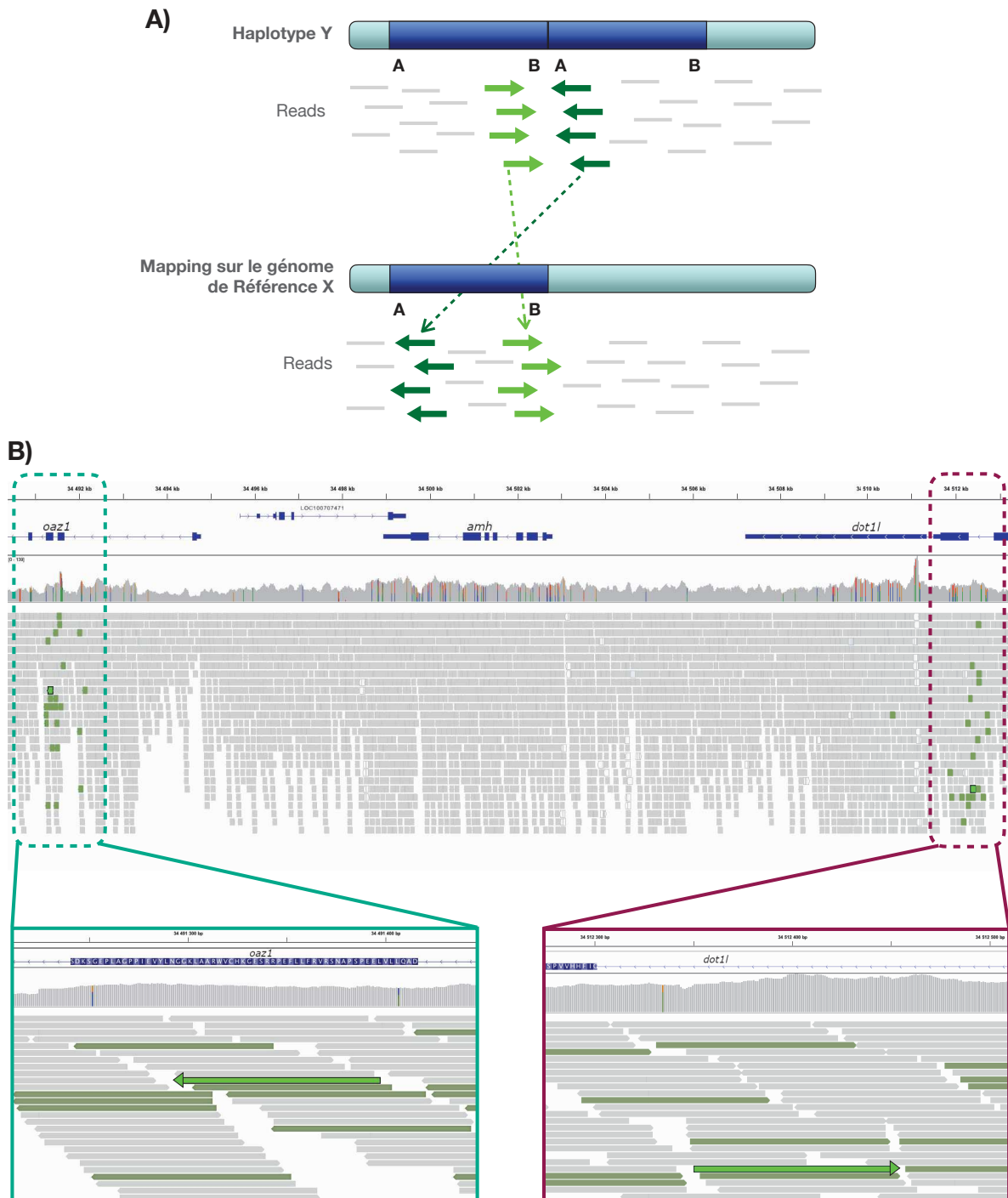


Figure 2.1. Illustration de la détection d'une duplication par l'étude de lectures courtes sur un génome complet, visionné dans IGV (Integrative Genome Viewer).

A) La partie supérieure représente des reads (lectures) paires en gris et en vert échantillonnées le long de l'haplotype Y. La partie inférieure montre la réalité observée lorsque les reads appartenant à l'haplotype Y sont alignées sur une référence qui ne possède pas la duplication. Les lectures chevauchant la région BA de l'haplotype Y sont éloignées et ne sont pas orientées correctement car elles viennent s'aligner sur les régions avec lesquelles elles ont le plus de similarités. **B)** Captures d'écrans de IGV de la région dupliquée dans les données de WGS du pool de mâles Kpandu. La couleur verte des lectures dans IGV signifie que leurs paires correspondantes sont éloignées plus qu'attendu et qu'elles sont mal orientées, indiquant la présence d'une duplication de cette région.

iv. Suspicion de la présence d'un chromosome B dans les populations naturelles

Finalement, un point essentiel de ces analyses a été la détection dans le génome des populations sauvages de patrons de couvertures laissant penser qu'un chromosome B pourrait ségréger dans le milieu naturel. Il s'agit de la première étude à notre connaissance mettant en évidence des indices de la présence d'un tel élément génomique dans les populations sauvages. Le chromosome B est un chromosome surnuméraire et non essentiel dans le génome d'un individu ou d'une espèce. Des chromosomes B ont déjà été identifiés chez des Cichlidés et ils ne sont pas toujours portés par l'ensemble des individus d'une population (Valente et al., 2014; Clark et al., 2017; Fantinatti & Martins, 2016; Perazzo et al., 2018). Certains peuvent parfois agir comme un chromosome sexuel s'ils contiennent un locus agissant comme un déterminant du sexe (Clark & Kocher, 2019). Ce chromosome non-essentiel peut être constitué de régions appartenant à différents groupes de liaisons dans le génome qui sont copiées et dupliquées (parfois en des centaines de copies et donc particulièrement polymorphiques) contribuant alors à son allongement (Valente et al., 2014, Figure 2.2). Or, lorsque l'on utilise une méthode d'alignement sur un génome de référence, le chromosome B ne figure pas dans la référence et les différentes lectures appartenant au chromosome B sont alignées sur le gène le plus similaire du génome de référence, créant ainsi des régions de très fortes couvertures fortement polymorphiques (Valente et al., 2014; Clark et al., 2017, 2018).

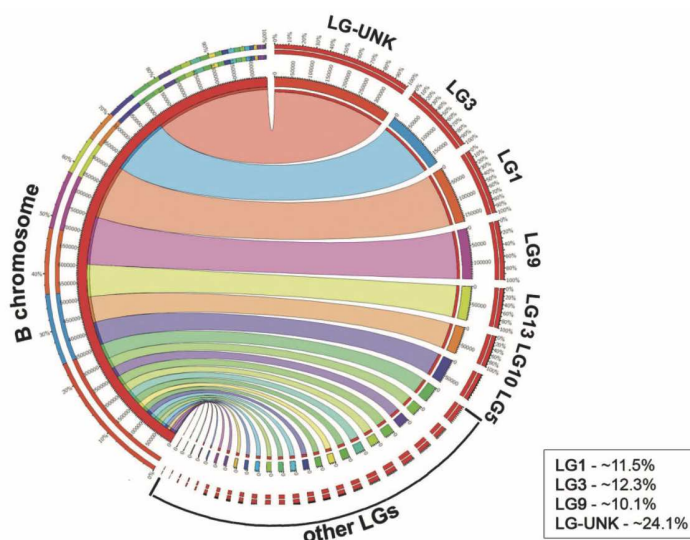


Figure 2.2. Relations entre les régions du chromosome B d'*Astatotilapia latifasciata* et les LGs de *Metriaclicma zebra*.

Les régions du B sont représentées à gauche et les LGs correspondants à droite. Le pourcentage du génome des principaux LGs (1, 3, et 9) figurant sur le B est mis en évidence dans l'encadré. Issue et traduit de Valente et al., 2014.

Ce chromosome B n'a pas montré d'association avec le sexe à l'échelle de l'espèce puisqu'il est retrouvé chez les mâles dans une population et chez les femelles dans l'autre. Cependant, sa présence a été une réelle contrainte à l'exploitation du WGS puisque nous avons effectué les analyses pour chacune des populations (Kpandu et Koka) séparément et qu'il n'est présent que dans l'un des deux pools par population. Ainsi, la présence de ces régions apparaît spécifique du sexe et la forte densité de SNPs présents mène à l'identification de forts pics d'association au sexe sur l'ensemble des groupes de liaisons.

La suspicion de ce chromosome B dans les pools a donc limité nos analyses à la région associée au sexe préalablement identifiée par les analyses de ddRAD. S'il ne figure pas de caractérisation propre du chromosome B dans l'article, certaines analyses complémentaires ont été effectuées afin de mieux appréhender sa présence dans nos populations sauvages ainsi que son contenu en gènes. Ces analyses sont explicitées à la fin de ce chapitre dans la section « Analyses Complémentaires ».

Conclusion

Cette étude a été la première caractérisation génomique du déterminisme du sexe utilisant des populations de tilapia du Nil issues du milieu naturel. Nous avons pu mettre en évidence que le déterminisme du sexe chez cette espèce est en lien étroit avec la région de l'*amh* sur LG23 et suit un patron hétérogamétique mâle dans les populations sauvages. Nous avons également pu confirmer l'existence d'une duplication de l'*amh* spécifique aux mâles dans les deux populations naturelles, ce qui n'avait pas pu être fait par le passé en utilisant des approches de marqueurs PCR chez les populations Ethiopiennes. Ces analyses ont également pu mettre en lumière la structure partagée de cette duplication sur le Y en précisant le contenu, les éléments spécifiques au Y ainsi que les limites de cette duplication malgré les conditions écologiques et les aires géographiques différentes de ces populations. Des SNPs spécifiquement mâles, et partagés par d'autres souches, ont été identifiés comme candidats pour le déterminisme du sexe et des variants spécifiques à chacune des populations ont également été mis en évidence. Enfin, il s'agit de la première étude qui fait état d'indices suggérant la présence d'un chromosome B, non lié au sexe à l'échelle des populations, qui ségrégerait en milieu naturel.

Références

- Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Trivisany, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., et al. (2019). Fine mapping using whole-genome sequencing confirms anti-müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3 (Bethesda)* 9, 3213–3223.
- Carrasco, L.A.P., Penman, D.J., and Bromage, N. (1999). Evidence for the presence of sex chromosomes in the Nile tilapia (*Oreochromis niloticus*) from synaptonemal complex analysis of XX, XY and YY genotypes. *Aquaculture* 173, 207–218.
- Clark, F.E., and Kocher, T.D. (2019). Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. *Scientific Reports* 9, 1–10.
- Clark, F.E., Conte, M.A., Ferreira-Bravo, I.A., Poletto, A.B., Martins, C., and Kocher, T.D. (2017). Dynamic sequence evolution of a sex-associated B chromosome in Lake Malawi cichlid fish. *JHERED* 108, 53–62.
- Clark, F.E., Conte, M.A., and Kocher, T.D. (2018). Genomic characterization of a B chromosome in Lake Malawi cichlid fishes. *Genes* 9, 610.
- Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., D'Hont, A., Baroiller, J.-F., D'Cotta, H., Penman, D.J., et al. (2008). Genetics of sex determination in tilapiine species. *Sexual Development* 2, 43–54.
- Conte, M.A., Joshi, R., Moore, E.C., Nandamuri, S.P., Gammerdinger, W.J., Roberts, R.B., Carleton, K.L., Lien, S., and Kocher, T.D. (2019). Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *Gigascience* 8.
- Eshel, O., Shirak, A., Weller, J.I., Hulata, G., and Ron, M. (2012). Linkage and physical mapping of sex Region on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3: Genes, Genomes, Genetics* 2, 35–42.
- Eshel, O., Shirak, A., Dor, L., Band, M., Zak, T., Markovich-Gordon, M., Chalifa-Caspi, V., Feldmesser, E., Weller, J.I., Seroussi, E., et al. (2014). Identification of male-specific *amh* duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 15, 774.
- Fantinatti, B.E.A., and Martins, C. (2016). Development of chromosomal markers based on next-generation sequencing: the B chromosome of the cichlid fish *Astatotilapia latifasciata* as a model. *BMC Genetics* 17, 119.
- Gammerdinger, W.J., Conte, M.A., Acquah, E.A., Roberts, R.B., and Kocher, T.D. (2014). Structure and decay of a proto-Y region in Tilapia, *Oreochromis niloticus*. *BMC Genomics* 15, 975.
- Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., and Strüssmann, C.A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *PNAS* 109, 2955–2959.
- Jalabert, B., Moreau, J., Planquette, P., and Billard, R. (1974). Déterminisme du sexe chez *tilapia macrochir* et *tilapia nilotica* : action de la méthyltestostérone dans l'alimentation des

- alevins sur la différenciation sexuelle; proportion des sexes dans la descendance des mâles «inversés». *Annales de Biologie Animale Biochimie Biophysique Vol. 14*, 729–739.
- Lee, B.-Y., Penman, D.J., and Kocher, T.D. (2003). Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics* 34, 379–383.
- Lee, B.-Y., Coutanceau, J.-P., Ozouf-Costaz, C., D’Cotta, H., Baroiller, J.-F., and Kocher, T.D. (2011). Genetic and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus*). *Mar Biotechnol (NY)* 13, 557–562.
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genetics* 11, e1005678.
- Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2013). Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS ONE* 8, e68389.
- Palaiokostas, C., Bekaert, M., Khan, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2015). A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 16, 171.
- Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.-M., Bobe, J., Concordet, J.-P., et al. (2019). Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genetics* 15, e1008013.
- Perazzo, G.X., Noleto, R.B., Vicari, M.R., Gava, A., and Cestari, M.M. (2018). B chromosome polymorphism in South American cichlid. *Neotropical Biodiversity* 4, 3–9.
- Pfennig, F., Standke, A., and Gutzeit, H.O. (2015). The role of *Amh* signaling in teleost fish – Multiple functions not restricted to the gonads. *General and Comparative Endocrinology* 223, 87–107.
- Poonlaphdecha, S., Pepey, E., Huang, S.-H., Canonne, M., Soler, L., Mortaji, S., Morand, S., Pfennig, F., Mélard, C., Baroiller, J.F., et al. (2011). Elevated *amh* Gene Expression in the Brain of Male Tilapia (*Oreochromis niloticus*) during Testis Differentiation. *Sex Dev* 5, 33–47.
- Sissao, R., D’Cotta, H., Baroiller, J.-F., and Toguyeni, A. (2019). Mismatches between the genetic and phenotypic sex in the wild Kou population of Nile tilapia *Oreochromis niloticus*. *PeerJ* 7.
- Taslima, K., Wehner, S., Taggart, J.B., de Verdal, H., Benzie, J.A.H., Bekaert, M., McAndrew, B.J., and Penman, D.J. (2020). Sex determination in the GIFT strain of tilapia is controlled by a locus in linkage group 23. *BMC Genet* 21, 49.
- Taslima, K., Khan, M.G.Q., McAndrew, B.J., and Penman, D.J. (2021). Evidence of two XX/XY sex-determining loci in the Stirling stock of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 532, 735995.
- Valente, G.T., Conte, M.A., Fantinatti, B.E.A., Cabral-de-Mello, D.C., Carvalho, R.F., Vicari, M.R., Kocher, T.D., and Martins, C. (2014). Origin and evolution of B chromosomes in the

- cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Mol Biol Evol* 31, 2061–2072.
- Wessels, S., Krause, I., Floren, C., Schütz, E., Beck, J., and Knorr, C. (2017). ddRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC Genomics* 18, 531.
- Wilson, C.A., High, S.K., McCluskey, B.M., Amores, A., Yan, Y., Titus, T.A., Anderson, J.L., Batzel, P., Carvan, M.J., Scharl, M., et al. (2014). Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. *Genetics* 198, 1291–1308.

PUBLICATION ASSOCIÉE

Article

Structure and Sequence of the Sex Determining Locus in Two Wild Populations of Nile Tilapia

Cécile Triay ^{1,2}, Matthew A. Conte ³, Jean-François Baroiller ^{1,2}, Etienne Bezault ⁴ ,
Frances E. Clark ³ , David J. Penman ⁵, Thomas D. Kocher ³  and Helena D'Cotta ^{1,2,*}

¹ UMR-Institut des Sciences de l'Evolution de Montpellier, Centre National de la Recherche Scientifique, Institut de Recherche pour le Développement, Ecole Pratique des Hautes Etudes, University of Montpellier, 34090 Montpellier, France; cecile.triay@cirad.fr (C.T.); baroiller@cirad.fr (J.-F.B.)

² UMR-Institut des Sciences de l'Evolution de Montpellier, Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Campus Int. Baillarguet, 34398 Montpellier, France

³ Department of Biology, University of Maryland, College Park, MD 20742, USA; mconte@umd.edu (M.A.C.); betsy.clark@nih.gov (F.E.C.); tdk@umd.edu (T.D.K.)

⁴ UMR BOREA, CNRS-7208/MNH/UPMC/IRD-207/UCN/UA, Université des Antilles, 97159 Guadeloupe, France; etienne.bezault@univ-antilles.fr

⁵ Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling, Scotland FK9 4LA, UK; davidpenman59@gmail.com

* Correspondence: dcotta@cirad.fr; Tel.: +33-4-67593994

Received: 17 July 2020; Accepted: 27 August 2020; Published: 29 August 2020



Abstract: In domesticated strains of the Nile tilapia, phenotypic sex has been linked to genetic variants on linkage groups 1, 20 and 23. This diversity of sex-loci might reflect a naturally polymorphic sex determination system in Nile tilapia, or it might be an artefact arising from the process of domestication. Here, we searched for sex-determiners in wild populations from Kpandu, Lake Volta (Ghana-West Africa), and from Lake Koka (Ethiopia-East Africa) that have not been subjected to any genetic manipulation. We analysed lab-reared families using double-digest Restriction Associated DNA sequencing (ddRAD) and analysed wild-caught males and females with pooled whole-genome sequencing (WGS). Strong sex-linked signals were found on LG23 in both populations, and sex-linked signals with LG3 were observed in Kpandu samples. WGS uncovered blocks of high sequence coverage, suggesting the presence of B chromosomes. We confirmed the existence of a tandem *amh* duplication in LG23 in both populations and determined its breakpoints between the *oaz1* and *dot11* genes. We found two common deletions of ~5 kb in males and confirmed the presence of both *amhY* and *amhΔY* genes. Males from Lake Koka lack both the previously reported 234 bp deletion and the 5 bp frameshift-insertion that creates a premature stop codon in *amhΔY*.

Keywords: sex-determinant; Y-chromosome; *amh*; male duplication; Nile tilapia

1. Introduction

The male heterogametic (XY) system of sex determination in mammals, and the female heterogametic (ZW) system in birds, are evolutionarily conserved in most species of their respective clades [1,2]. Furthermore, the master sex-determining gene on the Y chromosome, *sry*, is almost universal across the mammalian class [3]. The sex chromosomes of most mammals and birds are highly differentiated due to the suppression of recombination, and the subsequent accumulation of deleterious mutations which has resulted in the degeneration of the Y (or W) chromosomes (reviewed by Bachtrog [4]). Fish in contrast, usually have homomorphic sex chromosomes, so the sex of an individual typically cannot be distinguished by karyotyping. Genetic sex determination (GSD) has been established for many fish species by progeny testing, sex-reversal treatments and by

the association of genetic markers with phenotypic sex (see reviews of Baroiller et al. [5]; Devlin and Nagahama [6]). These studies have demonstrated a wide range of GSD mechanisms, including both mono- and polygenic systems. Sex determiners have been described on undifferentiated autosomes and well-differentiated sex chromosomes, and in some cases interact with environmental factors to determine phenotypic sex [7].

Frequent replacement of fish sex-determining systems is inferred from the fact that both male XY and female ZW heterogametic systems can be found between and even within fish lineages [8,9]. The existence of multiple sex chromosomes (e.g., *Xiphophorus*) or polygenic systems (e.g., *Dicentrarchus labrax*) within species also suggest the dynamic evolution of fish sex chromosomes [10–12]. The diversity and rapid turnover of fish sex chromosomes have been made more apparent with the identification of sex-determination genes. The master sex-determining gene of the medaka *Oryzias latipes* evolved from a duplication of *dmrt1* [13,14]. In another *Oryzias* species, the sex-determinant is a Y-specific copy of the *gsdf* gene [15]. The sex-determining gene in the Patagonian pejerrey *Odontesthes hatcheri* is a duplication of the anti-Müllerian hormone *amh* [16,17]. *Amh* has also been duplicated to become the sex-determining gene in the Northern pike *Esox lucius* [18]. A male-specific allele of its receptor *amhr2* was identified as the sex-determinant in *Takifugu* species [19]. In contrast, the master sex-determination gene *sdv* of salmonids arose from the duplication of an immune-related gene [20].

African cichlids, particularly those that have undergone adaptive radiation in the large East African Lakes, show an extraordinary diversity of sex chromosomes. Studies of cichlids from Lake Tanganyika and Lake Malawi have identified more than twelve different sex-determining systems that have been mapped to at least ten distinct chromosomes, including ZW systems on LG5 and LG7 and XY systems mapping to LG7, LG19 and LG20 [9,21–23]. A complex scenario has been found for various populations of *Astatotilapia burtoni*, including an XYW system mapped to LG13, an XY system on LG18 and an XY system on a fusion comprising LG5 and LG14 [24,25]. In addition, some species from Lake Malawi possess female-specific B chromosomes that carry a female sex determiner, that is epistatically dominant to the LG7 XY system [26,27].

Due to economic interest in producing fast-growing all-male stocks for aquaculture, numerous studies have examined the sex-determining system in the Nile tilapia (*Oreochromis niloticus*) [28]. These studies have revealed a complex sex-determining system, involving both genetic and environmental factors [12,29–31]. Many of these studies have been performed on domesticated or laboratory strains, which may have experienced several rounds of selection, introgressive hybridizations, and/or significant inbreeding [32,33].

Cytogenetic observations of the synaptonemal complex in Nile tilapia revealed that the terminal ends of the largest chromosome (LG3) remained unpaired in males, suggesting this was the sex chromosome [34]. However, the genetic mapping of families showed a sex linkage of markers on a smaller chromosome (LG1) [29,30,35]. In other strains of *O. niloticus* sex was linked to LG23 [36,37]. Genomic studies also showed that sex was associated with either LG1 [31,38–40] or LG23 [40–44] depending on the strain. The *amh* gene, on LG23, has since been identified as the sex-determining gene in a Japanese strain [43]. The Y has two tandem copies, *amhY* [43] and *amhΔY*, which has a premature stop codon due to an insertion [42,43], while the X chromosome has a single copy, here called *amhX*. Knock-out studies showed that *amhY* is necessary for determining maleness [43]. When not a sex determinant, the *amh* still intervenes in gonad development, in germ cell regulation and is even expressed in the brains of some teleost species (reviewed by Pfenning et al. [45]).

It is not clear to what extent these differences in sex-determination of Nile tilapia are due to natural diversity in the mechanisms of sex determination or due to processes of domestication (including introgressive hybridizations whether incidental or not). Domestication of zebrafish (*Danio rerio*) caused the loss of the strong sex determiner on chromosome 4 present in wild fish [46]. Studies of wild Nile tilapia from Lake Kou showed variability in the Y-linked *amh* sequences on LG23 and also suggested that another locus was also involved in sex determination [47]. The objective of the present study was to characterize the sex determiner(s) in wild populations of Nile tilapia that have not been subjected to

any form of domestication or genetic manipulations. We investigated a population from West Africa (Ghana) and another from East Africa (Ethiopia), using two complementary genomic approaches to analyse lab-reared families as well as pools of wild-caught males and females.

2. Materials and Methods

2.1. Fish and DNA Samples

Nile tilapia samples originated from two wild populations caught in Lake Volta (Kpandu, Ghana) in 2002 and 2003 and Lake Koka (Ethiopia) in 2002. The fish from Kpandu belong to the subspecies *Oreochromis niloticus niloticus*, whereas those from Lake Koka are the subspecies *Oreochromis niloticus cancellatus*. They were sexually mature and could be phenotypically sexed by their genital papilla, as described in Bezault et al. (2007) [48]. The collection site near Kpandu is a dendritic expansion on the eastern side of Lake Volta in Ghana (Figure 1) where temperature varies throughout the year between 27 °C and 32 °C. Lake Koka is a reservoir of the Awash River in the Ethiopian Highlands of East Africa, where temperatures vary throughout the year between 21–26 °C and therefore it can be considered a “cold lake” (Figure 1). Wild individuals were transferred to the CIRAD experimental facilities in Montpellier (France), where they were progressively acclimated to a standard temperature of 27 °C $\pm$ 1 °C in a thermo-regulated re-circulating system and individually tagged.

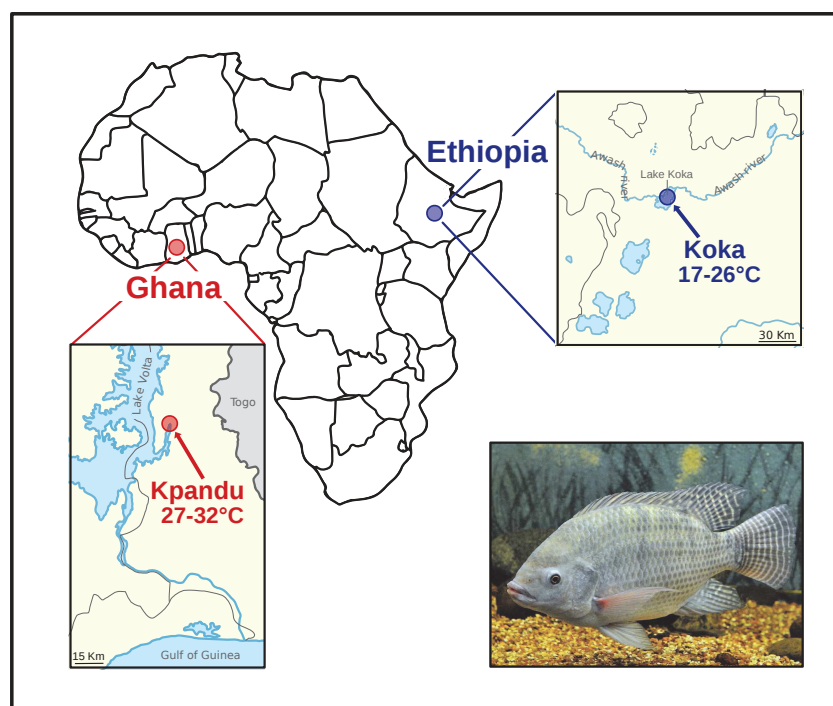


Figure 1. Map of Africa showing Ghana in West Africa with the location of the Kpandu population in Lake Volta (with medium to hot temperatures) and Lake Koka (a cold lake) located in the Ethiopian highlands of the East African Rift.

For the double-digest Restriction Associated DNA (ddRAD) experiment, we produced an F1 family from each population by crossing the wild-caught individuals in 240 L aquaria. The Kpandu Kp20 progeny was obtained by crossing male KpM17 to female KpF32. The Koka Kk-6 progeny was obtained by crossing male KkM39 and female KkF4 (Table 1). Fry were collected from the mother’s mouth at 3 days post-fertilisation (dpf), transferred to McDonald jars for incubation until 9 or 10 dpf, and then reared in 40 L aquaria at 27 $\pm$ 1 °C. At 90 dpf, fish were euthanised with 2-phenoxyethanol (1 mL/L) and 100 randomly sampled individuals per group were sexed by microscopic observation of the gonads [28] to determine the phenotypic sex and define the sex ratios of the progenies. Fin clips

from the parents, and from 14 females and 15 males per family were collected and stored in 100% ethanol until DNA extraction. We deliberately selected families that gave balanced sex-ratios when reared at 27 °C. This minimized the effects of environmental or minor genetic factors that might bias the proportion of males. The Kpandu Kp-20 family showed a sex-ratio of 46% males at 27 °C which was not significantly different from the standard balanced (1:1) sex ratio (Chi-square test NS (Non Significant) $p > 0.05$) (Table 1). The Koka Kk-6 progeny also gave a balanced sex ratio of 55% males NS for $p > 0.05$ at 27 °C.

Fin clips for the Whole Genome Sequencing (WGS) were taken from the wild-caught individuals, consisting of 14 females and 20 males from Koka, and 27 females and 27 males from Kpandu (Table 1). To validate the sex chromosome markers we used fin clips from fish of known sexual genotype from the Manzala-Tihange strain (Belgium) kept at CIRAD. They were an XX female, an XY male and a YY supermale.

Table 1. Fish samples were used in this study.

Approach	Population	Family	Sire	Dam	Sex-Ratio	Sampled Males	Sampled Females	Total Analyzed
ddRAD sequencing	Kpandu	Kp-20	KpM17	KpF32	49% (NS)	15	14	31
	Koka	Kk-6	KkM39	KkF4	55% (NS)	15	14	31
Whole Genome Sequencing	Kpandu					27	27	54
	Koka					20	14	34

Chi-square tests of observed vs. expected (1:1). NS for non-significant.

2.2. Ethics Statement

All animal and experimental procedures were performed at the CIRAD facilities in Montpellier in accordance with the French protocol N°2016101810463 and the author's personal authorization for animal experimentation N° 35–15, both delivered by the French Government. Fish captures followed the laws and veterinary agreements of each country as stipulated in 2003.

2.3. DNA Extraction

Genomic DNA (gDNA) was extracted in 96 well plates from fin clips cut to 0.5 × 0.5 cm size, digested in 300 µL lysis solution containing 0.3 M NaCl, 50 mM Tris-base, 0.2 mM EDTA, 0.2 mM EGTA, 0.356 mM spermidine, 0.256 mM spermine, 4.8% SDS and 10 µg proteinase K at 55 °C overnight. This was followed by an inactivation at 70 °C for 10 min and then by treatment with 6.6 µg RNase for 1 h at 37 °C. After brief centrifugation, the clear supernatant was extracted with the King Fisher Flex robot using the Nucleomag 96 Tissue Kit (Macherey-Nagel, Dueren, Germany) following the commercial protocol. We only changed the buffer quantities, using 200 µL of MBE and MB4 per sample, 300 µL of MB5/sample, and 335 µL of MB2 + 25 µL of magnetic beads (diluted at 1/2 with Ambion water). The gDNA was eluted in 100 µL of MB6 (5 mM Tris) and after the residual magnetic beads were settled on a magnetic rack, a clear eluate was retained. DNA was quantified using the Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) dsDNA BR kit and the quality was checked with a Nanodrop spectrophotometer (ThermoFisher, Waltham, MA, USA) and on a 0.8% agarose gel. Finally, the gDNA was diluted to 50 ng/µL.

2.4. Sexual Genotyping with *Amh* X and Y Chromosome Markers

We genotyped the sex of the wild individuals (11 females and 15 males from Koka and 27 females and 27 males from Kpandu) using PCR amplification of four markers on LG23 shown to be specific for the X and/or the Y chromosome in domestic strains of Nile tilapia. These markers are located within the *amh* genes and adjacent promoter regions. Due to the high sequence homology of the 3 *amh* genes, the markers were designed to cover indels. The procedures were performed as described in Sissao et al. (2019) [47]. The first marker, *amh*X₊₃₆ (primers F1-GTTTGCAATAGTTAGGGTGCTGCTG;

R- GGAAATGCAGCCATTCCTGAG) which amplifies a 1000 base pair (bp) fragment present on the promoter region of the *amh* gene located on the X chromosome [43]. The second marker, *amh* Δ Y₊₅ (primers F2-AAACCTCCTTCCTTTGTGAATGTC; R2-CTAGCGGCATCCACACTCCCTCAC) amplifies a 1500 bp fragment that corresponds to a 5 bp insertion in exon6 that is present in the *amh* Δ Y gene, located on the Y chromosome. This insertion causes a change in the reading frame resulting in a truncated protein [43]. The third marker, *amh* Δ Y₋₂₃₃ (primers F3-CGGTCCCAGTGACCTATGAG; R3-AAGTACACGTGGTGTATTGTAATTGA) gives two fragments; one of 1000 bp corresponding to either *amh*X and/or *amh*Y genes, and another fragment at 767 bp, which corresponds to the *amh* Δ Y gene [42]. The fourth marker, *amh*Y₋₅₆₀₈ (primers F4-GAAAGGGGTGTTTTGGTGCTGGC; R4-ACCCAGGAAGCGTTTCATCTCA) amplifies two fragments: a 2414 bp band that specifically amplifies the *amh*Y gene present on the Y chromosome, and a fragment of 8022 bp which specifically amplifies the *amh*X gene on the X chromosome [43]. We used this marker to detect the presence of the *amh*Y gene in males, because the large *amh*X fragment does not amplify consistently. An XX female, XY male and YY male of the Manzala-Tihange (Belgium) strain were included as positive controls.

2.5. Double Digest RAD Sequencing

2.5.1. Library Construction

The ddRAD library contained 64 individuals from two F1 families obtained from wild-caught crossings (Table 1) and was constructed following a protocol modified from Palaiokostas et al. (2015) [39], using 100 ng of gDNA per individual and a no-DNA control. Briefly, the gDNA was double digested for 3 h at 37 °C with 0.5 μ L of *Sbf*I and 0.5 μ L of *Sph*I (20 U/ μ L each, New England Biolabs (NEB), Ipswich, MA, USA) with 2 μ L of 10X Cutsmart Buffer in a reaction volume of 20 μ L. Restriction enzymes were then heat-inactivated at 65 °C for 15 min. Unique barcodes with specific combinations of P1 and P2 adaptors were used to identify all samples individually. We used 4 μ L/sample of a 1:16 ratio consisting in 120 nM *Sbf*I compatible P1 adaptor/1920 nM *Sph*I compatible P2 adaptor with 0.5 μ L of T4 Ligase (1000 U, 2000 U/ μ L, NEB), 0.3 μ L rATP (100 nM, Promega, Madison, WI, USA), 1 μ L 10X Cutsmart Buffer in a total volume of 6 μ L/sample. Samples were incubated for 2 h at 22 °C. Subsequently, all samples were combined into a single pool, purified with the Nucleospin Gel & PCR clean-up kit (Macherey-Nagel, Dueren, Germany) and then eluted in 40 μ L Elution Buffer (5 mM Tris). Size selection was performed by separating the eluate on a 1.2% agarose gel and excising fragments between 300 and 700 bp. These gel fragments were then purified with the QIAquick gel extraction kit (Qiagen, Dueren, Germany) and eluted in 3 $\times$ 30 μ L Elution Buffer. The library was subsequently PCR amplified in 22 replicates, each containing 4 μ L template, 1.35 μ L of each PCR P1/P2 primers and 12.54 μ L of Q5 Hot start HF 2X Master mix (NEB). The PCR program was: initial denaturation at 98 °C for 40 s, followed by 13 cycles of 10 s at 98 °C, 30 s at 65 °C, 45 s at 72 °C with a final elongation for 2 min at 72 °C. All PCR products were pooled and purified using the MinElute PCR purification kit (Qiagen). They were subsequently cleaned twice with AMPure XP magnetic beads (1 $\times$ vol. followed by 0.7 $\times$ vol.) to eliminate residual fragments under 300 bp and above 700 bp. The final library concentration was quantified with the HS dsDNA Qubit kit, and the quality and size estimated on a Bioanalyzer 2100 using a High Sensitivity DNA Chip (Agilent, Santa Clara, Californie, CA, USA), which confirmed that the majority of fragments were between 300–600 bp. The ddRAD library was sequenced at the GeT-PlaGe platform (Toulouse, Auzerville, France) on one lane of 150 bp paired-end reads on a HiSeq3000 (Illumina, San Diego, CA, USA).

2.5.2. Data Processing

The data is accessible under the BioProject accession number PRJNA657179. The quality of the reads was checked using FastQC (0.11.7) [49]. Raw data were demultiplexed using the STACKS ‘process_radtags.pl’ script for paired-end reads [50]. The *Sbf*I-*Sph*I restriction enzymes and adapter sequences were specified and trimmed. Reads were then mapped to the *O. niloticus* reference genome

O_niloticus_UMD_NMBU [51] using the BWA-MEM mapper (version 0.7.15) [52]. Output files from BWA-MEM (SAM files) were converted to BAM files, and then sorted and indexed with SAMtools (version 1.9) [53]. We then used STACKS (version 2.2) to complete the single nucleotide polymorphism (SNP) calling and population genetic statistics for each population separately [50]. STACKS implements a maximum likelihood statistical model algorithm to identify loci and call genotypes, thus reducing false polymorphic sites (e.g., due to sequencing errors). The STACKS script 'ref_map.pl' combines 'pstacks' to build loci, 'cstacks' to assemble a catalogue of loci, and 'gstacks' to match samples reads against the catalog. Finally, it runs 'populations' which computes the population genetics statistic AMOVA F_{ST} (derived from [54]) and Fisher's Exact test between sexes at each nucleotide position. Relevant options passed to 'populations' were -M specifying a 'population map' with 2 groups, splitting samples according to their phenotypic sex; -smooth; -hwe; -fstats -f p_value -vcf. The output VCF files were filtered using VCF tools (0.1.14) [55] to keep sites with a maximum of 2 alleles (-max-alleles 2), and minimum read depth of 5 (-min-meanDP 5). We removed one phenotypic male of the Kpandu and one phenotypic female of the Koka offspring that had 2098 and 3261 times fewer reads than the mean of all individuals (-remove-indiv) in VCF files. Plain text table outputs were filtered using custom python and bash scripts. Only loci that had reads for at least half of the males and half of the females (per family) were kept.

2.6. Whole-Genome Sequencing

2.6.1. Library Construction

Two libraries per population were produced for the WGS consisting of a pool of wild-caught females and a pool of wild-caught males. The two Kpandu libraries contained a pool of 27 females and a pool of 27 males, whilst the Koka libraries consisted of a pool of 14 females and a pool of 20 males (Table 1). Each pool contained an equal amount of individual gDNA, for a total of 2200 ng in 55 µL. Pooled DNA were sonicated using a Covaris S220 (Woburn, Massachusetts, MA, USA) to obtain 550 bp fragments. Libraries were subsequently prepared using the TruSeq DNA PCR-Free kit (Illumina, San Diego, CA, USA). Each pool was tagged with unique barcodes. Library quality and fragment size range were assessed using a High Sensitivity DNA Chip on the Bioanalyzer (Agilent, Santa Clara, CA, USA). Libraries were quantified by qPCR with the Library Quantification Kit (Takara Bio USA, Inc, Mountain View, CA, USA) on a Lightcycler 96 with the following program: initial denaturation of 2 min at 98 °C; 35 cycles of 10 s at 98 °C, 15 s at 60 °C and 45 s at 68 °C, finishing with a melting curve. The 4 libraries were sequenced on 2 lanes of a HiSeq 4000 (Illumina), using paired-end read lengths of 100 bp at the GenomEast Platform (Strasbourg, Illkirch, France).

2.6.2. Data Processing

The data is accessible under the BioProject accession number PRJNA657179. The quality of the reads was checked using FastQC (version 0.11.7) [49]. Illumina TruSeq Adapters were trimmed using Trimmomatic (version 0.33) [56] for paired-end reads. The reads were mapped onto the XX female *O. niloticus* reference genome O_niloticus_UMD_NMBU using BWA mem mapper (version 0.7.15) [52]. SAM files were then converted to BAM files and sorted using SAMtools (version 1.9) [53]. PCR duplicates were identified using GATK MarkDuplicates (version 2.1.0) [57]. Files were subsequently indexed using Samtools index, and mpileup files were generated using the Samtools mpileup program. The output mpileup files were converted to sync files from Popoolation2 [58] with the 'mpileup2sync.pl' script and used as input for the next step. Fisher's Exact Test were computed for both populations using popoolation2, designed for comparing allele frequencies between pooled sequencing data. F_{ST} was computed using the SEX_SNP_Finder_GA.pl pipeline [38] to identify SNPs at intermediate frequencies in males and fixed (or nearly fixed) in females. This was achieved by looking at both XY and ZW patterns for each population independently. The relevant parameters used were the following: -fixed_threshold = 0.9; -minimum_polymorphic_frequency = 0.3;

–maximum_polymorphic_frequency = 0.7; –minimum_read_depth = 10; –maximum_read_depth = 100; –minimum_read_count = 2; –sex_SNP_finder_window_size = 10000. Structural variants were called using DELLY (0.8.1) [59], which focuses on paired-ends, split-reads, and read-depth to predict and discover genomic rearrangement. Small variants in the whole genome data were then called using FreeBayes (1.2.0) [60], a haplotype-based variant detector designed for small indels and polymorphisms. A filter was set for each pool independently to keep only variants for which the number of occurrences was at least 5% of the overall coverage. All graphs were made using R [61] with the qqman package for the Manhattan plot [62].

3. Results

3.1. Genotyping with *Amh* X and Y Chromosome Markers

The genotypic sex of the parents of the two families, and the other wild-caught fish from the Koka and Kpandu populations, was predicted by using four *amh* markers that have distinguished the X and Y chromosomes in the domesticated Manzala and Swansea strains (Israel) [42], the Japanese strain [43], and the Manzala-Tihange strain (Belgium) [47]. The presence of the Y chromosome in the majority of Kpandu males was proven with the successful amplification of the two *amh*ΔY markers as well as the *amh*Y marker. Indeed we amplified the 1400 bp band of the *amh*ΔY₊₅ marker and the 767 bp fragment of the *amh*ΔY₋₂₃₃ marker in the Kpandu males, like the control Manzala-Tihange males that possess the LG23 Y Chromosome (Figure 2). These results indicate the presence of the *amh*ΔY gene and that these males, including the KPM17 sire, possess a Y chromosome and consequently are XY individuals. We found three Kpandu males without these bands suggesting they were XX males (Table 2; Figure 2 shows one of these males “M23”). All Kpandu females, including the KpF32 dam, were negative for these bands, indicating they were all XX females. None of the Koka males (or females) amplified the two *amh*ΔY gene markers.

The presence of the *amh*Y gene was confirmed using the *amh*Y₋₅₆₀₈ marker which gives a 2414 bp fragment belonging to the *amh*Y and an 8022 bp fragment for *amh*X [43]. We used this marker to verify the presence of *amh*Y, but the 8 kb *amh*X band amplifies poorly. We saw only the X fragment in the Manzala-Tihange XX female used as PCR control (Figure 2). We amplified the *amh*Y fragment in all Kpandu males except the three XX males and did not amplify this fragment in Kpandu females. Weak amplification of the *amh*Y fragment was observed in some Koka males, but not in any Koka females. All of the fish from Koka, Kpandu and the Manzala-Tihange strain (excluding the Manzala-Tihange YY male) amplified the 1000 bp *amh*X₊₃₆ marker and the X-specific 1000 bp band from the *amh*ΔY₋₂₃₃ marker.

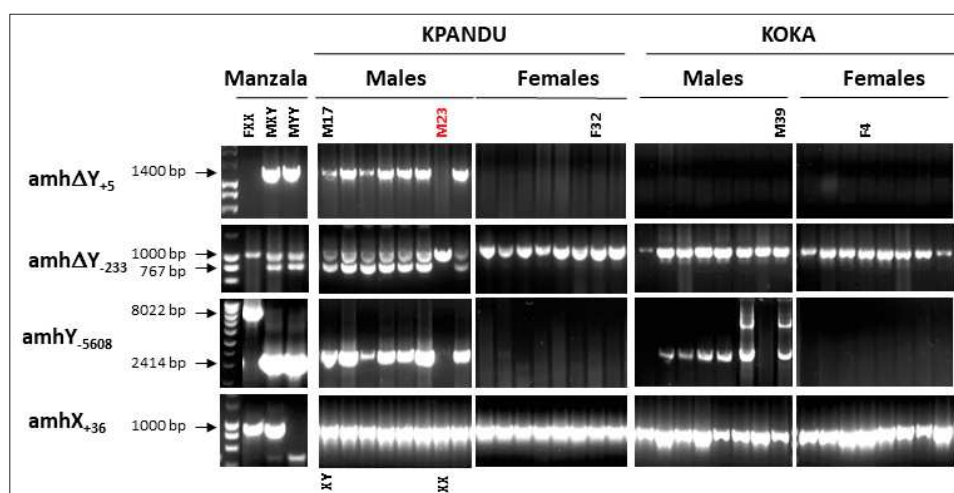


Figure 2. *Amh* genotyping of the Kpandu and Koka breeders used in ddRAD and a subset of individuals used in the whole-genome sequencing (WGS) pools. The complete genotyping of all individuals is shown in Figures S1 and S2.

Table 2. Sex chromosomal LG23 *amh* markers analysed in the Kpandu and Koka wild-caught fish to predict the sexual genotype. The Manzala strain with known sex genotypes was used to validate the amplification.

Population or Strain		Phenotype	N Analyzed	Genotype	<i>amhX</i> Genotype	<i>amhΔY</i> Genotype	<i>amhY</i> Genotype	Predicted Genotype
Domestic strain	Manzala	Females	1	XX	<i>amhX</i> ⁺	<i>amhΔY</i> [−]	<i>amhY</i> [−]	XX
		Males	1	XY	<i>amhX</i> ⁺	<i>amhΔY</i> ⁺	<i>amhY</i> ⁺	XY
			1	YY	<i>amhX</i> [−]	<i>amhΔY</i> ⁺	<i>amhY</i> ⁺	YY
Wild populations	Kpandu	Females	27		<i>amhX</i> ⁺	<i>amhΔY</i> [−]	<i>amhY</i> [−]	XX
		Males	24		<i>amhX</i> ⁺	<i>amhΔY</i> ⁺	<i>amhY</i> ⁺	XY
			3		<i>amhX</i> ⁺	<i>amhΔY</i> [−]	<i>amhY</i> [−]	XX
	Koka	Females	14		<i>amhX</i> ⁺	<i>amhΔY</i> [−]	<i>amhY</i> [−]	XX
		Males	8		<i>amhX</i> ⁺	<i>amhΔY</i> [−]	<i>amhY</i> ⁺	XY
			9		<i>amhX</i> ⁺	<i>amhΔY</i> [−]	ND	ND

3.2. Read Quality and Mapping

3.2.1. ddRAD Sequencing Libraries

We obtained a total of ~221 million clean reads for the Kpandu family after removing ambiguous RAD barcodes, orphaned paired-ends and low-quality reads. From these reads, an average of 94.9% per sample mapped properly paired to the *O. niloticus* reference genome (corresponding to an average of 7.1 million reads per individual). A total of 145,258 loci were built in STACKS, with an average insert length of 336.2 bp, and a mean of 7.6 males and 7.0 females mapped per locus. For the Koka family, a total of ~161 million clean reads were obtained, of which an average of 95.3% per sample mapped properly paired on the *O. niloticus* reference genome (corresponding to an average of 5.2 million reads per individual). This allowed us to detect 129,388 loci with a mean insert length of 329.3 bp. An average of 8.2 males and 7.4 females were mapped per locus. The number of reads per tag and other demultiplexing statistics for each sample are reported in Table S1.

3.2.2. Whole-Genome Sequencing Libraries

We performed a pooled WGS strategy on the wild-caught samples of males and females (Table 1). We obtained ~115 million paired reads for the Koka female pool, ~93 million for Koka males, ~104 million for Kpandu females and ~226 million for Kpandu males. After trimming, we retained between 98 and 99% of the reads from each library. The mapping success onto the *O. niloticus* reference genome was over 94% for each library, with a least 83% correct paired reads. The average coverage obtained for Kpandu males was 39X, which is more than twice the average coverage for the other pools (Kpandu females: 18X, Koka females: 19X, Koka males: 16X).

3.3. Sex-Associated Loci Identified by ddRAD

3.3.1. Sex-Associated Signal is Observed on LG23 and LG3

For the Kpandu family, we identified 14,073 polymorphic sites in phenotypic males and females located in 6437 RAD loci. For the Koka family, we identified 10,314 polymorphic sites in phenotypic males and females located in 4890 RAD loci. In both families, we found a significant peak of differentiation between phenotypic males and females on LG23, with 29 SNPs associated with sex for Koka and 125 SNPs for Kpandu (Fisher's Exact Test p -value < 0.01) (Figure 3A and Table S2). The Manhattan plots of the F_{ST} values are similar for the two families. For the Kpandu family, 220 sites had a higher F_{ST} than the mean F_{ST} of 0.028. For the Koka family, there were 151 sites above the mean F_{ST} of 0.024.

The region of LG23 between 24 Mb and 40 Mb contains the majority of the signal for both families (Figure 3B). The Kpandu family showed a strong sex association of markers in this region. This region contained 62% of the significant sites, although we can observe significant sex-associated SNPs all along this linkage group. In the Koka family, sex was even more strongly and specifically associated with this region, which included 97% of the significantly differentiated sites. The overall lower number of significantly differentiated sites in the Koka family might be due to the lower level of sequence coverage (Table S1).

In the Kpandu family, we also obtained a sex-associated signal on LG3 (Figure 3C and Table S2). The highest signals on LG3 were located at 20Mb according to Fisher's exact test. It corresponds to two RAD loci, one localised in a non-coding region and the other is in the "Fc fragment of IgG receptor IIb" gene (*fcgr2b*) (Table S3 and Figure 3C). More than 40 other significant sex-patterned SNPs on LG3 (Fisher's Exact Test (FET) p -value < 0.001) occur in a region spanning ~40 Mb between 40 to 80 Mb (Table S3 and Figure 3C).

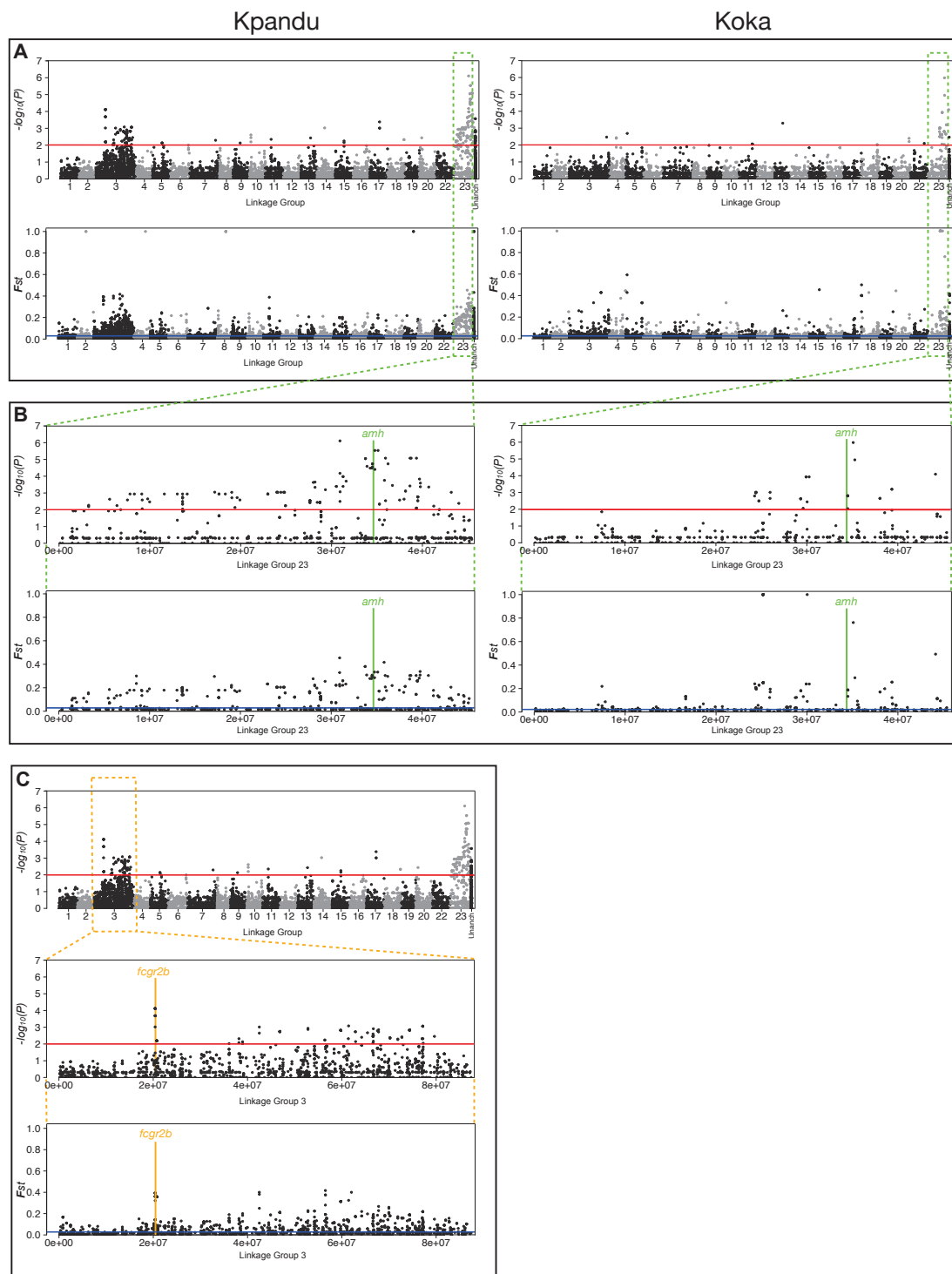


Figure 3. Manhattan plots of log₁₀-transformed *p*-values from Fisher's Exact Test (FET) and AMOVA *F_{ST}* of the association of single nucleotide polymorphisms (SNPs) with phenotypic sex. (A) Plots for the Kpandu and the Koka families for all LGs. Unanchored contigs/scaffolds are grouped under the label "Unanch". (B) Plots for the LG23 region showing SNP sites that were associated with sex. Most of the stronger sex-signals are found between 24 and 40 Mb. (C) Plots for Kpandu family with a close up of the LG3 region showing a very localised signal at 20 Mb due to a RAD locus in the *fcgr2b* gene. The rest of the signal observed on LG3 comes from the "long arm" of the chromosome known to carry repeated elements. Red lines correspond to *p*-values of 0.01 in the FET plots while blue lines correspond to mean AMOVA *F_{ST}* between males and females of each family.

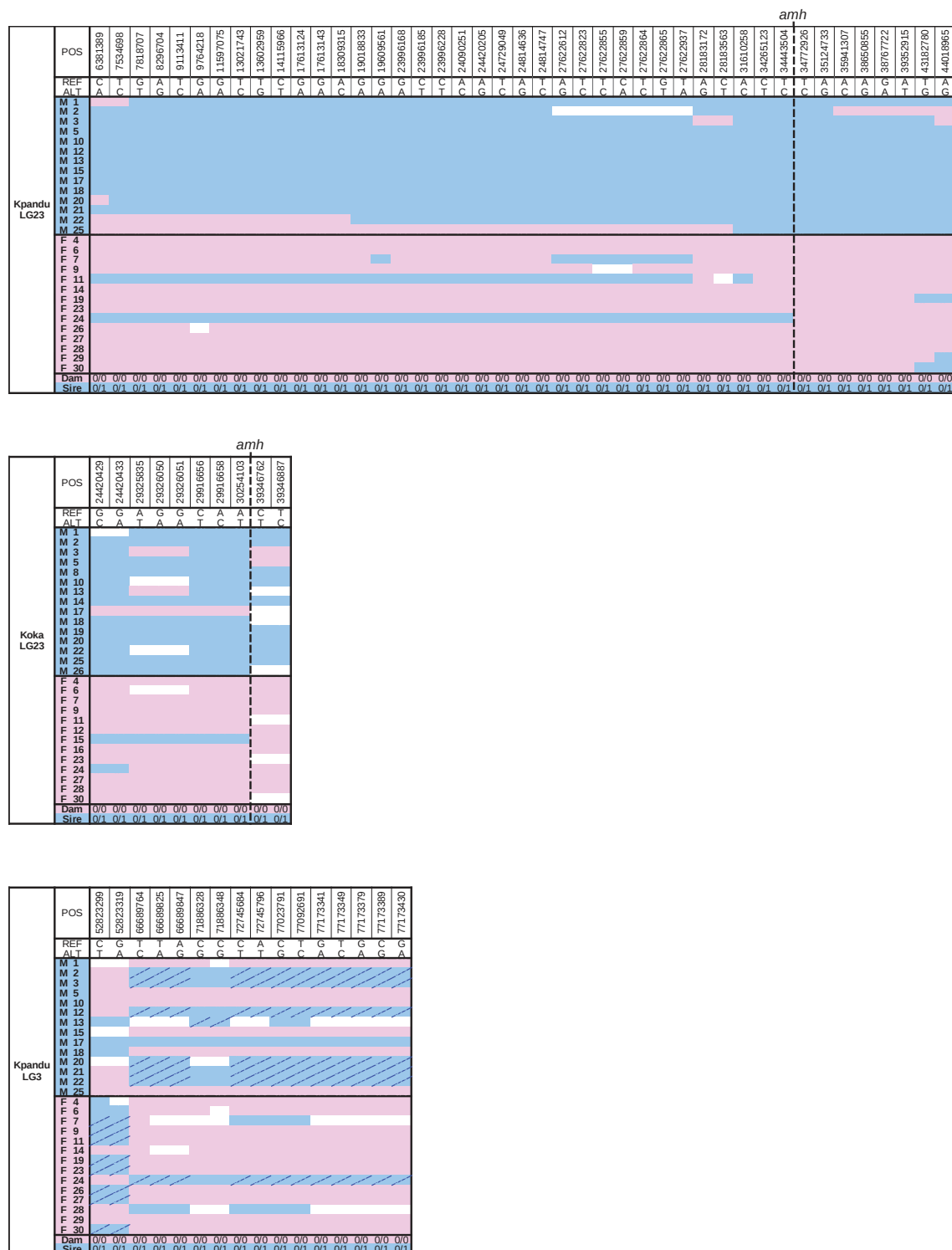
3.3.2. Inherited Patterns of Offspring on LG23 Confirm a Male Heterogametic System

One advantage of using families for the ddRAD sequencing is that we can compare the SNPs of father and sons, and likewise mother and daughters, to determine whether the pattern of inheritance on LG23 corresponds to an XY or a ZW sex-determining system. We, therefore, focussed on sites for which both parents were sequenced. We found that most of the LG23 sites with a significant FET were consistent with an XX/XY pattern of inheritance (Figure 4).

In the Kpandu family, there was an increase in the concordance between genotype and phenotypic sex as we approached the *amh* gene on LG23 (Figure 4). The *amh* gene, which remains the strongest candidate for sex determination [43], is centred at 34.501 Mb. At the closest flanking SNPs (34,772,926 bp and 35,124,733 bp) all the offspring have genotypes that match their phenotypic sex, following an XY pattern of inheritance from their parents. The 7 SNP sites between 34,265,123 and 39,352,915 bp show a maximum of 3.6% (1/28) discordance between an XY genotype and a male phenotype in the offspring (Figure 4). Only one female shows a heterozygous genotype, at 34,443,504 bp, but she is also heterozygous for all previous sites on this chromosome. This pattern suggests a recombination breakpoint very close to the *amh* gene in this individual. At the 4 SNP sites after the *amh* gene (at 35,941,307, 38,650,855, 38,767,722 and 39,352,915 bp) there is a perfect match between genotype and phenotype, with the exception of one male presenting an XX genotype (Figure 4). Here again, we suspect a recombination breakpoint just after the *amh* gene in this individual.

The heterozygous signals are less obvious in the Koka fish due to the overall lower coverage and lower number of sites that passed the filters to build the genotypic table of parents and offspring. Nevertheless, there is a good correspondence between the genotypes of markers near *amh* and the sex of the offspring. The data are again consistent with an XX/XY pattern of inheritance (Figure 4). The genotypes immediately adjacent to the *amh* gene on LG23 follow an XX/XY pattern and are sufficient to predict the phenotypic sex of all but two individuals of the Kpandu and Koka families.

Although some markers on LG3 were statistically associated with phenotypic sex in the Kpandu family, the genotypes of individual markers were not sufficient to accurately predict the sex of every individual (Figure 4). Breakpoints between sites 52,823,319 and 66,689,764 may correspond to a region of increased recombination or a genomic rearrangement so that these markers are further apart than the reference genome suggests (Figure 4). However, the fact that females can be heterozygous or homozygous on either side of these breakpoints, and the same for males, indicates that this interval on LG3 probably does not contain a sex determiner. We also tested the Kpandu and Koka SNPs for a female ZW pattern but failed to detect a female heterozygous pattern (Figure S3). The statistical signal on LG3 may be an artifact arising from the highly repetitive sequences on this chromosome.



3.4. Whole-Genome Sequencing Analysis

3.4.1. B Chromosome Presence Creates Noise in the WGS Data

In the WGS data, every linkage group had many SNPs with unexpectedly high F_{ST} and significant Fisher's Exact Test p -values between phenotypic males and females (Figure 5A). Taking a closer look at each of these high F_{ST} loci independently, we observed that this surprising signal was found in both populations and at the same positions, but it is clearer in the Koka population (Figure 5B). These high F_{ST} peaks correspond to regions with particularly high coverage and are typically observed in exons (Figure 5C). Similar high coverage polymorphic regions scattered in the genome have been attributed to the presence of highly repetitive B chromosomes in other African cichlids [26,63,64]. Thus, the presence of B chromosomes in our samples might explain the highly polymorphic regions of high sequence coverage observed in our study. At the species level, B chromosomes were not sex-specific since they were observed only in the Koka female pool and the Kpandu male pool. Consequently, the noise in F_{ST} masks the polymorphism linked to the genetic sex-determinant.

3.4.2. Male-Specific Structural Variants in the Amh Region of LG23

The whole-genome data allowed us to study in more depth the genome structure around the *amh* region in wild Kpandu and Koka males. The most striking structure we detected is a tandem duplication of 21.5kb located on the LG23 Y chromosome (Figure 6A) comparable to what Li et al. (2015) [43] found in a domesticated strain of Nile tilapia. Here we have found that the breakpoints on the X chromosome that led to the Y chromosome duplication are between two genes that flank *amh*. The ornithine decarboxylase antizyme 1 gene (*oaz1*) is found upstream of *amh* and the disruptor of telomeric silencing 1-like gene (*dot1l*), also known as dot1-like histone lysine methyltransferase, is found downstream (Figure 6, Table 3). Both genes are transcribed on the same strand as the *amh* gene (antisense on the reference genome). The breakpoints we identified differ by only 4 bp between the Koka and Kpandu males. In the Koka males, the duplication spans the region from 34,491,225 bp and 34,512,737 bp, and in the Kpandu males spans the region from 34,491,225 bp to 34,512,741 bp on the XX female reference genome (Table 3). The confidence interval (CI) around the start and the end of the region that was duplicated is only 50 bp (Table 3) which allows us to say that the beginning of the duplication occurs after the third exon of the *oaz1* gene and ends before the last exon of *dot1l* (Figure 6A). The breakpoints take place within introns, and consequently, the duplication does not include the whole coding sequence for either of the two boundary genes (Figure 6B). Neither *oaz1* nor *dot1l* gene is fully duplicated on the Y haplotype. Instead, it creates a "chimera" comprised of a part of *dot1l* and *oaz1* genes, which occurs between the two copies of the *amh* gene on the Y chromosome. The insertion of this duplication on the Y chromosome could have taken place either upstream of the original Y structure within the *oaz1* gene or downstream within the *dot1l* gene (Figure 6B). It is important to note that Li et al. (2015) [43] positioned the *amh* Δ Y gene upstream of the *amh*Y gene according to a sense strand. We, on the other hand, have defined the Y structure based on the new reference genome assembly where the reads in this region correspond to the anti-sense strand (Figure 6).

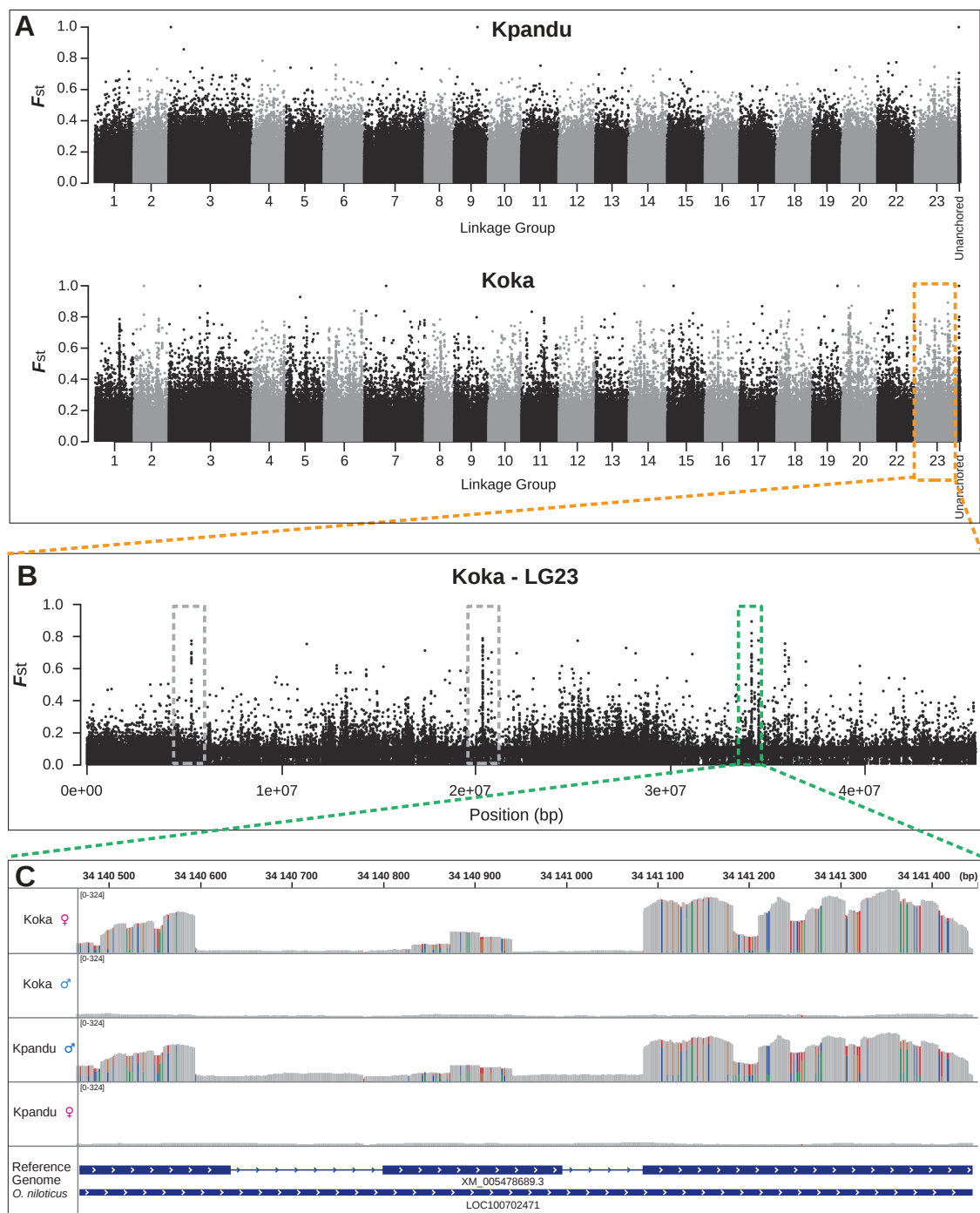


Figure 5. SNPs associated with phenotypic sex obtained from the WGS. **(A)** Manhattan plots of F_{ST} for Kpandu and Koka populations obtained with Sex SNP Finder. High F_{ST} peaks are observed on every LG meaning there is a signal associated to sex on all chromosomes; **(B)** Close up of F_{ST} on LG23 for Koka showing three high F_{ST} regions in grey and green boxes; **(C)** Integrative Genomics Viewer (IGV) plot of sequence coverage at ~34.14 Mb on LG23 corresponding to the green box above for both sexes of the two wild populations. High F_{ST} values in this region are due to a very high sequence coverage (>10 times the mean coverage) likely corresponding to repetitive sequences on the B chromosome. SNPs are represented by colour traits with a bicolor trait being a polymorphic site, with the size of the colour band proportional to the nucleotide frequency. Frequencies of SNPs are conserved between Kpandu males and Koka females.

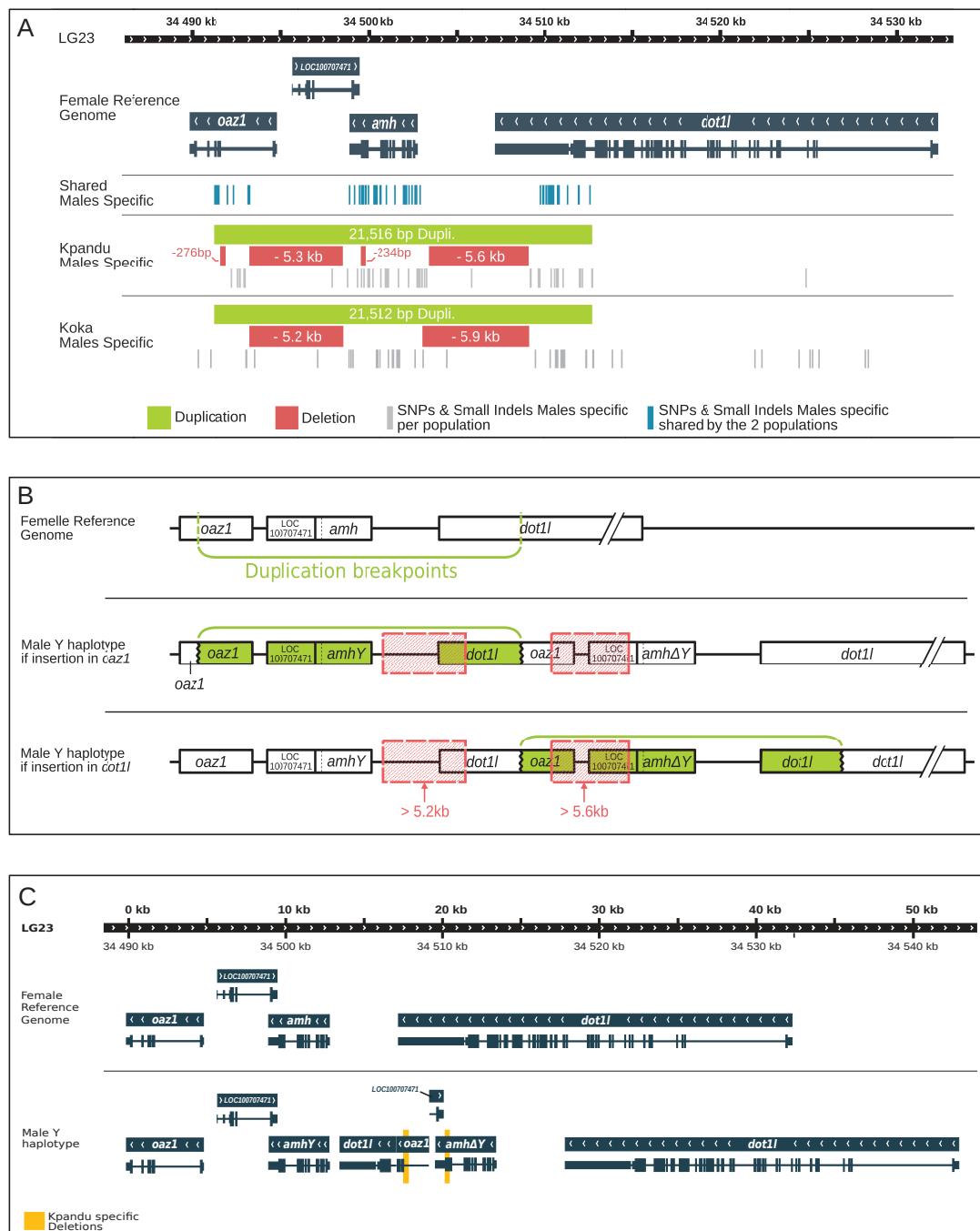


Figure 6. Schematic representation of the duplicated region of LG23 and its structural variants and polymorphic sites found with DELLY and FreeBayes. The tandem duplication includes the *amh* gene and part of the *oaz1* and *dot1l* genes. The duplicated regions in each population (in green) are shown with the corresponding deletions and suggested size (in red). The shared male SNPs are shown in blue bars while the population male-specific SNPs are shown in light grey bars. (A) Schematic view of mapping results of the Y chromosome by WGS of the Kpandu and Koka males against the female XX reference genome. (B) Schematic representation of the two possible places of insertion of the tandem duplication in *oaz1* and *dot1l* genes. Both large deletions found with DELLY are likely to be placed in the *oaz1-dot1l* chimera as the genes are already truncated on the Y haplotype. (C) Schematic representation of the inferred Y haplotype of Kpandu and Koka males.

We also detected two large deletions of more than 5 kb that were common in males from both populations (Table 3 and Figure 6A,B). The first deletion was 5273 bp in length in Koka males and

5,233bp in length in Kpandu males ($CI \pm 390$ bp and ± 150 bp, respectively) (Table 3). This male-specific variant encompasses a region between the first intron of *oaz1* and the *amh* gene according to the female XX reference genome (from 34,493,315 bp to 34,498,588/34,498,548 bp) (Figure 6). In the Y chromosome, the location of this large deletion depends on where the insertion of the duplicated sequences occurred. Either it starts in the complete copy of *oaz1* and ends close to the 5'UTR region of the *amhY* copy, or it starts in the truncated copy of *oaz1* and goes nearly up to the 5'UTR region of *amhY* (Figure 6A). The first hypothesis implies that there would not be a complete *oaz1* copy on the Y chromosome. Thus, the second option where the deletion comprises an already truncated (and putatively non-functional) gene seems more plausible (Figure 6B).

Table 3. Structural variants located in the *oaz1-dot1l* region of LG23 in wild populations of Nile tilapia found with DELLY on whole-genome sequencing data.

Populations	Start Position	End Position	Confidence Interval around Start & End	Type	Size of Structural Variant	Shared between Populations
Koka	34491225	34512737	−50,50	Duplication	21512	Yes
	34493315	34498588	−390,390	Deletion	5273	Yes
	34503117	34509103	−852,852	Deletion	5986	Yes
Kpandu	34491225	34512741	−50,50	Duplication	21516	Yes
	34491681	34491957	−2,2	Deletion	276	No
	34493315	34498548	−150,150	Deletion	5233	Yes
	34499596	34499830	−3,3	Deletion	234	No
	34503494	34509103	−2,2	Deletion	5609	Yes

The second common male deletion is 5,986 bp ($CI \pm 852$ bp) in Koka, and 5,609 bp in Kpandu ($CI \pm 2$ bp). It starts around 34,503,000 bp and ends at the same position (34,509,103 bp) in both populations (Figure 6A). This corresponds to the region downstream of the *amh* gene and upstream the last exon of the *dot1l* gene in the X reference genome (Figure 6A). Here again, there are two hypotheses for the position of this deletion on the Y chromosome. It could be located immediately after the complete copy of the *amh* gene and overlap the truncated *dot1l* copy, or it could be placed after the *amhY* and overlap the complete copy of *dot1l* gene. For the same reasons as previously stated, the first option where the deletion comprises an already truncated gene seems more likely (Figure 6B).

From these data, we were able to describe the SD region of the Y haplotype compared to the X chromosome of the reference genome. After accounting for duplication and deletions sizes, the overall length of this region in males (starting from the first copy of *oaz1* to the last copy of *dot1l* genes) was estimated to be ~52 Kb whereas it is ~42 Kb in length in the female reference genome (Figure 6C).

3.4.3. SNPs and Small Indels in the Amh Region in Males

A total of 60 SNPs and indels were found between males and females within the region encompassing *oaz1* to *dot1l* when mapping on the female reference genome (Table 4). However, all those SNPs and indels were described based on the female reference genome and using Illumina short reads. We are not able to confirm in which copies of genes (truncated or complete) these variants take place within the male's duplication. A multiple nucleotide polymorphism (mnp) in the 3'UTR region, and 1 bp insertion in an intron within the *amh* gene, were observed in males of both populations. Five deletions are also shared in males from both wild populations in this region, which all occur in the 3'UTR part of the *dot1l* gene. A mnp is also detected in an intron of the *dot1l* gene. All other 52 shared polymorphisms are SNPs (Table 4), amongst which 9 are non-synonymous mutations: 2 in *oaz1*, 6 in *amh*, and 1 in the *dot1l* gene. Among the 52 SNPs found, the 3 in the *amh* gene and the 1 in *dot1l* are shared between the males of our 2 wild populations and the commercial stocks used in Cáceres et al. (2019) [44]. Among the 3 SNPs in *amh*, there is a T/C in exon 3, which corresponds to an A/G mutation in the transcript, changing the amino acid from threonine into alanine. The second SNP is a T/C occurring in exon 6, which corresponds to an A/G in the transcript, changing the amino acid from an asparagine into a serine.

Table 4. SNPs and small indels found in the sex-determining *oaz1-dot1l* region on the Y haplotype between males and females of both populations.

Position	Reference	Alternative	Length	Type	Gene	Intron/Exon	Change of Amino Acid	Found in Literature
34491252	C	G	1	snp	<i>oaz1</i>	Exon		
34491407	A	C	1	snp	<i>oaz1</i>	Exon	Leu → Gly	
34491477	C	T	1	snp	<i>oaz1</i>	Intron		
34491574	T	C	1	snp	<i>oaz1</i>	Exon	Lys → Glu	Cáceres et al., 2019 [44]
34492052	A	G	1	snp	<i>oaz1</i>	Intron		
34492058	T	C	1	snp	<i>oaz1</i>	Intron		
34492381	A	G	1	snp	<i>oaz1</i>	Intron		
34493190	G	C	1	snp	<i>oaz1</i>	Intron		
34493277	A	G	1	snp	<i>oaz1</i>	Intron		
34498917	A	G	1	snp	LOC100707471	Intron		
34499201	CTAT	TTAC	4	complex	<i>amh</i>	3' UTR		
34499475	A	G	1	snp	<i>amh</i>	3' UTR		
34499597	C	T	1	snp	<i>amh</i>	Exon 7		
34499645	A	C	1	snp	<i>amh</i>	Exon 7		
34499675	A	G	1	snp	<i>amh</i>	Exon 7		
34499706	G	C	1	snp	<i>amh</i>	Exon 7	Thr → Ser	
34499810	A	G	1	snp	<i>amh</i>	Exon 7		
34499839	C	T	1	snp	<i>amh</i>	Exon 7	Ala → Thr	
34499987	C	T	1	snp	<i>amh</i>	Intron		
34499994	A	C	1	snp	<i>amh</i>	Intron		
34500307	C	T	1	snp	<i>amh</i>	Intron		
34500348	T	C	1	snp	<i>amh</i>	Intron		
34500364	G	C	1	snp	<i>amh</i>	Intron		
34500464	G	A	1	snp	<i>amh</i>	Intron		
34500668	C	A	1	snp	<i>amh</i>	Intron		
34500750	G	A	1	snp	<i>amh</i>	Intron		
34500773	C	T	1	snp	<i>amh</i>	Exon 6	Glu → Lys	
34501082	T	C	1	snp	<i>amh</i>	Exon 6	Thr → Ala	Cáceres et al., 2019 [44]
34501555	C	T	1	snp	<i>amh</i>	Intron		
34502034	T	C	1	snp	<i>amh</i>	Exon 3	Asn → Ser	Cáceres et al., 2019 [44]; Li et al., 2015 [43]
34502075	T	A	1	snp	<i>amh</i>	Exon 3		Li et al., 2015 [43]
34502169	ATG	AGTG	1	insertion	<i>amh</i>	Intron		
34502196	G	A	1	snp	<i>amh</i>	Intron		
34502353	A	C	1	snp	<i>amh</i>	Exon 2	Asp → Glu	Li et al., 2015 [43]
34502501	T	C	1	snp	<i>amh</i>	Intron		
34502686	T	C	1	snp	<i>amh</i>	5' UTR		
34502748	T	C	1	snp	<i>amh</i>	5' UTR		Cáceres et al., 2019 [44]
34502756	C	T	1	snp	<i>amh</i>	5' UTR		Cáceres et al., 2019 [44]

Table 4. Cont.

Position	Reference	Alternative	Length	Type	Gene	Intron/Exon	Change of Amino Acid	Found in Literature
34502954	G	A	1	snp	Non Coding DNA			
34509735	T	C	1	snp	<i>dot1l</i>	3' UTR		
34509898	T	A	1	snp	<i>dot1l</i>	3' UTR		
34509976	T	C	1	snp	<i>dot1l</i>	3' UTR		
34510100	G	T	1	snp	<i>dot1l</i>	3' UTR		
34510131	C	G	1	snp	<i>dot1l</i>	3' UTR		
34510239	TTTAACT	TT	5	deletion	<i>dot1l</i>	3' UTR		
34510250	A	C	1	snp	<i>dot1l</i>	3' UTR		
34510260	A	G	1	snp	<i>dot1l</i>	3' UTR		
34510309	A	G	1	snp	<i>dot1l</i>	3' UTR		
34510392	A	G	1	snp	<i>dot1l</i>	3' UTR		
34510428	GACA	GA	2	deletion	<i>dot1l</i>	3' UTR		
34510495	T	G	1	snp	<i>dot1l</i>	3' UTR		
34510529	C	T	1	snp	<i>dot1l</i>	3' UTR		Cáceres et al., 2019 [44]
34510747	G	C	1	snp	<i>dot1l</i>	3' UTR		
34510768	TGTGTGCG	TG	6	deletion	<i>dot1l</i>	3' UTR		
34510842	CTTTTTTTTTTTTTC	CTTTTTTTTTC, CTTTTTTTTC	5,6	deletion	<i>dot1l</i>	3' UTR		
34511324	TAAATG	TAATG	1	deletion	<i>dot1l</i>	3' UTR		
34511937	T	C	1	snp	<i>dot1l</i>	Exon		
34511950	A	T	1	snp	<i>dot1l</i>	Exon	Leu → His	
34511973	T	C	1	snp	<i>dot1l</i>	Exon		
34512585	ACAA	GCAG	4	complex	<i>dot1l</i>	Intron		

3.4.4. Population-Specific Sex Variants

We found other small deletions that were specific to the Kpandu population (Table 3). One of these deletions is 276 bp in length ($CI \pm 2$ bp) and located in the first intron of the *oaz1* gene towards the limit of the second exon. There are two possibilities for the exact positioning of this deletion since it could be located either in the intact *oaz1* gene or in the truncated copy (Figure 6A,B). A second deletion of 234 bp in length ($CI \pm 3$ bp) is located in exon 7 of the *amh* gene. It is placed in an exon, although it does not induce a frameshift. This means that at least one of the two copies of the *amh* located in the duplication is heavily modified (Figure 6C and Table 3) but from our short Illumina reads we are not able to say which (*amhY* or *amhΔY*). Nevertheless, this 234 bp deletion confirms our PCR results when using the *amhΔY*₋₂₃₃ marker (Figure 2) indicating that the *amhΔY* gene is present in nearly all Kpandu males.

By using FreeBayes to look for small indels and SNPs, we observed 41 nucleotide polymorphisms and 6 small indels that are specific to Kpandu males (Table S4). All 6 non-synonymous mutations detected were within the *amh* gene, whereas indels were found in every gene in the *oaz1-dot11* region, according to the female reference genome. In the Kpandu males, we found the 5bp insertion of exon 6 previously described by Eshel et al. (2014) [42] and Li et al., (2015) [43] which considered specific to the *amhΔY* gene. Here again, the sequencing data confirms our PCR results obtained with the *amhΔY*₊₅ marker (Figure 2). Other male-specific indels present in the Kpandu population are found in introns or the 3'UTR regions.

Koka males have 43 specific SNPs, 1 deletion and 1 insertion both in introns of the *amh* gene (Table S5). These include two non-synonymous mutations. A change of C/A at position 34,501,472 (corresponding to a G/T in exon 4 of the *amh* transcript), induces a change from arginine to serine. The second mutation is a C/A at position 34,521,953 is in an exon of the *dot11* gene and causes a change from glycine to valine. The male-specific variants of Kpandu and Koka are distributed mainly over the same region as the shared males' SNPs (Figure 6A). There are still some sites presenting male-specific variants that overlap with the deleted regions which means that they must appear in the intact copies of the genes. Finally, we found many more SNPs in Koka males than in Kpandu males for the *dot11* region especially outside the duplication (Figure 6A, Table S5).

The WGS confirmed our PCR results (Figure 2) that Koka males do not have the 234 bp deletion found in Kpandu males, and which was reported to be specific to the *amhΔY* copy in the Japanese strain [43]. Likewise, Koka males did not have the 5 bp insertion also considered to be *amhΔY* specific according to the analyses of the Japanese strain [43], confirming our negative PCR amplifications (Figure 2).

4. Discussion

In the Nile tilapia, markers on LGs 1, 3, 20 and 23 have all been found associated with phenotypic sex in various domesticated strains [29,30,34–36,39,40,44,65]. To overcome possible artefacts due to the process of domestication, we have searched for the genetic basis of sex determination (SD) in wild populations of Nile tilapia and studied whether there are population differences. We chose two populations of Nile tilapia from geographically very distant locations, one from Kpandu (Lake Volta) in Ghana (West Africa) and the other from Lake Koka in the highlands of Ethiopia (East Africa). We opted to use a combination of two genomic approaches: double digest Restriction Associated DNA (ddRAD) sequencing on families with balanced sex-ratios and pooled Whole-Genome Sequencing (WGS) of wild-caught males and females. We found strong sex-associations with LG23 in both the Kpandu and Koka populations and a Kpandu-specific signal on LG3. We did not observe any association of sex with markers on LG1 or LG20.

4.1. Sex-Linked Signals Found on LG3

Our F1 ddRAD results showed a weak signal on LG3 that was found exclusively in Kpandu fish. LG3 is the largest chromosome pair of the *O. niloticus* karyotype [30]. It was presumed to be the sex

chromosome after studies of chromosome pairing in the synaptonemal complex showed unpaired terminal ends of the largest chromosome in XY males [34]. Subsequent studies showed stronger linkage to LG1 [31,35,38,39] or LG23 [40,42–44]. In the sister species *O. aureus*, LG3 carries a female ZW determiner [30,66,67]. Several sex-patterned variants were found along a ~50 Mbp region in the long arm of LG3 in *O. aureus* when mapped onto the *O. niloticus* genome [67]. There is a very strong signal in the Kpandu family at 2 RAD loci located around 20Mb. One is located in a region of non-coding DNA. The other is found in the *fcgr2b* gene which codes for the receptor for Fc low-affinity II-b gamma immunoglobulin. Curiously, *sdv* which arises from a duplication of the immune system gene *irf9*, has become the master sex determinant in salmonid species [20]. The immune system showed higher sexual dimorphism for the *fcgr2a* gene, another gene of the “low-affinity immunoglobulin gamma Fc region receptor II” group in cichlids, a difference that was significantly greater in maternal mouth-brooders compared to biparental mouth-brooders [68]. The Nile tilapia is a maternal mouth-brooder, and we speculate that the difference in expression in *fcgr2b* might arise from sexual selection rather than being implicated in sex-determination.

The other sex-patterned SNPs on LG3 were found all along the “long-arm” of LG3. This linkage group is considered to have all the characteristics of an old sex chromosome in Nile tilapia, with signs of differentiation due to its large size and extensive heterochromatic regions with the accumulation of many repetitive sequences that may prevent recombination [51]. The SNPs we observe may result from the history of this pair of chromosomes and the recent accumulation of polymorphisms that do not follow sex-related patterns. This could explain why we did not find a definitive pattern of an XY or a ZW system for the LG3 SNP sites, even though a large number of highly differentiated SNPs were observed. Due to the high amount of repeated sequences in LG3 (54.7% repetitive vs. 37% genome-wide) only ~87 Mb of the chromosome could be assembled instead of the estimated ~130 Mb, and there might be incorrect anchoring to the *O. niloticus* genome [51]. We found on LG3 a very high proportion of immune-related genes and they were present in multiple copies. For example, 96 genes of the “Fc receptor” family are found on this linkage group, and among them 35 are “Low-affinity immunoglobulin gamma Fc region receptor II”. Consequently, the sex differences we observed may also result from mapping errors of reads and a lower quality of the assembly of this chromosome amplified by the random sampling of the ddRAD method.

4.2. Strong XY Patterns on LG23

Our family ddRAD results reveal that LG23 is an XY sex chromosome in both the Kpandu and Koka populations. LG23 has already been reported to be the sex chromosome in some domesticated Nile tilapia [41,43,44,65]. Among the significant sex-variants between 24 and 40 Mb on LG23, one is common to both populations: the anti-Müllerian hormone gene (*amh*). The *amh* gene has a primary role in the sex differentiation pathway of vertebrates including fish [69] and is implicated in germ cell proliferation in teleost fish (reviewed by Mullen et al. [70]). *Amhy* also appears to be the sex-determiner in several domesticated strains of Nile tilapia [40–44,65]. A recent study of a wild Lake Kou population (Burkina Faso) of Nile tilapia also showed that sex was linked to LG23 but the *amh* gene alone was not sufficient to explain all the sex-ratios observed [47].

We also looked at the other sex-significant sites found on LG23 and found a male-specific SNP for the *lingo3* gene in Kpandu fish. Variants have also been identified in *lingo3* in other Nile tilapia genomic data when searching for sex-determining regions [41,44]. The close linkage of the *lingo3* gene to the *amh* region probably explains the signal found in those previous studies and could be enough to explain the signal we found in the Kpandu individuals. In our two wild populations, we did not find any sex-signals on LG1 nor on LG20 which have both been found in the Manzala-Stirling strain [31,38,39].

Genotyping at the region with the most significant sex-pattern variants on LG23 was enough to perfectly predict the phenotypic sex of individuals in both populations. The Kpandu population belongs to the widespread Sudan-Sahelian cluster that separated from the Nile basin through multi paleo-geographic events [71]. The Koka population belonging to the Ethiopian Rift Valley showed

large genetic divergence from the Sudan-Sahelian but also from the Nile basin from which it was isolated 12 to 8 thousand years ago [71]. The LG23 sex chromosome seems ancestral in *O. niloticus* since it appeared in the two wild populations studied here. This observation however contrasts with the apparent rapid evolution of the sex-determination in domesticated strains since at least three different LGs have been reported to be sex chromosomes. However, wild populations might segregate multiple sex determiners among which LG23 Y would be the most frequent and dominant. This may also explain why we found only LG23 in both the Koka and Kpandu families we selected. In wild Kou fish, for instance, sex was mainly associated with LG23 but it did not explain the phenotypic sex in 18% of the individuals, suggesting segregation of another sex-determiner [47]. This has already been shown for the Manzala-Stirling strain in which sex is strongly associated with LG1 and LG20 in some families [31,38] but it is also linked to LG23 in low frequency (Penman, unpublished data). Founding effects or domestication events have led to sex being determined exclusively by LG23 in the Manzala-Göttingen [65] and the Manzala-Tihange strains ([47], and this study) although they both originated from Manzala-Stirling breeders. Even stronger differences have been observed in the case of zebrafish where laboratory strains have lost the sex-chromosome found in wild populations [46]. Our study further highlights the importance of characterising the sex determiner in wild fish populations.

4.3. WGS Uncovers a Putative B Chromosome in *O. niloticus*

Blocks of high sequence coverage have been described previously in cichlid genomic data when B chromosomes were present, such as in *Astatotilapia latifasciata* [63,72], *Metriaclicma lombardoi* from Lake Malawi [26], and the South American cichlid *Crenicichla lepidota* [73]. In these studies, high coverage regions were due to A chromosome fragments copied onto a B chromosome, where they are highly amplified and highly polymorphic. Because the reference assembly does not contain a B chromosome, the whole genome shotgun reads from all the different copies are mapped at the same position on the reference genome. The presence of these “B blocks” in our data, therefore, suggests the presence of a B chromosome in these two wild populations of Nile tilapia. In our study, the high coverage blocks mostly overlapped exons, which is a major difference from the B chromosomes previously described in other cichlids. This could be because the B chromosome is ancient, and the sequences corresponding to introns have diverged too much to map correctly onto the reference genome, unlike the exons that would be more conserved over time. Another hypothesis would be that genes accumulate on the B chromosome via an RNA intermediate, for instance through retrotransposition of mRNA, so the introns are not copied. Because the B blocks were found in Kpandu males and Koka females, they created strong differentiation between the sexes in each population, but these signals were independent of the true genetic sex-determining locus. It appears unlikely that the putative B chromosomes play a role as a female W sex-determiner such as in some Lake Malawi cichlids [27]. However, we are not able to determine the frequency of this B chromosome in our wild populations since we pooled the individual DNAs prior to sequencing. Even if the frequency is equal between sexes, if it is low in the wild, then a random sampling of individuals in our study could make it look like the B chromosome is sex-associated. B chromosomes have never before been described in Nile tilapia. Further studies will be needed to thoroughly characterize them in these populations. If B chromosomes are present in any of the domesticated strains, this can significantly impact genetic and genomic studies, from de novo reference assembly to identify sex-determining regions, as seen in this study. As such it would be beneficial to incorporate cytogenetic analysis as a control into the experimental design process.

4.4. Confirmation of the Tandem Duplication of the *Amh* Region on LG23

The Y chromosome *amhY* gene located on LG23 was shown to be necessary for maleness and is possibly the primary gene in a Japanese strain [43]. These authors also found the presence of a tandem *amh* duplication on the Y chromosome, with the existence of a truncated *amhΔY* gene together with the *amhY* gene. Our work confirms the existence of this tandem duplication on the Y chromosome of LG23

in the two wild populations studied. We previously used four *amh* markers that distinguish the X and the Y-chromosomes that predicted the sexual phenotype with 100% certitude in the domesticated Manzala strain [47]. These markers also predicted sex correctly for 82% of the individuals in the wild Lake Kou population, while the remaining fish showed evidence of other epistatic loci [47]. We used these markers in both the Kpandu and Koka populations, but they were discriminative only in the Kpandu population. They allowed the identification of 3 XX males in the wild-caught fish, which were probably sex-reversed by high temperatures [48]. In contrast, only 2 markers corresponding to the *amh* gene amplified in the Koka population. Those corresponding to *amh*ΔY did not amplify. These results initially suggested that the Koka males lacked the *amh*ΔY gene or that the Koka sequences varied too much for our primers to amplify. However, our WGS allowed us to affirm with certainty the presence of this tandem duplication of the *amh* region with the presence of both *amh*Y and *amh*ΔY genes in the Kpandu and Koka males.

4.5. Position of the Structural Variants in the Tandem Duplication of the Y Chromosome

We found two large deletions (>5 Kb) within the Y chromosome duplication that are shared by both wild populations. This could be evidence for a common origin of the male-specific *oaz1-dot11* region between the two wild populations of Ethiopia and Ghana. One of these deletions overlaps *oaz1* and ends just before the last exon of the *amh*, whereas the second deletion occurs just before the first exon of the *amh* and overlaps the last exon of *dot11*. There are two ways in which the deletion could overlap *oaz1* gene within the duplicated region: either it could overlap the intact copy of *oaz1*, or it could overlap the *dot11-oaz1* chimera. The first scenario implies that there would not be a complete copy of *oaz1* on the Y chromosome. Thus, we suggest that the second scenario seems more plausible because the deletion removes an already truncated (and putatively non-functional) copy of the gene. For the same reasons, we propose that the deletion overlapping the *dot11* gene takes place within the *dot11-oaz1* chimera on the Y haplotype. These conclusions are consistent with Li et al. (2015) [43] who describe a large deletion (5.6 Kb) in the *amh*Y promoter. We suggest that this deletion is the one overlapping the truncated copy of the *dot11* gene.

Amongst the various male-specific SNPs we found in the duplication region, 8 are shared with other laboratory and commercial strains previously described [43,44]. All the SNPs described in the Japanese strain and which we found in the *amh* gene of our wild populations, were reported by Li et al. (2015) [43] as being *amh*ΔY specific. However, we did not find the missense SNP in exon 2 specific for the *amh*Y gene possibly contributing to male sex determination in the Japanese strain [43], in neither the Kpandu nor the Koka males. Nonetheless, we detected 52 new male-specific mutations in the *oaz1* to *dot11* region in our two natural populations. Among them, six are non-synonymous mutations found in the *amh* gene that are common to both Koka and Kpandu. Two of them are located in exon 7, changing a threonine to a serine (both hydroxyl-containing amino acids) and an alanine to a threonine substitution which implies a change from a non-polar hydrophobic to a polar hydrophilic amino acid. Another important missense mutation in both populations is present in exon 6 which changes a glutamic acid to lysine, a basic amino acid. The other *amh* missense located in exon 6 is also common with the Latin American strains studied by Cáceres et al. (2019) [44]. Their fish are all from commercial stocks originating from the GIFT strain, itself originating from crosses between five domesticated and five wild populations [33]. The threonine to alanine mutation of exon 6 (34,501,082 bp) is, therefore, a strong candidate for maleness as it is conserved across populations and strains and might induce a major difference between the *amh*Y and the female *amh* gene in wild populations.

Because our study was conducted using Illumina short reads sequencing, we are not able to determine which copy of the genes (truncated or complete) in the Y duplication our SNPs are from. Thus, it is impossible to conclude whether the SNPs affect the presumed sex-determining *amh*Y gene or if they occur in the supposedly truncated *amh*ΔY copy. As shown in Figure 6A, none of the shared male-specific SNPs or indels (blue bars) overlap the two large deletions (red bars). This indicates that

sex-specific SNPs are not found in the conserved copies of these regions. We suggest that the majority of the SNPs and small indels we found are on the truncated versions of the genes.

4.6. Polymorphism between Wild Populations in this SD Region

For the same reasons explained previously, it is likely that the majority of the population sex-specific variants are found in the truncated copies of genes within the duplication. Nevertheless, some SNPs specific to Koka males or Kpandu males overlap the large deletions. These variants must, therefore, be found in the intact copies of the tandem duplication. The Nile tilapia Kpandu (Ghana) and Koka (Ethiopia) populations have evolved divergently, partly because of their geographical isolation, but perhaps also because of their ecology [71]. Therefore, it is not surprising to see population-specific polymorphisms.

None of the variants specific to Koka males have previously been reported in the literature. In contrast, we found several genomic variants (small and large indels) in Kpandu males that have been identified in the past in domesticated strains. This is the case for the 5 bp insertion in exon 6 and the 234 bp ($CI \pm 3$ bp) deletion in exon 7 of the *amh* gene which are specificities of the *amh* Δ Y according to Eshel et al. (2014) (who called it initially *amhy*) [42] and Li et al. (2015) [43]. The 5 bp insertion causes a pre-mature stop codon in *amh* Δ Y, which would indicate that it is a truncated copy. We determined that this 5 bp insertion both by using PCR markers and by WGS was absent in the Koka males. This raises the question of whether the *amh* Δ Y might be functional in Koka males.

Due to our mapping onto the latest reference Nile tilapia genome and by specifically looking at the *amh* region, we highlighted a high nucleotide diversity in both wild populations. We cannot define which versions of the *oaz1*, *dot1l* and *amh* genes the SNPs belong to, however, the number of variants between X and Y haplotypes is greater than what has previously been described in domesticated strains [40,43,44]. We also found that the SNPs are numerous in the sex-determining region in wild populations. We did not find amongst these the sex-specific SNP in exon 2 described by Li et al. (2015) [43], and therefore, maleness is not dependent on this SNP in either population. Because of this unexpected diversity in this sex-determining region, it is possible that previously designed *amh* markers might not be efficient in wild populations. This appears to be the case for the Koka population which requires the analyses of new markers to discriminate better the *amh*Y, *amh* Δ Y and *amh*X genes. This also raises the question of whether using the existing *amh* markers are sufficient to sex individuals of wild populations and could explain some of the results of Sissao et al. (2019) [47]. Designing primers close to structural variants might be an efficient way to sex both wild populations and domesticated fish, as our study suggests that the sex region of farmed GIFT strains is still closely related to the Kpandu population [40,44].

4.7. Limits of the Methods Used

Although ddRAD sequencing is a relevant method to quickly identify potential sex-signals, we showed it might be influenced by repetitive elements and “old sex-chromosome” sequences, producing confounding artefactual signals. The ability of ddRAD to detect a signal will vary depending on the region analysed and the position of the ddRAD loci. Problems associated with a low density of markers in “RAD” studies (incorrect estimation of nucleotide diversity across the genome, missing sites...) have been addressed previously (reviewed by Lowry et al. [74]) and are accentuated when using ddRAD methods which sequence even fewer loci across the genome. Our results highlighted the role of the *amh* gene as a major sex-determining gene, but by using ddRAD sequencing we might have missed smaller regions implicated in sex determination acting as minor genetic factors. A recent study on a wild population of Burkina Faso showed that *amh* genotyping predicted the phenotypic sex with only 82% success, suggesting the existence of other minor sex-determinants [47]. The existence of such minor genetic factors influencing sex determination has previously been shown [12]. Here we combined our ddRAD analyses with whole-genome sequencing (WGS). Pooling by sex is a good method to ensure high coverage with reduced costs. However, we found that the presence of B

chromosomes with highly repeated sequences is an issue when using the sequencing of pooled wild individuals to look for sex-specific signals with WGS. One should make sure of the absence of such genomic structures, especially when studying wild populations as we found evidence for B chromosomes in two distant populations across Africa. Nevertheless, WGS did enable us to study in-depth the genome structure around the *amh* region in the genome of wild Kpandu and Koka males.

5. Conclusions

Our study is the first to search for the sex determiner in wild populations of Nile tilapia using genomic approaches. F1 family analyses with ddRAD sequencing indicated that the major Y male determinant was located in LG23, in both Lake Volta (Kpandu) in West Africa and Lake Koka from East Africa. While the sex-variants covered a region of about 16 Mb in Koka, it encompassed nearly the whole of LG23 in Kpandu. A weaker sex association was identified on LG3 only in the Kpandu population, although neither XY nor ZW signals were conclusive. We confirmed the existence of a tandem duplication of the sex determiner candidate *amh* gene, in the Y chromosome of both Kpandu and Koka males. The breakpoints of the duplicated region were shown to be between the *oaz1* and *dot11* genes. We found two common male deletions of ~5 kb, and the presence of both *amhY* and *amhΔY* genes in both populations. We did not find the missense SNP in exon 2 which was thought to be important for maleness in a domesticated strain but found other male-specific variants. The *amhΔY* gene in Koka males presented no structural variation, which raises the question of whether it is still a functional gene. Finally, WGS uncovered B blocks of high coverage suggesting for the first time in Nile tilapia the presence of B chromosomes in both populations that were not sex-related at the species level.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/11/9/1017/s1>, Figure S1: Genotyping of wild-caught Kpandu fish of the AmhDY+5, AmhDY-233, AmhY-5608 and AmhX+36 markers; Figure S2: Genotyping of wild-caught Koka fish of the AmhDY+5, AmhDY-233, AmhY-5608 and AmhX+36 markers; Figure S3: Genotypes of the offspring compared to their parents inferred from ddRAD FET significant sites following a ZZ/ZW pattern; Table S1: Demultiplexing statistics for the ddRAD sequencing; Table S2: Details of the ddRAD FET-significant sites for LG23 found for both Kpandu and Koka populations; Table S3: Details of the ddRAD FET-significant sites for LG3 found for the Kpandu population; Table S4: SNPs and small indels found in the sex-determining *oaz1-dot11* region on the Y haplotype-specific to the Kpandu population between males and females; Table S5: SNPs and small indels found in the sex-determining *oaz1-dot11* region on the Y haplotype-specific to the Koka population between males and females.

Author Contributions: Conceptualization H.D., T.D.K. and J.-F.B.; methodology C.T., H.D., D.J.P. and T.D.K.; software C.T. and M.A.C.; validation C.T., H.D., T.D.K. and M.A.C.; formal analysis C.T., H.D., T.D.K., M.A.C. and J.-F.B.; investigation C.T., H.D. and F.E.C.; resources E.B. and J.-F.B.; data curation C.T., M.A.C., H.D., J.-F.B. and T.D.K.; writing—original draft preparation C.T., H.D. and J.-F.B.; writing—review and editing C.T., H.D., J.-F.B., T.D.K., M.A.C., F.E.C., D.J.P. and E.B.; visualization C.T., H.D., J.-F.B. and T.D.K.; supervision, H.D. and T.D.K.; project administration J.-F.B.; funding acquisition J.-F.B. and H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research and Cécile Triay's PhD was supported by the CLIMSEX project funded by the Agence Nationale de la Recherche (ANR,) grant number ANR-15-CE02-0012. CIRAD gave an additional help for Cécile Triay PhD. CIRAD and LabEx CEMEB contributed financially to Cécile Triay's trip to work at the University of Maryland.

Acknowledgments: We are very grateful to John Taggart for his help with the ddRAD primers and protocol. We also thank Kerry Bartie for instructing Cécile Triay in ddRAD library preparation. The authors thank Charles Mélard for the Manzala-Tihange XX, XY and YY fish. gDNA extraction on the KingFisher robot was performed at the GenSeq platform. The sequencing of the ddRAD library was performed by the GeT-PlaGe platform. The sequencing of WG libraries was performed by the GenomEast platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Ellegren, H. Evolutionary stasis: The stable chromosomes of birds. *Trends Ecol. Evol.* **2010**, *25*, 283–291. [[CrossRef](#)] [[PubMed](#)]
- Cortez, D.; Marin, R.; Toledo-Flores, D.; Froidevaux, L.; Liechti, A.; Waters, P.D.; Grützner, F.; Kaessmann, H. Origins and functional evolution of Y chromosomes across mammals. *Nature* **2014**, *508*, 488–493. [[CrossRef](#)] [[PubMed](#)]
- Koopman, P. The molecular biology of SRY and its role in sex determination in mammals. *Reprod. Fertil. Dev.* **1995**, *7*, 713–722. [[CrossRef](#)]
- Bachtrog, D. Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **2013**, *14*, 113–124. [[CrossRef](#)] [[PubMed](#)]
- Baroiller, J.-F.; Guiguen, Y.; Fostier, A. Endocrine and environmental aspects of sex differentiation in fish. *Cell. Mol. Life Sci.* **1999**, *55*, 910. [[CrossRef](#)]
- Devlin, R.H.; Nagahama, Y. Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *Aquaculture* **2002**, *208*, 191–364. [[CrossRef](#)]
- Baroiller, J.-F.; D’Cotta, H.; Saillant, E. Environmental Effects on Fish Sex Determination and Differentiation. *Sex. Dev.* **2009**, *3*, 118–135. [[CrossRef](#)]
- Bachtrog, D.; Mank, J.E.; Peichel, C.L.; Kirkpatrick, M.; Otto, S.P.; Ashman, T.-L.; Hahn, M.W.; Kitano, J.; Mayrose, V.B.; Ming, R.; et al. Sex Determination: Why So Many Ways of Doing It? *PLoS Biol.* **2014**, *12*, e1001899. [[CrossRef](#)]
- Gammerding, W.J.; Kocher, T.D. Unusual Diversity of Sex Chromosomes in African Cichlid Fishes. *Genes* **2018**, *9*, 480. [[CrossRef](#)]
- Volff, J.-N.; Scharl, M. Variability of genetic sex determination in poeciliid fishes. *Genetica* **2001**, *111*, 101–110. [[CrossRef](#)]
- Vandeputte, M.; Dupont-Nivet, M.; Chavanne, H.; Chatain, B. A Polygenic Hypothesis for Sex Determination in the European Sea Bass *Dicentrarchus labrax*. *Genetica* **2007**, *176*, 1049–1057. [[CrossRef](#)]
- Baroiller, J.-F.; D’Cotta, H.; Bezault, E.; Wessels, S.; Hoerstgen-Schwark, G. Tilapia sex determination: Where temperature and genetics meet. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **2009**, *153*, 30–38. [[CrossRef](#)] [[PubMed](#)]
- Matsuda, M.; Nagahama, Y.; Shinomiya, A.; Sato, T.; Matsuda, C.; Kobayashi, T.; Morrey, C.E.; Shibata, N.; Asakawa, S.; Shimizu, N.; et al. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **2002**, *417*, 559–563. [[CrossRef](#)] [[PubMed](#)]
- Nanda, I.; Kondo, M.; Hornung, U.; Asakawa, S.; Winkler, C.; Shimizu, A.; Shan, Z.; Haaf, T.; Shimizu, N.; Shima, A.; et al. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11778–11783. [[CrossRef](#)] [[PubMed](#)]
- Myosho, T.; Otake, H.; Masuyama, H.; Matsuda, M.; Kuroki, Y.; Fujiyama, A.; Naruse, K.; Hamaguchi, S.; Sakaizumi, M. Tracing the Emergence of a Novel Sex-Determining Gene in Medaka, *Oryzias latipes*. *Genetics* **2012**, *191*, 163–170. [[CrossRef](#)]
- Hattori, R.S.; Murai, Y.; Oura, M.; Masuda, S.; Majhi, S.K.; Sakamoto, T.; Fernandino, J.I.; Somoza, G.M.; Yokota, M.; Strüßmann, C.A. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2955–2959. [[CrossRef](#)]
- Hattori, R.S.; Somoza, G.M.; Fernandino, J.I.; Colautti, D.; Miyoshi, K.; Gong, Z.; Yamamoto, Y.; Strüßmann, C.A. The Duplicated Y-specific amhy Gene Is Conserved and Linked to Maleness in Silversides of the Genus *Odontesthes*. *Genes* **2019**, *10*, 679. [[CrossRef](#)]
- Pan, Q.; Feron, R.; Yano, A.; Guyomard, R.; Jouanno, E.; Vigouroux, E.; Wen, M.; Busnel, J.-M.; Bobe, J.; Concordet, J.-P.; et al. Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLoS Genet.* **2019**, *15*, e1008013. [[CrossRef](#)]
- Kamiya, T.; Kai, W.; Tasumi, S.; Oka, A.; Matsunaga, T.; Mizuno, N.; Fujita, M.; Suetake, H.; Suzuki, S.; Hosoya, S.; et al. A Trans-Species Missense SNP in Amhr2 Is Associated with Sex Determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genet.* **2012**, *8*, e1002798. [[CrossRef](#)]
- Yano, A.; Guyomard, R.; Nicol, B.; Jouanno, E.; Quillet, E.; Klopp, C.; Cabau, C.; Bouchez, O.; Fostier, A.; Guiguen, Y. An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Curr. Biol.* **2012**, *22*, 1423–1428. [[CrossRef](#)]

21. Ser, J.R.; Roberts, R.B.; Kocher, T.D. Multiple Interacting Loci Control Sex Determination in Lake Malawi Cichlid Fish. *Evolution* **2009**, *64*, 486–501. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Parnell, N.F.; Streelman, J.T. Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. *Heredity* **2012**, *110*, 239–246. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Gammerdinger, W.J.; Conte, M.A.; Sandkam, B.A.; Ziegelbecker, A.; Koblmüller, S.; Kocher, T.D. Novel Sex Chromosomes in 3 Cichlid Fishes from Lake Tanganyika. *J. Hered.* **2018**, *109*, 489–500. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Böhne, A.; Wilson, C.; Postlethwait, J.H.; Salzburger, W. Variations on a theme: Genomics of sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genom.* **2016**, *17*, 1–12. [\[CrossRef\]](#)
25. Roberts, N.B.; Juntti, S.A.; Coyle, K.P.; Dumont, B.L.; Stanley, M.K.; Ryan, A.Q.; Fernald, R.D.; Roberts, R.B. Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genom.* **2016**, *17*, 835. [\[CrossRef\]](#)
26. Clark, F.B.; Conte, M.; Ferreira-Bravo, I.A.; Poletto, A.B.; Martins, C.; Kocher, T.D. Dynamic Sequence Evolution of a Sex-Associated B Chromosome in Lake Malawi Cichlid Fish. *J. Hered.* **2016**, *108*, 53–62. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Clark, F.B.; Kocher, T.D. Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. *Sci. Rep.* **2019**, *9*, 1–10. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Baroiller, J.-F.; D’Cotta, H. Sex Control in Tilapias. In *Sex Control in Aquaculture*; John Wiley & Sons Ltd.: Chichester, UK, 2018; pp. 189–234.
29. Lee, B.-Y.; Penman, D.J.; Kocher, T.D. Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Anim. Genet.* **2003**, *34*, 379–383. [\[CrossRef\]](#)
30. Cnaani, A.; Lee, B.Y.; Zilberman, N.; Ozouf-Costaz, C.; Hulata, G.; Ron, M.; D’Hont, A.; Baroiller, J.-F.; D’Cotta, H.; Penman, D.; et al. Genetics of Sex Determination in Tilapiine Species. *Sex. Dev.* **2008**, *2*, 43–54. [\[CrossRef\]](#)
31. Palaiokostas, C.; Bekaert, M.; Khan, M.G.Q.; Taggart, J.B.; Gharbi, K.; McAndrew, B.J.; Penman, D.J. Mapping and Validation of the Major Sex-Determining Region in Nile Tilapia (*Oreochromis niloticus* L.) Using RAD Sequencing. *PLoS ONE* **2013**, *8*, e68389. [\[CrossRef\]](#)
32. Mair, G.C.; Scott, A.G.; Penman, D.J.; Beardmore, J.A.; Skibinski, D.O.F. Sex determination in the genus *Oreochromis*. *Theor. Appl. Genet.* **1991**, *82*, 144–152. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Ponzoni, R.W.; Nguyen, N.H.; Khaw, H.L.; Hamzah, A.; Abu Bakar, K.R.; Yee, H.Y. Genetic improvement of Nile tilapia (*Oreochromis niloticus*) with special reference to the work conducted by the WorldFish Center with the GIFT strain. *Rev. Aquac.* **2011**, *3*, 27–41. [\[CrossRef\]](#)
34. Carrasco, L.A.; Penman, D.J.; Bromage, N. Evidence for the presence of sex chromosomes in the Nile tilapia (*Oreochromis niloticus*) from synaptonemal complex analysis of XX, XY and YY genotypes. *Aquaculture* **1999**, *173*, 207–218. [\[CrossRef\]](#)
35. Lee, B.-Y.; Coutanceau, J.-P.; Ozouf-Costaz, C.; D’Cotta, H.; Baroiller, J.-F.; Kocher, T.D. Genetic and Physical Mapping of Sex-Linked AFLP Markers in Nile Tilapia (*Oreochromis niloticus*). *Mar. Biotechnol.* **2010**, *13*, 557–562. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Shirak, A.; Seroussi, E.; Cnaani, A.; Howe, A.E.; Domokhovskiy, R.; Zilberman, N.; Kocher, T.D.; Hulata, G.; Ron, M. AmhandMrta2Genes Map to Tilapia (*Oreochromis* spp.) Linkage Group 23 Within Quantitative Trait Locus Regions for Sex Determination. *Genetics* **2006**, *174*, 1573–1581. [\[CrossRef\]](#)
37. Eshel, O.; Weller, J.I.; Slossman, T.; Ron, M.; Shirak, A.; Hulata, G.; Cnaani, A. Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*). *Anim. Genet.* **2010**, *42*, 222–224. [\[CrossRef\]](#)
38. Gammerdinger, W.J.; Conte, M.A.; Acquah, E.A.; Roberts, R.B.; Kocher, T.D. Structure and decay of a proto-Y region in Tilapia, *Oreochromis niloticus*. *BMC Genom.* **2014**, *15*, 975. [\[CrossRef\]](#)
39. Palaiokostas, C.; Bekaert, M.; Khan, M.G.Q.; Taggart, J.B.; Gharbi, K.; McAndrew, B.J.; Penman, D.J. A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genom.* **2015**, *16*, 171. [\[CrossRef\]](#)
40. Taslima, K.; Wehner, S.; Taggart, J.B.; de Verdal, H.; Benzie, J.A.H.; Bekaert, M.; McAndrew, B.J.; Penman, D.J. Sex determination in the GIFT strain of tilapia is controlled by a locus in linkage group 23. *BMC Genet.* **2020**, *21*, 1–15. [\[CrossRef\]](#)
41. Eshel, O.; Shirak, A.; Weller, J.I.; Hulata, G.; Ron, M. Linkage and Physical Mapping of Sex Region on LG23 of Nile Tilapia (*Oreochromis niloticus*). *G3 Genes Genomes Genet.* **2012**, *2*, 35–42. [\[CrossRef\]](#)

42. Eshel, O.; Shirak, A.; Dor, L.; Band, M.; Zak, T.; Markovich-Gordon, M.; Chalifa-Caspi, V.; Feldmesser, E.; Weller, J.I.; Seroussi, E.; et al. Identification of male-specific amh duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genom.* **2014**, *15*, 774. [CrossRef] [PubMed]
43. Li, M.; Sun, Y.; Zhao, J.; Shi, H.; Zeng, S.; Ye, K.; Jiang, D.; Zhou, L.; Sun, L.; Tao, W.; et al. A Tandem Duplicate of Anti-Müllerian Hormone with a Missense SNP on the Y Chromosome Is Essential for Male Sex Determination in Nile Tilapia, *Oreochromis niloticus*. *PLoS Genet.* **2015**, *11*, e1005678. [CrossRef] [PubMed]
44. Cáceres, G.; López, M.E.; Cádiz, M.I.; Yoshida, G.M.; Jedlicki, A.M.; Palma-Véjares, R.; Travisany, D.; Díaz-Domínguez, D.; Maass, A.; Lhorente, J.P.; et al. Fine Mapping Using Whole-Genome Sequencing Confirms Anti-Müllerian Hormone as a Major Gene for Sex Determination in Farmed Nile Tilapia (*Oreochromis niloticus* L.). *G3 Genes Genomes Genet.* **2019**, *9*, 3213–3223. [CrossRef] [PubMed]
45. Pfennig, F.; Standke, A.; Gutzeit, H.O. The role of Amh signaling in teleost fish—Multiple functions not restricted to the gonads. *Gen. Comp. Endocrinol.* **2015**, *223*, 87–107. [CrossRef] [PubMed]
46. Wilson, C.A.; High, S.K.; McCluskey, B.M.; Amores, A.; Yan, Y.-L.; Titus, T.A.; Anderson, J.L.; Batzel, P.; Carvan, M.J.; Scharl, M.; et al. Wild Sex in Zebrafish: Loss of the Natural Sex Determinant in Domesticated Strains. *Genetics* **2014**, *198*, 1291–1308. [CrossRef]
47. Sissao, R.; D'Cotta, H.; Baroiller, J.-F.; Toguyeni, A. Mismatches between the genetic and phenotypic sex in the wild Kou population of Nile tilapia *Oreochromis niloticus*. *PeerJ* **2019**, *7*, e7709. [CrossRef]
48. Bezault, E.; Clota, F.; Derivaz, M.; Chevassus, B.; Baroiller, J.-F. Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture* **2007**, *272*, S3–S16. [CrossRef]
49. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed on 27 August 2020).
50. Catchen, J.M.; Amores, A.; Hohenlohe, P.; Cresko, W.A.; Postlethwait, J.H. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3 Genes Genomes Genet.* **2011**, *1*, 171–182. [CrossRef]
51. A Conte, M.; Joshi, R.; Moore, E.C.; Nandamuri, S.P.; Gammerdinger, W.J.; Roberts, R.B.; Carleton, K.L.; Lien, S.; Kocher, T.D. Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *GigaScience* **2019**, *8*, 8. [CrossRef]
52. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
53. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.R.; Durbin, R. 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
54. Weir, B.S. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*; Sinauer Assoc. Inc.: Sunderland, MA, USA, 1996; ISBN 0878939024.
55. Danecek, P.; Auton, A.; Abecasis, G.R.; Albers, C.A.; Banks, E.; Depristo, M.A.; Ha, G.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]
56. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
57. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; Del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [CrossRef]
58. Kofler, R.; Pandey, R.V.; Schlötterer, C. PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **2011**, *27*, 3435–3436. [CrossRef] [PubMed]
59. Rausch, T.; Zichner, T.; Schlattl, A.; Stütz, A.M.; Benes, V.; Korbel, J.O. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **2012**, *28*, i333–i339. [CrossRef]
60. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**, arXiv:1207.3907.
61. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
62. Turner, S.D. QQman: An R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots BioRxiv. Available online: <https://www.biorxiv.org/content/10.1101/005165v1> (accessed on 8 June 2020).

63. Valente, G.T.; Conte, M.A.; Fantinatti, B.E.; Cabral-de-Mello, D.C.; Carvalho, R.; Vicari, M.R.; Kocher, T.D.; Martins, C. Origin and Evolution of B Chromosomes in the Cichlid Fish *Astatotilapia latifasciata* Based on Integrated Genomic Analyses. *Mol. Boil. Evol.* **2014**, *31*, 2061–2072. [[CrossRef](#)]
64. Clark, F.B.; Conte, M.A.; Kocher, T.D. Genomic Characterization of a B Chromosome in Lake Malawi Cichlid Fishes. *Genes* **2018**, *9*, 610. [[CrossRef](#)]
65. Wessels, S.; Krause, I.; Floren, C.; Schütz, E.; Beck, J.; Knorr, C. ddRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC Genom.* **2017**, *18*, 531. [[CrossRef](#)]
66. Lee, B.-Y.; Hulata, G.; Kocher, T.D. Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity* **2004**, *92*, 543–549. [[CrossRef](#)] [[PubMed](#)]
67. A Conte, M.; Gammerdinger, W.J.; Bartie, K.L.; Penman, D.J.; Kocher, T.D. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genom.* **2017**, *18*, 341. [[CrossRef](#)]
68. Tanger, I.S.; Bayer, T.; Salzburger, W.; Roth, O. Effects of parental care on resource allocation into immune defense and buccal microbiota in mouthbrooding cichlid fishes*. *Evolution* **2018**, *72*, 1109–1123. [[CrossRef](#)]
69. Poonlaphdecha, S.; Pepey, E.; Canonne, M.; de Verdal, H.; Baroiller, J.-F.; D’Cotta, H. Temperature induced-masculinisation in the Nile tilapia causes rapid up-regulation of both dmrt1 and amh expressions. *Gen. Comp. Endocrinol.* **2013**, *193*, 234–242. [[CrossRef](#)] [[PubMed](#)]
70. Mullen, R.D.; Ontiveros, A.E.; Moses, M.M.; Behringer, R.R. AMH and AMHR2 mutations: A spectrum of reproductive phenotypes across vertebrate species. *Dev. Boil.* **2019**, *455*, 1–9. [[CrossRef](#)] [[PubMed](#)]
71. Bezault, E.; Balaesque, P.; Toguyeni, A.; Fermon, Y.; Araki, H.; Baroiller, J.-F.; Rognon, X. Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genet.* **2011**, *12*, 102. [[CrossRef](#)]
72. Fantinatti, B.E.A.; Martins, C. Development of chromosomal markers based on next-generation sequencing: The B chromosome of the cichlid fish *Astatotilapia latifasciata* as a model. *BMC Genet.* **2016**, *17*, 119. [[CrossRef](#)]
73. Perazzo, G.X.; Noleto, R.B.; Vicari, M.R.; Gava, A.; Cestari, M.M. B chromosome polymorphism in South American cichlid. *Neotrop. Biodivers.* **2018**, *4*, 3–9. [[CrossRef](#)]
74. Lowry, D.B.; Hoban, S.; Kelley, J.L.; Lotterhos, K.E.; Reed, L.K.; Antolin, M.F.; Storfer, A. Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.* **2016**, *17*, 142–152. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

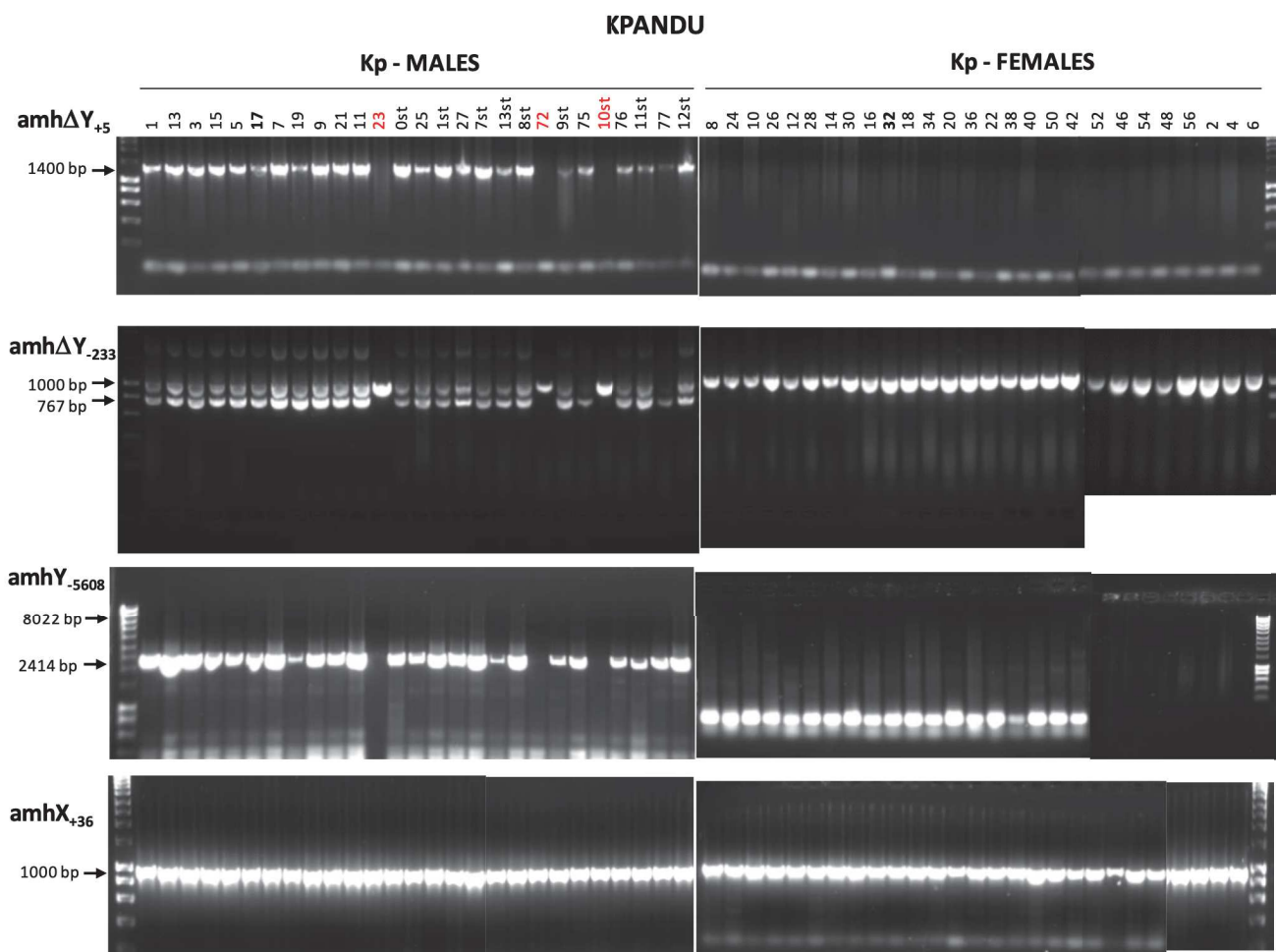


Figure S1: Genotyping of wild-caught Koka fish with the AmhDY+5, AmhDY- 233, AmhY-5608 and AmhX+36 markers

KOKA

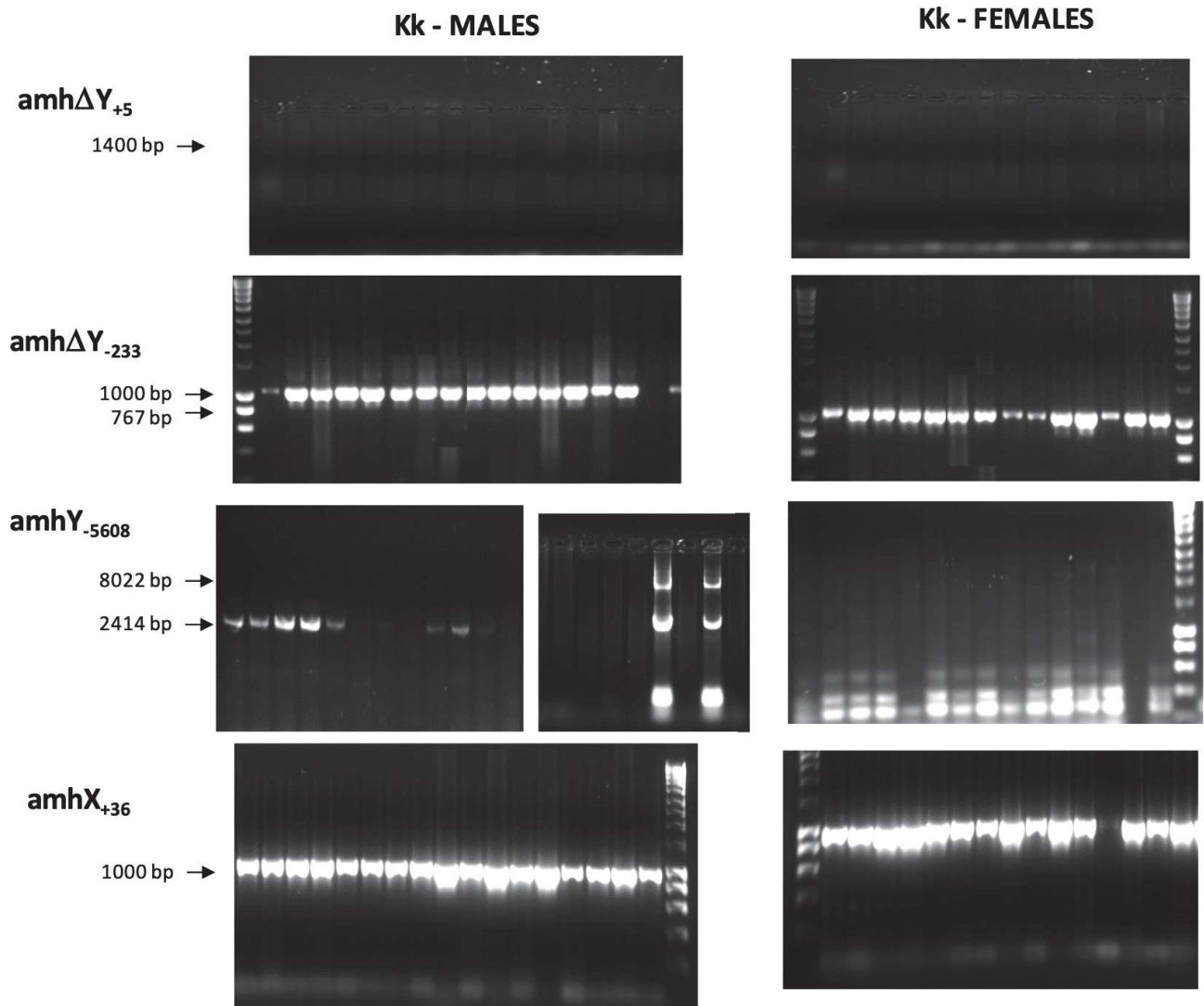


Figure S2: Genotyping of wild-caught Koka fish with the AmhDY+5, AmhDY- 233, AmhY-5608 and AmhX+36 markers

Kpandu LG23	POS	35271872	35271930
	REF	C	A
	ALT	T	T
	M 1		
	M 2		
	M 3		
	M 5		
	M 10		
	M 12		
	M 13		
	M 15		
	M 17		
	M 18		
	M 20		
	M 21		
	M 22		
	M 25		
	F 4		
	F 6		
	F 7		
	F 9		
	F 11		
	F 14		
	F 19		
	F 23		
	F 24		
	F 26		
	F 27		
	F 28		
	F 29		
	F 30		
	Dam	0/1	0/1
	Sire	0/0	0/0

Koka LG3	POS	25957411	25957414	38037497	38037538
	REF	G	A	C	T
	ALT	A	G	A	C
	M 1				
	M 2				
	M 3				
	M 5				
	M 8				
	M 10				
	M 13				
	M 14				
	M 17				
	M 18				
	M 19				
	M 20				
	M 22				
	M 25				
	M 26				
	F 4				
	F 6				
	F 7				
	F 9				
	F 11				
	F 12				
	F 15				
	F 16				
	F 23				
	F 24				
	F 27				
	F 28				
	F 30				
	Dam	0/1	0/1	0/1	0/1
	Sire	0/0	0/0	0/0	0/0

Kpandu LG03	POS	38955081	38955177	79657728
	REF	C	C	A
	ALT	T	T	T
	M 1			
	M 2			
	M 3			
	M 5			
	M 10			
	M 12			
	M 13			
	M 15			
	M 17			
	M 18			
	M 20			
	M 21			
	M 22			
	M 25			
	F 4			
	F 6			
	F 7			
	F 9			
	F 11			
	F 14			
	F 19			
	F 23			
	F 24			
	F 26			
	F 27			
	F 28			
	F 29			
	F 30			
	Dam	0/1	0/1	0/1
	Sire	0/0	0/0	0/0

Figure S3. Genotypes of the offspring compared to their parents inferred from ddRAD FET significant sites following a ZZ/ZW pattern. Blue cells indicates an individual having the same genotype as it’s father at this site whereas pink cells indicate an individual having the same genotype as it's mother (dashed cells indicates homozygous WW sites) at this site. White cells are missing data. Only sites having both parents and at least half males and females of the offspring sequenced were kept here

Table S1. Demultiplexing statistics for the ddRAD sequencing

Filename	Barcode	Total	No RadTag	Low Quality	Retained
Kk06_27-01	TCAGA-TAGCA	6519498	62393	10	6432161
Kk06_27-02	TGCAACA-AGCTGTC	6488002	42432	10	6421987
Kk06_27-03	GATCG-AGTCA	6130008	59742	12	6046134
Kk06_27-04	CGTATCA-TACGTGT	4786780	30645	11	4738595
Kk06_27-05	CATGA-GCATA	9155348	118757	17	9000160
Kk06_27-06	CACAGAC-GAGATGT	4304474	65547	8	4219379
Kk06_27-07	ATCGA-CGATC	6587634	72103	7	6492862
Kk06_27-08	ACTGCAC-CATCTGT	4087738	56399	2	4016341
Kk06_27-09	TCGAG-CTGGT	6728002	69386	11	6633318
Kk06_27-10	TCTCTCA-GTCAAGT	3997246	31884	7	3948143
Kk06_27-11	GTCAC-GAAGC	5398064	73342	3	5302251
Kk06_27-12	GTACACA-ATACGGT	5160480	37936	8	5104071
Kk06_27-13	GCATT-TAGCA	4329904	63290	8	4251195
Kk06_27-14	CTCTTCA-AGCTGTC	5150804	63152	12	5066501
Kk06_27-15	CGATA-AGTCA	6795198	67075	10	6702247
Kk06_27-16	CTAGGAC-TACGTGT	4458124	25108	8	4416516
Kk06_27-17	ACGTA-GCATA	8356696	150957	20	8170910
Kk06_27-18	CAGTCAC-GAGATGT	4402590	52193	9	4333027
Kk06_27-19	AGAGT-CGATC	4436860	57079	7	4363275
Kk06_27-20	GCTAACA-CATCTGT	6733466	59885	10	6650901
Kk06_27-21	ATGCT-CTGGT	45828	30344	0	15428
Kk06_27-22	ACACGAG-GTCAAGT	2569408	24098	3	2534472
Kk06_27-23	GACTA-GAAGC	3222708	63755	2	3145377
Kk06_27-24	AGGACAC-ATACGGT	3725288	58448	3	3652006
Kk06_27-25	TCAGA-AGCTGTC	2807592	36100	4	2760146
Kk06_27-26	TGCAACA-AGTCA	7424726	71086	9	7326480
Kk06_27-27	GATCG-TACGTGT	3747604	24415	3	3707727
Kk06_27-28	CGTATCA-GCATA	10693924	146273	20	10502658
Kk06_27-30	CATGA-GAGATGT	5827138	90528	4	5712289
KkGF04_27	GTACACA-AGCTGTC	4455130	44283	7	4392216
KkGM39_27	GCATT-AGTCA	3663614	43555	3	3606067
Kp20_27-01	CTCTTCA-TACGTGT	9118810	64960	19	9022280
Kp20_27-02	CGATA-GCATA	12392034	156916	21	12189103
Kp20_27-03	CTAGGAC-GAGATGT	7313068	69919	11	7218594
Kp20_27-04	ACGTA-CGATC	8325070	75518	20	8223412
Kp20_27-05	CAGTCAC-CATCTGT	7442942	72497	11	7344093
Kp20_27-06	AGAGT-CTGGT	5390846	113234	4	5259540
Kp20_27-07	GCTAACA-GTCAAGT	4501978	41902	9	4443363
Kp20_27-08	ATGCT-GAAGC	38146	28303	0	9777
Kp20_27-09	ACACGAG-ATACGGT	4740246	66828	2	4657183
Kp20_27-10	GACTA-TAGCA	6519380	92477	16	6403155
Kp20_27-11	AGGACAC-AGCTGTC	3317674	26002	9	3280807
Kp20_27-12	TCAGA-TACGTGT	5258990	48510	4	5192469
Kp20_27-13	TGCAACA-GCATA	10216096	71647	12	10106968
Kp20_27-14	GATCG-GAGATGT	6292344	36782	14	6232033
Kp20_27-15	CGTATCA-CGATC	7721200	71407	17	7624258
Kp20_27-17	CATGA-CATCTGT	11741310	81447	21	11625140
Kp20_27-18	CACAGAC-CTGGT	8255206	90916	12	8135409
Kp20_27-19	ATCGA-GTCAAGT	4667892	25809	5	4625598
Kp20_27-20	ACTGCAC-GAAGC	6783430	146471	12	6612036
Kp20_27-21	TCGAG-ATACGGT	7855194	42422	14	7787581
Kp20_27-22	TCTCTCA-TAGCA	7878918	111920	11	7741960
Kp20_27-23	GTCAC-AGCTGTC	7542616	44289	8	7472232
Kp20_27-24	GTACACA-AGTCA	8141530	65761	13	8047445
Kp20_27-25	GCATT-TACGTGT	5461980	34237	6	5408628
Kp20_27-26	CTCTTCA-GCATA	15576420	125831	22	15395325
Kp20_27-27	CGATA-GAGATGT	4719534	39821	6	4662161
Kp20_27-28	CTAGGAC-CGATC	10168514	70911	13	10064095
Kp20_27-29	ACGTA-CATCTGT	8314098	50526	16	8237185
Kp20_27-30	CAGTCAC-CTGGT	7810766	115129	8	7669778
KpGF32_27	AGAGT-GTCAAGT	2776938	36878	4	2729496
KpM17_27	GCTAACA-GAAGC	6458654	42653	7	6393873
Neg Control	ATGCT-ATACGGT	4408	2433	0	1965

Table S2. Details of ddRAD FET-significant sites for LG23 found for both Kpandu and Koka families

	BP	Fisher's Exact Test p-value	Amova Fst	Gene ID	Gene Start	Gene End	Product description
Kpandu LG23	3268204	5.42E-03	0.156	LOC109197015	3145390	4078398	ATP-dependent DNA helicase PIF1
	3268469	6.52E-03	0.148				
	5578395	2.36E-03	0.185	LOC100702162	5544973	5584661	multidrug resistance-associated protein 4
	6381389	5.65E-03	0.133	trak2	6373940	6389882	trafficking protein, kinesin binding 2
	6790753	9.93E-03	0.136	zranb3	6724257	6842613	zinc finger RANBP2-type containing 3
	7534698	2.60E-03	0.155	LOC100689728	7501806	7649676	gamma-aminobutyric acid receptor subunit gamma-3
	7818707	1.15E-03	0.179				
	8296704	1.15E-03	0.179				
	9113411	1.15E-03	0.179				
	9161445	2.61E-03	0.174	Non Coding			
	9161653	8.86E-03	0.134				
	9764218	2.25E-03	0.171				
	9982398	3.39E-03	0.190				
	11597075	1.15E-03	0.179				
	13021743	1.15E-03	0.179	LOC100699606	13020746	13022102	gamma-crystallin M3-like
	13602491	1.77E-03	0.203				
	13602504	4.45E-03	0.189				
	13602569	9.65E-03	0.165				
	13602575	1.95E-03	0.200				
	13602590	4.70E-03	0.171				
	13602591	1.95E-03	0.200	Non Coding			
	13602694	1.95E-03	0.200				
	13602701	9.19E-03	0.182				
	13602732	3.19E-03	0.196				
	13602959	1.15E-03	0.179				
	14115966	1.15E-03	0.179				
	17141915	8.95E-04	0.296				
	17613124	1.15E-03	0.179				
	17613143	1.15E-03	0.179	als2	17587289	17614314	alsin Rho guanine nucleotide exchange factor
	18309315	1.15E-03	0.179				
	19018833	9.10E-04	0.205				
	19031065	1.93E-03	0.221				
	19373150	7.34E-03	0.176				
	19609561	1.90E-03	0.166				
	19915347	2.61E-03	0.174	Non Coding			
	22988019	1.15E-03	0.232				
	22988087	1.15E-03	0.232				
	22988389	1.15E-03	0.232				
	23996168	9.10E-04	0.205				
	23996185	9.10E-04	0.205				
	23996228	9.10E-04	0.205				
	24090251	9.10E-04	0.205	LOC100690062	24067428	24119616	CLIP-associating protein 1-A
	24420205	9.10E-04	0.205	Non Coding			
	24729049	9.10E-04	0.205				
	24814636	9.10E-04	0.205	agap1	24688557	24858705	arf-GAP with GTPase ANK repeat and PH domain-containing protein 1
	24814747	9.10E-04	0.205				
	24932340	5.77E-03	0.156	LOC109196997	24910543	25412345	NLR family CARD domain-containing protein 3
	24932396	5.67E-03	0.156				
	27622612	2.88E-03	0.166				
	27622823	2.88E-03	0.166				
	27622855	3.73E-03	0.156	Non Coding			
	27622859	3.73E-03	0.156				
	27622864	2.88E-03	0.166				
	27622865	2.88E-03	0.166				
	27622937	2.88E-03	0.166				
	28006648	2.34E-04	0.276	LOC102077606	27999712	28030875	uncharacterized LOC102077606
	28183172	2.24E-04	0.224	LOC102077510	28169381	28190713	adhesion G-protein coupled receptor G7
	28183563	4.60E-04	0.216				
	28427026	3.64E-03	0.197	Non Coding			
	28720524	3.64E-03	0.156				
	28720537	3.64E-03	0.156	LOC100701816	28694684	28746200	NACHT LRR and PYD domains-containing protein 3
	28720563	3.64E-03	0.156				
	28720615	3.64E-03	0.156				
	28720616	3.64E-03	0.156				
	28720801	3.64E-03	0.156	LOC100701273	28706535	28721156	protein NLRC3-like
	28720868	3.64E-03	0.156				
	29326015	6.40E-04	0.255	mcf2l2	29307709	29404131	guanine nucleotide exchange factor DBS
	30529202	4.05E-04	0.221	LOC102076398	30446019	30539111	uncharacterized LOC102076398
	30529404	4.05E-04	0.221				
	30929265	7.83E-07	0.455				
	30929418	6.70E-05	0.329	LOC109196929	30927917	30936814	aldehyde dehydrogenase family 9 member A1-like
	30929419	3.19E-03	0.229				
	30944555	8.49E-04	0.214	LOC100707441	30940790	30950350	aldehyde dehydrogenase family 9 member A1
	31271536	4.05E-04	0.221				
	31271549	1.06E-04	0.260	evi5	31250049	31293025	ecotropic viral integration site 5
	31272084	1.06E-04	0.260				
	31610258	1.97E-04	0.231	Non Coding			
	33734321	8.90E-06	0.381	cyfip1	33706514	33746187	cytoplasmic FMR1 interacting protein 1
	33734328	8.90E-06	0.381				
	33907878	2.57E-05	0.309	celf5	33898573	34074566	CUGBP Elav-like family member 5
	34265123	3.30E-05	0.278	LOC100701388	34260519	34285917	zinc finger and BTB domain-containing protein 7A

Table S2. Continued

	BP	Fisher's Exact Test p-value	Amova Fst	Gene ID	Gene Start	Gene End	Product description
Kpandui LG23	34443504	3.30E-05	0.278	lingo3	34417222	34451018	leucine rich repeat and Ig domain containing 3
	34502301	2.82E-05	0.308	amh	34498945	34502800	anti-Mullerian hormone
	34502353	1.83E-05	0.297				
	34502501	1.83E-05	0.297				
	34772926	2.92E-06	0.333	tmem59l	34764172	34775066	transmembrane protein 59 like
	34774781	3.94E-05	0.286	sema6b	35024697	35148164	semaphorin 6B
	35124733	2.92E-06	0.333				
	35271872	4.63E-03	0.171				
	35271930	4.63E-03	0.171	Non Coding			
	35825054	1.76E-04	0.417	nwd1	35824755	35843686	NACHT and WD repeat domain containing 1
	35941307	8.38E-06	0.304	dbt	35940945	35947903	dihydrolipoamide branched chain transacylase E2
	35953054	2.89E-03	0.207	lrrc39	35948014	35953096	leucine rich repeat containing 39
	36092189	9.85E-03	0.176	grin3b	36077197	36130419	glutamate ionotropic receptor NMDA type subunit 3B
	36219951	1.83E-05	0.297	LOC102079807	36219513	36228399	leukemia NUP98 fusion partner 1
	36702603	3.54E-04	0.219	Non Coding			
	37501005	2.63E-03	0.180	lnx1	37471068	37507778	ligand of numb-protein X 1
	37989174	2.33E-03	0.161	Non Coding			
	38142119	7.70E-04	0.211				
	38142263	7.70E-04	0.211				
	38635419	3.81E-04	0.274				
	38650855	8.38E-06	0.304				
	38767722	8.38E-06	0.304	vamp4	38761399	38771204	vesicle associated membrane protein 4
	39345791	8.00E-03	0.143	LOC106097024	39345054	39350322	cytochrome P450 2J6
	39345873	8.00E-03	0.143				
	39345898	8.00E-03	0.143				
	39352915	8.38E-06	0.304	LOC102080408	39352504	39355741	cytochrome P450 2J6
	39495317	1.71E-03	0.259	LOC100701255	39408394	39510887	cytochrome P450 2J2
	39495320	2.90E-03	0.258				
	39495321	1.71E-03	0.259				
	39495351	1.71E-03	0.259				
	39495368	1.71E-03	0.259				
	39495382	1.71E-03	0.259				
	39495394	1.71E-03	0.259				
	39495409	2.90E-03	0.258				
	39687063	1.48E-04	0.282	ddr2	39674574	39709288	discoidin domain receptor tyrosine kinase 2
	39785274	4.25E-04	0.338	fsd1	39782941	39792956	fibronectin type III and SPRY domain containing 1
	39785344	2.58E-04	0.308				
	40703079	4.42E-04	0.304				
	41258852	6.04E-03	0.178	LOC100691436	40679580	40708220	midnolin
	41863916	9.86E-03	0.121	saftb2	41860641	41868251	scaffold attachment factor B2
	41864056	9.86E-03	0.121				
	43182780	9.10E-04	0.205				
	44018965	4.16E-03	0.142	Non Coding			
				LOC100706720	43894847	44022814	voltage-dependent R-type calcium channel subunit alpha-1E
	BP	Fisher's Exact Test p-value	Amova Fst	Gene ID	Gene Start	Gene End	Product description
Koka LG23	24256201	1.62E-03	0.240	Non Coding			
	24377090	1.00E-03	0.208	LOC106098969	24373894	24379069	E3 ubiquitin-protein ligase TRIM21-like
	24377195	1.00E-03	0.208				
	24377630	1.07E-03	0.208				
	24420429	9.76E-04	0.195	Non Coding			
	24420433	9.76E-04	0.195				
	24728973	3.19E-03	0.229				
	24729049	3.19E-03	0.229	agap1	24688557	24858705	arf-GAP with GTPase, ANK repeat and PH domain-containing protein 17
	25957411	9.76E-04	0.195	Non Coding			
	25957414	2.25E-03	0.171				
	29325835	2.38E-03	0.181				
	29326050	2.38E-03	0.181	mcf2l2	29307709	29404131	guanine nucleotide exchange factor DBS5
	29326051	2.38E-03	0.181	Non Coding			
	29595691	8.81E-03	0.131				
	29916656	1.18E-04	0.239				
	29916658	1.18E-04	0.239	nlg1	29685708	29978507	neuroligin 11
	30030174	3.62E-03	1.000	naaladl2	29983730	30443760	N-acetylated alpha-linked acidic dipeptidase like 22
	30254103	1.18E-04	0.239				
	34502301	1.56E-03	0.188				
	34502339	1.62E-03	0.188	amh	34498945	34502800	anti-Mullerian hormone
	34502353	9.24E-03	0.133				
	34502501	1.56E-03	0.188				
	35097073	1.05E-06	0.762	sema6b	35024697	35148164	semaphorin 6B1
	35271920	1.13E-05	0.292	Non Coding			
	38037497	2.25E-03	0.171	impact	38027394	38052075	impact RWD domain protein
	38037538	2.25E-03	0.171				
	39346762	6.40E-04	0.255				
	39346887	6.40E-04	0.255	LOC106097024	39345054	39350322	cytochrome P450 2J6
	44145775	8.10E-05	0.493	LOC100702067	44143983	44228029	xenotropic and polytropic retrovirus receptor 1 homolog

Table S3. Details of ddRAD FET-significant sites for LG3 found for the Kpandu family

	BP	Fisher's Exact Test p-value	Amova Fst	Gene ID	Gene Start	Gene End	Product description
Kpandu LG3	20413224	7.68E-05	0.392	Non Coding			
	20413329	2.07E-04	0.361				
	20413333	2.07E-04	0.361				
	20413350	9.58E-04	0.322				
	20413371	7.68E-05	0.392				
	20413372	2.07E-04	0.361				
	20413385	2.07E-04	0.361				
	20413387	7.68E-05	0.392				
	20413395	7.68E-05	0.392				
	20413397	7.68E-05	0.392				
	20413407	7.68E-05	0.392				
	20766014	6.43E-03	0.357	LOC109198185	20483081	21480470	low affinity immunoglobulin gamma Fc region receptor II-b
	20766054	6.43E-03	0.357				
	20766055	6.43E-03	0.357				
	20766056	6.43E-03	0.357				
	20766057	6.43E-03	0.357				
	20766122	6.43E-03	0.357				
	20766236	6.43E-03	0.357				
	36091489	9.47E-03	0.190	LOC109198941	36064938	36268996	polymeric immunoglobulin receptor
	38181758	8.30E-03	0.154	Non Coding			
	38187085	4.82E-03	0.138				
	38955081	7.39E-03	0.183				
	38955177	9.53E-03	0.200				
	42505631	2.24E-03	0.381	LOC100695786	42240535	42871489	zinc finger protein 239
	42505666	9.66E-04	0.400	LOC102079722	42486440	42551726	uncharacterized LOC102079722
	46867058	1.80E-03	0.210	LOC102075779	46859269	46890096	interferon-induced protein 44
	46867125	1.80E-03	0.210				
	46867130	1.80E-03	0.210				
	46867131	1.80E-03	0.210				
	46867155	1.80E-03	0.210				
	46867199	1.80E-03	0.210				
	46867207	1.80E-03	0.210				
	46867229	1.80E-03	0.210				
	46867235	1.80E-03	0.210				
	46867244	1.80E-03	0.210				
	46867260	1.80E-03	0.210				
	46867275	1.80E-03	0.210				
	52823299	1.35E-03	0.218	Non Coding			
	52823319	1.16E-03	0.228				
	56398606	5.81E-03	0.132				
	56398607	5.81E-03	0.132				
	59781296	1.54E-03	0.314	LOC102081234	59776051	59874428	interferon-induced protein with tetratricopeptide repeats 5
	59781378	1.54E-03	0.314				
	59781461	1.54E-03	0.314				
	59781466	1.54E-03	0.314				
	59781493	1.54E-03	0.314				
	59781499	1.54E-03	0.314				
	59781506	1.54E-03	0.314				
	59781526	1.54E-03	0.314				
	61393476	4.82E-03	0.138	LOC100703578	60783576	61755773	nuclear factor 7, ovary
	61443409	8.31E-04	0.325	LOC100708849	61376869	61661012	scavenger receptor cysteine-rich type 1 protein M130
				LOC109196347	61441293	61444643	zinc finger BED domain-containing protein 1-like
	62886249	1.85E-03	0.174	Non Coding			
	62886306	1.85E-03	0.174				
	62925690	7.20E-03	0.169				
	64212311	3.97E-03	0.220	LOC112846310	64212138	64216441	selection and upkeep of intraepithelial T-cells protein 8-like
	65461749	2.14E-03	0.197	Non Coding			
	66629972	8.45E-03	0.130	LOC109199604	66617589	66632971	uncharacterized LOC109199604
	66630036	3.46E-03	0.159				
	66630046	3.46E-03	0.159				
	66630065	2.52E-03	0.178				
	66630117	3.46E-03	0.159				
	66630142	3.46E-03	0.159				
	66630215	3.46E-03	0.159				
	66689731	6.85E-03	0.159	LOC102076984	66675533	66697030	immunoglobulin superfamily member 2
	66689764	1.23E-03	0.193				
	66689825	3.22E-03	0.171				
	66689847	3.22E-03	0.171				
	67568787	1.62E-03	0.185	Non Coding			
	67799323	1.41E-03	0.268				
	67799376	1.91E-03	0.228				
	70202835	4.34E-03	0.179	LOC106096754	70191359	70259626	coxsackievirus and adenovirus receptor
	70202841	4.34E-03	0.179				
	70202862	4.34E-03	0.179				
	70202908	4.34E-03	0.179				
	70202916	4.34E-03	0.179				
	71886328	5.42E-03	0.156	Non Coding			
	71886348	5.05E-03	0.171				
	72745678	3.90E-03	0.158				
	72745684	3.90E-03	0.158				
	72745796	3.90E-03	0.158				

Table S3. Continued

	BP	Fisher's Exact Test p-value	Amova Fst	Gene ID	Gene Start	Gene End	Product description
Kpandu LG3	74381680	1.48E-03	0.188	LOC102082603	74377971	74409771	uncharacterized LOC102082603
	74381681	1.48E-03	0.188				
	74381692	1.48E-03	0.188				
	74381744	1.48E-03	0.188				
	74381746	1.48E-03	0.188				
	74381750	1.48E-03	0.188				
	74381754	1.48E-03	0.188				
	74381780	1.48E-03	0.188				
	74381798	1.48E-03	0.188				
	77023791	4.79E-03	0.155	Non Coding			
	77092691	4.79E-03	0.155				
	77173341	8.70E-04	0.218	cnga1	77170311	77177913	cyclic nucleotide gated channel alpha 1
	77173349	8.70E-04	0.218				
	77173379	8.70E-04	0.218				
	77173389	8.70E-04	0.218				
	77173430	8.70E-04	0.218				
	77201534	4.82E-03	0.138	nfxl1	77178317	77205828	nuclear transcription factor, X-box binding like 1
	77266911	9.03E-03	0.125	LOC109194306	77255116	77269156	sodium/hydrogen exchanger 9B2
	79657728	3.51E-03	0.148	LOC100699677	79653968	79662438	Beta-enolase

Table S4. SNPs and small indels found between males and females in the sex-determining *oaz1-dot1l* region on the Y haplotype specific to the Kpandu population

Position	Reference	Alternative	Length	Type	Gene	Intron / Exon	Change of amino acid	Found in literature
34492258	C	G	1	snp	<i>oaz1</i>	Intron		
34492617	G	A	1	snp	<i>oaz1</i>	Intron		
34492713	C	T	1	snp	<i>oaz1</i>	Intron		
34492746	ATAGTGTTT GTAGTGTTT GT	ATAGTGTTT GT	9	deletion	<i>oaz1</i>	Intron		
34492977	AAG	AACAG	2	insertion	<i>oaz1</i>	Intron		
34493031	A	G	1	snp	<i>oaz1</i>	Intron		
34497933	G	C	1	snp	LOC1007074 71	Intron		
34498818	A	C	1	snp	LOC1007074 71	Intron		
34499366	CAAAATATC	CAAATATC	1	deletion	LOC1007074 71 <i>amh</i>	5'UTR / 3'UTR		
34499592	G	A	1	snp	<i>amh</i>	Exon 7	Ala → Val	Eshel et al., 2014
34499740	T	C	1	snp	<i>amh</i>	Exon 7	Asn → Asp	
34499957	CA	TG,CG	2,1	mnp,snp	<i>amh</i>	Exon 7		Eshel et al., 2014
34500039	T	C	1	snp	<i>amh</i>	Intron		
34500045	C	G	1	snp	<i>amh</i>	Intron		
34500046	G	A	1	snp	<i>amh</i>	Intron		
34500246	A	C	1	snp	<i>amh</i>	Intron		
34500780	C	T	1	snp	<i>amh</i>	Exon 6		
34501028	C	T	1	snp	<i>amh</i>	Exon 6	Gly → Ser	
34501098	G	C	1	snp	<i>amh</i>	Exon 6		
34501125	CTC	CGACATTC	5	insertion	<i>amh</i>	Exon 6	Frame Shift	Li et al., 2015
34501180	G	A	1	snp	<i>amh</i>	Intron		
34501759	C	T	1	snp	<i>amh</i>	Intron		
34502301	A	G	1	snp	<i>amh</i>	Exon 2	Phe → Leu	
34502780	G	A	1	snp	<i>amh</i>	Intron		
34502864	G	C	1	snp	Non coding region			
34502934	CCG	TCA	3	complex	Non coding region			
34503031	C	T	1	snp	Non coding region			
34503062	G	C	1	snp	Non coding region			
34503220	A	G	1	snp	Non coding region			
34503326	CTC	CTCTTC	3	insertion	Non coding region			
34505875	A	C	1	snp	Non coding region			
34509196	T	C	1	snp	<i>dot1l</i>	3'UTR		
34509215	T	C	1	snp	<i>dot1l</i>	3'UTR		
34509624	A	T	1	snp	<i>dot1l</i>	3'UTR		
34509723	C	T	1	snp	<i>dot1l</i>	3'UTR		
34510382	GAAATATTT	GAAATAGC TACAAAATA TTT	11	insertion	<i>dot1l</i>	3'UTR		
34510584	C	T	1	snp	<i>dot1l</i>	3'UTR		
34510698	C	T	1	snp	<i>dot1l</i>	3'UTR		
34510781	CATG	CGTC,CGTG	4,1	complex,snp	<i>dot1l</i>	3'UTR		
34511070	G	T	1	snp	<i>dot1l</i>	3'UTR		
34512003	CA	TG,CG	2,1	mnp,snp	<i>dot1l</i>	Exon		
34512023	A	G	1	snp	<i>dot1l</i>	Exon		
34512038	G	A	1	snp	<i>dot1l</i>	Exon		
34512159	A	G	1	snp	<i>dot1l</i>	Exon		
34512701	G	A	1	snp	<i>dot1l</i>	Intron		
34512721	CAG	TAT	3	complex	<i>dot1l</i>	Intron		
34524877	G	A	1	snp	<i>dot1l</i>	Intron		

Table S5. SNPs and small indels found between males and females in the sex-determining *oaz1-dot1l* region on the Y haplotype specific to the Koka population

Position	Reference	Alternative	Length	Type	Gene	Intron / Exon	Change of amino acid	Found in literature
34490339	G	A	1	snp	<i>oaz1</i>	Intron		
34491022	G	T	1	snp	<i>oaz1</i>	Intron		
34493089	G	C	1	snp	<i>oaz1</i>	Intron		
34493108	G	A	1	snp	<i>oaz1</i>	Intron		
34493560	T	C	1	snp	<i>oaz1</i>	Intron		
34497130	G	T	1	snp	LOC100707471	Intron		
34498890	G	T	1	snp	LOC100707471	Intron		
34499005	G	A	1	snp	LOC100707471	Intron		
34499106	G	A	1	snp	LOC100707471 <i>amh</i>	Exon / 3'UTR		
34500441	T	G	1	snp	<i>amh</i>	Intron		
34500513	G	T	1	snp	<i>amh</i>	Intron		
34500704	TTA	TA	1	deletion	<i>amh</i>	Intron		
34501083	G	A	1	snp	<i>amh</i>	Exon 6		
34501186	C	A	1	snp	<i>amh</i>	Intron		
34501405	G	A	1	snp	<i>amh</i>	Intron		
34501472	C	A	1	snp	<i>amh</i>	Exon 4	Arg → Ser	
34501667	G	T	1	snp	<i>amh</i>	Intron		
34501748	GAAAAAAA AAACAAGT	GAAAAAAA GAAACAAG T	1	insertion	<i>amh</i>	Intron		
34501797	T	A	1	snp	<i>amh</i>	Intron		
34502670	G	A	1	snp	<i>amh</i>	5'UTR		
34503139	G	T	1	snp	Non coding region			
34504441	C	A	1	snp	Non coding region			
34509462	C	A	1	snp	<i>dot1l</i>	3'UTR		
34510327	C	T	1	snp	<i>dot1l</i>	3'UTR		
34510652	T	C	1	snp	<i>dot1l</i>	3'UTR		
34510677	A	G	1	snp	<i>dot1l</i>	3'UTR		
34510759	C	G	1	snp	<i>dot1l</i>	3'UTR		
34510948	G	A	1	snp	<i>dot1l</i>	3'UTR		
34511044	A	G	1	snp	<i>dot1l</i>	3'UTR		
34511089	T	A	1	snp	<i>dot1l</i>	3'UTR		
34511147	T	A	1	snp	<i>dot1l</i>	3'UTR		
34512360	C	A	1	snp	<i>dot1l</i>	Intron		
34512391	T	A	1	snp	<i>dot1l</i>	Intron		
34512745	A	C	1	snp	<i>dot1l</i>	Intron		
34512777	T	G	1	snp	<i>dot1l</i>	Intron		
34513830	G	A	1	snp	<i>dot1l</i>	Intron		
34514369	C	T	1	snp	<i>dot1l</i>	Intron		
34521953	C	A	1	snp	<i>dot1l</i>	Exon	Gly → Val	
34522348	T	G	1	snp	<i>dot1l</i>	Intron		
34524490	C	A	1	snp	<i>dot1l</i>	Intron		
34525095	G	A	1	snp	<i>dot1l</i>	Intron		
34525239	G	A	1	snp	<i>dot1l</i>	Intron		
34525597	A	T	1	snp	<i>dot1l</i>	Intron		
34528214	G	A	1	snp	<i>dot1l</i>	Intron		
34528383	C	T	1	snp	<i>dot1l</i>	Intron		

Analyses complémentaires

Certaines analyses complémentaires ont été effectuées en complément de celles présentées précédemment et ne figurent pas dans l'article publié. Ces analyses sont restées à un stade préliminaire et n'ont pas permis d'aboutir à des résultats suffisamment concluants pour donner lieu à une publication scientifique.

Cependant, certains aspects de ces études complémentaires donnent des indices supplémentaires sur les faits présentés dans l'article et sont donc présentés succinctement ci-après.

- **Analyse ddRAD préliminaire de la population naturelle Kou (Burkina-Faso)**

L'ensemble des échantillons utilisés pour ces analyses proviennent de collaborateurs du projet ANR Climsex, Aboubacar Toguyeni et Rokyatou Sissao, qui ont conduit l'ensemble des élevages et prélèvements de tissus ainsi que l'analyse de marqueurs PCR. Les banques de ddRAD ont été produites conjointement par Helena D'Cotta et Rokyatou Sissao. Le protocole utilisé est exactement le même que celui publié dans Triay et al., (2020).

Cette population est intéressante car d'après Sissao et al., 2019, le génotypage des individus par les marqueurs PCR de l'*amh* et l'*amhΔY* ne permet d'expliquer que près de 80% des sex-ratios obtenus dans les descendance après des croisements en conditions contrôlés. Il est attendu d'après Li et al., 2019 que la présence de l'*amhΔY* implique forcément celle de l'*amhY* puisque les deux copies sont en tandem. Cet agencement est également suggéré par les analyses que nous avons conduits pour ce chapitre et publiés dans Triay et al., 2020, même si elles sont en réalité sûrement séparées par des copies tronquées des gènes *dot11* et *oaz1*. Cependant, il existe des incongruences dans le génotypage des individus de la population Kou, et certains individus peuvent amplifier l'*amhΔY* mais pas l'*amhY* par exemple, laissant un doute sur le génotype sexuel réel de ces individus. Ces résultats surprenants semblent suggérer l'existence d'un second locus impliqué dans le déterminisme du sexe chez cette population (Sissao et al., 2019). Ainsi, nous avons utilisé exactement la même approche de ddRAD que dans Triay et al., 2020 sur une famille (D14) issue de la population Kou (Lac Kou, Burkina-Faso). Pour le séquençage ont été utilisés : les deux géniteurs dont une femelle XX et mâles XY d'après les génotypes avec les marqueurs PCR à l'*amh* ainsi que 58 individus de la descendance. Parmi les descendants, 6 mâles ont été

identifiés comme des mâles XX d'après ces mêmes marqueurs PCR. La méthodologie utilisée pour le traitement bioinformatique et la production des graphiques est exactement la même que celle publiée dans Triay et al., 2020.

Les résultats de ces analyses montrent à nouveau une association du sexe avec LG23 chez cette famille de la population Kou, et la majorité du signal est centrée autour de la position de *l'amh* (Figure 2.3). Cette analyse confirme donc que le locus principal du déterminisme du sexe, au moins dans cette famille, est bien le gène de *l'amh*, comme suggéré dans Sissao et al., 2019 et de même que dans de nombreuses souches et populations de tilapia du Nil à présent (Eshel et al., 2014; Li et al., 2015; Cáceres et al., 2019; Curzon et al., 2020; Triay et al., 2020; Taslima et al., 2021). Cependant, un second pic de SNPs spécifiques du sexe est observé sur LG19 (Figure 2.3). Ce signal est concentré sur une courte région (environ 1Mb) d'après les positions du génome de référence. Ce signal ressemble beaucoup en apparence à celui trouvé dans la population Kpandu sur LG3 dans Triay et al., 2020.

En effet, sur LG19 plusieurs SNPs montrent une forte association au sexe mais tous dans un seul locus de ddRAD : l'Olfactomedin-4-like (Figure 2.4). Il est en tandem avec une seconde copie de ce même gène et appartient à une grande famille multigénique (Olfactomedin). Ainsi, il est difficile de savoir si le signal observé sur LG19 est effectivement impliqué dans le déterminisme du sexe ou s'il s'agit en réalité d'un artefact dû à un mauvais assemblage des lectures sur cette région créant artificiellement du polymorphisme. Il est tout de même pertinent de noter que LG19 a été recruté au moins deux fois indépendamment comme chromosome sexuel avec un patron XX/XY chez des Cichlidés (Haplochromini/Tropheini) du lac Tanganyika (ElTaher et al., 2020). Cependant, les patrons de différenciation entre mâles et femelles sur LG19 chez ces cichlidés ne sont pas similaires à ceux observés chez la famille D14 de Kou puisque les régions décrites couvrent l'ensemble de 22Mb du début du chromosome ou la fin de LG19, ce qui n'est pas le cas ici.

La comparaison des génotypes et phénotypes de la descendance à leur géniteur ne permet pas d'éclaircir vraiment cette question (Figure 2.5). En effet, pour cette analyse ne sont gardés que les sites pour lesquels : i) la *p-value* du Test Exact de Fisher est inférieur à 0.05 ; ii) au moins la moitié des mâles phénotypiques et la moitié des femelles phénotypiques de la descendance ; iii) les deux géniteurs sont séquencés et suivent un patron XX/XY (site hétérozygote pour le mâle et fixé pour la femelle) ou inversement pour un patron ZZ/ZW.

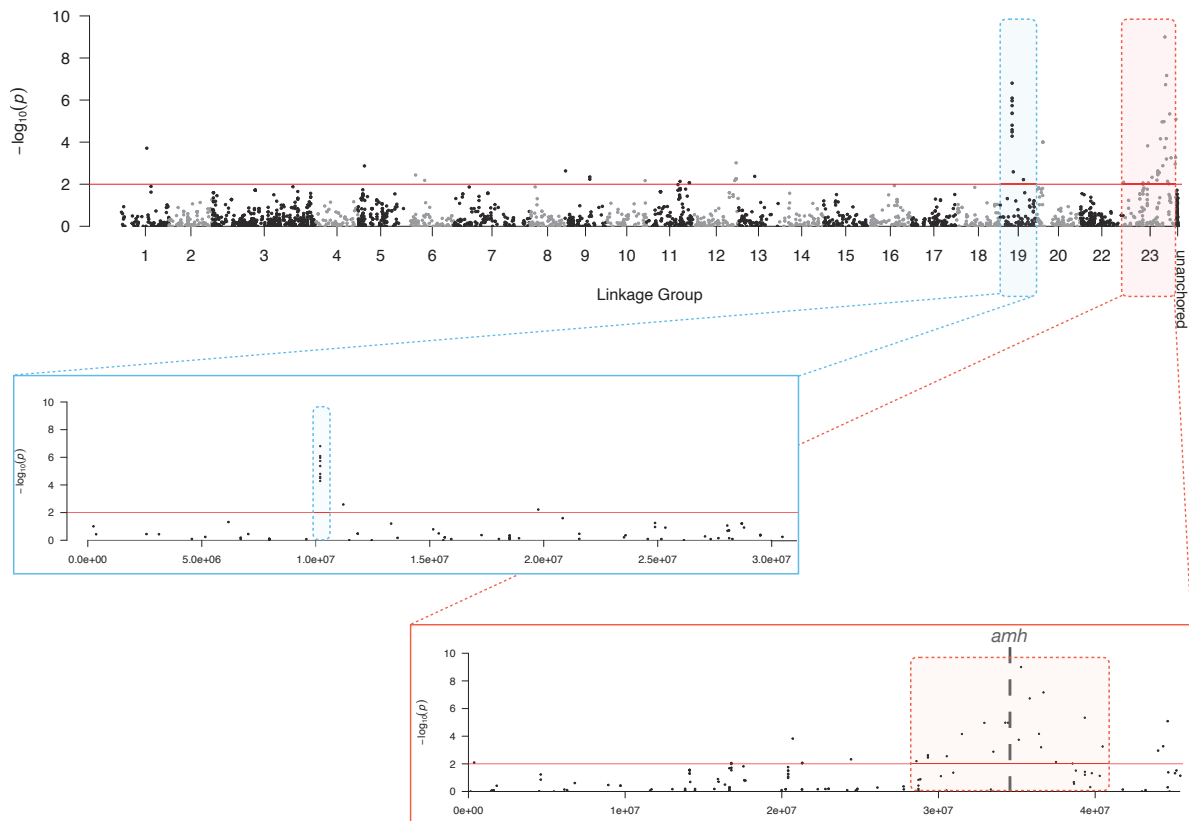


Figure 2.3. Manhattan plots des $-\log_{10}(p\text{-value})$ issues des Test Exact de Fisher pour l'association des polymorphisme nucléotidique et sexe phénotypique.

L'étiquette « Unanchored » réfère à l'ensemble des sites qui sont alignés sur les contigs non ancrés dans les principaux groupes de liaisons d'après le génome de référence. Sur le graphique présentant l'ensemble des groupes de liaisons (Linkage Group), deux groupes de liaisons portent des signaux d'une association au sexe : LG19 et LG23. Sur l'agrandissement de LG19, on constate que l'ensemble des SNPs très significativement liés au sexe (dans l'encadré en pointillé bleu) sont très proches les uns des autres. Ils appartiennent tous au même locus de ddRAD. Pour ce qui est de l'agrandissement de LG23, on constate une région montrant un fort dimorphisme entre mâles et femelles phénotypique entre 30 et 40Mb environ. Au centre de cette région est indiqué le gène de l'*amhY* d'après sa position dans le génome de référence. Les points situés au dessus de ligne rouge correspondent à une $p\text{-value} < 0,01$.

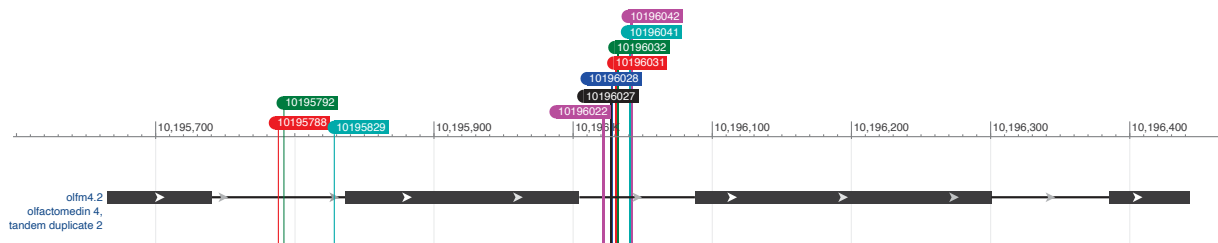


Figure 2.4. Emplacement des polymorphismes nucléotidiques simples (SNPs) significativement dimorphiques entre mâles et femelles sur LG19 d'après l'alignement sur le génome de référence d'*O. niloticus*.

La visualisation est effectuée à l'aide du Genome Browser de NCBI en utilisant la même version du génome de référence que celle utilisée pour l'alignement lors des analyses. Les sites fortement significatifs de LG19 appartiennent à un même locus de ddRAD correspondant au gène de l'Olfactomedin-4-like (seconde copie de la duplication en tandem de ce gène). Les SNPs prennent place dans les introns 1 et 2 de ce gène.

Or, pour LG19, seuls deux sites correspondent à ces critères, un pour le patron XX/XY et l'autre pour ZZ/ZW. Il est donc impossible de tirer des conclusions en se basant sur ces deux sites uniquement. En revanche, dans le cas de LG23, plus de sites sont disponibles. L'interprétation pour ce groupe de liaison n'est cependant pas beaucoup plus claire car pour le site le plus proche de l'*amh*, toutes les femelles possèdent le même génotype, et tous les mâles (sauf les supposés XX d'après les marqueurs de l'*amh*) possèdent également le même génotype ; seulement ceux-ci sont l'opposé des génotypes attendus. En effet, toutes les femelles portent le génotype hétérozygote du père alors que les mâles portent le génotype homozygote de la mère, et les mâles XX sont également porteurs du génotype paternel. Un tel phénomène paraît difficile à expliquer biologiquement et il est possible qu'une erreur de manipulation lors de la préparation de la banque soit en réalité la cause de ces patrons surprenants.

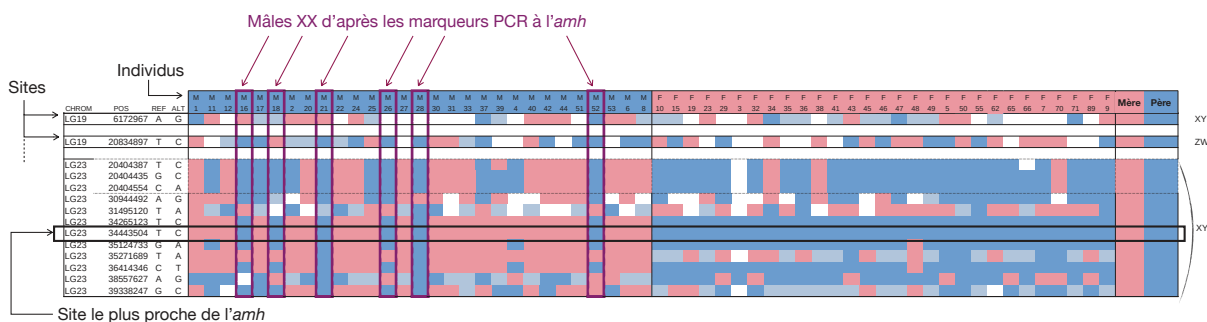


Figure 2.5. Génotypes de la descendance par rapport à leurs parents inférés à partir des sites significatifs du FET du ddRAD.

Les cases roses indiquent un individu ayant le même génotype que sa mère sur ce site tandis que les cases bleues foncées indiquent un individu ayant le même génotype que son père sur ce site et les cellules bleues claires indiquent un homozygote à l'allèle alternatif. Les cellules blanches correspondent à des données manquantes. Les sites conservés sont ceux pour lesquels i) la *p*-value du Test Exact de Fisher est inférieure à 0.05 ; ii) les deux parents sont séquencés et ils correspondent à un patron XX/XY pour LG23 et XX/XY ou ZZ/ZW pour LG19 ; iii) au moins la moitié des mâles et des femelles de la descendance ont été séquencés.

Comme nous n'avons pas eu le temps de re-séquencer une banque pour cette famille, nous n'avons pas prolongé les analyses au-delà de ce qui a été présenté ici. Cependant, cette analyse préliminaire confirme une forte association du sexe sur LG23 dans cette famille de la population Kou. Des analyses complémentaires seront nécessaires pour comprendre l'implication ou non de LG19 dans le déterminisme du sexe dans cette population. De plus, il s'agit ici seulement d'une famille et il est nécessaire de séquencer un plus grand nombre de géniteurs et de familles afin de pouvoir asseoir l'existence ou non d'un déterminant mineur,

qui pourrait notamment expliquer l'existence de mâles XX à l'*amh* alors que les individus ont été élevés en conditions contrôlées et à une température constamment inférieure à la température de masculinisation du tilapia du Nil.

- **Exploration de la composition du chromosome B**

Afin de mieux caractériser le chromosome B suspecté d'être présent dans le pool mâles Kpandu et le pool femelles Koka, des analyses complémentaires de biologie moléculaire ainsi que de bioinformatique ont été effectuées.

La première curiosité à laquelle nous voulions répondre était de savoir si ce que nous avons identifié comme un potentiel chromosome B présentait des régions similaires à celui des chromosomes B des cichlidés Africain déjà identifiés et notamment celui de *Metriaclima lombardoi* (qui agit comme un chromosome W, dominant sur le système XX/XY déjà existant). Pour ce faire, nous avons appliqué la méthodologie de détection des régions de chromosome B décrite dans Clark et al., 2018 qui se base sur les différentiels de couverture entre le pool d'individus contenant le chromosome B et celui ne le contenant pas. En utilisant un test binomial, et en tenant compte du nombre total de lectures dans les deux pools, le but est de définir la probabilité que le nombre observé de lectures alignés à une position soit seulement le fruit du hasard. Pour cette analyse, seules les régions de plus de 500kb ont été conservées. Cinquante et un segments du génome chez Koka et 54 segments du génome chez Kpandu ont été identifiés comme appartenant au chromosome B (Figure 2.6). Ces segments forment un total de 41 régions distinctes chez Koka et 23 régions distinctes chez Kpandu (en sachant que plusieurs segments peuvent appartenir à la même région). Parmi l'ensemble de ces régions identifiées chez Koka et Kpandu, 16 sont communes aux deux populations et présentent des caractéristiques extrêmement similaires de couverture et de fréquences alléliques des SNPs. Aucun de ces 16 locus identifiés ne correspond à des séquences du chromosome B de *Metriaclima lombardoi*.

Afin de pouvoir identifier combien d'individus sont porteurs de chromosomes B dans chacun des pools, des analyses de biologie moléculaires ont été menées. Malheureusement, malgré le design de nombreux couples de primers sur des régions spécifiques de nombreuses régions B spécifiques, nous n'avons pas réussi à obtenir un couple capable d'amplifier les régions correspondantes.

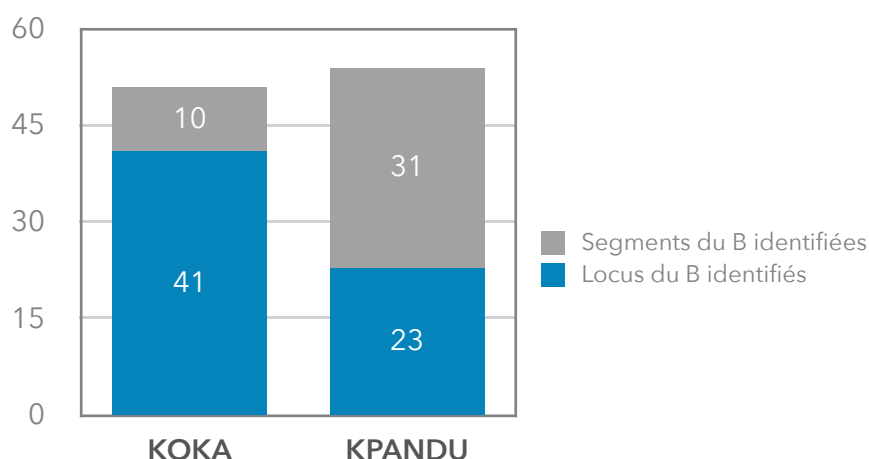


Figure 2.6. Diagramme en barres du nombre de segments et du nombre de locus identifiés comme appartenant au chromosome B dans le pool de femelles Koka et le pool de mâles Kpandu utilisés dans Triay et al., 2020.

Plusieurs segments peuvent appartenir à un même locus. Au total, 51 segments ont été identifiés chez Koka, appartenant à 41 locus différents et 54 segments appartenant à 23 locus différents ont été identifiés chez Kpandu

Ce problème est notamment dû au fait que les copies présentes sur le chromosome B et la version originelle sur le A peuvent avoir accumulé beaucoup de divergences et rendent fastidieux le design de primers. De plus, le chromosome B, s'il est en une seule copie comme chez les autres cichlidés africain, ne peut pas recombinaison et profiter du balayage sélectif. Ainsi, ce chromosome est sujet à l'accumulation de mutations et de variants structuraux qui peuvent être différents entre tous les paralogues d'un même gène sur ce chromosome. Or, en utilisant les méthodes d'alignement sur un génome de référence, l'ensemble des lectures de toutes les différentes copies viennent s'aligner sur l'unique copie existante dans le génome de référence et créent ainsi un polymorphisme nucléotidique qui reflète en réalité les SNPs trouvés possiblement sur différentes copies de gènes sur le B. De plus, les locus identifiés comme appartenant au B et partagés entre les deux populations peuvent correspondre à des régions répétées du génome et être confondus avec des segments de chromosome B authentiques. Il est donc laborieux de designer des primers permettant d'amplifier des régions spécifiques du chromosome B et la caractérisation en détails de cet élément génétique demandera des optimisations à la fois de la méthodologie bioinformatique pour détecter plus de régions possiblement appartenant au chromosome B ainsi que des optimisations à la paillasse pour essayer d'identifier les individus du pool porteur du B et ainsi, peut-être se donner une idée de la prévalence de ce chromosome B dans les populations sauvages.

Références

- Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Travisan, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., et al. (2019). Fine mapping using whole-genome sequencing confirms anti-müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3 (Bethesda)* 9, 3213–3223.
- Clark, F.E., Conte, M.A., and Kocher, T.D. (2018). Genomic characterization of a B chromosome in Lake Malawi cichlid fishes. *Genes* 9, 610.
- Curzon, A.Y., Shirak, A., Dor, L., Zak, T., Perelberg, A., Seroussi, E., and Ron, M. (2020). A duplication of the Anti-Müllerian hormone gene is associated with genetic sex determination of different *Oreochromis niloticus* strains. *Heredity* 125, 317–327.
- El Taher, A., Ronco, F., Matschiner, M., Salzburger, W., and Böhne, A. (2020). Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes (Evolutionary Biology). Unpublished. Available at <http://biorxiv.org/lookup/doi/10.1101/2020.10.23.335596>
- Eshel, O., Shirak, A., Dor, L., Band, M., Zak, T., Markovich-Gordon, M., Chalifa-Caspi, V., Feldmesser, E., Weller, J.I., Seroussi, E., et al. (2014). Identification of male-specific amh duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 15, 774.
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genetics* 11, e1005678.
- Sissao, R., D’Cotta, H., Baroiller, J.-F., and Toguyeni, A. (2019). Mismatches between the genetic and phenotypic sex in the wild Kou population of Nile tilapia *Oreochromis niloticus*. *PeerJ* 7.
- Taslima, K., Khan, M.G.Q., McAndrew, B.J., and Penman, D.J. (2021). Evidence of two XX/XY sex-determining loci in the Stirling stock of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 532, 735995.
- Triay, C., Conte, M.A., Baroiller, J.-F., Bezault, E., Clark, F.E., Penman, D.J., Kocher, T.D., and D’Cotta, H. (2020). Structure and Sequence of the Sex Determining Locus in Two Wild Populations of Nile Tilapia. *Genes* 11, 1017.

CHAPITRE 2 : DIVERSITÉ ET PERTE DE L'HAPLOTYPE Y ANCESTRAL EN POPULATIONS NATURELLES

Objectifs

Dans le chapitre précédent, nous avons conclu que le déterminisme du sexe chez le tilapia du Nil en population sauvage est régi par une région dupliquée sur LG23 contenant deux copies du gène de l'*amh*. Comme les deux populations étudiées précédemment sont géographiquement éloignées, nous avons émis l'hypothèse que ce déterminisme sur LG23 Y soit ancestral à la divergence des populations de l'Est et de l'Ouest de l'Afrique. Pour ce chapitre l'objectif était donc de mieux caractériser l'haplotype LG23 Y. Pour passer outre les limites des méthodes précédemment utilisées, nous avons utilisé des séquençages longs fragments (Long Reads) afin de traverser l'ensemble de la structure complexe. De plus, comme le déterminisme est *a priori* ancestral, nous avons choisi d'ajouter une seconde population Ethiopienne à nos analyses, pour laquelle les marqueurs PCR de l'*amh* n'étaient pas fiables.

Définir au mieux la structure de la région du déterminisme du sexe dans les populations sauvages (similarité et différences) est essentiel pour pouvoir identifier des régions propices au design de primers dans les populations sauvages et appréhender au mieux l'histoire évolutive de ce déterminisme LG23-Y.

Introduction et contexte

Chez le tilapia du Nil, les analyses du déterminisme du sexe sur les souches domestiques (commerciales ou de laboratoires) identifient le déterminisme du sexe sur différents groupes de liaisons selon la souche étudiée. Ainsi, des associations génétiques au sexe phénotypique ont été trouvées sur LG1 chez une souche commerciale des Etats-Unis ou encore sur la souche Manzala-Stirling de l'Université de Stirling (Ecosse) par exemple (Gammerdinger et al., 2014; Palaiokostas et al., 2015; Taslima et al., 2021). En revanche, aucun locus précis n'a été proposé pour comme déterminant sur ce LG. D'autre part, des associations génétiques au sexe ont été montrées sur LG23 dans la région du gène de l'Hormone Anti-Müllerienne (*amh*) sur des stocks commerciaux d'Amérique Latine, sur la souche Manzala-Stirling (en parallèle de LG1, mais en plus faible fréquence), sur la souche GIFT, la souche Japonaise ou encore les souches israélienne et belge originellement issues de la souche Manzala-Stirling (Eshel et al., 2014; Li et al., 2015; Wessels et al., 2017; Cáceres et al., 2020; Taslima et al., 2020). Une forte association au sexe est également montrée dans la population naturelle Kou du Burkina-

Faso, mais un second déterminant non identifié semble co-ségréger pour déterminer le sexe (Sissao et al., 2019). Nous avons montré dans nos analyses précédentes de deux populations naturelles que le déterminisme était lié à la présence d'une duplication du gène de *l'amh* sur LG23-Y (Triay et al., 2020), comme suggéré sur la souche Japonaise par Li et al., (2015). Cependant, sur la souche Japonaise, les deux copies de *l'amh* sur le Y sont divergentes, et l'une d'entre elle présente un indel qui change le cadre de lecture et provoque un codon stop prématuré (Li et al., 2015). Si dans la population sauvage de Kpandu (Ghana) nous avons bien identifié cette copie tronquée de *l'amh*, en revanche, chez la population Koka (Ethiopie) les deux copies semblent être relativement conservées (Triay et al., 2020). De plus, dans cette étude précédente, nous avons pu définir les limites de cette duplication mâle-spécifique sur LG23-Y. Cependant, si les séquençages de génomes complets en lectures courtes nous ont permis de proposer un haplotype Y attendu, des doutes subsistaient sur la structure de cette région, et notamment quant à l'ordre successif des deux copies de *l'amh* ou encore la position des variants structuraux identifiés dans les gènes dupliqués.

Matériels et méthodes

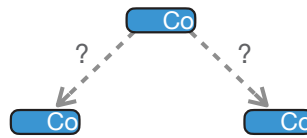
Afin de préciser la structure de la duplication spécifique des mâles sur LG23-Y, nous avons de nouveau étudié les deux populations naturelles utilisées auparavant : Kpandu (Lac Volta, Ghana) et Koka (Lac Koka, Ethiopie). Comme les mâles de la population Koka présentent la duplication dans la région liée au sexe sur LG23-Y alors que les marqueurs PCR à *l'amh* ne permettaient pas de la mettre en évidence, nous avons ajouté une seconde population Ethiopienne (Hora, Lac Hora) à notre analyse avec *l'a priori* de détecter à nouveau cette duplication.

Trois séquençages de génomes complets en lectures longues (MinION Oxford Nanopore Long reads) ont été effectués sur un mâle (F0) de chacune des populations sauvages. Cette technologie permet d'obtenir des lectures suffisamment longues pour traverser plusieurs régions qui resteraient litigieuses avec des lectures courtes. Comme nous nous attendions à une duplication dans la région du sexe, utiliser des lectures longues permet de reconstruire l'ordre des gènes sans faire face aux difficultés rencontrées dans Triay et al., 2020 (Figure 3.1). Afin d'optimiser les temps d'assemblages, nous avons concentré notre analyse uniquement sur la région du déterminisme du sexe identifiée préalablement sur LG23. Ainsi, un filtre a

été appliqué sur les lectures pour ne conserver que celles qui étaient composées des gènes identifiés précédemment dans les haplotypes X et Y de LG23. Ces reads ont ensuite été assemblés de deux manières : alignés à un génome de référence, soit l'haplotype Y de référence (d'après l'haplotype Y attendu de Triay et al., 2020), soit le génome de référence (XX) ; et assemblés de manière *de novo* afin de reconstruire la séquence du Y sans *a priori* sur sa structure.

A)

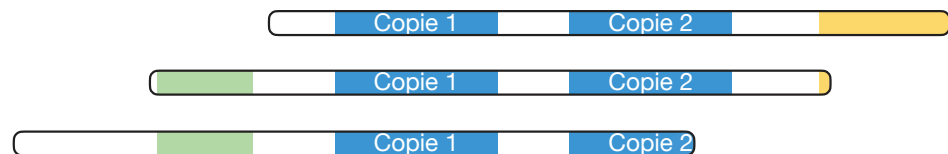
Lectures courtes



Région avec duplication



Lecture longues



B)

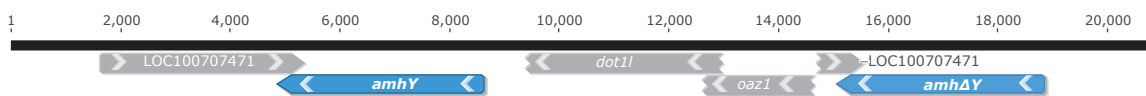


Figure 3.1. Intérêt du séquençage lectures longues (long reads) pour correctement assembler une région dupliquée.

A) Une lecture courte peut être alignée sur les deux copies dupliquées. Les lectures longues à l'inverse traversent la totalité de la région dupliquée, ne laissant pas de doute quant à l'ordre des séquences. B) Exemple d'une lecture longue d'un mâle Koka séquence pendant cette thèse, contenant les deux copies du gène de l'*amh* et permettant de définir la structure de la région du sexe sur LG23-Y.

Aucun haplotype Y n'a pu être reconstruit avec ces méthodes pour le mâle Hora utilisé (*cf.* Principaux résultats et discussion). Ce résultat surprenant pouvait être dû à 2 facteurs : soit le mâle utilisé est un mâle XX spontané du milieu naturel, puisque nous avons utilisé des mâles F0 ; soit les mâles Hora ne présentent pas (ou pas systématiquement) la duplication de la

région de l'*amh* sur LG23. Afin de pouvoir élucider ce fait, nous avons également produit un séquençage de génome complet en lecture courtes pour pouvoir exploiter la méthode exploratoire déjà utilisée dans Triay et al., 2020 et vérifier si l'on observe un signal spécifique du sexe sur LG23 entre un pool de mâles phénotypiques et un pool de femelles phénotypiques (tous F0). Les analyses bioinformatiques conduites sur ces données sont les mêmes que celles publiées dans Triay et al., 2020, c'est à dire un alignement du génome complet sur le génome de référence, puis une identification des sites polymorphiques et une comparaison des signaux entre le pool de mâles et de femelles pour rechercher des régions dimorphiques entre les deux sexes.

Enfin, des analyses phylogénétiques nucléaires, mitochondriales et spécifiques à la région du sexe chez les femelles ont été réalisées. Pour ce faire, nous avons utilisé les séquençages WGS du chapitre précédant ainsi que les lectures courtes de Hora produites pour ce chapitre afin de définir les relations de parentés entre les trois populations.

Principaux résultats et discussion

i. Reconstruction de la duplication dans la région du sexe sur LG23-Y

Nous avons pu reconstruire l'haplotype Y chez les deux mâles Kpandu et Koka. Ces résultats ont confirmé la structure attendue du Y décrite précédemment en détaillant l'ordre des gènes au sein de cette duplication ainsi que les copies des gènes portant les indels mâles spécifiques identifiés dans Triay et al., 2020.

D'une part ces résultats confirment à nouveau l'existence d'un système XX/XY sur LG23 avec une duplication du gène de l'*amh*. De plus, ils permettent de constater que les mâles Kpandu ont, dans cette région, une structure très similaire à celle décrite par Li et al., 2015, avec la présence d'une copie conservée de l'*amh* (*amhY*) et une copie tronquée (*amhΔY*). Chez Koka en revanche, les deux copies de l'*amh* semblent conservées. Nous avons pu confirmer également que l'*amhΔY* se situe en amont de l'*amhY* (dans le sens de la transcription), séparés par une chimère composée de copies tronquées des gènes *oaz1* et *dot11*. Cependant, la manière dont cette région régit le déterminisme du sexe reste incomprise car nous n'avons pas pu exploiter la présence de variants nucléotidiques du fait d'une trop faible couverture des données de MinION.

ii. Une perte du déterminant ancestral (LG23 XX/XY) chez les mâles Hora

Nous avons pu mettre en évidence que les mâles de la population Hora ne présentent pas la duplication de la région de l'*amh* sur LG23-Y, contrairement aux mâles des populations Koka et Kpandu. De plus, aucun polymorphisme nucléotidique n'a été identifié entre les mâles et les femelles sur l'*amh* suggérant que cette région ne détermine pas le sexe chez la population Hora. Les analyses phylogénétiques ont pourtant confirmé que les individus Hora et Koka sont plus proches entre eux qu'ils ne le sont des individus de la population Kpandu. Ce résultat était attendu puisque les populations Hora et Koka appartiennent à la sous-espèce *O. niloticus cancelatus* et sont proches géographiquement alors que la population Kpandu est une population d'*O. niloticus niloticus* et que des analyses de données de microsatellites insinuaient déjà de telles relations (Bezault et al., 2011). Le fait que les mâles Koka et Kpandu présentent le déterminant LG23-Y suggère que ce déterminant est ancestral à la divergence de ces populations. Ainsi, nous avons interprété l'absence du déterminisme LG23-Y chez Hora comme une perte de ce déterminant ancestral.

ii.a. LG23, une paire de chromosomes sexuels « ever-young »

Si le déterminisme LG23 XX/XY de Kpandu et Koka est certainement ancestral à la divergence des populations, les haplotypes X et Y ne sont pour autant pas très divergents. Nous avons montré que la différence entre les deux chromosomes pour la région spécifique du sexe est de moins de 10kb. Ce résultat, ainsi que l'observation de recombinaisons proches de la région de l'*amh* dans les données de ddRAD du premier chapitre (Triay et al., 2020) semblent indiquer que les chromosomes X et Y sont encore très similaires. Dans l'article nous discutons donc de la possibilité qu'il s'agit de chromosomes sexuels qui peu importe l'âge n'entament pas de processus de différenciation (arrêt de la recombinaison, diminution du contenu en gène etc...), et nous proposons quelques mécanismes et processus qui pourraient mener à un tel phénomène.

ii.b. Pas d'indice du déterminant ayant pris le relais

Aucun autre groupe de liaison n'a pu être identifié comme potentiel chromosome sexuel chez Hora avec les analyses menées jusqu'à présent. Nous discutons donc dans cet article deux questions : pourquoi la population Hora a-t-elle connu un turnover du déterminisme du sexe ? Et quel est le système actuel régissant le déterminisme du sexe dans cette population ? Même si nous n'avons pas de piste concrète actuellement, les possibilités

écologiques et évolutives pouvant mener à un changement de déterminisme du sexe dans cette population sont envisagées. Nous discutons également des limites méthodologiques qui pourraient expliquer que nous n'ayons pas trouvé de signal clair sur un autre groupe de liaison avec les approches utilisées dans l'article.

ii.c. Diversité intraspécifique du déterminisme du sexe en milieu naturel

La diversité du déterminisme du sexe dans les souches domestiques de tilapia du Nil et maintenant la démonstration de sa variabilité dans les populations naturelles avec les publications de Sissao et al., (2019) et Triay et al., (2020 et actuelle) démontrent l'importance d'étudier les systèmes de déterminisme à l'échelle des populations. En effet, comme suggéré par Furman et al., (2020), la diversité des chromosomes sexuels est peut-être parfois aussi importante au niveau intraspécifique qu'interspécifique.

iii. Une introgression entre *O. niloticus* et son espèce sœur *O. aureus*

Enfin, nos analyses phylogénétiques ont permis de mettre en évidence une incongruence entre le signal nucléaire et le signal mitochondrial. La position de la population de l'Ouest Africain (Kpandu) varie entre les arbres : les données mitochondriales rapprochent les individus Kpandu d'*O. niloticus* de l'espèce sœur *O. aureus*. Dans la phylogénie nucléaire en revanche les trois populations sauvages de *O. niloticus* forment un groupe monophylétique. Ces résultats vont dans le sens de ceux de Rognon & Guyomard (2003), qui avaient émis l'hypothèse d'une introgression complète du génome mitochondrial de *O. aureus* vers *O. niloticus* dans les populations d'Afriques de l'ouest et des lacs égyptiens.

Conclusion

Dans cet article nous avons défini en détails la structure (taille, ordre des gènes, positionnement des indels identifiés préalablement comme spécifiques des mâles) de la région du déterminisme du sexe chez le tilapia du Nil sur LG23-Y. Nous avons également établi les relations de parenté entre les trois populations sauvages, suggérant une origine ancestrale du déterminisme XX/XY de LG23. Curieusement, nous avons montré que les deux populations éthiopiennes, malgré qu'elles soient proches parents ne présentent pas le même système de déterminisme que les populations Koka et Kpandu pourtant plus éloignées phylogénétiquement. Nous interprétons cette absence de LG23-Y chez Hora

comme une perte du déterminant ancestral, mais n'avons cependant pas pu identifier de locus ni même de groupe de liaison en lien avec le sexe chez cette population.

Enfin nous avons confirmé une introgression mitochondriale entre *O. aureus* et *O. niloticus*.

Références

- Bezault, E., Balaesque, P., Toguyeni, A., Fermon, Y., Araki, H., Baroiller, J.-F., and Rognon, X. (2011). Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genetics* 12, 102.
- Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Travisany, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., et al. (2019). Fine mapping using whole-genome sequencing confirms anti-müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3 (Bethesda)* 9, 3213–3223.
- Carrasco, L.A.P., Penman, D.J., and Bromage, N. (1999). Evidence for the presence of sex chromosomes in the Nile tilapia (*Oreochromis niloticus*) from synaptonemal complex analysis of XX, XY and YY genotypes. *Aquaculture* 173, 207–218.
- Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., D'Hont, A., Baroiller, J.-F., D'Cotta, H., Penman, D.J., et al. (2008). Genetics of sex determination in tilapiine species. *Sexual Development* 2, 43–54.
- Eshel, O., Shirak, A., Weller, J.I., Hulata, G., and Ron, M. (2012). Linkage and physical mapping of sex Rregion on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3: Genes, Genomes, Genetics* 2, 35–42.
- Eshel, O., Shirak, A., Dor, L., Band, M., Zak, T., Markovich-Gordon, M., Chalifa-Caspi, V., Feldmesser, E., Weller, J.I., Seroussi, E., et al. (2014). Identification of male-specific amh duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 15, 774.
- Furman, B.L.S., Metzger, D.C.H., Darolti, I., Wright, A.E., Sandkam, B.A., Almeida, P., Shu, J.J., and Mank, J.E. (2020). Sex chromosome evolution: so many exceptions to the rules. *Genome Biol Evol* 12, 750–763.
- Gammerdinger, W.J., Conte, M.A., Acquah, E.A., Roberts, R.B., and Kocher, T.D. (2014). Structure and decay of a proto-Y region in Tilapia, *Oreochromis niloticus*. *BMC Genomics* 15, 975.
- Jalabert, B., Moreau, J., Planquette, P., and Billard, R. (1974). Déterminisme du sexe chez tilapia macrochir et tilapia nilotica : action de la méthyltestostérone dans l'alimentation des alevins sur la différenciation sexuelle; proportion des sexes dans la descendance des mâles «inversés». *Annales de Biologie Animale Biochimie Biophysique Vol. 14*, 729–739.
- Lee, B.-Y., Penman, D.J., and Kocher, T.D. (2003). Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics* 34, 379–383.
- Lee, B.-Y., Coutanceau, J.-P., Ozouf-Costaz, C., D'Cotta, H., Baroiller, J.-F., and Kocher, T.D. (2011). Genetic and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus*). *Mar Biotechnol (NY)* 13, 557–562.
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y

- chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genetics* 11, e1005678.
- Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2013). Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS ONE* 8, e68389.
- Palaiokostas, C., Bekaert, M., Khan, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2015). A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 16, 171.
- Pfennig, F., Standke, A., and Gutzeit, H.O. (2015). The role of Amh signaling in teleost fish – Multiple functions not restricted to the gonads. *General and Comparative Endocrinology* 223, 87–107.
- Poonlaphdecha, S., Pepey, E., Huang, S.-H., Canonne, M., Soler, L., Mortaji, S., Morand, S., Pfennig, F., Mélard, C., Baroiller, J.F., et al. (2011). Elevated amh Gene Expression in the Brain of Male Tilapia (*Oreochromis niloticus*) during Testis Differentiation. *Sex Dev* 5, 33–47.
- Rognon, X., and Guyomard, R. (2003). Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. *Mol Ecol* 12, 435–445.
- Sissao, R., D'Cotta, H., Baroiller, J.-F., and Toguyeni, A. (2019). Mismatches between the genetic and phenotypic sex in the wild Kou population of Nile tilapia *Oreochromis niloticus*. *PeerJ* 7.
- Taslina, K., Wehner, S., Taggart, J.B., de Verdal, H., Benzie, J.A.H., Bekaert, M., McAndrew, B.J., and Penman, D.J. (2020). Sex determination in the GIFT strain of tilapia is controlled by a locus in linkage group 23. *BMC Genet* 21, 49.
- Taslina, K., Khan, M.G.Q., McAndrew, B.J., and Penman, D.J. (2021). Evidence of two XX/XY sex-determining loci in the Stirling stock of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 532, 735995.
- Triay, C., Conte, M.A., Baroiller, J.-F., Bezault, E., Clark, F.E., Penman, D.J., Kocher, T.D., and D'Cotta, H. (2020). Structure and Sequence of the Sex Determining Locus in Two Wild Populations of Nile Tilapia. *Genes* 11, 1017.
- Wessels, S., Krause, I., Floren, C., Schütz, E., Beck, J., and Knorr, C. (2017). ddRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC Genomics* 18, 531.

PUBLICATION ASSOCIÉE

Polymorphism of Y haplotypes or loss of ancestral LG23 sex determinant in three wild populations of Nile tilapia

Cécile Triay ^{1,2}, **Maxime Courcelle** ¹, **Pierre Caminade** ¹, **Etienne Bezault** ³, **Jean-François Baroiller** ^{1,2}, **Thomas D. Kocher** ⁴ & **Helena D'Cotta** ^{1,2}

1. UMR-Institut des Sciences de l'Evolution de Montpellier, Centre National de la Recherche Scientifique, Institut de Recherche pour le Développement, Ecole Pratique des Hautes Etudes, University of Montpellier, 34090 Montpellier, France; cecile.triay@umontpellier.fr (C.T), maxime.courcelle@umontpellier.fr (M.C); baroiller@cirad.fr (J-F.B.); dcotta@cirad.fr (H.D)
2. UMR-Institut des Sciences de l'Evolution de Montpellier, Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Campus Int. Baillarguet, 34398 Montpellier, France
3. UMR BOREA, CNRS-7208/MNHN/UPMC/IRD-207/UCN/UA, Université des Antilles, 97159 Guadeloupe, France; etienne.bezault@univ-antilles.fr
4. Department of Biology, University of Maryland, College Park, MD 20742, USA; tdk@umd.edu (T.D.K.)

Abstract: Sex-determining regions have been identified on linkage groups (LG) 1 and 23 in the Nile tilapia, depending on the domesticated strains used. Sex determining studies on wild populations are extremely scarce. We previously worked on two wild populations originating from Lake Volta (Ghana) and from Lake Koka (Ethiopia) and showed that LG23 also carries the sex-determining region in these wild populations, and more precisely they have a Y tandem duplication containing two copies of the Anti-Müllerian Hormone *amh* gene (named *amhY* and *amhΔY*). Here, we extend our focus on this Y chromosome duplicated region in the two populations of Lake Volta and Lake Koka and added a new Ethiopian population to the analysis, from Lake Hora. Using long reads of whole genome sequencing of a male from each population, we confirmed the duplicated structure, its breakpoints as well as genes' order on the Y chromosome of LG23 in both Lake Volta and Lake Koka populations. We did not, however, identify the LG23 Y haplotype in the Hora male. A whole-genome short-reads analysis, carried out between pools of phenotypic males and females of Hora confirmed the absence of a Y *amh* duplication on LG23 Y. The phylogenies conducted on the nuclear and mitochondrial genomes, nevertheless show a closer relationship between the two Ethiopian populations compared to the Ghanaian population. In lake Hora population the absence of the LG23Y sex-determination region is interpreted as a loss. The potential reasons for this loss and the implications it may have on the evolution of sex chromosomes are discussed in this paper.

Keywords: Nile tilapia; Sex chromosomes; *amh*; Duplication; Y-haplotype

1. Introduction

Sex determination (SD) initiates the development of males or females. Genetic sex determination (GSD) inherited through sex chromosomes has been highly conserved and stable in mammals and birds for millions of years (Cortez et al., 2014). In mammals, sex chromosomes are heteromorphic and follow a XX/XY system. The degenerated Y chromosome carries in most mammals the *Sry* gene, the master gene that triggers the development of male gonads (Sinclair et al., 1990). In birds, ZW embryos become females. The W chromosome is also small in carinate birds but testis develop in the presence of a double dose of the *dmrt1* gene on the Z chromosome (Smith et al., 2009; Graves, 2014). However, SD stability does not appear to be the norm in vertebrates (Bachtrog et al., 2014). Reptiles, amphibians and fish have a remarkable diversity of sex chromosomes, heterogamety and sex-determination genes found even among closely related species (Ezaz et al., 2009; Myosho et al., 2015; Jeffries et al., 2018). In some of these vertebrates, sex can be triggered by environmental cues (Environmental Sex Determination ESD) such as temperature (*e.g.* reptiles), pH or even density, or through a combination of both GSD and the environment which is the case in many teleost fishes (Mank and Avise, 2009; Baroiller and D'Cotta, 2016).

Sex chromosomes do not evolve like autosomes. They undergo differential sexual forces in the rates of mutations, selection and recombination (Bachtrog et al., 2011). The classical canonical model based mostly on the XY system of mammals considers that a sex chromosome originated on an autosome pair that acquired a sex-determining mutation. This sex-locus, particularly when close to male-beneficial alleles that are detriment for females (sexually antagonistic alleles), is expected to spread in the population because of their Y linkage and therefore, transmitted to sons and not daughters (Rice, 1987; Charlesworth et al., 2005; van Doorn and Kirkpatrick, 2010). Suppressed recombination will isolate this sex-limited region, which expands by mechanisms such as accumulation of mutations, inversions, repetitive elements, deletions and insertions (Kirkpatrick, 2010; Bachtrog, 2013; Blaser et al., 2014). Over time the differences between the X and Y are such that they become morphologically distinct, reaching a degree of degeneration where turnover is no longer possible (Bachtrog, 2013; Abbott et al., 2017).

In contrast, 90% of teleost fishes have homomorphic sex chromosomes (Devlin and Nagahama, 2002), implying that recombination suppression is probably recent. Differences between sexes can be reduced to a single missense SNP such as in pufferfish (Kamiya et al., 2012). The identification of undifferentiated sex chromosomes which have similar gene content is however a challenging process. Genomic sequencing has shed light in fish lineages evidencing the diversity of sex systems and sex chromosomes. A striking characteristic of teleost sex chromosomes is their rapid turnover. There are transitions in the sex chromosome system (Roberts et al., 2009; El Taher et al., 2020) even within the same genus (Cnaani et al., 2008; Takehana et al., 2008; Böhne et al., 2019). Furthermore, many sex-determining genes have been conserved but translocated into different chromosomes

(Guiguen et al., 2018) or have emerged in a short evolutionary time with switches in the sex chromosome pair (Gammerdinger and Kocher, 2018). In ricefishes, the sex-determination gene *dmY* (also called *dmrt1bY*) was found on Linkage Group (LG) 1 in the medaka *Oryzias latipes* (Nanda et al., 2002), whereas in *O. luzenensis* it is a Y-copy of the *gsdf* gene located on LG12 (Myosho et al., 2012) and in another *Oryzias* it is *sox3* on LG10 (Takehana et al., 2014). Turnovers in sex chromosomes are remarkable in the large Cichlid family that has undergone adaptive radiation, where over 20 sex-determinants have been identified on 17 different LGs as well as on B chromosomes, with the turnover driven possibly by sexual antagonistic selection (Gammerdinger and Kocher, 2018; El Taher et al., 2020). Some sex chromosomes appeared multiple times (e.g. LG5, LG7 and LG19) (Böhne et al., 2019; El Taher et al., 2020). Differentiation of the sex chromosomes can be small, or comprise the whole LG which suggests widespread recombination suppression (El Taher et al., 2020). Some “usual suspects” are found amongst the master genes and they tend to play musical chairs to endorse the role of major sex-determinant (Böhne et al., 2016).

The high diversity observed in teleost fish (approximately 27,000 living species) has been proposed to be related to the genomic plasticity of this clade, matching their extensive ecology (Helfman et al., 2009). Fishes have a higher rate of evolution and gene duplications than other vertebrates (Mank and Avise, 2009). Indeed, gene duplications are a substrate for the evolution of innovations by giving the duplicated copy (paralog) the possibility to complete the function of the original gene (subfunctionalization), or by acquiring a new function (neofunctionalization) (Sandve et al., 2018). Several sex-determining genes of teleosts have emerged from duplications of sex-differentiating downstream genes. The duplication of a gene already involved in the sex determination pathway can thus result in a new locus that can potentially carry mutations significantly increasing the fitness of one of the sexes and thus become a major determinant (Mank and Avise, 2009; Ortega-Recalde et al., 2020). *Dmy* was duplicated from the male-differentiating *dmrt1* gene found on the autosomal LG9, that was translocated to LG1 in the medaka (Matsuda et al., 2002; Nanda et al., 2002). In the Patagonian pejerrey (*Odontesthes hatcheri*) and the Northern pike (*Esox lucius*) duplicated copies of the *amh* gene have taken on the role of master sex-determinant (Hattori et al., 2012; Pan et al., 2019). Duplication of genes outside the sex pathway can also result in a major sex-determinant such as the *sdY* gene in salmonids that is a neofunctionalized paralog of an immune-related gene (*irf9*) (Yano et al., 2012).

The Nile tilapia is a XX/XY farmed cichlid where the GSD can be overridden by temperature generating XX males (Baroiller et al., 2009). This has complicated the finding of the sex chromosomes added by the use of intermixed domesticated strains. In these strains sex has been linked either to LG1 (Cnaani et al., 2008; Gammerdinger et al., 2014; Palaikostas et al., 2015) or LG23 (Eshel et al., 2014; Li et al., 2015; Wessels et al., 2017; Cáceres et al., 2019) or both (Taslima et al., 2021). No causal gene has been found on LG1. In the Japanese domesticated strain of Nile tilapia, the *amh* gene on LG23 has been identified as the sex-determination gene, which evolved through a tandem duplication within the same LG. Hence, males have three copies of the *amh* gene, two copies located in tandem on the Y

chromosome, one with a missense SNP (*amhY*) and the other theoretically with an insertion that causes a premature stop codon (*amhΔY*), and the third copy on the X chromosome (Li et al., 2015). We found that sex in various wild populations was also linked to LG23 and to the *amh* gene (Sissao et al., 2019; Triay et al., 2020). The Y chromosome sex-region which contains this *amh* duplication was estimated to be ~51 kb using short reads from Whole Genome sequencing (WGS) and comprised the adjacent *oaz1* and *dot1l* genes (Triay et al., 2020). However, we were not able to know how the duplication was inserted because there were two possible scenarios and the order of the genes within the duplication. We also found major differences within the duplicated region and in the *amhΔY* copy between the two populations compared with domesticated strains. *AmhΔY* in the Ethiopian Koka population seemed to lack the 5 bp insertion which causes premature stop codon (Triay et al., 2020). In addition, the *amh* markers we used for genotyping the sex did not discriminate between XX and XY individuals in the Ethiopian population (Sissao et al., 2019; Triay et al., 2020).

Molecular and genomic analyses have helped to better detect sex linked loci and identify sex chromosomes, particularly in facilitating the reconstruction of Y or W chromosomes (Abbott et al., 2017). Calibration and choice of different methods of new generation sequencing (NGS) technologies are aiming to better detect sex chromosomes depending on the divergence between the two sex chromosomes, from strongly heteromorphic to homomorphic (Palmer et al., 2019). However, the reconstruction of a duplicated region with short reads is tricky if two homologous paralogs exist, and even more when working with reads that are shorter than the similar regions.

Long reads, if they allow a better description of a duplicated structure, are still prone to sequencing error with rates up to 5-15% (Carneiro et al., 2012; Menegon et al., 2017). To counter this issue and still be able to interpret SNPs, the coverage has to be high enough to compensate for uncertainties and sequencing errors. Moreover, usual assembly tools are not developed specifically for sex chromosomes and thus neither mapping methods nor *de novo* assemblies are optimized to reconstruct regions specific to the heterogametic sex.

In this study, we wanted to define more accurately the structure and sequences of the duplicated sex-region on the Y chromosome, and compare the differences between populations, which would also allow sexing of Ethiopian wild individuals for population genetics. To build the Y haplotypes, we performed long reads Nanopore sequencing on males from two East African populations (originating from Lake Koka and Lake Hora in Ethiopia) and one West African population (originating from Lake Volta in Ghana). We expected the Hora male to be similar to the Koka male. However, the Lake Hora male appeared to lack the Y *amh* duplication. This could be due to this male being an XX male. Therefore, we performed WGS on several Hora males and females to study the SD and see whether males lacked the *amh* duplication on LG23. We confirmed that Hora males did not have the *amh* duplication thus looked for new sex-association in the whole genome of these samples.

2. Material and Methods

2.1. Sample collection

2.1.1. Fish samples and DNA extractions

Three wild populations of Nile tilapia *Oreochromis niloticus* were caught in 2002 and 2003 from the Sudano-Sahelian basin in West Africa and the Ethiopian Rift Valley basin of East Africa (Bezault, 2005). Lake Koka and Lake Hora populations belong to the endemic *O. niloticus cancellatus* subspecies and were collected from two cold-temperature (between 17°C and 26°C) reservoirs of the Awash River in the Ethiopian highlands. Lake Koka is a large artificial reservoir of the Awash River of 255 km² with an average depth of 9 m while Lake Hora is an isolated lake of about 1 km² with a depth of up to 35 m. The Kpandu population on the other hand belong to the *O. niloticus niloticus* subspecies captured in Ghana from a dentritic expansion of the eastern side of Lake Volta (8,482 km² with an average depth of 19 m) where temperatures fluctuate between 27°C to 32°C. Fully mature fish were collected using several seine net captures and the phenotypic sex was then defined. Fish were anaesthetized with 2-phenoxyethanol to sample the fin clips which were preserved in 96-100% ethanol.

High-molecular weight genomic DNA (gDNA) was lysed from the fin clips in TNES-Urea buffer (4 M urea; 10 mM Tris-HCl, pH 7.5; 125 mM NaCl; 10 mM EDTA; 1% SDS) and extracted using the phenol-chloroform procedure (Barker, 1998). The gDNA was stored in Tris-EDTA buffer at -20°C. DNA was quantified first on a NanoDrop spectrophotometer (ThermoFisher, France) to estimate the absorbance ratios and the concentration was then measured on a Qubit 2.0 fluorometer (Invitrogen Carlsbad, USA). gDNA quality was also verified on a 0.8% agarose gel.

2.1.2. Ethics Statement

Fish captures complied with the veterinary laws of each country as stipulated in 2003. All animal and experimental procedures were performed in accordance with the French protocol N°2016101810463 and the French Government's authorization for animal experimentation N° 35-15.

2.2. Whole genome short reads sequencing analysis of Hora

2.2.1. Libraries & sequencing

Twenty WGS libraries were prepared at Macrogen (Seoul, Korea) from ten males and ten females' gDNA which were sonicated to obtain 550 bp fragments. All libraries were prepared using the TruSeq DNA PCR-Free kit (Illumina) except one male library where

TruSeq Nano DNA Kit (Illumina) was used due to limited amounts of gDNA. Libraries were 2 x 150 bp paired-end sequenced on a Illumina NovaSeq 6000 (Flowcell S4).

2.2.2. Data processing

Raw data quality was checked using FastQC (0.11.7). Illumina TruSeq adapters were trimmed and short reads (< 15bp) were removed using Fastp (version 0.21.0). Samples were pooled according to their phenotypic sex, resulting in a pool of females (with 10 samples) and a pool of males (with 10 samples). The two pools were mapped to the latest *O. niloticus* reference genome available on NCBI (O_niloticus_UMD_NMBU, Genbank ID: GCA_001858045.3) using BWA mem mapper (version 0.7.15). Sam files were then converted to Bam files and sorted using SAMtools (version 1.9) and read groups were added to the two pools using Picard AddOrReplaceReadGroups (2.7.0). The output files were then combined with SAMtools mpileup (version 1.9). Using Popoolation2 mpileup2sync.pl (version 1.2.2), the mpileup file was transformed to a synchronized file (sync). Finally, the sync file was processed with Sex_SNP_finder_GA.pl to obtain *Fst* between males and females. This pipeline allows to identify sites that are fixed (or nearly fixed) in one sex and at intermediate frequencies in the other sex. The following parameters were used in Sex_SNP_finder: *fixed_threshold* = 0.9; *minimum_polymorphic_frequency* = 0.3; *maximum_polymorphic_frequency* = 0.7; *minimum_read_depth* = 10; *maximum_read_depth* = 100; *minimum_read_count* = 2; *sex_SNP_finder_window_size* = 10000. Finally, the *Fst* were plotted on a manhattan plot using R and the qqman package (version 0.1.4).

2.2.3. Library construction for MinION sequencing

One Kpandu male, one Koka male and one Hora male were used for the MinION sequencing (Oxford Nanopore Technology). The male genomic DNA libraries were constructed with 3 to 4 µg of high molecular weight gDNA by MGX (Montpellier, France) using the Ligation Sequencing Kit 1D from (ONT). Samples were fragmented, the extremities repaired and adapters ligated. The flow cell FLO-MIN106 (flow cell 9.4.1) was charged twice, with a DNase treatment performed on the flow cell to reactivate the pores when half of the library was sequenced.

2.2.4. Subsetting of read datasets and identification of Y specific reads

Using the results of the Illumina short reads (2 x 150 bp) whole genome sequencing published in Triay et al., (2020), we built the theoretical Y haplotype for the LG23 sex region (Figure 1). This theoretical haplotype is based on the X haplotype from the reference genome sequences (an XX female), duplicated according to the breakpoints described in Triay et al., (2020). We voluntarily did not include the small deletions (<500bp) found using the short reads WGS data, as we expect long reads to be long enough to cross the deleted regions of more than 500bp so that we can detect and thus confirm these « small » deletions. Although

some of the long reads are less than 2Kb in length, we included the two large (>5kb) deletions to obtain the final sequence of the theoretical Y.

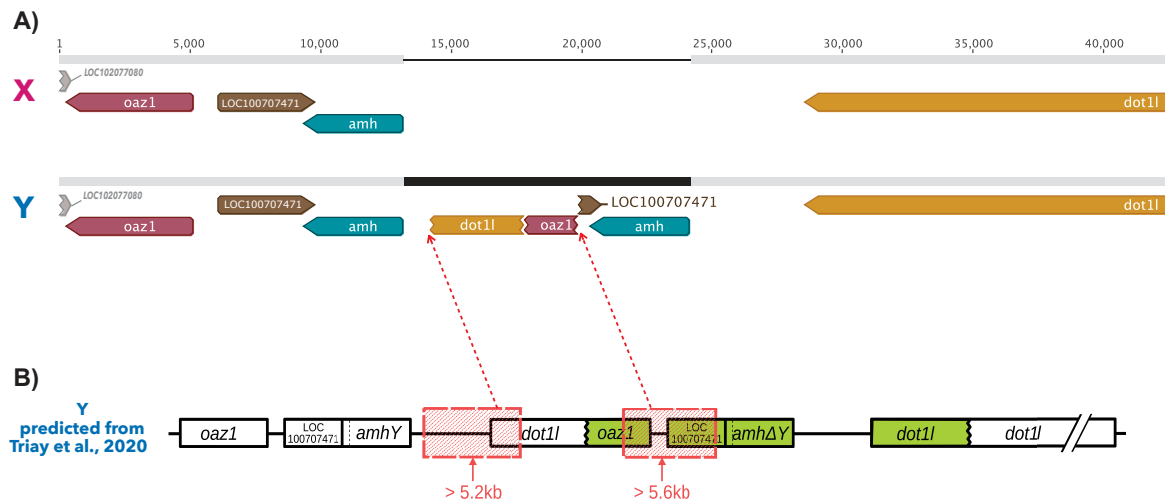


Figure 1. Alignment of both the X haplotype and the Y theoretical haplotype built according to the descriptions of the tandem duplication in Triay et al., (2020). A) Geneious alignment of both the X haplotype from the reference genome and the Y theoretical haplotype. Grey bars represent homologous sequences between the X and the Y haplotypes whereas the black region in the Y haplotype represents Y specific regions that appeared due to a duplication of the *dot1l*, *oaz1*, LOC100707471 and *amh* genes. A chimera of the *dot1l* and *oaz1* genes is seen as the result of large deletions that occurred after the duplication. The second copy of the *amh* gene however is well conserved in this inserted region. B) Y haplotype prediction published in Triay et al., (2020) with the position of the large deleted regions (in red boxes) removed from the tandem duplication to build the Y of the alignment.

Oxford Nanopore MinION raw reads were used to build a BLAST database against which were blasted the sequences of *oaz1*, *amh* and *dot1l* genes extracted from the latest available *O. niloticus* XX female reference genome (O_niloticus_UMD_NMBU, Genbank ID: GCA_001858045.3). Since the first intron of the *dot1l* gene is very similar to the sequence of the unrelated *vasa* gene, we split the sequence of *dot1l* in order to remove this intron and limit its influences in blast results.

All raw reads for which at least one of the sequences presented blast hits were kept. This subset of reads was then trimmed for 23bp at 3' and 5' extremities, corresponding to regions with low quality bases.

This set of trimmed reads was mapped onto the theoretical Y haplotype in order to identify Y specific reads according to the presence of Y specific indels described in Triay et al., (2020). We mainly focused on the deletion of about 270bp in length found only in *oaz1* known to be Y specific in both Kpandu and Koka males (Triay et al., 2020). All reads presenting this deletion were thus labelled as « Y-specific » reads. Conversely, all other reads were labelled as « Non Y-specific ». The sequential steps for this process are illustrated in Figure 2.

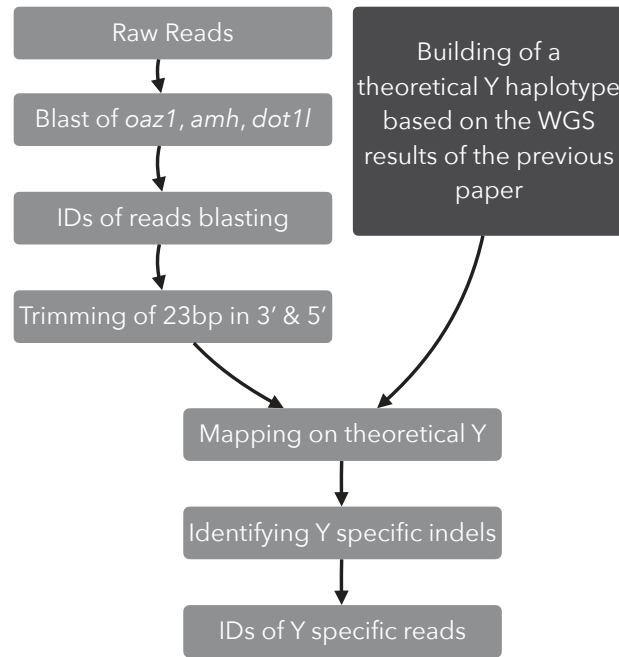


Figure 2. Schematic pipeline designed to label Y specific reads.

2.3. Assembly of subset of reads

Reads of each subset (All Reads, Y specific, Non Y-specific) were *de novo* assembled using Canu (version 2.1.1). The resulting contigs were automatically annotated using Geneious (version 9.0.5), based on the annotation of the reference genome for the defined region. Contigs that could not be annotated were checked using BLAST (version 2.11.0) and were removed when not belonging to the LG23 region of interest. This was the case mainly for the « Non Y-specific » subset of reads and when all reads were used to build *de novo* assemblies. This is a downside of the BLAST approach used to identify genes belonging to the LG23 sex region, as it led us to keep some reads of the *oaz1b* gene (a paralog of *oaz1*) and *vasa* (of which part of the sequence is similar to the *dot1l* gene). However, contigs containing those genes were easily removed from the final set of contigs.

Contigs that were confirmed to belong to LG23 were mapped to the *O. niloticus* XX female reference genome and to the Y theoretical haplotype using the Geneious assembler (version 9.0.5) with high sensitivity settings to check for the haplotype they correspond to and whether the structure is the expected one.

In order to increase the contiguity of the Koka Y haplotype, we did a pairwise alignment of only the Y-specific reads using MAFFT with a scoring matrix of 100PAM/K=2 and an automatically chosen algorithm. Among the Y-specific reads, some contained both the *amh* copies or both the *dot1l* copies for example (Figure 3). Thus, the pairwise alignment allowed us to compute a consensus sequence with the Y specific reads, increasing the length of the Y haplotype contig for the Koka male, even though the whole region could not be constructed

using long reads assembly softwares. Unfortunately, this method of consensus sequence with a low coverage of reads results in many ambiguities at several sites (Figure 3).

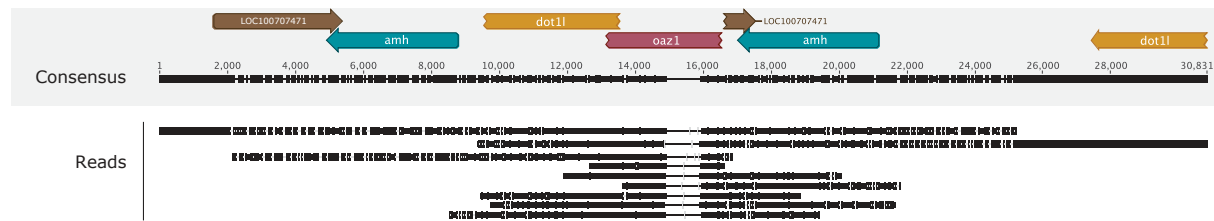


Figure 3. Pairwise alignments of reads identified as Y specific. This method was used to create the consensus sequence in Geneious. The consensus was annotated using Geneious automated annotations based on the annotations of the reference genome of *O. niloticus*.

2.4. Alignments and phylogeny

2.4.1. Nuclear phylogeny

The 42 loci used by Meyer et al., (2015) to construct their cichlid phylogeny were retrieved in the trimmed data of our wild populations using HMMER (version 3.3.2) (Supplementary Table 1). In order to maximize accuracy of nucleotide sequences comparisons, we decided to use only the Illumina short reads data obtained from Hora in the current study, and those for Koka and Kpandu from our previous study (Triay et al., 2020). Because a B chromosome is suspected to be present in the WGS short reads from the female pool of Koka as well as in the male pool of Kpandu individuals, we selected for the phylogeny the Koka male pool, the Kpandu female pool and the Hora female pool. We did not expect strong bias between sexes. The targeted regions of the mapping alignments were converted from BAM file to consensus fasta files after a SNP Calling using bcftools (version 1.9). For the two outgroups *Etia nguti* and *Hemichromis elongatus*, and for *O. niloticus* reference genome, we used the fasta files available on NCBI (Genbank IDs : GCA_015106755.1, GCA_015108515.1 and GCA_001858045.3 respectively). When needed, we produced a reverse complement in order to obtain all sequences in identical sense to the sequences of Meyer et al., (2015). The forty two sequences of all samples were then combined into a unique fasta and we added some of the species of Meyer et al., (2015) in order to check if relationships between these species were identical to their paper. All sequences were then aligned using MAFFT (v7.475) and sites presenting more than 50% missing data among all species/samples, usually corresponding to species-specific insertions, were removed using trimAl (version 1.4.rev22). Finally, the Maximum-Likelihood phylogeny was produced using IQtree (1.6.12) with 100 bootstraps for node support. The sequential steps to produce this phylogeny are shown in Figure 4.

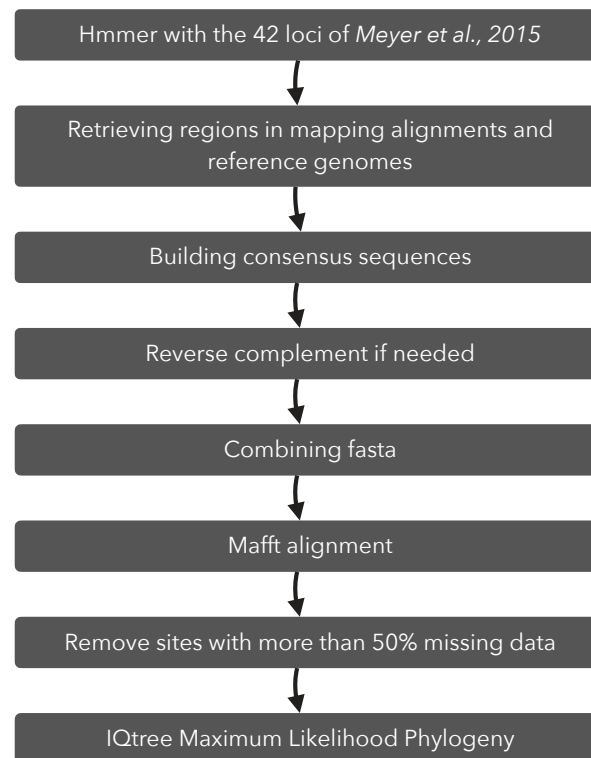


Figure 4. Schematic representation the sequential steps used to produce the nuclear phylogeny.

2.4.2. Mitochondrial phylogeny

We assembled the mitochondrial genomes from Koka, Hora and Kpandu *O. niloticus* wild populations obtained from the short reads of our Whole Genome Sequencing trimmed data using MitoFinder (version 1.4) (Allio et al., 2020). We used the reference mitochondrial genome as the basis for annotation (Genbank ID: GU238433.1). In addition, we used the closely related species *Haplochromis burtoni* as outgroup (NCBI ID: NC_027289.1). In order to use the same basis as the one used for gene annotation, we also ran the reference mitochondrial genome and *Haplochromis burtoni* through Mitofinder to extract individual genes. The sequences of the 13 mitochondrial protein-coding genes and the two ribosomal subunits (Supplementary Table 2) were then aligned using MAFFT (v7.475) (Katoh and Standley, 2013). A maximum-likelihood phylogeny was inferred from the concatenated sequence alignments using IQtree 1.6.12 (Nguyen et al., 2015) with 100 bootstraps for node support.

2.4.3. Phylogeny of Sex haplotypes

Because of the low coverage of long reads data and due to the high error rate of Oxford Nanopore MinION data, we could not use directly the Y haplotypes built from the *de novo* analyses. Thus, we focused on the *amh* gene region and more precisely on the LG23 region

from 34,489,620 to 34,532,693 bp according to the XX reference genome and thus only analysed females individuals. This corresponds to the region defined as carrying the major sex-determinant in previous studies (Li et al., 2015; Cáceres et al., 2019; Triay et al., 2020; Taslima et al., 2021). This region was extracted in Kpandu, Koka and Hora females mapping files using SAMtools (version 1.9) (Li et al., 2009). These regions were then converted to consensus sequences using Geneious (version 9.0.5) (Kearse et al., 2012). The sequences, along with fasta files of this region from *O. niloticus* and *O. aureus* were aligned using MAFFT (v7.475) (Kato and Standley, 2013). The subsequent alignments were then used to build a phylogeny of Females X haplotypes with 100 bootstraps with IQtree (1.6.12) (Nguyen et al., 2015). The final visualization of all the trees (nuclear, mitochondrion and X haplotypes) were formatted using Geneious (version 9.0.5) (Kearse et al., 2012).

3. Results

3.1. Identification of Y specific reads in the LG23 duplicated sex-region

3.1.1. Mapping of long reads confirms the existence of Y specific structures in Kpandu and Koka but not in Hora natural populations

For the Kpandu male we obtained a total of 4,474,365 Nanopore reads corresponding to 22,830,560,109 nucleotides. A total of 5,266,388 reads corresponding to 14,776,650,374 nucleotides were obtained for the Koka male and a total of 777,765 reads for the Hora male corresponding to 4,034,783,321 nucleotides were used. The average error rate was 6,44%. This resulted in 228 of Kpandu, 287 of Koka and 236 reads of Hora that passed the Blast filter using the *oaz1*, *amh* and *dot11* genes (Table 1).

The coverage is different along this region for the three samples. If it drops from the end of the left *amh* copy in the Kpandu, it appears to be almost constant along the region for the Koka male and finally approaches 0 in the Y specific region in the Hora male (Figure 5). This result indicates the theoretical Y haplotype structure we constructed is correct since reads are mapping with high quality to the Y haplotype, confirming the structure previously described using WGS short reads data for the Kpandu and Koka populations (Triay et al., 2020).

Coverage drop over the Y specific structures (such as seen over the *dot11-oaz1* chimera, e.g. dashed arrows in Figure 5) in Kpandu is due to the mapping of only male-specific reads, whereas, reads of both X and Y chromosomes are mapping over the *amh* gene, common to both X and Y. In the Koka male assembly, the coverage pattern might be explained by the absence of an X chromosome, thus Y reads show a good mapping all along the region, corresponding to the constant coverage seen over and outside the Y-specific regions. However, we still observe some reads that present a bad mapping quality. They could be sequencing artefacts, or the sequenced male could be an XY male with low coverage for the X resulting in the visual aspect homogeneous coverage along the region.

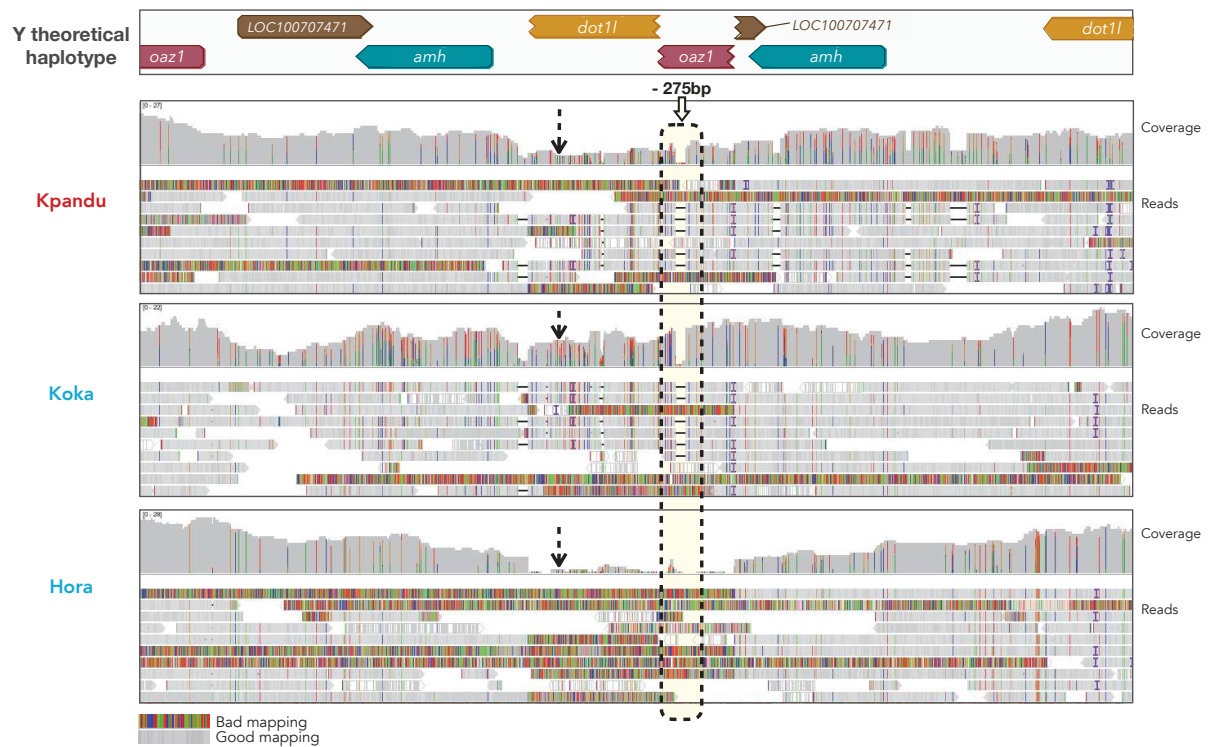


Figure 5. IGV snapshot of long reads mapping obtained from the three males Kpandu (Lake Volta, Ghana), Koka (Lake Koka, Ethiopia) and Hora (Lake Hora, Ethiopia). Reads were mapped onto a Y theoretical haplotype. Dashed arrows are showing the coverage level over the beginning of the Y specific region. Reads mapping to the *oaz1-dot1l* chimera presenting a 275bp deletion found in the *oaz1* truncated copy (within the dashed rectangle) were labelled as Y-specific.

The structure of the Y haplotype was confirmed for Koka and Kpandu males, but this was not the case for the Hora male. This male individual presents no reads with a good mapping score over the Y specific regions. Moreover, it presents none of the expected Y specific indels (such as the -275 bp in the truncated copy of the *oaz1* gene). Either this male could be an XX male or males of the Hora population do not carry any tandem duplication in this region. This raised the question of the *amh* being the sex-determining region in the Hora population and lead to the short reads whole genome sequencing of several other samples of this population.

3.1.2. Indels usage to identify Y specific reads

Reads presenting the previously described deletion in the truncated copy of *oaz1* gene of ~275 bp were labelled as Y specific, whereas those not showing it were presumed to belong to the X chromosome. This step was carried out only for Kpandu and Koka males as the Hora male did not present the expected Y haplotype. This resulted in 15 long reads for Kpandu and 11 for Koka labelled as Y-specific (Table 1 and Figure 5).

Table 1. Number of reads in the output of the filtering pipeline for MinION long reads sequencing.

	Reads passing the filtering and blast steps	Reads labeled as Y haplotypes
Kpandu	228	15
Koka	287	11
Hora	236	–

3.2. Assembly of sex haplotypes

3.2.1. *De novo* assemblies and sex haplotypes constructions

In order to get the Y haplotype and study it in more detail, our first approach was to do a *de novo* assembly of the long reads from the sex-region. We obtained very different results depending on the male's origin (Figure 6). We compared the obtained contigs to the *O. niloticus* XX reference genome (female from the Stirling-Manzala strain). The automatic annotation of contigs based on these two references allowed us to label the different contigs as X or Y haplotypes based on the presence or not of Y specific duplicated structures.

We observed only one long contig for the Hora male, fitting with an X haplotype. No Y haplotype was assembled, as was expected according to the mapping results where no reads presented Y-specific structures. On the contrary, two contigs were assembled for the Kpandu male: one perfectly corresponding to the X reference genome haplotype, and another one presenting the Y-specific structures. The second contig however is much smaller than the X haplotype and is centered around the *oaz1-dot11* chimera and a second copy of the *amh* gene (corresponding to the *amh* Δ Y described in the Japanese strain (Li et al., 2015)). Finally, three contigs were built from the Koka reads. The longest contig carries Y-specific structures, such as the *dot11-oaz1* chimera and the Y-specific indels corresponding to the duplicated region. However, the first copy of the *amh* gene (which is the *amh*Y in the Japanese strain) is not as fully reconstructed as for the Kpandu male. Two smaller contigs contain respectively the conserved version of the *oaz1* gene and the beginning of the *amh*Y gene. No clear X haplotype was built from the *de novo* using all reads for the Koka male. This absence of an X contig is in line with the observed coverage of the mapping on the Y theoretical haplotype. *De novo* assemblies of « Non-Y specific » reads did not produce any contig long enough to be associated to the X haplotype either. Moreover, the longest contig assembled using the non Y-specific reads was identified as belonging to the Y haplotype as part of the chimera that could be identified with the automated annotation. As described in section « 2.4. Subsetting of read datasets and identification of Y specific reads », the longest Y haplotype produced for the Koka male was processed through pairwise alignment of Y-specific reads only. This haplotype was used for further analyses.

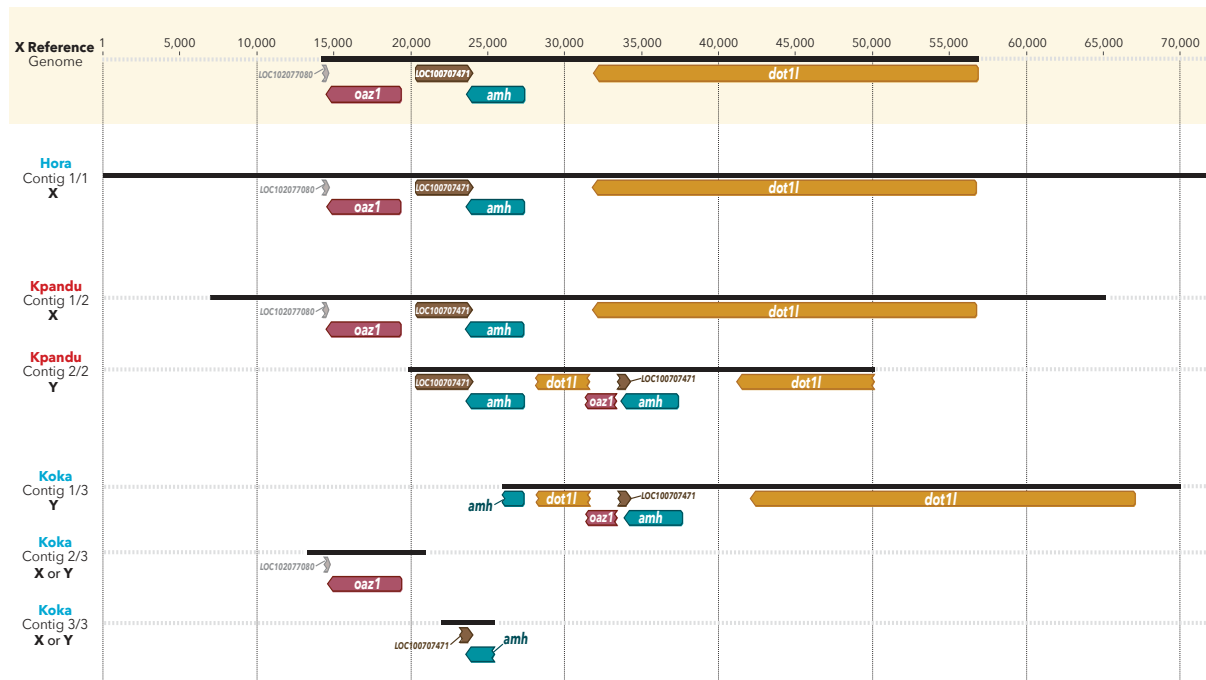


Figure 6. Contigs output of the Canu *de novo* assembly of the three males belonging to the wild populations of Hora, Koka and Kpandu. Koka and Hora (in blue) are Ethiopian males from lakes of the same names, whereas Kpandu (in red) is originating from Lake Volta in Ghana. All contigs are visually aligned to the *oaz1* gene of the reference genome, based on Geneious automated annotations from the XX reference genome and from the theoretical Y haplotype built previously. X and Y labels of contigs are deduced according to the structure of contigs, presenting or not the *dot1l-oaz1* chimera, specific to the Y haplotype as described in Triay et al., (2020).

3.2.2. Limits of using the long reads data of Kpandu and Koka

The alignment of X sex haplotypes (LG23 *amh* region) obtained with the WGS short reads data were aligned along with the Y sex haplotypes of the same region produced with the MinION Long Reads. The goal of this alignment was to use it to build a sex-haplotype phylogeny. In this alignment, X is presenting exactly the same structure with minor SNPs differences between each population. However, a large structure corresponding to the tandem duplication of *oaz1-amh-dot1l* genes is shared between the two Y haplotypes and confirms once again the existence of the Y structure as suggested in the built theoretical Y. A close up of this Y specific region helps to understand the misleading visual alignments. Even though the regions do not seem to align correctly between the two Ys, they actually are for the most part similar. This visual artefact is due to the low coverage with which the Koka Y haplotype was built, leading to several ambiguities or uncertainties in the consensus sequence (as shown in the blue highlights in Figure 7). This situation is extended to the whole region for the Koka Y haplotype, such as the loci shared with the X. This highlights very well the limits of our assembly methods. The low coverage of our data does not allow to produce a confident haplotype, corrected for sequencing errors. Thus, the Koka Y

haplotype is very informative about the duplication but it is on the other hand mostly useless for population genetic and phylogenetic analyses.



Figure 7. Alignments of all sex haplotypes. X were constructed using the whole genome sequencing data after a mapping on the reference genome. Y haplotypes were built from the Long reads data using Canu *de novo* alignments and pairwise alignments. The annotations are obtained using Geneious automatic annotations based on the annotations of the reference genome of *O. niloticus*. The lower part of this figure is a closeup of the alignment of both Y haplotypes of Kpandu and Koka males. The Koka Y haplotype is showing sites that are ambiguous (K, R, S, T) and sites with N indicating uncertain nucleotides.

3.3. Absence of sex-association on LG23 and no strong sex-pattern across the whole genome in Hora

Our initial long read analyses on a Hora male showed that the Y haplotype was absent. Either this could be due to no Y haplotype being present in Hora males and therefore questioning the role of the *amh* as a sex-determinant in this population, or alternatively it could be explained if the Hora male was in fact an XX individual. To answer these questions we performed a short reads WGS of Hora males and females to search for a strong sex-determinant pattern between sexes through a Genome wide association study (GWAS). We obtained from this sequencing between 108,570,962 and 129,028,644 total reads representing a mean of 17X per sample. For our assembly, after having pooled all the females' reads and all the males' reads, we used a random subset of 20% of the reads for each pool.

The *Fst* Manhattan plot of Hora (Figure 8) does not present any obvious pattern on LG23 in contrast to what we found from our Koka and Kpandu previous results (Triay et al., 2020). We only found one SNP between males and females identified in a non-coding region in the *oaz1* to *dot1l* region. However, we did not see any reliable signal on other linkage groups for this population either. There were three regions showing elevated sex-patterns peaks in LG5. A focus on these regions did not allow us to identify any sex-candidate gene in the genome annotations (Supplementary Table 3). The reason for this high *Fst* region might be the numerous paralogs encountered that could lead to mismapping with short reads analyses.

Although the signal is overall confusing, we did not find any of the 16 shared regions of high coverage that potentially belong to a B chromosome, which we found in Kpandu males and Koka females. Thus, the cause of this noise along the genome is not easy to explain and further study of males-females comparisons will be needed to conclude what are the causes of the high F_{st} peaks along the genome.

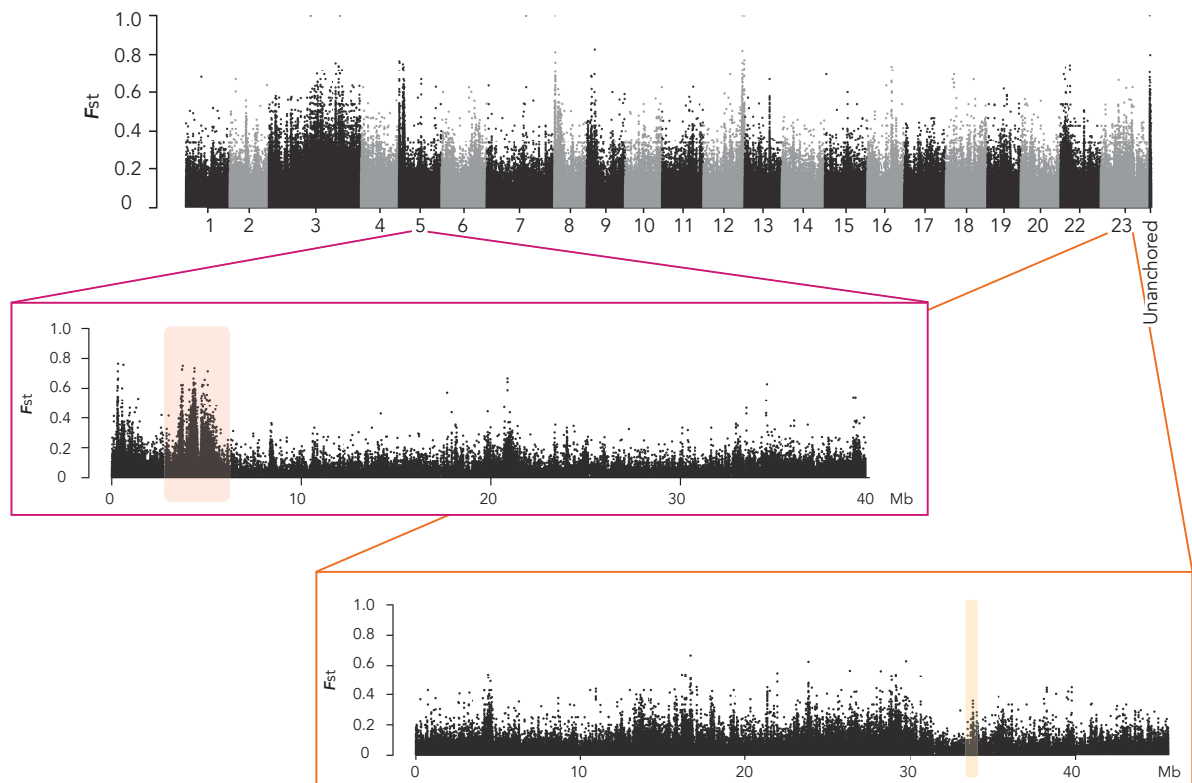


Figure 8. Manhattan plots of F_{st} for sites fixed or nearly fixed for one sex and at intermediate frequencies for the other sex in Hora population. A) Plot for all Linkage Groups. Unanchored contigs are grouped into a single label. A close up of LG05 present a region of divergence between males and females highlighted, that was not convincing after a check for genes in this region. The close up of LG23 only has the *amh* region highlighted. No particular high peak is observed in this region where there were expected.

3.4. Comparing sex haplotypes to population phylogeny

Both the nuclear and the mitochondrion phylogenies show that the two Koka and Hora, populations from Ethiopia, are more closely related to each other than to the Kpandu population originating from Ghana (Figure 9). These results are coherent with their subspecies taxonomy as *Oreochromis niloticus cancellatus* for Koka and Hora versus *Oreochromis niloticus niloticus* for the Kpandu population as well as the geographical location of these three populations (Trewavas, 1983; Bezault et al., 2011). The individual of *O. niloticus* used for the reference genome is placed as a sister clade of the Ethiopian populations in the mitochondrion phylogeny, whereas the Kpandu male is placed as a sister group in the nuclear phylogeny. The mitochondrion phylogeny also strongly supports a sister clade

relationship between the Kpandu *O. niloticus* population and the *O. aureus* sample. The X haplotypes of Ethiopian populations are also forming a monophyletic group, with the Ghanaian Kpandu population as a sister group. In X haplotypes, the divergence between Hora and Koka females is hardly noticeable.

The position of the reference genome is changing in the three trees as it forms a monophyletic clade with Kpandu in the X haplotype phylogeny, but it is the sister group of the Ethiopian populations in the Mitochondrion phylogeny and finally it is placed as a sister group of Koka, Hora and Kpandu in the nuclear phylogeny. The reference genome is showing a longer branch than all other samples in the three trees, with it being the longest branch of the tree in the nuclear phylogeny, apart from the outgroups.

These phylogenies results are however surprising considering no Y haplotype could be constructed in the Hora male using long reads sequencing, even if this structure is shared between Kpandu and Koka.

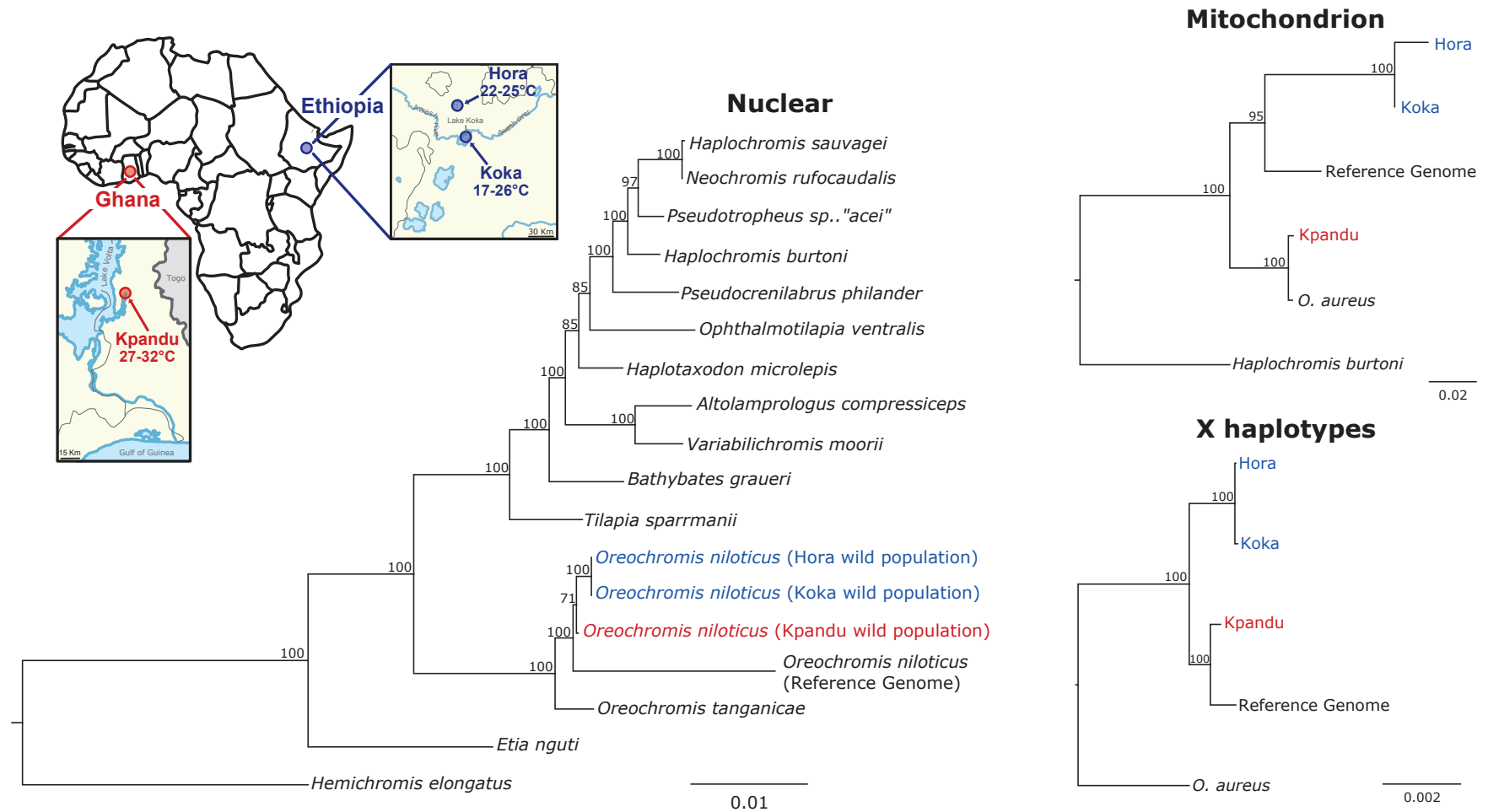


Figure 9. Maximum likelihood phylogenies of Nile tilapia populations and related cichlids. Node support correspond to bootstrap values. Koka and Hora (in blue) are Ethiopian populations of *O. niloticus cancellatus* and Kpandu (in red) is a Lake Volta population of *O. niloticus niloticus* originating from Ghana. **A)** Nuclear DNA phylogeny built using the 42 loci from Meyer et al., (2015). The *Oreochromis* species are forming a monophyletic clade in which the three wild populations are more closely related to each other than to the individual used from the reference genome. The tree is rooted with *Hemichromis elongatus*. The species from Meyer et al., 2015 are positioned identically to their phylogeny. These results are also showing that the *O. niloticus cancellatus* Koka and Hora populations are more closely related to each other than to the *O. niloticus niloticus* of Kpandu population. **B)** Mitochondrial DNA phylogeny built from 13 mitochondrial protein-coding genes and the two ribosomal subunits. The tree is rooted with *Haplochromis burtoni*. This phylogeny is confirming that Koka and Hora populations are more closely related to each other than to Kpandu population. *O. niloticus* samples from Kpandu population are forming a monophyletic clade with *O. aureus*. **C)** LG23-X haplotype phylogeny from LG23 *oaz1* to *dot11* genes. The tree is rooted with *O. aureus*. Koka and Hora *O. niloticus* populations are forming a monophyletic clade.

4. Discussion

An ancestral XX/XY system on LG23 with a tandem duplication of the *amh* gene

In the past, we had thought that the Ethiopian wild populations lacked this duplication because X and Y discriminating markers based on *amh* indels (Eshel et al., 2014; Li et al., 2015; Sissao et al., 2019) did not work in the Koka Ethiopian population (Triay et al., 2020). PCR markers have shown the presence of *amhY* and/or *amhΔY* on LG23 now in several laboratory and aquaculture strains of Nile tilapia (Eshel et al., 2014; Li et al., 2015; Wessels et al., 2017; Curzon et al., 2020; Taslima et al., 2020) as well as in the wild populations of Lake Kou, Burkina Faso (Sissao et al., 2019). Subsequently, we found that two wild populations of Nile tilapia *O. niloticus*, including an Ethiopian population (Lake Koka), shared the same sex chromosome (Linkage group 23) as several domesticated strains (Triay et al., 2020). These previous results used short Illumina reads obtained from genome-wide sequencing, and suggested that males had a Y specific duplication of the *amh* gene.

Here, we have conducted a study using Nanopore long reads to build Y haplotypes in wild populations. Because the minION reads are long enough to cross complex regions of the duplication encompassing the *amh* genes on LG23, we were able to confirm the presence of this Y-specific duplication and its structure, in wild populations from Lake Koka in Ethiopia (East Africa) and Lake Volta in Ghana (West Africa). In both populations, the duplication on the Y haplotype contains two *amh* genes and the boundaries comprise the *oaz1* and *dot11* genes, with the duplication harbouring truncated copies of *dot11* and *oaz1* found upstream (in the - strand) of the *amhY* promoter. Kpandu males from Lake Volta have a similar structure and the same *indels* to the males from the Japanese strain analysed by Li et al., (2015). One of the *amh* genes in Kpandu males resembles the truncated *amhΔY* copy found in the males from the Japanese strain (Li et al., 2015) since it also has the characteristic indels of 5 bp insertion that causes a premature stop codon and the deletion of about 233bp in the last exon. Therefore, the other Kpandu *amh* gene is presumably the SD *amhY* gene but it does not have the same missense SNP that may cause maleness (Li et al., 2015). In contrast, the two *amh* genes in Koka males do not have neither the missense SNP of *amhY*, nor the indels within or surrounding *amhΔY*, particularly the 5 bp insertion (Triay et al., 2020). Consequently, we cannot say which *amh* gene might be the master SD *amhY* gene because both might be functional.

Amh is a gene of the sex-differentiating network of vertebrates, and despite teleosts lacking Müllerian ducts, it plays a role in testis development, regulating germ cell proliferation (Poonlaphdecha et al., 2011; Pfennig et al., 2015; Adolfi et al., 2019). Duplicated *amh* genes have emerged repeatedly from autosomes in several fish clades, taking on the role of master male-determinant. In silversides, which like Nile tilapia, have the coexistence of genetics and temperature effects on sex determination, Y chromosome *amhy* genes have appeared in 2 species studied (Hattori et al., 2012; Yamamoto et al., 2014) with a third, surprisingly truncated, *amhy* that evolved independently in another clade (Bej et al., 2017).

Northern Pike's *amhby* is also an *amh* duplication that diverged from the autosomal *amha* gene (Pan et al., 2019) in the common ancestor of the Esocidae family, but *amhby* and its surrounding Y region was subsequently lost in a Northern American Pike population (Pan et al., 2021). In two Esocidae species the predicted *amhby* protein is truncated either at the N-terminal or C-terminal regions, but their tight *amhby* sex linkage suggests that they remain functional genes that have altered protein sequences (Pan et al., 2021).

We were unable to estimate the divergence between the Koka and Kpandu Y haplotypes. Nevertheless, it is parsimonious to think that the Y-*amh* duplication appearance on LG23 is ancestral since it is found in Nile tilapia populations from two distant African hydrographic basins. Microsatellites' structuring showed that the Koka population which belongs to the Ethiopian Awash basin was genetically different at the macrogeographical level from the Kpandu population located in the Volta basin that comprises the large Sudano-Sahelian region (Bezault et al., 2011). The phylogeny we inferred is in accordance with this previous study: the two subspecies *O. n. cancellatus* found in Lake Koka and Lake Hora in the Ethiopian basin are sister populations, reflecting the genetic proximity seen at the microgeographic level (Bezault et al., 2011), while Kpandu is more divergent. It would be interesting to analyse Y haplotypes in more wild populations in order to time the evolution of the Y *amh* duplication and when LG23 became a sex chromosome.

LG23 has also been recruited as the pair of sex chromosome in the *Pundamilia* cichlids of Lake Victoria in which a QTL approach identified a sex-linked region of 1.9 Mb, encompassing the *amh* gene (Feulner et al., 2018). In addition, previous QTL presenting sex-associated markers on LG23 were found in a meiogynogenetic line of *O. aureus* and in the F2 offspring of interspecific cross between *O. aureus* and *O. mossambicus* (Cnaani et al., 2004). This suggest that LG23 of cichlids might have a high potential to become a sex chromosome pair.

We could not construct a Y haplotype using the long reads of the Hora male, suggesting this haplotype does not exist in the male used. This conclusion is reinforced by the fact that no sex-specific pattern could be identified on LG23 *amh* region in the Hora genome short-reads analysis. Even if we suppose that the male used for the long read sequencing was a sex-reversed XX male, it is unlikely that all of the 10 males used for the WGS short reads, were XX sex-reversed males. This is surprising considering that Lake Hora is closely related to Lake Koka and belongs to the same subspecies. Therefore, it raises the question of whether *amh* is the sex-determining gene in the Hora population (discussed in more detail below).

Diversity of sex-determination system at the intraspecific level

Within Nile tilapia, some strains and stocks present their major sex-determining region not on LG23 but on LG1, found in some USA commercial stocks (Gammerdinger et al., 2014), in a Ghanaian family (Cnaani et al., 2008) and in the Manzala-Stirling strain, where LG23 Y is still segregating at low frequencies (Taslina et al., 2021). Surprisingly though, sex

segregates to LG23 in both the Manzala-Göttingen stock (Wessels et al., 2017) and the Manzala-Tihange stock (Sissao et al., 2019) which originated from the Manzala-Stirling stock. Likewise, the ARO-Israeli stock, originating from the Swansea strain, which itself originated from the Stirling stock, have the LG23 locus controlling sex (Eshel et al., 2014). This could suggest that LG1 with potential sex-loci is segregating in Lake Manzala (Egypt) and that LG1 could under some conditions, such as domestication inducing gene loss or fixation by drift, become the pair of sex chromosome. It would be intriguing to study SD in the wild population of Lake Manzala in Egypt but also in other subspecies, and populations belonging notably to the Nile basin to see if sex-patterns appear in LG1 and/or LG23. Interestingly, the reference genome used in our study is an individual originating from the Manzala-Stirling lab stock. The fact that this sample presents such a long branch in our phylogenies reinforce the idea of a strong molecular divergence of this stock, probably after its domestication. The domestication could be the reason for these strong changes as effects of domestications on stocks have already been associated to strong differences between lab stocks and natural population genetics (including sex-determination systems) (Wilson et al., 2014). Alternatively, this molecular divergence could result from a hybridization in this stock that might have happened during the process of domestication, which could explain the large divergence observed with other *O. niloticus* fish. This could even be the cause of the emergence of a new sex-determining system on LG1. For instance, the Thai-Chiltralada hybrid, a strain originating from crosses of *O. niloticus*, *O. mossambicus* and *O. aureus*, presents sex-associated markers on LG1, LG3 and LG23 (Curzon et al., 2019). An introgression of closely related species of *O. niloticus* could thus favor the turnover of sex-chromosomes so that LG1 became fixed as a new sex chromosome.

Loss of the Y *amh* duplication in the Hora Lake population reflects a rapid change in SD within Nile tilapia populations (discussed below). Indeed, more than 20 different major sex-determining genes, on more than 17 LGs, have now been described in cichlids, highlighting the fact that in this family there are a great diversity of heterogamety systems but also that many sex chromosomes have emerged between closely related species (Gammerdinger and Kocher, 2018; El Taher et al., 2020). To date, cichlids from Lake Tanganyika, appear to have the highest rates of sex chromosome turnover as well as heterogamety transitions (El Taher et al., 2020).

Studies to look at sex-determination systems in many species, typically only examine a few individuals of each species. Furman et al., (2020) suggested the intraspecific diversity of sex chromosomes might be as high as what is observed at the interspecific level. Variance in sexual conflict and recombination patterns, two important factors possibly for the evolution of sex chromosomes, could induce divergence between sex chromosomes of different populations.

Our study again emphasizes the need to analyse the evolution of sex chromosomes at the population level. Studies in Nile tilapia have also demonstrated that in domesticated strains (Wessels et al., 2014; Taslima et al., 2021) and wild populations (Bezault et al., 2007; Sissao et al., 2019) the sex-determination system can be more complex than just the *amh* gene on

LG23. Sex in the Nile tilapia is genetically inherited through the sex chromosomes but there is also a genetic inheritance of autosomal factors and of temperature sensitivity (Baroiller and D'Cotta, 2001; Bezault et al., 2007; Baroiller et al., 2009; Wessels et al., 2017). Temperature could be the trigger causing the turnover, towards a new sex chromosome system (Grossen et al., 2011; Baroiller and D'Cotta, 2016). This, along with the results of the Hora population lacking the LG23 Y sex-determinant suggest that the turnover in sex chromosomes and sex-determination systems might not be rare events at the intraspecific level.

Ever-young sex-chromosomes ?

The X and Y chromosomes of wild populations of *O. niloticus* studied here are visually homomorphic, like all cichlids sex chromosomes karyotyped to date (reviewed in Gammerdinger and Kocher, (2018)). Reduced sex chromosome differentiation is evidenced by the small size of the sex-specific region (<0.5Mb) of LG23 suggesting that recombination still occurs around the sex-region. The study of genotype-phenotype associations using ddRAD methods on families of Kpandu and Koka populations already suggested that recombination could occur close to the male-specific region (Triay et al., 2020). This reduced recombination further highlights the fact that LG23 has remained a « young » pair of sex chromosomes. The shared structure between males of the two wild populations suggests this XX/XY system predates the divergence of populations of *O. niloticus*. Thus, even if chronologically old, Nile Tilapia sex chromosomes do not seem to have evolved to a non-recombinant Y that starts to differentiate, but is maintained at a state of low recombination for long evolutionary times.

This finding further reinforces the idea that sex chromosomes do not always cease recombination and begin degenerating after acquiring a major sex-determinant. There are more and more examples of pairs of sex chromosomes that are homomorphic, and sometimes even so similar at the molecular level that they differ only by a SNP, as in *Takifugu* (Kamiya et al., 2012). Two alternative models (which might not be exclusive) were proposed to occur in cold-blooded vertebrates with homomorphic sex chromosomes. The “high-turnover hypothesis” suggests that the master-sex-determining genes switch frequently, and the regions not recombining around the new SD gene do not have time to decay (Volf et al., 2007). In cichlids this model has indeed been evidenced (Gammerdinger and Kocher, 2018; El Taher et al., 2020). The “fountain-of-youth model” on the other hand proposes that the sex chromosomes remain stable over long periods with some recombination occurring in XY females (Perrin, 2009). European tree frogs and ratites also have very old sex chromosomes at the species level with no reduction of recombination outside the sex region (Stöck et al., 2011). The authors suggested that the presence of supposedly young sex chromosomes in these frogs that contain nevertheless old SD genes, occurred due to occasional recombination between the X and Y chromosomes that took place in males or in sex-reversed XY females (Stöck et al., 2011).

We do not know to what extent sex-reversed XY females may exist in wild populations of Nile tilapia, although preliminary studies suggested they existed (Baroiller and D'Cotta, 2016). We still do not know what mechanisms exist in the Nile tilapia for the prevalence of LG23 as a sex chromosome. Turnover of sex chromosomes (appearance of LG1?) or in the SD system exemplified here with the Hora population, might as evoked previously, be more frequent than what is apparent. We need to date the divergence between Koka and Kpandu populations compared to the divergence of their Y haplotypes in order to understand if this « young » pair of sex chromosome is in an evolutionary pathway to stay « ever-young ».

Introgression of wild populations?

We noted an incongruence for Kpandu position between the nuclear and mitochondrial phylogenies. Our mitochondrial DNA analysis demonstrates a clustering of Kpandu *O. niloticus* population with the sample of *O. aureus*, also a Nilo-Sudanic species. The formation of a monophyletic group is consistent with the study of Rognon and Guyomard, (2003) in which they hypothesize a complete introgression of the mitochondrial DNA from *O. aureus* to *O. niloticus* in Western African populations, but also in Egyptian lakes. Clustering of Kpandu Nile tilapia and *O. aureus* was also observed when using mtDNA but they were differentiated in phylogenetic trees derived from nuclear DNA markers (SNPs) (Syaifudin et al., 2019). Discordance between mtDNA and nuclear DNA has been shown in several studies within the *Oreochromine* cichlids (Ford et al., 2019). Mitochondria DNA reflects only maternal inheritance, whereas species topology is probably more reliable when using multi-nuclear markers (Meyer et al., 2015; Ford et al., 2019). Ancient hybridization was suggested as being the most likely explanation for the discordance between mtDNA and nuDNA (Dunz and Schlieuwen, 2013).

Our results suggest that at some point in the evolutionary history of *O. niloticus* introgression with other closely related cichlids species might have occurred. Ancient hybridization of *O. niloticus* with other closely related species would not be surprising since several species coexist in many African fresh water basins. *O. aureus* and *O. niloticus* divergence was estimated to be 2.93 MYA (Xiao et al., 2015). They cohabitate in several river basins, both being Nilo-Sudano species (Trewavas, 1983), but they are not known to hybridize naturally (Payne and Collinson, 1983). In Lake Manzala there is a question on whether hybridization might have occurred between Nile tilapia and *S. galilaeus* confused perhaps as *O. aureus* (Ford et al., 2019). This might be the explanation of a SD on LG1 appearing in several *O. niloticus* strains (Cnaani et al., 2008). Thus, particularly when studying sex chromosomes, it might be essential to consider the other sex-determining loci segregating in the sympatric species of studied populations.

Loss of the Y-specific region on LG23 in the Ethiopian population of Lake Hora

We did not find a Y duplication of the *amh* sex region on the LG23 in the Hora males nor were we able to detect any SD system in this population. The phylogenetic studies

confirmed that the Ethiopian Koka and Hora populations, both from the subspecies *Oreochromis niloticus cancellatus*, belonged to a monophyletic clade with the Kpandu population *O. niloticus niloticus* branching as a sister group. These results appear contradictory with a shared ancestral Y chromosome between Koka and Kpandu males that is not shared between Koka and Hora. A better estimation of the divergence between Koka and Kpandu would have been the use of the Y haplotypes for phylogeny construction, helped by additional populations, to understand the evolutionary history of the LG23 Y chromosome. This would also enable us to date the event that led to the loss of LG23Y in the Hora population. We constructed the phylogeny using multinuclear markers analysed with a concatenation method, which generates artefacts since we assume the same evolution history for all loci. A better estimation of the evolution of LG23 would be obtained using the Y haplotypes, which we were unable to do. It would be necessary to first increase the long-reads sequencing to improve the coverage of the Y haplotypes of Kpandu and Koka males, correct for sequencing errors to avoid ambiguities and uncertainties in final sequences, and then rerun the phylogenies using coalescent methods. This would improve the phylogenies and would presumably allow us to know if this pair of sex chromosomes is chronologically young or if it is in a stable state and will possibly remain forever young.

Hybridization with other species would certainly give different results in the phylogeny and potentially a different position of Hora samples in the trees or at least a long branch for this population. From the numerous studies where LG23 Y is found as the major sex-determinant in *O. niloticus*, we can assume that there has been a loss of the ancestral Y haplotype in the Hora population. This loss raises two questions: Why has the Hora population experienced a turnover in the system of sex determination? What is the current system of sex determination in this population?

Possible causes of the loss of LG23-Y

Causes for sex chromosome turnovers can be numerous, and among all, a loss of the sex-determinant by drift, sex ratio bias (e.g. environmental effects inducing sex-reversal), or pleiotropic effects of a sexually antagonistic mutation can facilitate the fixation of a new sex-determinant (reviewed in Tao et al., (2021)). Demographic events might have caused changes in Lake Hora SD. Lake Hora is thought to have been restocked at the beginning of the twentieth century with the endemic subspecies *O. n. cancellatus* (Bezault, 2005). Lake Hora is, nevertheless, a small lake (~1,15 Km²) and the effective population size was supposedly small. The Hora population might have experienced a genetic bottleneck that led to the loss of the LG23-Y by drift. Intraspecific differences in SD was also observed for the haplochromine cichlid *Pseudocrenilabrus philander* for which WGS detected a strong XY locus in LG7 in a lake population whereas no GSD was found in the outflow creek population which experienced genetic bottleneck (Böhne et al., 2019).

Sex ratio bias can also occur in *O. niloticus* species because of the environmental effect of temperature. This species can undergo a sex reversal over the threshold temperature of 32°C

causing XX individuals to develop as viable and fertile males, and this temperature sensitivity is a variable and heritable trait (Baroiller and D'Cotta, 2001; Bezault et al., 2007; Wessels and Hörstgen-Schwark, 2007; Baroiller et al., 2009; Wessels and Hörstgen-Schwark, 2011). In the worst case scenario, the increasing number of XX individuals (males and females) in a population could progressively lead to the loss of the Y chromosome and permit a new sex-determinant master gene to ensure the sex-determination of the population. However, Lake Hora (like Lake Koka) is referred to as a « Cold » Lake because its temperature varies between 22 to 25°C during the year. This hypothesis does not seem relevant to explain the loss of the Y chromosome according to the current environment of the Hora population. Still, the history of past temperatures in the lake have not been investigated. Hence, the loss of an ancestral LG23 Y system in the Hora fish could have been facilitated by demographic events of this small population, with colonisation, and perhaps further influenced if part of the founders were XX neo-males.

Finally, we cannot exclude the possibility that there has been a transition in the Hora population from a genetic sex determination system to an environmental sex determination system.

What could be the current sex determining system(s) of the Hora population?

When a major sex-determinant is lost, it can be replaced by a new « master » gene, possibly already implied in the sex-determination and differentiation pathway. In the case of the Hora population, we did not identify any obvious sex-determinant loci between the pools of wild males and females. This could be due to the new major sex-determinant region being too recent to have accumulated enough SNPs to be detected through windowed Genome wide *Fst* analysis. It is also possible that after the loss of the major determinant LG23 Y, other loci involved in sex determination and differentiation pathway acquire a more important role and coexist during a transition period until the fixation and potential appearance of a new major determinant (Furman et al., 2020; Tao et al., 2020). In this case, more than one loci is segregating on several chromosomes, appearing as a polygenic sex determining system (Mank and Avise, 2009; Roberts et al., 2016). Although WGS of pooled sexes is a very efficient method to quickly detect sex determination locus and the linkage group acting as a sex chromosome when the system is monogenic, it is however not as efficient to detect polygenic systems (Gammerdinger and Kocher, 2018). We can not exclude the scenario of two sex-determining loci (or more?) segregating in the Hora population as recently suggested by (Taslina et al., 2021) for the Manzala-Stirling stock.

Ideally when sex chromosomes are strongly homomorphic, as it is expected in the Hora population, Palmer et al., (2019) suggest studying the segregation of SNPs in families between parents and progeny. Tools such as SEX-DETECTOR perform this kind of analysis, ideally on RNAseq data (Muyle et al., 2016). The downside of this method is that it is not optimized for strongly inbred populations and that it requires live fish to produce families data.

5. Conclusion

In conclusion, we have been able to reconstruct using long reads the Y haplotype corresponding to a duplicated Y-specific region on the LG23 sex chromosome in two wild populations of Nile tilapia. The Y haplotypes contain a duplication in tandem of two *amh* genes, a truncated chimera comprising the *oaz1* and *dotl1* genes, and the complete genes found in the breakpoints of the duplication. We evidenced polymorphisms in the Y haplotype between the two wild populations. The West African population of Kpandu, (Volta basin), has similar indels and sequences as previously described in domesticated strains of Nile tilapia, containing the master sex determinant *amhy* gene and a possibly truncated *amh* Δ Y gene. In sharp contrast, the East African Ethiopian population of Lake Koka does not possess these indels, and we found no distinctions between the duplicated *amh* genes (supposedly *amhy* and *amh* Δ Y) since both may be functional. We studied an additional Ethiopian population (Hora Lake), which has lost the Y *amh* duplication. Sex no longer segregates on LG23 and we were unable to find a GSD in this population. We suggest that the SD has changed to become either a polygenic SD or it has acquired an ESD system. Alternatively a new SD locus might be still too reduced to be identified. Further analysis will be required to understand the sex-determination system in Hora population and potentially find the major sex-determination gene. An analysis of the SNPs density between males and females at a genome wide scale could already bring some clues about key regions in linkage disequilibrium with sex.

References

- Abbott, J.K., Nordén, A.K., and Hansson, B. (2017). Sex chromosome evolution: historical insights and future perspectives. *Proceedings of the Royal Society B: Biological Sciences* 284, 20162806.
- Adolfi, M.C., Nakajima, R.T., Nóbrega, R.H., and Schartl, M. (2019). Intersex, Hermaphroditism, and Gonadal Plasticity in Vertebrates: Evolution of the Müllerian Duct and Amh/Amhr2 Signaling. *Annu. Rev. Anim. Biosci.* 7, 149–172.
- Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., and Delsuc, F. (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour* 20, 892–905.
- Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics* 14, 113–124.
- Bachtrog, D., Kirkpatrick, M., Mank, J.E., McDaniel, S.F., Pires, J.C., Rice, W., and Valenzuela, N. (2011). Are all sex chromosomes created equal? *Trends in Genetics* 27, 350–357.
- Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., et al. (2014). Sex determination: why so many ways of doing it? *PLOS Biology* 12, e1001899.
- Baroiller, J.F., and D’Cotta, H. (2001). Environment and sex determination in farmed fish. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 130, 399–409.
- Baroiller, J.-F., and D’Cotta, H. (2016). The Reversible Sex of Gonochoristic Fish: Insights and Consequences. *Sexual Development* 10, 242–266.
- Baroiller, J.F., D’Cotta, H., Bezault, E., Wessels, S., and Hoerstgen-Schwark, G. (2009). Tilapia sex determination: where temperature and genetics meet. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 153, 30–38.
- Bej, D.K., Miyoshi, K., Hattori, R.S., Strüssmann, C.A., and Yamamoto, Y. (2017). A Duplicated, Truncated *amh* Gene Is Involved in Male Sex Determination in an Old World Silverside. *G3: Genes, Genomes, Genetics* 7.
- Bezault, E. (2005). Etude Du Systeme De Determinisme Du Sexe au Sein De Populations Naturelles De Tilapia Du Nil, *Oreochromis niloticus* (linnaeus, 1758) : importance Des Composantes Genetiques Et Environnementales. UNIVERSITE DE PARIS-SUD.
- Bezault, E., Clota, F., Derivaz, M., Chevassus, B., and Baroiller, J.-F. (2007). Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture* 272, S3–S16.
- Bezault, E., Balaesque, P., Toguyeni, A., Fermon, Y., Araki, H., Baroiller, J.-F., and Rognon, X. (2011). Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genetics* 12, 102.
- Blaser, O., Neuenschwander, S., and Perrin, N. (2014). Sex-Chromosome Turnovers: The Hot-Potato Model. *The American Naturalist* 183, 140–146.
- Böhne, A., Wilson, C.A., Postlethwait, J.H., and Salzburger, W. (2016). Variations on a theme: Genomics of sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 17.
- Böhne, A., Weber, A.A.-T., Rajkov, J., Rechsteiner, M., Riss, A., Egger, B., and Salzburger, W. (2019). Repeated Evolution Versus Common Ancestry: Sex Chromosome Evolution in the Haplochromine Cichlid *Pseudocrenilabrus philander*. *Genome Biology and Evolution* 11, 439–458.

- Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Travisany, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., et al. (2019). Fine mapping using whole-genome sequencing confirms anti-müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3 (Bethesda)* 9, 3213–3223.
- Carneiro, M.O., Russ, C., Ross, M.G., Gabriel, S.B., Nusbaum, C., and DePristo, M.A. (2012). Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* 13, 375.
- Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95, 118–128.
- Cnaani, A., Zilberman, N., Tinman, S., Hulata, G., and Ron, M. (2004). Genome-scan analysis for quantitative trait loci in an F2 tilapia hybrid. *Mol Genet Genomics* 272, 162–172.
- Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., D'Hont, A., Baroiller, J.-F., D'Cotta, H., Penman, D.J., et al. (2008). Genetics of sex determination in tilapiine species. *Sexual Development* 2, 43–54.
- Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P.D., Grützner, F., and Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature* 508, 488–493.
- Curzon, A.Y., Seroussi, E., Benet-Perelberg, A., Zak, T., Dor, L., Shirak, A., and Ron, M. (2019). Hybrid Origin of the Thai-Chitralada Tilapia Strain Using DNA Barcoding and Microsatellite Analysis. *Israeli Journal of Aquaculture - Bamidgeh*.
- Curzon, A.Y., Shirak, A., Dor, L., Zak, T., Perelberg, A., Seroussi, E., and Ron, M. (2020). A duplication of the Anti-Müllerian hormone gene is associated with genetic sex determination of different *Oreochromis niloticus* strains. *Heredity* 125, 317–327.
- Devlin, R.H., and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208, 191–364.
- van Doorn, G.S., and Kirkpatrick, M. (2010). Transitions Between Male and Female Heterogamety Caused by Sex-Antagonistic Selection. *Genetics* 186, 629–645.
- Dunz, A.R., and Schlieven, U.K. (2013). Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as “Tilapia.” *Molecular Phylogenetics and Evolution* 68, 64–80.
- El Taher, A., Ronco, F., Matschiner, M., Salzburger, W., and Böhne, A. (2020). Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes (*Evolutionary Biology*).
- Eshel, O., Shirak, A., Dor, L., Band, M., Zak, T., Markovich-Gordon, M., Chalifa-Caspi, V., Feldmesser, E., Weller, J.I., Seroussi, E., et al. (2014). Identification of male-specific amh duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 15, 774.
- Ezaz, T., Sarre, S.D., O'Meally, D., Marshall Graves, J.A., and Georges, A. (2009). Sex Chromosome Evolution in Lizards: Independent Origins and Rapid Transitions. *Cytogenet Genome Res* 127, 249–260.
- Feulner, P.G.D., Schwarzer, J., Haesler, M.P., Meier, J.I., and Seehausen, O. (2018). A Dense Linkage Map of Lake Victoria Cichlids Improved the *Pundamilia* Genome Assembly and Revealed a Major QTL for Sex-Determination. *G3* 8, 2411–2420.
- Ford, A.G.P., Bullen, T.R., Pang, L., Genner, M.J., Bills, R., Flouri, T., Ngatunga, B.P., Rüber, L., Schlieven, U.K., Seehausen, O., et al. (2019). Molecular phylogeny of *Oreochromis*

- (Cichlidae: *Oreochromini*) reveals mito-nuclear discordance and multiple colonisation of adverse aquatic environments. *Molecular Phylogenetics and Evolution* 136, 215–226.
- Furman, B.L.S., Metzger, D.C.H., Darolti, I., Wright, A.E., Sandkam, B.A., Almeida, P., Shu, J.J., and Mank, J.E. (2020). Sex chromosome evolution: so many exceptions to the rules. *Genome Biol Evol* 12, 750–763.
- Gammerdinger, W.J., and Kocher, T.D. (2018). Unusual diversity of sex chromosomes in African cichlid fishes. *Genes* 9, 480.
- Gammerdinger, W.J., Conte, M.A., Acquah, E.A., Roberts, R.B., and Kocher, T.D. (2014). Structure and decay of a proto-Y region in *Tilapia*, *Oreochromis niloticus*. *BMC Genomics* 15, 975.
- Graves, J.A.M. (2014). Avian sex, sex chromosomes, and dosage compensation in the age of genomics. *Chromosome Res* 22, 45–57.
- Grossen, C., Neuenschwander, S., and Perrin, N. (2011). Temperature-dependent turnovers in sex-determination mechanisms: a quantitative model. *Evolution* 65, 64–78.
- Guiguen, Y., Bertho, S., Herpin, A., and Fostier, A. (2018). Sex Determination and Sex Control in *Salmonidae*. In *Sex Control in Aquaculture*, (Chichester, UK: John Wiley & Sons, Ltd), pp. 251–281.
- Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., and Strüssmann, C.A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *PNAS* 109, 2955–2959.
- Helfman, G.S., Colette, B.B., Facey, D.E., and Bowen, B.W. (2009). *The diversity of fishes: biology, evolution, and ecology* (Chichester, UK ; Hoboken, NJ: Blackwell).
- Jeffries, D.L., Lavanchy, G., Sermier, R., Sredl, M.J., Miura, I., Borzée, A., Barrow, L.N., Canestrelli, D., Crochet, P.-A., Dufresnes, C., et al. (2018). A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nature Communications* 9.
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., et al. (2012). A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genet* 8.
- Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30, 772–780.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649.
- Kirkpatrick, M. (2010). How and Why Chromosome Inversions Evolve. *PLoS Biol* 8, e1000501.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genetics* 11, e1005678.
- Mank, J.E., and Avise, J.C. (2009). Evolutionary Diversity and Turn-Over of Sex Determination in Teleost Fishes. *SXD* 3, 60–67.

- Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., et al. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417, 559–563.
- Menegon, M., Cantaloni, C., Rodriguez-Prieto, A., Centomo, C., Abdelfattah, A., Rossato, M., Bernardi, M., Xumerle, L., Loader, S., and Delledonne, M. (2017). On site DNA barcoding by nanopore sequencing. *PLoS ONE* 12, e0184741.
- Meyer, B.S., Matschiner, M., and Salzburger, W. (2015). A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. *Molecular Phylogenetics and Evolution* 83, 56–71.
- Muyle, A., Käfer, J., Zemp, N., Mousset, S., Picard, F., and Marais, G.A. (2016). SEX-DETECTOR: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biol Evol* 8, 2530–2543.
- Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., Naruse, K., Hamaguchi, S., and Sakaizumi, M. (2012). Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics* 191, 163–170.
- Myosho, T., Takehana, Y., Hamaguchi, S., and Sakaizumi, M. (2015). Turnover of Sex Chromosomes in Celebensis Group Medaka Fishes. *G3 Genes | Genomes | Genetics* 5, 2685–2691.
- Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A., et al. (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *PNAS* 99, 11778–11783.
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32, 268–274.
- Ortega-Recalde, O., Goikoetxea, A., Hore, T.A., Todd, E.V., and Gemmell, N.J. (2020). The Genetics and Epigenetics of Sex Change in Fish. *Annual Review of Animal Biosciences* 8, null.
- Palaiokostas, C., Bekaert, M., Khan, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2015). A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 16, 171.
- Palmer, D.H., Rogers, T.F., Dean, R., and Wright, A.E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology* 28, 4709–4724.
- Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.-M., Bobe, J., Concordet, J.-P., et al. (2019). Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genetics* 15, e1008013.
- Pan, Q., Feron, R., Jouanno, E., Darras, H., Herpin, A., Koop, B., Rondeau, E., Goetz, F.W., Larson, W.A., Bernatchez, L., et al. (2021). The rise and fall of the ancient northern pike master sex-determining gene. *ELife* 10, e62858.
- Payne, A.I., and Collinson, R.I. (1983). A comparison of the biological characteristics of *Sarotherodon niloticus* (L.) with those of *S. aureus* (Steindachner) and other tilapia of the delta and lower Nile. *Aquaculture* 30, 335–351.
- Perrin, N. (2009). Sex Reversal: A Fountain of Youth for Sex Chromosomes? *Evolution* 63, 3043–3049.
- Pfennig, F., Standke, A., and Gutzeit, H.O. (2015). The role of *Amh* signaling in teleost fish – Multiple functions not restricted to the gonads. *General and Comparative Endocrinology* 223, 87–107.

- Poonlaphdecha, S., Pepey, E., Huang, S.-H., Canonne, M., Soler, L., Mortaji, S., Morand, S., Pfennig, F., Mélard, C., Baroiller, J.F., et al. (2011). Elevated *amh* Gene Expression in the Brain of Male Tilapia (*Oreochromis niloticus*) during Testis Differentiation. *Sex Dev* 5, 33–47.
- Rice, W.R. (1987). The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. *Evolution* 41, 911–914.
- Roberts, N.B., Juntti, S.A., Coyle, K.P., Dumont, B.L., Stanley, M.K., Ryan, A.Q., Fernald, R.D., and Roberts, R.B. (2016). Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 17, 835.
- Roberts, R.B., Ser, J.R., and Kocher, T.D. (2009). Sexual Conflict Resolved by Invasion of a Novel Sex Determiner in Lake Malawi Cichlid Fishes. *Science* 326, 998–1001.
- Rognon, X., and Guyomard, R. (2003). Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. *Mol Ecol* 12, 435–445.
- Sandve, S.R., Rohlfs, R.V., and Hvidsten, T.R. (2018). Subfunctionalization versus neofunctionalization after whole-genome duplication. *Nat Genet* 50, 908–909.
- Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.-M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346, 240–244.
- Sissao, R., D’Cotta, H., Baroiller, J.-F., and Toguyeni, A. (2019). Mismatches between the genetic and phenotypic sex in the wild Kou population of Nile tilapia *Oreochromis niloticus*. *PeerJ* 7.
- Smith, C.A., Roeszler, K.N., Ohnesorg, T., Cummins, D.M., Farlie, P.G., Doran, T.J., and Sinclair, A.H. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature* 461, 267–271.
- Stöck, M., Horn, A., Grossen, C., Lindtke, D., Sermier, R., Betto-Colliard, C., Dufresnes, C., Bonjour, E., Dumas, Z., Luquet, E., et al. (2011). Ever-Young Sex Chromosomes in European Tree Frogs. *PLoS Biol* 9, e1001062.
- Syaifudin, M., Bekaert, M., Taggart, J.B., Bartie, K.L., Wehner, S., Palaikostas, C., Khan, M.G.Q., Selly, S.-L.C., Hulata, G., D’Cotta, H., et al. (2019). Species-Specific Marker Discovery in Tilapia. *Sci Rep* 9, 13001.
- Takehana, Y., Hamaguchi, S., and Sakaizumi, M. (2008). Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, *Oryzias javanicus* and *O. hubbsi*. *Chromosome Res* 16, 801–811.
- Takehana, Y., Matsuda, M., Myosho, T., Suster, M.L., Kawakami, K., Shin-I, T., Kohara, Y., Kuroki, Y., Toyoda, A., Fujiyama, A., et al. (2014). Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nat Commun* 5, 4157.
- Tao, W., Xu, L., Zhao, L., Zhu, Z., Wu, X., Min, Q., Wang, D., and Zhou, Q. (2020). High-quality chromosome-level genomes of two tilapia species reveal their evolution of repeat sequences and sex chromosomes (Genomics).
- Tao, W., Conte, M.A., Wang, D., and Kocher, T.D. (2021). Network architecture and sex chromosome turnovers: Do epistatic interactions shape patterns of sex chromosome replacement? *BioEssays* 43, 2000161.

- Taslina, K., Wehner, S., Taggart, J.B., de Verdal, H., Benzie, J.A.H., Bekaert, M., McAndrew, B.J., and Penman, D.J. (2020). Sex determination in the GIFT strain of tilapia is controlled by a locus in linkage group 23. *BMC Genet* 21, 49.
- Taslina, K., Khan, M.G.Q., McAndrew, B.J., and Penman, D.J. (2021). Evidence of two XX/XY sex-determining loci in the Stirling stock of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 532, 735995.
- Trewavas, E. (1983). Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*.
- Triay, C., Conte, M.A., Baroiller, J.-F., Bezault, E., Clark, F.E., Penman, D.J., Kocher, T.D., and D'Cotta, H. (2020). Structure and Sequence of the Sex Determining Locus in Two Wild Populations of Nile Tilapia. *Genes* 11, 1017.
- Volff, J.-N., Nanda, I., Schmid, M., and Scharf, M. (2007). Governing Sex Determination in Fish: Regulatory Putsches and Ephemeral Dictators. *Sex Dev* 1, 85–99.
- Wessels, S., and Hörstgen-Schwark, G. (2007). Selection experiments to increase the proportion of males in Nile tilapia (*Oreochromis niloticus*) by means of temperature treatment. *Aquaculture* 272, S80–S87.
- Wessels, S., and Hörstgen-Schwark, G. (2011). Temperature dependent sex ratios in selected lines and crosses with a YY-male in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 318, 79–84.
- Wessels, S., Sharifi, R.A., Luehmann, L.M., Rueangsri, S., Krause, I., Pach, S., Hörstgen-Schwark, G., and Knorr, C. (2014). Allelic Variant in the Anti-Müllerian Hormone Gene Leads to Autosomal and Temperature-Dependent Sex Reversal in a Selected Nile Tilapia Line. *PLoS ONE* 9, e104795.
- Wessels, S., Krause, I., Floren, C., Schütz, E., Beck, J., and Knorr, C. (2017). ddRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC Genomics* 18, 531.
- Wilson, C.A., High, S.K., McCluskey, B.M., Amores, A., Yan, Y., Titus, T.A., Anderson, J.L., Batzel, P., Carvan, M.J., Scharf, M., et al. (2014). Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. *Genetics* 198, 1291–1308.
- Xiao, J., Zhong, H., Liu, Z., Yu, F., Luo, Y., Gan, X., and Zhou, Y. (2015). Transcriptome analysis revealed positive selection of immune-related genes in tilapia. *Fish & Shellfish Immunology* 44, 60–65.
- Yamamoto, Y., Zhang, Y., Sarida, M., Hattori, R.S., and Strüssmann, C.A. (2014). Coexistence of Genotypic and Temperature-Dependent Sex Determination in Pejerrey *Odontesthes bonariensis*. *PLOS ONE* 9, e102574.
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., Cabau, C., Bouchez, O., Fostier, A., and Guiguen, Y. (2012). An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Current Biology* 22, 1423–1428.

Supplementary Table 1. Genes used to infer the nuclear phylogeny. Gene nomenclature, species and accession numbers are excerpted from Meyer et al. 2015.

Gene_ID	<i>Altolamprologus compressiceps</i>	<i>Bathybates graueri</i>	<i>Haplochromis burtoni</i> *	<i>Haplotaxodon microlepis</i>	<i>Neochromis rufocaudalis</i>	<i>Ophthalmotilapia ventralis</i>	<i>Oreochromis tanganicae</i>
rag1	KP131329	KP131317	KM263618	KP131326	KP131331	KM263621	KP131316
b2m	KP129986	KP129970	KP129955	KP129983	KP129989	KP129959	KP129969
gapdhs	KP130246	KP130234	KM263633	KP130243	KP130248	KM263636	KP130233
Ptchd4	KP131371	KP131354	KP131338	KP131368	KP131374	KP131342	KP131353
enc1	KP131215	KP131198	KP131182	KP131212	KP131218	KP131186	KP131197
phpt1	KP130420	KP130403	KP130388	KP130417	KP130422	KP130392	KP130402
rps7	KP129946	KP129934	KM263648	KP129943	KP129948	KM263651	KP129933
tbr1	KP130121	KP130104	KP130088	KP130118	KP130124	KP130092	KP130103
aqp1a.1	KP131036	KP131020	KP131004	KP131033	KP131039	KP131008	KP131019
hprt1	KP130992	KP130976	KP130960	KP130989	KP130995	KP130964	KP130975
anxa4	KP130031	KP130014	KP129998	KP130028	KP130034	KP130002	KP130013
pgk1	KP129879	KP129862	KP129846	KP129876	KP129882	KP129850	KP129861
bmp4	KP130912	KP130900	KM263663	KP130909	KP130914	KM263666	KP130899
bmp2	KP129756	KP129739	KP129723	KP129753	KP129759	KP129727	KP129738
TMO-4C4	KP130705	KP130688	KP130672	KP130702	KP130708	KP130676	KP130687
fgf6b	KP130076	KP130059	KP130043	KP130073	KP130079	KP130047	KP130058
runx2	KP131416	KP131399	KP131383	KP131413	KP131419	KP131387	KP131398
furina	KP130750	KP130733	KP130717	KP130747	KP130753	KP130721	KP130732
wnt7b	KP130166	KP130149	KP130133	KP130163	KP130169	KP130137	KP130148
pax9	KP130333	KP130316	KP130300	KP130330	KP130335	KP130304	KP130315
sox10b	KP129711	KP129695	KP129679	KP129708	KP129714	KP129683	KP129694
otx2	KP130464	KP130447	KP130431	KP130461	KP130467	KP130435	KP130446
otx1	KP131081	KP131064	KP131048	KP131078	KP131084	KP131052	KP131063
dlx2a	KP130551	KP130535	KP130519	KP130548	KP130554	KP130523	KP130534
dlx4b	KP131293	KP131276	KP131260	KP131290	KP131296	KP131264	KP131275
barx1	KP130628	KP130612	KP130596	KP130626	KP130631	KP130600	KP130611
ednrb1a	KP131251	KP131239	KM263678	KP131248	KP131253	KM263681	KP131238
mc1r	KP130288	KP130271	KP130255	KP130285	KP130291	KP130259	KP130270
skia	KP130376	KP130359	KP130343	KP130373	KP130379	KP130347	KP130358
kita	KP129801	KP129784	KP129768	KP129798	KP129804	KP129772	KP129783
mitfa	KP129837	KP129825	KM263693	KP129834	KP129839	KM263696	KP129824
tyr	KP129914	-	KM263708	KP129911	KP129916	KM263711	KP129902
hagoromo (fbxw4)	KP130587	KP130575	KM263723	KP130584	KP130589	KM263726	KP130574
slc45a2 (aim)	KP131126	KP131109	KP131093	KP131123	KP131129	KP131097	KP131108
rh1	KP130833	KP130816	KP130800	KP130830	KP130836	KP130804	KP130815
opn1mw (lws)	KP130509	KP130492	KP130476	KP130506	KP130511	KP130480	KP130491
opn1sw (sws)	KP130877	-	KP130845	KP130874	KP130880	KP130849	KP130860
ccng1	KP130950	KP130934	KP130920	KP130947	KP130953	KP130923	-
snx33	-	KP130776	KP130762	KP130788	KP130793	KP130766	KP130775
rpl13a	KP130210	KP130193	KP130178	KP130207	KP130213	KP130182	-
edar	KP131170	KP131154	KP131138	KP131167	KP131173	KP131142	KP131153
csf1ra	KP130663	KP130651	KM263738	KP130660	KP130665	KM263741	KP130650

Gene_ID	Pseudocrenilabrus_philander	Pseudotropheus_sp	Tilapia_sparrmanii	Tylochromis_polylepis	Variabilichromis_moorii
rag1	KM263622	KP131335	KP131337	KP131325	KP131314
b2m	KP129973	KP129994	KP129997	-	KP129968
gapdhs	KM263637	KP130252	KP130254	KP130242	KP130231
Ptchd4	KP131357	KP131379	KP131382	KP131366	KP131351
enc1	KP131201	KP131223	KP131226	KP131210	KP131195
phpt1	KP130406	KP130427	KP130430	KP130415	-
rps7	KM263652	KP129952	KP129954	KP129942	KP129931
tbr1	KP130107	KP130129	KP130132	KP130116	KP130101
aqp1a.1	KP131023	KP131044	KP131047	-	KP131017
hprt1	KP130979	KP131000	KP131003	-	KP130973
anxa4	KP130017	KP130039	KP130042	KP130026	KP130011
pgk1	KP129865	KP129887	KP129890	KP129874	KP129859
bmp4	KM263674	KP130918	-	KP130908	KP130897
bmp2	KP129742	KP129764	KP129767	KP129751	KP129736
TMO-4C4	KP130691	KP130713	KP130716	KP130700	KP130685
fgf6b	KP130062	KP130084	KP130087	KP130071	KP130056
runx2	KP131402	KP131424	KP131427	KP131411	KP131396
furina	KP130736	KP130758	KP130761	KP130745	KP130730
wnt7b	KP130152	KP130174	KP130177	KP130161	KP130146
pax9	KP130319	KP130340	KP130342	KP130328	KP130313
sox10b	KP129698	KP129719	KP129722	KP129706	KP129692
otx2	KP130450	KP130472	KP130475	KP130459	KP130444
otx1	KP131067	KP131089	KP131092	KP131076	KP131061
dlx2a	KP130538	KP130559	KP130562	-	KP130532
dlx4b	KP131279	KP131301	KP131304	KP131288	KP131273
barx1	KP130615	KP130636	KP130638	KP130624	KP130609
ednrb1a	KM263682	KP131257	KP131259	KP131247	KP131236
mc1r	KP130274	KP130296	KP130299	KP130283	KP130268
skia	KP130362	KP130384	KP130387	KP130371	KP130356
kita	KP129787	KP129809	KP129812	KP129796	KP129781
mitfa	KM263697	KP129843	KP129845	KP129833	KP129822
tyr	KM263712	KP129919	KP129921	KP129910	KP129900
hagoromo (fbxw4)	KM263727	KP130593	KP130595	KP130583	KP130572
slc45a2 (aim)	KP131112	KP131134	KP131137	KP131121	KP131106
rh1	KP130819	KP130841	KP130844	KP130828	KP130813
opn1mw (lws)	KP130495	KP130516	KP130518	KP130504	KP130489
opn1sw (sws)	KP130863	KP130885	KP130887	KP130872	KP130858
ccng1	KP130937	KP130958	-	-	KP130932
snx33	KP130779	KP130798	-	-	KP130774
rpl13a	KP130196	KP130218	KP130221	KP130205	KP130191
edar	KP131157	KP131178	KP131181	KP131165	KP131151
csf1ra	KM263742	KP130669	KP130671	KP130659	KP130648

Supplementary Table 2. Mitochondrial sequences used to infer the phylogeny. The 13 mitochondrial protein-coding genes and large and small ribosomal subunits (rrnL and rrnS respectively) alignments length are shown.

Gene	Alignment length (bp)
ATP6	184
ATP8	168
COX1	1596
COX2	699
COX3	783
CYTB	1140
ND1	975
ND2	1047
ND3	351
ND4	1386
ND4L	297
ND5	1959
ND6	642
rrnL	1821
rrnS	1005

Supplementary Table 3. Annotations of LG05 Significant Sites Highlighted in Figure 8

NCBI LOC ID	Name	Start bp	Stop bp	Length	Direction	Full Name
100707136		3468286	3765853	297568	forward	dedicator of cytokinesis protein 3
102081134		3697255	3703302	6048	reverse	putative selection and upkeep of intraepithelial T-cells protein 1 homolog
102081399		3770176	3781937	11762	reverse	GTPase IMAP family member 8
109200463		3832863	3835200	2338	forward	uncharacterized
109200464		3839140	3862031	22892	reverse	probable polypeptide N-acetylgalactosaminyltransferase 8
100700530		3883158	3903851	20694	reverse	probable polypeptide N-acetylgalactosaminyltransferase 8
109200465		3925744	3926951	1208	forward	uncharacterized
109194354		3927032	3929537	2506	forward	sentrin-specific protease 1
102081731		3941553	3965998	24446	reverse	butyrophilin subfamily 2 member A2
109202277		4044100	4046879	2780	reverse	uncharacterized
102081639		4116682	4127317	10636	reverse	uncharacterized
109202276		4183981	4187797	3817	reverse	uncharacterized
102078600		4191284	4197462	6179	reverse	butyrophilin-like protein 10
112846868		4202815	4207032	4218	reverse	uncharacterized
109202275		4215918	4235119	19202	reverse	selection and upkeep of intraepithelial T-cells protein 3-like
102078378		4242417	4262471	20055	reverse	uncharacterized
102078311		4300049	4304861	4813	reverse	V-set domain-containing T-cell activation inhibitor 1
106097578		4316770	4363784	47015	reverse	selection and upkeep of intraepithelial T-cells protein 1
109202273		4348528	4350516	1989	reverse	uncharacterized
112841760		4350829	4355367	4539	reverse	uncharacterized
100707106	rbpl	4386732	4426087	39356	reverse	recombination signal binding protein for immunoglobulin kappa J region-like
102077798		4392623	4395300	2678	forward	uncharacterized
100706844	matn4	4417058	4451696	34639	forward	matrilin 4
100706572		4460524	4461763	1240	forward	odorant receptor 131-2-like
112846848		4468404	4471950	3547	reverse	proline-rich protein 36-like
100706303		4494720	4497167	2448	forward	odorant receptor 131-2-like
102081278		4502781	4504678	1898	forward	odorant receptor 131-2-like
112847007		4504734	4505294	561	forward	uncharacterized
100705779		4511779	4513181	1403	forward	odorant receptor 131-2-like
100705248		4526167	4528340	2174	forward	odorant receptor 131-2-like
109202272		4535516	4538000	2485	forward	uncharacterized
112846983		4542115	4543448	1334	reverse	uncharacterized
109202134		4563274	4565691	2418	forward	odorant receptor 131-2-like
106098729		4574434	4575750	1317	forward	odorant receptor 131-2-like
100704178		4576361	4585370	9010	forward	odorant receptor 131-2-like
100703909		4586935	4589688	2754	forward	odorant receptor 131-2-like
100703638		4591947	4593762	1816	forward	odorant receptor 131-2-like
100703363		4596175	4613618	17444	forward	potassium voltage-gated channel subfamily S member 2
100695319		4619676	4657720	38045	reverse	serine/threonine-protein kinase 4
100695052		4658952	4663817	4866	reverse	cytochrome c oxidase subunit 6C-1
100703093	angpt4	4667803	4738828	71026	reverse	angiopoietin 4
109202154		4752852	4760352	7501	forward	uncharacterized
102080461		4764244	4769887	5644	forward	coxsackievirus and adenovirus receptor homolog
112846841		4775808	4777642	1835	forward	uncharacterized
102080365		4786930	4793877	6948	forward	polymeric immunoglobulin receptor-like
109202271		4796866	4810566	13701	forward	uncharacterized
102080253		4814583	4824219	9637	forward	uncharacterized
102080165		4832547	4844009	11463	forward	uncharacterized
100694792		4847268	4866754	19487	reverse	cullin-associated and neddylation-dissociated 2 (putative)
100702828	sec13	4868840	4878343	9504	forward	SEC13 homolog, nuclear pore and COPII coat complex component
112846869		4876720	4877263	544	reverse	uncharacterized
109194281		4879605	4894911	15307	forward	deoxyribonuclease-1
100694524		4898581	4914010	15430	forward	uncharacterized
100702560		4914008	4915485	1478	reverse	granzyme B
100702290		4931470	4945346	13877	forward	deoxyribonuclease-1
100702027		4965422	4967405	1984	reverse	granzyme B(G,H)
100701753		5002826	5061770	58945	reverse	mast cell protease 1A-like
109194506		5099114	5100702	1589	reverse	anionic trypsin-1-like
102076094		5114496	5119647	5152	forward	deoxyribonuclease-1-like
100700928		5121186	5129749	8564	forward	deoxyribonuclease-1
109194541		5131125	5166160	35036	forward	deoxyribonuclease-1
100701205		5142734	5144527	1794	reverse	granzyme B(G,H)
106098728		5166548	5168054	1507	reverse	granzyme B-like
100534529		5180256	5181654	1399	reverse	granzyme-1
100700119		5185253	5194810	9558	forward	keratin, type II cytoskeletal 8
112846870		5187238	5191437	4200	reverse	uncharacterized
102075824		5192671	5195932	3262	reverse	uncharacterized
100699851	rb1	5196203	5214460	18258	forward	retinoblastoma-like 1 (p107)
100699587	tmem74b	5217848	5228706	10859	forward	transmembrane protein 74B
100693714	psmf1	5242815	5255349	12535	reverse	proteasome inhibitor subunit 1
100693452	pigu	5256940	5267329	10390	forward	phosphatidylinositol glycan anchor biosynthesis, class U
100699316		5269816	5271363	1548	forward	tripartite motif-containing protein 35-like
100699045		5275154	5278246	3093	forward	zinc-binding protein A33-like
109202133		5314207	5318951	4745	forward	uncharacterized
109202270		5327569	5329694	2126	forward	uncharacterized
106098731		5333278	5337245	3968	forward	L-selectin
102082447		5353400	5364128	10729	forward	protein shisa-5
102082295		5369009	5381572	12564	forward	protein shisa-5
102082209		5383592	5387008	3417	forward	protein shisa-5-like
102082092		5389594	5392140	2547	forward	protein shisa-4-like
112846826		5394440	5410812	16373	reverse	cell wall adhesin EAP1-like
100693178	itga7	5419577	5457235	37659	forward	integrin, alpha 7
100692906	npff	5458991	5461382	2392	forward	neuropeptide FF-amide peptide precursor like
106098726	ppp1r1a	5466362	5502747	36386	forward	protein phosphatase 1 regulatory inhibitor subunit 1A
100692636	pde1b	5589607	5634041	44435	reverse	phosphodiesterase 1B

Le challenge de l'assemblage des chromosomes sexuels

De nouvelles méthodes de séquençage pour étudier les chromosomes sexuels

L'identification des chromosomes sexuels peut être un défi, surtout quand ceux-ci sont homomorphes car ils ne peuvent pas être différenciés par des analyses cytogénétiques. Les analyses moléculaires et génomiques offrent de bonnes opportunités pour mieux détecter les loci liés au sexe et identifier les chromosomes sexuels. L'amélioration des technologies de séquençage de nouvelle génération (NGS) facilitera particulièrement la reconstruction des chromosomes Y ou W grâce à l'utilisation des lectures courtes d'Illumina et plus récemment grâce au séquençage des lectures longues (Abbott et al., 2017). Au fil des années, différentes méthodes ont été optimisées pour étudier les chromosomes sexuels selon leur homomorphie, leur hétéromorphie ou le long d'un gradient d'intermédiaire (Palmer et al., 2019).

Chaque technologie de séquençage présente ses avantages et ses inconvénients. Le séquençage Illumina short reads, bien qu'ayant l'avantage de produire une très grande quantité de lectures avec peu d'erreurs, ne produit que de courtes séquences. Elles permettent d'identifier efficacement les groupes de liaison portant des marqueurs liés au sexe. Mais, en raison de leur longueur de 100-250pb en général, les short reads ne sont pas toujours assez longs pour distinguer les régions répétées. Ainsi, les régions dupliquées demeurent non assemblées ou les lectures des deux copies sont fusionnées comme un seul locus lorsque les paralogues ne sont pas suffisamment divergents.

Les longs reads peuvent cependant surmonter ce problème. La longueur des lectures permet de franchir non seulement les limites des architectures complexes mais aussi de chevaucher entièrement ces structures lorsqu'elles sont présentes dans le génome. Elles permettent de distinguer plus facilement les régions dupliquées, les régions répétées ou d'autres types de variants structuraux, comme le montrent diverses études (Audano et al., 2019 ; Wenger et al., 2019 ; Nattestad et al., 2018 ; Sedlazeck et al., 2018). Cependant, l'inconvénient du séquençage de longs reads est le taux d'erreur plus élevé dans les séquences. Les lectures longues peuvent être affectées par 5 à 15% d'erreurs de séquençage (Carneiro et al. 2012 ; Menegon et al. 2017). La seule façon de contrer cet inconvénient est d'augmenter la couverture, ce qui peut être coûteux en fonction de la taille du génome étudié

Certaines contraintes ont un impact sur les deux technologies. Lors de l'étude du sexe hétérogamétique, la première contrainte est la présence de deux haplotypes divergents dans le même génome. Par conséquent, la couverture de chaque haplotype (par exemple X et Y) sur la région spécifique au sexe est seulement de la moitié de celle des autosomes. Ainsi, l'effort de séquençage doit être doublé pour obtenir une couverture équivalente dans la région du sexe d'un individu hétérogamétique. Une conséquence directe est que lors du séquençage d'un individu pour produire le génome de référence d'une espèce, c'est généralement le sexe homogamétique qui est préférentiellement utilisé (Tomaszkiewicz et al., 2017).

Mapping sur un génome de référence

Lorsqu'un génome de référence est disponible, utiliser un mapping sur cette référence est généralement plus rapide que de générer un assemblage *de novo*. Cependant, sauf dans le cas d'organismes modèles, les génomes de référence sont généralement réalisés sur des échantillons homogamétiques car ils ont tendance à produire de meilleurs assemblages.

Ceci constitue une limite lors de l'étude des chromosomes sexuels. Si le locus déterminant le sexe est partagé entre les deux haplotypes (par exemple X et Y ou Z et W) et qu'ils cumulent des polymorphismes nucléotidiques liés au sexe, alors un simple mapping, accompagné d'une détection des SNPs permet de décrire correctement la région. Mais dans l'éventualité où les chromosomes sexuels ne sont pas totalement homologues, différents scénarios peuvent se produire et avoir un impact sur la reconstruction des chromosomes Y ou W.

i) S'il existe des loci spécifiques au X (ou Z) et absents du Y (ou W), lors du mapping d'un individu XY (ou ZW) sur un génome de référence homogamétique, aucune lecture provenant du Y (ou W) ne traversera cette région spécifique. Seule la moitié de la couverture attendue sera observée sur les régions X (ou Z) (Figure 3.2 A).

ii) Si une région présente dans le Y (ou le W) n'a pas de copie homologue sur le X (ou le Z), le mapping d'un individu du sexe hétérogamétique donnera lieu à un mapping satisfaisant dans l'ensemble. Cependant, les lectures chevauchant la région homologue et l'insertion ne seront pas mappées sur le génome de référence. Une baisse de couverture est donc attendue pour la région étroite correspondant aux limites de l'insertion sur le Y (ou le W) (Figure 3.2 B).

iii) S'il existe une duplication spécifique au chromosome Y (ou W) et que les deux copies sont relativement similaires à la seule région homologe sur le chromosome X (ou Z), la duplication des chromosomes Y ne sera pas entièrement assemblée. Les lectures appartenant aux deux copies du Y (ou W) vont mapper sur l'unique région homologe du X (ou Z), augmentant la couverture de cette région en comparaison des régions autosomales (Figure 3.2 C).

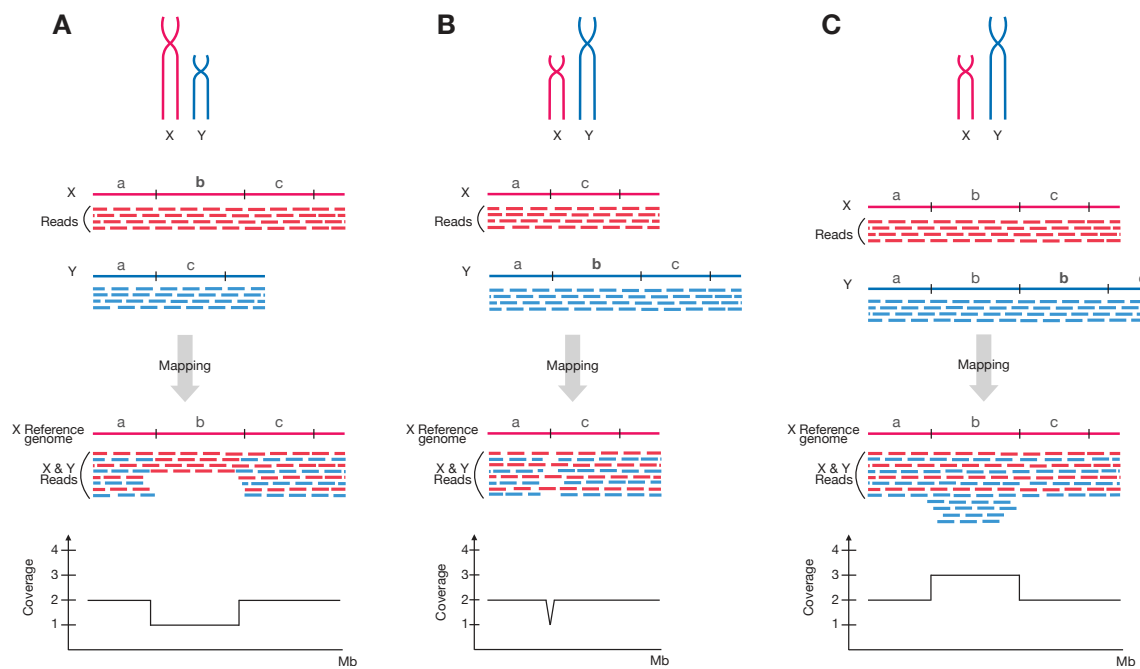


Figure 3.2. Problèmes rencontrés lors de la cartographie d'un individu hétérogamétique sur un génome de référence homogamétique.

L'impact attendu sur la couverture des lectures est présenté pour trois situations. Toutes les situations avec un système de déterminisme du sexe XY sont équivalentes pour un système ZW. **A)** Mapping attendu lorsque le chromosome Y a perdu une partie de son contenu en gènes (région b du X). Seules les lectures du chromosome X (en rouge) mappent sur le génome de référence, ce qui entraîne une baisse de la couverture sur cette région. **B)** mapping attendu lorsque le chromosome Y a acquis un locus absent de l'haplotype X (région b du Y). Les lectures spécifiques à la région b du Y ne mapperont pas sur le génome de référence. Seules les lectures du chromosome X (en rouge) mapperont sur les limites de l'insertion (région de transition a à c), ce qui entraîne une baisse locale de la couverture. **C)** mapping attendu lorsqu'une duplication s'est produite uniquement sur le chromosome Y (deux régions b sur le Y contre une région sur le X) et que les paralogues copiés ne sont pas fortement divergents avec la copie X. Les lectures appartenant aux deux copies b sur les chromosomes Y mapperont sur l'unique région b du génome de référence. Cette situation se traduira par une couverture accrue au niveau de la région correspondante.

Ces différents cas ne permettent pas de correctement assembler les deux haplotypes du sexe hétérogamétique, cependant, les variations de couverture résultant de ces situations

peuvent à leur tour être utilisées pour distinguer les loci liés au sexe de ceux des régions autosomales (Vicoso et al., 2013 ; Darolti et al., 2019).

En utilisant des lectures courtes paires, la détection de tels variants structuraux peut être améliorée. Certains logiciels, tels que DELLY, LUMPY ou GRIDSS (Rausch et al., 2012 ; Layer et al., 2014 ; Cameron et al., 2017) comparent la taille de l'insert entre les couples de lectures ainsi que leurs directions et la couverture par rapport aux attendus et peuvent ainsi déterminer le type et la longueur des variants structuraux le long du génome. Néanmoins, les structures complexes spécifiques aux seuls haplotypes Y ou W ne peuvent être décrites avec précision et, à titre d'exemple, il est difficile de déterminer à quelles copies appartiennent les SNP identifiés lorsque toutes les lectures mappent sur une unique copie dans le génome de référence.

Il existe également des méthodes pour améliorer la détection des variants structuraux dans les données de long reads (basées sur l'analyse de la couverture entre autres paramètres d'assemblage) telles que Sniffles, SMRT-SV (pour les données PacBio) ou NanoSV (pour les données Oxford Nanopore) (Huddleston et al., 2017 ; Cretu Stancu et al., 2017 ; Sedlazeck et al., 2018). La détection de variants structuraux à l'aide de longs reads permet une meilleure détection de ces régions complexes même si le taux d'erreur de séquençage est plus élevé que pour le séquençage de short reads. Pourtant, certaines limites apparaissent en présence de très grandes insertions ou inversions (par exemple, plus de 5kb). Cela est dû au fait que les lectures sont plus courtes que l'allèle de la variation. Pour une revue complète de la méthode de détection des variants structuraux avec des lectures courtes et longues suite à un mapping, voir Mahmoud et al, (2019).

Brève introduction à l'assemblage de graphes

L'utilisation de l'assemblage *de novo* pour reconstruire les haplotypes des deux sexes est une approche prometteuse lorsqu'on veut étudier des haplotypes avec des variants structuraux tels que des duplications en tandem qui sont spécifiquement liés à un des sexes.

La majorité des logiciels d'assemblage récents utilisent la méthode des graphes. En quelques mots, les graphes sont composés d'un ensemble de sommets reliés par des arêtes (qui peuvent avoir une direction) reflétant l'association entre les objets. Dans le cas de l'assemblage de long reads, le type de graphe utilisé est souvent un "graphe de De Bruijn". Ils sont construits à partir des lectures et des informations relatives à leur chevauchement. Pour

une revue détaillée de la théorie et de la pratique de l'assemblage de séquences génomiques et des graphes, voir Simpson & Pop (2015).

En présence d'un variant structurel, le graphe peut ne plus être linéaire et certaines arêtes peuvent être liées dans la même direction à de multiples arêtes. C'est le cas par exemple lorsqu'une répétition en tandem est présente. La formation non linéaire peut être appelée une "bulle". Le but de ces logiciels d'assemblage est de linéariser les graphes obtenus et ceci notamment en résolvant les bulles le long du graphe.

La qualité d'un assemblage se mesure principalement sur deux critères : la fidélité ("accuracy") et la contiguïté de l'assemblage. La contiguïté est accrue lorsque les graphes linéaires résultent en des chaînes les plus longues possibles avec le moins de ruptures possible. D'autre part, la précision est plus élevée lorsque l'hétérozygotie est faible et donc que la séquence consensus résultant du graphe contient le moins d'ambiguïtés possible. Pour produire les meilleurs assemblages possibles, des outils sont développés afin d'éliminer les variants structuraux, tels que les duplications, qui sont spécifiques à un haplotype dans un génome diploïde (Guan et al., 2020).

Le cas des chromosomes sexuels

Les chromosomes sexuels sont par définition des régions de forte hétérozygotie car ils contiennent des loci qui peuvent être spécifiques à un sexe et donc susceptibles de ne plus recombiner et d'accumuler des SNP spécifiques au sexe. De plus, ces régions sont sujettes à des variants structurels spécifiques des haplotypes. Chez les poissons, par exemple, plusieurs espèces présentent une duplication du gène déterminant le sexe sur un seul des deux chromosomes sexuels (par exemple, la duplication du gène de l'*amh* sur le Y chez le Pejerrey (*Odontesthes hatcheri*) et le Brochet (*Esox lucius*) (Hattori et al., 2012 ; Pan et al., 2019)). La présence de telles structures créerait des bulles dans le graphe (figure 3.3 A). Lors de la résolution de cette bulle, le compromis entre fidélité et contiguïté peut conduire à différents scénarios. Si la région spécifique au sexe est suffisamment divergente entre les deux haplotypes, alors l'assemblage pourrait aboutir à deux contigs, ce qui est le cas idéal lorsqu'on étudie les chromosomes sexuels (figure 3.3 B). Mais, si la région sexuelle est petite par rapport à la longueur totale du graphe du contig, alors un hybride des deux haplotypes sexuels pourrait être construit (Figure 3.3 C). Cela pourrait notamment être le cas pour une région sexuelle de petite taille portée par des chromosomes homomorphes. La fusion des

deux haplotypes pour de telles régions conduit à une reconstruction incorrecte des régions X et Y. Une bulle peut également être ignorée si le soutien en faveur de son existence est faible. Huddleston et al. (2017) ont démontré que la détection des variants structuraux est facilitée si le génome est haploïde. En revanche l'analyse d'un génome diploïde ne permet que d'identifier que 41% des variants structuraux identifier sur les génomes haploïdes (Huddleston et al. 2017).

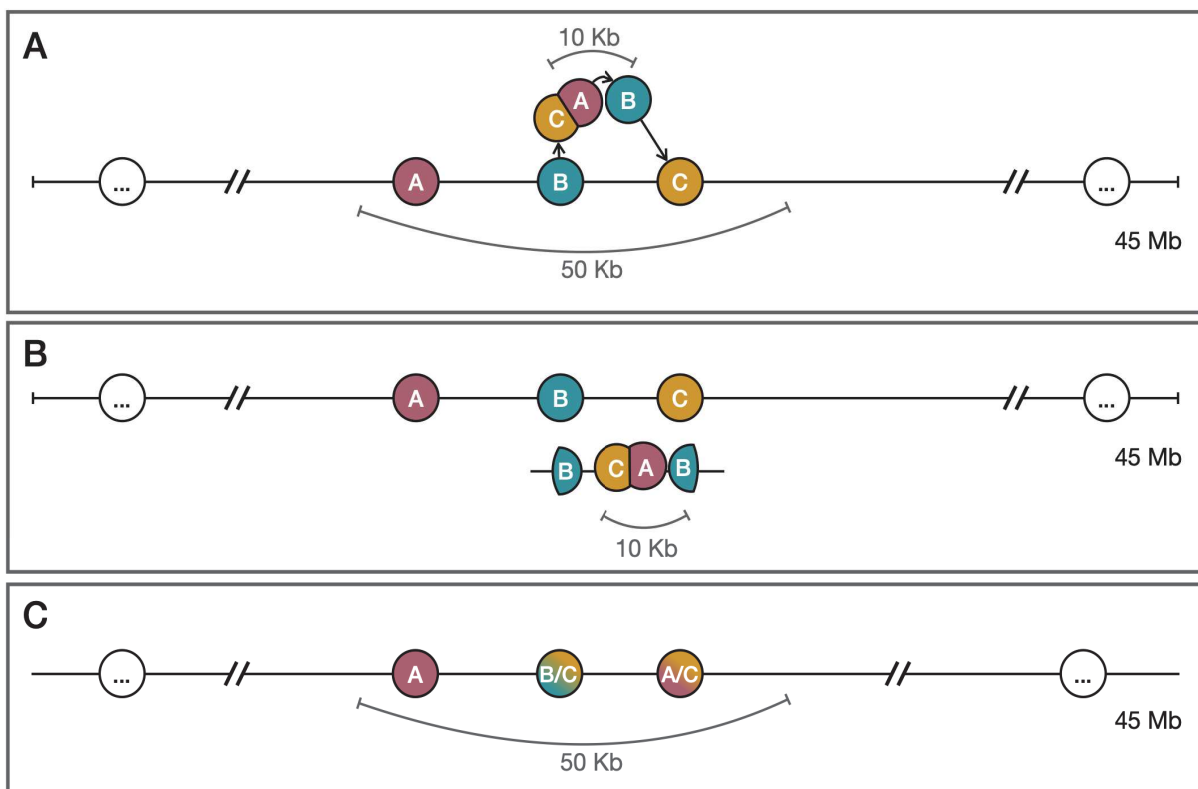


Figure 3.3. Représentation schématique de graphe et de « bulles » de long reads.

Le panneau A) est une illustration graphique d'une région dupliquée. Lors de la construction du graphe d'un groupe de liaison entier, les duplications sont interprétées comme des tracés alternatifs reliant les gènes/séquences. Cette situation est appelée une bulle. Afin de maximiser la contiguïté des assemblages, les assembleurs modernes essaient de résoudre ou de supprimer les bulles en fonction de leur taille. B) Deux contigs qui devraient être assemblés afin de reconstruire les haplotypes sexuels lorsque les chromosomes sexuels sont encore homologues à l'exception d'une duplication en tandem. Ce cas est peu probable car il nécessite de construire un nouveau contig pour la région dupliquée. C'est le contraire de ce que les assembleurs à graphes visent dans le cas d'un grand groupe de liaison. C) Résultats attendus mais erronés de la construction de contigs dans la situation du panneau A. Les petites bulles sont interprétées comme du polymorphisme et résultent donc en la construction de contigs consensus « hybrides ».

En conclusion ?

Idéalement, lors de l'assemblage d'une région spécifique au sexe, on s'attend à ce que deux haplotypes soient reconstruits. Mais il existe des cas, tel que la présence d'une région

du déterminisme du sexe réduite par rapport à l'ensemble de la taille du contig, dans lesquels il peut être difficile de reconstruire correctement les deux haplotypes. Une bonne option serait de séquencer un génome homozygote pour la région sexuelle, comme des individus XX et YY ou des individus ZZ et WW. Cependant, cela n'est pas toujours possible car les génotypes YY ou WW ne sont pas toujours viables ou facile à produire.

Pour compenser l'absence d'échantillons YY ou WW, l'étude du sexe hétérogamétique requiert des efforts plus importants. La profondeur de séquençage doit être doublée pour obtenir une couverture de la région du sexe aussi satisfaisante que celle des régions autosomiques. Les paramètres d'assemblage devront également faire l'objet d'une attention particulière pour calibrer les logiciels d'assemblage afin de produire deux contigs différents correspondant aux deux haplotypes sexuels en réalisant un compromis entre la fidélité et la contiguité dans ces régions en particulier.

Finalement, ces problèmes ne perdureront pas éternellement car les technologies de séquençage deviennent moins chères, plus efficaces et produisent des lectures plus longues et en plus grande quantité au fil du temps. Pour citer Phillippy (2017) qui passe en revue les nouvelles avancées dans l'assemblage de séquences : "dès qu'une technologie de séquençage pourra produire des lectures suffisamment bonnes, de sorte que toutes les régions du génome puissent être distinguées de manière unique, l'assemblage du génome deviendra trivial". Espérons qu'à ce stade, les séquences de chromosomes sexuels seront plus faciles à démasquer et amélioreront considérablement notre compréhension de l'évolution des chromosomes sexuels.

Références

- Abbott, J.K., Nordén, A.K., and Hansson, B. (2017). Sex chromosome evolution: historical insights and future perspectives. *Proceedings of the Royal Society B: Biological Sciences* 284, 20162806.
- Audano, P.A., Sulovari, A., Graves-Lindsay, T.A., Cantsilieris, S., Sorensen, M., Welch, A.E., Dougherty, M.L., Nelson, B.J., Shah, A., Dutcher, S.K., et al. (2019). Characterizing the Major Structural Variant Alleles of the Human Genome. *Cell* 176, 663-675.e19.
- Cameron, D.L., Schröder, J., Penington, J.S., Do, H., Molania, R., Dobrovic, A., Speed, T.P., and Papenfuss, A.T. (2017). GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res.* 27, 2050–2060.
- Carneiro, M.O., Russ, C., Ross, M.G., Gabriel, S.B., Nusbaum, C., and DePristo, M.A. (2012). Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* 13, 375.
- Cretu Stancu, M., van Roosmalen, M.J., Renkens, I., Nieboer, M.M., Middelkamp, S., de Ligt, J., Pregno, G., Giachino, D., Mandrile, G., Espejo Valle-Inclan, J., et al. (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun* 8, 1326.
- Darolti, I., Wright, A.E., Sandkam, B.A., Morris, J., Bloch, N.I., Farré, M., Fuller, R.C., Bourne, G.R., Larkin, D.M., Breden, F., et al. (2019). Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proc Natl Acad Sci USA* 116, 19031–19036.
- Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36, 2896–2898.
- Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., and Strüssmann, C.A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *PNAS* 109, 2955–2959.
- Huddleston, J., Chaisson, M.J.P., Steinberg, K.M., Warren, W., Hoekzema, K., Gordon, D., Graves-Lindsay, T.A., Munson, K.M., Kronenberg, Z.N., Vives, L., et al. (2017). Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.* 27, 677–685.
- Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* 15, R84.
- Mahmoud, M., Gobet, N., Cruz-Dávalos, D.I., Mounier, N., Dessimoz, C., and Sedlazeck, F.J. (2019). Structural variant calling: the long and the short of it. *Genome Biol* 20, 246.
- Menegon, M., Cantaloni, C., Rodriguez-Prieto, A., Centomo, C., Abdelfattah, A., Rossato, M., Bernardi, M., Xumerle, L., Loader, S., and Delledonne, M. (2017). On site DNA barcoding by nanopore sequencing. *PLoS ONE* 12, e0184741.
- Nattestad, M., Goodwin, S., Ng, K., Baslan, T., Sedlazeck, F.J., Rescheneder, P., Garvin, T., Fang, H., Gurtowski, J., Hutton, E., et al. (2018). Complex rearrangements and oncogene

- amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. *Genome Res.* 28, 1126–1135.
- Palmer, D.H., Rogers, T.F., Dean, R., and Wright, A.E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology* 28, 4709–4724.
- Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.-M., Bobe, J., Concordet, J.-P., et al. (2019). Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genetics* 15, e1008013.
- Phillippy, A.M. (2017). New advances in sequence assembly. *Genome Res.* 27, xi–xiii.
- Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339.
- Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 15, 461–468.
- Simpson, J.T., and Pop, M. (2015). *The Theory and Practice of Genome Sequence Assembly*. 22.
- Tomaszkiewicz, M., Medvedev, P., and Makova, K.D. (2017). Y and W Chromosome Assemblies: Approaches and Discoveries. *Trends in Genetics* 33, 266–282.
- Vicoso, B., Emerson, J.J., Zektser, Y., Mahajan, S., and Bachtrog, D. (2013). Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. *PLoS Biol* 11, e1001643.
- Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Functammasan, A., Kolesnikov, A., Olson, N.D., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 37, 1155–1162.

The challenge of sex chromosome assembly

New sequencing methods to study sex chromosomes

Identifying sex chromosomes can be challenging especially when they are homomorphic as they can not be differentiated by cytogenetic visualization. Molecular and genomic analyses offer great opportunities to better detect sex-linked loci and identify sex chromosomes. The improvement of new generation sequencing (NGS) technologies can facilitate particularly the reconstruction of Y or W chromosomes through the use of Illumina short reads, and more recently through long reads sequencing (Abbott et al., 2017). Over the years different methods have been optimised to study sex chromosomes according to their homomorphy, heteromorphy or along an intermediate gradient (Palmer et al., 2019).

Each sequencing technology presents its advantages and disadvantages. Illumina short reads sequencing, while having the advantage of producing a very large amount of reads with few errors, only produces small size sequences. They allow to efficiently identify linkage groups carrying markers linked to sex. But, because of their usual length of 100-250bp, short reads are not always long enough to distinguish between repeat regions. As such, duplicated regions are left unassembled or reads of both copies are merged as a single locus when the paralogs are not divergent enough.

Long reads, however, can overcome this issue. The length of the reads allows to cross not only the breakpoints but they can also overlap whole complex architectures when these occur in the genome. Consequently, they untangle duplicated regions, repeated regions or other types of structural variants more easily, as shown in various studies (Audano et al., 2019; Nattestad et al., 2018; Sedlazeck et al., 2018; Wenger et al., 2019). However, the downside of long reads sequencing is the higher error rate in the sequences. Long reads can be impacted with 5 to 15% sequencing errors (Carneiro et al., 2012; Menegon et al., 2017). The only way to counterbalance this downside is to increase the coverage, which can be expensive depending on the size of the genome studied.

Some constraints impact both technologies. When studying the heterogametic sex, the first concern is the presence of two divergent haplotypes in the same genome. As a result, the coverage of each haplotype (e.g. X and Y) over a sex-specific region is only half of that of the autosomes. Thus, the sequencing effort has to be doubled to obtain an equivalent coverage in

the sex-region of an heterogametic individual. This is why in the majority of cases the reference genome of a species is chosen preferentially from an individual corresponding to the homogametic sex (Tomaszkiewicz et al., 2017).

Mapping on a reference genome

When a reference genome is available, it is usually faster to map onto this reference than generating a *de novo* assembly. However, except in the case of model organisms, reference genomes, as previously mentioned, are usually conducted on homogametic individuals as they tend to produce better assemblies. This is indeed a limitation when studying sex chromosomes. If the sex determining locus is shared between the two haplotypes (*e.g.* X and Y or Z and W) and these haplotypes have accumulated sex-specific nucleotide polymorphisms, then a simple mapping, along with a SNPs detection will allow to correctly describe the region. But in the eventuality where the sex chromosomes are not fully homologous, different scenario can occur and impact the reconstruction of the Y or W chromosomes.

i) If there are loci that are specific to the X (or Z) and absent in the Y (or W), when mapping an XY (or ZW) individual over an homogametic based reference genome, there will be no reads from the Y (or W) that cross this specific region. Consequently, only half of the expected coverage will be observed over the X (or Z) regions (Figure 3.2 A).

ii) If there is a region present in the Y (or W) that has no homologous copy on the X (or Z), a mapping of an individual of the heterogametic sex will result in an overall good mapping. However, reads overlapping the homologous region and the insertion will not map to the reference genome. A drop of coverage is thus expected for the narrow region corresponding to the breakpoints of the insertion on the Y (or W) (Figure 3.2 B).

iii) If there is a duplication that is specific to the Y (or W) chromosome and both copies are relatively similar to the single homologous region on the X (or Z) chromosome, the Y chromosome duplication will not be fully assembled. The reads belonging to both copies present on the Y (or W) will map to the single copy of the X (or Z), which will increase the coverage of this region compared to autosomal regions (Figure 3.2 C).

These three different cases will not enable to properly assemble the two sex haplotypes of the heterogametic sex but, nevertheless, the coverage variations resulting from these

situations can in turn be helpful to distinguish sex-linked loci from autosomal regions (Vicoso et al., 2013; Darolti et al., 2019).

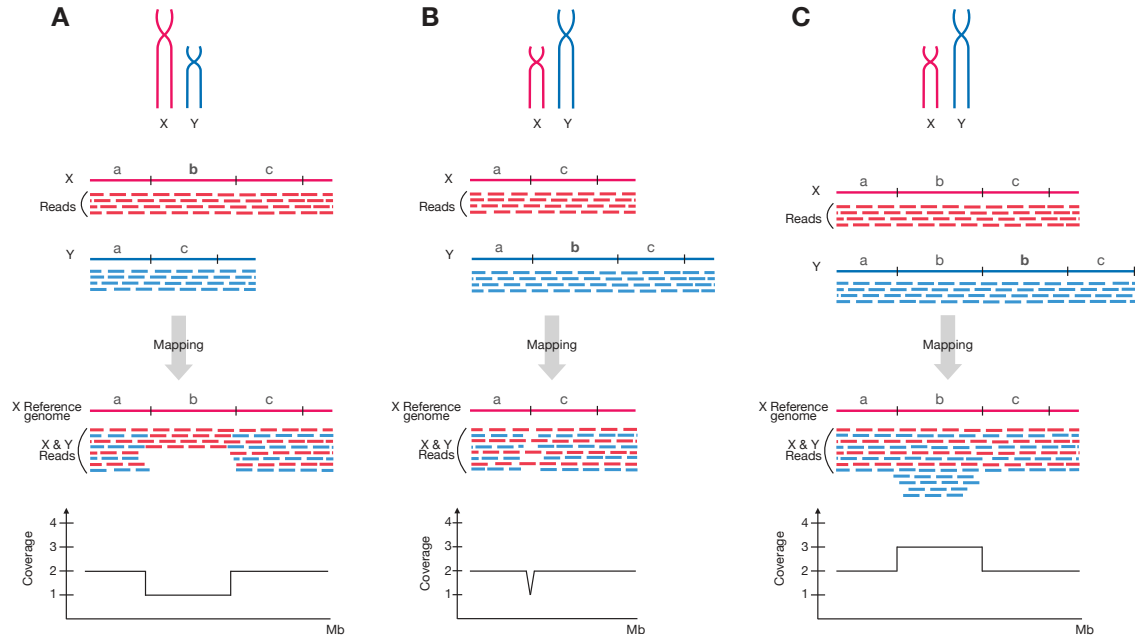


Figure 3.2. Problems encountered when mapping a heterogametic individual on a homogametic reference genome.

The expected impact on reads coverage is displayed for three situations. All situations with a XY sex system are equivalent for a ZW system. **A)** Expected mapping when the Y chromosome lost part of its genes content (region b of the X). Only reads from the X chromosome (in red) map on the reference genome resulting in a drop of coverage over this region. **B)** Expected mapping when the Y chromosome acquired a locus absent from the X haplotype (region b of the Y). Reads specific to the b region of the Y will not map on the reference genome. Only reads from the X chromosome (in red) will map on the breakpoints of the insertion (a to c transition region) resulting in a local drop of coverage. **C)** Expected mapping when a duplication occurred only on the Y chromosome (two region b on the Y vs one region on the X) and copied are paralogs are not strongly divergent with the X copy. Reads belonging to both b copies on the Y chromosomes will map on the single b region on the reference genome. This situation will result in an increased coverage over the corresponding region.

By using pair-end short reads, the detection of such structural variants can be improved. Some software, such as DELLY, LUMPY, GRIDSS (Rausch et al., 2012; Layer et al., 2014; Cameron et al., 2017), compare the insert size between mate pairs along with their direction and the coverage to what is expected and thus can determine the type and the length of structural variants along the genome. Nevertheless, complex structures that are specific to only Y or W haplotypes cannot be precisely described and as an illustration, it is hard to determine to which copies belong the SNPs identified when all reads map to a single copy on the reference genome.

Methods also exist to improve the detection of structural variants in long reads data (based on coverage analysis among other assembly parameters) such as Sniffles, SMRT-SV (for PacBio data) or NanoSV (for Oxford Nanopore data) (Cretu Stancu et al., 2017; Huddleston et al., 2017; Sedlazeck et al., 2018). The detection of structural variants using long reads results in better detection of these complexe regions even if the sequencing error rate is higher than for short reads sequencing. Still, some limits appear in the presence of very large insertions or inversions (eg. larger than 5kb). This is due to the reads being shorter than the allele of the variant. For a complete review of the structural variants calling with short and long reads following a mapping process, see Mahmoud et al., (2019).

Quick introduction to graph assembly

Using *de novo* assembly to reconstruct both sex haplotypes is thus a promising approach when one wants to study haplotypes with structural variants such as tandem duplications that are specific to one sex.

The majority of recent assembly software uses graph theory. Briefly, graphs are composed of a set of vertices connected by edges (that can have a direction) reflecting association between objects. In the case of long reads assembly, the type of graph used is often a “De Bruijn Graph”. They are built based on reads and their overlapping information. For a detailed review of theory and practice of genome sequence assembly and graph, see Simpson and Pop, (2015).

When a structural variant occurs, the graph might not be linear anymore and some edges might be linked in the same direction to several edges. It is the case for example when a tandem repeat is present. The non-linear formation can be called a “bubble”. The aim of these assembly software is to linearise the graphs obtained and this notably by resolving the bubbles along the graph.

The quality of the resulting assembly is measured mainly on two criteria: the accuracy and the contiguity of the assembly. The contiguity is increased when the linear graphs are resulting in the longest chains possible with the least breaks possible. The accuracy on the other hand is higher when the heterozygosity is low and thus the consensus sequence resulting from the graph contains the least ambiguities possible. To produce the best assemblies possible, tools are being developed in order to remove structural variants, such as duplications, that are specific to an haplotype in a diploid genome (Guan et al., 2020).

The case of sex chromosomes

Sex chromosomes are by definition regions of high heterozygosity because they contain loci that can be sex-specific and thus might not recombine anymore and accumulate sex-specific SNPs. In addition, these regions are prone to haplotype-specific structural variants. In fish for example, several species present a duplication of the sex-determinant gene in only one of the two sex chromosomes (*e.g. amh* duplication on the Y in the Patagonian pejerrey (*Odontesthes hatcheri*) and the Northern pike (*Esox lucius*) (Hattori et al., 2012; Pan et al., 2019)). The presence of such structures would create bubbles in a graph (Figure 3.3 A). When resolving this bubble, the compromise between accuracy and contiguity can lead to different scenarios. If the sex specific region is divergent enough between the two haplotypes, then the assembly might result in two contigs, which is the ideal case when studying sex chromosomes (Figure 3.3 B). But, if the sex region is small compared to the overall length of the contig graph, then a hybrid of the two sex haplotypes might be built (Figure 3.3 C). This might notably be the case for a short sex-specific region carried by homomorphic chromosomes. Merging the two haplotypes for such regions lead to false reconstruction of both X and Y regions. A bubble can also be ignored if the support for its existence is low. Huddleston et al., (2017) demonstrated that haploid genomes help to better detect structural variants. They performed that same analysis on a pseudo-diploid genome (merging two haploids) and detected 59% less structural variants than on the haploid dataset.

What are the options?

Ideally, when assembling a sex region, two haplotypes are expected to be reconstructed. In the case of a tandem duplication where the divergent region between the two chromosomes is small compared to the whole contig, it might be a challenge to correctly reconstruct the two sequences. A good option would be to sequence homozygous genomes for the sex region, such as XX and YY individuals or ZZ and WW individuals. However, this is not always possible as YY or WW genotypes are not always viable or easy to obtain.

To compensate for the absence of YY or WW samples, the study of heterogametic sex will require greater efforts. The sequencing depth has to be doubled to obtain a coverage over the sex-region that is as good as over autosomal regions. Assembly parameters will also require a close attention to calibrate assembly software in order to produce two different contigs

corresponding to sex-specific loci with a compromise between accuracy and contiguity in these regions particularly.

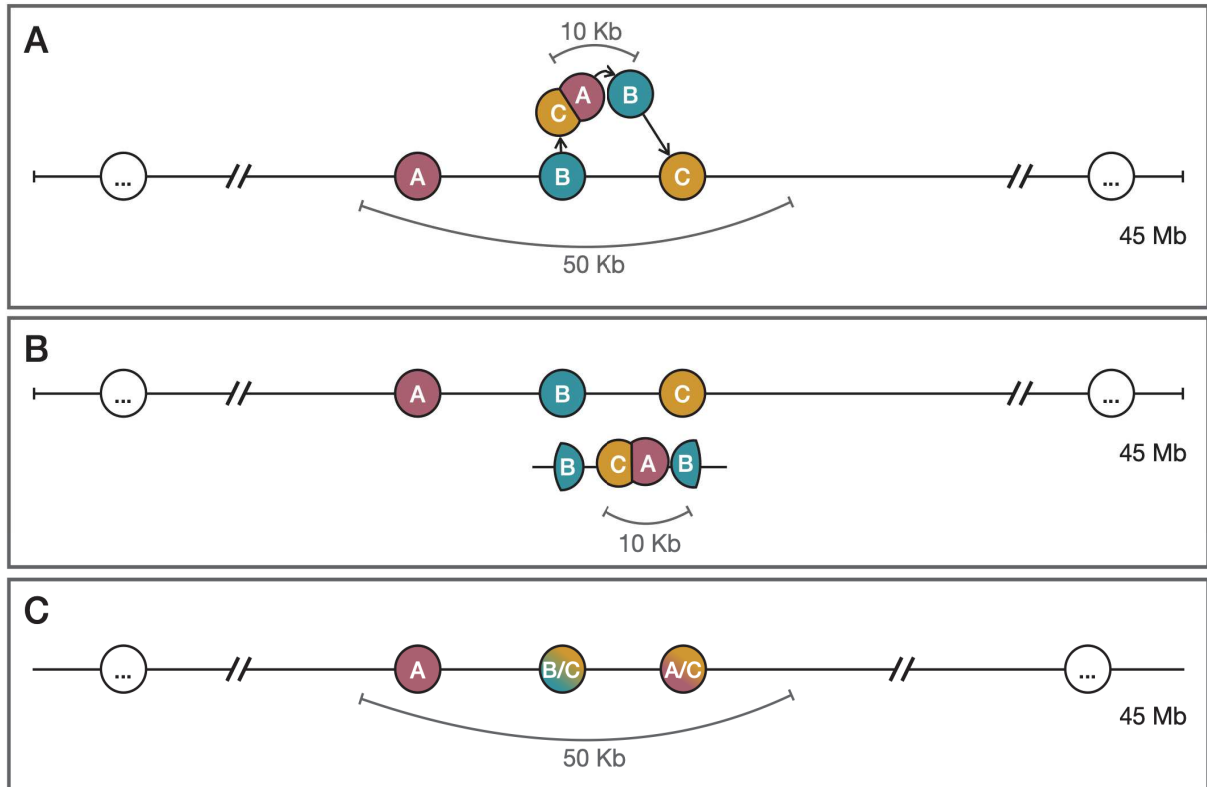


Figure 3.3. Schematic representation of graph and « bubbles » of long reads and hybrid assembly.

Panel A) is a graph illustration of a duplicated region. While constructing the graph of a whole linkage group, duplications are interpreted as alternative pathways linking genes/sequences. This situation is called a bubble. In order to maximise the contiguity of assemblies, modern graph assemblers are trying to resolve or suppress bubbles depending on their sizes. B) Two contigs that should be built in order to re-construct sex haplotypes when sex chromosomes are still homologous with exception of a tandem duplication. This case is unlikely to happen as it requires to build a new contig for the duplicated region. This is the opposite of what graphs assemblers are aiming for in the case of a large linkage group. C) Expected but erroneous results of contig construction with graph assembler in the panel A situation. Small bubbles are interpreted as polymorphism and thus result in the construction of « hybrid » consensus contig.

Eventually, these problems will not last forever as sequencing technologies are becoming cheaper, more efficient and are producing longer reads and in larger amounts through time. To quote Phillippy, (2017) reviewing the new advances in sequence assembly: “as soon as a sequencing technology can produce good enough reads, such that all regions of the genome can be uniquely distinguished, genome assembly will become trivial”. Hopefully, at this point in time, the sequences of sex chromosomes will be easier to uncover, greatly improving our comprehension of sex chromosome evolution.

References

- Abbott, J.K., Nordén, A.K., and Hansson, B. (2017). Sex chromosome evolution: historical insights and future perspectives. *Proceedings of the Royal Society B: Biological Sciences* 284, 20162806.
- Audano, P.A., Sulovari, A., Graves-Lindsay, T.A., Cantsilieris, S., Sorensen, M., Welch, A.E., Dougherty, M.L., Nelson, B.J., Shah, A., Dutcher, S.K., et al. (2019). Characterizing the Major Structural Variant Alleles of the Human Genome. *Cell* 176, 663-675.e19.
- Cameron, D.L., Schröder, J., Penington, J.S., Do, H., Molania, R., Dobrovic, A., Speed, T.P., and Papenfuss, A.T. (2017). GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res.* 27, 2050–2060.
- Carneiro, M.O., Russ, C., Ross, M.G., Gabriel, S.B., Nusbaum, C., and DePristo, M.A. (2012). Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* 13, 375.
- Cretu Stancu, M., van Roosmalen, M.J., Renkens, I., Nieboer, M.M., Middelkamp, S., de Lig, J., Pregno, G., Giachino, D., Mandrile, G., Espejo Valle-Inclan, J., et al. (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun* 8, 1326.
- Darolti, I., Wright, A.E., Sandkam, B.A., Morris, J., Bloch, N.I., Farré, M., Fuller, R.C., Bourne, G.R., Larkin, D.M., Breden, F., et al. (2019). Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proc Natl Acad Sci USA* 116, 19031–19036.
- Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36, 2896–2898.
- Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., and Strüssmann, C.A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *PNAS* 109, 2955–2959.
- Huddleston, J., Chaisson, M.J.P., Steinberg, K.M., Warren, W., Hoekzema, K., Gordon, D., Graves-Lindsay, T.A., Munson, K.M., Kronenberg, Z.N., Vives, L., et al. (2017). Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.* 27, 677–685.
- Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* 15, R84.
- Mahmoud, M., Gobet, N., Cruz-Dávalos, D.I., Mounier, N., Dessimoz, C., and Sedlazeck, F.J. (2019). Structural variant calling: the long and the short of it. *Genome Biol* 20, 246.
- Menegon, M., Cantaloni, C., Rodriguez-Prieto, A., Centomo, C., Abdelfattah, A., Rossato, M., Bernardi, M., Xumerle, L., Loader, S., and Delledonne, M. (2017). On site DNA barcoding by nanopore sequencing. *PLoS ONE* 12, e0184741.
- Nattestad, M., Goodwin, S., Ng, K., Baslan, T., Sedlazeck, F.J., Rescheneder, P., Garvin, T., Fang, H., Gurtowski, J., Hutton, E., et al. (2018). Complex rearrangements and oncogene

- amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. *Genome Res.* 28, 1126–1135.
- Palmer, D.H., Rogers, T.F., Dean, R., and Wright, A.E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology* 28, 4709–4724.
- Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.-M., Bobe, J., Concordet, J.-P., et al. (2019). Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genetics* 15, e1008013.
- Phillippy, A.M. (2017). New advances in sequence assembly. *Genome Res.* 27, xi–xiii.
- Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339.
- Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 15, 461–468.
- Simpson, J.T., and Pop, M. (2015). *The Theory and Practice of Genome Sequence Assembly*. 22.
- Tomaszkiewicz, M., Medvedev, P., and Makova, K.D. (2017). Y and W Chromosome Assemblies: Approaches and Discoveries. *Trends in Genetics* 33, 266–282.
- Vicoso, B., Emerson, J.J., Zektser, Y., Mahajan, S., and Bachtrog, D. (2013). Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. *PLoS Biol* 11, e1001643.
- Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Functammasan, A., Kolesnikov, A., Olson, N.D., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 37, 1155–1162

CHAPITRE 3 : DIFFÉRENCIATION DU SEXE CHEZ UNE POPULATION QUI NE PRÉSENTE PAS LE CHROMOSOME Y ANCESTRAL

Objectifs

Dans la population Ethiopienne du lac Hora, étudiée précédemment, aucune duplication de l'*amh* n'a pu être identifiée sur LG23-Y et aucun autre groupe de liaison n'a pour le moment été identifié comme la nouvelle paire de chromosomes sexuels. Dans ce chapitre, l'objectif était donc d'explorer le transcriptome de cette population Ethiopienne Hora pour le comparer à celui des souches domestiques dont le sexe est déterminé par LG23-Y. En effet, à un stade précoce de la différenciation, il peut déjà exister un dimorphisme dans l'expression des gènes clés de la cascade du sexe. Nous avons aussi analysé l'effet de la température sur le transcriptome de cette population.

Introduction et contexte

Le tilapia du Nil (*Oreochromis niloticus*) est une espèce de cichlidé Africain d'intérêt économique (FAO, 2019). Son aire de répartition est grande puisqu'on la retrouve du Sénégal à l'ouest de l'Afrique, jusqu'à l'est de l'Afrique dans le Rift Ethiopien, en passant par les bassins du Nil (Trewavas, 1983; Bezault et al., 2011). Cette espèce présente une grande capacité d'adaptation à différentes conditions écologiques et notamment thermiques avec des populations vivant dans des eaux « froides » (17 à 26°C), des eaux tempérées (28 à 34°C) mais aussi des sources chaudes (36 à 40°C) (Bezault et al., 2007; Ndiwa et al., 2014). Comme nous l'avons vu dans les chapitres précédents, le déterminisme du sexe chez cette espèce est complexe et reste mal compris. En effet, des signaux associés au sexe ont été identifiés sur les groupes de liaison (LG) 1 (Lee et al., 2003; Palaiokostas et al., 2013; Gammerdinger et al., 2014; Palaiokostas et al., 2015; Taslima et al., 2020) ou 23 (Eshel et al., 2012, 2014; Li et al., 2015; Wessels et al., 2017; Cáceres et al., 2019), ou même les deux en parallèle (Taslima et al., 2021), selon la provenance des souches domestiques étudiées. Dans les populations naturelles étudiées jusqu'alors, LG23 a aussi été identifié comme fortement associé au sexe dans les individus du lac Kou (Burkina-Faso), Kpandu du lac Volta (Ghana) et Koka du lac Koka (Ethiopie) (Sissao et al., 2019; Triay et al., 2020). La région du sexe identifiée sur LG23 correspond à une région dupliquée donnant lieu à deux copies du gène de l'*amh* dans l'haplotype Y (Li et al., 2015; Triay et al., 2020). Des analyses phylogénétiques sur les populations sauvages de Kpandu, Koka et Hora (Lac Hora, Ethiopie) ont permis de conclure que ce déterminisme est certainement ancestral à la divergence des populations de l'Est et de

l'Ouest de l'Afrique. Cependant, dans la population Hora aucune duplication de l'*amh* n'a pu être identifiée sur LG23-Y et aucun autre groupe de liaison n'a pour le moment été identifié comme la nouvelle paire de chromosomes sexuels.

Les cichlidés présentent tout particulièrement une grande diversité des déterminants du sexe avec près de 20 systèmes de déterminismes différents sur 17 LGs (Gammerdinger & Kocher, 2018; El Taher et al., 2020). Cependant, certains gènes candidats (« usual suspects ») reviennent régulièrement et « jouent aux chaises musicales » chez les cichlidés et plus largement chez les téléostéens (Böhne et al., 2016; Pan et al., 2016). Ainsi, même si le déterminant du sexe peut changer, les gènes impliqués en aval de la cascade de la différenciation du sexe sont généralement conservés chez les téléostéens (revues de Todd et al., 2016 et Ortega-Recalde et al., 2020) et pourraient jouer un rôle chez la population Hora (Figure 4.1).

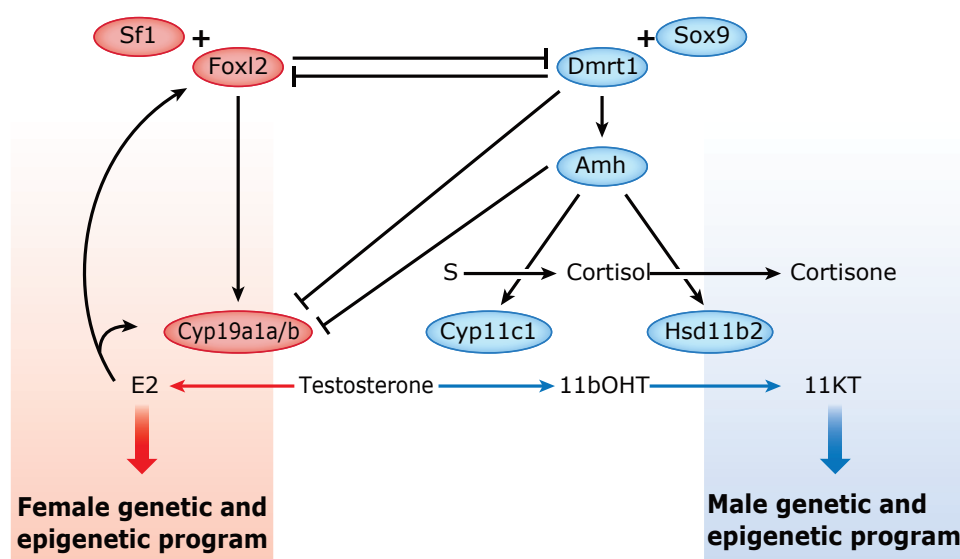


Figure 4.1. Modèle de réseaux de gènes antagonistes spécifiques à chaque sexe contrôlant la différenciation sexuelle chez les poissons.

Des facteurs conservés en aval favorisent la voie de féminisation ou de masculinisation, respectivement, tout en inhibant activement la voie sexuelle opposée. La testostérone peut être convertie soit en 17β -œstradiol (*E2*, l'œstrogène le plus actif chez le poisson) par l'aromatase gonadique (*Cyp19a1a*), soit en 11KT (l'androgène le plus actif chez les poissons) par les enzymes 11β -hydroxylase (*Cyp11c1*) et 11β -hydroxystéroïde déshydrogénase de type 2 (*Hsd11b2*). *Cyp11c1* et *Hsd11b2* sont critiques non seulement pour la synthèse des androgènes 11 -oxygénés mais aussi pour le métabolisme des glucocorticoïdes (par exemple, le cortisol). Abréviations : 11bOHT, 11β -hydroxytestostérone; 11KT, 11-ketotestostérone; *Amh*, anti-Müllerian hormone; *Dmrt1*, doublesex and mab-3 related transcription factor 1; *Foxl2*, forkhead transcriptional factor L2; *S*, 11-déoxycortisol; *Sf1*, steroidogenic factor 1; *Sox9*, SRY-related HMG box 9. Issu et traduit de Ortega-Recalde et al., (2020), déjà adapté de Todd et al., (2016).

De plus, à un stade précoce de la différenciation, il peut déjà exister un dimorphisme dans l'expression des gènes clés de la cascade du sexe chez le tilapia du Nil (Poonlaphdecha et al., 2011; Zhao et al., 2019). L'étude de l'expression des gènes candidats pourrait donc nous permettre de comparer le transcriptome de cette population Ethiopienne Hora avec celui des souches domestiques dont le sexe est déterminé par LG23-Y. Enfin, nous avons également analysé l'effet de la température sur le transcriptome de cette population pendant le développement des alevins.

Matériels et méthodes

Des individus d'*O. niloticus cancellatus* originaires du lac Hora (Ethiopie) ont été maintenus pendant plusieurs années au CIRAD à Montpellier, sans sélection particulière ou domestication particulières. En 2016, une descendance a été produite par reproduction naturelle entre deux géniteurs. Les œufs fertilisés se sont développés en conditions contrôlées et au stade de 10 jours après fécondation (10 dpf) les alevins ont été séparés en deux groupes : un groupe contrôle, maintenu dans un bassin à 27°C et un groupe traité par la température, maintenu dans un bassin à 36°C. À 14 dpf et 23 dpf pendant la période de différenciation du sexe où notamment l'expression de *l'amh* est déjà dimorphique (Poonlaphdecha et al., 2011), les gonades (toujours attachées au tronc), ainsi que les cerveaux (contenus dans la tête entière) ont été prélevés (n=16/lot) des deux groupes expérimentaux. A 14 dpf, 8 individus (8 têtes et 8 troncs) à 27°C et 9 troncs d'individus à 36°C (contenant potentiellement des mâles XX inversés par la température) ont servi à produire des banques de RNAseq, séquencées en lectures paires de 100bp.

L'ensemble des données de RNAseq de chaque individu a suivi le même traitement bioinformatique, à savoir : un nettoyage des lectures, un alignement sur le génome de référence du tilapia du Nil, les lectures alignées sur chaque locus pour chaque individu sont dénombrées et analysées sur R pour comparer les différences d'expressions entre tissus (cerveaux vs troncs) et traitements températures (27°C vs 36°C). Nous avons également procédé à une identification des voies métaboliques sur- ou sous-représentées selon les tissus et les températures d'exposition.

Enfin, l'ensemble de ces analyses (choix des individus et étapes bioinformatiques) ont dû être menées sans *a priori* sur le sexe phénotypique des individus puisqu'ils sont trop petits à

14 dpf pour être sexés sur la base de la morphologie des gonades, et sans *a priori* génotypiques puisque les marqueurs de la duplication de l'*amh* chez les mâles ne sont pas fiables pour les populations Ethiopiennes (même celle présentant la duplication de la région de l'*amh* sur LG23-Y).

Afin d'identifier de potentielles modifications de la cascade du déterminisme du sexe chez Hora nous avons également porté une attention particulière aux niveaux d'expressions dans le cerveau et les gonades de certains gènes préalablement identifiés comme différentiellement exprimés entre mâles et femelles ou appartenant à la cascade du déterminisme du sexe chez des espèces proches.

Des analyses de PCR quantitative en temps réelles (qPCR) ont aussi été mené sur les troncs à 14 et 23 dpf en utilisant des amorces de gènes clefs de la différenciation testiculaire (*gsdf*, *dmrt1*, *amh* et *arα*) et de la différenciation ovarienne (*foxl2*, *cyp19a1a*, *cyp19alb*). Nous avons aussi vérifié la correspondance entre les analyses RNASeq et qPCR.

Principaux résultats et discussion

i. Les fortes températures impactent les voies métaboliques de la synthèse des stéroïdes dans les gonades

Nous avons observés des différences d'expressions des cascades de production de stéroïdes dans les gonades chez les individus traités à la température, en comparaison de ceux maintenus à température contrôle. Les voies de biosynthèse des stéroïdes étaient et de la biosynthèse des hormones stéroïdiennes était particulièrement impactées par les hautes températures.

Les hormones stéroïdiennes sont impliquées dans le développement embryonnaire, la différenciation du sexe, la reproduction ainsi que la réponse au stress chez les poissons téléostéens (Goikoetxea et al., 2017). On retrouve entre autre les œstrogènes, qui sont synthétisés à partir des androgènes par l'enzyme aromatase codée par les gènes *cyp19a1a* et *cyp19a1b* (Baroiller et al., 1999; Guiguen et al., 2010). Les ovaires pendant la différenciation chez le tilapia du Nil ont des niveaux plus élevés d'estradiol-17 et de *cyp19a1a* par rapport aux testicules (Baroiller et D'Cotta, 2001; D'Cotta et al., 2001). Nous avons observé que la température avait un effet répressif sur l'expression de *cyp19a1a* à 23 dpf dans les troncs de Hora à 36°C ce qui n'était pas le cas dans les troncs à 27°C. Une répression similaire de

l'expression de *cyp19a1a* (affectant les niveaux d'œstrogènes), a été montrée dans les gonades de tilapia XX masculinisé par la température (traitées à 36°C) (D'Cotta et al., 2001; Poonlaphdecha et al., 2013; Sun et al., 2018).

Parmi ces hormones stéroïdiennes on retrouve également le cortisol, un glucocorticoïde qui est considéré comme « l'hormone du stress » chez les vertébrés. L'augmentation de la production de cortisol a déjà été mise en lien avec des conditions de fortes températures chez le Pejerrey et le Medaka notamment (Hattori et al., 2009; Yamaguchi et al., 2010). De plus, l'augmentation du cortisol peut impacter la cascade de différenciation du sexe en influant par exemple sur la régulation du gène de l'aromatase (*Cyp19a1a*) ou encore du gène de la 11 β -hydroxysteroid dehydrogenase type 2 (*Hsd11b2*) (Wang et al., 2007; Quek & Chan, 2009; Rajakumar & Senthilkumaran, 2016; Goikoetxea et al., 2017). Or, la 11 β -hydroxylase, codée par le gène *cyp11c1*, est impliquée dans la synthèse du cortisol mais aussi de la 11-Kétotestostérone, le principal androgène chez les téléostéens (Zheng et al., 2020). Enfin, les niveaux d'expression de *cyp11c1* et *11 β -hsd* montrent des augmentations considérables chez de nombreux individus à 36°C, suggérant que de forte température puissent induire une masculinisation liée aux niveaux de cortisol ou de 11-kétotestostérone. En revanche, les températures du milieu ne dépassent pas la température connue de masculinisation du sexe (32°C) et ce phénomène n'est peut-être pas représentatif des conditions naturelles (Baroiller & D'Cotta, 2001).

ii. Certains « suspects habituels » de la différenciation du sexe sont différenciellement exprimés dans les gonades exposées à de fortes températures

Entre les gonades dans les deux conditions de température, nous avons observé des expressions différentielles de certains gènes qui s'avèrent également faire partie des locus régulièrement retrouvés dans la cascade du déterminisme et de la différenciation du sexe (Böhne et al., 2016; Pan et al., 2016). Parmi eux, des membres de la famille des « Jumonji », des « Heatshock protein ».

Certains gènes de la famille des Jumonji ont déjà été identifiés chez les reptiles dans les processus de réversion du sexe (Deveson et al., 2017). Des changements d'expression de gènes de la famille des « Heatshock proteins » ont lieu chez certains poissons pendant les périodes les plus chaudes de l'année en milieu naturel (Domingos et al., 2013). Elles sont

également liées à la prolifération plus ou moins altérée des cellules germinales dans la région des gonades chez le danio notamment (Pfeiffer et al., 2018) et de fortes températures entraînent des modifications d'expression des heatshock protein dans les gonades mâles chez l'Anguille de marais asiatique (He et al. 2010). Enfin, on retrouve aussi *Kdm6bb*, qui chez la tortue de Floride (dont le déterminisme du sexe est dépendant de la température) favorise la transcription du gène du déterminisme du sexe mâle *dmrt1* (Ge et al., 2018). La répression de l'expression de l'aromatase *cyp19a1a* à 23dpf dans les troncs à 36°C ainsi que de l'augmentation des gènes de *gsdf*, *dmrt1*, *cyp11c1*, *11β-hsd* et de *kdm6bb* et *jarid2*, tend également à montrer que les individus Hora sont sensibles à la températures et sujets à la masculinisation.

iii. *Gsdf* permet d'identifier deux groupes au sein de chaque traitement température, suggérant un rôle important dans la cascade de différenciation du sexe

Gsdf (gonadal soma-derived factor) a déjà régulièrement été identifié dans les cascades de différenciation mâle chez les poissons et notamment chez le médaka, le Danio, le Bar ou le Cardeau hrame (Crespo et al., 2013; Zhang et al., 2016; Horie et al., 2016; Yan et al., 2017; Yang et al., 2019). L'expression de *gsdf* chez le tilapia du Nil est dimorphique entre mâles et femelles avant même la différenciation morphologique des gonades (Kaneko et al. 2015) et a été mis en lien avec la masculinisation dans les souches de laboratoires (Jiang et al., 2016).

Parmi les gènes candidats du déterminisme et de la différenciation du sexe, *gsdf* se détache des autres dans nos analyses sur les gonades.

Nous avons constaté que les différences de niveaux d'expressions pour ce gène permettent de former deux groupes : les individus à 27 et 36°C présentant une forte expression de *gsdf* et ceux à 27 et 36°C ne présentant pas une très forte expression de *gsdf* et que ce profil plus marqué encore à 23 dpf. Ces deux groupes observés pourraient correspondre aux sexes phénotypiques et les individus présentant de fortes expressions seraient des mâles en suivant cette logique. Ce résultat indiquerait donc un rôle important et conservé de *gsdf* au stade précoce du développement chez la population Hora même si celle-ci ne présente pas le même déterminant majeur du sexe que les autres populations sauvages analysées.

De plus, on note que la variance des expressions entre les individus à 36°C est bien supérieure à celle des individus à température contrôle. Il est possible que cela soit dû à la présence d'individus dont le sexe est inversé par la température et ainsi que le groupe des mâles 36°C soit en réalité composé d'individus XX et XY qui n'auraient pas exactement les mêmes expressions. *Gsdf* est d'ailleurs plus fortement exprimé chez les individus à 36°C que chez les individus à 27°C, or une augmentation de son expression par la température a déjà été identifiée comme élément déclencheur du développement de mâles XX chez le tilapia du Nil (Yang et al., 2019, 2020). Malheureusement, comme nous ne disposons pas de marqueurs pour sexer génotypiquement les individus, nous n'avons pas été en mesure d'identifier ou non la présence de mâles XX dans le groupe des potentiels mâles 36°C.

iv. Parmi les gènes candidats, *l'amh* ne contribue pas à former les groupes phénotypiques potentiels identifiés

Contrairement aux résultats de Poonlaphdecha et al., (2011) montrant un clair dimorphisme de l'expression de *l'amh* dans le cerveau entre 10 et 15 dpf entre les mâles XY et les femelles XX, nous n'avons pas pu observer de clustering dans les cerveaux des individus. *L'amh* n'était donc pas suffisamment dimorphique pour pouvoir former deux groupes d'individus correspondant aux sexes phénotypiques. Comme *l'amh* ne semble pas être le déterminant du sexe dans la population Hora, le fait qu'elle n'apparaisse pas comme fortement dimorphique n'est peut-être pas si surprenant. Cependant, nous n'avons pas pu observer de clustering qui pourrait correspondre au sexe phénotypique en utilisant un set entier de gènes distinguant mâles et femelles dans les cerveaux chez le tilapia du Nil selon Zhao et al., (2019). Une raison pourrait être que l'analyse sur les cerveaux que nous avons effectuée était biaisée par le faible nombre d'individus, et le fort nombre de gènes analysés en parallèle.

Conclusion

Cette étude de la différenciation sexuelle chez le tilapia du Nil Hora a montré que *gsdf* est déjà exprimé à 14 dpf et son expression semble distinguer deux groupes d'individus identifiés comme potentiels mâles et potentielles femelles. En revanche, ni *dmrt1* ni *l'amh* ne semble contribuer à identifier ces potentiels groupes phénotypiques. Ceci indique que la différenciation sexuelle chez le tilapia Hora suit une voie différente de celle que l'on trouve

généralement chez les souches domestiques de tilapia du Nil. Des différences ont également été observées dans les têtes, puisque l'expression de *l'amh* ne semble pas dimorphique à 14 dpf. Nous avons également trouvé que les fortes températures dans les troncs influençait l'expression de plusieurs gènes identifiés comme des régulateurs épigénétiques tels que les modificateurs d'histone *kdm6bb* et les gènes *jarid2*, liés à la synthèse du cortisol et de la 11-ketotestostérone. Ces résultats, ainsi que l'augmentation des niveaux d'expression de *gsdf*, *dmrt1* et la répression du *cyp19a1a* ovarien dans les gonades potentiellement mâles, suggèrent que des températures élevées induisent une masculinisation dans cette famille Hora. Enfin, cette famille semble thermosensible bien qu'elle vive dans un lac froid.

Références

- Baroiller, J.F., and D'Cotta, H. (2001). Environment and sex determination in farmed fish. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 130, 399–409.
- Baroiller, J.-F., Guiguen, Y., and Fostier, A. (1999). Endocrine and environmental aspects of sex differentiation in fish. *CMLS, Cell. Mol. Life Sci.* 55, 910–931.
- Bezault, E., Clota, F., Derivaz, M., Chevassus, B., and Baroiller, J.-F. (2007). Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture* 272, S3–S16.
- Bezault, E., Balaesque, P., Toguyeni, A., Fermon, Y., Araki, H., Baroiller, J.-F., and Rognon, X. (2011). Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genetics* 12, 102.
- Böhne, A., Wilson, C.A., Postlethwait, J.H., and Salzburger, W. (2016). Variations on a theme: Genomics of sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 17.
- Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Travisan, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., et al. (2019). Fine mapping using whole-genome sequencing confirms anti-müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3 (Bethesda)* 9, 3213–3223.
- Crespo, B., Gómez, A., Mazón, M.J., Carrillo, M., and Zanuy, S. (2013). Isolation and characterization of *Ff1* and *Gsdf* family genes in European sea bass and identification of early gonadal markers of precocious puberty in males. *General and Comparative Endocrinology* 191, 155–167.
- D'Cotta, H., Fostier, A., Guiguen, Y., Govoroun, M., and Baroiller, J.-F. (2001). Aromatase plays a key role during normal and temperature-induced sex differentiation of tilapia *Oreochromis niloticus*. *Molecular Reproduction and Development* 59, 265–276.
- Deveson, I.W., Holleley, C.E., Blackburn, J., Marshall Graves, J.A., Mattick, J.S., Waters, P.D., and Georges, A. (2017). Differential intron retention in *Jumonji* chromatin modifier genes is implicated in reptile temperature-dependent sex determination. *Sci. Adv.* 3, e1700731.
- Domingos, F.F.T., Thomé, R.G., Martinelli, P.M., Sato, Y., Bazzoli, N., and Rizzo, E. (2013). Role of *HSP70* in the regulation of the testicular apoptosis in a seasonal breeding teleost *Prochilodus argenteus* from the São Francisco River, Brazil. *Microsc Res Tech* 76, 350–356.
- El Taher, A., Ronco, F., Matschiner, M., Salzburger, W., and Böhne, A. (2020). Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes (Evolutionary Biology).
- Eshel, O., Shirak, A., Weller, J.I., Hulata, G., and Ron, M. (2012). Linkage and physical mapping of sex Rregion on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3: Genes, Genomes, Genetics* 2, 35–42.
- Eshel, O., Shirak, A., Dor, L., Band, M., Zak, T., Markovich-Gordon, M., Chalifa-Caspi, V., Feldmesser, E., Weller, J.I., Seroussi, E., et al. (2014). Identification of male-specific *amh*

- duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 15, 774.
- FAO (2021). The State of World Fisheries and Aquaculture 2021 (FAO).
- Gammerdinger, W.J., and Kocher, T.D. (2018). Unusual diversity of sex chromosomes in African cichlid fishes. *Genes* 9, 480.
- Gammerdinger, W.J., Conte, M.A., Acquah, E.A., Roberts, R.B., and Kocher, T.D. (2014). Structure and decay of a proto-Y region in Tilapia, *Oreochromis niloticus*. *BMC Genomics* 15, 975.
- Ge, C., Ye, J., Weber, C., Sun, W., Zhang, H., Zhou, Y., Cai, C., Qian, G., and Capel, B. (2018). The histone demethylase *KDM6B* regulates temperature-dependent sex determination in a turtle species. *Science*.
- Goikoetxea, A., Todd, E.V., and Gemmell, N.J. (2017). Stress and sex: does cortisol mediate sex change in fish? *Reproduction* 154, R149–R160.
- Guiguen, Y., Fostier, A., Piferrer, F., and Chang, C.-F. (2010). Ovarian aromatase and estrogens: A pivotal role for gonadal sex differentiation and sex change in fish. *General and Comparative Endocrinology* 165, 352–366.
- Hattori, R.S., Fernandino, J.I., Kishii, A., Kimura, H., Kinno, T., Oura, M., Somoza, G.M., Yokota, M., Strüßmann, C.A., and Watanabe, S. (2009). Cortisol-Induced Masculinization: Does Thermal Stress Affect Gonadal Fate in Pejerrey, a Teleost Fish with Temperature-Dependent Sex Determination? *PLOS ONE* 4, e6548.
- He, Y., Shang, X., Sun, J., Zhang, L., Zhao, W., Tian, Y., Cheng, H., and Zhou, R. (2010). Gonadal apoptosis during sex reversal of the rice field eel: implications for an evolutionarily conserved role of the molecular chaperone heat shock protein 10. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 314B, 257–266.
- Horie, Y., Myosho, T., Sato, T., Sakaizumi, M., Hamaguchi, S., and Kobayashi, T. (2016). Androgen induces gonadal soma-derived factor, *Gsdf*, in XX gonads correlated to sex-reversal but not *Dmrt1* directly, in the teleost fish, northern medaka (*Oryzias sakaizumii*). *Molecular and Cellular Endocrinology* 436, 141–149.
- Jiang, D.-N., Yang, H.-H., Li, M.-H., Shi, H.-J., Zhang, X.-B., and Wang, D.-S. (2016). *gsdf* is a downstream gene of *dmrt1* that functions in the male sex determination pathway of the Nile tilapia. *Molecular Reproduction and Development* 83, 497–508.
- Kaneko, H., Ijiri, S., Kobayashi, T., Izumi, H., Kuramochi, Y., Wang, D.-S., Mizuno, S., and Nagahama, Y. (2015). Gonadal soma-derived factor (*gsdf*), a TGF-beta superfamily gene, induces testis differentiation in the teleost fish *Oreochromis niloticus*. *Molecular and Cellular Endocrinology* 415, 87–99.
- Lee, B.-Y., Penman, D.J., and Kocher, T.D. (2003). Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics* 34, 379–383.
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y

- chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. PLOS Genetics 11, e1005678.
- Ndiwa, T.C., Nyingi, D.W., and Agnese, J.-F. (2014). An Important Natural Genetic Resource of *Oreochromis niloticus* (Linnaeus, 1758) Threatened by Aquaculture Activities in Lobo Drainage, Kenya. PLOS ONE 9, e106972.
- Ortega-Recalde, O., Goikoetxea, A., Hore, T.A., Todd, E.V., and Gemmell, N.J. (2020). The Genetics and Epigenetics of Sex Change in Fish. Annual Review of Animal Biosciences 8, null.
- Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2013). Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. PLoS ONE 8, e68389.
- Palaiokostas, C., Bekaert, M., Khan, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2015). A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). BMC Genomics 16, 171.
- Pan, Q., Anderson, J., Bertho, S., Herpin, A., Wilson, C., Postlethwait, J.H., Scharl, M., and Guiguen, Y. (2016). Vertebrate sex-determining genes play musical chairs. Comptes Rendus Biologies 339, 258–262.
- Pfeiffer, J. (2018). Rapid progression through the cell cycle ensures efficient migration of primordial germ cells – The role of *Hsp90*. Developmental Biology 10.
- Poonlaphdecha, S., Pepey, E., Huang, S.-H., Canonne, M., Soler, L., Mortaji, S., Morand, S., Pfennig, F., Mélard, C., Baroiller, J.F., et al. (2011). Elevated *amh* Gene Expression in the Brain of Male Tilapia (*Oreochromis niloticus*) during Testis Differentiation. Sex Dev 5, 33–47.
- Poonlaphdecha, S., Pepey, E., Canonne, M., de Verdal, H., Baroiller, J.-F., and D'Cotta, H. (2013). Temperature induced-masculinisation in the Nile tilapia causes rapid up-regulation of both *dmrt1* and *amh* expressions. General and Comparative Endocrinology 193, 234–242.
- Quek, S.-I., and Chan, W.-K. (2009). Transcriptional activation of zebrafish *cyp11a1* promoter is dependent on the nuclear receptor *Ff1b*. Journal of Molecular Endocrinology 43, 121–130.
- Rajakumar, A., and Senthilkumaran, B. (2016). *Sox3* binds to 11 β -hydroxysteroid dehydrogenase gene promoter suggesting transcriptional interaction in catfish. The Journal of Steroid Biochemistry and Molecular Biology 158, 90–103.
- Sissao, R., D'Cotta, H., Baroiller, J.-F., and Toguyeni, A. (2019). Mismatches between the genetic and phenotypic sex in the wild Kou population of Nile tilapia *Oreochromis niloticus*. PeerJ 7.
- Sun, L.X., Teng, J., Zhao, Y., Li, N., Wang, H., and Ji, X.S. (2018). Gonad Transcriptome Analysis of High-Temperature-Treated Females and High-Temperature-Induced Sex-Reversed Neomales in Nile Tilapia. International Journal of Molecular Sciences 19, 689.
- Taslina, K., Wehner, S., Taggart, J.B., de Verdal, H., Benzie, J.A.H., Bekaert, M., McAndrew, B.J., and Penman, D.J. (2020). Sex determination in the GIFT strain of tilapia is controlled by a locus in linkage group 23. BMC Genet 21, 49.

- Taslina, K., Khan, M.G.Q., McAndrew, B.J., and Penman, D.J. (2021). Evidence of two XX/XY sex-determining loci in the Stirling stock of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 532, 735995.
- Todd, E.V., Liu, H., Muncaster, S., and Gemmell, N.J. (2016). Bending Genders: The Biology of Natural Sex Change in Fish. *Sex Dev* 10, 223–241.
- Trewavas, E. (1983). Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*.
- Triay, C., Conte, M.A., Baroiller, J.-F., Bezault, E., Clark, F.E., Penman, D.J., Kocher, T.D., and D’Cotta, H. (2020). Structure and Sequence of the Sex Determining Locus in Two Wild Populations of Nile Tilapia. *Genes* 11, 1017.
- Wang, D.-S., Kobayashi, T., Zhou, L.-Y., Paul-Prasanth, B., Ijiri, S., Sakai, F., Okubo, K., Morohashi, K., and Nagahama, Y. (2007). *Foxl2* up-regulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1. *Mol Endocrinol* 21, 712–725.
- Wessels, S., Krause, I., Floren, C., Schütz, E., Beck, J., and Knorr, C. (2017). ddRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC Genomics* 18, 531.
- Yamaguchi, T., Yoshinaga, N., Yazawa, T., Gen, K., and Kitano, T. (2010). Cortisol Is Involved in Temperature-Dependent Sex Determination in the Japanese Flounder. *Endocrinology* 151, 3900–3908.
- Yan, Y.-L., Desvignes, T., Bremiller, R., Wilson, C., Dillon, D., High, S., Draper, B., Buck, C.L., and Postlethwait, J. (2017). Gonadal soma controls ovarian follicle proliferation through *Gsdf* in zebrafish. *Dev. Dyn.* 246, 925–945.
- Yang, Y., Liu, Q., Xiao, Y., Xu, S., Wang, X., Yang, J., Song, Z., You, F., and Li, J. (2019). High temperature increases the *gsdf* expression in masculinization of genetically female Japanese flounder (*Paralichthys olivaceus*). *General and Comparative Endocrinology* 274, 17–25.
- Yang, Y., Liu, Q., Xiao, Y., Xu, S., Wang, X., Yang, J., Song, Z., You, F., and Li, J. (2020). Effects of environmental stress (sex steroids and heat) during sex differentiation in Japanese flounder (*Paralichthys olivaceus*): insight from germ cell proliferation and *gsdf-amh-cyp19a1a* expression. *Aquaculture* 515, 734536.
- Zhang, X., Guan, G., Li, M., Zhu, F., Liu, Q., Naruse, K., Herpin, A., Nagahama, Y., Li, J., and Hong, Y. (2016). Autosomal *gsdf* acts as a male sex initiator in the fish medaka. *Scientific Reports* 6, 19738.
- Zhao, Y., Chen, H.J., Wang, Y.Y., Mei, Y., Huang, L.B., Wang, H., and Ji, X.S. (2019). Gonad development examination of high-temperature-treated genetically female Nile tilapia. *Aquaculture* 515, 734535.
- Zheng, Q., Xiao, H., Shi, H., Wang, T., Sun, L., Tao, W., Kocher, T.D., Li, M., and Wang, D. (2020). Loss of *Cyp11c1* causes delayed spermatogenesis due to the absence of 11-ketotestosterone. *Journal of Endocrinology* 244, 487–499.

PUBLICATION ASSOCIÉE

Transcriptome analyses of a wild population of Nile tilapia lacking the ancestral LG23 Y suggests *gsdf* plays a key role in their sex differentiation

Cécile Triay^{1,2}, Astrid Böhne³, Jean-François Baroiller^{1,2}, Matteo Barbarini^{1,2}, Thomas D. Kocher⁴ & Helena D'Cotta^{1,2}

1. UMR-Institut des Sciences de l'Evolution de Montpellier, Centre National de la Recherche Scientifique, Institut de Recherche pour le Développement, Ecole Pratique des Hautes Etudes, University of Montpellier, 34090 Montpellier, France; cecile.triay@umontpellier.fr (C.T); baroiller@cirad.fr (J-F.B.); dcotta@cirad.fr (H.D)
2. UMR-Institut des Sciences de l'Evolution de Montpellier, Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Campus Int. Baillarguet, 34398 Montpellier, France
3. Center for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, Bonn, Germany; a.boehne@leibniz-zfmk.de (A.B)
4. Department of Biology, University of Maryland, College Park, MD 20742, USA; tdk@umd.edu (T.D.K.)

Abstract: In Nile tilapia, the sex-determining region is composed of two copies of the *amh* gene (*amhY* and *amhΔY*) and has been found on Linkage group (LG) 23-Y in many domesticated strains and recently also in an Ethiopian wild population of Lake Koka. However, Lake Hora, another Ethiopian wild population, does not have the *amh* duplication and no association to sex was found on LG23. Here we studied the downstream gene network of sex-differentiation at 27°C in a Lake Hora offspring and compared it to what is described in domesticated tilapia strains. Part of the offspring was treated to 36°C to study early effects of high-temperature on the sex-differentiating pathway in order to analyze the persistence of thermosensitivity occurring in this cold lake (22-24°C). Heads and trunks (containing gonads) were sampled at the pivotal stage of 14 days post-fertilization (dpf) and at 23 dpf. RNASeq of 14 dpf heads showed no cluster division of sexual phenotypes. In trunks in contrast, we identified 3018 genes up-regulated in 36°C and 1319 genes down-regulated compared to 27°C with impacts on the pathways of the steroid synthesis (including sex and stress hormones) as well as of replication and repair. Subsets of genes, previously identified in sex-differentiation pathways were used to look for a putative sex clustering. Expression levels of *gsdf*, *nr3c1* and *kdm6bb* formed clusters interpreted as phenotypic sex. However, neither *dmrt1* nor *amh* expressions were dimorphic in 27°C trunks of putative males and females. This suggests that the sex-differentiating pathway is not fully conserved in Hora tilapia. qPCR detected low expression of the ovarian *cyp19a1a* gene at 14 dpf with increases observed in some trunks at 23 dpf. The *cyp19a1a* repression in the 36°C trunks at 23 dpf together with the increase of *gsdf*, *dmrt1*, *jarid2*, *kdm6bb*, *nr3c1*, *11β-hsd* and *cyp11c1* (involved in 11-ketotestosterone synthesis), all suggest that temperature induced masculinization of this Hora family. The Hora population is hence probably thermosensitive but sex must be determined genetically since we obtained a 50:50 sex ratio at 27°C.

Keywords: Sex-differentiation; *gsdf*; transcriptomic; Nile tilapia

1. Introduction

The Nile tilapia (*Oreochromis niloticus*), is a cichlid fish with a broad distribution across Africa, found from Senegal to the Nile basins, covering the Valleys of the Ethiopian and part of the Northern East African Rifts, as well as the Kivu and North Tanganyika provinces (Trewavas, 1983; Bezault et al., 2011). It has adapted to freshwaters with a wide range of temperatures from cold (17 to 26°C), over temperate (28 to 34°C) to hot springs (36 to 40°C) (Bezault et al., 2007; Ndiwa et al., 2014). Nile tilapia has become an aquaculture fish of global importance with 5.5 million tons produced in 2019, (FAO, 2021). Despite the economic interest in producing males as the larger and hence more profitable sex, sex determination in this species is complex, and the regulation of its sex-differentiating network is still unclear. While this species has a genetic sex-determination with an XX/XY system (Jalabert et al., 1974), minor genetic factors brought by both parents also contribute to offspring sex-ratios (Baroiller & D'Cotta, 2001). In addition, temperature can also influence sex differentiation and override the genetic sex determination signal. Water temperatures above 32°C during the time window from 10 to 20 days post fertilization (dpf) prior to the beginning of gonadal sex differentiation can induce sex-reversal of XX individuals into viable and fertile phenotypic males (Baroiller et al., 1995, 2009a). This thermosensitivity, considered as the proportion of males induced by temperature, is a heritable trait that can be selected for in a few generations (Baroiller and D'Cotta, 2001; Wessels and Hörstgen-Schwark, 2007). It has been found in all wild Nile tilapia populations studied, whatever their thermal regime, but presenting different intensity of response (Bezault et al., 2007; Baroiller et al., 2009a). Furthermore, putative XX males (siring all-female progenies) have been reported in several wild populations (Bezault et al., 2007; Sissao et al., 2019).

Sex determination (SD) has mainly been studied in domesticated (laboratory or commercial) strains of Nile tilapia. A sex-determining XY locus has been found either on linkage group (LG)1 (Lee et al., 2003; Palaikostas et al., 2013, 2015; Gammerdinger et al., 2014; Taslima et al., 2020) or LG23 (Eshel et al., 2012, 2014; Li et al., 2015; Wessels et al., 2017; Cáceres et al., 2019; Taslima et al., 2020), depending on the strain. The master sex-determining gene on LG23 corresponds to a tandem duplication of the *anti-Müllerian hormone* (*amh*) gene, consisting in the *amhy* gene and a supposedly truncated *amhΔY* gene (Eshel et al., 2014; Li et al., 2015).

Recruitment of new master SD genes in teleosts is not unusual and has occurred mainly through the duplication of genes from the sex-differentiating network (Guiguen et al., 2018). *Amh* belonging to the TGF-β signaling pathway has been duplicated not only in the Nile tilapia but also in three *Atherinidae* (Hattori et al., 2012; Yamamoto et al., 2014; Bej et al., 2017) and in some pikes from the esocids family (Pan et al., 2019, 2021). Other genes from the TGF-β superfamily have also risen to become master SD alleles or genes, such as *amhr2* (anti-Müllerian hormone receptor type 2) (Kamiya et al., 2012; Feron et al., 2020), *gsdf* (gonadal soma derived factor) (Myosho et al., 2012; Fowler and Buonaccorsi, 2016) and *gdf6* (growth differentiation factor 6) (Reichwald et al., 2015). Transcription factors involved in the sex-

differentiating network have also become Y-linked master genes such as *dmy/dmrt1by* which originated from the duplication of the *dmrt1* (doublesex and mab-3 related transcription factor 1) gene in the medaka *Oryzias latipes* (Matsuda et al., 2002; Nanda et al., 2002) and the cis-regulator of *sox3Y* (SRY sex determining region Y-box 3) in *Oryzias dancena* (Takehana et al., 2014). In the genus *Seriola* the female master gene is *hsd17b1* (17-beta-steroid dehydrogenase) involved in steroid biosynthesis (Purcell et al., 2018; Koyama et al., 2019). The reuse of the same and of a small number of sex-differentiating genes has led to the “limited options” hypothesis (Marshall Graves and Peichel, 2010). An exception is the salmonid *sdY* (sexually dimorphic on the Y) gene, which is probably a modified paralog of an immune gene (Yano et al., 2012). *SdY* integrated the SD network by disrupting the usual female differentiating pathway (preventing activation of the aromatase *cyp19a1a* promoter), causing testicular differentiation (Bertho et al., 2018).

We have initiated the study of SD in wild populations of Nile tilapia. Findings that sex-determining loci are associated to either LG1 or LG23 in domesticated strains may stem from wild populations’ differences. In addition, because several populations of Nile tilapia have already adapted to high temperatures, it is an ideal model to understand how climate changes are affecting fish sex determination and differentiation, and to follow the dynamics of sex-ratios and population resilience to future climate changes. We used on several wild populations (Baroiller and D’Cotta, 2016) PCR-based *amh* markers that discriminated the X and the Y chromosomes successfully in several domesticated strains (Eshel et al., 2014; Sun et al., 2014; Li et al., 2015; Sissao et al., 2019). However, they proved inefficient in some populations (Sissao et al., 2019; Triay et al., 2020). The *amh* genotyping in Lake Kou (Burkina Faso) assigned the sex correctly in only 82% of wild individuals (Sissao et al., 2019). Furthermore, some of the sex ratios obtained with Lake Kou broodstock could not be explained by LG23 only (Sissao et al., 2019). In Ethiopian populations, we were unable to genotype the sex of individuals, which we subsequently proved during this thesis, were due to the absence of indels and other sequence differences in the *amh* duplicated region between males and females (Triay et al., 2020; Triay et al., unpublished – Chapter 2).

We therefore, decided to investigate further the Ethiopian populations at the beginning of this thesis, studying their sex determination in Lake Koka and Lake Hora (Chapter 1, & 2). In parallel, we carried out here a transcriptome analysis to study the sex-differentiating network in the Ethiopian population of Lake Hora, to see whether differences in the regulation of sex-differentiating genes might be occurring. Lake Hora is an isolated small crater lake, that like Lake Koka are part of the reservoir of the Awash River in the Ethiopian Highlands (altitude of 1700 m). Both Lakes are considered to be “cold” as temperatures vary throughout the year between 17-26°C for Lake Koka and between 22-24.5 °C for Lake Hora (Bezault, 2005; Bezault et al., 2007). Despite these cold temperature habitats, Lake Koka displayed a thermosensitivity of sex-differentiation with high temperatures increasing the proportion of males (Bezault et al., 2007).

In this work, we used a RNASeq approach to study the sex-differentiating pathway at 27°C and compared it to the “genetic” pathway described in domesticated tilapia strains

(D’Cotta et al., 2001; Ijiri et al., 2008; Baroiller et al., 2009a; Tao et al., 2013; Sun et al., 2018; Wang et al., 2019; Teng et al., 2020; Yao et al., 2021). We wanted also to search for early sexually-dimorphic expressed genes that could highlight a putative master sex determinant in the Hora Ethiopian population. In addition, we wanted to study early effects of high-temperature on the sex-differentiating pathway in order to analyze the persistence of thermosensitivity in populations living under cold temperature regime as already evidenced in the Koka population (Bezault et al., 2007). Based on the hypothesis of “limited options”, we searched amongst the expressed genes those considered “usual suspects” (Herpin and Scharf, 2015). Our search was performed in the gonads but we extended it to the brain since we have found that it harbors numerous sexually dimorphic genes at early stages in Nile tilapia (Poonlaphdecha et al., 2011; Gennotte et al., 2014).

2. Material and Methods

2.1. Biological Material

Wild individuals of Nile tilapia (*Oreochromis niloticus*, subspecies *cancellatus*) were caught in 2002 in Lake Hora (Ethiopia) from the Awash drainage (Figure 1).

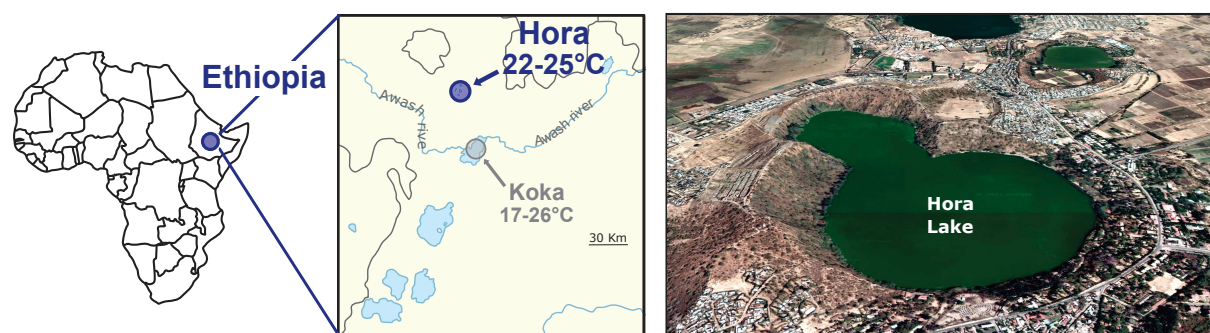


Figure 1. Localization of the Hora wild population and Google Earth aerial view of the Hora Lake.

Sexually mature individuals (phenotypically sexed by their genital papilla), were captured in the wild in 2002 and fin clips taken at capture (n=47). Live fish were transferred to the CIRAD experimental facilities in Montpellier (France) and progressively acclimated to a standard temperature of $27\text{ °C} \pm 1\text{ °C}$ in a thermo-regulated re-circulating system in 200L aquaria. This stock was then maintained for 14 years.

In 2016, a progeny was produced by natural reproduction. Fertilized eggs were taken from the mother’s mouth at 2 days post-fertilization (dpf), placed in McDonald bottles until hatching (4-5 dpf at 27 °C) and then transferred into aquaria. The high temperature treatment of 36 °C was started at 10 dpf, when the gonads in Nile tilapia are bipotential and sensitive to temperature treatments (Baroiller et al., 2009b). For this, we divided the offspring at 10 dpf

into two batches: one batch reared at the control temperature of 27°C and the second batch at 36°C, for 30 days until 40 dpf. Both batches were subsequently reared at 27°C until 160 dpf when the phenotypic sex was defined. Fish were kept under a stable 12h light:12h dark photoperiod and fed *ad libitum*.

For the sampling, the fish were first euthanized with a lethal dose of Eugenol according to legal regulations (directive 2010-63-UE and French decree 2013-118). Fish (n=16/batch) were sampled during the period of sex differentiation at 14 and 23 dpf (D'cotta et al., 2001; Kobayashi and Nagahama, 2009). For each individual, the trunk containing the gonads (after removing the viscera) and the head were dissected under a stereo microscope, placed in a solution of RNAsave (Biological Industries, Israel) and stocked at -20°C until RNA extraction. Caudal ends of the fingerlings were placed in 100% ethanol for genetic analysis. At 160 dpf fish were sexed by macroscopic gonad visualization or by gonadal squash under a microscope. Fin clips were collected from all the fish, placed in 100% ethanol until genomic DNA extraction.

2.2. Ethics Statement

Fish captures were performed with the Ethiopian Law and Veterinary 2002 Agreements. All animal and experimental procedures were performed at the CIRAD facilities in Montpellier in accordance with the French and European legislations on animal research, under the protocol N_2016101810463 and JB personal authorization for animal experimentation N_35–15, both delivered by the French Government.

2.3. RNA extraction

Total RNA was extracted from heads (aiming for brain transcripts) and trunks (aiming for gonad transcripts) with 1 ml TRIzol (Life Technologies, Cergy-Pontoise, France) after homogenization with glass beads on a TissueLyser (Qiagen, Germany) as described in Poonlaphdecha et al., (2011). Extracted RNA was solubilized in 30 µl RNAase free water, and then further purified with the RNA clean-up and concentrator kit (Zymo, Ozyme, France) following the manufacturer's protocol. Removal of residual gDNA was performed with 10 µl of rDNase (Macherey-Nagel, France) on the column during this purification. RNA was eluted in 25 µl of water and then quantified on a Nanodrop (Thermo scientific, Wilmington, Delaware). RNA quality was also verified on a 2100 Bioanalyzer (Agilent, France) using a Total RNA Nano kit (Agilent, France).

2.4. Production of RNA Sequencing libraries

We selected the RNA from 8 fish (heads and trunks) from the 27°C batch, sampled at 14 dpf. From the 36°C batch we sampled 9 individuals, which potentially included early temperature-induced sex reversed males. All RNAs had RINs (RNA integrity number) of > 8.5 on the Bioanalyzer. RNASeq libraries were generated by the GenoEast platform (CERBM

GIE – IGBMC, Illkirch, France) with the Directional mRNASeq samplePrep (Illumina, France) and sequenced on two lanes of a HiSeq4000 (100 bp paired-end).

2.5. Bioinformatic analysis

RNAseq raw reads were quality filtered using Trimmomatic (Bolger et al., 2014) for paired-end reads with the following parameters: *ILLUMINACLIP* 2:30:10:2, *SLIDINGWINDOW* 4:15, *MINLEN* 30 to remove illumina adapters and filter reads shorter than 30bp with a sliding-window approach. Clean reads were mapped with STAR (version 2.5.0b) (Dobin et al., 2013) adding a '*read_len* 50' argument to the latest reference genome available to us at the time (*O. niloticus*_UMD_NMBU, Genbank ID: GCA_001858045.3). Alignment files in SAM output were transformed to BAM, sorted and indexed using SAMtools (version 1.9) (Li et al., 2009).

Mapped reads were transformed into counts per gene with HTSeq-counts (version 0.11.0) (Anders et al., 2015) based on the reference genome annotations of *O. niloticus*. Gene level counts were then processed with R using the Bioconductor DESeq2 package (version 1.28.1) (Love et al., 2014) to analyze differentially expressed (DE) genes between temperature conditions (27°C and 36°C, assuming 27°C as the control temperature) and between head and trunk tissues. At this step, genes with count levels equal to zero for all individuals were removed. All following analyses were carried out with no *a priori* on genetic nor phenotypic sex of samples. Counts were normalized using the rlog function of DESeq2 to transform counts to a log2 scale, minimizing differences between samples for genes with small counts, with respect to library size. Transcripts were considered DE with an absolute value of log2-fold change > 1.

We visualized global gene expression clustering with the Pheatmap package (version 1.0.12) (Kolde, 2015) using a ward.D2 clustering method. PCA plot representations were also used to exploit variance observed between and within trunks' expressions (package Factoextra version 1.0.7) (Kassambara and Mundt, 2020). Analyses were made with count data of all types of tissues (heads and trunks) and temperature treatments (27°C and 36°C) combined and on a subset of trunk data only. We further used count data of all genes, as well as a subset of 141 candidate genes implied in sex determination or sex differentiation pathways in vertebrates (Supplementary Table 1). For heads only, a third subset of genes was analyzed based on the results of Zhao et al., (2019). Among the genes differentially expressed in brains of males and females raised under control temperature in their Chinese GIFT stock of Nile tilapia at 21 dpf, 2887 were properly annotated and could be retrieved to compose this third subset which could be a marker set for phenotypic sex.

KEGG analyses were carried out using the KEGGprofiler package in R (version 1.30.0) in order to identify genes and pathways over-represented at 36°C compared to 27°C control temperature among the differentially expressed genes. GO identifiers of genes were retrieved using the BiomaRt bioconductor package (version 2.44.1) (Durinck et al., 2009).

2.6. Expression analysis by Quantitative real-time PCR

Complementary DNA (cDNA) was generated from 3 μ g of total purified RNA from the heads' and 2 μ g from the trunks. RNA was first DNase treated with 0.5 U Turbo DNase (Ambion, Austin Texas, USA) at 37°C for 30 min, and then inactivated at 70°C for 10 min. RNA was then reverse transcribed by the addition of 0.5 μ g (Oligo dT) and 10 mM dNTP, with the Superscript II enzyme (Life Technologies, France) incubated at 42°C for 1 hr. The resulting cDNA was then diluted five-fold (heads) and four-fold (trunks), (estimated to be 30 ng/ μ l based on initial RNA amounts) for quantitative real-time PCR (qPCR) analysis. A total of twelve genes were analyzed, seven were sex-differentiating or gonad developing pathway genes and five were used to validate RNASeq data. Primers used for these genes' analyses are listed in Supplementary Table 5. Transcript quantification was performed in duplicates in a 10 μ l reaction containing 400nM of each primer, 2 μ l of diluted cDNA and 2x Forget-Me-Not™ EvaGreen® qPCR Master Mix (Biotium, USA), on a Stratagene MX3005 QPCR machine (Stratagene, La Jolla, USA) with the following program: Activation at 95°C for 2 min, and then 40 cycles of 5 sec at 95°C and 30 sec at 60°C, with a final melting curve. A reaction without DNA was systematically used. Normalized gene expression was the ratio of the housekeeping gene *ef1* (Elongation factor 1 alpha) expression calculated as $(1 + E_{ef1})^{C_{ef1}}$, versus the target gene expressions $(1 + E_{target})^{C_{target}}$, where E is the efficiency which was calculated using a standard curve. Values of relative gene expressions were then calculated as a function of the maximum normalized expression (100%), and as Log2 of this maximum expression.

3. Results

We chose to analyze the Hora progeny at 14 dpf as an early stage of differentiation selected for the RNASeq analyses and 23 dpf as a mid-stage of sex-differentiation. Studies on Nile tilapia sex-differentiation performed by our group using the Bouaké strain (D'Cotta et al., 2001) and the Manzala strain (Berishvili et al., 2006; Poonlaphdecha et al., 2013) and by other groups on the Japanese strain (Ijiri et al., 2008; Kobayashi and Nagahama, 2009), have shown that the period comprised between ~9 and 15 dpf corresponds to a stage when the gonads start to be sensitive to estrogen or androgen hormones as well as to temperature treatments. We consider it the critical period when the fate of the gonad is decided, and although the gonads are morphologically undifferentiated, molecular differences are already observed between the future ovary and testis (Ijiri et al., 2008; Baroiller et al., 2009a). At around 17-18 dpf there is active mitosis and primordial germ cells start to multiply in the future ovary, clearly visible at 23 dpf (D'Cotta et al., 2001).

3.1. Genotype and Sex-ratios of the Hora progeny

We had to wait until 160 dpf in order to sex the phenotypes of the 27°C and 36°C

treated batches because of their slow growth, which led to a large mortality seen in both groups due to the aggressive dominance behavior common for this species (Gonçalves-de-Freitas et al., 2019). In the 27°C group the sex-ratio was of 50% males out of 34 sexed individuals, while in the 36°C all the individuals were phenotypic males but this group was reduced to only 11 fish, so this sex-ratio has to be taken with much caution. In order to distinguish between the phenotypic males, the genetic XY males *vs* the temperature-induced XX males, we needed to have X and Y chromosome markers. All the sex PCR markers based on the *amhY* and *amhΔY* indels (Figure 2) that were previously used to genotype the sex of individuals of laboratory stocks (Eshel et al., 2014; Sun et al., 2014; Li et al., 2015; Sissao et al., 2019) and wild populations (Baroiller and D'Cotta, 2016; Sissao et al., 2019) did not discriminate the genetic sexes in Hora adult or fry individuals raised under the experimental temperatures of 27°C and 36°C (Figure 2). Sex-associated SNPs recently described in the other Ethiopian population of Lake Koka (Triay et al., 2020) were not sex-related in the Hora population (Supplementary Figure 1).

We have previously been able to sex the phenotype of several Nile tilapia strains in the heads at 14 dpf because *amh* expression is sexually dimorphic (Poonlaphdecha et al., 2011; Baroiller and D'Cotta, 2016). This analysis was also carried in the Hora progeny before RNA Sequencing, but it proved inefficient. We therefore, randomly selected the individuals for the RNASeq libraries.

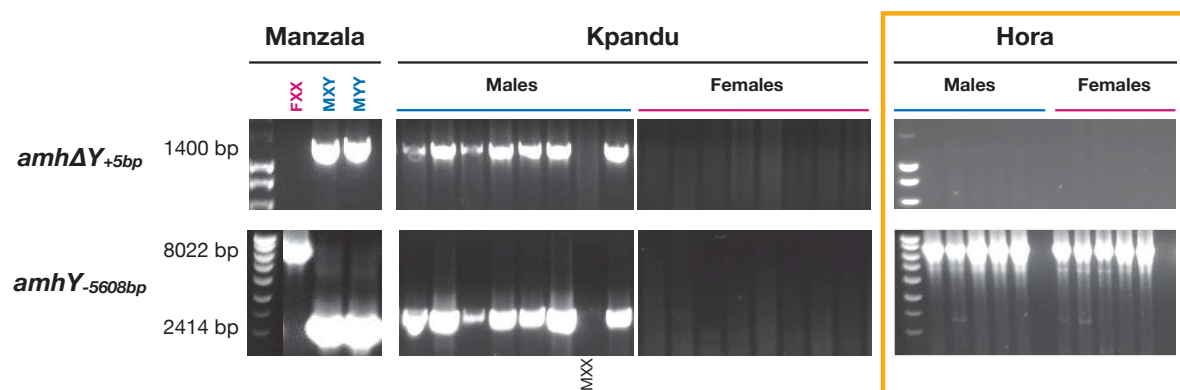


Figure 2. *Amh* genotyping of males and females of the Manzala-Tihange strain, the Kpandu (wild Ghanaian population) and Hora (wild Ethiopian population) adult fish, using two markers. Manzala of known sexual genotypes were used as controls, showing the marker discrimination of the sexes in both this strain and in the Kpandu population however, they do not discriminate between males and females of the Hora samples, studied here.

3.2. Cleaning and mapping statistics

An average of 28 million raw reads per sample were obtained by RNA sequencing. After quality filtering, an average of 26 million reads per individual were retained and mapped on the Nile tilapia reference genome with a mean mapping rate of 94.6% per sample. The

details for each sample can be found in Supplementary Table 2.

3.3. Analysis of all heads and trunks treated to 27°C and 36°C temperatures

3.3.1. Global gene expression clustering

We first did a global analysis of all genes and all samples combined which showed clear clustering with regard to temperature treatment and organ type when using a distance matrix approach (Figure 3). The impact of the organ was more important than temperature since trunks treated at 27°C (T27) or 36°C (T36) clustered together, with the 27°C heads (H27) clustering separately. Clustering varied at the individual level between heads and trunks of the 27°C control temperature treatment *ie*: H27_4 clustered with samples H27_1 and H27_2, whereas T27_4 clustered with samples T27_1, T27_3 and T27_5 (Figure 3). We did not find an additional sub-clustering within the trunks or heads that would correspond to putatively expressed genes between the sexual phenotypes.

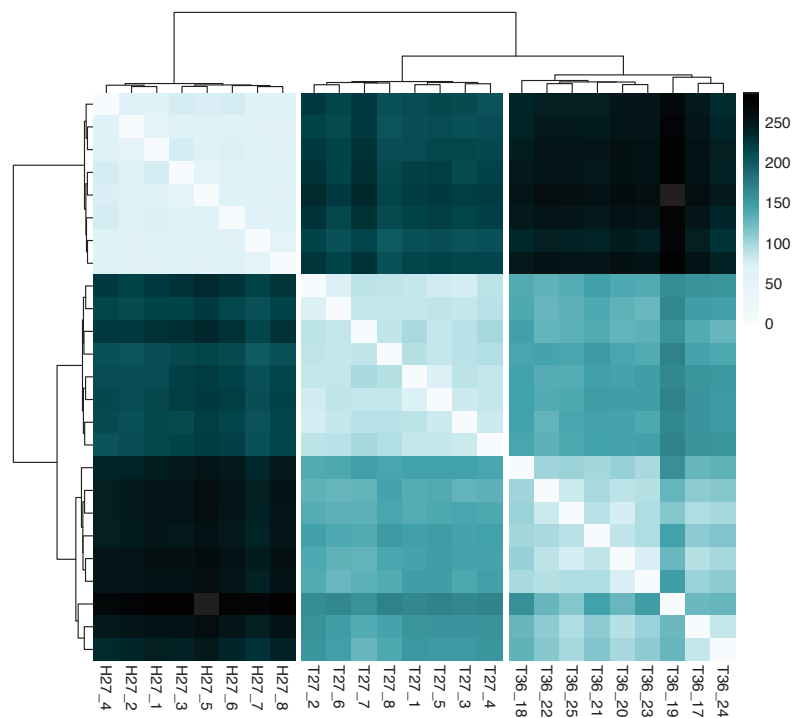


Figure 3. Global expression heatmap using a hierarchical clustering derived from Euclidean distance from normalized and log2 transformed counts and a Ward's distance as agglomeration method.

3.3.2. Vision, endocrine and digestive systems are strongly influencing gene expression under temperature treatment

We worked on the individual trunks because during sex-differentiation the gonads are too small to allow individual analyses, and likewise for the brain, working on whole heads.

This implies that we are “diluting” the transcripts aimed for (gonads and brain) in both these organs, with the transcripts that are more abundant in heads and trunks. We therefore, analyzed the top 50 as well as the top 500 most variable genes of heads, and trunks treated to the two temperature conditions in all the samples to see if the clustering was different. We obtained a similar sample clustering as the global expression matrix with three main clusters corresponding to H27, T27 and T36 (Figure 4 and Supplementary Figure 2). Here again, we did not observe within each replicate group any clear sub-clustering which could correspond to the putative sex but found defined clusters among genes by expression levels. We obtained for the heads a higher gene diversity showing a large cluster of over-expressed genes. All genes identified in this clustering were related to vision (such as *rhodopsin*, or the *cry* gene family) (Figure 4). In the trunks the over-expressed genes’ cluster belonged to endocrine and digestive functions (such as the elastase-1 gene). The comparison between T27 vs T36 showed a cluster of over-expressed genes in 36°C trunks that appeared to be linked to hemoglobin production and oxygen binding.

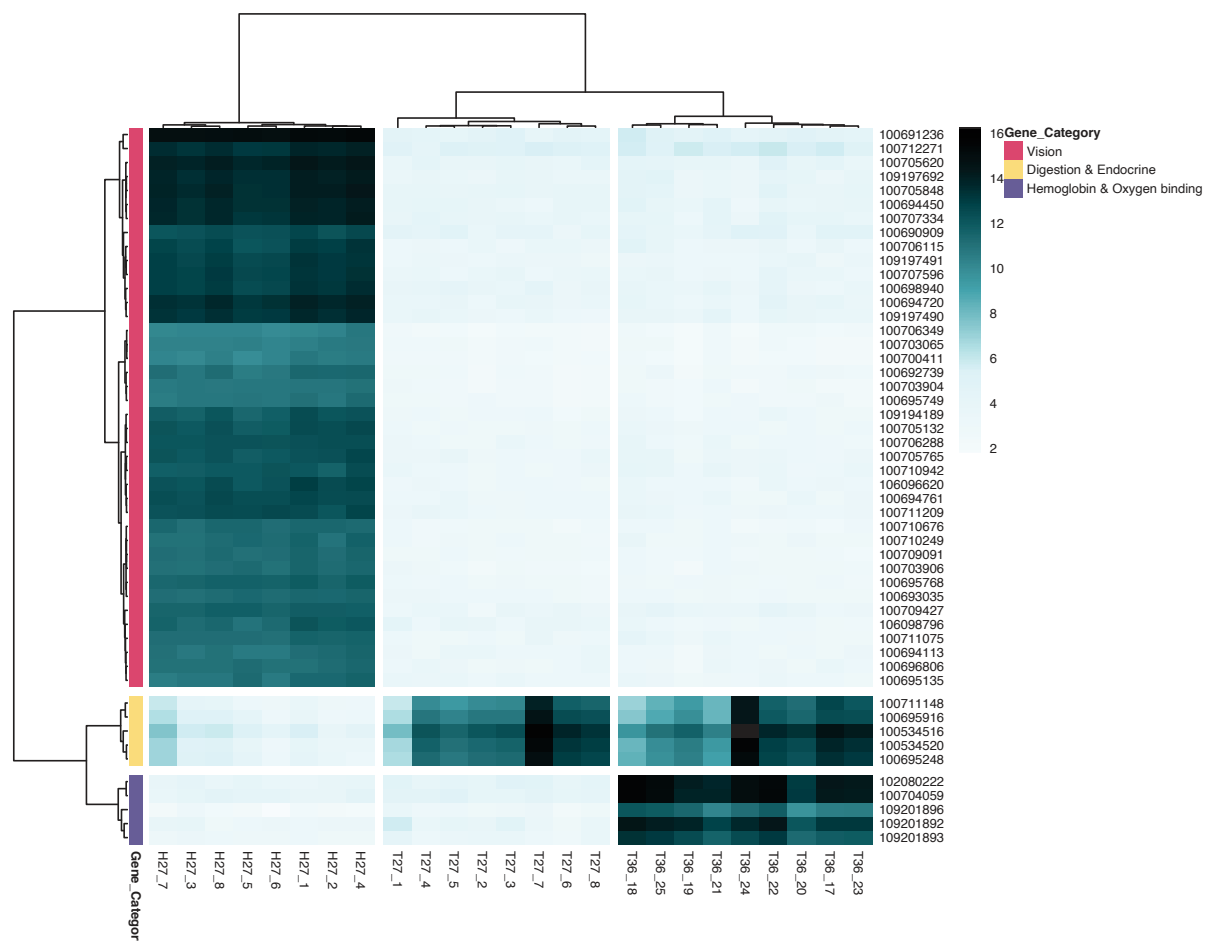


Figure 4. Heatmap of the top 50 most variable genes using a hierarchical clustering derived from Euclidean distance from normalized and log2 transformed counts and a Ward's distance as agglomeration method.

Manzala Each row represents a gene with its NCBI gene ID on the right. The expression of genes is color coded from low (light) to high (dark).

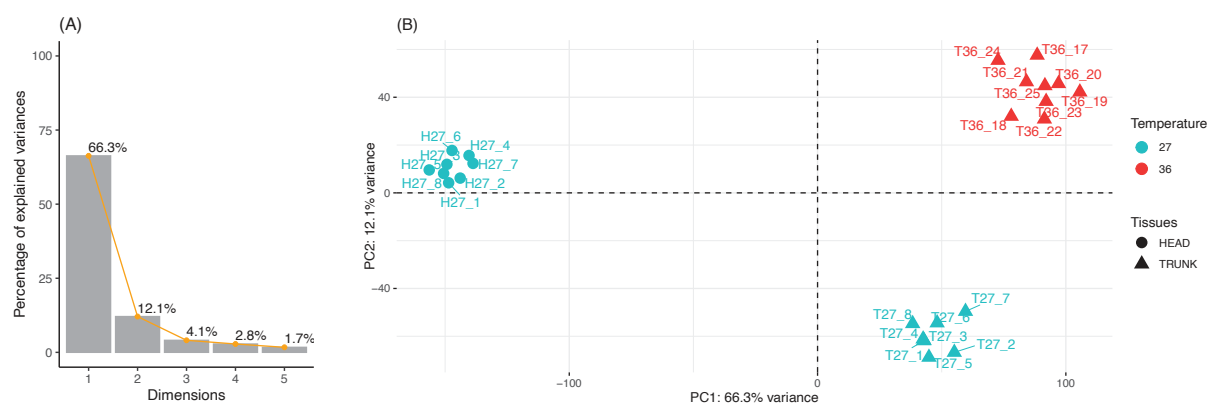
We performed complementary GO (Supplementary Figure 3) and KEGG Pathway enrichment (Table 1) analyses of the top 500 most variable genes amongst all samples which confirmed the strong representation of signaling processes connected to vision. As such, 62% genes of the phototransduction pathway are present. There was also, an overrepresentation of purine metabolism (7%), Arachidonic acid metabolism (7%), and starch and sucrose metabolism (13%) pathways seen among the top 500 variable genes. Nine out of the 48 genes (~18%) of the phototransduction pathway were also found among the top 50 most variable genes.

Table 1. KEGG Pathway output of the top 500 most variable genes among all conditions.

KEGG ID	Pathway Name	Gene Found	Gene in Pathway	Percentage	pvalue
04744	Phototransduction	30	48	0.62	5.344551e-44
00500	Starch and sucrose metabolism	6	45	0.13	1.207691e-05
00590	Arachidonic acid metabolism	6	87	0.07	8.135076e-04
00230	Purine metabolism	13	188	0.07	4.411590e-06
01100	Metabolic pathways	54	1978	0.03	1.596558e-07

3.3.3. PCA analysis confirms KEGG and heatmap conclusions

The PCA produced using all samples confirmed the results found in hierarchical clustering based on euclidean distances of gene expression and KEGG analysis (Figure 5). The first two dimensions of the PCA explained more than 78% of the variance among samples, with respectively 66.3% and 12.1%. PC1 clearly corresponds to a separation of the two organ types, whereas PC2 splits trunks according to temperature conditions. In addition, the top genes contributing to the PCA are confirming the GO and KEGG analysis as dimension 1 shows genes involved in vision (thus found only in heads), and in dimension 2 we find genes implicated in hemoglobin production, as shown by heatmaps.



(See legend on following page)

Figure 5. Principal component analysis of all samples and all transcripts.

A) Barplot of the percentage of explained variance for the first five dimensions of the PCA. **B)** Samples PCA plot. Dimension 1 is separating heads from trunks with a major contribution of the following genes to the PC1 axis: rhodopsin (LOC100691236), gamma-crystallin M2-like (LOC100705620), gamma-crystallin M3 family genes (LOC109197692, LOC100694720, LOC100694450, LOC109197490), crystallin, beta B1, like 1 (LOC100705848). Whereas dimension 2 is separating samples in 27°C control temperature from the 36°C treated trunk samples. The genes contributing strongly to the PC2 axis and thus to this clustering are the following: hemoglobin subunit beta-A (LOC100704059), hemoglobin subunit alpha-B (LOC102080222), hemoglobin subunit alpha-A (LOC109201892).

3.4. Gene expressions of sex differentiating candidates in the head

We analyzed three different sets of genes by PCA on head samples to search for potentially sex-dimorphic gene expressions in brains. The PCA using all genes showed no obvious clustering on the first dimension even though it represented 35.6% of the variance (Figure 6). The second set consisted in 2887 genes that we, selected based on their differential expression in the brain transcriptome of Nile tilapia at 21 dpf (Zhao et al., 2019), performed on all-females and all-males offspring of a Chinese GIFT strain. The PCA of these 2887 genes revealed a different clustering of the head samples (Figure 6) compared to the first PCA set but it did not separate the heads into two groups which could be attributed to a sexual phenotype.

The third PCA analysis was conducted on 141 candidate genes that have been shown to belong to the sex-determining or differentiating network. The samples showed no clear grouping (Figure 6), with samples positioned all along the PC1 axis (yet explaining more than 45% of the variance between individuals). Within the 141 genes, we found elevated expressions in heads for a glucocorticoid receptor gene *nr3c1*, *hsp70* (heatshock protein 70), adenylate cyclase type 3 (LOC100710007), genes related to epigenetic modifications such as *dnmt3a* (DNA methyltransferase 3 subfamily a) and the histone demethylase *kdm6bb* gene (lysine demethylase 6Bb) as well as the steroid biosynthesis enzyme *hsd11b2* gene (11-beta hydroxy steroid dehydrogenase) involved in the 11-ketotestosterone synthesis (Figure 6).

None of the sets of genes studied led to a dichotomous clustering in the heads that would correspond to females on the one hand and males on the other hand. Considering the disequilibrium between the number of genes and the number of individuals, the variance might be dragged by just one extreme sample. This seems to be the case as sample 4 in the “All genes” PCA is contributing more to the observed variance than the other samples. The variance explained on the first two axis of these PCAs are mostly due to sample 4 being very different from other individuals (Supplementary Figure 4).

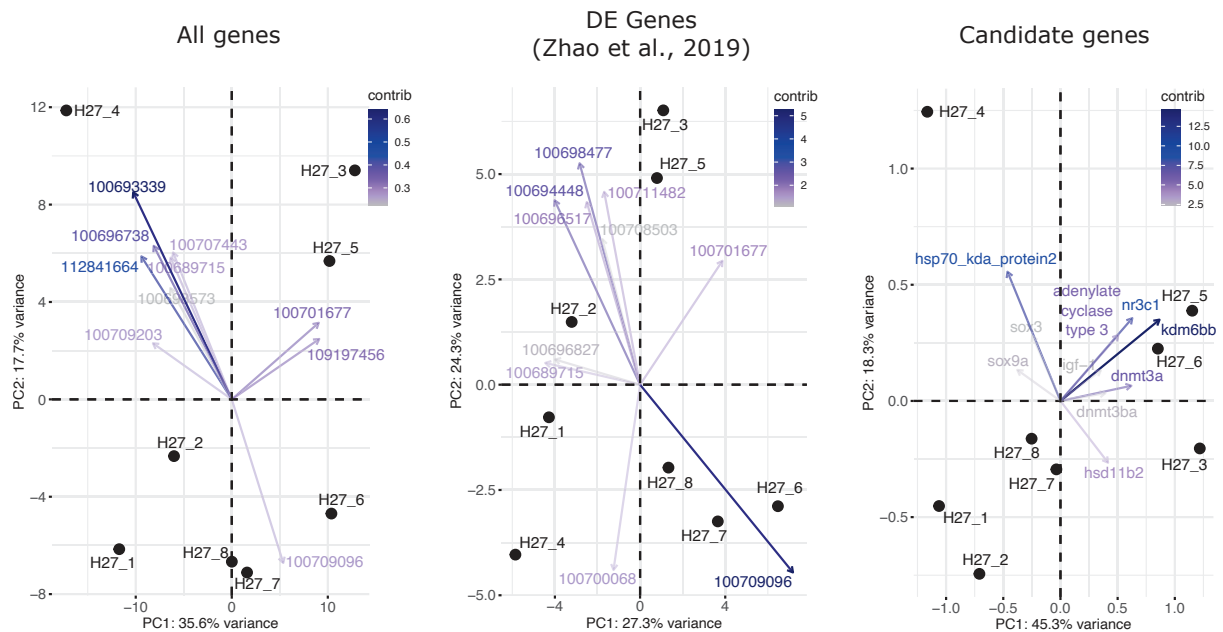


Figure 6. Comparison of three PCAs on head samples built with different gene sets.

Left panel: All transcripts; **Middle panel:** Genes identified by Zhao et al., (2019) as differentially expressed between males and females at control temperature; **Right panel:** Set of 141 sex-related candidate genes.

3.5. Effect of rearing temperature on gene expression in the developing gonad

We subsequently focused on the temperature effects of the trunks (dimension 2 of the previous PCA, Figure 5).

3.5.1. Global expression clustering

The global expression clustering of all trunks indicates a clear clustering of samples according to their temperature treatment (Figure 7). In addition, two sub clusters within the 27°C samples were detected with T27_1, T27_3, T27_4, T27_5 on one side, and T27_2, T27_6, T27_7, T27_8 on the other side. However, the length of the branch splitting of the two sub-clusters is really short and therefore they may not be relevant with regard to sex clusters.

3.5.2. Gene Ontology and KEGG analysis of Differentially Expressed genes

We identified 4337 genes to be differentially expressed (DE) between 27°C and 36°C conditions in trunks of which, 1319 genes were identified as down-regulated at 36°C and 3018 genes as up-regulated. Among the DE genes, “Metabolism genes” were over-represented according to the KEGG pathway analysis. Genes involved in lipid metabolism were particularly numerous. We found amongst them, more than 40% of these genes corresponding to the steroid biosynthesis pathways and 20% represented steroid hormones biosynthesis pathways genes (as named in the KEGG nomenclature) that were differentially

expressed in 36°C samples compared to 27°C control individuals (Figure 8 & Supplementary Table 3 and 4). We found an enrichment of genes implicated in the DNA replication and in the mismatch repair pathways, with more than half of these genes shown to be down-regulated at 36°C compared to the 27°C control temperature (Figure 9).

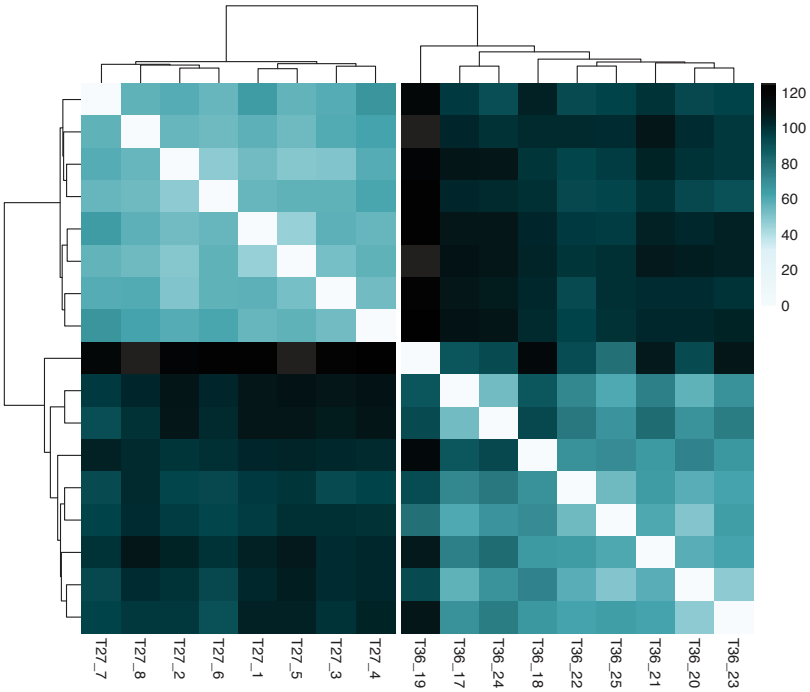


Figure 7. Global expression heatmap of trunk samples using a hierarchical clustering derived from Euclidean distance from normalized and log2 transformed counts and a Ward's distance as agglomeration method.

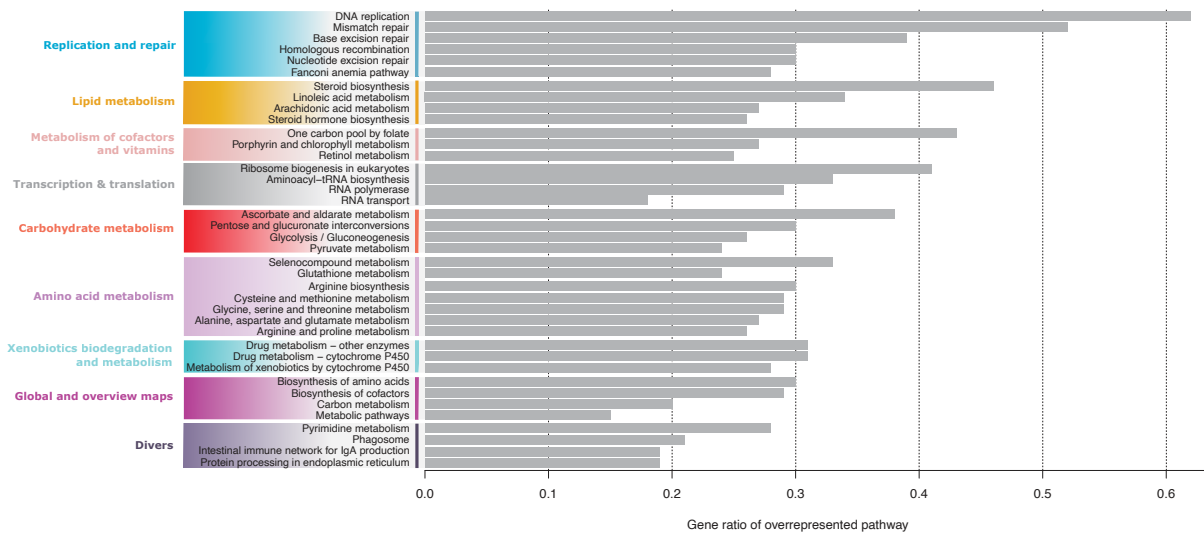


Figure 8. Barplot of the overrepresented pathways among the differentially expressed genes between the 27°C (control) and 36°C (treated) trunk samples.

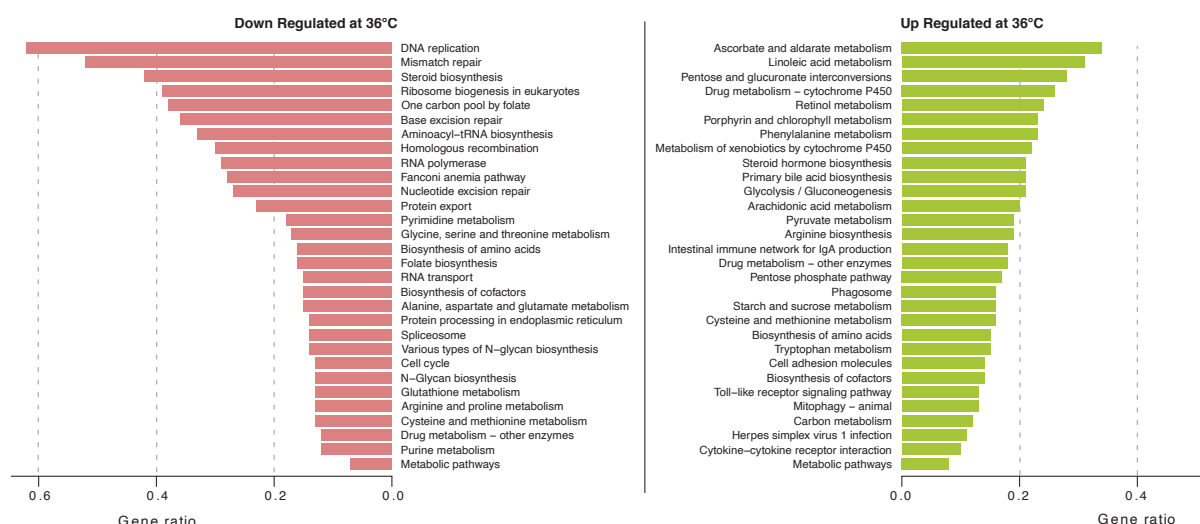


Figure 9. Barplot showing the gene ratios of the over-represented pathways among the differentially expressed genes between the 27°C (control) and 36°C (treated) trunk samples according to KEGG pathways analyses and categories.

3.5.3. PCA analysis

The PCA based on the expression of all the genes across all the trunk samples (27°C and 36°C) showed a strong clustering linked to the temperature treatments (Figure 10). There was a large individual variance observed in the 36°C samples, which was not the case for the 27°C samples on dimension 2. This is not surprising given that in the 36°C group we expect to find differentiating females, differentiating males and putative individuals for which the sex is being reversed by temperature. The details of genes contributing the most to the PC axis can be found in Supplementary Figure 5.



Figure 10. Principal component analysis of trunks with all genes.

- A) Barplot of the percentage of explained variance for the 5 first dimensions of the PCA.
 B) Samples PCA plot.

3.5.4. Volcano plot visualization of differentially expressed genes

Some of the genes identified as differentially expressed between trunks at 27°C and 36°C that have also been shown to be implicated in the sex-determining pathway in other teleost fish or more generally in vertebrates fish (Supplementary Table 1). Among these, we found the early male differentiating *dmrt1* transcription factor gene (found on LG12), several genes belonging to the *jmjd* family as well as the *hsp90aa1.1*, *kdm6bb*, *nr5a5*, and *dnmt1* genes (Figure 11).

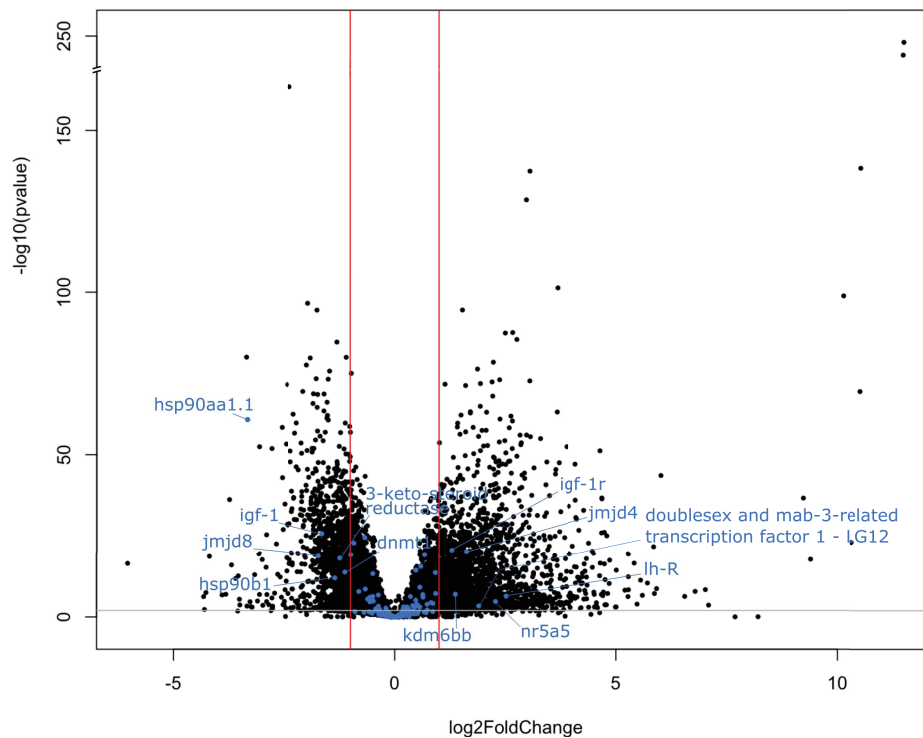


Figure 11. Volcano plot (log2 Fold Change vs. statistical significance) of transcriptome analysis of trunks.

Points outside the red lines are genes that are differentially expressed in 36°C samples with a Log2 Fold Change > 2 compared to 27°C trunks samples. Points over the grey line are significant points with a *p-value* < 0.01. Finally, blue points are genes that are differentially expressed at 36°C compared to the 27°C control and that are among the list of 141 sex-related candidate genes.

3.6. Candidate genes approach on the RNASeq data

We next focused our study on our list of 141 candidate genes involved in the sex-determining and sex-differentiating pathways in vertebrates.

3.6.1. PCA analysis of the 141 candidate genes shows 4 clusters

A PCA based on expression data of the 141 candidate genes (Figure 12 & Figure 13) showed that dimension 1 represents ~55% of the variance of the samples and draw a clear

separation between temperature treatments. Dimension 2, led mostly by *gsdf*, is strongly splitting T27 and T36 into sub-clusters. Because *gsdf* particularly (Kaneko et al., 2015), but also, *kdm6bb* and *gr* genes have been shown to be up-regulated in testis development in teleost fish or vertebrates (Fernandino et al., 2012; Ge et al., 2018) this could indicate that this division indeed separates differentiating males from differentiating females.

At control temperature (27°C), the clustering is clear and the variance in each sub cluster seems low. In general, we expect the variance of expression levels to be higher between 36°C samples as temperature alters several molecular mechanisms regulating the sex-differentiating pathways (Baroiller and D'Cotta, 2016). The higher variance observed in the two sub-clusters of the 36°C samples might be strengthened by the presence of XX female individuals, with some being sex-reversing individuals with levels of expression that are still not comparable to those of XY males.

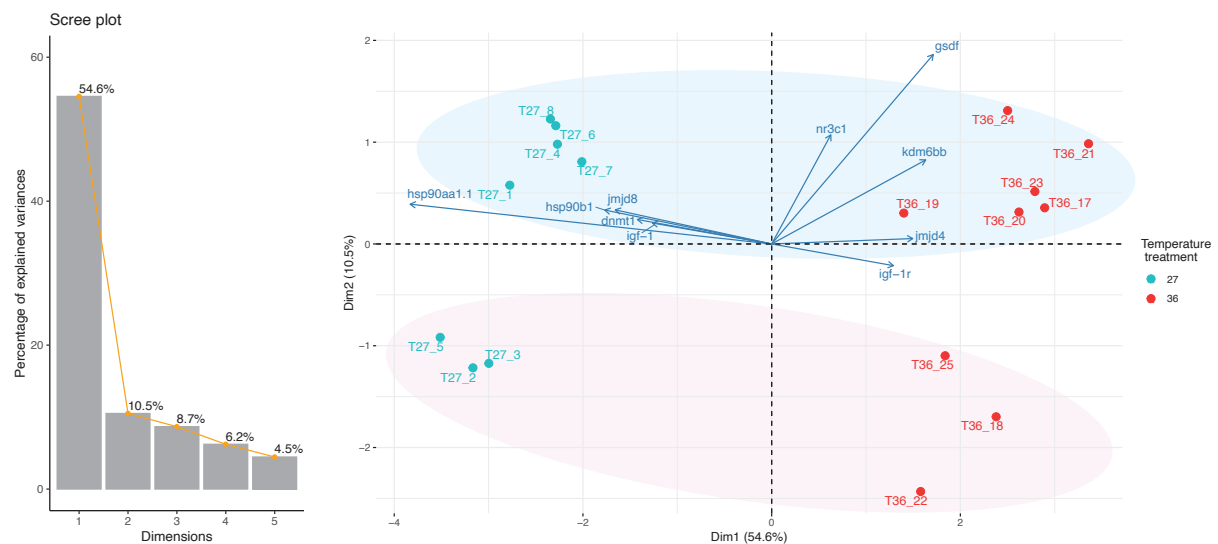


Figure 12. Principal component analysis of trunks samples performed using the 141 sex-differentiating candidate genes.

Left: Barplot of the percentage of explained variance for the first five dimensions of the PCA. **Right:** Biplot of samples with the top 10 genes contributing to the PCA. Because *gsdf*, *nr3c1* and *kdm6bb* are genes that are associated to masculinization in teleost fish, the upper part (in a blue circle) of the PCA might group the male samples, whereas the lower part (in a pink circle) might group the female samples.

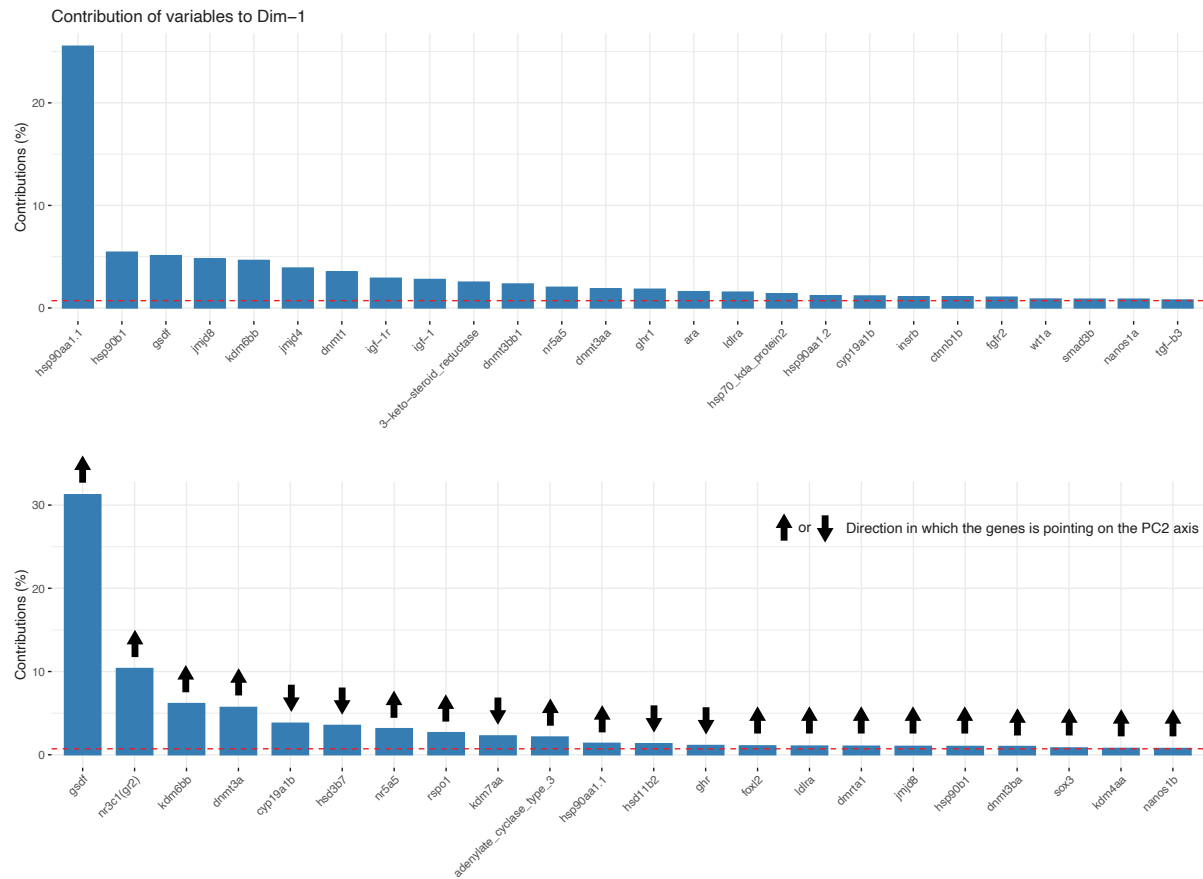


Figure 13. Barplot of genes contributing to dimension 1 (Top) and dimension 2 (Bottom) of the PCA run with 141 candidate genes for trunks samples.

Represented genes are only those contributing more than expected if all genes were contributing equally (red dashed line).

3.6.2. Variance analysis of “top Male genes”

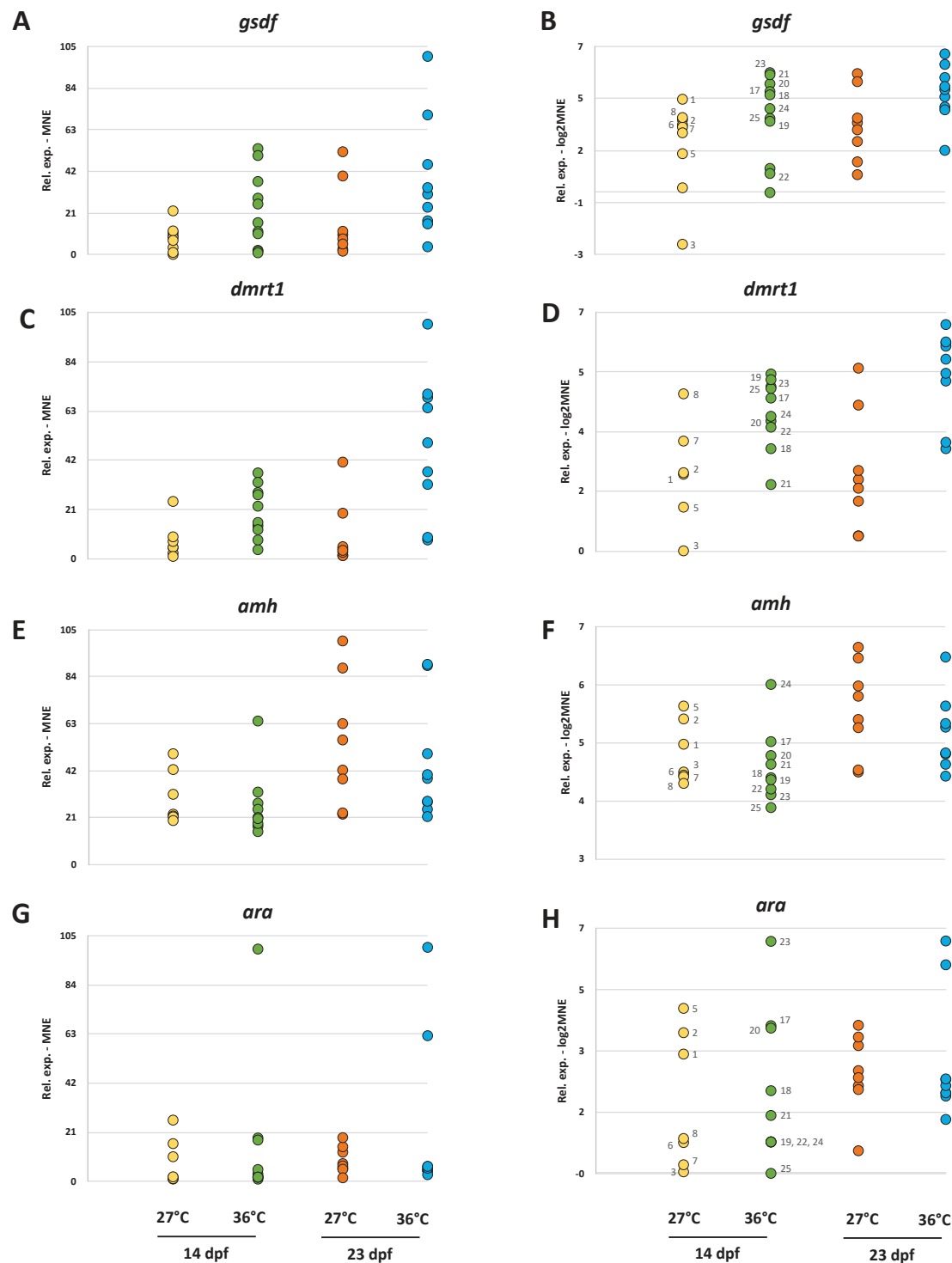
Since the PCA of candidate genes showed that *gsdf*, *nr3c1* and *kdm6bb* expression are strongly contributing to dimension 2 (potentially sex-clustering), we focused on these three genes and their expression among the samples. These genes are known to be involved in maleness in other species (Fernandino et al., 2012; Kaneko et al., 2015; Zhao et al., 2019). Thus, we checked in details the expression in each sample considering the potential sex-clustering given by the PCA (Table 2). The potential females at 27°C systematically have lower expressions than the mean for the three genes. *Kdm6bb* is over-expressed in 36°C putative males compared to all 27°C potential males and females. The *nr3c1* gene probably has less impact in males of 27°C as two males show expressions that are lower than the mean expression. The only “potential female” showing a *gsdf* expression higher than other females is T36_18. This could be due to the temperature treatment but as the two other “male” genes are not highly expressed this sample could still develop as a female. Overall, *gsdf* expression is higher in putative males than in putative females.

Table 2. Log2 transformed and normalized counts of the top 3 “male genes” of the PCA of 27°C and 36°C trunks samples.

Candidate genes for maleness	Potential Males 27°C					Potential Males 36°C						Potential Females 27°C			Potential Females 36°C			Mean of All samples	Mean of Potential Males	Mean of Potential Females
	T27	T27	T27	T27	T27	T36	T36	T36	T36	T36	T36	T27	T27	T27	T36	T36	T36			
	_1	_4	_6	_7	_8	_17	_19	_20	_21	_23	_24	_2	_3	_5	_18	_22	_25			
<i>nr3c1</i>	6.96	7.58	7.46	7.29	7.18	7.35	7.83	7.44	8.29	7.55	7.79	6.31	7.24	6.58	6.35	6.75	7.25	7.25	7.52	6.75
<i>kdm6bb</i>	11.29	11.69	11.94	11.57	11.65	12.51	12.44	12.52	13.37	12.68	13.19	11.16	11.59	10.97	11.62	11.80	12.32	12.02	12.26	11.58
<i>gsdf</i>	8.55	8.52	8.26	8.39	8.36	8.83	8.37	9.03	9.74	9.25	8.85	6.91	5.95	6.74	8.90	7.30	7.06	8.18	8.74	7.14

3.7. Real-time PCR of key sex-differentiating genes

In order to better define the putative phenotypic sex of the Hora samples used, we also analyzed the trunks by quantitative real-time PCR (qPCR), of well-known testicular differentiating genes (*gsdf*, *dmrt1*, *amh*, *ara*) and ovarian differentiating genes (*cyp19a1a*, *foxl2* and *cyp19a1b*). This analysis was performed at 14 dpf on the same samples used for the RNASeq study when these were still available with additional samples, as well as in 23 dpf individuals. Expression of *gsdf*, *dmrt1* and even *ara* showed higher gene expressions in T-36°C samples compared to T-27°C already at 14 dpf, with a similar pattern at 23 dpf but with higher individual variance (Figure 14). Surprisingly, the *amh* did not follow this profile neither at 14 nor at 23 dpf. Expression of the ovarian aromatase *cyp19a1a* gene was low at 14 dpf with some individuals showing higher expressions at 36°C (Figure 15). The elevated levels found for some T-27°C individuals at 23 dpf suggests a higher sexual dimorphism in the expression at this stage. In contrast, expression is down-regulated in the 36°C trunks at 23 dpf. *Foxl2* showed a somewhat similar pattern as the brain aromatase *cyp19a1b* gene that was in fact expressed at much stronger levels than the gonad *cyp19a1a* gene, although we found no inverse correlation with the expression of male genes (Supplementary Figure 6).



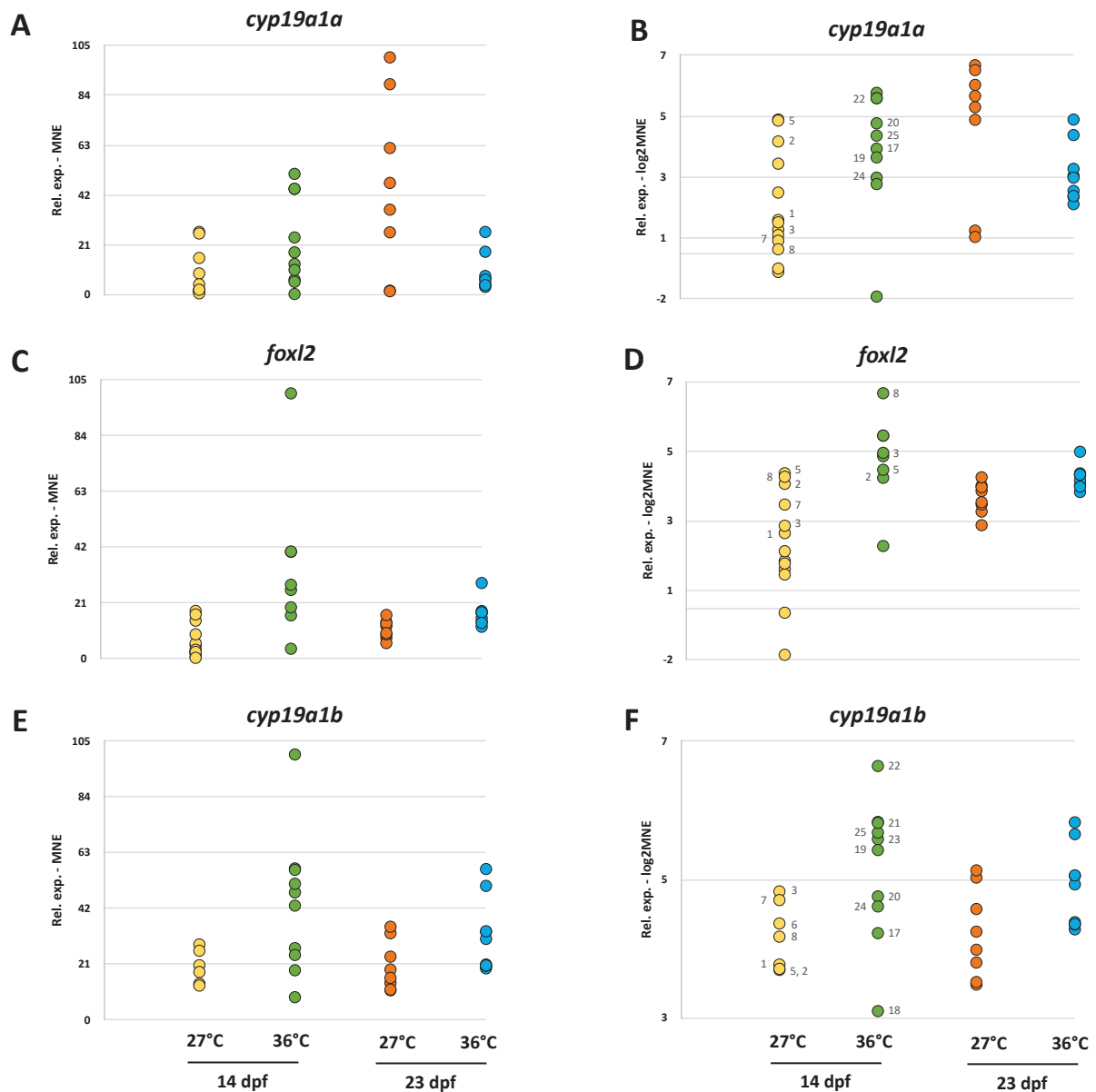


Figure 15. Relative expression of the Ovarian differentiating genes analyzed by qPCR in individual trunks shown as the maximum normalized expression (MNE) as well as the log2 of MNE of the *cyp19a1a* gene (A & B), *foxl2* gene (C & D) and the brain *cyp19a1b* gene (E & F). Individuals that were also analyzed by RNaseq are numbered.

To predict putative males and females, we plotted the expression levels of the gonad aromatase *cyp19a1a* relative to those of *gsdf*, *dmrt1* and *amh*. This separated relatively well two groups for the 27°C samples and the clustering observed was mostly consistent with the results of the RNaseq candidate gene approach (Figure 16). The sample T27-3 however showed low expressions levels for several of the genes studied by qPCR, including *cyp19a1a* (Figure 14 & 15). Thus, even if the RNaseq suggested this sample was a putative female, this sample does not cluster as expected (Figure 16).

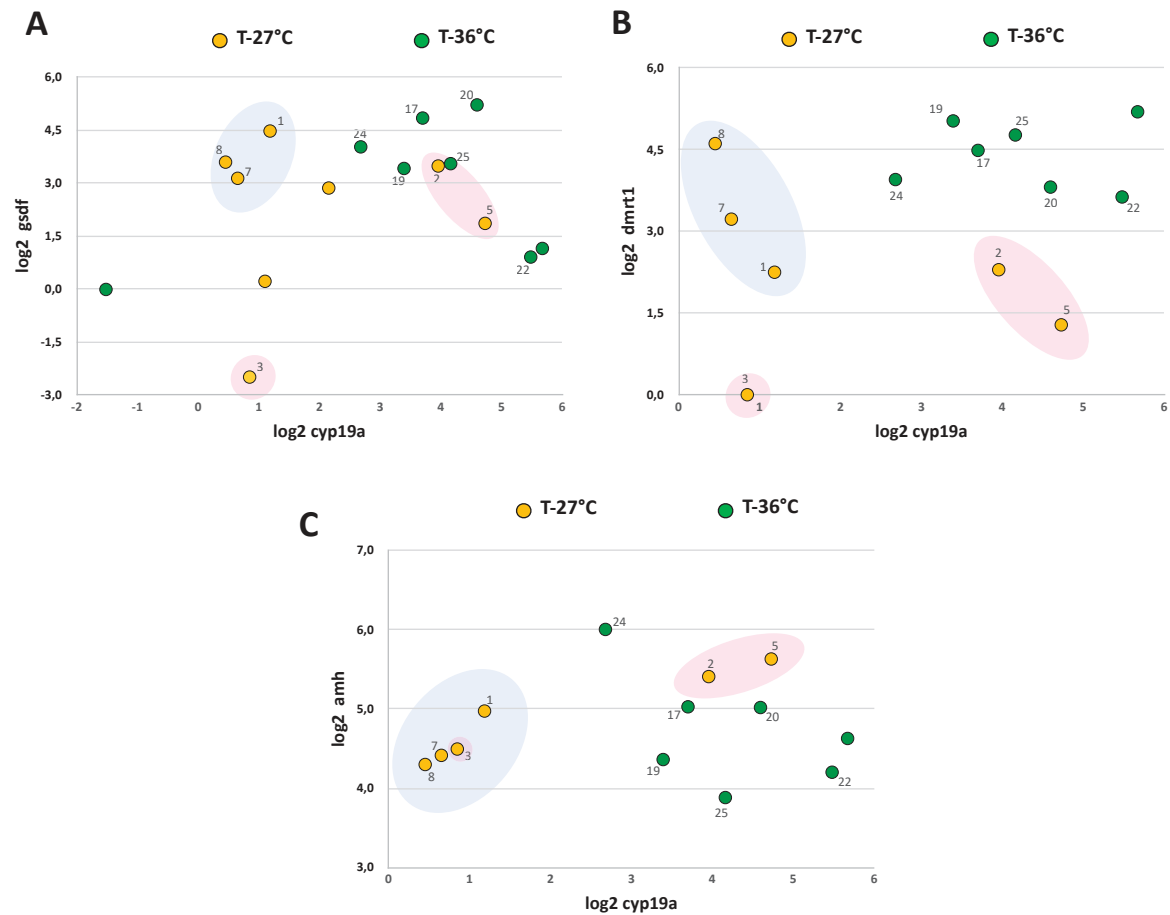


Figure 16. Relative expression of individual trunks of *cyp19a1a* plotted relative to *gsdf* (A), *dmrt1* (B) and *amh* (C) expressed as the log2 of maximum normalized expression (MNE).

Individuals that were also analyzed by RNASeq are numbered and putative males and females identified by the candidate gene approach the 27°C RNASeq data are shaded respectively in blue and pink.

3.8. Validation of differentially expressed genes by real-time PCR

To validate further some of our transcriptome results, we selected five genes shown to be differentially expressed between 27°C and 36°C trunks analyzed by qPCR. Similarly to the transcriptome analyses we detected higher expression for *kdm6bb* in the 36°C trunks compared to the 27°C at 14 dpf and this profile persisted at 23 dpf (Figure 17A). We observed a similar pattern of expression for *jarid2* (jumoni and AT-rich interaction domain containing 2) gene (Figure 17B). Mean expressions for the *hsd11β2* and *gr2* were likewise higher in the T-36°C samples compared to T-27°C and this trend was more marked at 23 dpf (Figure 17D & 17E). We also analyzed the steroidogenic 11β-hydroxylase *cyp11c1* gene involved in the synthesis of both 11-ketotestosterone and cortisol (Baroiller et al., 1999; Zheng et al., 2020). At 14 dpf *cyp11c1* expression levels remained low for both 27°C and 36°C trunks (Figure 17C) increasing strongly at 23 dpf in the trunks treated at 36°C.

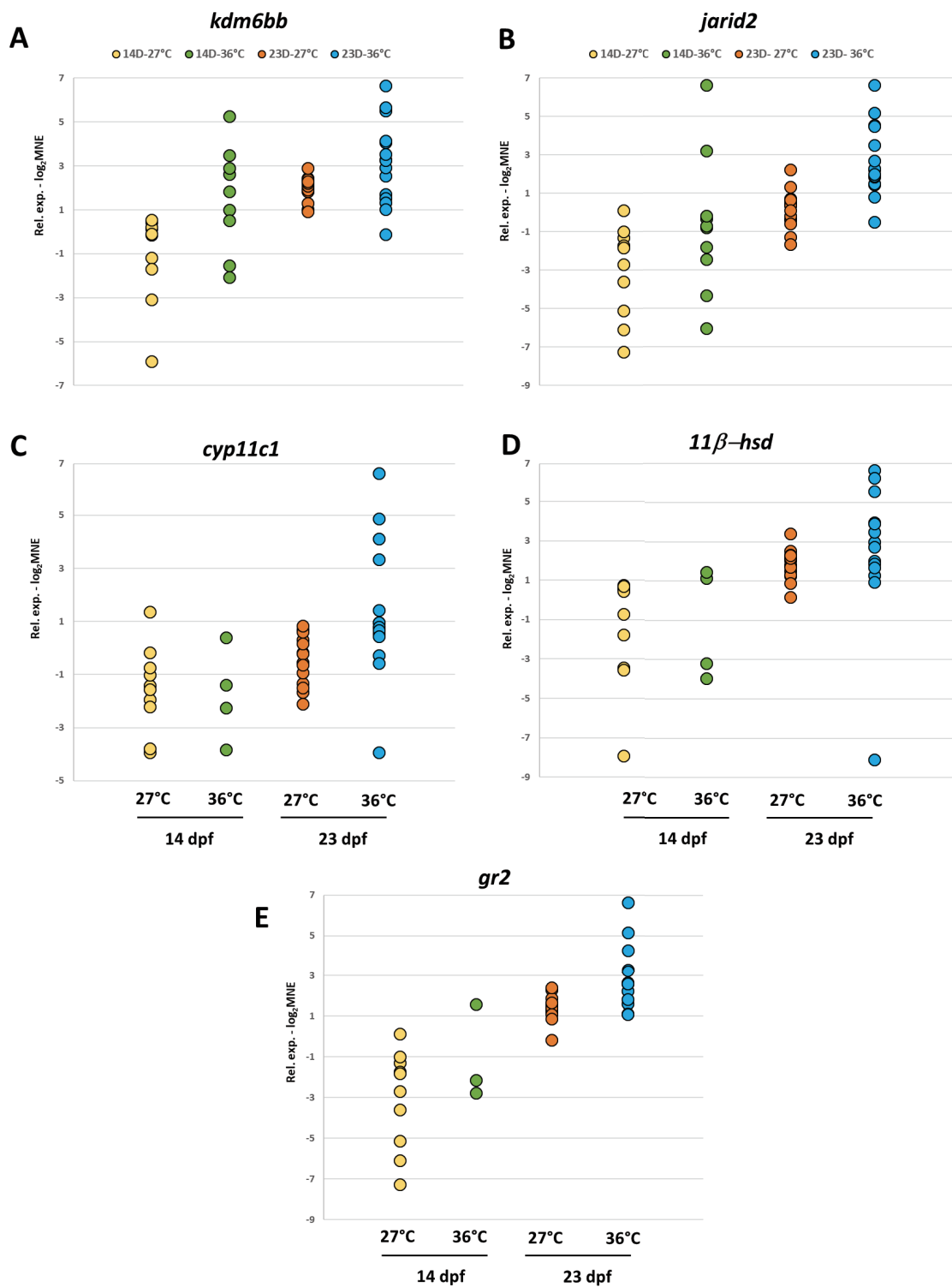


Figure 17. Relative expression of DE genes found in our RNASeq data analyzed by qPCR in individual trunks shown as the log₂ of the maximum normalized expression (MNE) of the *kdm6bb* gene (A), *jarid2* (B), *cyp11c1* (C), *11β-hsd* (D) and *gr2* (E).

4. Discussion

The Hora population of Nile tilapia belongs to the subspecies *O. niloticus cancellatus* like the Lake Koka population, both found in reservoirs of the Awash River in Ethiopian highlands. Although the sex-determining gene in the Koka population is centered around the *amh* gene on LG23 (Triay et al., unpublished – Chapter 2), this is not the case for Hora fish. No other linkage group was found segregating for sex. Despite this, we thought that genetic factors - perhaps through allelic variations - could be determining the sex in the Hora populations since we obtained in this family (but also in other families) a balanced phenotypic sex-ratio, when experimenting at control 27°C conditions. Although the trigger of sex in Hora is different from other Nile tilapia stocks, we wanted to know whether the downstream process of sex-differentiation was conserved, to understand better the early regulating network that is acting in the putative genetic and temperature-induced pathways of Hora fish.

Here we initiated this work by doing a transcriptome analysis at 14 dpf together with qPCR analyses at 14 and 23 dpf. We were hampered by the fact that we could not genetically sex the fry. Sex-differentiation of Nile tilapia has been studied mostly (if not only) in domesticated strains where the phenotypic sex is known before which greatly simplifies the search. Crossings of YY males that are fertile and viable, with genetic XX females gives all XY individuals (theoretically all males) while crossings of XX males with genetic XX females results in all XX individuals (presumably all females) (Baroiller and D'Cotta, 2018). Males can therefore, be easily detected when all-XX females are treated to masculinizing temperatures (Baroiller et al., 2009b). The critical period when undifferentiated gonads begin to differentiate into ovaries or testis in Nile tilapia is around 9 to 10 dpf, and this can be perceptible already at the molecular level (D'Cotta et al., 2001; Ijiri et al., 2008; Poonlaphdecha et al., 2011; Tao et al., 2013; Kaneko et al., 2015). At this stage gonads are already responsive to temperature (and hormone) treatments which can induce sex changes (Baroiller and D'Cotta, 2016). The gonads in domestic strains start to be morphologically distinguishable between testis and ovaries at ~12 dpf. The number of germ cells stop increasing in XY male gonads during 9 to 14 dah (13 to 18 dpf), in contrast to the ongoing increase in XX female gonads (Kobayashi et al., 2008). Primordial germ cells increase then drastically in ovaries at ~17 dpf, clearly visible at 23 dpf (Nakamura et al., 1998; D'Cotta et al., 2001; Ijiri et al., 2008; Kobayashi and Nagahama, 2009).

Our study showed that in Hora the molecular changes are subtle at 14 dpf. Sexual-dimorphism was not readily discernible by our global transcriptome analysis in neither heads nor trunks. Nevertheless, our enhanced analyses in trunks of 141 sex-related candidate genes showed that some could be differentially expressed, notably the *gsdf* gene, allowing us to define what could be putative males and females.

4.1. *gsdf* is differentially expressed in Hora trunks at 14dpf

We have found that the *gsdf* (gonadal soma-derived factor) gene is differentially expressed in Hora fish as early as 14 dpf in the 27°C trunks and even more in the 36°C trunks, and this profile became more marked at 23 dpf with high levels in presumably differentiating testis. The strong differences in *gsdf* expression levels at 14 dpf identified by RNASeq, two clusters among the 27°C and 36°C samples in trunks that could fit a sex-clustering pattern with males presenting a high *gsdf* expression in the upper cluster. *Gsdf* is a fish-specific gene of the TGF- β superfamily which also contains the *amh* gene (De Caestecker, 2004). It is the first target of the sex-determining *dmy/dmrt1Yb* gene in medaka (Zhang et al., 2016). *Gsdf* plays an important role in testis differentiation in many other teleosts such as in tilapia (Kaneko et al., 2015), the zebrafish (Yan et al., 2017), the European sea bass (Crespo et al., 2013), the wrasse (Horiguchi et al., 2013) and the Japanese flounder (Yang et al., 2019). The functional role of *gsdf* appears to be related to germ cells and spermatogonia proliferation in several of these species including the Nile tilapia (Sawatari et al., 2007; Horiguchi et al., 2013; Jiang et al., 2016; Yang et al., 2019). Higher expression of the *gsdf* gene might inhibit primordial germ cell proliferation in the Japanese flounder (Yang et al., 2020).

Kaneko et al., (2015) showed that *gsdf* expression and its protein are sexually dimorphic as early as 9 dpf even before the morphological differentiation in gonads of the Nile tilapia. They also highlighted that an over-expression of *gsdf* was enough to induce testis development in XX individuals. Knockout of *gsdf* in Nile tilapia caused a complete male-to-female sex reversal, establishing that *gsdf* expression in differentiating testis is downstream from *dmrt1* (Jiang et al., 2016). Estrogen synthesis by the aromatase enzyme Cyp19a1 is essential for ovarian differentiation (D'Cotta et al., 2001; Wang et al., 2007; Guiguen et al., 2010). In tilapia, *gsdf* might be linked to the repression of estrogen production, thus triggering testicular differentiation (Jiang et al., 2016). In Hora *gsdf* expression is overall higher than *dmrt1* at 14 dpf which resembles more the pathway proposed by Kaneko et al., (2015) that suggested *gsdf* preceded *dmrt1* expression in tilapia. Our results are suggesting that *gsdf* might also be expressed in a sexual dimorphic manner in Hora fish. *Gsdf* could be a good sex marker since Hora lack the *amhy/amΔY* genes (Sissao et al., 2019; Triay et al., 2020; Triay et al., unpublished – Chapter 2).

4.2. Temperature treatment affects gonad gene expression in Hora

In the Japanese flounder, the expression of *gsdf* is increased by high temperature treatments and results in XX female to male sex-reversal (Yang et al., 2019, 2020). Our results also confirm those observed in Yang et al., (2019, 2020) since we found that Hora fish treated to 36°C showed even higher *gsdf* expression levels than the 27°C trunks and this was more evident at 23 dpf suggesting a higher proportion of testis differentiating. The individual variation observed at 23 dpf might stem from the presence of genetic and temperature induced males as well as females.

4.2.1. Steroid and Steroid hormones biosynthesis

When comparing 27°C trunk samples to 36°C treated samples, we noticed pronounced differences in gene expression. Among these we observed a down regulation by the high temperature treatment of genes involved in steroid biosynthesis while genes of steroid hormone biosynthesis were up regulated. Steroid hormones are implicated in embryonic development, sex differentiation, reproduction as well as in stress response in teleost fish (Baroiller et al., 1999; Goikoetxea et al., 2017). Estrogens are synthesized from androgens by the aromatase enzyme encoded by the *cyp19a1a* and *cyp19a1b* genes (Baroiller et al., 1999; Guiguen et al., 2010). Differentiating ovaries of Nile tilapia were found to have higher levels of estradiol-17 β and *cyp19a1a* levels compared to differentiating testis (Baroiller and D'Cotta, 2001; D'Cotta et al., 2001). Hora trunks showed higher levels of the brain aromatase *cyp19a1b* expression at 14 dpf compared to the gonad aromatase *cyp19a1a*, particularly in the 36°C trunks. Elevated levels of *cyp19a1b* have also been detected in early testis differentiation of Nile tilapia but its role is unknown (Sudhakumari et al., 2005). In Hora we plotted the expression of *cyp19a1b* with respect to those of *gsdf* at 14 dpf however they gave no evidence of a dimorphism related to sex. In contrast, expressions of *cyp19a1a* together with expressions of *gsdf* or *dmrt1* or even *amh* did separate 27°C individuals into two, presumably putative males from putative females. Double plotting of expressed genes allowed the prediction of males and females of wild Southern flounder that have ESD (Honeycutt et al., 2019).

In the Hora wild population of Nile tilapia, temperature does produce differential expression of genes involved in the steroid hormone biosynthesis. We observed that temperature had a repressive effect on the *cyp19a1a* expression at 23 dpf in the 36°C Hora trunks which was not the case in the 27°C trunks. Similar repression of *cyp19a1a* expression (affecting estrogen levels), was shown in tilapia XX gonads treated to 36°C which were masculinized by temperature (D'Cotta et al., 2001; Poonlaphdecha et al., 2013; Sun et al., 2018), and one of the mechanisms of repression is through the increase in methylation levels on the *cyp19a1a* promoter (Wang et al., 2017).

Among the different categories of steroid hormones are the glucocorticoids of which cortisol is the most important. This hormone might be an important link between the environmental induction and the sex differentiation pathway, as cortisol production is increased by high temperatures in the Patagonian pejerrey, and the Japanese flounder leading to high male ratios (Hattori et al., 2009; Yamaguchi et al., 2010; Fernandino et al., 2013). In these fish, the increase of cortisol levels induces repression of the aromatase *cyp19a1a* expression and thus blocking estrogen synthesis and the ovarian differentiating pathway. This process is favoring the activation of the androgen pathway by regulating 11 β -hydroxylase (*Cyp11b*) and 11 β -hydroxysteroid dehydrogenase (*Hsd11 β*) enzymes, resulting in a female to male sex-reversal (reviewed in Goikoetxea et al., 2017). 11 β -hydroxylase, encoded by the *cyp11c1* gene, is involved in the synthesis of 11-Ketotestosterone the main fish androgen but also in cortisol synthesis (Zheng et al., 2020). Cortisol action, seen in the Japanese flounder, appeared to be through its glucocorticoid receptor, which regulated the

cyp19a1a gene expression by binding to the GRE site in the promoter (Yamaguchi et al., 2010). RNAseq in Hora showed elevated expression of the *gr2* gene in 27°C trunks (in presumed males) which was validated by qPCR, but levels increased thereafter in both 27°C and 36°C trunks. Expression levels of *cyp11c1* and *11 β -hsd* were similar between 27°C and 36°C trunks at 14 dpf but increased considerably in many individuals at 36°C, which suggests that temperature was inducing maleness, but whether this is due to 11-ketotestosterone or/and via cortisol remains to be verified.

4.2.2. DNA replication and mismatch repair is down-regulated at high temperature

Interestingly, more than half of the genes implicated in the DNA replication pathway and in the mismatch repair pathway were down-regulated at 36°C compared to the control temperature. This is surprising as high temperatures are generally stimulating more errors and thus more corrections over the mitosis period. If we do not have a strong clue to explain this phenomenon, it is interesting to note that it might result in higher mutation rates but lower rates of mitosis.

4.3. Some of the usual suspects are DE at 36°C in gonads

Among the DE genes at 36°C, we identified a dozen of genes that are implicated in the sex determination/sex-differentiation pathway. These identified genes might be good starting points to look for modifications of sex pathways between 36°C and control temperatures. Indeed, some of these genes might be tightly linked with temperature-induced sex-reversal in the Nile tilapia.

4.3.1. Heatshock proteins hsp90

Heatshock protein complex are linked to gamete quality in some fish. As such, lack of hsp protein in gonads leads to poorer quality of eggs for females or apoptosis of testicular cells in males (He et al., 2010; Chapman et al., 2014; Kohn et al., 2015). An explanation is the role of hsp proteins in cell migration during embryonic stages. In the zebrafish, Pfeiffer, (2018) demonstrated that a misregulation of Hsp90 proteins induces a slower migration rate of primordial germ cells to the gonadal region. In the more distant species of *Drosophila melanogaster*, a lack of regulation of Hsp90 proteins can lead to defects in cell proliferation and is thus essential for a proper progression through the cell cycle at early stages (Bandura et al., 2013). In addition, we found that the expression of *hsp90aa1.1* is higher in 27°C than in 36°C gonads of *O. niloticus* Hora population. This result is in agreement with those of Mahanty et al., (2019) that showed an increased temperature induced decrease of *hsp90* expression in the Barb (*Puntius sophore*). The rice field eel (*Monopterus albus*) also shows a down-regulation of *hsps* expression in testes of individuals exposed to heat-stress (He et al., 2010). Finally, in natural environments, some fish also present a down-regulation of

heatshock proteins during warmer seasons of the year (Domingos et al., 2013).

4.3.2. Epigenetic regulators

Genes of the Jumonji family are also showing a differential expression with high temperature in Hora individuals. In the Hora population of Nile tilapia the sexual dimorphism of expression of the histone demethylase *Kdm6bb* (lysine-specific demethylase 6Bb) gene in the control temperature is not obvious. According to the clusters designed primarily by *gsdf* expression, we expect to have only 3 females amongst the 27°C trunks so this lack of sexual dimorphic expression might only be due to the low number of samples used for the comparison. A strong difference, however, is seen between the 36°C putative males and all the other samples, including control temperature males. In the TDS Red-eared slider turtle, *Kdm6b* promotes transcription of the male sex-determining *dmrt1* gene (Ge et al., 2018). The expression of *Kdm6b* is sexually dimorphic even before gonads are distinct. *Kdm6b* can be regulated by the STAT3 gene that encodes for the signal transducer and activator of transcription 3. When temperature increases, STAT3 products get phosphorylated and bind to the *Kdm6b* gene, repressing its expression and blocking the male pathway (Weber et al., 2020). In the Chinese GIFT strain of the Nile tilapia however, high temperature treatments on females leads to increased expression of *kdm6bb* compared to control temperature females (Zhao et al., 2019). In addition, *kdm6bb* expression is in general higher in males than in control temperature females. Hora showed higher *kdm6bb* expression in 36°C trunks at 14 dpf and even more at 23 dpf. Thus, higher expression of *kdm6bb* might up-regulate *dmrt1* expression and be involved in XX sex-reversal individuals in thermosensitive families of Hora population.

Jarid 2 was shown to be expressed in Hora at 14 dpf and up-regulated at 23 dpf with higher levels of expression in both stages in the 36°C trunks. *Jarid2* has also been shown to be elevated in males and up-regulated with temperature in Nile tilapia brains (Zhao et al., 2019). *Jarid2* and *Jmjd3*, genes of this family, have been identified in reptiles with temperature-induced sex reversal (Deveson et al., 2017). These genes are tightly linked to epigenetics and perturbation of their function can occur through intron retention (which caused a premature stop codon) that triggered sex reversal under high temperature conditions.

Based on the downregulation of the aromatase *cyp19a1a* at 23dpf in the 36°C trunks together with the increase of *gsdf*, *dmrt1*, *cyp11c1*, *11b-hsd* as well as *kdm6bb* and *jarid2*, we consider that temperature was probably inducing masculinization of the Hora individuals.

4.4. Head RNA analyses

Our transcriptome analysis of heads at 14 dpf showed an enrichment of genes involved in vision, which is not surprising since the heads contained the eyes. The brain harbors many sexually dimorphic genes. In a previous study of brain transcripts of *O. niloticus* (Chinese GIFT strain, with *amh* SD) during the sex-differentiation period, a large number of genes

were found to be differentially expressed between males and females at 21 dpf (Zhao et al., 2019). The analyses of 2887 genes found by Zhao et al., (2019) on Hora did not highlight the same genes, which might suggest different regulating pathways in the brain sexualisation of Hora fish. The first clear difference is that in Hora brains at 14 dpf *amh* is not sexually dimorphic. For strains of *O. niloticus* in which the *amh* gene is the major SD, some specific genes are dimorphically expressed at 14 dpf. As such, the *amh* gene itself is found highly expressed in brains of XY samples whereas its expression is practically undetectable in XX individuals (Poonlaphdecha et al., 2011; Baroiller and D'Cotta, 2018). This gene, though present in all three datasets used for the PCA (all genes, DE genes according to Zhao et al., (2019) and 141 candidate genes), is not among the genes influencing the brain clustering.

In conclusion, this study of sex-differentiation in Hora Nile tilapia showed that *gsdf* is already expressed at 14 dpf with higher expression in putative differentiating testis, while neither *dmrt1* nor *amh* are. This indicates that sex-differentiation in Hora tilapia follows a different pathway from that found commonly in domesticated Nile tilapia strains. Differences were also observed in Hora heads, since *amh* expression was not dimorphic at 14 dpf. We found in the trunks several genes that are epigenetic regulators such as the histone modifiers *kdm6bb* and *jarid2* genes, and genes related to cortisol and 11-ketotestosterone synthesis in putative testis that were up-regulated in trunks treated to 36°C. These results together with the increase of the testis-differentiating genes *gsdf*, *dmrt1* and the repression of the ovarian *cyp19a1a* suggests that high-temperature induced masculinization in this Hora family. The Hora population is hence probably thermosensitive despite living in a cold lake.

References

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Bandura, J.L., Jiang, H., Nickerson, D.W., and Edgar, B.A. (2013). The Molecular Chaperone Hsp90 Is Required for Cell Cycle Exit in *Drosophila melanogaster*. *PLoS Genet* 9, e1003835.
- Baroiller, J.-F., and D’Cotta, H. (2001). Environment and sex determination in farmed fish. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 130, 399–409.
- Baroiller, J.-F., and D’Cotta, H. (2016). The Reversible Sex of Gonochoristic Fish: Insights and Consequences. *Sexual Development* 10, 242–266.
- Baroiller, J.-F., and D’Cotta, H. (2018). Sex Control in Tilapias. In *Sex Control in Aquaculture*, (John Wiley & Sons, Ltd), pp. 189–234.
- Baroiller, J.F., Chourrout, D., Fostier, A., and Jalabert, B. (1995). Temperature and sex chromosomes govern sex ratios of the mouthbrooding Cichlid fish *Oreochromis niloticus*. *J. Exp. Zool.* 273, 216–223.
- Baroiller, J.-F., Guiguen, Y., and Fostier, A. (1999). Endocrine and environmental aspects of sex differentiation in fish. *CMLS, Cell. Mol. Life Sci.* 55, 910–931.
- Baroiller, J.F., D’Cotta, H., and Saillant, E. (2009a). Environmental effects on fish sex determination and differentiation. *SXD* 3, 118–135.
- Baroiller, J.F., D’Cotta, H., Bezault, E., Wessels, S., and Hoerstgen-Schwark, G. (2009b). Tilapia sex determination: where temperature and genetics meet. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 153, 30–38.
- Bej, D.K., Miyoshi, K., Hattori, R.S., Strüssmann, C.A., and Yamamoto, Y. (2017). A Duplicated, Truncated *amh* Gene Is Involved in Male Sex Determination in an Old World Silver-side. *G3: Genes, Genomes, Genetics* 7.
- Berishvili, G., D’Cotta, H., Baroiller, J.-F., Segner, H., and Reinecke, M. (2006). Differential expression of *IGF-I* mRNA and peptide in the male and female gonad during early development of a bony fish, the tilapia *Oreochromis niloticus*. *General and Comparative Endocrinology* 146, 204–210.
- Bertho, S., Herpin, A., Branthonne, A., Jouanno, E., Yano, A., Nicol, B., Muller, T., Pannetier, M., Pailhoux, E., Miwa, M., et al. (2018). The unusual rainbow trout sex determination gene hijacked the canonical vertebrate gonadal differentiation pathway. *Proc Natl Acad Sci USA* 115, 12781–12786.
- Bezault, E. (2005). Etude Du Systeme De Determinisme Du Sexe au Sein De Populations Naturelles De Tilapia Du Nil, *Oreochromis Niloticus* (linnaeus, 1758) : importance Des Composantes Genetiques Et Environnementales. UNIVERSITE DE PARIS-SUD.
- Bezault, E., Clota, F., Derivaz, M., Chevassus, B., and Baroiller, J.-F. (2007). Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture* 272, S3–S16.
- Bezault, E., Balaesque, P., Toguyeni, A., Fermon, Y., Araki, H., Baroiller, J.-F., and Rognon, X. (2011). Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genetics* 12, 102.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

- Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Travisan, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., et al. (2019). Fine mapping using whole-genome sequencing confirms anti-müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3 (Bethesda)* 9, 3213–3223.
- de Caestecker, M. (2004). The transforming growth factor- β superfamily of receptors. *Cytokine & Growth Factor Reviews* 15, 1–11.
- Chapman, R.W., Reading, B.J., and Sullivan, C.V. (2014). Ovary Transcriptome Profiling via Artificial Intelligence Reveals a Transcriptomic Fingerprint Predicting Egg Quality in Striped Bass, *Morone saxatilis*. *PLOS ONE* 9, e96818.
- Crespo, B., Gómez, A., Mazón, M.J., Carrillo, M., and Zanuy, S. (2013). Isolation and characterization of *Ff1* and *Gsdf* family genes in European sea bass and identification of early gonadal markers of precocious puberty in males. *General and Comparative Endocrinology* 191, 155–167.
- D’Cotta, H., Fostier, A., Guiguen, Y., Govoroun, M., and Baroiller, J.-F. (2001). Aromatase plays a key role during normal and temperature-induced sex differentiation of tilapia *Oreochromis niloticus*. *Molecular Reproduction and Development* 59, 265–276.
- Deveson, I.W., Holleley, C.E., Blackburn, J., Marshall Graves, J.A., Mattick, J.S., Waters, P.D., and Georges, A. (2017). Differential intron retention in *Jumonji* chromatin modifier genes is implicated in reptile temperature-dependent sex determination. *Sci. Adv.* 3, e1700731.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Domingos, F.F.T., Thomé, R.G., Martinelli, P.M., Sato, Y., Bazzoli, N., and Rizzo, E. (2013). Role of HSP70 in the regulation of the testicular apoptosis in a seasonal breeding teleost *Prochilodus argenteus* from the São Francisco River, Brazil. *Microsc Res Tech* 76, 350–356.
- Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 4, 1184–1191.
- Eshel, O., Shirak, A., Weller, J.I., Hulata, G., and Ron, M. (2012). Linkage and physical mapping of sex Region on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3: Genes, Genomes, Genetics* 2, 35–42.
- Eshel, O., Shirak, A., Dor, L., Band, M., Zak, T., Markovich-Gordon, M., Chalifa-Caspi, V., Feldmesser, E., Weller, J.I., Seroussi, E., et al. (2014). Identification of male-specific *amh* duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 15, 774.
- FAO (2021). The State of World Fisheries and Aquaculture 2021 (FAO).
- Fernandino, J.I., Hattori, R.S., Kishii, A., Strüssmann, C.A., and Somoza, G.M. (2012). The Cortisol and Androgen Pathways Cross Talk in High Temperature-Induced Masculinization: The 11 β -Hydroxysteroid Dehydrogenase as a Key Enzyme. *Endocrinology* 153, 6003–6011.
- Fernandino, J.I., Hattori, R.S., Moreno Acosta, O.D., Strüssmann, C.A., and Somoza, G.M. (2013). Environmental stress-induced testis differentiation: Androgen as a by-product of cortisol inactivation. *General and Comparative Endocrinology* 192, 36–44.
- Feron, R., Zahm, M., Cabau, C., Klopp, C., Roques, C., Bouchez, O., Ech , C., Vali re, S., Donnadi u, C., Haffray, P., et al. (2020). Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale

- genome assembly of yellow perch, *Perca flavescens*. *Molecular Ecology Resources* 20, 531–543.
- Fowler, B.L.S., and Buonaccorsi, V.P. (2016). Genomic characterization of sex-identification markers in *Sebastes carnatus* and *Sebastes chrysomelas* rockfishes. *Molecular Ecology* 25, 2165–2175.
- Gammerdinger, W.J., Conte, M.A., Acquah, E.A., Roberts, R.B., and Kocher, T.D. (2014). Structure and decay of a proto-Y region in Tilapia, *Oreochromis niloticus*. *BMC Genomics* 15, 975.
- Ge, C., Ye, J., Weber, C., Sun, W., Zhang, H., Zhou, Y., Cai, C., Qian, G., and Capel, B. (2018). The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science*.
- Gennotte, V., Mélard, C., D’Cotta, H., Baroiller, J.-F., and Rougeot, C. (2014). The sensitive period for male-to-female sex reversal begins at the embryonic stage in the Nile tilapia and is associated with the sexual genotype. *Molecular Reproduction and Development* 81, 1146–1158.
- Goikoetxea, A., Todd, E.V., and Gemmell, N.J. (2017). Stress and sex: does cortisol mediate sex change in fish? *Reproduction* 154, R149–R160.
- Gonçalves-de-Freitas, E., Bolognesi, M.C., Gauy, A.C. dos S., Brandão, M.L., Giaquinto, P.C., and Fernandes-Castilho, M. (2019). Social Behavior and Welfare in Nile Tilapia. *Fishes* 4, 23.
- Guiguen, Y., Fostier, A., Piferrer, F., and Chang, C.-F. (2010). Ovarian aromatase and estrogens: A pivotal role for gonadal sex differentiation and sex change in fish. *General and Comparative Endocrinology* 165, 352–366.
- Guiguen, Y., Fostier, A., and Herpin, A. (2018). Sex Determination and Differentiation in Fish: Genetic, Genomic, and Endocrine Aspects. In *Sex Control in Aquaculture*, (John Wiley & Sons), pp. 35–64.
- Hattori, R.S., Fernandino, J.I., Kishii, A., Kimura, H., Kinno, T., Oura, M., Somoza, G.M., Yokota, M., Strüssmann, C.A., and Watanabe, S. (2009). Cortisol-Induced Masculinization: Does Thermal Stress Affect Gonadal Fate in Pejerrey, a Teleost Fish with Temperature-Dependent Sex Determination? *PLOS ONE* 4, e6548.
- Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., and Strüssmann, C.A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *PNAS* 109, 2955–2959.
- He, Y., Shang, X., Sun, J., Zhang, L., Zhao, W., Tian, Y., Cheng, H., and Zhou, R. (2010). Gonadal apoptosis during sex reversal of the rice field eel: implications for an evolutionarily conserved role of the molecular chaperone heat shock protein 10. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 314B, 257–266.
- Herpin, A., and Schartl, M. (2015). Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Reports* 16, 1260–1274.
- Honeycutt, J.L., Deck, C.A., Miller, S.C., Severance, M.E., Atkins, E.B., Luckenbach, J.A., Buckel, J.A., Daniels, H.V., Rice, J.A., Borski, R.J., et al. (2019). Warmer waters masculinize wild populations of a fish with temperature-dependent sex determination. *Scientific Reports* 9, 6527.
- Horiguchi, R., Nozu, R., Hirai, T., Kobayashi, Y., Nagahama, Y., and Nakamura, M. (2013). Characterization of gonadal soma-derived factor expression during sex change in the protogynous wrasse, *Halichoeres trimaculatus*. *Developmental Dynamics* 242, 388–399.

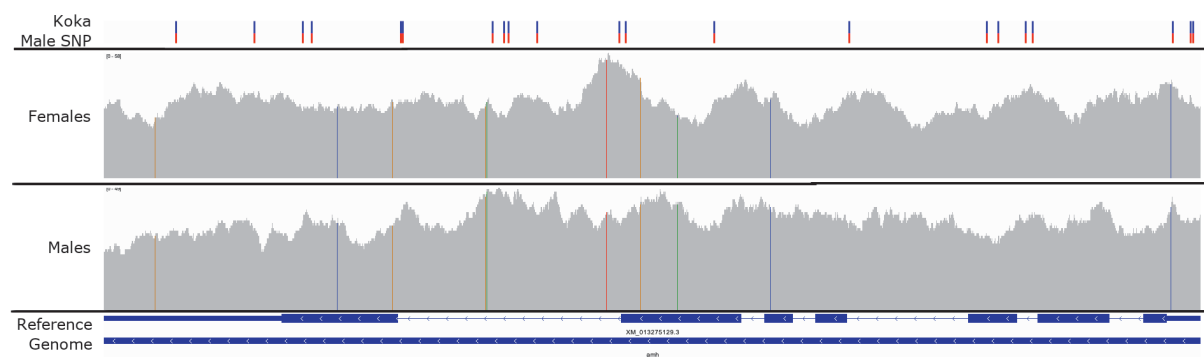
- Ijiri, S., Kaneko, H., Kobayashi, T., Wang, D.-S., Sakai, F., Paul-Prasanth, B., Nakamura, M., and Nagahama, Y. (2008). Sexual Dimorphic Expression of Genes in Gonads During Early Differentiation of a Teleost Fish, the Nile Tilapia *Oreochromis niloticus*. *Biol Reprod* 78, 333–341.
- Jalabert, B., Moreau, J., Planquette, P., and Billard, R. (1974). Déterminisme du sexe chez *tilapia macrochir* et *tilapia nilotica* : action de la méthyltestostérone dans l'alimentation des alevins sur la différenciation sexuelle; proportion des sexes dans la descendance des mâles «inversés». *Annales de Biologie Animale Biochimie Biophysique Vol. 14*, 729–739.
- Jiang, D.-N., Yang, H.-H., Li, M.-H., Shi, H.-J., Zhang, X.-B., and Wang, D.-S. (2016). *gsdf* is a downstream gene of *dmrt1* that functions in the male sex determination pathway of the Nile tilapia. *Molecular Reproduction and Development* 83, 497–508.
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., et al. (2012). A trans-species missense SNP in *Amlhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genet* 8.
- Kaneko, H., Ijiri, S., Kobayashi, T., Izumi, H., Kuramochi, Y., Wang, D.-S., Mizuno, S., and Nagahama, Y. (2015). Gonadal soma-derived factor (*gsdf*), a TGF-beta superfamily gene, induces testis differentiation in the teleost fish *Oreochromis niloticus*. *Molecular and Cellular Endocrinology* 415, 87–99.
- Kassambara, A., and Mundt, F. (2020). Package 'factoextra'. Extract and visualize the results of multivariate data analyses.
- Kobayashi, T., and Nagahama, Y. (2009). Molecular aspects of gonadal differentiation in a teleost fish, the Nile tilapia. *Sex Dev* 3, 108–117.
- Kobayashi, T., Kajiura-Kobayashi, H., Guan, G., and Nagahama, Y. (2008). Sexual dimorphic expression of *DMRT1* and *Sox9a* during gonadal differentiation and hormone-induced sex reversal in the teleost fish Nile tilapia (*Oreochromis niloticus*). *Developmental Dynamics* 237, 297–306.
- Kohn, Y.Y., Symonds, J.E., Kleffmann, T., Nakagawa, S., Lagisz, M., and Lokman, P.M. (2015). Proteomic analysis of early-stage embryos: implications for egg quality in hapuku (*Polyprius oxygeneios*). *Fish Physiol Biochem* 41, 1403–1417.
- Kolde, R. (2015). Package 'pheatmap', R package.
- Koyama, T., Nakamoto, M., Morishima, K., Yamashita, R., Yamashita, T., Sasaki, K., Kuruma, Y., Mizuno, N., Suzuki, M., Okada, Y., et al. (2019). A SNP in a Steroidogenic Enzyme Is Associated with Phenotypic Sex in *Seriola* Fishes. *Current Biology* 29, 1901–1909.e8.
- Lee, B.-Y., Penman, D.J., and Kocher, T.D. (2003). Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics* 34, 379–383.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genetics* 11, e1005678.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 550.

- Mahanty, A., Purohit, G.K., Mohanty, S., and Mohanty, B.P. (2019). Heat stress-induced alterations in the expression of genes associated with gonadal integrity of the teleost *Puntius sophore*. *Fish Physiol Biochem* 45, 1409–1417.
- Marshall Graves, J.A., and Peichel, C.L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biol* 11, 205.
- Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., et al. (2002). *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417, 559–563.
- Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., Naruse, K., Hamaguchi, S., and Sakaizumi, M. (2012). Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics* 191, 163–170.
- Nakamura, M., Kobayashi, T., Chang, X.-T., and Nagahama, Y. (1998). Gonadal Sex Differentiation in Teleost Fish. *The Journal of Experimental Zoology* 281, 11.
- Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A., et al. (2002). A duplicated copy of *DMRT1* in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *PNAS* 99, 11778–11783.
- Ndiwa, T.C., Nyingi, D.W., and Agnese, J.-F. (2014). An Important Natural Genetic Resource of *Oreochromis niloticus* (Linnaeus, 1758) Threatened by Aquaculture Activities in Lobo Drainage, Kenya. *PLOS ONE* 9, e106972.
- Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2013). Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS ONE* 8, e68389.
- Palaiokostas, C., Bekaert, M., Khan, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2015). A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 16, 171.
- Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.-M., Bobe, J., Concordet, J.-P., et al. (2019). Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genetics* 15, e1008013.
- Pan, Q., Feron, R., Jouanno, E., Darras, H., Herpin, A., Koop, B., Rondeau, E., Goetz, F.W., Larson, W.A., Bernatchez, L., et al. (2021). The rise and fall of the ancient northern pike master sex-determining gene. *ELife* 10, e62858.
- Pfeiffer, J. (2018). Rapid progression through the cell cycle ensures efficient migration of primordial germ cells – The role of *Hsp90*. *Developmental Biology* 10.
- Poonlaphdecha, S., Pepey, E., Huang, S.-H., Canonne, M., Soler, L., Mortaji, S., Morand, S., Pfennig, F., Mélard, C., Baroiller, J.F., et al. (2011). Elevated *amh* Gene Expression in the Brain of Male Tilapia (*Oreochromis niloticus*) during Testis Differentiation. *Sex Dev* 5, 33–47.
- Poonlaphdecha, S., Pepey, E., Canonne, M., de Verdal, H., Baroiller, J.-F., and D’Cotta, H. (2013). Temperature induced-masculinisation in the Nile tilapia causes rapid up-regulation of both *dmrt1* and *amh* expressions. *General and Comparative Endocrinology* 193, 234–242.
- Purcell, C.M., Seetharam, A.S., Snodgrass, O., Ortega-García, S., Hyde, J.R., and Severin, A.J. (2018). Insights into teleost sex determination from the *Seriola dorsalis* genome assembly. *BMC Genomics* 19, 31.

- Reichwald, K., Petzold, A., Koch, P., Downie, B.R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., et al. (2015). Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell* 163, 1527–1538.
- Sawatari, E., Shikina, S., Takeuchi, T., and Yoshizaki, G. (2007). A novel transforming growth factor- β superfamily member expressed in gonadal somatic cells enhances primordial germ cell and spermatogonial proliferation in rainbow trout (*Oncorhynchus mykiss*). *Developmental Biology* 301, 266–275.
- Sissao, R., D’Cotta, H., Baroiller, J.-F., and Toguyeni, A. (2019). Mismatches between the genetic and phenotypic sex in the wild Kou population of Nile tilapia *Oreochromis niloticus*. *PeerJ* 7.
- Sudhakumari, C.C., Senthilkumaran, B., Kobayashi, T., Kajiura-Kobayashi, H., Wang, D.S., Yoshikuni, M., and Nagahama, Y. (2005). Ontogenic expression patterns of several nuclear receptors and cytochrome P450 aromatases in brain and gonads of the Nile tilapia *Oreochromis niloticus* suggests their involvement in sex differentiation. *Fish Physiol Biochem* 31, 129.
- Sun, L.X., Teng, J., Zhao, Y., Li, N., Wang, H., and Ji, X.S. (2018). Gonad Transcriptome Analysis of High-Temperature-Treated Females and High-Temperature-Induced Sex-Reversed Neomales in Nile Tilapia. *International Journal of Molecular Sciences* 19, 689.
- Sun, Y.-L., Jiang, D.-N., Zeng, S., Hu, C.-J., Ye, K., Yang, C., Yang, S.-J., Li, M.-H., and Wang, D.-S. (2014). Screening and characterization of sex-linked DNA markers and marker-assisted selection in the Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 433, 19–27.
- Takehana, Y., Matsuda, M., Myosho, T., Suster, M.L., Kawakami, K., Shin-I, T., Kohara, Y., Kuroki, Y., Toyoda, A., Fujiyama, A., et al. (2014). Co-option of *Sox3* as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nat Commun* 5, 4157.
- Tao, W., Yuan, J., Zhou, L., Sun, L., Sun, Y., Yang, S., Li, M., Zeng, S., Huang, B., and Wang, D. (2013). Characterization of Gonadal Transcriptomes from Nile Tilapia (*Oreochromis niloticus*) Reveals Differentially Expressed Genes. *PLOS ONE* 8, e63604.
- Taslina, K., Wehner, S., Taggart, J.B., de Verdal, H., Benzie, J.A.H., Bekaert, M., McAndrew, B.J., and Penman, D.J. (2020). Sex determination in the GIFT strain of tilapia is controlled by a locus in linkage group 23. *BMC Genet* 21, 49.
- Teng, J., Zhao, Y., Chen, H.J., Wang, H., and Ji, X.S. (2020). Transcriptome profiling and analysis of genes associated with high temperature-induced masculinization in sex-undifferentiated Nile tilapia gonad. *Mar Biotechnol* 22, 367–379.
- Trewavas, E. (1983). Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*.
- Triay, C., Conte, M.A., Baroiller, J.-F., Bezault, E., Clark, F.E., Penman, D.J., Kocher, T.D., and D’Cotta, H. (2020). Structure and Sequence of the Sex Determining Locus in Two Wild Populations of Nile Tilapia. *Genes* 11, 1017.
- Wang, D.-S., Kobayashi, T., Zhou, L.-Y., Paul-Prasanth, B., Ijiri, S., Sakai, F., Okubo, K., Morohashi, K., and Nagahama, Y. (2007). *Foxl2* up-regulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1. *Mol Endocrinol* 21, 712–725.
- Wang, J., Liu, Y., Jiang, S., Li, W., Gui, L., Zhou, T., Zhai, W., Lin, Z., Lu, J., and Chen, L. (2019). Transcriptomic and epigenomic alterations of Nile tilapia gonads sexually reversed by high temperature. *Aquaculture* 508, 167–177.
- Wang, Y.Y., Sun, L.X., Zhu, J.J., Zhao, Y., Wang, H., Liu, H.J., and Ji, X.S. (2017). Epigenetic control of *cyp19a1a* expression is critical for high temperature induced Nile tilapia masculinization. *J. Therm. Biol.* 69, 76–84.

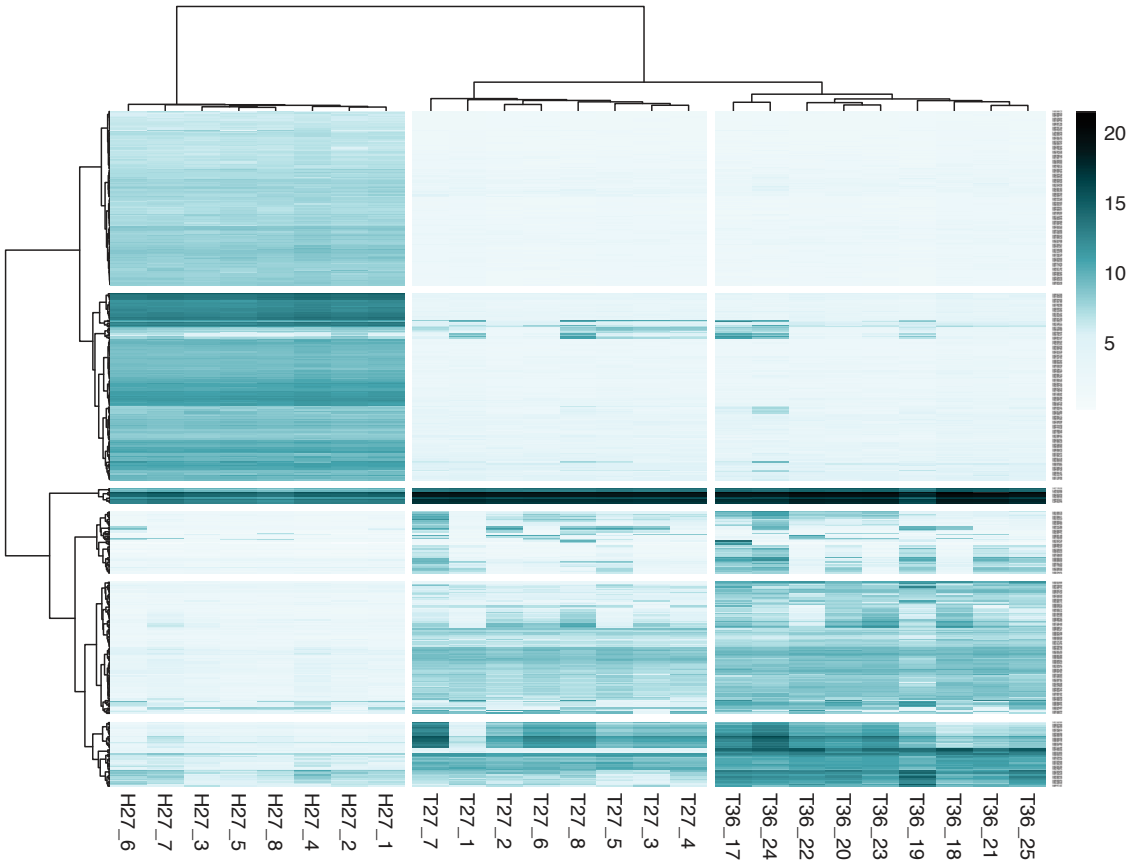
- Weber, C., Zhou, Y., Lee, J.G., Looger, L.L., Qian, G., Ge, C., and Capel, B. (2020). Temperature-dependent sex determination is mediated by *pSTAT3* repression of *Kdm6b*. *Science* 368, 303–306.
- Wessels, S., and Hörstgen-Schwark, G. (2007). Selection experiments to increase the proportion of males in Nile tilapia (*Oreochromis niloticus*) by means of temperature treatment. *Aquaculture* 272, S80–S87.
- Wessels, S., Krause, I., Floren, C., Schütz, E., Beck, J., and Knorr, C. (2017). ddRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC Genomics* 18, 531.
- Yamaguchi, T., Yoshinaga, N., Yazawa, T., Gen, K., and Kitano, T. (2010). Cortisol Is Involved in Temperature-Dependent Sex Determination in the Japanese Flounder. *Endocrinology* 151, 3900–3908.
- Yamamoto, Y., Zhang, Y., Sarida, M., Hattori, R.S., and Strüssmann, C.A. (2014). Coexistence of Genotypic and Temperature-Dependent Sex Determination in *Pejerrey Odontesthes bonariensis*. *PLOS ONE* 9, e102574.
- Yan, Y.-L., Desvignes, T., Bremiller, R., Wilson, C., Dillon, D., High, S., Draper, B., Buck, C.L., and Postlethwait, J. (2017). Gonadal soma controls ovarian follicle proliferation through *Gsdf* in zebrafish. *Dev. Dyn.* 246, 925–945.
- Yang, Y., Liu, Q., Xiao, Y., Xu, S., Wang, X., Yang, J., Song, Z., You, F., and Li, J. (2019). High temperature increases the *gsdf* expression in masculinization of genetically female Japanese flounder (*Paralichthys olivaceus*). *General and Comparative Endocrinology* 274, 17–25.
- Yang, Y., Liu, Q., Xiao, Y., Xu, S., Wang, X., Yang, J., Song, Z., You, F., and Li, J. (2020). Effects of environmental stress (sex steroids and heat) during sex differentiation in Japanese flounder (*Paralichthys olivaceus*): insight from germ cell proliferation and *gsdf-amhcyp19a1a* expression. *Aquaculture* 515, 734536.
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., Cabau, C., Bouchez, O., Fostier, A., and Guiguen, Y. (2012). An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Current Biology* 22, 1423–1428.
- Yao, Z.L., Chen, H.J., Zhao, Y., Cao, Z.J., Wang, H., and Ji, X.S. (2021). A time course transcriptome analysis of brains from sex-undifferentiated Nile tilapia discloses genes associated with high-temperature-induced masculinization. *Aquaculture* 530, 735762.
- Zhang, X., Guan, G., Li, M., Zhu, F., Liu, Q., Naruse, K., Herpin, A., Nagahama, Y., Li, J., and Hong, Y. (2016). Autosomal *gsdf* acts as a male sex initiator in the fish medaka. *Scientific Reports* 6, 19738.
- Zhao, Y., Mei, Y., Chen, H.J., Zhang, L.T., Wang, H., and Ji, X.S. (2019). Profiling expression changes of genes associated with temperature and sex during high temperature-induced masculinization in the Nile tilapia brain. *Physiological Genomics* 51, 159–168.
- Zheng, Q., Xiao, H., Shi, H., Wang, T., Sun, L., Tao, W., Kocher, T.D., Li, M., and Wang, D. (2020). Loss of *Cyp11c1* causes delayed spermatogenesis due to the absence of 11-ketotestosterone. *Journal of Endocrinology* 244, 487–499.

Supplementary Materials



Supplementary Figure 1. Comparison of males pool and females pool *amh* gene mapping of Hora short reads whole genome sequencing to the *O. niloticus* reference genome.

No sex-specific SNP can be found between Hora males and females over the *amh* gene, and none of the Koka Male specific SNP are existing in the Hora males.

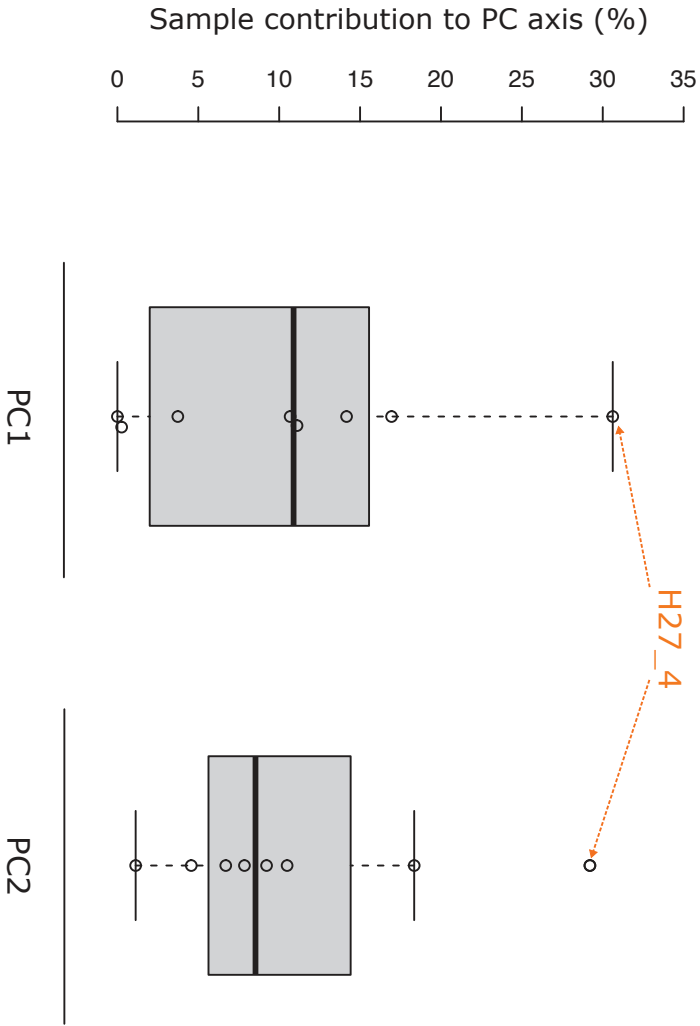


Supplementary Figure 2. Heatmap of the top 500 most variable genes using a hierarchical clustering derived from Euclidean distance from normalized and log2 transformed counts and a Ward's distance as agglomeration method. Each line represent a gene.



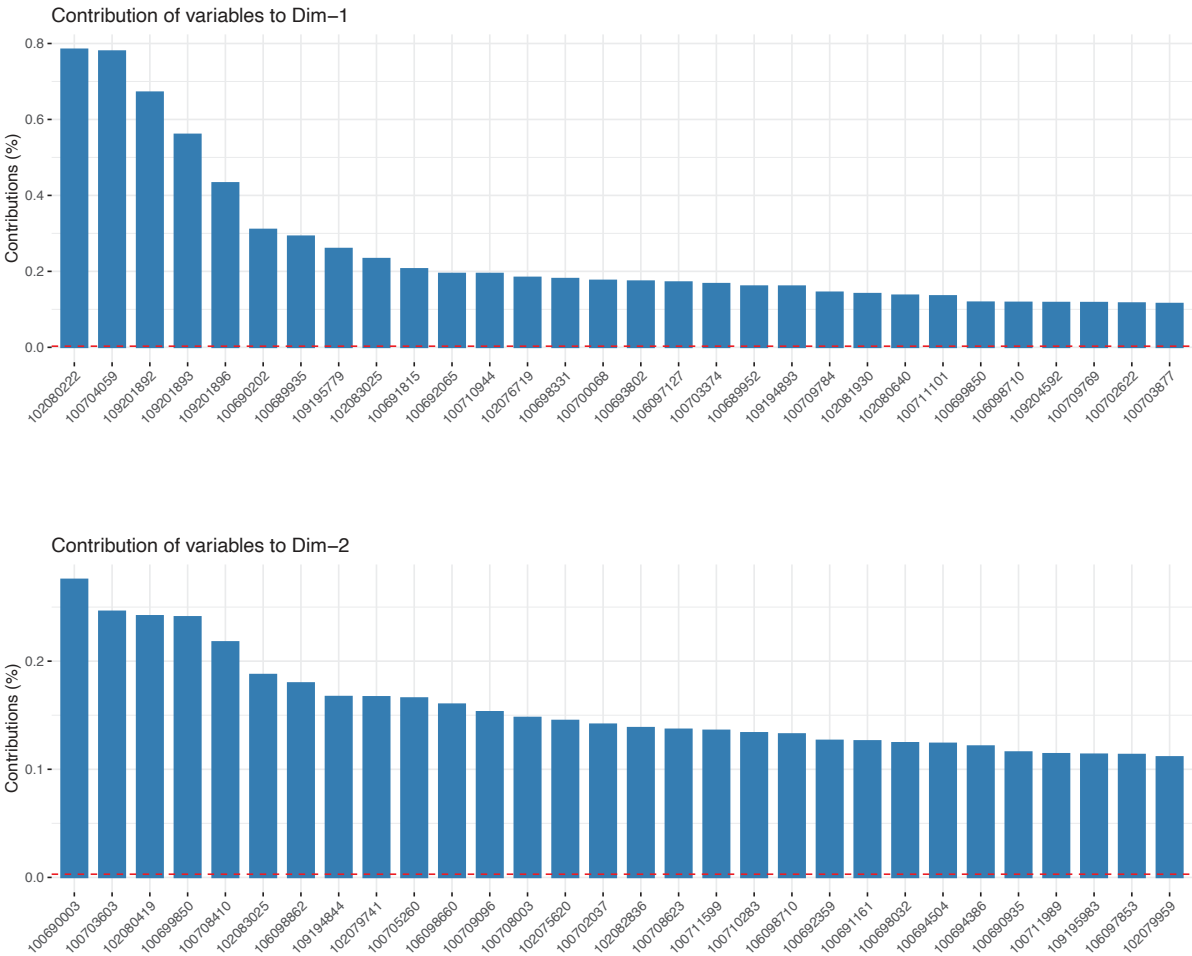
Supplementary Figure 3. GO categories barplot of the top 500 most variable genes among all tissues and samples.

Only 208 genes out of the 500 identified had annotations in the GO database and are thus represented in this barplot.

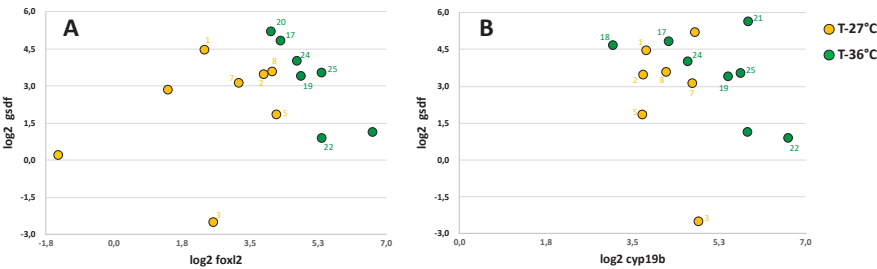


Supplementary Figure 4. Boxplot of the contribution of each sample to the PCA.

The sample's contributions to PC1 (left) and PC2 (right) show that sample 4 is the most extreme point for these two PC axes and explains the pattern seen on our biplot PCA.



Supplementary Figure 5. Barplot of genes contribution to dimension 1 (Top) and dimension 2 (Bottom) of the PCA run with all genes for trunks samples.
Only the top 30 genes contributing to PC1 and PC2 are represented.



Supplementary Figure 6. Relative expression of individual trunks of *gsdf* plotted relative to *foxl2* (A), and *cyp19a1b* (B) expressed as the log2 of maximum normalized expression (MNE).
Individuals that were also analyzed by RNASeq are numbered.

Supplementary Table 1. List of 141 candidate genes already involved in the sex differentiation pathway in teleosts or other vertebrates used in our analyses.

Description Names	Gene	LOC ID	LG
P450(11beta) hydroxylase type I	cyp11b1	100533461	
sox30 Sox30 protein	sox30	100533980	
Follicle stimulating hormone receptor (=gth-r1)	fsh-R	100534395	LG8
brain aromatase	cyp19a1B	100534396	
gonadotropin receptor 2	Gth-RII	100534397	LG13
glucocorticoid R (gr1 =nr3c1)	nr3c1	100534398	LG2
growth hormone receptor a (=ghr1a)	ghr1	100534400	LG12
androgen receptor alpha(ara)	arA	100534409	
androgen receptor (ar)	arB	100534410	
gata4 GATA binding protein 4	gata4	100534420	
insulin-like growth factor II	igf-2	100534429	LG7
heat shock protein 70	hsp70	100534448	LG4
hydroxy-delta-5-steroid dehydrogenase	hsd3b7	100534450	
Growth hormone 1	gh1	100534452	LG4
nr0b1a orphan nuclear receptor Dax-1	dax1A	100534476	
gonadotropin releasing hormone 1	gnrh1	100534486	LG12
gonadotropin releasing hormone 3	gnrh3	100534487	LG13
progesterone receptor	pgr	100534498	LG14
Follicle stimulating hormone beta subunit (=gth1)	Fsh	100534500	LG1
Luteinizing hormone (beta subunit)	Lh	100534501	LG13
hydroxysteroid 11-beta-dehydrogenase 1-like	hsd11b2	100534505	LG11
Putative gonadotropin-releasing hormone II receptor		100534508	LG1
putative gonadotropin releasing hormone 2	GnRHR2	100534509	
dmrt2 doublesex and mab-3 related transcription factor 2a	dmrt2	100534511	LG12
Estrogen receptor (nr3a1)	nr3a1	100534514	LG15
Estrogen receptor (nr3a2)	nr3a2	100534515	LG15
foxl2 forkhead box L2	foxl2	100534517	
putative gonadotropin-releasing hormone II receptor		100534518	LG1
DNA meiotic recombinase1	dmc1	100534532	
Growth hormone receptor (=ghr2)	ghr	100534534	LG7
SRY-box transcription factor 10	Sox10	100534551	LG4
nr0b1b orphan nuclear receptor DAX2	dax1B	100534552	
doublesex- and mab-3-related transcription factor 1-LG12	dmrt1	100534555	LG12
Estrogen receptor (Erbeta2)	esr2b	100534556	LG19
DMRT like family A1 (=dmrt4)	dmrt4	100534559	LG3
insulin-like growth factor I	igf-1	100534565	
hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1	hsd3b1	100534568	LG16
putative gonadotropin releasing hormone 2	gnrh2	100534574	LG5
glucocorticoid R (gr2)	gr2	100534588	LG10
nr5a1 nuclear receptor subfamily 5 group A member 1	sf-1	100628563	
cytochrome P450 1B1	cyp1b1	100653403	
rspo1 R-spondin 1	rspo1	100653404	
catenin beta 1 (ctnnb1)	ctnnb1A	100690187	
nr5a2 nuclear receptor subfamily 5 group A member 2	nr5a2	100690205	
LOC100690744 nanos homolog 1	nanos1B	100690744	
lysine (K)-specific demethylase 3B	kdm3b	100690767	
cytochrome P450 26B1-like	cyp26b1	100690880	
bone morphogenetic protein receptor type-2-LG23	bmpr2	100690976	LG23
testis expressed 2	Tex2	100691039	LG4
growth differentiation factor 6b	Gdf6	100691097	LG22
SRY-box transcription factor 1a	Sox1a	100691128	LG16
SMAD family member 7	Smad7	100691821	
wnt4 Wnt family member 4	wnt4A	100691883	

Description Names	Gene	LOC ID	LG
SMAD family member 3a	Smad3a	100692385	LG1
DMRT-like family A2 (=dmrt5)	dmrt2	100692595	LG23
DNA (cytosine-5-)-methyltransferase 3 beta, duplicate b.1	dnmt3bb1	100692621	
DNA (cytosine-5)-methyltransferase 3A	dnmt3a	100693331	
LOC100693400 transcription factor Sox-9-A	sox9A	100693400	
SRY-box transcription factor 19b	sox19b	100693580	LG3
Smad2 mothers against decapentaplegic homolog 2	Smad2	100693777	LG7
mothers against decapentaplegic homolog 4		100693833	LG11
LOC100693957 nanos homolog 1	nanos1A	100693957	
catenin beta-1 (LOC100693960)	ctnnb1B	100693960	
lysine-specific demethylase 6A		100694035	LG23
cytosolic phospholipase A2	pla2g4a	100694325	
factor in the germline alpha-like (LOC100694426)	figla	100694426	
LOC100694473 Wilms tumor protein homolog	wt1A	100694473	
Kisspeptin 2	kiss2	100694550	LG15
SRY-box transcription factor 6	sox6	100694759	LG1
lysine (K)-specific demethylase 4A, genome duplicate a	kdm4aa	100694970	
bone morphogenetic protein receptor type-2-LG16		100694986	LG16
SMAD family member 6b	Smad6b	100695008	LG7
SRY-box transcription factor 2	Sox2	100695175	LG17
SMAD family member 3b	Smad3b	100695277	LG7
jumonji domain containing 1Cb	jmjd1cb	100695422	
methionine adenosyltransferase II, beta	mat2b	100695434	
lysine (K)-specific demethylase 5Bb	kdm5bb	100695530	
lysine (K)-specific demethylase 2Ab	kdm2ab	100695561	
DNA (cytosine-5-)-methyltransferase 3 beta, duplicate a	dnmt3ba	100696323	
heat shock protein 90, alpha (cytosolic), class B member 1	hsp90ab1	100696358	
lysine-specific demethylase 4A		100696578	
nuclear receptor subfamily 5 group A member 2	nr5a5	100696603	
SRY-box transcription factor 1b	Sox1b	100697857	LG23
jumonji domain containing 4	jmjd4	100698106	
fgfr2 fibroblast growth factor receptor 1a	FgFr1a	100698136	
Vasa (3 different genes see with Matt)	vas	100698352	
SRY-box transcription factor 3	sox3	100698575	LG2
jumonji domain containing 6, arginine demethylase and lysine hydroxylase	jmjd6	100698738	
lysine (K)-specific demethylase 8	kdm8	100699006	
activator of hsp90	ahsa1	100699088	
lysine (K)-specific demethylase 7Aa	kdm7aa	100699317	
fgfr2 fibroblast growth factor receptor 2	FgFr2	100699731	
LOC100699759 transcription factor Sox-9 B	sox9B	100699759	
lysine (K)-specific demethylase 6B, b	kdm6bb	100699848	
heat shock 70 kDa protein 1		100700083	
cAMP-dependent protein kinase catalytic subunit alpha		100700452	
heat shock protein 90	hsp90b1	100700552	
mothers against decapentaplegic homolog 4		100700558	LG12
jumonji domain containing 7	jmjd7	100700949	
synaptonemal complex protein 3	sycp3	100700972	
LOC100701078 Wilms tumor protein homolog	wt1B	100701078	
jumonji domain containing 8	jmjd8	100701174	
heat shock protein 8	hsp9	100701377	
lysine (K)-specific demethylase 5A	kdm5a	100701637	
lysine (K)-specific demethylase 5C	kdm5c	100701700	
heat shock protein 90, alpha (cytosolic), class A member 1, tandem duplicate 2	hsp90aa1.2	100701982	

Description Names	Gene	LOC ID	LG
heat shock protein 90, alpha (cytosolic), class A member 1, tandem duplicate 1	hsp90aa1.1	100702247	
DNA (cytosine-5)-methyltransferase 3A	dnmt3aa	100702269	LG15
adenylate cyclase type 6		100702460	
lysine (K)-specific demethylase 4B	kdm4b	100703178	
ribonuclease P/MRP 30 subunit	rpp30	100703274	
trdm1 tRNA aspartic acid methyltransferase 1	dnmt2	100703654	LG9
cathepsin L1		100705087	
PHD finger protein 2	phf2	100705594	
Luteinizing hormone receptor (lhcr)	lh-R	100705740	LG13
adenylate cyclase type 8		100705805	
methyltransferases (dnmt)	dnmt1	100706458	LG4
gonadotropin releasing hormone 4	gnrh4	100707068	LG7
anti-Mullerian hormone (amh)	amh	100707206	
lysine (K)-specific demethylase 4C	kdm4c	100707211	
insulin-like growth factor 1 receptor-like	igf-1r	100707581	
PHD finger protein 8	phf8	100707999	
LOC100708217 cytochrome P450 19A1-like	cyp19a1A	100708217	
SMAD family member 4a	Smad4a	100708381	LG7
3-keto-steroid reductase		100708654	
tdrd1 tudor domain containing 1 [	tdrd1	100709682	
adenylate cyclase type 3		100710007	
guanine nucleotide-binding protein G(s) subunit alpha		100710082	
aldehyde dehydrogenase 1 family member A1	Aldh1a1	100710237	
gsdf growth/differentiation factor 6-B-like	gsdf	100710262	LG7
SMAD family member 2	Smad2	100710325	LG12
Wnt2	Wnt2	100710509	LG7
low density lipoprotein receptor a	ldlra	100711286	
minealocorticoid R (=nr3c2)	mr	100712208	LG6
lysine demethylase 6B	kdm6ba	100712234	LG3
LOC100712297 protein Wnt-4	wnt4B	100712297	
doublesex- and mab-3-related transcription factor 1		100712376	LG22
lysine (K)-specific demethylase 6A	kdm6a	100712385	
lysine (K)-specific demethylase 5Ba	kdm5ba	100712531	
TGF-b3	TGF-b3	100712568	
insulin receptor-like		102777679	

Supplementary Table 2. Number of raw, clean and mapped reads of Hora RNA sequencing.

Sample	Sample ID	Raw reads	Trimming output	Mapping output	
				Uniquely mapped reads	% of uniquely mapped reads
HN DT1	T27_1	23491158	21013911	18153700	86.39%
HN DT2	T27_2	31899564	29712909	28322680	95.32%
HN DT3	T27_3	21748948	20058428	19175366	95.60%
HN DT4	T27_4	26674375	24838808	23723420	95.51%
HN DT5	T27_5	22874995	21246013	20340942	95.74%
HN DT6	T27_6	31514924	29306575	28001583	95.55%
HN DT7	T27_7	24391045	22855747	21700772	94.95%
HN DT8	T27_8	27149842	25250651	24159627	95.68%
HN DT9	H27_1	30828164	28278223	26518845	93.78%
HN DT10	H27_2	33991742	31463225	29795137	94.70%
HN DT11	H27_3	26573407	24029331	22782663	94.81%
HN DT12	H27_4	33281072	30884281	29242232	94.68%
HN DT13	H27_5	24380059	22288264	21163260	94.95%
HN DT14	H27_6	39037928	35601928	33507930	94.12%
HN DT15	H27_7	26910813	24684283	23517533	95.27%
HN DT16	H27_8	27455526	24829974	23519105	94.72%
HN DT17	T36_17	28704221	26597398	25114469	94.42%
HN DT18	T36_18	28542948	26548881	25433502	95.80%
HN DT19	T36_19	30027239	27977437	25867525	92.46%
HN DT20	T36_20	29128715	27011150	25806943	95.54%
HN DT21	T36_21	27473253	25374350	24431107	96.28%
HN DT22	T36_22	31871493	29846203	28403104	95.16%
HN DT23	T36_23	36067322	33508249	31975889	95.43%
HN DT24	T36_24	30094046	27882642	26321388	94.40%
HN DT25	T36_25	27625046	24979069	23692622	94.85%

Supplementary Table 3. KEGG pathways overrepresented among the down-regulated genes at 36°C compared to 27°C control temperature

KEGG ID	Pathway Name	Genes Found	Genes in Pathway	Percentage	pvalue	pvalueAdj	Subcategories	Categories
04110	Cell cycle	20	158	0.13	1.60E-04	1.42E-03	Cell growth and death	Cellular Processes
03008	Ribosome biogenesis in eukaryotes	31	80	0.39	5.67E-20	4.76E-18		
00970	Aminoacyl-tRNA biosynthesis	14	43	0.33	4.97E-09	1.67E-07	Translation	
03013	RNA transport	28	187	0.15	3.63E-07	5.08E-06		
03020	RNA polymerase	9	31	0.29	4.38E-06	4.91E-05	Transcription	
03040	Spliceosome	22	157	0.14	1.65E-05	1.63E-04		
03030	DNA replication	23	37	0.62	1.04E-21	1.74E-19		Genetic Information Processing
03430	Mismatch repair	12	23	0.52	3.32E-11	1.86E-09		
03410	Base excision repair	13	36	0.36	3.16E-09	1.33E-07	Replication and repair	
03440	Homologous recombination	12	40	0.3	1.41E-07	2.62E-06		
03460	Fanconi anemia pathway	15	53	0.28	1.62E-08	3.90E-07		
03420	Nucleotide excision repair	12	44	0.27	4.91E-07	6.34E-06		
03060	Protein export	7	31	0.23	2.36E-04	1.98E-03	Folding, sorting and degradation	
04141	Protein processing in endoplasmic reticulum	31	219	0.14	3.34E-07	5.08E-06		
00983	Drug metabolism - other enzymes	13	104	0.12	1.81E-03	1.22E-02	Xenobiotics biodegradation and metabolism	
00240	Pyrimidine metabolism	14	76	0.18	1.59E-05	1.63E-04	Nucleotide metabolism	
00230	Purine metabolism	22	188	0.12	2.64E-04	2.11E-03		
00480	Glutathione metabolism	12	94	0.13	2.06E-03	1.33E-02	Metabolism of other amino acids	
00670	One carbon pool by folate	8	21	0.38	8.45E-07	1.01E-05	Metabolism of cofactors and vitamins	
00790	Folate biosynthesis	5	32	0.16	7.94E-03	4.23E-02		
00100	Steroid biosynthesis	10	24	0.42	2.06E-08	4.32E-07	Lipid metabolism	Metabolism
00513	Various types of N-glycan biosynthesis	7	50	0.14	6.19E-03	3.72E-02	Glycan biosynthesis and metabolism	
00510	N-Glycan biosynthesis	8	61	0.13	6.49E-03	3.76E-02		
01230	Biosynthesis of amino acids	15	96	0.16	7.24E-05	6.76E-04		
01240	Biosynthesis of cofactors	30	201	0.15	1.56E-07	2.62E-06	Global and overview maps	
01100	Metabolic pathways	139	1973	0.07	1.08E-08	3.01E-07		
00260	Glycine, serine and threonine metabolism	9	52	0.17	5.27E-04	4.03E-03	Amino acid metabolism	
00250	Alanine, aspartate and glutamate metabolism	9	59	0.15	1.46E-03	1.02E-02		
00270	Cysteine and methionine metabolism	8	63	0.13	8.01E-03	4.23E-02		
00330	Arginine and proline metabolism	9	69	0.13	4.76E-03	2.96E-02		

Supplementary Table 4. KEGG pathways overrepresented among the up-regulated genes at 36°C compared to 27°C control temperature

KEGG ID	Pathway Name	Genes Found	Genes in Pathway	Percentage	pvalue	pvalueAdj	Subcategories	Categories
04145	Phagosome	52	330	0.16	3.18E-10	2.90E-08	Transport and catabolism	Cellular Processes
04514	Cell adhesion molecules	66	480	0.14	3.46E-10	2.90E-08	Signaling molecules and interaction	Environmental Information Processing
04060	Cytokine-cytokine receptor interaction	35	342	0.1	2.10E-03	1.60E-02		
05168	Herpes simplex virus 1 infection	58	531	0.11	9.62E-06	1.24E-04	Infectious disease: viral	Human Diseases
00360	Phenylalanine metabolism	5	22	0.23	2.49E-03	1.76E-02		
00220	Arginine biosynthesis	5	27	0.19	7.42E-03	4.30E-02	Amino acid metabolism	
00270	Cysteine and methionine metabolism	10	63	0.16	2.52E-03	1.76E-02		
00380	Tryptophan metabolism	9	59	0.15	4.88E-03	3.15E-02		
00053	Ascorbate and aldarate metabolism	16	47	0.34	3.05E-09	1.71E-07		
00040	Pentose and glucuronate interconversions	13	47	0.28	1.18E-06	2.48E-05	Carbohydrate metabolism	
00010	Glycolysis / Gluconeogenesis	19	91	0.21	1.35E-06	2.48E-05		
00030	Pentose phosphate pathway	6	35	0.17	7.16E-03	4.29E-02		
00500	Starch and sucrose metabolism	7	45	0.16	8.58E-03	4.81E-02		
00620	Pyruvate metabolism	8	51	0.16	5.69E-03	3.54E-02		
01230	Biosynthesis of amino acids	14	96	0.15	1.42E-03	1.19E-02		
01240	Biosynthesis of cofactors	28	201	0.14	3.97E-05	3.92E-04	Global and overview maps	Metabolism
01200	Carbon metabolism	19	160	0.12	3.77E-03	2.54E-02		
01100	Metabolic pathways	152	1973	0.08	5.74E-07	1.61E-05		
00591	Linoleic acid metabolism	11	35	0.31	1.31E-06	2.48E-05	Lipid metabolism	
00120	Primary bile acid biosynthesis	6	28	0.21	1.90E-03	1.52E-02		
00140	Steroid hormone biosynthesis	17	80	0.21	3.24E-06	4.95E-05		
00590	Arachidonic acid metabolism	16	79	0.2	1.15E-05	1.29E-04		
00830	Retinol metabolism	19	80	0.24	1.48E-07	4.97E-06	Metabolism of cofactors and vitamins	
00860	Porphyrin and chlorophyll metabolism	12	52	0.23	2.35E-05	2.47E-04		
00982	Drug metabolism - cytochrome P450	19	74	0.26	3.63E-08	1.53E-06	Xenobiotics biodegradation and metabolism	
00980	Metabolism of xenobiotics by cytochrome P450	17	76	0.22	1.48E-06	2.48E-05		
00983	Drug metabolism - other enzymes	19	104	0.18	1.15E-05	1.29E-04		
04672	Intestinal immune network for IgA production	22	124	0.18	4.68E-06	6.55E-05	Immune system	Organismal Systems
04620	Toll-like receptor signaling pathway	21	163	0.13	8.93E-04	7.89E-03		

Supplementary Table 5: Primers used for quantitative realtime PCR

Gene name		Primer name (F=forward) (R=reverse)	Primer sequence (5' - 3')	Tm
amh	anti-Müllerian hormone	Amh-F358	CCACAGTAACCGAGTACTCAC	60°C
		Amh-R640	GAACCAAAGTCTGTTTGTCTAG	
ara	androgen receptor alpha	Ara-F840	GCCTTCGGTGTGGATGAATTGC	60°C
		Ara-R976	AGGTGACTGCTCCATAGTGACAAC	
cyp19a1A	aromatase cytochrome P450 family 19 subfamily A polypeptide 1 A	Cyp19A-F1462	CTACTTTCAGCCGTTTCGGTTCAG	60°C
		Cyp19A-R1744	GTTTCGGGTCTCGGAGGGTTTG	
cyp19a1B	aromatase cytochrome P450 family 19 subfamily A polypeptide 1 B	Cyp19B-F957	GATTCATGAAGCCGAGAAGC	60°C
		Cyp19B-R1184	TTCAAGATGGTGTTCATCATCTCCT	
cyp11c1	cytochrome P450 -11 beta hydroxylase	cyp11c1-F-DW	AAAGAAGTCCTCAGGTTGTACC	60°C
		cyp11c1-R-DW	GACCAAAGTTCACAGCAGGTATG	
dmrt1	doublesex and mab-3 related transcription factor 1	Dmrt1-F972	CGGATTGCAGCGGACCGA	63°C
		Dmrt1-R1232	GGACAGAGACACAGGACTAC	
ef1	enlongation factor 1 alpha	Ef1-F1014	AGAACGTCTCCGTCAAGGAAAT	60°C
		Ef1-R1115	GATGATGACCTGAGCGTTGAAG	
foxl2	forkhead transcription factor L2	Foxl2-F310	AAGAGGAGCCGGTTCAGGACAA	60°C
		Foxl2-R396	GCTCTCCCGGATAGCCATGG	
gr2	Glucocorticoid receptor 2	GR2-F	GCCCCAGTAGATGATCTCTGGTT	60°C
		GR2-R	CAGGAGCATGCCCAACT	
gsdf	Gonadal soma-derived factor	Gsdf-F194	CCACCATGGCCTTTCCATTC	60°C
		Gsdf-R930	GACTGCTGGGGTTGCAGTAT	
hsd11b2	hydroxysteroid 11-beta dehydrogenase	Hsd11B2-F546	CTGGGCTTTGAAGTGTTCGC	60°C
		Hsd11B2-R847	AGAAGGCCTTGGTGATGCTC	
jarid2	jumonji and AT-rich interaction domain containing 2	Jarid2-F233	AGAGGAAGTTTGCCAGTCG	60°C
		Jarid2-R389	CTGGAGAGGTCTCGGATGGA	
Kdm6bb	lysine demethylase 6Bb	Kdm6bb-F	GCAGTTTGAGGTCCAGAACTGGCA	62°C
		Kdm6bb-R	GAAGCTCCCTGTTGTTGCGATGCA	

DISCUSSION GÉNÉRALE ET PERSPECTIVES

Un système ancestral qui garde des zones d'ombres

Les études antérieures portant sur le déterminisme du sexe chez le tilapia du Nil ont été menées sur des souches domestiques de laboratoire ou des souches commerciales. Ces prospections ont conclu à un déterminisme génétique du sexe suivant un patron XX/XY pouvant être influencé par la température ainsi que des facteurs génétiques parentaux mineurs (Baroiller et al., 2009). Les régions responsables du déterminisme du sexe dans le génome de cette espèce ont été identifiées préalablement sur deux groupes de liaison (LG) : LG1 et/ou LG23 selon les souches utilisées (Lee et al., 2003; Cnaani et al., 2008; Lee et al., 2011; Eshel et al., 2012; Palaikostas et al., 2013; Eshel et al., 2014; Gammerdinger et al., 2014; Li et al., 2015; Palaikostas et al., 2015; Wessels et al., 2017; Cáceres et al., 2019; Curzon et al., 2020; Taslima et al., 2020, 2021). Sur LG23, une duplication en tandem du gène de l'Hormone Anti-Müllerienne (*amh*) sur le Y est considérée comme la région du déterminisme du sexe (Li et al., 2015). Une première copie de l'*amh*, nommée *amhY* possède une mutation non-synonyme par rapport au X, responsable de la masculinisation (Li et al., 2015). La seconde copie présente quant à elle une insertion de 5bp changeant le cadre de lecture et induisant un codon stop prématuré dans la séquence (Li et al., 2015). Aucun locus n'a été identifié précisément sur LG1 (Gammerdinger et al., 2014).

Nos études génomiques sur les populations naturelles du lac Volta (Afrique de l'Ouest) et du lac Koka en Ethiopie (Afrique de l'Est) ont permis d'établir que la région de la duplication de l'*amh* sur LG23-Y ségrège également pour déterminer le sexe dans les populations naturelles. De plus, nos résultats indiquent que ce système de déterminisme semble être ancestral à la divergence des populations d'Afrique de l'Est et de l'Ouest. Nos analyses ont révélé une structure conservée dont nous avons pu définir en détails les limites ainsi que la composition en gènes. Nous avons également pu identifier des mutations spécifiques aux mâles et communes aux deux populations et à certaines souches commerciales, particulièrement au sein de l'*amh*, qui pourraient induire des changements d'acides aminés dans la séquence protéique. En revanche, parmi ces polymorphismes nucléotidiques entre mâles et femelles, nous n'avons pas observé la mutation non-synonyme décrite comme responsable de la masculinisation dans la souche japonaise (Li et al., 2015).

Une autre différence majeure mise en lumière par nos analyses est le fait qu'une des copies de l'*amh* est tronquée dans une population de l'Ouest, comme décrit sur la souche

japonaise (Li et al., 2015), en revanche, les deux copies de ce gène sont intactes (bien qu'avec quelques polymorphismes nucléotidiques) au sein de la région du sexe chez une des populations de l'Est (Koka, Ethiopie). Ainsi, bien que nous ayons pu identifier avec précision la région associée au sexe dans ces deux populations, nos analyses ne permettent pas de conclure quant à l'aspect fonctionnel des deux copies de l'*amh* et nous ne sommes pas à même de définir la manière dont cette région détermine le sexe. Il semble tout de même que cela ne tienne pas au SNP décrit précédemment dans l'*amhY* par Li et al., (2015). Une étude récente a indiqué que l'*amh* agissait d'une manière dose-dépendante chez le tilapia du Nil pour réguler la folliculogénèse en se liant à l'*amhr2* (Liu et al., 2020). Il est possible que la présence/expression de deux ou bien de trois copies de ce gène induise d'une manière dose-dépendante le développement des mâles. Des études fonctionnelles sont donc nécessaires pour déterminer si la dose a une importance ou s'il s'agit par exemple d'une régulation différentielle au niveau des promoteurs.

Des études améliorant la couverture des données « long reads » permettraient de préciser les séquences nucléotidiques des deux copies de l'*amh* sur le Y et ainsi d'avoir des séquences de références pour chacune des copies de l'*amh* (X, Y et ΔY). Avec les données dont nous disposons actuellement, il est déjà possible de produire ce genre de références en corrigeant les contigs construits *de novo* à partir des lectures longues à l'aide des lectures courtes de ces mêmes populations en passant par des logiciels comme LoRDEC ou Jabba par exemple (Salmela and Rivals, 2014; Miclotte et al., 2016). Il serait également intéressant de produire de nouveaux assemblages *de novo* par des méthodes dites « hybrides » qui utilisent directement une combinaison des lectures courtes (fiables et avec une grande couverture) et des lectures longues (traversant les régions complexes comme les duplications) pour optimiser la construction des contigs. Des logiciels comme WENGAN ou HASLR sont disponibles et présentent l'avantage de prendre peu de temps pour les assemblages car ils se basent sur la construction de graphes (Haghshenas et al., 2020; Di Genova et al., 2021).

Pour compléter ces analyses et comprendre le fonctionnement de cette duplication dans le déterminisme du sexe, il est important de savoir si les deux copies du Y sont transcrites, et si oui, dans quelles mesures. Une analyse du transcriptome pourrait dans un premier temps permettre de définir si les deux copies sont traduites dans les deux populations en les comparant aux transcrits attendus d'après les séquences de référence (X, Y et ΔY). Si la copie

dupliquée produit des protéines, même tronquées, qui sont en mesure de se fixer sur le récepteur de l'*amhr2* alors on peut imaginer que cela ait un impact fonctionnel sur la cascade du déterminisme du sexe.

Enfin, des approches d'édition du génome, notamment par CRISPR/Cas9, peuvent permettre d'étudier les conséquences de modifications ou de knockout de certains gènes sur le phénotype et cela a déjà été utilisé chez le tilapia du Nil (Jiang et al., 2017; Li et al., 2019, 2020; Liu et al., 2020). Cela a été réalisé notamment par Li et al., (2015) pour identifier le SNP non-synonyme de l'*amhY*, considéré par les auteurs comme essentielle pour la masculinisation. Cependant, il n'est formellement précisé si cela a été conduit uniquement sur l'*amhY* ou bien une région plus vaste de forte homologie entre l'*amhY* et *amhΔY*. Mettre en place de tels protocoles pour tenter de déduire l'effet du nombre de copies de l'*amh* sur le phénotype, ou encore étudier leurs promoteurs est certainement prometteur, cependant la similarité qui existe entre les différentes copies de l'*amh* du Y mais aussi du X, notamment chez les mâles Koka par exemple, imposera une difficulté supplémentaire pour cibler une unique copie avec précision.

La variabilité au niveau intra-spécifique, un atout pour comprendre l'évolution des chromosomes sexuels

Malgré la présence d'un déterminisme ancestral qui ségrège dans les populations domestiques comme dans les populations sauvages, nos résultats montrent que ce système n'est pas immuable. Nous avons mis en évidence qu'une des populations éthiopiennes (Hora) ne présente pas la duplication de la région de l'*amh* sur LG23-Y et plus ne présente aucune association de l'*amh* avec le sexe. Ce système sur LG23-Y peut donc être perdu et suggère qu'un polymorphisme du déterminisme du sexe puisse exister chez cette espèce puisque différentes populations n'ont pas systématiquement le même déterminant. D'autres systèmes de déterminismes du sexe ont déjà été mis en évidence chez le tilapia du Nil puisqu'un QTL sur LG1 est associé au sexe chez certaines souches commerciales Américaines ou encore dans la souche de laboratoire Manzala-Stirling (Palaïokostas et al., 2013, 2015; Gammerdinger et al., 2014). Il serait maintenant intéressant d'étudier davantage de populations sauvages pour déterminer la prévalence des systèmes alternatifs à LG23-Y, et s'ils font intervenir les mêmes gènes ou les mêmes conditions environnementales par

exemple. Il serait notamment intéressant de savoir si certaines populations sauvages présentent un déterminisme porté par LG1 ou s'il s'agit d'un artefact de la domestication. Il paraît notamment intéressant d'étudier le(s) déterminisme(s) génétique(s) qui existe(nt) dans le lac Manzala (Basin du Nil, Egypte) d'où sont originaires les souches de Manzala-Stirling qui présentent une association au sexe sur LG1.

De par les motivations économiques pour optimiser les productions aquacoles, les études sur le déterminisme du sexe ont principalement été conduites sur des souches domestiquées. Or, le tilapia du Nil est présent dans des environnements variés en terme de température, de surface, de profondeur et de salinité ce qui en fait une espèce de choix pour comprendre l'effet des facteurs environnementaux sur l'évolution des chromosomes sexuels. De plus, travailler à l'échelle populationnelle permettrait de limiter l'impact de la variation entre les taxons dû aux temps de divergences importants entre espèces et ainsi diminuer les facteurs confondants lors de l'analyse de l'évolution des chromosomes sexuels (Furman et al., 2020). Par exemple, travailler au niveau intra-spécifique a permis de montrer que les réarrangements chromosomiques peuvent être une conséquence et non une cause de la diminution de la recombinaison comme chez le champignon *Neurospora tetrasperma* et le killi turquoise, petit poisson africain (Reichwald et al., 2015; Sun et al., 2017).

Identifier le système de déterminisme du sexe de Hora

Pour mémoire, nous avons vu que pour l'une des populations étudiées (Hora), la région du déterminisme du sexe semble avoir changé. Différents scénarios pourraient expliquer un changement de déterminant et parmi eux, nous avons particulièrement évoqué la possibilité que la disparition de LG23-Y soit simplement due à la dérive génétique. Comme le lac Hora est particulièrement petit (à peine plus de 1 km²) par rapport aux autres étudiés, et qu'il y a certainement eu un repeuplement de tilapia du Nil au début du vingtième siècle avec des individus d'*O. niloticus cancellatus* endémique d'Ethiopie dans ce lac (Bezault, 2005), on peut penser que cette population a une petite taille efficace. Les goulots d'étranglement des individus fondateurs et introduits, ainsi que ceux causés par la superficie du lac rendent la population plus sensible à la stochasticité et ont pu conduire à une perte du déterminant majeur du sexe par dérive. Ce phénomène aurait pu être d'autant plus important si parmi les individus introduits se trouvaient des mâles XX contribuant à produire des descendance

100 % XX. Si nous privilégions cette hypothèse, nous ne pouvons en revanche pas exclure les autres à ce stade. Il est possible également que cette population ait accumulé des locus faiblement délétères près de la région du sexe qui apparaît ne pas recombiner avec le X sur quelques kb. Pour les mêmes raisons de taille efficace et d'effet de la dérive génétique, il est possible que le Y ait accumulé rapidement des allèles (faiblement) délétères, entraînant la contre-sélection de cette région. Nos analyses n'ont pas permis d'identifier de locus qui endosserait le rôle de déterminant, ni plus largement de groupe de liaison, en association avec le sexe dans la population Hora. Il semble peu probable que le déterminant du sexe (la région de *l'amh*) se soit déplacé sur une nouvelle paire de chromosomes puisque l'assemblage *de novo* des lectures longues a construit exactement l'haplotype X et les locus voisins attendus d'après le génome de référence. Cependant, s'il s'agit d'une translocation d'une région plus large que 50kb autour de *l'amh* alors cette méthode n'aurait pas été en mesure de la détecter. De même, les analyses utilisées sur les génomes complets des pools ne permettent pas d'identifier efficacement les systèmes polygéniques. Ainsi, si plusieurs locus interagissent pour définir le sexe suite à la perte du rôle de LG23-Y, nous ne serions pas en mesure de détecter de tels signaux. Des systèmes polygéniques ont déjà été mis en évidence chez les vertébrés, incluant des cichlidés des grands lacs Africains (Vandeputte et al., 2007; Liew et al., 2012; Moore and Roberts, 2013; Roberts et al., 2016; Clark and Kocher, 2019). Une difficulté réside alors dans le choix de la méthode à utiliser pour détecter un système polygénique car les analyses classiquement utilisées ne font souvent ressortir que le facteur majeur ou attendu (Moore and Roberts, 2013; Wellenreuther and Hansson, 2016). Dans notre cas, il serait possible d'analyser les génomes individuellement pour rechercher des déséquilibres de liaison récurrents entre certains allèles et une association avec les sexes phénotypiques. Ces données sont déjà disponibles puisque les pools par sexe des individus Hora ont été faits par informatique, après production des génomes complets individuels. Une approche intéressante pourrait être le « polygenic score » qui permet de déterminer les associations entre des combinaisons de SNPs et un trait phénotypique, habituellement utilisé en médecine pour détecter le risque de développer une maladie particulière d'après les mutations dans le génome (Choi et al., 2020; Khera et al., 2019).

Nous savons également que l'expression précoce de *gsdf* (gonadal soma-derived factor, membre de la super-famille des TGF- β) joue certainement un rôle dans le développement des gonades chez cette population, comme cela a déjà pu être reporté chez d'autres souches de

tilapia du Nil (Kaneko et al., 2015; Jiang et al., 2016). Si nous sommes en mesure d'identifier des locus spécifiques ou au moins en fort déséquilibre de liaison avec le sexe, nous pouvons imaginer génotyper les individus utilisés dans les banques de RNAseq et s'assurer que les groupes identifiés correspondent bien aux mâles et aux femelles au moins en température contrôle. On pourrait alors exploiter beaucoup plus en profondeur les données de ce séquençage et en déduire plus finement la cascade précocement impliquée dans la différenciation du sexe dans la famille analysée.

D'autre part, il existe des méthodes telles que Sex-DEtector qui permettent de détecter des chromosomes sexuels (particulièrement adapté aux chromosomes sexuels encore peu divergents) à partir de séquençage d'ARN en se basant sur l'analyse de familles avec idéalement les 2 géniteurs et au moins 5 individus de chaque sexe issus de la descendance (Muyle et al., 2016). En étudiant les génotypes de la descendance et ceux des parents, les méthodes probabilistes de ce logiciel permettent de grouper des contigs en fonction de leur type de ségrégation (XY ou ZW dans notre cas par exemple). On pourrait envisager d'utiliser cette méthode également pour identifier les chromosomes sexuels dans la population Hora puisque des croisements ont été produits et que certains individus ont été maintenus jusqu'à un âge suffisamment avancé pour permettre le sexage par observation des gonades. Ainsi, des banques de RNAseq sur des adultes Hora, de la même famille que celle utilisée dans le Chapitre 3, sont en cours de séquençage.

Enfin, en lien avec la thermosensibilité, il serait intéressant d'étudier les modifications épigénétiques puisque celles-ci peuvent entraîner des régulations d'expression, par des mécanismes tels que l'épissage alternatif. Par exemple ces mécanismes sont impliqués dans la réversion du sexe chez le poisson plat *Cynoglossus semilaevis* (Shao et al., 2014) et chez l'Agame barbu (*Pogona vitticeps*) (Deveson et al., 2017).

Les limites de nos études

Un focus sur certaines régions et certains gènes

Une limite récurrente dans nos analyses est l'utilisation de sous-ensembles de régions précises du génome. Dans le cas des séquençages ddRAD, le concept même de la méthode consiste à cibler uniquement les régions du génome comprises entre deux sites d'enzymes de restrictions et dont l'insert entre les deux sites est de 300 à 700bp environ. Cette approche

permet une exploration d'un grand nombre d'individus simultanément, à un coût réduit, puisque peu de sites sont séquencés contrairement à un génome complet. En revanche, les assemblages obtenus sont un peu à l'image d'un texte à trous et si l'information recherchée se situe dans une région non séquencée, alors ce type d'analyses ne permet pas de répondre complètement aux questions posées. De plus, en se focalisant sur des sites de restrictions, on prend le risque de séquencer de manière non aléatoire certaines familles multigéniques qui présenteraient ces sites spécifiques. Il est possible que nous ayons rencontré ce biais à la fois pour la famille Kpandu (Ghana) et pour la famille Kou (Burkina-Faso) analysées. La présence de polymorphisme entre les différents gènes d'une même famille crée artificiellement des régions particulièrement polymorphiques pouvant induire un signal artefactuel.

Lors des analyses de « long reads », une fois encore nous nous sommes focalisés sur une partie précise du génome. Parce que les analyses précédentes ont suggéré que la région de la duplication de *l'amh* sur LG23-Y était le principal gène déterminant le sexe sur les populations naturelles étudiées, nous avons sélectionné un sous-ensemble de lectures correspondant à ces gènes spécifiques parmi l'ensemble de données brutes afin de gagner du temps lors des assemblages en réduisant la taille de la région à assembler.

Ces approches peuvent se comprendre puisque le même ensemble de gènes « joue souvent aux chaises musicales » pour endosser le rôle de déterminant du sexe chez les poissons téléostéens (Böhne et al., 2016; Pan et al., 2016; Ortega-Recalde et al., 2020; Tao et al., 2021), justifiant que certains suspects « habituels » soient recherchés en priorité. Toutefois, certains gènes *a priori* inattendus peuvent être impliqués dans le déterminisme du sexe. C'est le cas chez la truite arc-en-ciel pour laquelle une néo-fonctionnalisation d'un gène immunitaire à la suite d'une duplication a conduit à en faire le gène majeur du déterminisme du sexe (Yano et al., 2012). Des systèmes polygéniques moins évidents peuvent aussi ne pas être détectés (Moore and Roberts, 2013; Wellenreuther and Hansson, 2016). En ne regardant que les gènes candidats et les suspects habituels, nous nous sommes concentrés sur les gènes attendus de la voie du sexe mais avons peut-être ignorés d'autres gènes qui pourraient jouer un rôle dans le déterminisme du sexe.

Identification d'un chromosome B : une difficulté supplémentaire

Dans le cas des analyses de génomes complets des individus sauvages des populations Kpandu (Ghana) et Koka (Ethiopie), le soupçon de la présence d'un chromosome B nous a

conduit à nous focaliser uniquement sur une petite région du génome, identifiée préalablement par les analyses de ddRAD. Or, comme nous avons utilisé des familles pour le ddRAD, il est possible que, si plusieurs locus du déterminisme du sexe ségrègent dans les populations, nous ayons échantillonné une famille portant le déterminant majoritaire et que nous passions à côté de déterminant qui ségrégerait en faible fréquences dans les populations.

Cette approche nous empêche donc de tirer profit de l'étude de l'ensemble des génomes complets d'individus du milieu naturel qui auraient pu contenir cette information. Idéalement, il faudrait être en mesure soit : i) d'identifier quels individus sont porteurs du chromosome B dans chacun des pools et séquencer à nouveaux en séparant ces individus des autres ; ii) de caractériser la vaste majorité des régions appartenant au chromosome B et les retirer du jeu de données pour pouvoir avoir une approche globale du génome.

i) La mise en place d'amorces PCR pour détecter les individus porteurs de chromosome(s) B a été un travail important dans ce projet de thèse afin de pouvoir identifier les individus porteurs dans les pools. Cependant, il est particulièrement compliqué de concevoir des amorces PCR dans des régions de chromosome B sans en avoir la séquence exacte. Le design des amorces a été effectué sur la base de la séquence consensus de ces régions de forte couverture en ciblant les zones riches en SNPs spécifiques au pool contenant le B. Seulement, une région de forte couverture indique en réalité que de nombreuses copies de ce gène existent sur le chromosome B et il est difficile de savoir si elles possèdent toutes exactement le même enchainement de SNPs. Ainsi, en dessinant des amorces sur le consensus, on dessine peut-être des amorces sur la base d'un hybride de 10 copies de ce locus qui ne possèdent pas toutes les mêmes SNPs. De plus, l'utilisation d'un mapping sur un génome de référence ne nous permet pas de voir avec certitude l'agencement des régions spécifiques B les unes par rapports aux autres. Il est possible que nous ayons essayé d'amplifier des régions dont nous estimions la taille à 300bp qui s'avéraient être bien au-delà de la taille raisonnable amplifiable en PCR. D'autres amorces peuvent être dessinées, mais cette approche restera problématique tant que nous ne connaissons pas exactement la séquence du chromosome B.

ii) Caractériser un chromosome B et définir précisément son contenu est un travail fastidieux qui peut être conduit de deux manières : directe ou indirecte (revue dans Ruban et

al., 2017). Cela passe idéalement par le séquençage du chromosome B préalablement isolé des autres chromosomes par exemple par micro-dissection ou par « flow sorting » (tri de cellules et de chromosomes). Ce type d'analyse cytogénétique permet d'une part de confirmer la présence d'un tel élément génétique et d'autre part d'obtenir un assemblage du chromosome B et donc son contenu en gènes (Vij et al., 2016). Cependant, la microdissection nécessite d'avoir des chromosomes en métaphase bien étalés et, de plus, est prône aux contaminations par les autres chromosomes (Valente et al., 2014 ; D'Cotta et al., 2014). Dans notre cas, aucun individu vivant n'était disponible pour procéder à de tels analyses. Néanmoins, nous avons des suspensions chromosomiques de certains individus sauvages qui pourrait éventuellement être étalées s'ils s'avèrent porteur d'un chromosome B. Cependant, il existe également des méthodes indirectes pour caractériser le contenu d'un chromosome B lorsque que les approches cytogénétiques ne sont pas envisageables. À partir de séquençages de génomes complets d'individus portant le B et d'individus sans B, il est possible d'identifier des régions appartenant potentiellement au chromosome B. Dans les analyses complémentaires du Chapitre 1, nous avons utilisé une de ces méthodes indirectes, basée sur les ratios de couvertures entre individus présentant le B ou non (Valente et al., 2014; Fantinatti & Martins, 2016; Clark et al., 2018). Il est aussi possible de se baser sur les ratios des fréquences de *k*-mer (Kurtz et al., 2008; Marçais & Kingsford, 2011; Navarro-Domínguez et al., 2019) ou encore sur le clustering des lectures d'après leur similarités (un peu à la manière d'un assembleur à graphe) (Novák et al., 2010). La méthode de clustering par graphe contrairement aux deux autres ne nécessite pas de génome de référence. Une utilisation de différentes méthodes en parallèle est préférable puisque cela permet d'assoir l'identification de certaines régions si elles sont retrouvées avec plusieurs approches. Les jeux de données dont nous disposons actuellement répondent aux prérequis nécessaires pour pouvoir caractériser ce B et font parties des ambitions de la suite de ce projet de thèse.

Enfin, les chromosomes B chez les cichlidés ne sont pas forcément présents (ni en même nombre de copies) chez tous les individus d'une population ou d'une population à l'autre (Poletto et al., 2010; Perazzo et al., 2018; Clark & Kocher, 2019). Si à terme il est possible d'identifier ce chromosome B facilement (comme avec des marqueurs PCR) dans l'ensemble des échantillons encore disponibles au laboratoire on pourrait estimer la fréquence à laquelle celui-ci ségrège dans les populations naturelles et si on le retrouve dans les deux sexes dans les populations Koka et Kpandu. En effet, nous avons soupçonné un chromosome B dans le

pool mâles de la population Kpandu et dans le pool femelles de la population Koka, ce qui laisse penser qu'il n'agit pas comme déterminant du sexe à l'échelle de l'espèce. Cependant, il n'est pas impossible que dans l'une des populations il agisse comme un déterminant du sexe comme cela a déjà été vu chez *Metriaclima lombardoi* (Clark and Kocher, 2019). L'acquisition de ces informations repose donc sur les deux points (i) et (ii) explicités précédemment.

L'Hybridation : une alternative pour expliquer nos résultats ?

Les espèces du genre *Oreochromis* ont une grande capacité d'hybridation interspécifique et les descendances produites sont, pour la plupart des combinaisons, viables et fertiles (Hickling, 1960; Jalabert et al., 1971; revue dans Baroiller and D'Cotta, 2018). Pour l'aquaculture, de nombreux croisements ont été effectués pour profiter entre autre de la vigueur hybride et/ou obtenir des souches piscicoles ayant les traits les plus favorables de deux espèces différentes (résistances aux maladies, à la salinité, croissance rapide etc...) (Yehya et al., 2012; de Verdal et al., 2014; Rahman et al., 2019; Nzohabonayo et al., 2021). Ces croisements entre espèces proches ont également été utilisés pour étudier le déterminisme du sexe et rechercher les paires de chromosomes sexuels chez les cichlidés du genre *Oreochromis* (Jalabert et al., 1971; Cnaani et al., 2004). De plus, les systèmes hétérogamétiques différents ne sont pas une barrière à ces croisements et permettent entre autre de produire des souches monosexes mâles en pisciculture (Hickling, 1960; Baroiller and D'Cotta, 2018).

Les cichlidés utilisés en aquaculture, particulièrement des genres *Tilapia*, *Sarotherodon* et *Oreochromis*, ont fait l'objet de nombreux transferts par la main de l'homme en Afrique durant la seconde moitié du XXe siècle (revue dans Bezault, 2005). Comme le tilapia du Nil (*O. niloticus*) représente la très grande majorité de la production piscicole parmi ces différentes espèces, ses transferts ont été particulièrement importants à la fois au sein de son aire de répartition mais également en dehors de celle-ci (Canonico et al., 2005; D'Amato et al., 2007; Firmat et al., 2013; Deines et al., 2014). La souche d'*O. niloticus* Bouaké (aussi appelée « Souche Côte d'Ivoire ») est une souche synthétique issue de deux populations d'Afrique de l'Ouest qui a par exemple été expédiée dans différents pays d'Afrique (Burkina Faso, Sénégal, Bénin...) mais aussi en France, aux Etats-Unis et en Amérique du Sud, au Brésil (Bezault, 2005). Un autre exemple marquant est l'utilisation commerciale de la souche GIFT (Genetically Improved Farmed Tilapia) qui a été très largement distribuée à travers le monde

et qui est une souche synthétique produite à partir de quatre populations provenant d'Égypte, du Ghana, du Kenya et du Sénégal et de quatre souches d'aquaculture provenant d'Israël, de Singapour, de Taiwan et de Thaïlande (Eknath et al., 1998).

Dans le milieu naturel, ces introductions ont pu mener à de l'hybridation avec les espèces endémiques comme cela a été le cas dans le lac Naivasha (Kenya) entre *O. niloticus* et *O. variabilis* et dans le lac Itasy (Madagascar) entre *O. macrochir* et *O. niloticus* (tous deux introduits) (Lévêque and Paugy, 2006). Des études ont également suggéré l'existence d'hybridations entre *O. aureus* et *O. niloticus* suite à l'observation d'introgression de génome mitochondrial dans les populations d'Afrique de l'Ouest (Rognon and Guyomard, 2003) ainsi que dans le génome mitochondrial de *O. leucosticus* vers *O. niloticus* au Kenya dans le lac Baringo (Nyingi and Agnèse, 2007). Des hybridations ont également été rapportées dans le sud de l'Afrique entre *O. mossambicus* (endémique) et *O. niloticus* (introduit) (D'Amato et al., 2007; Firmat et al., 2013).

En sachant que les aires de répartitions de certaines de ces espèces peuvent se recouper, comme *O. niloticus* et *O. aureus* en Afrique de l'Ouest (Rognon and Guyomard, 2003) (Figure 5.1) et que des hybridations/introgressions ont lieu dans le milieu naturel, il est légitime de se demander si les populations que nous avons étudiées ont pu être confrontées à de tels événements. Par exemple, il est surprenant de retrouver la présence de chromosomes B et le déterminisme du sexe par *l'amh* sur LG23-Y dans les populations Kpandu et Koka alors que celles-ci sont éloignées géographiquement, et que Koka est phylogénétiquement plus proche de Hora. Cependant, les analyses phylogénétiques nucléaires et mitochondriales ne semblent pas soutenir l'hypothèse d'hybridations. En effet, si des croisements avaient eu lieu entre des individus de Kpandu et Koka, on pourrait s'attendre à ce que ces populations s'en retrouvent rapprochées dans la phylogénie nucléaire. A contrario, on observe une plus grande proximité entre Koka et Hora dans l'ensemble des analyses phylogénétiques. Cependant, dans le cas où l'évènement d'introgression serait suffisamment lointain pour que la majorité du génome ait subi un balayage sélectif ou l'effet de la dérive génétique, alors nous ne pourrions pas forcément observer de conséquences de cette introgression en se basant seulement sur l'analyse de 42 marqueurs nucléaires (Figure 5.2) (Folk et al., 2018). Tout de même, l'analyse de la région de *l'amh* sur LG23 montre que les haplotypes X sont plus similaires entre Koka et Hora qu'entre Koka et Kpandu.

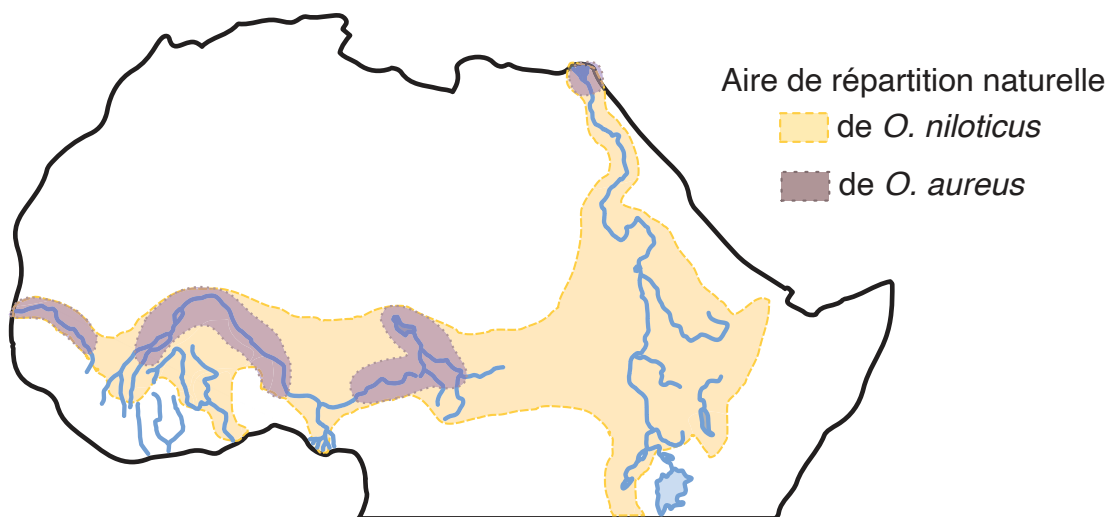


Figure 5.1. Aire de répartition géographique se chevauchant pour deux espèces d'*Oreochromis*.
En jaune : aire de répartition de *O. niloticus* ; En mauve : aire de répartition géographique d'*O. aureus*. Ces deux espèces ont des aires de répartition chevauchantes à l'Ouest de l'Afrique.

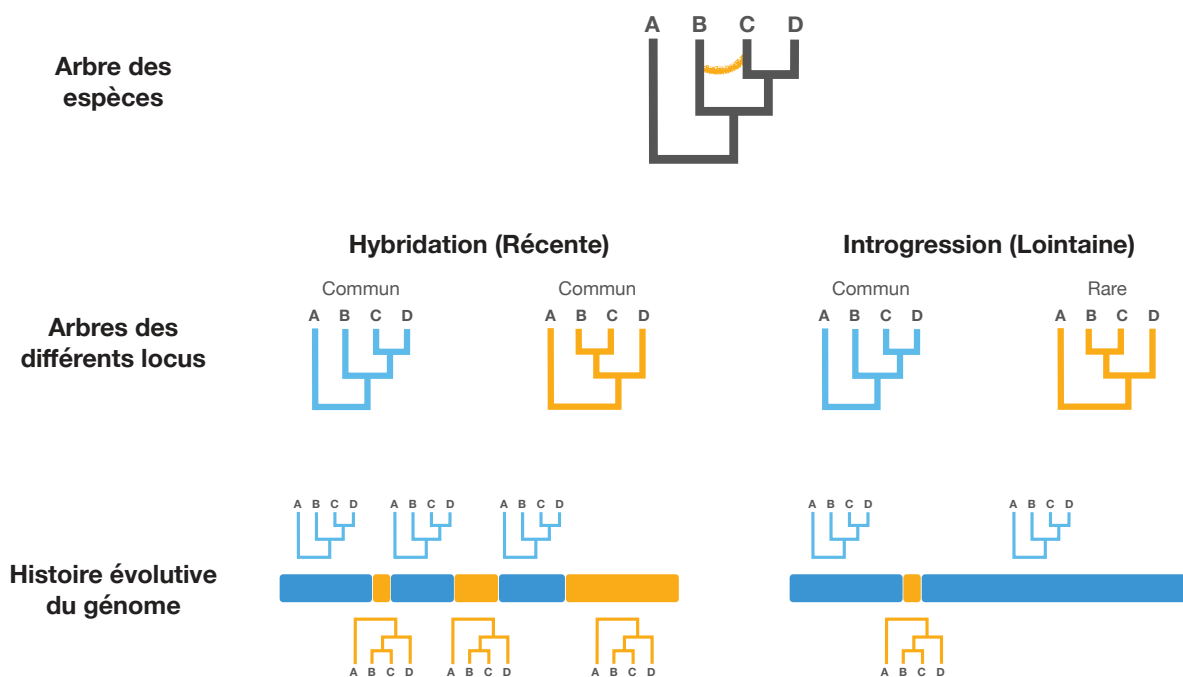


Figure 5.2. Proportion du génome introgressé en fonction du temps passé depuis l'hybridation.

L'arbre des espèces représente les relations de parentés entre espèces étudiées et un évènement d'hybridation entre les espèces B et C est illustré par un arc orange. Les portions non introgressées du génome suivent l'arbre des espèces (arbre bleu) alors que dans les régions introgressées les séquences des espèces B et C sont très proches (arbre orange). Plus l'évènement d'hybridation est ancien, plus les locus témoins sont susceptibles d'être perdu par dérive génétique ou balayage sélectif. Ils deviennent donc progressivement plus rares dans le génome. La détection de l'hybridation sur la base d'une phylogénie de peu de marqueurs devient donc moins probable. Modifié d'après Folk et al., (2018).

Si les deux haplotypes ont cessé de recombiner sur cette région, il n'y a pas de raisons qu'après l'introggression il ne reste pas de traces dans les haplotypes X de Kpandu et Koka, ni que les populations éthiopiennes aient le schéma d'une histoire évolutive commune avec des SNPs partagés dans cette région.

Il serait également envisageable que la population Hora ne présente plus l'*amh* comme déterminant du sexe sur LG23-Y suite à une hybridation qui aurait entraîné la colonisation de cette région par l'haplotype de l'espèce introgressante. Cependant, les relations de parenté décrites par les analyses phylogénétiques montrent Koka et Hora en groupe monophylétique et indiquent que ces espèces sont peu divergentes en comparaison de Kpandu ou de l'individu séquencé pour produire le génome de référence par exemple. Ici encore, si l'hybridation est suffisamment lointaine, la majeure partie du génome « introgressant » a pu être nettoyé et nos analyses phylogénétiques pourraient ne pas en rendre compte. De plus, des SNPs spécifiques à Koka et Hora sur l'haplotype X semblent décrire une histoire évolutive commune de ces régions et ne supportent pas à première vue l'hypothèse d'une introgression. Enfin, les espèces sympatriques de *O. niloticus* dans le lac Hora sont essentiellement des poissons-chats (*Clarias sp.*) ainsi que de Cyprinodontidae (*Paropanchax sp.*) qui ne pourraient pas faire l'objets de flux de gènes avec le tilapia du Nil et il n'y a pas non plus de traces de transferts d'autres espèces de tilapia en Ethiopie qui soit en mesure de s'hybrider avec le tilapia du Nil (Bezault, 2005). L'ensemble de ces indices nous laisse donc penser que l'introggression de LG23-Y par une autre espèce n'est probablement pas la raison de l'absence de la région du déterminisme du sexe chez cette population.

Déterminer l'état ancestral du déterminisme du sexe chez le tilapia du Nil pourrait nous aider à trancher entre les hypothèses d'une acquisition de la duplication de l'*amh* sur LG23-Y par flux de gènes entre Koka et Kpandu (probablement par la main de l'homme dans ce cas) ou bien d'une perte secondaire de cette structure chez Hora. Des améliorations de nos analyses phylogénétiques pourraient par exemple aider à éclaircir les différents scénarios décrits ci-dessus. D'une part, nous pourrions intégrer un plus grand nombre de locus, qui idéalement possèderaient des caractéristiques variées d'évolution des séquences, afin de mieux retracer l'histoire évolutive des génomes et des populations. D'autre part, il semble essentiel d'ajouter des populations supplémentaires pour pouvoir retracer au mieux l'histoire évolutive de la paire de chromosome sexuel LG23. Nous pourrions également

envisager d'utiliser des méthodes de coalescence dans ces nouvelles analyses phylogénétiques (Hibbins and Hahn, 2019). Cela permettrait notamment de limiter les biais induits par la concaténation, qui présuppose une histoire évolutive commune des différents locus analysés (Gatesy et al., 2017). Enfin, l'ensemble de ces améliorations permettrait d'obtenir des données solides pour comprendre l'évolution des paires de chromosomes sexuels si LG23 est une « vieille » paire de chromosomes à l'échelle de l'espèce, avec le X et le Y peu divergents. Malheureusement, nos données ne permettent pas en l'état d'effectuer ces comparaisons et un re-séquençage des lectures longues pour augmenter la couverture sera nécessaire pour reconstruire les haplotypes Y avec fiabilité. Les lectures courtes pourront ensuite aider à corriger finement ces haplotypes Y reconstruits.

Conclusion

Ce manuscrit de thèse représente le premier complexe d'études de génétique et génomique sur le déterminisme du sexe chez le tilapia du Nil en populations naturelles. Malgré les difficultés techniques que peuvent entraîner l'étude des chromosomes sexuels, en couplant différentes approches méthodologiques nous avons mis en évidence un polymorphisme du déterminisme du sexe dans trois populations naturelles de cette espèce. Ce résultat met en lumière un patron d'évolution complexe de ce système que nous pensons être lié à la perte du déterminisme ancestral dans l'une des populations étudiées. Ces travaux montrent donc l'intérêt d'analyser le déterminisme du sexe à une échelle populationnelle pour estimer l'importance de la diversité qui peut exister en milieu naturel. De plus, étudier le déterminisme du sexe au niveau intraspécifique chez des espèces adaptées à des milieux variés représente une excellente opportunité pour comprendre les mécanismes à l'origine de l'évolution des chromosomes sexuels. Nous sommes également parvenus à détailler une région du déterminisme du sexe dont l'architecture complexe (et variable entre populations) n'avait jusqu'ici pas été décrite. En particulier nous avons proposé pour la première fois un haplotype du chromosome Y dans cette région. Ces analyses représentent un socle sur lequel il sera possible de s'appuyer pour faciliter la reconstruction des chromosomes sexuels lors de l'assemblage de génome de tilapia. Cependant, la manière exacte, dont cet haplotype Y agit pour déterminer le sexe et notamment le rôle des paralogues de *amh*, reste à définir. La question du mécanisme du déterminisme du sexe reste également ouverte chez la population

ne présentant pas la région ancestrale sur le Y et des analyses complémentaires à la fois de génomique et de transcriptomique seront nécessaires. Enfin, ces travaux suggèrent aussi pour la première fois l'existence d'un chromosome B (chromosome surnuméraire, non essentiel et au comportement égoïste) qui ségrégerait dans certaines populations naturelles de tilapia du Nil.

Références

- Baroiller, J.-F., and D'Cotta, H. (2018). Sex Control in Tilapias. In Sex Control in Aquaculture, (John Wiley & Sons, Ltd), pp. 189–234.
- Baroiller, J.F., D'Cotta, H., Bezault, E., Wessels, S., and Hoerstgen-Schwark, G. (2009). Tilapia sex determination: where temperature and genetics meet. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 153, 30–38.
- Bezault, E. (2005). Etude Du Systeme De Determinisme Du Sexe au Sein De Populations Naturelles De Tilapia Du Nil, *Oreochromis niloticus* (linnaeus, 1758) : importance Des Composantes Genetiques Et Environnementales. UNIVERSITE DE PARIS-SUD.
- Böhne, A., Wilson, C.A., Postlethwait, J.H., and Salzburger, W. (2016). Variations on a theme: Genomics of sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 17.
- Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Travisany, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., et al. (2019). Fine mapping using whole-genome sequencing confirms anti-müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3 (Bethesda)* 9, 3213–3223.
- Canonica, G.C., Arthington, A., McCrary, J.K., and Thieme, M.L. (2005). The effects of introduced tilapias on native biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems* 15, 463–483.
- Choi, S.W., Mak, T.S.-H., and O'Reilly, P.F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols* 15, 2759–2772.
- Clark, F.E., and Kocher, T.D. (2019). Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. *Scientific Reports* 9, 1–10.
- Clark, F.E., Conte, M.A., and Kocher, T.D. (2018). Genomic characterization of a B chromosome in Lake Malawi cichlid fishes. *Genes* 9, 610.
- Cnaani, A., Zilberman, N., Tinman, S., Hulata, G., and Ron, M. (2004). Genome-scan analysis for quantitative trait loci in an F2 tilapia hybrid. *Mol Genet Genomics* 272, 162–172.
- Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., D'Hont, A., Baroiller, J.-F., D'Cotta, H., Penman, D.J., et al. (2008). Genetics of sex determination in tilapiine species. *Sexual Development* 2, 43–54.
- Curzon, A.Y., Shirak, A., Dor, L., Zak, T., Perelberg, A., Seroussi, E., and Ron, M. (2020). A duplication of the Anti-Müllerian hormone gene is associated with genetic sex determination of different *Oreochromis niloticus* strains. *Heredity* 125, 317–327.
- D'Amato, M.E., Esterhuyse, M.M., van der Waal, B.C.W., Brink, D., and Volckaert, F.A.M. (2007). Hybridization and phylogeography of the Mozambique tilapia *Oreochromis mossambicus* in southern Africa evidenced by mitochondrial and microsatellite DNA genotyping. *Conserv Genet* 8, 475–488.
- D'Cotta H., Coutanceau J-P., Pepey E., Baaziz S.B., Coriton O., Bonillo C., Rodier-Goud M., Soler L., D'Hont A., Kocher T.D., Ozouf-Costaz C., Baroiller J-F. (2014). Fishing for sex genes by chromosome microdissection in two tilapia species with different sex

- determining systems. 21st International Colloquium on Animal Cytogenetics and Gene Mapping (ICACGM).
- Deines, A., Bhole, I., Katongo, C., Feder, J., and Lodge, D. (2014). Hybridisation between native *Oreochromis* species and introduced Nile tilapia *O. niloticus* in the Kafue River, Zambia. *African Journal of Aquatic Science* 39, 23–34.
- Deveson, I.W., Holleley, C.E., Blackburn, J., Marshall Graves, J.A., Mattick, J.S., Waters, P.D., and Georges, A. (2017). Differential intron retention in *Jumonji* chromatin modifier genes is implicated in reptile temperature-dependent sex determination. *Sci. Adv.* 3, e1700731.
- Di Genova, A., Buena-Atienza, E., Ossowski, S., and Sagot, M.-F. (2021). Efficient hybrid de novo assembly of human genomes with WENGAN. *Nature Biotechnology* 39, 422–430.
- Eknath, A., Dey, M., Rye, M., Gjerde, B., Abella, T.A., Sevilleja, R., Tayamen, M.M., Reyes, R.A., and Bentsen, H. (1998). Selective breeding of Nile tilapia for Asia.
- Eshel, O., Shirak, A., Weller, J.I., Hulata, G., and Ron, M. (2012). Linkage and physical mapping of sex Region on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3: Genes, Genomes, Genetics* 2, 35–42.
- Eshel, O., Shirak, A., Dor, L., Band, M., Zak, T., Markovich-Gordon, M., Chalifa-Caspi, V., Feldmesser, E., Weller, J.I., Seroussi, E., et al. (2014). Identification of male-specific amh duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 15, 774.
- Fantinatti, B.E.A., and Martins, C. (2016). Development of chromosomal markers based on next-generation sequencing: the B chromosome of the cichlid fish *Astatotilapia latifasciata* as a model. *BMC Genetics* 17, 119.
- Firmat, C., Alibert, P., Losseau, M., Baroiller, J.-F., and Schlieuwen, U.K. (2013). Successive Invasion-Mediated Interspecific Hybridizations and Population Structure in the Endangered Cichlid *Oreochromis mossambicus*. *PLOS ONE* 8, e63880.
- Folk, R.A., Soltis, P.S., Soltis, D.E., and Guralnick, R. (2018). New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany* 105, 364–375.
- Furman, B.L.S., Metzger, D.C.H., Darolti, I., Wright, A.E., Sandkam, B.A., Almeida, P., Shu, J.J., and Mank, J.E. (2020). Sex chromosome evolution: so many exceptions to the rules. *Genome Biol Evol* 12, 750–763.
- Gammerdinger, W.J., Conte, M.A., Acquah, E.A., Roberts, R.B., and Kocher, T.D. (2014). Structure and decay of a proto-Y region in Tilapia, *Oreochromis niloticus*. *BMC Genomics* 15, 975.
- Gatesy, J., Meredith, R.W., Janecka, J.E., Simmons, M.P., Murphy, W.J., and Springer, M.S. (2017). Resolution of a concatenation/coalescence kerfuffle: partitioned coalescence support and a robust family-level tree for Mammalia. *Cladistics* 33, 295–332.
- Haghshenas, E., Asghari, H., Stoye, J., Chauve, C., and Hach, F. (2020). HASLR: Fast Hybrid Assembly of Long Reads. *IScience* 23, 101389.
- Hibbins, M.S., and Hahn, M.W. (2019). The Timing and Direction of Introgression Under the Multispecies Network Coalescent. *Genetics* 211, 1059–1073.

- Hickling, C.F. (1960). The Malacca tilapia hybrids. *J Genet* 57, 1–10.
- Jalabert, B., Kammacher, P., and Lessent, P. (1971). Déterminisme du sexe chez les hybrides entre tilapia macrochir et tilapia nilotica. étude de la sex-ratio dans les croisements des hybrides de première génération par les espèces parentes. *Annales de Biologie Animale, Biochimie, Biophysique* 11, 155–165.
- Jiang, D., Chen, J., Fan, Z., Tan, D., Zhao, J., Shi, H., Liu, Z., Tao, W., Li, M., and Wang, D. (2017). CRISPR/Cas9-induced disruption of *wt1a* and *wt1b* reveals their different roles in kidney and gonad development in Nile tilapia. *Developmental Biology* 428, 63–73.
- Jiang, D.-N., Yang, H.-H., Li, M.-H., Shi, H.-J., Zhang, X.-B., and Wang, D.-S. (2016). *gsdf* is a downstream gene of *dmrt1* that functions in the male sex determination pathway of the Nile tilapia. *Molecular Reproduction and Development* 83, 497–508.
- Kaneko, H., Ijiri, S., Kobayashi, T., Izumi, H., Kuramochi, Y., Wang, D.-S., Mizuno, S., and Nagahama, Y. (2015). Gonadal soma-derived factor (*gsdf*), a TGF-beta superfamily gene, induces testis differentiation in the teleost fish *Oreochromis niloticus*. *Molecular and Cellular Endocrinology* 415, 87–99.
- Khera, A.V., Chaffin, M., Zekavat, S.M., Collins, R.L., Carolina, R., Natarajan, P., Lichtman, J.H., D'Onofrio, G., Mattera, J., Dreyer, R., et al. (2019). Whole-Genome Sequencing to Characterize Monogenic and Polygenic Contributions in Patients Hospitalized With Early-Onset Myocardial Infarction. *Circulation* 139, 1593–1602.
- Kurtz, S., Narechania, A., Stein, J.C., and Ware, D. (2008). A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. *BMC Genomics* 9, 517.
- Lee, B.-Y., Penman, D.J., and Kocher, T.D. (2003). Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics* 34, 379–383.
- Lee, B.-Y., Coutanceau, J.-P., Ozouf-Costaz, C., D'Cotta, H., Baroiller, J.-F., and Kocher, T.D. (2011). Genetic and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus*). *Mar Biotechnol (NY)* 13, 557–562.
- Lévêque, C., and Paugy, D. (2006). Les poissons des eaux continentales africaines: diversité, écologie, utilisation par l'homme (Paris: IRD éd.).
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genetics* 11, e1005678.
- Li, M., Liu, X., Dai, S., Xiao, H., and Wang, D. (2019). High Efficiency Targeting of Non-coding Sequences Using CRISPR/Cas9 System in Tilapia. *G3 Genes | Genomes | Genetics* 9, 287–295.
- Li, M., Dai, S., Liu, X., Xiao, H., and Wang, D. (2020). A detailed procedure for CRISPR/Cas9-mediated gene editing in tilapia. *Hydrobiologia*.
- Liew, W.C., Bartfai, R., Lim, Z., Sreenivasan, R., Siegfried, K.R., and Orban, L. (2012). Polygenic Sex Determination System in Zebrafish. *PLoS ONE* 7, e34397.

- Liu, X., Xiao, H., Jie, M., Dai, S., Wu, X., Li, M., and Wang, D. (2020). Amh regulate female folliculogenesis and fertility in a dose-dependent manner through Amhr2 in Nile tilapia. *Molecular and Cellular Endocrinology* 499, 110593.
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770.
- Miclotte, G., Heydari, M., Demeester, P., Rombauts, S., Van de Peer, Y., Audenaert, P., and Fostier, J. (2016). Jabba: hybrid error correction for long sequencing reads. *Algorithms Mol Biol* 11, 10.
- Moore, E.C., and Roberts, R.B. (2013). Polygenic sex determination. *Current Biology* 23, R510–R512.
- Muyle, A., Käfer, J., Zemp, N., Mousset, S., Picard, F., and Marais, G.A. (2016). SEX-DETECTOR: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biol Evol* 8, 2530–2543.
- Navarro-Domínguez, B., Martín-Peciña, M., Ruiz-Ruano, F.J., Cabrero, J., Corral, J.M., López-León, M.D., Sharbel, T.F., and Camacho, J.P.M. (2019). Gene expression changes elicited by a parasitic B chromosome in the grasshopper *Eyprepocnemis plorans* are consistent with its phenotypic effects. *Chromosoma* 128, 53–67.
- Novák, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11, 378.
- Nyingi, D.W., and Agnèse, J.-F. (2007). Recent introgressive hybridization revealed by exclusive mtDNA transfer from *Oreochromis leucostictus* (Trewavas, 1933) to *Oreochromis niloticus* (Linnaeus, 1758) in Lake Baringo, Kenya. *Journal of Fish Biology* 70, 148–154.
- Nzohabonayo, E., Manyala, J., Kang'ombe, J., and Kassam, D. (2021). Effect of hybridization on reproductive performance of *Oreochromis karongae*, *Oreochromis shiranus* and *Oreochromis mossambicus*. *Aquac. Res.* 52, 302–308.
- Ortega-Recalde, O., Goikoetxea, A., Hore, T.A., Todd, E.V., and Gemmell, N.J. (2020). The Genetics and Epigenetics of Sex Change in Fish. *Annual Review of Animal Biosciences* 8, null.
- Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2013). Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS ONE* 8, e68389.
- Palaiokostas, C., Bekaert, M., Khan, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., and Penman, D.J. (2015). A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 16, 171.
- Pan, Q., Anderson, J., Bertho, S., Herpin, A., Wilson, C., Postlethwait, J.H., Scharl, M., and Guiguen, Y. (2016). Vertebrate sex-determining genes play musical chairs. *Comptes Rendus Biologies* 339, 258–262.
- Perazzo, G.X., Noleto, R.B., Vicari, M.R., Gava, A., and Cestari, M.M. (2018). B chromosome polymorphism in South American cichlid. *Neotropical Biodiversity* 4, 3–9.

- Poletto, A.B., Ferreira, I.A., Cabral-de-Mello, D.C., Nakajima, R.T., Mazzuchelli, J., Ribeiro, H.B., Venere, P.C., Nirchio, M., Kocher, T.D., and Martins, C. (2010). Chromosome differentiation patterns during cichlid fish evolution. *BMC Genet* 11, 50.
- Rahman, M.L., Zahangir, Md.M., Kitahashi, T., Shahjahan, Md., and Ando, H. (2019). Effects of high and low temperature on expression of GnIH, GnIH receptor, GH and PRL genes in the male grass puffer during breeding season. *General and Comparative Endocrinology* 282, 113200.
- Reichwald, K., Petzold, A., Koch, P., Downie, B.R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., et al. (2015). Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell* 163, 1527–1538.
- Roberts, N.B., Juntti, S.A., Coyle, K.P., Dumont, B.L., Stanley, M.K., Ryan, A.Q., Fernald, R.D., and Roberts, R.B. (2016). Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 17, 835.
- Rognon, X., and Guyomard, R. (2003). Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. *Mol Ecol* 12, 435–445.
- Ruban, A., Schmutzer, T., Scholz, U., and Houben, A. (2017). How Next-Generation Sequencing Has Aided Our Understanding of the Sequence Composition and Origin of B Chromosomes. *Genes* 8, 294.
- Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514.
- Shao, C., Li, Q., Chen, S., Zhang, P., Lian, J., Hu, Q., Sun, B., Jin, L., Liu, S., Wang, Z., et al. (2014). Epigenetic modification and inheritance in sexual reversal of fish. *Genome Res* 24, 604–615.
- Sun, Y., Svedberg, J., Hiltunen, M., Corcoran, P., and Johannesson, H. (2017). Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nature Communications* 8, 1140.
- Tao, W., Conte, M.A., Wang, D., and Kocher, T.D. (2021). Network architecture and sex chromosome turnovers: Do epistatic interactions shape patterns of sex chromosome replacement? *BioEssays* 43, 2000161.
- Taslima, K., Wehner, S., Taggart, J.B., de Verdal, H., Benzie, J.A.H., Bekaert, M., McAndrew, B.J., and Penman, D.J. (2020). Sex determination in the GIFT strain of tilapia is controlled by a locus in linkage group 23. *BMC Genet* 21, 49.
- Taslima, K., Khan, M.G.Q., McAndrew, B.J., and Penman, D.J. (2021). Evidence of two XX/XY sex-determining loci in the Stirling stock of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 532, 735995.
- Valente, G.T., Conte, M.A., Fantinatti, B.E.A., Cabral-de-Mello, D.C., Carvalho, R.F., Vicari, M.R., Kocher, T.D., and Martins, C. (2014). Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Mol Biol Evol* 31, 2061–2072.

- Vandeputte, M., Dupont-Nivet, M., Chavanne, H., and Chatain, B. (2007). A polygenic hypothesis for sex determination in the european sea bass *Dicentrarchus labrax*. *Genetics* 176, 1049–1057.
- de Verdal, H., Rosario, W., Vandeputte, M., Muyalde, N., Morissens, P., Baroiller, J.-F., and Chevassus, B. (2014). Response to selection for growth in an interspecific hybrid between *Oreochromis mossambicus* and *O. niloticus* in two distinct environments. *Aquaculture* 430, 159–165.
- Vij, S., Kuhl, H., Kuznetsova, I.S., Komissarov, A., Yurchenko, A.A., Van Heusden, P., Singh, S., Thevasagayam, N.M., Prakki, S.R.S., Purushothaman, K., et al. (2016). Chromosomal-Level Assembly of the Asian Seabass Genome Using Long Sequence Reads and Multi-layered Scaffolding. *PLoS Genet* 12, e1005954.
- Wellenreuther, M., and Hansson, B. (2016). Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. *Trends in Genetics* 32, 155–164.
- Wessels, S., Krause, I., Floren, C., Schütz, E., Beck, J., and Knorr, C. (2017). ddRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC Genomics* 18, 531.
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., Cabau, C., Bouchez, O., Fostier, A., and Guiguen, Y. (2012). An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Current Biology* 22, 1423–1428.
- Yehya, S., El-Zaeem, S., Morsi, M., Ahmed, M., El, M., Mohamed, D., and Darwesh, F. (2012). Production of salinity tolerant tilapia through interspecific hybridization between Nile tilapia (*Oreochromis niloticus*) and red tilapia (*Oreochromis sp.*). *African Journal of Agricultural Research* 7, 2955–2961.

CONTRIBUTIONS SCIENTIFIQUES ET ENSEIGNEMENTS

Articles scientifiques

1. Triay Cécile, Conte M.A., Baroiller J.-F., Bezault E., Clark F.E., Penman D.J., Kocher T.D. and D'Cotta H. (2020). Structure and Sequence of the Sex Determining Locus in Two Wild Populations of Nile Tilapia. *Genes* 11, 1017

Contribution personnelle : Paillasse ; Bioinformatique ; Rédaction

2. Triay Cécile, Courcelle M., Caminade P., Bezault E., Baroiller J.-F., Kocher T.D. & D'Cotta H. (*Unpublished*). Polymorphism of Y haplotypes or loss of ancestral LG23 sex determinant in three wild populations of Nile tilapia.

Contribution personnelle : Paillasse ; Bioinformatique ; Rédaction

3. Triay Cécile, Böhne A., Baroiller J.-F., Barbarini M., Kocher T.D. & D'Cotta H. (*Unpublished*). Transcriptome analyses of a wild population of Nile tilapia lacking the ancestral LG23 Y suggests *gsdf* plays a key role in their sex differentiation.

Contribution personnelle : Bioinformatique ; Rédaction

Communications scientifiques internationales

Poster

Triay Cécile, Conte M.A., Baroiller J.-F., Bezault E., Clark F.E., Penman D.J., Kocher T.D. and D'Cotta H. (2019). Sex determination of the Nile tilapia shows polymorphism between two wild populations from East and West Africa. *Paradigm shift in sex chromosome evolution Workshop, Berlin*.

Communication orales

Les communications orales listées ci-dessous sont acceptées mais n'ont pas pu être présentées en 2020 comme prévu du fait des conditions sanitaires. La date affichée correspond à la date de la nouvelle tenue de ces congrès.

1. Triay Cécile, Conte M.A., Baroiller J.-F., Bezault E., Clark F.E., Penman D.J., Kocher T.D. and D'Cotta H. (Automne, 2021). Exploring the sex determining region and its structural variants between wild populations of Nile tilapia. *2nd European Symposium on Sex Determination in Vertebrates*.
2. Triay Cécile, Sissao R., Conte M.A., Baroiller J.-F., Bezault E., Toguyeni A., Kocher T.D. and D'Cotta H. (Mai, 2021). Mapping for Males: sustainable sex-control in Nile tilapia. *6th International Symposium on Genomics in Aquaculture*.

Enseignements

Bio-statistiques appliquées à la biologie – Niveau Master 1 Biologie Santé à l'Université de Montpellier (2017–2019, 2 semestres)

Cette thèse a été financée par le projet CLIMSEX de l'Agence Nationale de la Recherche (ANR) sous le numéro de subvention ANR-15-CE02-0012 et le CIRAD et le LabEx CEMEB ont également contribué à une aide financièrement.

Résumé :

Le déterminisme du sexe chez le tilapia du Nil dépend de facteurs génétiques de type XX/XY et peut être influencé par la température ainsi que par des facteurs parentaux mineurs. Des associations avec le sexe ont été identifiées précédemment sur plusieurs groupes de liaison (LG) et notamment LG1 et LG23 selon la souche étudiée. Les travaux précédents n'ont été réalisés que sur des souches domestiquées (de laboratoire ou commerciales) potentiellement fortement consanguines qui ont fait l'objet de nombreux transferts et de goulots d'étranglement génétiques. Ainsi, il est difficile de savoir si la complexité du déterminisme du sexe génétique est due aux processus de domestication ou si elle reflète la diversité rencontrée dans les populations naturelles. Afin d'étudier si les populations sauvages suivent un système XY hétérogamétique mâle tel que détecté dans les souches domestiquées, et de trouver quels LGs sont impliqués dans le sexe phénotypique, nous avons étudié trois populations sauvages : deux d'Afrique de l'Est (Ethiopie) et une d'Afrique de l'Ouest (Ghana). Différentes analyses génomiques ont été menées au moyen de séquençages de type ddRAD, de séquençages de génomes complets et de séquençage d'ARN en utilisant des technologies de lectures courtes et/ou longues. Nos résultats ont confirmé l'existence du système hétérogamétique mâle sur LG23-Y avec une région spécifique du sexe correspondant à une duplication en tandem qui comporte deux copies du gène de l'Hormone Anti-Müllérienne (*amh*) pour deux des populations sauvages. Nous avons pu définir précisément cette région spécifique du sexe et décrire le contenu et l'ordre des gènes ainsi que les limites de cette duplication. Nos études ont également suggéré que ce système est ancestral à la divergence des populations de l'Est et de l'Ouest. Cependant, nous présentons des preuves de la perte de ce système sexuel ancestral dans l'une des populations éthiopiennes, suggérant un polymorphisme du déterminisme du sexe non seulement dans les souches domestiquées mais aussi dans les populations sauvages de Tilapia du Nil. L'analyse du transcriptome de cette population éthiopienne a donné des indices quant à l'importance de l'expression de *gsdf* dans les stades précoces de la différenciation sexuelle. Aucune nouvelle paire de chromosomes sexuels n'a été mise en évidence pour cette population dépourvue de la duplication de l'*amh* sur LG23-Y, laissant en suspens la question de l'identité ou même de l'existence d'un nouveau déterminant génétique du sexe. Ce travail de thèse met en évidence la possibilité que différents systèmes de déterminisme du sexe ségrègent dans les populations sauvages de tilapia du Nil et, plus généralement, renforce l'idée qu'il est nécessaire de considérer la diversité des déterminismes sexuels au niveau intraspécifique, car celle-ci pourrait être aussi importante qu'à l'échelle interspécifique.

Abstract:

Sex determination in the Nile tilapia depends on genetic factors following an XX/XY pattern and can be influenced by temperature as well as minor parental factors. Associations with sex have been identified previously on several linkage groups (LG) and especially LG1 and LG23 depending on the strain studied. These studies have been performed only on domesticated (laboratory or commercial) strains which have undergone considerable stock transfers, inbreeding and genetic bottlenecks. It is unclear to what extent the complexity of the genetic sex determination is due to the processes of domestication or if it reflects a diversity of sex determination systems encountered in natural populations. To study whether wild populations followed a male heterogametic XY system as detected in domesticated strains, and find what LGs were segregating for the phenotypic sex we studied three different wild populations: two from East Africa (Ethiopia) and one from West Africa (Ghana). A broad range of genomic studies have been conducted through ddRAD sequencing, Whole Genome sequencing and RNA sequencing using both short and long reads technologies. Our results confirmed a male heterogametic sex determination system on LG23-Y with a sex region composed of a tandem duplication which involved two copies of the Anti-Müllerian Hormone gene (*amh*) for two of the wild populations. We were able to define precisely the sex-region and describe the gene content and order, along with the boundaries of the duplication. Our studies also suggested that this system is ancestral to the divergence of Eastern and Western populations. However, we produced evidence for the loss of this ancestral sex-system in one of the Ethiopian population, suggesting polymorphism of sex determination system in wild populations, similarly to what occurs in domesticated strains. Transcriptome analysis of this Ethiopian population gave clues for the early involvement of the *gsdf* gene in sex-differentiation. No evidence for a new pair of sex chromosomes were found for this population lacking the LG23-Y *amh* duplication, leaving open the question of the identity or even the existence of a new genetic sex determinant that would assume this role. This thesis work highlights the potential for different sex-determination system segregating in wild populations of Nile tilapia and more generally emphasizes the need to look for the diversity of sex-system at intra-specific level as it might be as important as at the interspecific scale.