

Analyse et visualisation du gène *TATc*

III.1. 1 – Comparaison entre la carte génétique publiée en 2002 et la carte actuelle disponible sur le serveur Ensembl

Dans une étude génétique préliminaire, 5 marqueurs polymorphes de type microsatellite couvrant la région du gène *TATc* ont été sélectionnés sur la base de données bibliographiques, de leurs pourcentages respectifs d'hétérozygotie et de la distance physique et génétique séparant le locus morbide et le marqueur étudié. L'interrogation du serveur Ensembl en 2002 a permis de déterminer les positions physiques de ces différents marqueurs (Figure 7).

Nous avons alors ré-analysé la région chromosomique du gène *TATc* par le serveur Ensembl afin d'identifier de nouveaux marqueurs microsatellites à tester (Figure 8). Le résultat de l'interrogation du serveur Ensembl dans sa version actuelle a montré un changement des positions physiques des 5 marqueurs microsatellites mentionnés.

Nous avons également observé qu'il existe très peu de marqueurs microsatellites à proximité du gène *TATc*. La fenêtre du serveur Ensembl illustrant notre requête montre plusieurs marqueurs qui correspondent à des marqueurs non polymorphes de type STS.

III 1. 2 – Visualisation de la région du gène *TATc* par le logiciel Artemis

Nous avons pu accéder aux différentes structures géniques : exons, introns, CDS, STS, les régions répétées... etc. Ainsi que leurs positions physiques sur le gène (Figure 9).

CEN

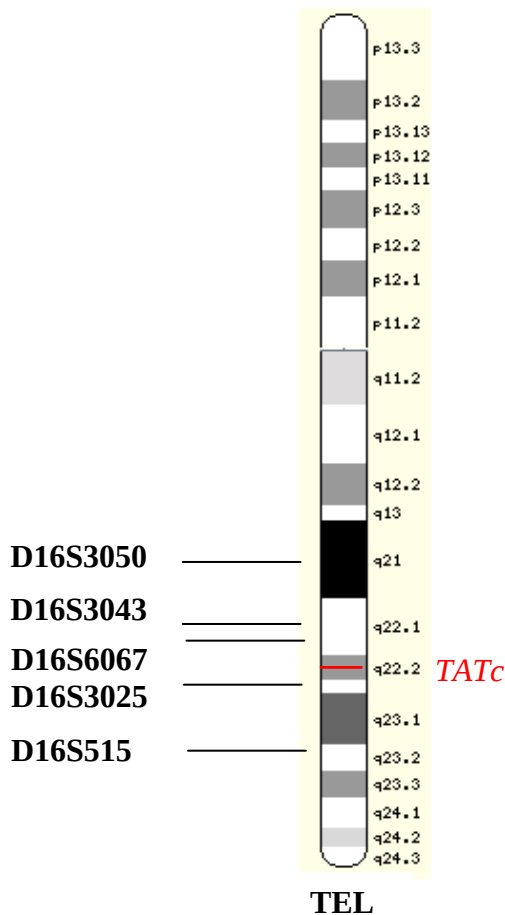


Figure 7: Représentation schématique de l'intervalle chromosomique du gène *TATc* sur le chromosome 16 avec les marqueurs microsatellites proches du gène *TATc*. Les marqueurs sélectionnés par la carte génétique d'Ensembl ne peuvent pas être représentés à cause de la taille relativement élevée de la région à visualiser.



Figure 8: la carte génétique et physique de la région du gène *TATc* fournis par Ensembl

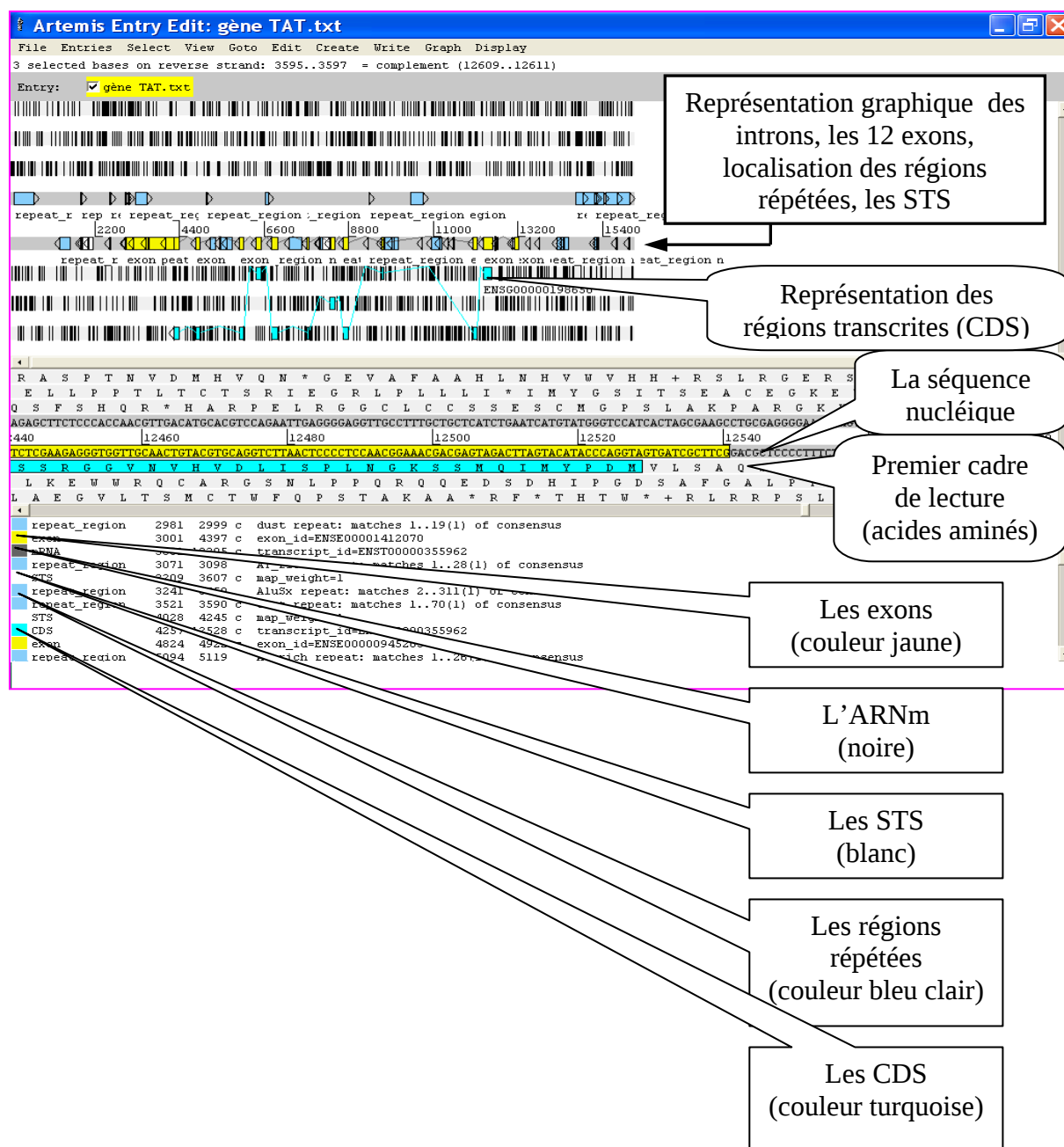


Figure 9: Visualisation du gène *TATc* et illustration d'une «fenêtre» standard du logiciel Artemis.

III. 2 – Identification d'un nouveau polymorphisme intragénique

III. 2. 1 – Mise en évidence de régions répétées de type microsatellites

Deux motifs nucléotidiques répétés en tandem de plus de 10 fois ont été sélectionnés. L'une de ces régions est située dans l'intron 4 et correspond à un motif répété imparfait tétranucléotidique de type (CAAA)₁₀. L'autre région est localisée juste en amont du gène *TATc* (-2Kb) et correspond à un motif dinucléotidique répété de type (TA)₁₇ (Figure 10). Nous avons également, observé avec l'annotation proposée par la GeneBank un ensemble d'éléments répétés essentiellement autour de la séquence microsatellite du type (CAAA)₁₀.

III. 2. 2 – Comparaison entre la séquence du gène *TATc* publié en 1990 et la séquence actuelle

En utilisant le logiciel ALIGN (www.infobiogen.fr), nous avons aligné deux séquences nucléotidiques s'étendant de -2,3Kb à -1,9 Kb du gène *TATc*, l'une provenant de l'article publié pour la première fois par Rettenmeier et al. (Rettenmeier *et al.*, 1990) et l'autre extraite des banques de données actuelles. Cet alignement a mis en évidence une insertion de 8 dinucléotides (TA) dans la séquence de Rettenmeier et al. par rapport à celle disponible dans les banques de données (Figure 11).

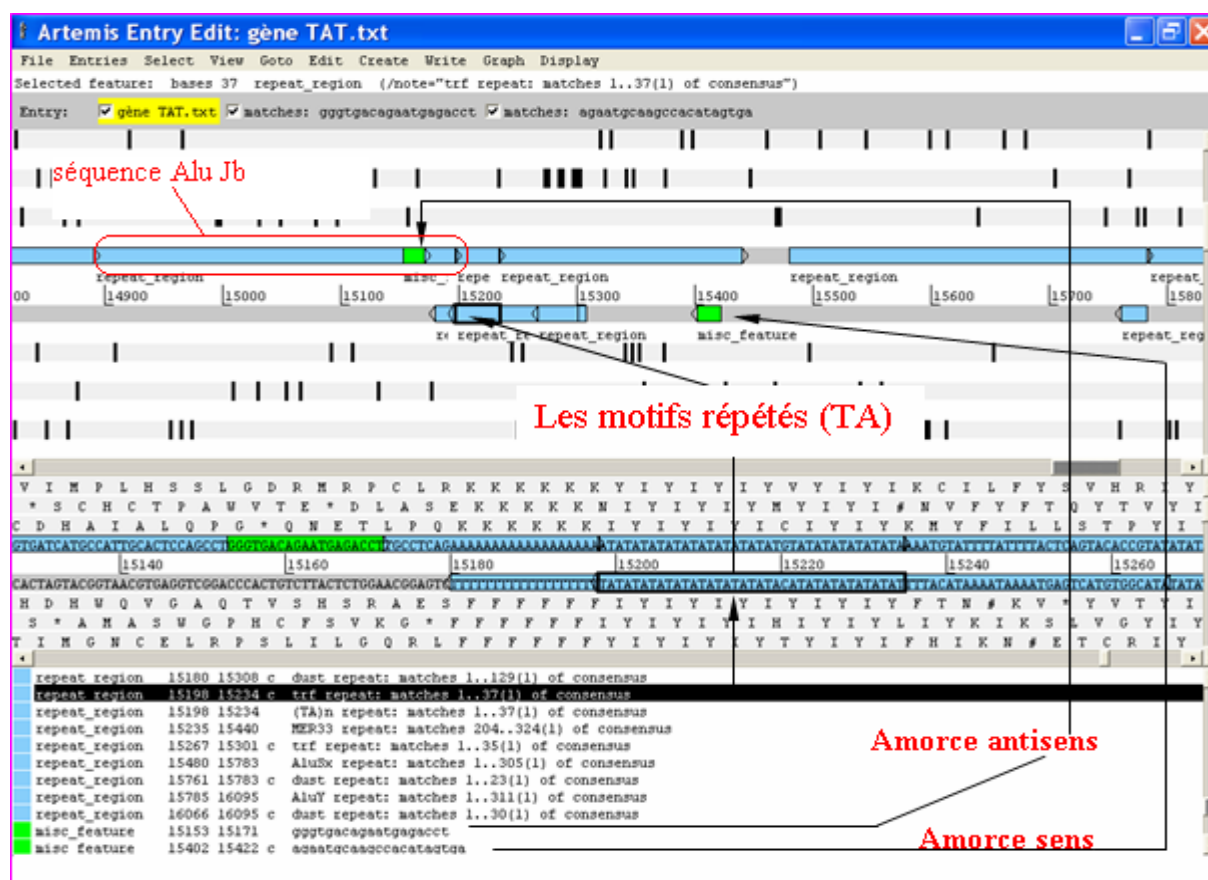
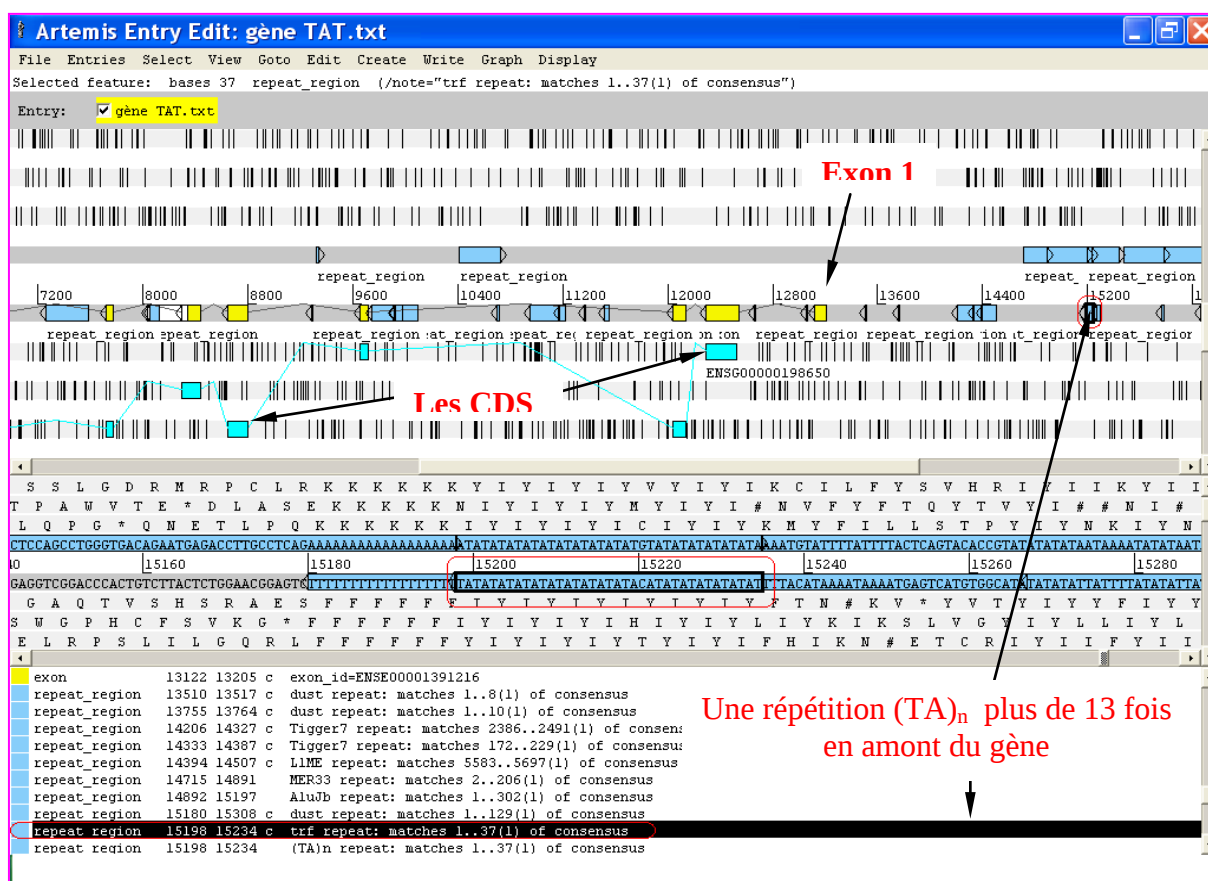


Figure 10 : Mise en évidence d'une répétition de type microsatellite en amont du gène TATc. En bas : Visualisation de la séquence du microsatellite avec les amorces choisis.

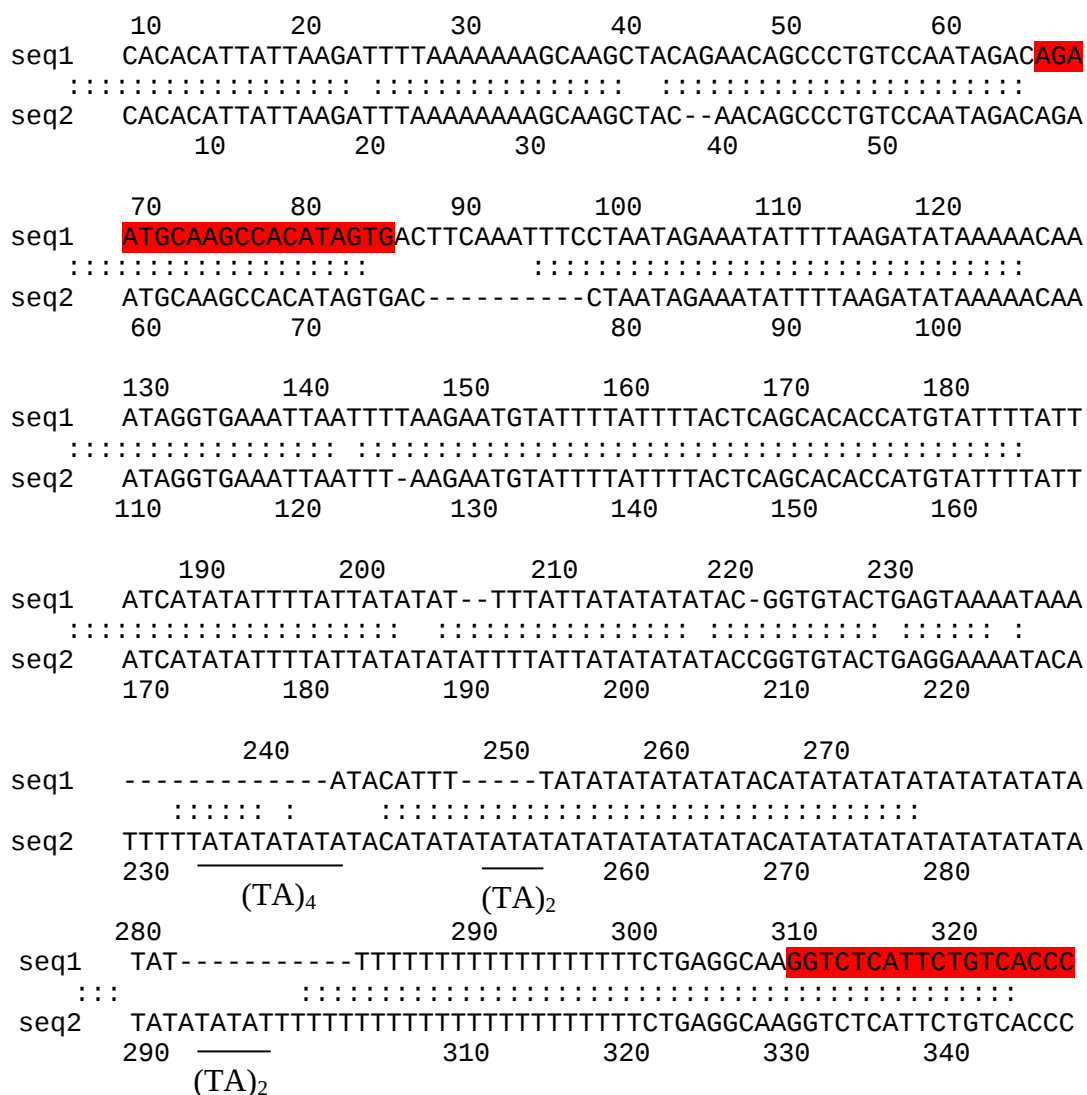


Figure 11: Comparaison entre les séquences génomiques de -2,3 Kb à -1,9 Kb en amont du gène *TATc* :

seq1 : la séquence publiée dans les banques de données.

seq2 : la séquence publiée du gène *TATc* pour la première fois par Rettenmeier *et al.*, 1990.

(:) : correspond aux similitudes des nucléotides entre les deux séquences.

(-) : correspond aux discordances des nucléotides entre les deux séquences.

Les espaces correspondent aux "brèches" (*gaps*) ; les chiffres correspondent à la position des nucléotides.

(TA)_n : correspondent aux répétitions du dinucléotides (TA) insérées dans la séquence de Rettenmeier *et al.*, 1990.

AGAATGCAAGCCACATAGTG : séquences oligonucléotidiques choisis pour amplifier le motif répété.

III. 2. 3 – Recherche de similarité et Alignement de séquences

Pour chacune des deux régions répétées sélectionnées, la séquence requête est une courte séquence de quelques centaines de paires de bases contenant le motif répété entouré par les séquences voisines. La base de donnée servant comme matrice est celle des séquences humaines de type *Alu* «*alu_repeats*». En utilisant l'option «*Limit by entrez query*», nous avons sélectionné l'espèce *Homo sapiens*.

Les résultats obtenus ont montré que la séquence répétée (ACCC)₁₀ est inclus dans une répétition de type Alu-J₀. La séquence correspondant au motif répété (TA)₁₇ quant à elle, ne montre aucune similarité avec l'ensemble des répétitions de type *Alu* dans le génome humain malgré qu'elle est entourée de deux régions à forte similarité (81-85%) avec les répétitions Alu-Sb (Figure 12).

La recherche de similarité et l'alignement des courtes séquences (19 à 21 nucléotides) flanquant les régions répétées, avec les séquences *Alu* par le logiciel BLAST, est significatif pour la spécificité de chaque paire d'amorces (Figure 13).

Figure 12 : Résultats Blast significatifs avec les banques de séquences *Alu*
Query : séquence requête ; sbjct : séquence cible ; les numéros à droite et à gauche correspondent aux positions de la séquence respectivement.

5' ATTTTATTTTATTTTATTTTGGAGACGGAGTCTCACTCTGTGCGCCAGACTGGAGTGCAGTGGCACG
ATCTCGGCTCACCACAACCTCCACCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTTGC
TGGGATTACAGGTGTGCACCACTGTGCCTGGCTAATTGTTGTATTTTAGTAGAGATGGGGTTTTGCC
ATGTTGGCCAGGTTGGTCTTGAACCTCTGATCTCAGATGATCTGCCTGCCTTGGCCTCCAAAGTGCTG
GGGTTACAGGCATGAGCTGCCACGCCCAGCCACACATTATTAAGATTTTAAAAAAGCAAGCTACAG
AACAGCCCTGTCCAATAGACAGAATGCAAGCCACATAGTGACTTCAAATTTCTAATAGAAATATTTT
AAGATATAAAAACAAATAGGTGAAATTAATTTTAAGAATGTATTTTATTTTACTCAGCACACCATGTA
TTTTATTATCATATATTTTATTATATATTTTATTATATATATACGGTGTACTGAGTAAAATAAAATAC
ATTTTATATATATATATACATATATATATATATATATATATTTTTTTTTTTTTTTTTTCTGAGGCAA
GGTCTCAATCTGTCACCCAGGCTGGAGTGCAATGGCATGATCACAGCTCTCTGCAGCCTTGACCTCTT
GGGTTCAAGGTGATCTTCTACCTCAGCCTCCCAAGTAGCTGGGACTATAGGCTCATGCCACCATGCCT
GGCTAGTTTTTGTATCTTTTTGTATAGACGGGGTTTCACTATGTTGCCCAAGCTGGTCTTGAATTCCT
TAGCTCAAGTGATCCATCCACCTTGGCCTCCCGAAAGGCTGGGATTACCAAAGGTGTGAGCCACCATG
CCGAGCCCCAAAATATTC 3'

Figure 13: Séquence partielle du gène *TATc* montrant la répétition (TA)₁₇.

La région répétée est représentée en caractères gras ; les régions correspondant à des séquences répétées de la famille Alu sont encadrées ; les oligonucléotides choisis pour mettre en évidence le polymorphisme de répétition par PCR sont soulignées.

III. 3 – Génotypage de 5 individus non apparentées

Pour le marqueur à motif répété (ACCC)₁₀, les 5 individus de la population générale testés présentent un seul allèle.

Pour le deuxième marqueur (TA)₁₇, nous avons observé un polymorphisme dans la taille des allèles (Figure 14). Cinq allèles ont été observés (notés de 1, 2, 3, 4 et 5) chez des individus non apparentés: 4 individus sont hétérozygotes et un individu homozygote. Les individus hétérozygotes représentent deux à deux le même génotype (tableau 5). Le pourcentage d'hétérozygotie pour ce marqueur est de 80%.

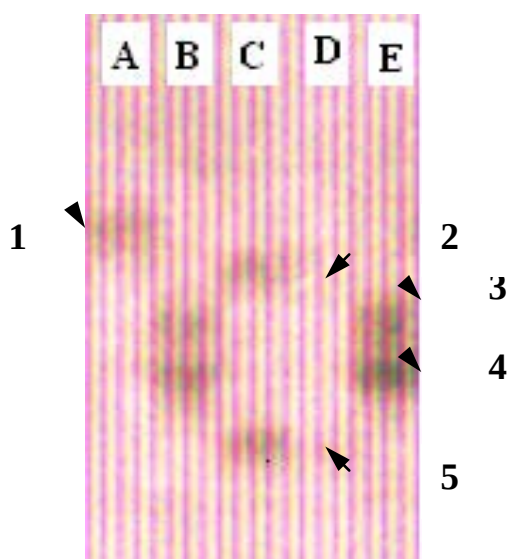


Figure 14: Autoradiographie montrant les différents allèles observés (1, 2, 3, 4 et 5) chez les individus (A, B, C, D et E) pour le marqueur du type (TA)₁₇.

Tableau 5: Génotypes des individus testés avec le marqueur comprenant la répétition (TA).

Individus	Le nombre d'allèles	Génotype	
A	1	[1 1]	Homozygote
B	2	[3 4]	Hétérozygote
C	2	[2 5]	Hétérozygote
D	2	[2 5]	Hétérozygote
E	2	[3 4]	Hétérozygote