

Table des matières

Résumé.....	iii
Abstract.....	iv
Table des matières.....	v
Liste des tableaux.....	vii
Liste des figures.....	viii
Liste des abréviations.....	ix
Remerciements.....	xii
Avant-propos.....	xiii
Chapitre 1 : Introduction générale.....	1
1.1 Le pergélisol.....	1
1.1.1 Les lacs de fonte de pergélisol.....	3
1.1.2 Caractéristiques limnologiques.....	4
1.1.3 Implication dans le réchauffement climatique.....	6
1.1.4 Écologie des lacs de thermokarst.....	7
1.2 Les virus.....	9
1.2.1 Rôles écologiques.....	10
1.2.1.1 Court-circuit viral.....	10
1.2.1.2 «Tuer le gagnant» (Kill-the-winner).....	11
1.2.1.3 «Reine Rouge» (Red Queen).....	11
1.2.1.4 «Chat du Cheshire» (Cheshire Cat).....	13
1.2.1.5 Transfert horizontal de gènes.....	14
1.2.1.6 Gènes auxiliaires métaboliques.....	15
1.2.1.7 Contrôle des populations microbiennes.....	17
1.2.2 Les virus infectant le phytoplancton.....	18
1.2.2.1 Les Phycodnaviridae - Chlorovirus.....	18
1.2.2.2 Les cyanophages.....	20
1.2.3 Écologie virale dans les écosystèmes d'eau douce polaires.....	21
1.2.4 Approches utilisées pour étudier les virus.....	22
1.3 Problématique.....	23
1.4 Objectifs spécifiques.....	24
1.5 Sites d'échantillonnage.....	24
1.5.1 Régions des villages de Whapmagoostui-Kuujuarapik et Umiujaq.....	25
1.5.2 Lacs à l'étude.....	26
Chapitre 2 : Contrôle des communautés virales dans les mares de fonte de pergélisol par le type de mare.....	28
2.1 Résumé.....	28
Chapter 2 : Landscape control of viral communities in permafrost thaw ponds.....	29
2.2 Abstract.....	29
2.3 Introduction.....	30
2.4 Materials and Methods.....	32
2.4.1 Study site and sampling.....	32
2.4.2 Sample processing and molecular analysis.....	33
2.4.3 Bioinformatics and statistical analyses.....	34
2.5 Results.....	36
2.5.1 Limnological conditions.....	36
2.5.2 Cellular communities.....	36
2.5.2.1 Community composition.....	36
2.5.2.2 Community comparisons.....	38

2.5.3	Viral communities	42
2.5.3.1	Community composition.....	42
2.5.3.2	Community comparisons	43
2.6	Discussion	45
2.7	Acknowledgments	49
2.7	Supporting information	49
Chapitre 3 : Isolement et séquençage de deux nouveaux cyanophages d'origine subarctique		51
3.1	Mise en contexte	51
3.1.1	Les cyanobactéries	51
3.1.2	Les cyanophages	53
3.2	Matériel et Méthodes	53
3.2.1	Amplification des cyanophages	54
3.2.1.1	Cultures de cyanobactéries	54
3.2.1.2	Inoculation avec les virus.....	55
3.2.1.3	Rondes d'amplification.....	55
3.2.2	Microscopie électronique	56
3.2.3	Isolement	57
3.2.4	Extraction d'ADN et séquençage	57
3.2.5	Analyses génomiques	58
3.2.6	Analyses phylogénétiques	58
3.3	Résultats	58
3.3.1	Lyse des cultures de cyanobactéries	58
3.3.2	Caractéristiques génomiques	59
3.3.3	Morphologie des particules virales	61
Chapitre 4 : Discussion générale et conclusion		63
4.1	Discussion	63
4.1.1	Objectif de recherche 1	63
4.1.2	Objectif de recherche 2	65
4.2	Conclusion	68
4.2.1	Perspectives	69
Bibliographie		71
Annexe		83

Liste des tableaux

Tableau 1.1 Principaux chlorovirus selon le symbiote de leur hôte (<i>Chlorella sp.</i>)	20
Table 2.1 Sets of primers used for viral amplification	34
Table 2.2 Limnological properties of the sampled lakes in August 2015	36
Table S2.1 Alpha-diversity statistics for samples grouped by organism, site and depth (S, surface; B, bottom)	49
Tableau 3.1 Description des cultures de cyanobactéries sélectionnées	54
Tableau A.1 Annotation génomique du cyanophage B3	84
Tableau A.2 Annotation génomique du cyanophage B23	86

Liste des figures

Figure 1.1 Distribution du pergélisol en Arctique	2
Figure 1.2 Différentes étapes de la formation d'un lac de thermokarst suite à la fonte d'une butte de pergélisol riche en glace	4
Figure 1.3 Profils limnologiques de trois lacs de thermokarst subarctiques	5
Figure 1.4 Schématisation du court-circuit viral	11
Figure 1.5 Localisation des trois vallées échantillonnées (SAS, BGR, SRB) au nord du Québec (Nunavik, Canada)	25
Figure 1.6 Anomalies de la température de l'air dans le village de W-K en comparaison à la valeur moyenne au 20e siècle	26
Figure 1.7 Photo de chaque site d'échantillonnage	27
Figure 2.1 Location of the three sampled valleys in the Canadian subarctic (Nunavik, Québec)	33
Figure 2.2 Relative abundance of each bacterial and microeukaryotic taxa in the samples grouped by lake and depth (surface -S or bottom -B)	37
Figure 2.3 Principal coordinate analysis (PCoA) generated with the unweighted UniFrac distances of the myoviral (A), chloroviral (B), bacterial (C), and microeukaryotic (D) communities from the three lakes	39
Figure 2.4 Unweighted UniFrac UPGMA clustering of the samples combined with a heatmap representing the relative abundance of the most abundant total myovirus (A), chlorovirus (B), bacterial (C) and eukaryotic (D) OTUs (>1% of the data set)	41
Figure 2.5 Venn diagrams representing the proportion of shared OTUs between lakes for each tested gene	42
Figure 2.6 Classification of viral OTUs according to the DIAMOND nearest match and the set of primers used (A, myovirus; B, chlorovirus)	43
Figure 2.7 Phylogenetic tree using the evolutionary placement algorithm (EPA) in RAxML of the chloroviral 41 OTUs	44
Figure S2.1 Box plots representing Chao1 richness and Shannon diversity of each sample, grouped by tested gene, according to the lake and the depth sampled	50
Figure 3.1 Les cyanobactéries peuvent posséder deux morphologies distinctes filamenteuse (A) ou unicellulaire (B)	52
Figure 3.2 Différences observées entre les cultures contrôles et infectées lors de l'amplification des cyanophages	59
Figure 3.3 Représentation du génome des cyanophages B3 et B23	60
Figure 3.4 Arbre phylogénétique sans racine basé sur l'alignement au niveau des acides aminés du gène de la primase des cyanophages B3 et B23 avec ceux de 275 autres phages	61
Figure 3.5 Photos obtenues en TEM des deux cyanophages isolés	62
Figure A.1 Alignement des génomes de B3 et B23	83

Liste des abréviations

Abi	Avortement de l'infection (<i>Abortive infection</i>)
ACTV-1	<i>Acanthocystis turfacea chlorella virus 1</i>
ADN	Acide désoxyribonucléique (<i>DNA</i>)
AMG	Gène auxiliaire métabolique (<i>Auxiliary metabolic gene</i>)
ARN	Acide ribonucléique
ARNr	Acide ribonucléique ribosomale (<i>rRNA</i>)
ARNt	Acide ribonucléique de transfert
BGR	Vallée de Bundesanstalt für Geowissenschaften und Rohstoffe
°C	Degré Celsius
CaCO ₃	Carbonate de calcium
Cas	Gène associé au locus CRISPR (<i>CRISPR associated gene</i>)
CDOM	Matière organique dissoute colorée (<i>Coloured dissolved organic matter</i>)
CDS	Séquence d'ADN codante (<i>Coding DNA sequence</i>)
CEN	Centre d'Études Nordiques
CH ₄	Méthane
Chl a	Chlorophylle a
CO ₂	Dioxyde de carbone
CRISPR	Petites répétitions palindromiques groupées et régulièrement espacées (<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>)
db	Double brin
dNTP	Désoxyribonucléotide triphosphate
DO	Oxygène dissout (<i>Dissolved oxygen</i>)
DOC	Carbone organique dissout (<i>Dissolved organic carbon</i>)
DOM	Matière organique dissoute (<i>Dissolved organic matter</i>)
e.g.	Par exemple (<i>exempli grātia</i>)
EhV	Virus d' <i>Emiliana huxleyi</i>
EhV1	<i>Emiliana huxleyi virus 1</i>
EPA	Algorithme de placement évolutif (<i>Evolutionary placement algorithm</i>)
FRQNT	Fond de Recherche du Québec – Nature et Technologies
G+C%	Pourcentage en G+C (du génome)
GES	Gaz à effet de serre
GTA	Agent de transfert de gène (<i>Gene transfer agent</i>)
ICTV	Comité international sur la taxonomie des virus (<i>International Committee on Taxonomy of Viruses</i>)
kb	Kilobase (10 ³ bases nucléiques)
km ²	Kilomètre carré
L4K	Lake 4K
M	Molaire
m ²	Mètre carré
Mb	Méga-base (10 ⁶ bases nucléiques)
MIMIVIRE	Élément de résistance au virophage du Mimivirus (<i>mimivirus virophage resistance element</i>)
mL	Millilitre
ML	Maximum-likelihood

mM	Millimolaire
µL	Microlitre
µm	Micromètre
µmol	Micromole
°N	Degré de latitude nord
NCLDV	Grands virus nucléocytoplasmiques (<i>Nucleocytoplasmic large DNA viruses</i>)
nm	Nanomètre
NSERC	Natural Sciences and Engineering Research Council of Canada
NSTP	Northern Scientific Training Program
°O	Degré de longitude ouest
O ₂	Dioxyène (ou simplement oxygène)
ORF	Cadre de lecture ouvert (<i>Open reading frame</i>)
OTU	Unité opérationnelle taxonomique (<i>Operational taxonomic unit</i>)
pb	Paire de base
PBCV-1	<i>Paramecium bursaria chlorella virus 1</i>
PCoA	<i>Principal coordinates analysis</i>
PCR	Réaction de polymérase en chaîne (<i>Polymerase chain reaction</i>)
PEG	Polyéthylène glycol
Pg	Pétagramme (10 ¹⁵)
POM	Matière organique particulaire (<i>Particular organic matter</i>)
ppGpp	Guanosine pentaphosphate
PPP	Voie des pentoses phosphates (<i>Phosphate pentose pathway</i>)
rDNA	ADN ribosomal
s	Seconde
SAS	Vallée de la rivière Sasapimakwananisikw
sb	Simple brin
SRB	Vallée de bassins rocheux peu profonds (<i>Shallow rock-basin</i>)
TEM	Microscopie électronique à transmission (<i>Transmission electronic microscopy</i>)
Tg	Téragramme (10 ¹²)
TN	Azote total (<i>Total nitrogen</i>)
TP	Phosphore total (<i>Total phosphorus</i>)
TSS	Solides totaux en suspension (<i>Total suspended solids</i>)
UPGMA	<i>Unweighted pair group method with arithmetic mean</i>
UV	Ultra-violet
W-K	Whapmagoostui-Kuujuarapik

*« Yesterday's the past,
Tomorrow's the future,
But today is the gift,
That's why it's called the present. »
- Bil Keane*

Rapport-Gratuit.com

Remerciements

Il y a tant de personnes que je voudrais remercier pour m'avoir aidée et accompagnée tout au long de ce projet de maîtrise. Tout d'abord, j'aimerais dire un énorme merci à mon directeur de recherche, Alexander Culley, qui m'a accueillie il y a déjà 3 ans dans son laboratoire dans le cadre d'un projet de recherche au baccalauréat. Il m'a ainsi transmis sa passion pour l'écologie virale et les environnements nordiques, des domaines de la microbiologie qui m'étaient à ce moment encore inconnus. Merci Alex d'avoir été si compréhensif, et toujours à l'écoute durant toutes ces années. Tes judicieux conseils m'ont grandement été utiles dans l'accomplissement de cette maîtrise. J'aimerais également remercier mon codirecteur de recherche, Warwick Vincent pour son positivisme et son enthousiasme peu importe la situation. Merci Warwick de m'avoir partagé ton intérêt pour les écosystèmes nordiques. Dorénavant, je perçois le Nord d'une toute autre manière.

Je tiens également à remercier les autres membres de mon comité d'encadrement, Sylvain Moineau et Connie Lovejoy pour leurs commentaires constructifs apportés lors de notre rencontre au tout début du projet.

Pour m'avoir initiée au travail de terrain à Kuujuarapik, je désirerais remercier Bethany Deshpande, Alex Matveev et Vani Mohit qui ont fait de cette première session d'échantillonnage une merveilleuse expérience. Concernant le travail en laboratoire, je voudrais remercier tous les membres des laboratoires Culley, Lovejoy, Moineau et Vincent qui ont toujours su répondre à mes nombreuses questions. D'ailleurs, je voudrais dire un merci tout particulier à Jérôme Comte pour sa patience et surtout tout le temps qu'il a dédié à m'éclaircir sur les mystères de la bio-informatique et des statistiques.

Enfin, je voudrais chaleureusement remercier mes ami(e)s et ma famille. Merci Catherine, Valérie et surtout Myriam pour toutes ces discussions passées dans le bureau. Merci à Jessica, Benjamin et Clara qui remplissent ma vie dans la ville de Québec de moments inoubliables. Merci à mes parents qui m'ont toujours encouragée et soutenue tout au long de mes études. Et un dernier merci à Cas pour faire partie de ma vie à tous les jours.

Avant-propos

Ce mémoire de maîtrise est divisé en quatre chapitres suivi d'une annexe. Le chapitre 1 consiste en une introduction générale des concepts traités à l'intérieur de ce mémoire. Le chapitre 2, rédigé sous forme d'un article scientifique, contient la méthodologie et les résultats obtenus au cours de mon premier objectif de recherche. Le chapitre 3 correspond à la méthodologie employée ainsi que les résultats générés dans le cadre de mon second objectif de recherche. Enfin, le chapitre 4 regroupe une discussion générale et une conclusion des résultats globaux tirés de ce projet de maîtrise.

Le deuxième chapitre de ce mémoire est rédigé en anglais sous la forme d'un article scientifique qui sera prochainement soumis au journal *Aquatic Microbial Ecology*. Je serai la première auteure de cet article avec comme coauteurs, Warwick F. Vincent, Jérôme Comte, Connie Lovejoy et Alexander I. Culley. J'ai contribué à la conception du projet avec Alexander I. Culley et Warwick F. Vincent, j'ai réalisé la prise d'échantillons sur le terrain avec l'aide d'Alexander I. Culley et Warwick F. Vincent, j'ai effectué l'ensemble des expérimentations en laboratoire, et j'ai réalisé les analyses bio-informatiques avec l'aide de Jérôme Comte. La base de données microbienne nordique utilisée pour l'assignation taxonomique était gérée par Connie Lovejoy. J'ai composé la première version du manuscrit, et tous les coauteurs ont participé à son amélioration par l'apport de commentaires.

Chapitre 1 : Introduction générale

1.1 Le pergélisol

Le pergélisol est défini comme étant le sol gelé pendant une période d'au moins deux années consécutives (Margesin, 2009). Celui-ci est composé généralement de roches, de sédiments, de matières organiques ou d'autres matériaux terrestres. Il existe un gradient de pergélisol en fonction de la latitude (Figure 1.1). En effet, plus on se rapproche du pôle nord (90°N), plus la concentration du sol en pergélisol est élevée. Cette logique est également applicable en fonction de l'altitude. On retrouve du pergélisol de type alpin à des latitudes plus basses en région montagneuse telles que dans les hauts plateaux des Alpes et du Tibet (Boeckli et al., 2012; Wu et al., 2010). Différents termes sont utilisés pour qualifier le pergélisol en fonction de sa concentration dans le sol : continu (90-100% du territoire), discontinu (50-90% du territoire), sporadique (10-50% du territoire) et isolé (2-10% du territoire) (Brown et al., 1998). Il est estimé que le pergélisol recouvre près du quart des terres émergées dans l'hémisphère nord, soit environ 26 millions de km² (Zhang et al., 2000). On constate rapidement que le pergélisol est réparti majoritairement entre la Russie et le Canada. L'Alaska, le Groenland et la Scandinavie en sont également pourvu mais avec une étendue moindre (Margesin, 2009).

Un sol pergélisolé comprend deux couches distinctes. Il y a d'abord à la surface, sous le couvert végétal, la couche active. Son épaisseur est dépendante de la capacité du sol, de la végétation et/ou de la neige à transférer la chaleur (ACIA, 2005). En effet, il s'agit de la couche du sol qui subit les variations saisonnières de température et qui va dégeler annuellement à chaque période estivale (Margesin, 2009). Sous la couche active, on retrouve le pergélisol, le sol qui demeure gelé en permanence et qui ne subit que très peu de variations de température au cours de l'année.



Figure 1.1 : Distribution du pergélisol en Arctique. Les différentes nuances de violet représentent les différents types de pergélisol. Les nuances plus foncées correspondent aux sols plus riches en pergélisol, tandis que les nuances plus pâles réfèrent aux sols pauvres en pergélisol (Crédits : Philippe Rekacewicz, IPA, 1998, Circumpolar Active-Layer Permafrost System).

En raison du réchauffement climatique, la limite sud du pergélisol ne cesse de graduellement remonter vers le nord lors des dégels estivaux (Smith and Riseborough, 2002). En effet, la hausse des températures entraîne la dégradation et la fonte du pergélisol. Il est également démontré que la couche active du sol est de plus en plus profonde dans plusieurs régions arctiques (Anisimov et al., 1997). Ces changements dans la distribution du pergélisol ont des impacts directs sur les processus hydrologiques, géomorphologiques et biologiques des environnements polaires (Nelson, 2003). De plus, il est estimé que le pergélisol arctique renferme 1672 Pg (10^{15} grammes) de carbone organique (Tarnocai et al., 2009). Ainsi, sa fonte risque de causer le relargage de grande quantité de carbone vers l'atmosphère sous forme de dioxyde de carbone (CO_2) et de méthane (CH_4). Ces deux gaz étant de puissants gaz à effets de serre (GES), ce relâchement engendrerait donc une boucle de rétroaction positive où les GES émis par la fonte du pergélisol contribueraient à accélérer le réchauffement planétaire, stimulant du même coup la fonte et l'activité microbienne dans les sols pergélisolés (Koven et al., 2011).

1.1.1 Les lacs de fonte de pergélisol

Les lacs de fonte de pergélisol, également appelés lacs (ou mares) de thermokarst, sont des écosystèmes d'eau douce formés suite à la fonte du pergélisol. Avec le réchauffement climatique, les régions polaires subissent une hausse des températures plus élevées que partout ailleurs sur la planète. On estime que les régions arctiques ont subi un réchauffement de 2,1°C depuis 1990, tandis que le reste du monde s'est réchauffé en moyenne de 0,4°C, soit cinq fois plus lentement que l'Arctique pour la même période (ACIA, 2005). Ainsi, le pergélisol fond de manière accélérée favorisant l'érosion du sol et la formation de dépressions où l'eau de fonte vient s'accumuler, ce qui crée un lac de fonte (Margesin, 2009). De ce fait, les lacs de thermokarst sont dorénavant parmi les écosystèmes aquatiques les plus répandus dans les régions arctiques et subarctiques canadiennes (Pienitz et al., 2008).

En zone de pergélisol sporadique à discontinu, la formation d'un lac de fonte peut se diviser en quatre étapes (Figure 1.2) (Bouchard et al., 2016). Tout d'abord, on retrouve dans le paysage une butte composée de pergélisol riche en glace. Puis, avec la hausse des températures, le pergélisol fond graduellement, ce qui favorise l'érosion et l'effondrement du sol à une extrémité de la butte. L'eau commence alors à s'accumuler dans cette petite dépression, créant petit à petit un étang en forme de croissant de lune. Par la suite, la fonte du pergélisol se poursuit entraînant l'élargissement et l'évolution de l'étang en petit lac. Enfin, la butte de pergélisol fond complètement et laisse place à un lac mature lorsque sa profondeur maximale est atteinte. Il est également important de remarquer une densification de la végétation autour de la butte tout au long du développement du lac (Bouchard et al., 2014; Lloyd et al., 2003). Avec l'évolution du lac de thermokarst, la végétation composée initialement de petits arbustes, de lichens et de mousses passe progressivement à une végétation arbustive. En zone de pergélisol continu, où l'on retrouve une grande quantité de coins de glace («ice wedges»), les lacs de fonte prennent plutôt une forme polygonale (Fortier and Allard, 2004).

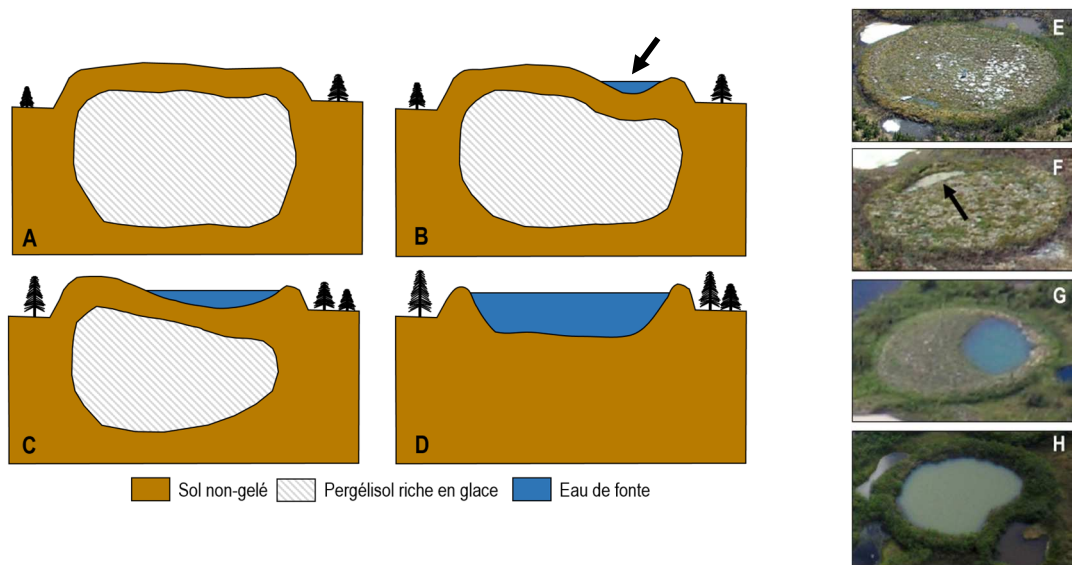


Figure 1.2 : Différentes étapes de la formation d'un lac de thermokarst suite à la fonte d'une butte de pergélisol riche en glace. (A) Butte de pergélisol (palse ou lithalse). (B) Début de la fonte de la butte de pergélisol et formation d'une mare où il y a affaissement du sol (indiquée par une flèche). (C) Agrandissement de la mare due à la fonte du pergélisol. (D) Formation d'un lac mature lorsque la butte a complètement fondue. (E-H) Évolution de la dégradation d'une lithalse suivant les mêmes étapes. Figure inspirée de Bouchard *et al.* (2016) et de Calmels *et al.* (2008).

Il existe deux types de lac de thermokarst déterminés en fonction de la concentration en matière organique de la butte de pergélisol dont le lac est originaire. Un lac de thermokarst formé à partir d'une butte riche en matière organique, comme dans le cas d'une tourbière, est appelé palse (Gurney, 2001). À l'opposée, la butte peut être composée majoritairement d'un substrat minéral et être pauvre en matière organique. Le lac résultant est alors appelé lithalse (Calmels *et al.*, 2008).

Dans une même région, les lacs de thermokarst ne sont pas tous au même stade de développement (Chen *et al.*, 2014; Manasyrov *et al.*, 2014). De manière générale, les lacs de thermokarst une fois mature vont finir par diminuer en taille et naturellement disparaître suite à des processus naturels tels que l'évaporation, la dégradation du pergélisol et le drainage (Jones *et al.*, 2011). Parfois, certains lacs vont même se reformer (van Huissteden *et al.*, 2011).

1.1.2 Caractéristiques limnologiques

Durant l'été, les lacs de thermokarst sont généralement caractérisés par une forte stratification verticale causée par des densités et des températures distinctes à la surface et au fond du lac. La grande

quantité de matière en suspension a tendance à absorber les rayonnements solaires, ce qui a pour effet de réchauffer davantage la couche de surface et d'entraîner une stratification thermique de la colonne d'eau; l'eau de fond demeurant plus froide. L'eau froide étant plus dense, celle-ci ne se mélange pas avec l'eau de surface moins dense, accentuant ainsi la stratification verticale. Par ailleurs, durant l'hiver, la colonne d'eau ne gèle pas complètement, ce qui concentre du même coup tous les ions au fond du lac augmentant la densité de la couche de fond (Deshpande et al., 2015). De plus, en raison de la respiration bactérienne et de l'absence de brassage, la couche de fond devient éventuellement hypoxique ou totalement anoxique ce qui crée des conditions diamétralement opposés au fond et à la surface (oxique) du lac (Deshpande et al., 2015). En conséquence, ces conditions distinctes seront propices pour le développement de différentes populations microbiennes tout au long de la colonne d'eau (Rossi et al., 2013). Ainsi, tous ces paramètres favorisent la création d'un gradient physico-chimique vertical malgré que les lacs de thermokarst soient très peu profonds (1-4 m) (Breton et al., 2009) (Figure 1.3).

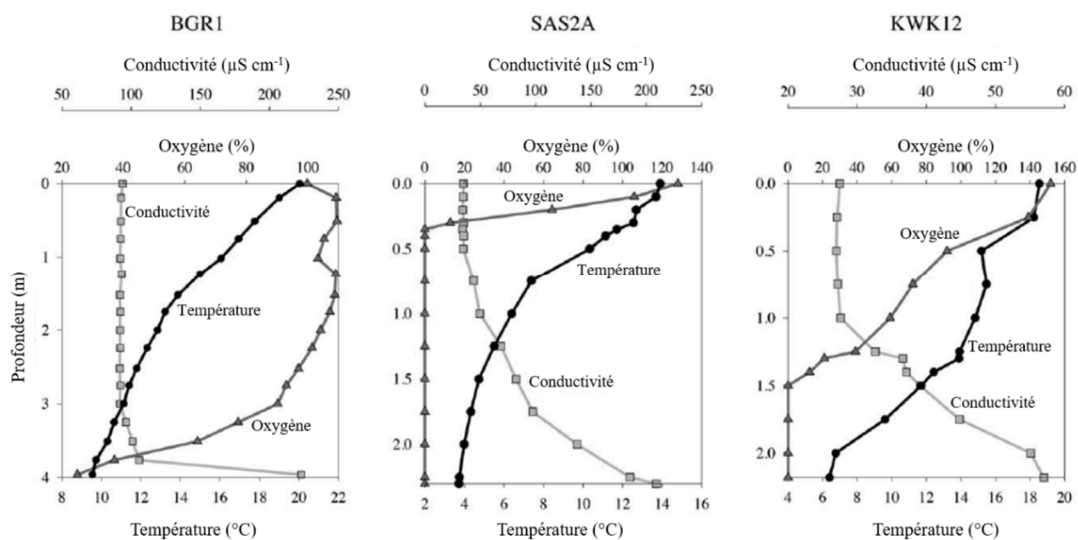


Figure 1.3 : Profils limnologiques de trois lacs de thermokarst subarctiques. Les conditions physico-chimiques sont totalement distinctes entre la surface et le fond des lacs. Dans les trois lacs, l'oxygène et la température diminuent avec la profondeur, tandis que la conductivité augmente drastiquement. BGR-1 provient d'une lithalse, SAS-2A d'un palse et KWK12 d'une lithalse complètement dégradée. Figure adaptée de Deshpande et al. (2015).

Dans une même vallée, chaque lac de fonte présente des propriétés biologiques, physiques et chimiques bien distinctes même s'ils sont très près les uns des autres (Breton et al., 2009). Cette grande variabilité peut être notamment observée par les différentes couleurs que peuvent arborer les

lacs (bleu clair, beige, brun, noir, etc.). Une étude a démontré que la coloration d'un lac est déterminée en fonction du rapport entre le carbone organique dissout (DOC) et la matière particulaire en suspension non-algale (argile, limon, etc.) (Watanabe et al., 2011). Avec l'augmentation de la quantité de DOC, la lumière bleue et verte est de plus en plus absorbée laissant la lumière jaune et rouge être perçue par l'œil humain. C'est pourquoi les lacs riches en matière organique possèdent généralement une coloration très foncée.

1.1.3 Implication dans le réchauffement climatique

Une étude publiée dans le prestigieux journal *Science* a démontré que les lacs de thermokarst ont joué un rôle majeur lors de la dernière déglaciation, il y a environ 9 000 à 14 000 ans (Walter et al., 2007). En effet, selon des estimations basées sur des observations actuelles et des analyses paléogéographiques, les lacs de fonte de pergélisol auraient émis par ébullition de 13 à 26 Tg de CH₄ par année, favorisant de manière importante le réchauffement du climat. Leurs données suggèrent que la formation de ce type de lac en Sibérie, au Canada et en Alaska coïncide avec une augmentation substantielle des concentrations de CH₄ atmosphérique, ce qui avait également été suggérée par une précédente étude (Chappellaz et al., 1997). Ainsi, l'apport en GES par les lacs de thermokarst ne date pas d'hier... De nos jours, le même phénomène semble se reproduire peu à peu. On estime que d'ici 60 ans, les lacs de fonte pourrait émettre soit par ébullition ou par diffusion jusqu'à 5.3 ± 2.8 Tg de CH₄ par année (Wik et al., 2016). Déjà, plusieurs études ont mesuré de hautes émissions en CH₄ de la part de ces lacs partout dans l'Arctique (Matheus Carnevali et al., 2015; Matveev et al., 2016). De plus, ces lacs sont la plupart du temps hyper saturés en CO₂ et en CH₄, démontrant à quel point ils constituent des sources potentielles de GES à considérer dans les prochains modèles climatiques (Breton et al., 2009).

À la base de cette production, on retrouve principalement des processus biologiques (Borrel et al., 2011). En effet, comme mentionné précédemment, le fond des lacs de thermokarst est généralement hypoxique ou simplement anoxique. De ce fait, les conditions y sont propices pour la croissance de micro-organismes pratiquant des métabolismes anaérobies, telle que la méthanogénèse. Le méthane produit en profondeur diffuse dans la colonne d'eau, et une partie est métabolisée à la surface par des micro-organismes aérobies de type méthanotrophe. Ces derniers consomment le méthane pour

relâcher du CO₂ qui s'accumule graduellement dans le lac. Éventuellement, ces deux gaz sont émis vers l'atmosphère par diffusion ou ébullition. Avec la hausse des températures et la fonte accélérée du pergélisol, l'activité microbienne est de plus en plus stimulée par l'apport continu de carbone organique auparavant séquestré dans le sol gelé, ce qui entraîne une augmentation de la production de CH₄ et de CO₂ (Negandhi et al., 2016). Jumelée à leur multiplication dans toutes les régions arctiques, la contribution des lacs de thermokarst aux émissions de GES augmente d'année en année et est désormais non-négligeable (Wik et al., 2016).

1.1.4 Écologie des lacs de thermokarst

Jusqu'à maintenant, peu d'études ont porté sur les communautés microbiennes vivant dans les lacs de fonte de pergélisol. Des recherches menées dans des lacs en régions arctique et subarctique ont démontré la richesse de ces écosystèmes en micro-organismes impliqués dans le cycle du méthane (Crevecoeur et al., 2015; Hultman et al., 2015). Les bactéries méthanotrophes (e.g. *Methylobacter*) deviennent parfois très dominantes dans la colonne d'eau avec des abondances relatives pouvant atteindre 25% dans certains échantillons de surface. En effet, des membres des ordres *Methylococcales* et *Methylacidiphilales* ont été identifiés dans tous les échantillons testés. Des archées méthanogènes, telles que *Methanoregula*, *Methanosaeta*, *Methanomicrobia* et *Methanobacterium*, ont été détectées au fond des lacs de thermokarst ainsi que des transcrits associés à la méthanogénèse (*mcrA*) (Crevecoeur et al., 2016; Mondav et al., 2014). Ceci peut paraître étonnant puisque, dans les écosystèmes d'eau douce la méthanogénèse est généralement limitée aux sédiments en raison de la possibilité de retrouver de l'oxygène dans la colonne d'eau, ce qui inhiberait la réaction (Briee et al., 2007). Par contre, la forte stratification en oxygène et l'absence de mélange dans les lacs de fonte permet la prolifération de ce type d'archée dans la couche de fond.

Les taxons bactériens qui dominent la colonne d'eau en période estivale sont : *Proteobacteria* (majoritairement de la classe des β -proteobactéries), *Actinobacteria*, *Verrucomicrobia*, *Bacteroidetes*, *Chlorobi* et *Cyanobacteria* (Comte et al., 2016b; Rossi et al., 2013). Des identifications taxonomiques plus raffinées ont permis d'identifier *Polynucleobacter*, *ACK-M1* et *Comamonadaceae* comme étant les micro-organismes les plus répandus à travers les lacs de thermokarst du subarctique canadien. Bien entendu, les conditions physico-chimiques de chaque lac déterminent la composition des

assemblages microbiens à l'intérieur de la colonne d'eau. Une hausse de la turbidité causée par de grandes concentrations en matière organique dissoute colorée (CDOM) désavantage la croissance des cyanobactéries, ou d'autres bactéries photosynthétiques, puisqu'il y a une diminution de la pénétration de la lumière (Rossi et al., 2013). Des facteurs biotiques et abiotiques (DOC, phosphore, chlorophylle a) entraînent également des modèles de cooccurrence dans les populations bactériennes (Comte et al., 2016a). Comme mentionné précédemment, les conditions oxiques et anoxiques amènent une différenciation des communautés dominantes à la surface et fond des lacs de fonte (Rossi et al., 2013). Or, une étude a démontré que la composition bactérienne des lacs est plus similaire intra-vallée qu'inter-vallées (Comte et al., 2016b). Les paises ont notamment des assemblages microbiens différents des lithaleses. En hiver, lorsque que la colonne d'eau devient complètement anoxique, un changement complet des taxons dominants peut être envisagé.

En été, les populations de zooplancton vivant dans les lacs de thermokarst peuvent être très diverses et atteindre des concentrations de 2 à 5 fois plus élevées que dans d'autres lacs arctiques en raison de l'absence de poisson (Rautio et al., 2011). Les espèces dominantes de macro zooplancton sont généralement *Cladocera* (*Daphnia* sp.) et *Copepoda* (*Leptodiaptomus minutus*, *Mixodiaptomus laciniatus*). Une étude réalisée dans le subarctique canadien a aussi démontré que les rotifères occupaient une grande partie du zooplancton en comptant pour 59% de la biomasse (Bégin and Vincent, 2017). Les taxons les plus répandus étaient *Polyarthra* sp., *Keratella* sp., *Anuraeopsis fissa* et *Conochilus unicornis*. L'analyse de l'ARNr 18S a aussi révélé que les ciliés sont parmi les organismes les plus abondants à la surface des lacs de fonte (Przytulska et al., 2016). De plus, plusieurs espèces d'eucaryotes unicellulaires autotrophes habitent les lacs de thermokarst, telles que des dinoflagellés, des chrysophytes et des chlorophytes (Przytulska et al., 2016). Dans une moindre proportion, on retrouve des cryptophytes et des diatomées, ces dernières étant défavorisées par le manque de mélange dans la colonne d'eau. Pour survivre aux conditions extrêmes en Arctique, certaines espèces ont développé des mécanismes astucieux : ajustement de la pigmentation (mélanine) chez les daphnies pour résister aux rayons UV (Hessen, 1996), production de cyste (équivalent des spores bactériennes) chez les chrysophytes (Firsova et al., 2008), et adoption d'un mode d'alimentation mixotrophe où la phagotrophie (hétérotrophie) et la photosynthèse (autotrophie) peuvent être pratiquées (Przytulska et al., 2016).

1.2 Les virus

Étant des parasites obligatoires, les virus ont besoin d'une cellule hôte (eucaryote, procaryote, archée) afin de pouvoir se répliquer. Sans hôte, un virus n'est qu'une particule inerte n'ayant aucune activité métabolique ou biochimique. En règle générale, les virus sont des particules extrêmement simples constituées d'une capsid de nature protéique renfermant un court génome composé d'ADN ou d'ARN simple (sb) et/ou double brin (db) (King et al., 2011). En ce qui concerne les génomes d'ARN sb, ils peuvent être soit de polarité négative (antisens) ou positive (sens). Un génome de polarité positive peut immédiatement être traduit par la machinerie cellulaire de l'hôte, tandis qu'un génome de polarité négative doit préalablement être converti en ARN sens par une enzyme virale, l'ARN polymérase ARN-dépendante. Les génomes viraux sont habituellement très petits (<500 kb) et contiennent un minimum de gènes nécessaires pour leur multiplication (King et al., 2011). Par contre, des virus possédant des génomes ayant des tailles comparables à ceux des bactéries ont été découverts dans les dernières années (Abergel et al., 2015). Le Pandoravirus étant celui possédant le plus grand génome, avec une longueur pouvant atteindre de 1,9 à 2,7 Mb (Philippe et al., 2013). Les virus ont des morphologies extrêmement variées. Par exemple, certains virus infectant exclusivement les bactéries (bactériophages, ou simplement phages) peuvent présenter une morphologie complexe en étant munis d'une queue et de fibres caudales. D'ailleurs, ces derniers ont été regroupés au sein de l'ordre des *Caudovirales* et classifiés en fonction de la taille de leur queue et de la capacité de celle-ci à se contracter (*Myoviridae*, *Podoviridae*, *Siphoviridae*) (Ackermann, 1999). D'autres virus infectant les procaryotes et les eucaryotes ont quant à eux une morphologie beaucoup plus simple en étant doté seulement d'une capsid de forme géométrique (*Microviridae*, *Paramyxoviridae*, *Adenoviridae*, etc.). La grande majorité des virus ont une taille inférieure à 250 nm (King et al., 2011). Or, il existe tout de même de plus gros virus appelés *virus géants*. Par exemple, parmi les virus infectant les amibes on retrouve les *Mimiviridae* et le *Pithovirus* qui peuvent atteindre des tailles impressionnantes de 400 nm et 1,5 µm respectivement, les rendant visible au microscope optique (La Scola et al., 2003; Legendre et al., 2014). Le cas du *Pithovirus* est particulièrement intéressant puisqu'il a été isolé à partir d'un échantillon de pergélisol sibérien datant de plus de 30 000 ans. Ainsi, cette découverte suggère que la fonte du pergélisol entraînée par le réchauffement climatique pourrait potentiellement «ramener à la vie» des virus préhistoriques prisonniers du sol...

Il est largement reconnu que les virus sont les entités biologiques les plus abondantes et diversifiées sur la planète (Rohwer et al., 2009; Suttle, 2005). En effet, on estime le nombre de particules virales sur Terre à 10^{31} , dont une majorité de phages (Rohwer et al., 2009). Seulement dans l'océan, on décompte de 10^6 - 10^7 particules virales par mL d'eau marine (Bergh et al., 1989). Comme ils ont la capacité d'infecter tous les organismes vivants connus, les virus sont également considérés ubiquitaires de tous les environnements, mêmes des plus extrêmes (Danovaro et al., 2016; Rachel et al., 2002).

1.2.1 Rôles écologiques

Il est désormais indéniable dans la communauté scientifique que les virus possèdent des rôles essentiels clés dans tous les écosystèmes (Rohwer et al., 2009; Roux et al., 2016). Dans les environnements aquatiques, les virus ont une importance capitale en agissant directement sur les cycles biogéochimiques et différents processus biologiques (Suttle, 2007). Pour ce faire, ils ont recours à diverses stratégies et de nombreuses théories écologiques existent pour tenter d'expliquer les dynamiques hôtes-virus.

1.2.1.1 Court-circuit viral

Dans les écosystèmes aquatiques, on retrouve à la base de la chaîne alimentaire un processus microbien appelé la boucle microbienne («*microbial loop*») qui a une importance majeure dans le recyclage des nutriments et de la matière organique dissoute et particulaire (DOM et POM) (Azam et al., 1983; Fenchel, 2008). Cet ensemble de micro-organismes comprend entre autres des hétérotrophes ainsi que des autotrophes (micro-algues et cyanobactéries) responsables de l'assimilation primaire de la matière organique (carbone) dans la chaîne alimentaire aquatique. Ces organismes sont d'ailleurs regroupés sous l'appellation de «producteurs primaires». Or, en venant infecter et lyser les cellules participant à la boucle microbienne, les virus exécutent le court-circuit viral (Wilhelm and Suttle, 1999). De cette manière, la matière organique ne passe pas par les niveaux trophiques supérieurs mais demeure uniquement disponible pour les micro-organismes hétérotrophes au bas de la chaîne alimentaire, d'où le terme de court-circuit (Figure 1.4) (Poorvin et al., 2004; Weitz and Wilhelm, 2012). Les virus ont donc une influence directe sur les flux de carbone et de nutriments à travers les niveaux trophiques.

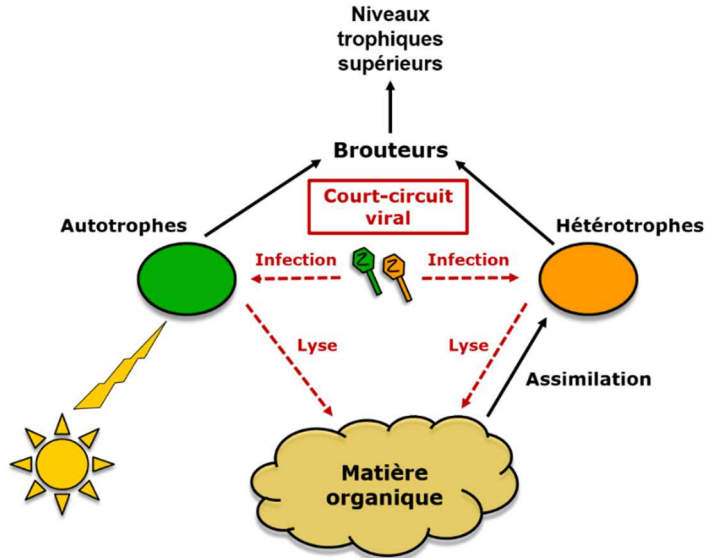


Figure 1.4 : Schématisation du court-circuit viral. En infectant les organismes à la base de la chaîne alimentaire, les virus participent au recyclage direct de la matière organique.

1.2.1.2 «Tuer le gagnant» (*Kill-the-winner*)

La théorie écologique *kill-the-winner* explique la dynamique hôte-virus dans un contexte où un micro-organisme (hôte) devient soudainement plus abondant dans un milieu (Thingstad and Lignell, 1997). Dans ce modèle, les virus préviennent qu'une seule espèce microbienne prenne le dessus au sein d'une communauté et ne monopolise tous les nutriments de l'écosystème. En effet, ce mécanisme est basé sur la spécificité hôte-virus, c'est-à-dire que lorsqu'un micro-organisme gagne rapidement en abondance, cela déclenche une production positive nette d'un virus infectant spécifiquement ce dernier (Maslov and Sneppen, 2017). Dans une communauté où la cellule hôte est en forte abondance, le taux de contact entre un virus et son hôte augmente, ce qui facilite la propagation du virus. Combiné à l'action des brouteurs, les virus participent ainsi à la régulation de la taille et au contrôle de la diversité des communautés microbiennes aquatiques. De cette manière, l'équilibre biologique de l'écosystème est maintenu.

1.2.1.3 «Reine Rouge» (*Red Queen*)

Selon cette théorie, les virus et leurs hôtes seraient impliqués dans une éternelle course à l'armement basée sur l'évolution (Van Valen, 1973). Chacun visant à s'adapter à son adversaire afin de pouvoir soit mieux infecter son hôte, soit mieux contrer l'infection. L'interaction de la proie et du prédateur

entraîne donc une coévolution continue des deux organismes, menant à une diversification constante de la biosphère (Comeau and Krisch, 2005). Divers exemples de cette coévolution ont pu être répertoriés dans la nature.

Les cellules bactériennes ont développé une panoplie de systèmes de défense visant à combattre les infections virales (Stern and Sorek, 2011). Pour n'en nommer que quelques-uns, il existe les systèmes *Abi* («*abortive infection*»), de restriction-modification et bien sûr, présentement le plus populaire d'entre tous, CRISPR-Cas («*clustered regularly interspaced short palindromic repeat*»). Ces mécanismes ne sont pas infaillibles étant donné qu'inévitablement un phage deviendra résistant, notamment en mutant son génome ou modifiant ses bases azotées, ce qui empêchera sa reconnaissance par le système de défense. De cette manière, la cellule est forcée de s'adapter à nouveau afin de pouvoir survivre.

Les eucaryotes unicellulaires ont eux aussi acquis au cours de l'évolution divers mécanismes de défense. Par exemple, la micro-algue *Emiliana huxleyi* possède une métacaspase pouvant enclencher la mort cellulaire programmée lors d'une infection virale (Sanmartin et al., 2005). Or, une étude a démontré que le virus EhV1 (*E. huxleyi* virus 1) recrutait cette protéase pour faire partie intégrante de son cycle de multiplication (Bidle et al., 2007). Si la production de cette enzyme était inhibée, le rendement de l'infection virale était significativement réduit. En effet, le virus se sert de cette protéase pour faciliter la lyse cellulaire à la fin de son cycle. Cette interaction hôte-virus illustre parfaitement comment les virus ont pu évoluer de manière fascinante pour tourner à leur avantage un système de défense développé par leur hôte.

Récemment, un système analogue au système CRISPR-Cas a été identifié chez les virus de la famille des *Mimiviridae*, soit le MIMIVIRE («*mimivirus virophage resistance element*») (Levasseur et al., 2016). Il existe trois lignées à l'intérieur de cette famille (A, B, C), mais seule la lignée A est résistante au virophage Zamilon. En investiguant le génome des souches appartenant cette lignée, une région de 28 nucléotides ayant une forte homologie avec le virophage a été découverte, et une séquence de 15 nucléotides dérivée de cette région était répétée 4 fois de suite, ces répétitions étant séparées de quelques nucléotides seulement. De plus, légèrement en aval de cette région, des cadres de lecture ouverts (ORFs) analogues aux gènes *cas* ont été identifiés (hélicase, nucléase, RNase III, etc.). Toute

cette région étant unique aux Mimivirus de la lignée A, les chercheurs ont estimé qu'elle avait un rôle essentiel dans la résistance contre Zamilon. Ils ont confirmé leur hypothèse en empêchant l'expression de ce locus, ce qui a permis la propagation du virophage. Ainsi, cette nouvelle découverte démontre que même les virus, lorsqu'ils font office de proies, peuvent développer des mécanismes de défense et suivre la théorie de la «Reine Rouge».

1.2.1.4 «Chat du Cheshire» (*Cheshire Cat*)

À ce jour, ce phénomène n'a été observé que chez *Emiliana huxleyi*, une espèce de coccolithophore (eucaryote unicellulaire photosynthétique) (Frada et al., 2008). Cette espèce est l'algue la plus abondante dans les océans de la planète (Tyrrell and Taylor, 1996). *E. huxleyi* est donc ubiquitaires des milieux marins excepté dans les régions polaires en raison des basses températures. La cellule d'*E. huxleyi* est protégée par une couche de plusieurs plaques de calcite en forme de disque appelées coccolithes (Paasche, 2001). À la mort de la cellule, l'enveloppe de calcite subsiste et sédimente au fond des océans. L'accumulation de milliards d'enveloppes va ainsi former de la craie, une roche sédimentaire de calcaire composée principalement de carbonate de calcium (CaCO_3). Cette capacité d'*E. huxleyi* démontre son importance écologique puisque les roches calcaires représentent des puits de carbone essentiels dans le cycle naturel du carbone (Godoi et al., 2009).

La théorie du *Cheshire Cat* décrit la stratégie utilisée par certaines cellules d'*E. huxleyi* pour échapper aux infections virales. Dans le cycle de vie d'*E. huxleyi*, il existe deux phases distinctes où le phénotype de la cellule change radicalement. La première phase consiste en une cellule diploïde non-motile contenue dans une enveloppe de calcite, et capable de former des efflorescences. Quant à la deuxième phase, celle-ci consiste en une cellule haploïde nue dotée d'un flagelle lui conférant de la motilité. Or, l'étude de Frada et al (2008) a pu démontrer que la présence de virus nommés *Emiliana huxleyi virus* (EhV) induisait chez l'algue une transition de la phase diploïde vers la phase haploïde. La phase haploïde n'étant pas reconnue par les EhV, ce mécanisme de transition rend donc la cellule résistante au virus. Il est supposé que l'absence de l'enveloppe de coccolithes à la surface des cellules haploïdes serait à la base de cette résistance. En effet, dans le cas des cellules haploïdes, le récepteur spécifique utilisé par le virus lors de l'attachement serait soit absent ou plus difficilement accessible. Ainsi, cette stratégie unique en son genre développée par *E. huxleyi* révèle l'importance des

interactions virus-hôte dans l'environnement, et que la «course aux armements» peut prendre littéralement plusieurs formes.

1.2.1.5 Transfert horizontal de gènes

Il est largement reconnu que le phénomène de transfert horizontal de gènes est parmi les facteurs majeurs qui dirigent l'évolution de toutes les espèces, surtout celle des bactéries (Keeling, 2009; Ochman et al., 2000; Polz et al., 2013). Ce processus est également appelé transduction lorsqu'un phage joue le rôle d'intermédiaire (Birge, 2006). En passant d'une cellule à une autre, les virus peuvent périodiquement transférer un fragment d'ADN ou un gène entier d'une cellule à une autre. La plupart du temps, il s'agit de gènes délétères qui sont éventuellement éliminés par sélection naturelle. Or, dans une moindre proportion, il arrive que le gène confère un avantage écologique ou métabolique à la cellule hôte. Le gène est alors conservé dans la cellule et transmis du même coup aux descendants. Il existe deux types de transduction basés sur le mécanisme dont le segment d'ADN cellulaire est acquis par le virus (Birge, 2006). Le transfert de gènes peut s'effectuer au moment de l'empaquetage des acides nucléiques où un fragment d'ADN cellulaire est également inclus dans la capsid (transduction généralisée). Le transfert peut également avoir lieu lors de l'excision du virus qui s'était préalablement intégré au génome de son hôte (transduction spécialisée). Dans ce dernier cas, il est nécessaire que le fragment transféré soit adjacent au prophage. De cette manière, un virus peut gagner de nouvelles caractéristiques, et par la suite, les propager à travers une population microbienne lors de l'infection virale (Sullivan et al., 2003; Wu and Zhang, 2011). En conséquence, des populations entières de bactéries, d'archées et d'eucaryotes peuvent acquérir de nouvelles capacités métaboliques ou encore étendre leur niche écologique grâce aux virus (Finke et al., 2017; Pierce et al., 1999). Le phage modèle utilisé pour l'étude de la transduction est le phage lysogénique lambda.

Il faut également noter l'existence d'un agent ayant une morphologie très similaire aux virus, mais de nature non-virale, pouvant également effectuer le transfert horizontal de gènes : les agents de transfert de gènes («*gene transfer agent*», GTA) (Lang et al., 2012). Un GTA contient un segment aléatoire du génome de sa cellule productrice, mais la quantité d'ADN qu'il renferme est insuffisante pour pouvoir coder pour ses propres protéines. Lors de la lyse cellulaire, les GTAs peuvent se propager dans l'environnement et ainsi transférer leur court fragment d'ADN à d'autres cellules réceptrices. Leur

abondance et les mécanismes par lesquels ils agissent demeurent peu connus, malgré qu'il y a quelques indices laissant croire qu'ils auraient un impact majeur dans les océans (McDaniel et al., 2010). *Rhodobacter capsulatus* et *Bartonella* sont parmi les bactéries les plus étudiées pour leur production de GTAs (Hynes et al., 2016; Quebatte et al., 2017). Compte tenu de l'importance du transfert horizontal de gènes dans l'évolution des espèces, il est probable que les GTAs occupent un rôle essentiel dans ce processus.

1.2.1.6 Gènes auxiliaires métaboliques

Comme mentionné précédemment, il a été démontré que certains virus, et particulièrement les cyanophages (virus infectant spécifiquement les cyanobactéries), transportent dans leur génome des gènes pouvant apporter un ou plusieurs avantages métaboliques à leur cellule hôte au moment de l'infection virale (Crummett et al., 2016). Le terme gène auxiliaire métabolique (AMG) a alors été suggéré pour désigner ce groupe de gènes (Breitbart et al., 2007). À l'origine, les AMGs ont été acquis par les virus à partir du génome de leur hôte par différents mécanismes tel que le transfert horizontal (Lindell et al., 2007; Mann et al., 2003).

On compte parmi les AMGs, des gènes impliqués dans divers processus cellulaires tels que la photosynthèse, le métabolisme du carbone, la synthèse et le métabolisme des acides nucléiques et la tolérance aux stress (Enav et al., 2014; Sullivan et al., 2005; Thompson et al., 2011). Ces gènes ont pour fonction de stimuler le métabolisme de la cellule puisque lors de l'infection virale, la transcription et la traduction des protéines cellulaires sont interrompues. Il est donc essentiel pour le virus de notamment prolonger et maintenir la photosynthèse durant l'infection afin de pouvoir compléter son cycle lytique (Lindell et al., 2005). Par exemple, la grande majorité des cyanophages infectant *Prochlorococcus* et *Synechococcus*, les deux espèces de cyanobactéries les plus abondantes dans les océans, transportent les gènes *psbA* et *psbD* qui codent respectivement pour les protéines D1 et D2 du noyau du photosystème II (Sullivan et al., 2006). Le gène *hli* («*high-light inducible*») est également souvent porté par les cyanophages (Lindell et al., 2004). Ainsi, il est estimé que le rendement des gènes de photosynthèse d'origine phagique génère suffisamment d'énergie pour la production de virions. De plus, il est évalué que 60% des gènes *psbA* identifiés à la surface des océans dans le cadre du projet international *Global Ocean Survey* seraient d'origine virale (Sharon et al., 2007).

D'autres gènes reliés à la photosynthèse, mais moins conservés que *psbAD*, ont sporadiquement été retrouvés chez plusieurs cyanophages (Dammeyer et al., 2008; Mann et al., 2005; Sullivan et al., 2010). Ces gènes sont notamment impliqués dans la biosynthèse de la phycobilline (*pebS*, *pcyA*, *ho1*) ou encore dans la chaîne de transport des électrons (*petE*, *ptox*). Par contre, le rôle du gène *cpeT* demeure inconnu chez les cyanophages, compte tenu que le reste de l'opéron (*cpeESR*) est absent du génome viral.

Concernant le métabolisme du carbone et des acides nucléiques, les AMGs portés par les virus sont généralement *talC* (transaldolase) qui faciliterait la mobilisation des réserves de carbone en période de noirceur ou encore CP12 qui inhibe le cycle de Calvin et favorise la voie des pentoses phosphates (PPP) et la biosynthèse de dNTPs (Sullivan et al., 2005; Thompson et al., 2011). Certains phages de *Synechococcus* transportent même deux gènes (*zwf* et *gnd*) codant pour des enzymes impliquées directement dans le PPP, ce qui suggère l'importance de cette voie métabolique dans la réplication des cyanophages (Sullivan et al., 2010). De cette manière, les virus ont la capacité de détourner à leur avantage le métabolisme du carbone dans la cellule. En effet, la quantité de dNTPs est souvent limitante lors de l'infection, ce qui résulte en une taille de la progéniture («*burst size*») plus petite. En aidant la cellule hôte à produire plus de dNTPs lors de l'infection, le virus peut donc augmenter la taille de sa progéniture (Brown et al., 2006). D'ailleurs, une étude a posé l'hypothèse que les cyanomyovirus posséderaient une cassette contenant quatre AMGs appartenant au métabolisme du carbone (*zwf*, *gnd*, *petE* et *ptox*) (Sullivan et al., 2010). Cette dernière serait retrouvée dans une région hypervariable du génome, entre les gènes *g15* et *g18* (Millard et al., 2009).

Certains cyanophages vont également transporter des AMGs permettant une meilleure tolérance aux stress environnementaux. Entre autres, on retrouve des gènes appartenant au métabolisme du phosphate (Sullivan et al., 2005). Le phosphate est essentiel au bon développement de tous les organismes, mais est souvent en concentration limitante dans les océans. Il est donc commun que les organismes vivant en milieu marin tels que les cyanobactéries subissent un stress en raison du manque de phosphate. Les virus en ont également besoin en grande concentration pour se répliquer. Ainsi, certains cyanophages possèdent le gène *pstS*, qui chez les bactéries code pour une protéine périplasmique liant le phosphate, favorisant du même coup son absorption par la cellule. D'autres phages infectant *Synechococcus* et *Prochlorococcus* contiennent également une phosphatase alcaline

(*phoA*) près de *pstS* (Sullivan et al., 2010). Cette enzyme permettrait de faciliter l'accès au phosphore organique. De ce fait, l'expression chez les phages de ces gènes lors de l'infection virale mènerait potentiellement à une augmentation de l'absorption du phosphate si les concentrations devenaient limitées. Il existe un autre AMG couramment retrouvé parmi les cyanophages, soit le gène *mazG* codant pour une pyrophosphatase. Or, son rôle précis lors de l'infection virale n'a pas encore été élucidé, tandis que chez *Escherichia coli*, cette enzyme est impliquée dans la diminution des taux de guanosine pentaphosphate (ppGpp). Le ppGpp est produit en conditions de carence nutritionnelle et sert à réguler à la hausse les gènes liés au métabolisme et à la biosynthèse d'acides aminés (Gross et al., 2006; Traxler et al., 2008).

Deux autres AMGs sont également largement répandus chez les cyanophages, soit un codant pour une cobalamine synthétase (*cobS*) et un autre codant pour une protéine de choc thermique (Crummett et al., 2016). En réalité, il existe plusieurs autres AMGs, mais ceux-ci sont conservés dans une moins grande mesure chez les virus séquencés à ce jour (Gao et al., 2016). En effet, les données de métaviromes récoltés lors de la *Global Ocean Expedition* ont démontré l'abondance et la grande diversité des AMGs en milieu marin (Williamson et al., 2008).

1.2.1.7 Contrôle des populations microbiennes

Un des exemples les plus concrets concernant le contrôle des populations par les virus est leur impact lors d'efflorescences algales. En conditions propices, il est commun que certains organismes photosynthétiques (cyanobactéries, micro-algues) créent des efflorescences en milieu aquatique (eau douce, salée et saumâtre), c'est-à-dire que la taille de leur population subit une augmentation importante, rapide, et soudaine. Ces efflorescences peuvent atteindre une ampleur impressionnante, et même être visibles depuis l'espace en raison de la grande concentration de pigments photosynthétiques. Or, ce phénomène n'est pas toujours bénéfique et peut nuire aux écosystèmes où il se produit et par conséquent à la santé publique (Grattan et al., 2016). En effet, les efflorescences algales peuvent être néfastes pour diverses raisons, telles que la libération de toxines (phyco/cyanotoxine) dans l'eau ou encore la consommation en totalité de l'oxygène dissout menant à une hypoxie de la colonne d'eau (Landsberg, 2002). Ainsi, les cas extrêmes de débalancement de l'écosystème conduisent parfois à la création d'une «zone morte» où la libération de toxine et/ou

l'absence d'oxygène entraînent la mort de plusieurs organismes aquatiques (poissons, crustacés, mollusques, coraux, etc.) (Paerl and Otten, 2013).

Le rôle des virus dans ces efflorescences algales est, comme proposée par la théorie *kill-the-winner*, d'infecter le micro-organisme dominant. Dans ce cas-ci, l'organisme ciblé serait le micro-organisme responsable de l'efflorescence. À cet effet, plusieurs études ont démontré l'effet de la lyse virale pour mettre fin à des efflorescences algales partout dans le monde (Brussaard et al., 1996; Fujimoto et al., 2013; Lawrence et al., 2001). Il est certain que dans un tel contexte il n'y a pas que les virus qui participent au contrôle par le haut («*top down control*»). Les organismes brouteurs (protozoaires, zooplanctons) et les bactéries produisant des composés algicides occupent également une place importante dans ce processus (Cram et al., 2016; Skerratt et al., 2002). Par contre, ces derniers peuvent être plus facilement affectés négativement par l'effet des toxines ou l'hypoxie (Buskey and Hyatt, 1995; Roncalli et al., 2016).

1.2.2 Les virus infectant le phytoplancton

Comme mentionné à la section «1.2.1.1 *Cour-circuit viral*», les virus infectant les producteurs primaires (micro-algues et cyanobactéries) influencent directement le recyclage des nutriments et de la matière organique au bas de la chaîne alimentaire. C'est donc dire que ce groupe de virus a un impact significatif dans les écosystèmes aquatiques. Parmi ceux-ci, on retrouve les virus faisant partie de la famille des *Phycodnaviridae* qui infectent une panoplie de micro-algues d'eau douce et d'eau marine, ainsi que les virus infectant les cyanobactéries, regroupés sous l'appellation cyanophages.

1.2.2.1 Les *Phycodnaviridae* - *Chlorovirus*

Les virus de la famille *Phycodnaviridae* possèdent tous un long génome à ADN db, et sont compris à l'intérieur de la super-famille des grands virus nucléocytoplasmiques (NCLDV). D'ailleurs, ce groupe est présenté comme le nouvel ordre des *Megavirales*, qui reste encore à être accepté par l'International Committee on Taxonomy of Viruses (ICTV). Actuellement, les *Phycodnaviridae* contiennent six genres : *Chlorovirus*, *Coccolithovirus*, *Prasinovirus*, *Phaeovirus*, *Prymnesiovirus*, et *Raphidovirus* (Wilson et al., 2011). Or, c'est le genre *Chlorovirus* qui possèdent le plus grand nombre de membres

dont la classification taxonomique a été acceptée par l'ICTV, soit 19. À ce jour, on dénombre 43 génomes complets de chlorovirus annotés (Van Etten et al., 2017). Tous les Chlorovirus découverts jusqu'à maintenant ont pour hôte une micro-algue unicellulaire appartenant à l'embranchement des *Chlorophyta* et au genre *Chlorella* («*chlorella-like alga*»). Ces cellules sont ubiquitaires des milieux d'eau douce, et peuvent exister sous forme individuelle ou de symbionte de protozoaires (cilié, hélizoaire, hydre) (Kang et al., 2005). À l'état de symbionte, l'algue est résistante à l'infection virale puisque la cellule est protégée par son hôte (Kawakami and Kawakami, 1978; Reisser et al., 1991). Or, il est essentiel de la cultiver individuellement afin de pouvoir étudier les chlorovirus en laboratoire. Quatre souches de *Chlorella* ont été identifiées comme hôte potentiel et sont utilisées à cette fin: *Chlorella variabilis* NC64A, *Chlorella variabilis* Syngen 2-3, *Chlorella heliozoae* SAG 3.83 et *Micractinium conductrix* (auparavant appelée *Chlorella* Pbi). L'espèce modèle de ce genre viral est *Paramecium bursaria chlorella virus-1* (PBCV-1) (Van Etten et al., 1982). Ce virus possède un génome linéaire de 330 744 pb conservé à l'intérieur d'une capsidie icosaédrique de 190 nm qui recouvre également une membrane lipidique nécessaire à l'infection (Yan et al., 2000). Une analyse protéomique a d'ailleurs démontré que PBCV-1 code pour 148 protéines virales et au moins une protéine de l'hôte lors de l'infection (Dunigan et al., 2012). Parmi les autres chlorovirus, on retrouve également *Acanthocystis turfacea chlorella virus 1* (ATCV-1) et *Hydra viridis chlorella virus 1* (Bubeck and Pfitzner, 2005; Meints et al., 1981). Un court résumé de ces trois virus est présenté au Tableau 1.1. Malgré qu'il n'existe pas de gène marqueur universel pour tous les virus, il est considéré que le gène de l'ADN polymérase, et plus particulièrement la région codant pour le site actif (motif YGDTDS), est fortement conservée chez les virus de la famille des *Phycodnaviridae*. Cette caractéristique permet donc de l'utiliser comme cible lors de la réaction de polymérase en chaîne (PCR) (Chen and Suttle, 1995). Dans les dernières années, une équipe a optimisé de nouvelles amorces dégénérées afin de viser spécifiquement les populations de chlorovirus dans des milieux d'eau douce en ciblant leur polymérase (*polB*) (Short et al., 2011). Par contre, l'écologie des chlorovirus demeure peu étudiée globalement malgré l'existence d'amorces spécifiques et l'ubiquité de leurs cellules hôtes dans l'environnement.

Tableau 1.1 : Principaux chlorovirus selon le symbionte de leur hôte (*Chlorella sp.*)

Virus	Hôte utilisée en laboratoire	Taille du génome (pb)	Référence
<i>Paramecium bursaria chlorella virus-1</i>	<i>Chlorella</i> NC64A	330 744	(Van Etten et al., 1982)
<i>Acanthocystis turfacea chlorella virus 1</i>	<i>Chlorella</i> SAG 3.83	288 047	(Bubeck and Pfitzner, 2005; Fitzgerald et al., 2007)
<i>Hydra viridis chlorella virus 1</i>	<i>Chlorella</i> NC64A	ND	(Meints et al., 1981)

1.2.2.2 Les cyanophages

Les cyanophages réfèrent à tous les phages infectant les cyanobactéries. La majorité des cyanophages répertoriés appartiennent à l'ordre des *Caudovirales*, et possèdent donc une capsidie icosaédrique munie d'une courte (*Podoviridae*) ou longue queue, contractile (*Myoviridae*) ou non (*Siphoviridae*) (Chenard et al., 2015; Coloma et al., 2017; Yoshida et al., 2006). Il est reconnu que les cyanophages sont très abondants en milieu aquatique et incroyablement riche sur le plan génétique avec la présence de multiples AMG dans leur génome (Gao et al., 2016; Matteson et al., 2013). Ces AMG suggèrent qu'il y aurait eu plusieurs évènements de transferts horizontaux entre les cyanobactéries et leur virus par le passé (Lindell et al., 2007). Les gènes *phoH* et *psbA* d'origine phagique ont notamment été utilisés à titre de gène marqueur afin d'étudier la diversité des cyanophages dans les écosystèmes d'eau douce et d'eau salée (Chenard and Suttle, 2008; Goldsmith et al., 2011). Il existe également d'autres gènes signatures utilisés pour cibler les cyanophages en biologie moléculaire: *gp23* codant pour la protéine majeure de la capsidie et *gp20* codant pour la protéine de portail des cyanomyovirus (Chow and Fuhrman, 2012; Filee et al., 2005; Zhong et al., 2002), et l'ADN polymérase des cyanopodovirus (Dekel-Bird et al., 2013). Pour l'instant, on peut observer un biais dans les bases de données en raison de la surreprésentation des phages de *Synechococcus* et de *Prochlorococcus*, les deux espèces de cyanobactéries les plus abondantes dans les océans. Pourtant, le cyanopodovirus SEIV-1 isolé d'un lac du Haut-Arctique a démontré une grande divergence génétique par rapport aux autres cyanophages recensés, suggérant l'importance écologique encore à découvrir des cyanophages d'eau douce (Chenard et al., 2015). Il apparaît donc nécessaire de réaliser davantage d'études pour approfondir nos connaissances sur la diversité des cyanophages dans les écosystèmes d'eau douce.

1.2.3 Écologie virale dans les écosystèmes d'eau douce polaires

Les environnements aquatiques polaires constituent, pour la plupart, des écosystèmes ultra-oligotrophes, c'est-à-dire très pauvre en nutriments, et sont généralement dominés par les organismes microbiens (Nuttall, 2005). De plus, étant donné les basses températures, les micro-organismes psychrophiles occupent une grande partie des populations. Or, dans les dernières années, des changements dans la composition des populations microbiennes arctiques ont été observés en raison des changements climatiques (Comeau et al., 2011). Cela indique que les populations microbiennes sont sensibles aux moindres modifications physico-chimiques entraînées par un réchauffement de l'eau, et qu'il pourrait y avoir des répercussions sur la chaîne alimentaire aquatique (Boras et al., 2010). Il est donc essentiel d'effectuer un suivi des communautés microbiennes polaires, et ce, particulièrement dans un contexte de réchauffement climatique.

Diverses études ont porté sur la composition des populations virales dans les écosystèmes d'eau douce en milieu tempéré (Clasen and Suttle, 2009; Wang et al., 2015; Zhong et al., 2014), mais beaucoup moins se sont penchés sur l'identité des virus en région polaire (Anesio et al., 2007; de Carcer et al., 2016; Lopez-Bueno et al., 2009). L'étude de Lopez-Bueno et al. (2009) s'est intéressée aux virus dans un lac d'eau douce situé en Antarctique, et a découvert une incroyable diversité génétique plus importante que dans tous les autres viromes provenant d'environnement aquatique. De plus, cette étude a aussi obtenu des assemblages viraux dominés étonnamment par des virus infectant les cellules eucaryotes, ce qui est inhabituel puisqu'en général ce sont les phages qui sont les plus abondants dans les échantillons environnementaux. Par contre, une étude effectuée dans des lacs subarctiques suédois est venue à la conclusion que les virus, malgré leur grande abondance, n'avaient qu'un impact mineur sur les communautés bactériennes benthiques et pélagiques (Sawstrom et al., 2009). D'un autre côté, les virus présents dans les trous de cryoconites à la surface des glaciers auraient un rôle majeur dans le recyclage du carbone et des nutriments dans les écosystèmes proglaciaires arctiques (Bellas et al., 2013). Malgré les conditions extrêmes retrouvées dans ces petites cavités, les virus ont un taux de productivité comparable à ceux mesurés dans les sédiments marins ($0.55-0.86 \times 10^8$ «virus-like particles» $g^{-1} h^{-1}$). De plus, les virus polaires auraient développé des stratégies intéressantes pour survivre en milieu inhospitalier. Par exemple, une étude menée à la fois dans des lacs arctiques et antarctiques a observé une grande fréquence de cellules bactériennes visiblement infectées par des virus, mais présentant un très faible nombre de virions intracellulaires (2-

15) (Sawstrom et al., 2007). Une des explications émises par les auteurs serait que les virus utiliseraient les bactéries pour y résider de longues périodes étant donné le long temps de génération cellulaire. Ce trait favoriserait la persistance des populations virales en milieu ultra-oligotrophique où les cellules hôtes sont peu abondantes et diverses.

Bref, bien que les virus arctiques aient été moins étudiés que ceux en milieux tempérés, de plus en plus de découvertes intéressantes ont pu être réalisées à leur sujet. L'exemple du cyanopodovirus SEIV-1 mentionné précédemment illustre très bien tout le mystère derrière les virus arctiques (Chenard et al., 2015). Malgré le nombre croissant de données sur la diversité virale dans l'Arctique, il n'en demeure pas moins qu'il en reste encore énormément à découvrir. D'ailleurs, l'étude des virus arctiques soulève de nouveaux défis en raison des difficultés logistiques pour obtenir les échantillons (sites très éloignés), les conserver (surtout lors du transport) et les analyser (souvent petite quantité d'acides nucléiques). Cela engendre des coûts relativement élevés reliés à la recherche scientifique sur les virus en région polaire, ce qui peut limiter les possibilités d'études.

1.2.4 Approches utilisées pour étudier les virus

Dans le but d'établir la composition et la diversité des assemblages viraux, deux techniques de biologie moléculaire sont couramment utilisées: une approche ciblée en amplifiant un gène marqueur par PCR (de Carcer et al., 2016), et une approche globale en séquençant le métavirome (Lopez-Bueno et al., 2009). L'approche utilisée dépendra bien sûr des questions de recherche puisque les deux techniques n'apportent pas les mêmes réponses. La première analyse en détails la diversité d'un ou des groupes de virus spécifiques possédant le gène amplifié, tandis que la seconde donne un portrait complet de la communauté virale (tous les virus), ce qui permet, par exemple, de découvrir de nouveaux gènes ou virus. La technique de séquençage est également à considérer puisque toutes les méthodes de séquençage haut-débit ne sont pas adéquates à toutes les situations. Par exemple, le pyroséquençage 454 produit de 1×10^5 à 1×10^6 séquences par cycle tandis que le séquençage Illumina™ permet d'en obtenir 7×10^6 à 6×10^9 (Shokralla et al., 2012). De plus, Illumina™ offre deux types de séquençage : MiSeq approprié pour les séquences de 400-500 pb et le HiSeq adapté pour les courtes séquences de 100 bp. Le HiSeq donne également une beaucoup plus grande quantité de lectures que le MiSeq. Par contre, il est nécessaire de fragmenter les séquences en vue du séquençage HiSeq, ce qui est

avantageux en ce qui concerne les métaviromes mais un inconvénient pour le séquençage d'amplicons PCR. En effet, la fragmentation d'amplicons pourrait engendrer des problèmes lors de l'analyse puisque la forte similarité de tous les fragments rendrait l'assemblage impossible. Le type d'approche à utiliser est donc défini par la question écologique à laquelle il faut répondre.

1.3 Problématique

En région arctique, on dénombre un nombre grandissant de lacs de fonte de pergélisol en raison du réchauffement climatique. Ces lacs, malgré qu'ils soient généralement peu profonds, représentent d'importants puits de carbone. Plusieurs études ont démontré que ces lacs très répandus en Alaska, en Sibérie et au Canada constituent de puissants producteurs et émetteurs de GES (CO₂ et CH₄). Or, à la base de toute cette production, on retrouve l'activité microbienne, fortement stimulée par l'apport élevé en DOM apporté par la dégradation du pergélisol. Une meilleure compréhension des mécanismes microbiens en jeu dans ces lacs est donc essentielle dans le but de mieux estimer l'impact potentiel de ces écosystèmes à l'échelle planétaire. Parmi les facteurs contrôlant les populations microbiennes, il existe le contrôle par le bas («*bottom-up*») qui va stimuler l'activité microbienne, telle que l'apport en nutriments, et le contrôle par le haut («*top-down*») effectué par les prédateurs. Ainsi, il est désormais reconnu que la lyse virale occupe une place majeure dans le contrôle par le haut dans les environnements aquatiques. Les virus en milieu aquatique agissent de manière spécifique sur la dynamique et la diversité génétique des populations microbiennes. De plus, en infectant les espèces de phytoplancton (micro-algues, cyanobactéries), les virus influencent significativement le recyclage des nutriments et de la matière organique au bas de la chaîne alimentaire aquatique. Toutefois, très peu d'études se sont penchées sur la diversité et l'impact des virus dans les lacs de thermokarst. Sachant que ces écosystèmes de plus en plus communs en Arctique sont très actifs sur le plan microbien, il est essentiel d'approfondir nos connaissances sur le rôle des virus dans ces milieux. Il est donc nécessaire de réaliser tout d'abord une étude approfondie sur la diversité des virus résidant dans la colonne d'eau des lacs de fonte afin d'en apprendre davantage sur les assemblages viraux, ce qui permettra ensuite de mieux évaluer leur influence dans ce type d'écosystème arctique.

1.4 Objectifs spécifiques

Ce projet de recherche visait à étudier les communautés virales dans les lacs de thermokarst du subarctique canadien. Ce dernier comporte deux objectifs principaux qui seront traités séparément à l'intérieur de ce mémoire, soit sous forme de deux chapitres distincts :

1. **Établir un portrait de la diversité virale dans différents types de lac de fonte en ciblant deux familles de virus spécifiques (*Myoviridae* et *Chlorovirus*).** Ces virus infectent les micro-organismes à la base de la chaîne alimentaire aquatique (cyanobactéries, micro-algues *Chlorella*). Une approche par PCR et du séquençage haut-débit (Illumina) ont été utilisés afin de détecter et d'identifier les assemblages viraux dans chaque lac. Cet objectif correspond au *Chapitre 2* de ce mémoire, et est présenté sous la forme d'un article scientifique.
2. **Isoler un cyanophage à partir de lacs subarctiques.** Au final, deux différents cyanophages ont été isolés avec succès à partir d'un lac subarctique (lac rocheux peu profond) en utilisant une approche par culture liquide. Le génome des deux cyanophages a été séquencé (Illumina) ce qui a permis d'identifier deux AMG's putatifs. Cet objectif correspond au *Chapitre 3* de ce mémoire.

1.5 Sites d'échantillonnage

Les sites à l'étude dans le cadre de ce projet se retrouvent dans la région subarctique canadienne. Ils se situent plus précisément au nord du Québec (Nunavik) sur la côte est de la Baie d'Hudson et à l'embouchure de la Grande Rivière à la Baleine (Figure 1.5). Cette région a été sélectionnée puisque le Centre d'Études Nordiques (CEN) y a établi une station de recherche, et plusieurs équipes scientifiques y récoltent des données environnementales depuis 1971 (Bhiry et al., 2011). La région présente un grand potentiel scientifique en raison de sa localisation géographique et de la géomorphologie du terrain. Elle se situe aux environs de trois types de pergélisol (isolé, sporadique et discontinu) et près de la frontière entre la taïga et la toundra arctique (Bhiry et al., 2011). Elle repose également sur les sédiments (limon) de l'ancienne Mer de Tyrrell ce qui rend le sous-sol de la région imperméable. Or, cette caractéristique rend les lacs de thermokarst de cette région plutôt uniques puisque leur drainage est limité en raison du limon dans le sol. Ainsi, ces lacs persistent d'années en

années, ce qui les distingue des lacs de thermokarst d'autres régions arctiques qui finissent par naturellement disparaître.

Une session d'échantillonnage d'une durée d'une semaine a eu lieu dans cette région entre le 24 août et le 1^{er} septembre 2015. Les lacs ont été échantillonnés du 26-30 août, et l'eau a été traitée au maximum 6 heures suite à sa récolte.

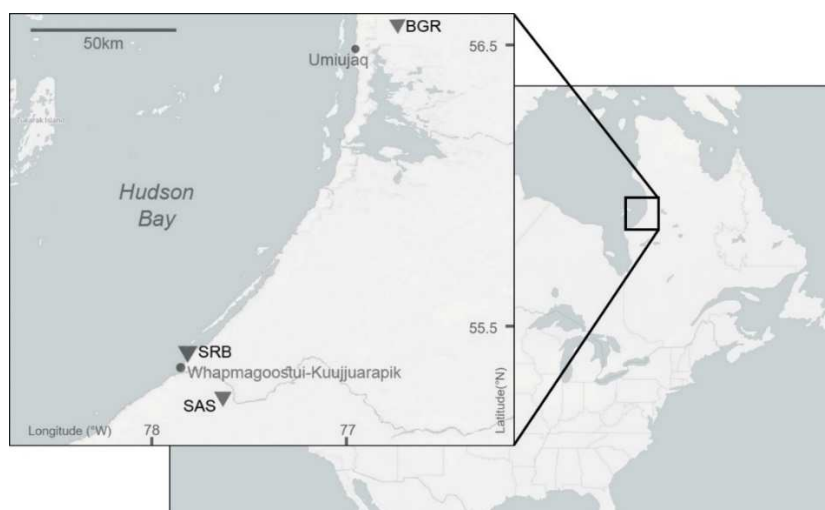


Figure 1.5 : Localisation des trois vallées échantillonnées (SAS, BGR, SRB) au nord du Québec (Nunavik, Canada). Quatre lacs ont été sélectionnés pour l'échantillonnage (SAS-2A, BGR-1, Olsha, Lake 4K).

1.5.1 Régions des villages de Whapmagoostui-Kuujuarapik et Umiujaq

La région autour du village de Whapmagoostui-Kuujuarapik (W-K) ($55^{\circ} 15' N$, $77^{\circ} 45' O$) est constituée de pergélisol de types sporadique et isolé puisqu'il compose moins de 10% du sol. Environ 100 km au nord se trouve le village d'Umiujaq ($56^{\circ} 33' N$, $76^{\circ} 32' O$) où le sol possède une concentration en pergélisol entre 10-50%, ce qui en fait une zone de pergélisol discontinu et sporadique (Allard and Seguin, 1987). Ainsi, les lacs de thermokarst sélectionnés dans le cadre de ce projet se situent le long d'un gradient de décomposition de pergélisol.

Une récente étude a démontré que la région de W-K subissait une période de changements rapides entraînée par le réchauffement climatique (Bhiry et al., 2011). La température de l'air a

considérablement augmentée dans les dernières décennies, passant d'une moyenne de $-4,3 \pm 1,6$ °C entre 1960-2001 à une moyenne de $-2,6 \pm 1,2$ °C entre 2001-2010 (Figure 1.6) (Bhiry et al., 2011).

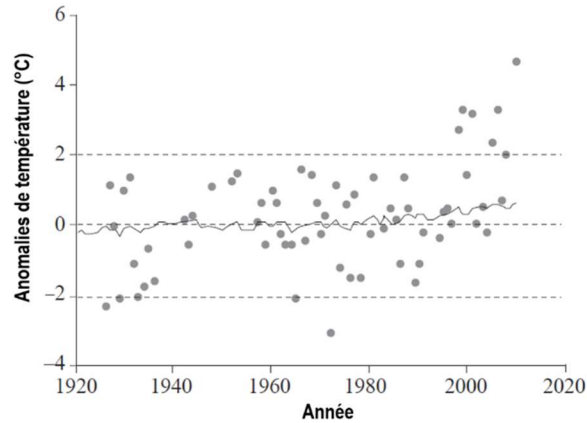


Figure 1.6 : Anomalies de la température de l'air dans le village de W-K en comparaison à la valeur moyenne au 20^e siècle. La ligne continue correspond à l'anomalie globale moyenne. Figure adaptée de Bhiry et al. (2011).

1.5.2 Lacs à l'étude

Au total, quatre lacs ont été échantillonnés (Figure 1.7). Un premier site d'échantillonnage se trouve 8 km au sud du village de W-K, soit dans la vallée de la rivière Sasapimakwananisikw (SAS 2; 55° 13' N, 77° 42' O). Celle-ci couvre une zone d'environ 5 km² et possède une élévation de 110 m au-dessus du niveau de la mer. Un second site, appelé Bundesanstalt für Geowissenschaften und Rohstoffe (BGR; 56° 37' N, 76° 13' W), se situe près du village d'Umiujaq dans la vallée de la rivière Sheldrake. Cette vallée se retrouve, quant à elle, à 185 m au-dessus du niveau de la mer. Étant une tourbière, le site de SAS 2 ne contient que des paises, tandis que le site de BGR comprend des paises et des lithales. La vallée de SAS 2 est composée majoritairement de plantes et de mousses telles que *Sphagnum*, *Carex* et *Drepanocladus* caractéristiques des milieux humides (tourbière) (Arlen-Pouliot and Bhiry, 2005). Au site de BGR, c'est plutôt un couvert de type arbustif qui domine (*Salix planifolia*, *Betula glandulosa*) avec quelques arbres répartis de manière dispersée et sporadique (*Picea mariana*) (Calmels et al., 2008).

Dans ce projet, un seul lac par région a été échantillonné (n=1/région). Le choix s'est donc porté vers SAS-2A (palse) et BGR-1 (lithalse) puisqu'une grande quantité de données biologiques et physico-

chimiques concernant ces lacs étaient disponibles (Crevecoeur et al., 2015; Deshpande et al., 2015; Matveev et al., 2016). De plus, il était question dans ce projet de maîtrise de comparer les assemblages viraux dans deux types de lac de fonte (palse vs lithalse). Deux lacs de référence (Olsha et Lake 4K) ont également été sélectionnés dans la région immédiate autour du village de W-K dans une zone de bassins rocheux peu profond (*Shallow rock-basin* (SRB)). Ces lacs sont considérés comme référentiels puisqu'il s'agit de lacs d'origine glaciaire datant de plusieurs milliers d'années ne résultant pas de la dégradation du pergélisol. En d'autres termes, Olsha et Lake 4K ne subissent pas directement les effets de la fonte du pergélisol, contrairement aux lacs de fonte. De plus, leur colonne d'eau n'est pas stratifiée verticalement en raison d'un mélange constant de la couche de surface avec la couche de fond. Il est donc juste de considérer Olsha et Lake 4K en tant que références car, malgré une origine et des caractéristiques limnologiques différentes, ces derniers sont exposés aux mêmes conditions subarctiques que les lacs de thermokarst.



Figure 1.7 : Photo de chaque site d'échantillonnage. Les deux lacs de thermokarst sont SAS-2A et BGR-1, et les deux lacs de référence sont Olsha et Lake 4K. Crédits photos : Alice Lévesque (SAS-2A et Olsha), Alex Matveev (BGR-1) et Alexander I. Culley (Lake 4K).

Chapitre 2 : Contrôle des communautés virales dans les mares de fonte de pergélisol par le type de mare

2.1 Résumé

En raison du réchauffement climatique, on dénombre un nombre croissant de mares de fonte de pergélisol (ou mare de thermokarst) en Arctique. Il existe deux types de mares de fonte (palse, lithalse) en fonction du sol où se forme la mare, puisque la composition du terrain a une grande influence sur leurs propriétés physico-chimiques. L'activité microbienne y est fortement stimulée par l'apport de carbone organique, mais la nature des forces la contrôlant demeure peu connue dans ces écosystèmes. Les virus ont un rôle essentiel dans tous les milieux aquatiques, c'est pourquoi l'objectif de ce projet était d'établir la diversité des virus (Myovirus, Chlorovirus) dans deux types de mare de fonte subarctique. Des analyses bio-informatiques ont révélé une différenciation des communautés virales et des interactions avec leur hôte selon le type de mare. Ainsi, cette étude permet d'en apprendre davantage sur l'écologie des virus en milieu nordique.

Chapter 2 : Landscape control of viral communities in permafrost thaw ponds

Alice V. Lévesque^{1,2,3}, Warwick F. Vincent^{2,5}, Jérôme Comte⁴, Connie Lovejoy^{3,5} and Alexander I. Culley^{1,2,3}

1. Département de biochimie, de microbiologie et de bio-Informatique, Université Laval, Québec, QC, Canada

2. Centre d'études nordiques (CEN), Université Laval, Québec, QC, Canada

3. Institut de biologie intégrative et des systèmes, Université Laval, QC, Canada

4. Water Science and Technology, Environment and Climate Change Canada, Canada Centre for Inland Waters, Burlington, ON, Canada

5. Département de biologie, Université Laval, Québec, QC, Canada

Keywords: microbial ecology, viral diversity, thaw ponds, arctic, permafrost

2.2 Abstract

Permafrost thaw ponds and lakes are currently among the most abundant freshwater ecosystem in Arctic landscapes. Thaw ponds can originate from two types of landscapes (palsa and lithalsa), which have an important impact on their physicochemical properties. Recent measurements taken from these water bodies showed that they are highly saturated in carbon dioxide and methane, gases that are a product of microbial activity. In order to ultimately understand the contribution of thaw ponds to the production of greenhouse gases, it is first necessary to characterize the dynamics of their microbiota. However, the mechanisms of top-down control of microbes in these aquatic ecosystems are poorly understood. In this study, we focused on viruses, largely unstudied agents of top-down control in subarctic thaw ponds. Based on DNA barcoding, we assessed the diversity of putative host and viral communities in the palsa and lithalsa ponds and found that the two communities were distinctly different. Our results suggest that host community composition is strongly influenced by environmental filtering, which in turn drives viral diversity in these ponds.

2.3 Introduction

Aquatic ecosystems harbour a remarkably abundant and diverse assemblage of viruses that infect microbes (Suttle, 2005). Aquatic viruses can regulate host abundance and influence population structure, catalyze evolution through the mediation of gene exchange, and ultimately modify biogeochemical cycles of fundamental importance (Brussaard, 2004; Gao et al., 2016; Rohwer and Thurber, 2009; Wilhelm and Suttle, 1999). In the northern high latitudes, viruses appear to be a persistent and ubiquitous component of the polar aquatic microbial community. Viruses have been reported from a diversity of Arctic and subarctic aquatic habitats including lakes (de Carcer et al., 2016; Zeigler Allen et al., 2017), cryoconites holes (Anesio et al., 2007), sea-ice brines (Wells and Deming, 2006), marine sediments (Colangelo-Lillis et al., 2016) and the Arctic Ocean (Guixa-Boixereu et al., 2002).

In some northern permafrost landscapes, global warming has contributed to the formation of an increasing number of ponds and lakes formed by the thawing of ice-rich soils (Pienitz et al., 2008). These thaw (or thermokarst) ponds and lakes, are strong emitters of the greenhouse gasses carbon dioxide and methane to the atmosphere (Laurion et al., 2010; Matveev et al., 2016). The production of these gases is ultimately due to the activity of microbes (Crevecoeur et al., 2015; Crevecoeur et al., 2016), which are subject to predation by grazers and viruses. Despite the fact that thermokarst ponds are among the most abundant freshwater ecosystems in the circumpolar North (Olefeldt et al., 2016), the forces that control community structure and dynamics (i.e. top-down control), including viruses, remain largely uncharacterized.

In the Hudson Bay region of subarctic Québec, thermokarst ponds are subjected to dramatic seasonal gradients in light, carbon, and oxygen. These ponds are typically vertically stratified, with a high-light oxygenated surface layer and a low-light hypoxic or anoxic stratum of bottom water in the summer. During winter, when the ponds are ice covered light is limited or absent and the water column becomes completely anoxic (Deshpande et al., 2015). Two types of permafrost landscapes occur in the Québec subarctic : palsa thaw ponds formed in hummocks of peatland, and lithalsa ponds formed in hummocks of mineral-rich soils (Calmels et al., 2008; Gurney, 2001). The nature of the landscape ultimately affects the quantity, nature and availability of organic material in the pond (Kiikkila et al., 2014; Watanabe et

al., 2011), and appears to be a principle driver of microbial community composition (Comte et al., 2016b).

Viruses are intracellular obligate parasites and therefore their dynamics are inextricably linked to that of their hosts. The diversity of bacteria (Comte et al., 2016a; Comte et al., 2016b), Archaea (Crevecoeur et al., 2015; Crevecoeur et al., 2016) and protists (Bégin and Vincent, 2017; Przytulska et al., 2017; Przytulska et al., 2016), the presumptive primary hosts of thermokarst viruses, have recently been characterized in several Québec subarctic ponds. Comte *et al* (2016) observed that the largest differences in bacterial diversity were between the two permafrost landscape types (e.g. palsa vs. lithalsa) and that the composition of these communities appears to be primarily determined by environmental filtering and dispersal limitation. Archaea from the orders *Methanomicrobiales* and *Methanosarcinales* dominated the active archaeal communities in the bottom waters of thaw ponds in subarctic Québec (Crevecoeur et al., 2016). Similar to bacteria, the species composition of archaeal communities between palsa and lithalsa ponds was significantly different (Crevecoeur et al., 2016). Potentially photosynthetic groups in these ponds include cyanobacteria, chlorophytes, cryptophytes and dinoflagellates, as well as photosynthetic sulfur bacteria (Przytulska et al., 2016). Unlike the archaeal and bacterial communities there was no clear difference in communities between palsa and lithalsa photosynthetic assemblages (Przytulska et al., 2017).

In an effort to characterize viruses that infect hosts that perform critical transformations in the carbon cycle, we targeted viruses that infect chlorophytes, a subgroup of oxygenic phototrophs, and bacteria. With one exception (Brussaard, 2004), chlorophytes appear to be primarily infected by viruses from a genus of large double-stranded DNA viruses (*Chlorovirus*) within the family *Phycodnaviridae* (Short, 2012). A set of primers targeting a fragment of the polymerase gene (*polB*) of this genus was developed by Short et al. (2011) to investigate the temporal and spatial dynamics of chloroviruses in freshwater. This approach resulted in the amplification of an assemblage of phylotypes that included both known and divergent viruses and demonstrated the method's capacity to capture divergent chloroviruses. Bacteriophage are abundant, ubiquitous and highly diverse in aquatic environments (Breitbart, 2012). There have been a number of surveys of myovirus (a type of bacteriophage) diversity from freshwater environments. These studies, based on primers targeting a conserved structural gene (Matteson et al., 2011), demonstrated that myoviruses are present in extreme, temperate and tropical environments

(Wilhelm and Matteson, 2008). These surveys also demonstrate that the number of bacteriophages in culture with sequenced genomes represent a small minority of the total diversity present in the environment.

The aims of the present study were to determine the diversity and potential environmental partitioning of chloroviruses and myoviruses in subarctic thaw ponds, in order to gain insight into one of the mechanisms of top-down control of microbial communities in 3 Québec sub-arctic ponds. Specifically, we investigated whether: (1) chloroviruses and myoviruses can be detected using known primers in thermokarst ponds; (2) thaw pond viruses are divergent from known types; (3) the permafrost landscape type (palsa versus lithalsa) influences viral community structure; and (4) there are distinct viral communities between surface and bottom layers of the ponds. Our approach was to analyze samples of total virus and extracellular whole virus communities taken from the surface and bottom regions of palsa, lithalsa and shallow rock-basin ponds in subarctic Quebec. Viral communities were identified using high-throughput sequencing (Illumina MiSeq) of the fragments of the gp23 structural gene of myoviruses and polB polymerase genes of chloroviruses. We also amplified the V6-V8 region of the bacterial and archaeal 16S rRNA gene and the V4 region of the eukaryotic 18S rRNA gene to identify informative patterns between viruses and their putative hosts.

2.4 Materials and Methods

2.4.1 Study site and sampling

Samples were collected from August 27 to 30, 2015 from the two classes of permafrost thaw ponds in northern Quebec (Nunavik, Canada) (Figure 2.1). The SAS-2A site is a thaw pond in a palsa valley adjacent to the Sasapimakwananisikw River (SAS-2A, 55°13'N; 77°42'W), located in the sporadic permafrost zone, where the extent of permafrost ranges from 10% to 50%. The second site, BGR-1, is a lithalsa pond (Bhiry et al., 2011) located in the Bundesanstalt für Geowissenschaften und Rohstoffe valley (BGR-1, 56°37'N; 76°13'W) near the Sheldrake River in a region of discontinuous permafrost where the extent of permafrost ranges from 50% to 90%. From now on, these two lakes will be referred to as SAS and BGR. For comparison to the thermokarst ponds, Lake Olsha (55°16'N; 77°44'W), a well-mixed shallow rock-basin lake (SRB) near Whapmagoostui-Kuujjuarapik was also sampled. Water samples from oxygenated (surface) and hypoxic/anoxic (bottom) depths were collected in triplicate.

Temperature, conductivity, dissolved oxygen (DO) and pH measured with a Hydrolab™ DS5X profiler (Loveland, CO, USA), were collected from an inflatable boat from the deepest region of each pond. Hydrolab data were not collected from the BGR site. Surface water was sampled with an acid-washed opaque bottle (Nalgene™) while bottom samples were collected with a horizontally mounted Van Dorn bottle (Wilco). Water samples were transferred into acid-washed Cubitainers™ that were stored in opaque bags, and transported back to the laboratory via helicopter for processing. DOC, TSS, TN and TP were determined as per Laurion et al. (2010).

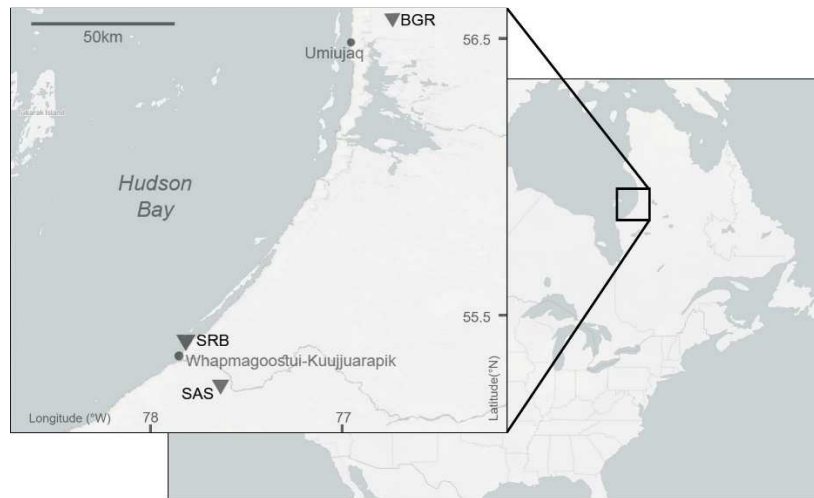


Figure 2.1: Location of the three sampled valley in the Canadian subarctic (Nunavik, Québec). SRB is outside the permafrost area, SAS occurs within sporadic permafrost, while BGR is located in the discontinuous area.

2.4.2 Sample processing and molecular analysis

For each replicate, total viruses (i.e samples that were not pre-filtered and presumably included extracellular viruses, intracellular viruses and viruses adsorbed to particles), were collected by direct filtration of the sample through a 0.02µm pore size AAO filter (Anotop 25, Whatman). Anotop filters were then stored at -80°C until extraction. Total nucleic acids were extracted from Anotop filters as described by Mueller et al. (2014). After extraction, samples were treated with the Power Clean Pro DNA Clean up kit (MoBio) to remove PCR inhibitors.

PCR was performed with the myovirus primers T4superF1-T4superR1 (Chow and Fuhrman, 2012) and the chlorovirus primers CHLVdF-CHLVdR (Short et al., 2011) listed in Table 2.1. Each reaction mixture

(final volume 25 μ l) contained 200 μ M of each dNTP (Bio Basics), 0.4 μ M of each primer, 1X Expand buffer containing MgCl₂, 2U of Expand High-fidelity enzyme blend (ROCHE) and 3 μ L of the purified DNA. The following thermocycler conditions were used: 94°C for 120 s, followed by 40 cycles of denaturation at 94°C for 30 s (CHLVd)/ 15 s (T4super), annealing at 52°C (CHLVd) / 54°C (T4super) for 30 s and extension at 72°C for 60 s and then a final extension at 72°C for 7 min. PCR products were separated on a 1% agarose gel, and DNA of the target size was purified using Axygen magnetic beads (Corning Life Sciences). In preparation for sequencing, Illumina TruSeq adaptors and unique barcodes were added with further rounds of PCR for each reaction. Sample barcoded amplicons were subsequently pooled in equimolar concentrations for sequencing on an Illumina MiSeq at the Plateforme d'Analyses Génomiques (IBIS, Université Laval, Québec, Canada). Sub-regions of the bacterial 16S rDNA (V6-V8 region, primers B969F-BA1406R) and eukaryotic 18S rDNA (V4 region; primers E572F-E1009R) were amplified as described in (Comeau et al., 2011) with template from extracted AAO filters. The raw Illumina sequences have been deposited in the NCBI Short Read Archive database with the following identifiers: SRP115531 (bacteria), SRP115532 (viruses), and SRP115533 (eukaryotes).

Table 2.1 : Sets of primers used for viral amplification

Virus	Gene	Primers	Sequence (5'→3')*	Amplicon length (bp)	Reference
Myovirus	<i>gp23</i>	T4superF1	GAYHTIKSIGGIGTICARCCATG	400-500	(Chow and Fuhrman, 2012)
		T4superR1	GCIYKIARRTCYTGICIARYTC		
Chlorovirus	<i>polB</i>	CHLVdF CHLVdR	CCWATCGCAGCWCTMGATTTTG ATCTCVCBGCVARCCACTT	560-575	(Short et al., 2011)

* : Degenerate nucleotides code

R = A, G ; Y = C, T ; M = A, C ; K = G, T ; S = G, T ; W = A, T ; H = A, C, T ; B = C, G, T ; V = A, C, G

2.4.3 Bioinformatics and statistical analyses

Sequence reads were first analyzed using the UPARSE quality-filtering pipeline to exclude all chimeras and singletons (Edgar, 2013). Operational taxonomic units (OTUs) were determined with >97% identity (16S rRNA), >98% identity (myovirus and 18S rRNA) and >99% identity (chlorovirus). The virus % identity cutoffs were based on a phylogenetic analysis of classified viruses. Bacterial and eukaryotic sequences were classified based on an alignment of OTUs with the SILVA reference database (Pruesse et al., 2007) and a curated reference database of Arctic protists (Lovejoy et al., 2016).

Sequences classified as fungi, streptophyta, metazoan, chloroplast and Archaea were removed from the data set. Viral OTUs were classified with DIAMOND (Buchfink et al., 2015) to identify the nearest match in the indexed nr BLAST database. Results were visualized in MEGAN6 (Huson et al., 2016). Non-viral OTUs (including unknown sequences) were subsequently removed from the dataset.

Statistical analyses were executed using QIIME (Caporaso et al., 2010) and RStudio v3.3.0 (R Core Team, 2016). Alpha-diversity metrics were calculated using the QIIME command `alpha_diversity.py` (Chao1, Shannon index). The assumptions of normality and homoscedasticity for an analysis of variance were confirmed using a Shapiro-Wilk test. After, a two-way ANOVA and an a posteriori TukeyHSD test were used to assess the significance of pond type or depth on α -diversity. Alternatively, a permutational MANOVA (PERMANOVA, 999 permutations) was performed with the `Adonis` function in the `Vegan` package (Oksanen et al., 2016) to test the effect of oxygen and depth on diversity. Venn diagrams were produced using the `VennDiagram` package in R (Chen and Boutros, 2011) to illustrate the percentages of unique and shared OTUs among ponds. A UniFrac distance matrix was generated (`beta_diversity.py`) after sub-sampling the dataset (75% of the reads per sample) and used to build a jackknife phylogenetic tree. Heatmaps representing the relative abundance of the most abundant OTUs (defined as >1% of reads in the data set) in each sample were produced to identify specific distribution patterns in the data. Principal coordinate analysis (PCoA) scatter plots based on unweighted UniFrac distances were generated to visualize the environmental variables potentially influencing diversity patterns among the pond microbial communities.

To gain further insight into the taxonomy of the thaw pond chlorovirus phylotypes, a maximum-likelihood (ML) reference tree (RAxML v8.2.9; (Stamatakis, 2014)) was produced based on an alignment (MUSCLE; (Edgar, 2004)) of representative viruses from the family *Phycodnaviridae*. Alignments were then transformed into distances with RAxML v8.2.9 (Stamatakis, 2014). The thaw pond phylotypes were subsequently added to the RAxML phylogeny with the evolutionary placement algorithm (EPA) (Berger et al., 2011).

2.5 Results

2.5.1 Limnological conditions

Table 2.2: Limnological properties of the sampled lakes in August 2015. Dissolved organic carbon (DOC), total suspended solids (TSS), dissolved oxygen (DO), dissolved oxygen saturation (DO %), total phosphore (TP), total nitrogen (TN) and chlorophyll a (Chl a).

Lake	Depth (m)	Temperature (°C)	DO (mg/L)	DO %	Conductivity (µS/cm)	TSS (mg/L)	pH	DOC (mg/L)	TP (µg/L)	TN (mg/L)	Chl a (µg/L)
SAS	0	12.75	4.13	47.1	44.4	2.17	6.46	13.7	13.45	0.66	1.91
	0.5	11.95	0.88	9.5	45.2	NA	5.51	NA	NA	NA	NA
	2	6.08	0	0	123.3	4.56*	5.64	9*	55.04	0.70	1.04
BGR	0	15.65*	6.8*	69*	85.4*	1.97	6.99*	2.5	23.50	0.41	2.00
	3.5	2.8*	0*	0*	150.7*	13.80	5.99*	2.8	31.39	0.43	15.85
Olsha	0	13.07	9.63	102.3	35.7	2.64	7.72	11.4	55.04	0.63	6.47
	1.5	14.01	9.16	96.5	35.9	2.32	7.14	12.2	19.76	0.44	5.69

* : Data measured in August 2014

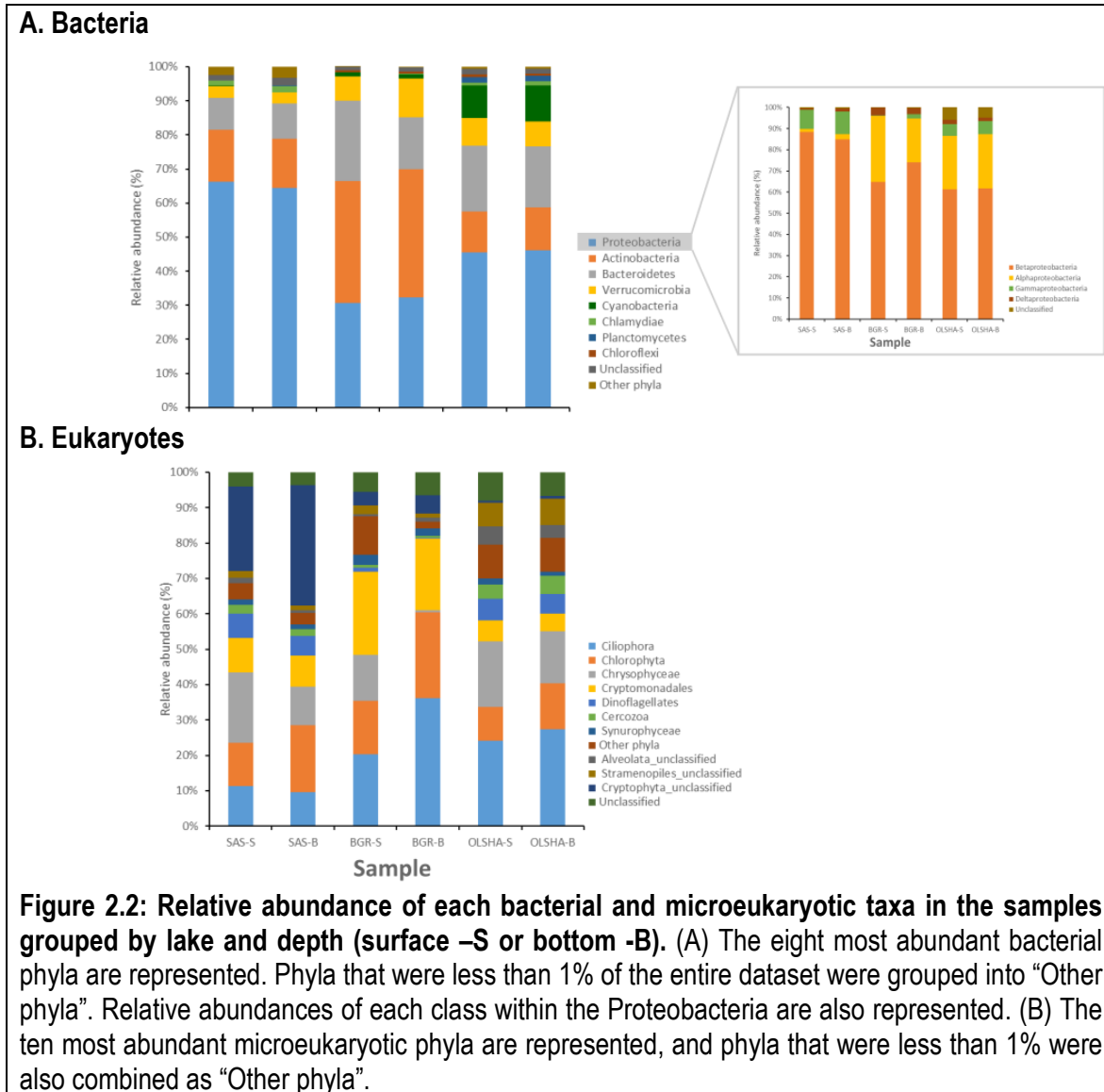
The data from the palsa thaw pond (SAS) were consistent with previous studies from the region and indicated that the thermokarst ponds are generally highly stratified with warmer, oxygenated water at the surface and cooler, anoxic, and more ion-rich water at the bottom. Conversely, the shallow rock-basin lake (Olsha) water column was well-mixed, oxygenated from surface to bottom, and had relatively low conductivity (Bégin and Vincent, 2017; Crevecoeur et al., 2016; Deshpande et al., 2015). SAS also had relatively lower pH values and higher DOC, total phosphorus and total nitrogen concentrations, while BGR had higher concentrations of TSS and bottom water Chl a (Crevecoeur et al., 2017; Crevecoeur et al., 2016; Deshpande et al., 2015; Przytulaska et al., 2016).

2.5.2 Cellular communities

2.5.2.1 Community composition

In descending order, *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* and *Verrucomicrobia* were the taxonomic groups with the highest relative abundance in all three ponds (Figure 2.2A). Within the phylum *Proteobacteria*, β -proteobacteria comprised between approximately 60 and 90% of the relative abundance. Members of the genus *Polynucleobacteria* had the highest relative abundance in SAS and

Olsha, while *Actinobacteria* were proportionally most abundant in BGR. Cyanobacteria were detected in all three ponds but had greater relative abundance in Olsha (10%) than in the thaw ponds (< 2%). Methanotrophic bacteria represented 5 and 3% of the bacterial community in SAS (*Methylobacter*) and Olsha (*Methylacidiphilales*), respectively.



At all three sampling sites, the eukaryote taxa with the highest relative abundances were *Ciliophora*, *Chlorophyta*, *Chrysophyceae* and *Cryptomonadales* (Figure 2.2B). Chlorophytes tended to have higher relative abundances in bottom waters while the surface waters had high relative abundances of dinoflagellates, cryptomonads and ciliates in SAS, BGR and Olsha, respectively. In general,

autotrophic protists represented a high percentage of the total community, ranging from 79% in SAS to 42% in Olsha.

2.5.2.2 Community comparisons

The PCoA analysis based on unweighted UniFrac distances showed distinct clustering of bacterial and eukaryotic communities by pond type, but no evident clustering by depth (Figure 2.3). SAS bacterial communities were distinct from BGR and Olsha along the X-axis (44% of the variance explained, Figure 2.3C). Although BGR and Olsha communities overlapped along the X-axis, these two communities were separate along the Y-axis (28% of the variance explained, Figure 2.3C). Olsha eukaryote communities were distinct from the thaw ponds both along the X-axis (36% of the variance explained) and Y-axis (32% of the variance explained) (Figure 2.3D). Although the thaw ponds were clustered along the X-axis, they were separate along the Y-axis.

UniFrac UPGMA trees based on the top 1% showed well-supported patterns in OTU distribution by pond type, but not by depth, with the exception of BGR bacterial and eukaryotic communities (Figure 2.4C-D). The dominant bacterial phylotype was a member of the *Puniceicoccaceae* and the dominant eukaryotic phylotype was a member of the *Chlorophyceae*.

Bacterial diversity patterns were generally consistent with those of the PCoA analysis. The Shannon Indices from all three pond types were significantly different ($p < 0.05$), while surface and bottom indices were not. Median values and ranges for the two indices are listed in Supplementary Table S2.1 Figure S2.1. Bacterial richness ranged from 577 +/- 37 phylotypes in the surface water of the SAS pond to 213 +/- 5 phylotypes in the surface water of the BGR pond.

Eukaryotic community Shannon indices were significantly different among all the ponds ($p < 0.05$). Surface and bottom indices from BGR were significantly different however the indices from Olsha and SAS were not. Eukaryotic richness was highest in the bottom waters of Lake Olsha (462 +/- 52) and lowest in the bottom waters of BGR (260 +/- 31). Median values and ranges for the two indices are listed in Supplementary Table S2.1 and Figure S2.1.

Venn diagrams (Figure 2.5) indicated that 22.2% of bacterial OTUs and 12.9% of eukaryotic OTUs were shared among the three ponds. The highest percentage of unique OTUs for bacteria (32.1%) and eukaryotes (25.6%) was in the SAS pond and the highest percentage of shared OTUs for bacteria (40.8%) and for eukaryotes (23.1%) was between the Olsha and SAS ponds.

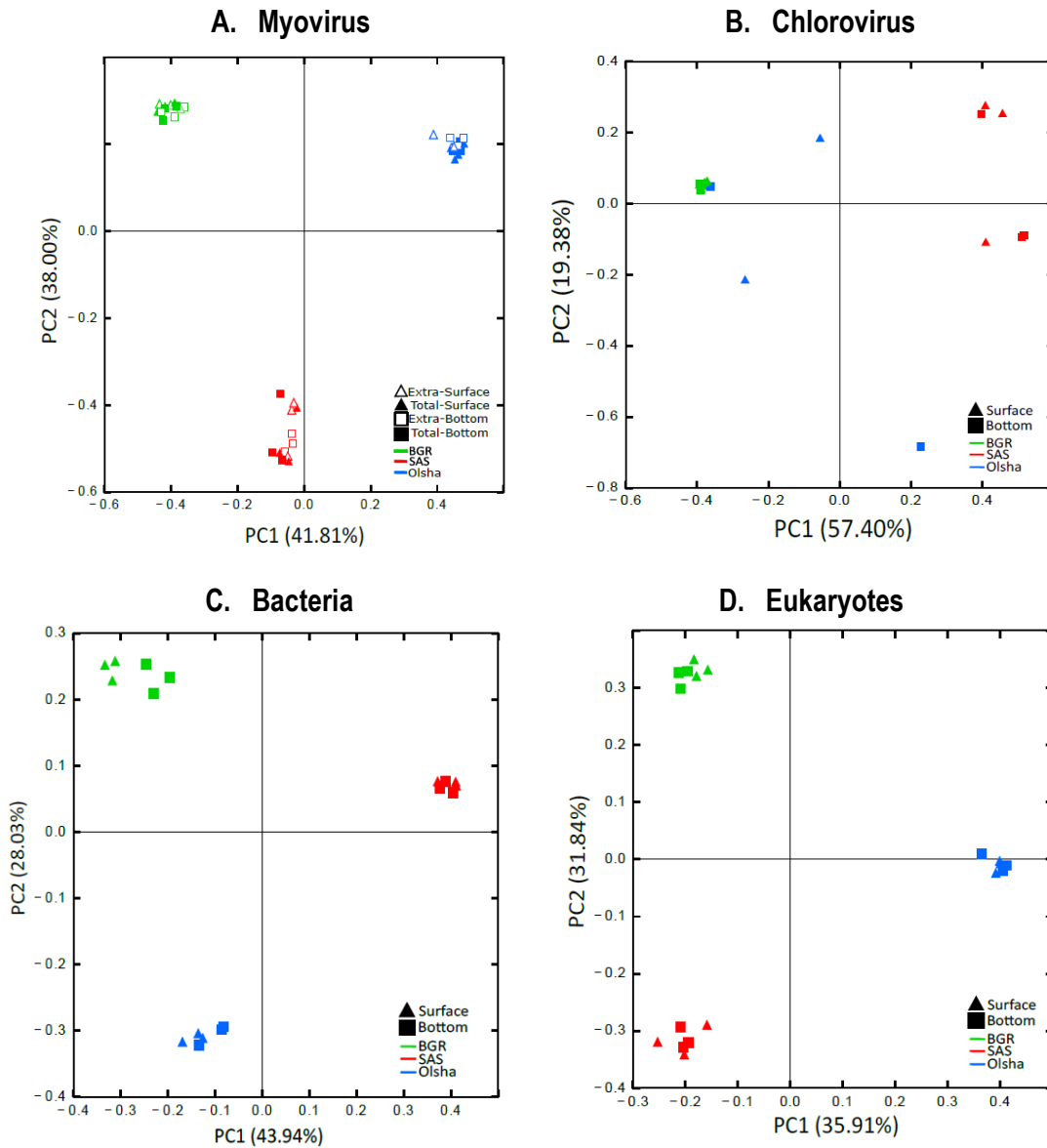
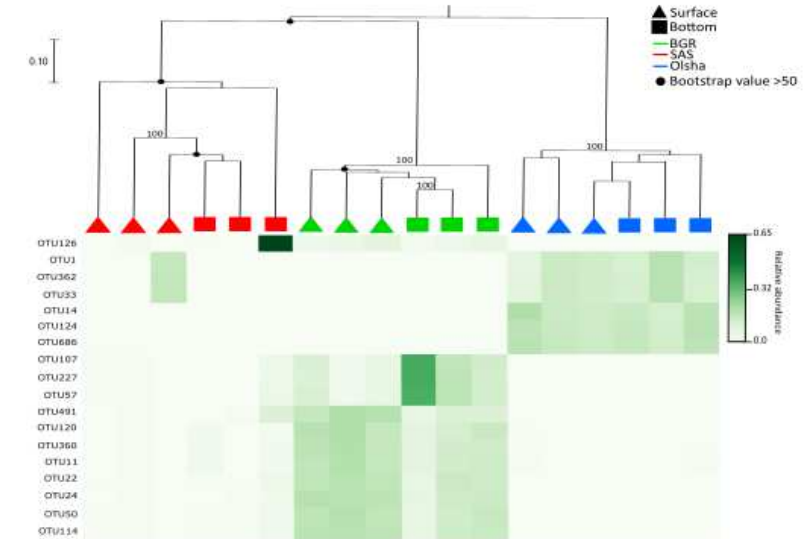
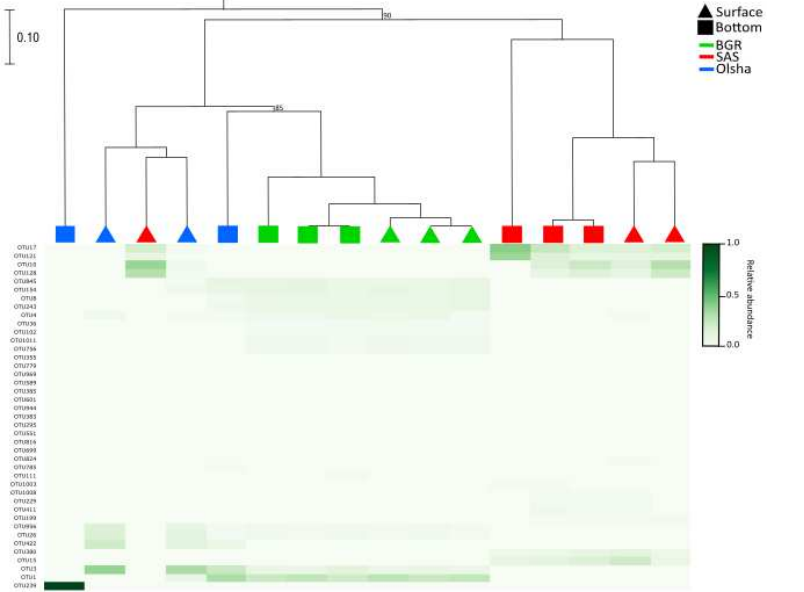


Figure 2.3: Principal coordinate analysis (PCoA) generated with the unweighted UniFrac distances of the myoviral (A), chloroviral (B), bacterial (C), and microeukaryotic (D) communities from the three lakes.

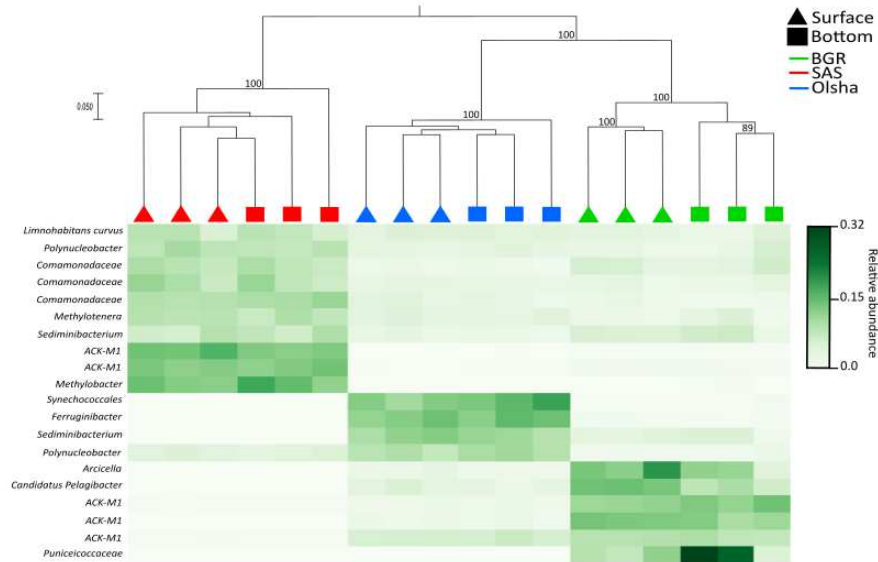
A. Myovirus



B. Chlorovirus



C. Bacteria



D. Eukaryotes

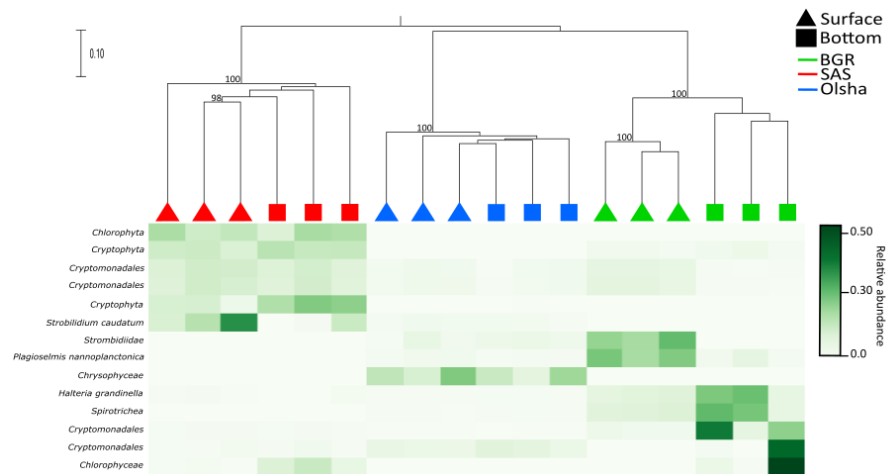


Figure 2.4: Unweighted UniFrac UPGMA clustering of the samples combined with a heatmap representing the relative abundance of the most abundant total myovirus (A), chlorovirus (B), bacterial (C) and eukaryotic (D) OTUs (>1% of the data set). Bacterial and eukaryotic OTUs were assigned at their lowest taxonomic level using the SILVA and the Lovejoy (2016) reference databases. Jackknife values over 50 (A, B) or 90 (C, D) are presented in the tree. Samples from the surface (triangle) and the bottom (square) were colored according to the lake: green, BGR; red, SAS; blue, Olsha.

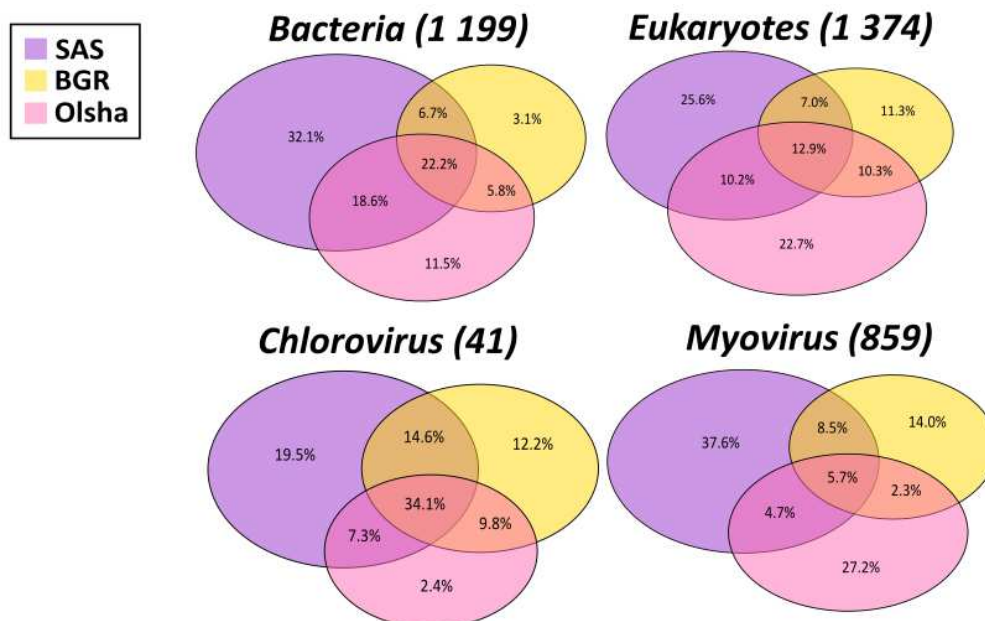
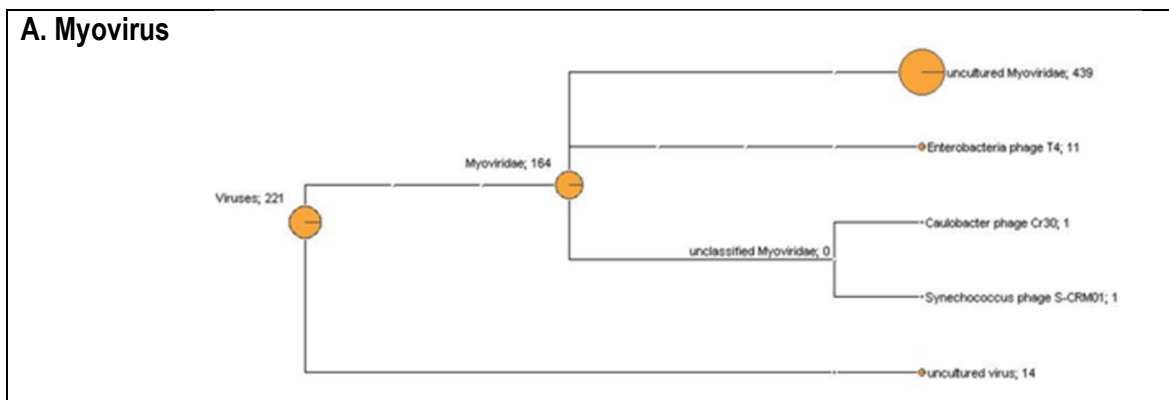


Figure 2.5: Venn diagrams representing the proportion of shared OTUs between lakes for each tested gene. The number in parentheses refers to the total number of OTUs obtained.

2.5.3 Viral communities

2.5.3.1 Community composition

Classification based on Diamond resulted in a majority of OTUs matching with sequences in the uncultured *Myoviridae* category (439) while fewer OTUs matched with the myoviruses *Enterobacteria* T4, and *Caulobacter* phage CR30. Only a single phylotype was identified as a cyanophage species (*Synechococcus* phage S-CRM01). Finally, a large number (221) of the phylotypes were binned in the general category of "Viruses".



B. Chlorovirus

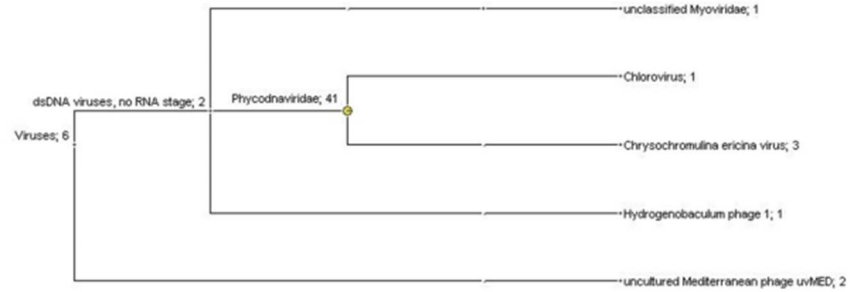


Figure 2.6: Classification of viral OTUs according to the DIAMOND nearest match and the set of primers used (A, myovirus; B, chlorovirus).

Forty-one OTUs matched viruses in the family *Phycodnaviridae* (Figure 2.6B). Within these 41 phlotypes, only one was assigned to the genus *Chlorovirus*. To further resolve the taxonomy of the chlorovirus phlotypes, we produced an ML phylogeny analysis based on an amino acid sequence alignment that resulted in a single large clade with a bootstrap value of 100 that contained all of the environmental OTUs from this study, sequences amplified with the same primer set from Lake Ontario (Short et al., 2011) and from the chlorovirus isolate *Acanthocystis turfacea chlorella virus 1* (Figure 2.7). This clade in turn, formed a well-supported clade (bootstrap value of 100) with the chloroviruses *Paramecium bursaria Chlorella virus 1* and *Paramecium bursaria Chlorella virus NY2A*. Established genera of phycodnaviruses with multiple representatives were resolved into monophyletic clusters with bootstrap values of 100.

2.5.3.2 Community comparisons

The PCoA analysis based on unweighted UniFrac distances showed distinct clustering of myovirus and chlorovirus communities by pond type (with one exception) but not by depth (Figure 2.3A-B). All three myovirus pond communities were separate along the X-axis (42% of the variance explained). BGR and Olsha myovirus communities overlapped along the Y-axis (38% of the variance explained) while SAS was distinct. The chlorovirus thaw pond communities were separate along the X-axis (57% of the variance explained), but overlapped along the Y-axis (19% of the variance explained) while the Olsha community showed no clear pattern (Figure 2.3B).

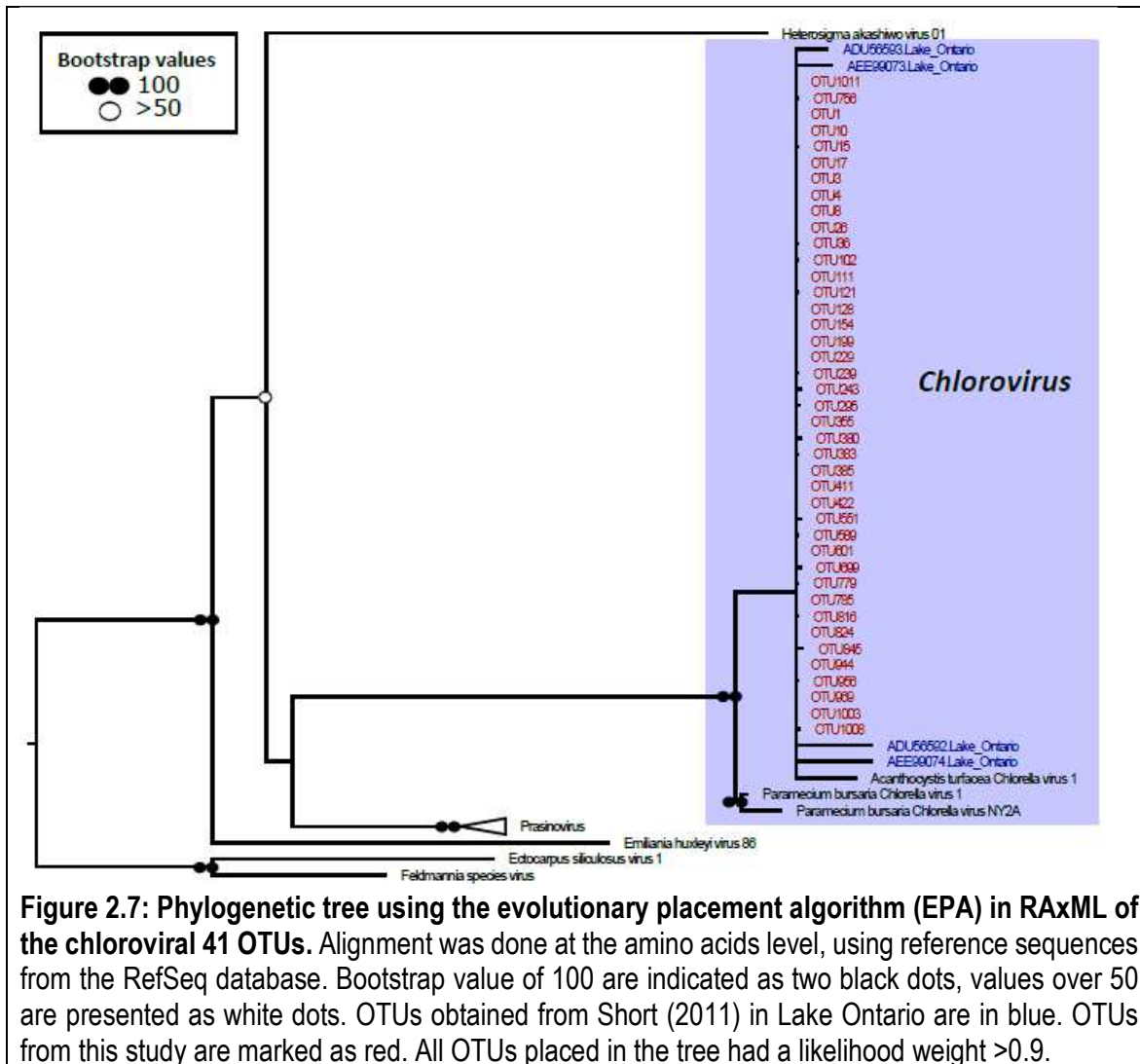


Figure 2.7: Phylogenetic tree using the evolutionary placement algorithm (EPA) in RAXML of the chloroviral 41 OTUs. Alignment was done at the amino acids level, using reference sequences from the RefSeq database. Bootstrap value of 100 are indicated as two black dots, values over 50 are presented as white dots. OTUs obtained from Short (2011) in Lake Ontario are in blue. OTUs from this study are marked as red. All OTUs placed in the tree had a likelihood weight >0.9.

The same trends exhibited by the PCoA analysis were evident in unweighted UniFrac UPGMA trees where myoviruses formed well-supported clades (bootstrap values = 100) by pond type (Figure 2.4A) but did not cluster by depth. The results were less clear for the chloroviruses and reflected the pattern of the PCoA. While SAS and BGR chlorovirus communities were distinct, Olsha and one SAS surface samples appeared to form a group, albeit a poorly supported one.

A heatmap showed regional patterns in myovirus and chlorovirus OTU distribution, (Figure 2.4A-B) and pinpointed particular phlotypes in individual samples. For example, one replicate of SAS from the bottom consisted mainly of the myovirus OTU126 that was practically absent in the other SAS samples.

Similarly, the chlorovirus OTU239 was dominant in one of the Olsha bottom replicates, but not elsewhere.

An analysis of variance showed that myovirus Shannon indices among all pond types were significantly different ($p < 0.05$), while surface and bottom myovirus indices were not. While the SAS vs. BGR (thaw ponds) and Olsha vs. BGR chlorovirus indices were significantly different, those from SAS vs. Olsha were not. Although Olsha chlorovirus surface and bottom indices were significantly different, the thaw pond surface and bottom samples were not. Myovirus richness ranged from 307 +/- 86 phylotypes in the bottom of the SAS pond to 193 +/- 17 phylotypes in the bottom of the BGR pond, while chlorovirus richness was highest in the BGR bottom sample (26 +/- 1) and lowest in the Olsha bottom sample (7 +/- 6). Alpha-diversity values are presented in Supplementary materials (Table S2.1 and Figure S2.1).

Venn diagrams (Figure 2.5) indicated that chloroviruses shared the highest percentage of OTUs (34.1%) among the three ponds, while myovirus shared the least (5.7%). The highest percentage of unique OTUs for chloroviruses (19.5%) and myoviruses (37.6%) was in the SAS pond, while the highest percentage of chlorovirus (48.7%) and myovirus (14.2%) OTUs was shared between the two thermokarst ponds (BGR and SAS).

2.6 Discussion

Consistent with the theory that viruses are an integral part of microbial ecosystems, we found myoviruses and chloroviruses in the water column of all three pond types in this study. Viruses have previously been detected in many cold-water environments including polar lakes and seas (Chenard et al., 2015; Lopez-Bueno et al., 2009) and it is therefore unsurprising that they are also present in thermokarst ponds. The presence of viral nucleic acids indicates that viral replication is taking place which in turn suggests that viruses are active and have an influence on the microbial ecology of Northern thaw ponds.

We found that the two classes of thermokarst ponds harboured distinct myovirus and chlorovirus communities and that landscape type likely ultimately drives these differences. Viral production and viral decay determine viral community composition and these processes are affected by biotic factors

such as host metabolism and community composition, direct predation, and extracellular enzymatic activity (Chow and Suttle, 2015), as well as by abiotic factors such as exposure to UV B, and the concentration of charged particles (Suttle and Chan, 1994; Suttle and Feng, 1992). Earlier work has demonstrated that rocky-bottom, palsa and lithalsa ponds are distinct, both in terms of biochemistry and biology (Bhiry et al., 2011). These differences have resulted in divergent sources of viral production (distinct host communities in each pond type) and loss factors (elements that influence viral decay), which in turn have resulted in significant differences in the diversity of the viral communities present in the three ponds tested.

It appears that DOC concentration may be the primary driver of viral diversity. Palsa ponds, such as SAS, are rich in biologically available DOC (Laurion and Mladenov, 2013), which has substantial effects on the structure and function of aquatic microbial communities (Judd et al., 2006; Kritzberg et al., 2006). In concordance with our findings, studies of bacterial biogeography that included the same sites sampled in this study (Comte et al., 2016b), concluded that DOC was among the principal determinants of community composition, and that bacterial diversity in palsa and lithalsa ponds differed significantly. Crevecoeur et al. (2016) also found a significant difference between SAS and BGR archaeal community diversity, although only communities from bottom waters were targeted.

The primers used in the present study are designed to capture a broad range of myoviruses (Chow and Fuhrman, 2012), which typically comprise a substantial fraction of the aquatic viroplankton (Brum et al., 2013) that infect bacteria from all major taxa (King et al., 2011). If we assume that the diversity captured by our primer sets is generally representative of the total diversity of the pond phage community and that loss factors are distinct among the ponds as discussed previously, it is unsurprising that the pronounced differences in community composition of the total bacterial community between the pond types is reflected in the community composition of the viruses that infect them.

The diversity patterns of chlorovirus and those of their putative hosts (Eukaryotes) were less synchronized than myoviruses and bacteria. The most probable explanation for the lack of concurrence is that while the universal 18S primers capture a large range of eukaryotes (including photosynthetic and non-photosynthetic organisms), the chlorovirus primers target a much more limited set of viruses: those phycodnaviruses that infect chlorophytes. The difference in Venn diagram patterns between

eukaryotes and chlorviruses also supports the idea that more of the diversity of the eukaryotic community is captured by the 18S primers than the diversity of the eukaryotic virus community by the chlorovirus primers. The narrower range of the chlorovirus primers thus results in more overlap among the three pond communities, which represents a presumably small fraction of the total virus community that infects eukaryotes, while the overlap of eukaryotic OTUs among the ponds is substantially less. It is therefore more likely that changes in chlorovirus diversity would correspond with changes in chlorella diversity among the ponds than with diversity dynamics of the entire eukaryotic community. Another factor to consider is that the manner by which viral OTUs were determined may have had a major impact on the final assessment of chlorovirus diversity. In our study, we chose a cut-off of 99% based on an alignment of the regions of the DNA polymerase amplified by the chlorovirus primers of representative species of the family *Phycodnaviridae*. However, as it is evident in the chlorovirus phylogeny, all the thaw pond virotypes are very closely related (although considered different OTUs according to our criteria). It remains unknown whether these OTUs represent different strains of viruses with different host ranges or if they are genetic variations of the same viral strain, and thus whether the differences in community diversity are linked with function or not.

We expected to find different viral communities in thaw pond surface and bottom waters, but our analyses indicated that this is generally not the case. In contrast to Olsha, a rock-basin pond with a well-mixed water column, thaw ponds are characterized by a stable, vertical physico-chemical gradient (Laurion et al., 2010). Previous work has demonstrated that warm, oxygenated surface waters overlying cold, hypoxic/anoxic, bottom waters favour different microbial populations over depth (Rossi et al., 2013). Thus, we were surprised to find a relatively similar bacterial, myovirus and chlorovirus community diversity between surface and bottom layers. A likely explanation for these results is that at the depth of our bottom samples the thaw ponds were not anoxic. We detected aerobic bacterial taxa (e.g. *Comamonadaceae*, *Polynucleobacter*, etc.) in our nominally bottom samples. This suggests that either oxygen was still present at the depths sampled or that aerobic communities were introduced into our samples during collection or through intermittent mixing associated with internal waves that have been shown to transport surface microbial communities to greater depths in the water column (Deshpande et al., 2015). It is also possible that we were sampling surface layer cells and viruses that were in bottom waters due to sinking.

The difference between surface and bottom eukaryotic diversity seems to be primarily driven by an increase in photosynthetic microbes in the bottom layer in BGR (Figure 2.2 , Table 2) and a reduction in the abundance of one species of ciliate (*Strobilidium caudatum*) in the bottom layer of SAS (Figure 2.4). One possibility is that we sampled a subsurface chlorophyll maximum present in BGR, a feature that has been previously found in small stratified shallow lakes from other regions (Noges and Kangro, 2005; Pannard et al., 2015). However, the typically low-oxygen concentration at this depth suggests that these cells were not actively photosynthesizing and may have been in the process of sinking to the bottom. The relative dominance of *Strobilidium caudatum* in SAS surface waters and its low abundance in the bottom layer suggests that this organism is poorly adapted to hypoxic/anoxic, low light conditions. On the other hand, ciliates have been shown to vertically migrate in small ponds (Fenchel and Finlay, 2008) and therefore we may have simply captured a moment in time when this particular organism was in the surface water.

The classification of viruses based on environmental sequences is challenging because of the paucity of relevant genomes in the reference database and the mutability and mosaicism of viral genomes. The classification of thaw pond OTUs from BLAST-based analyses was, with a few exceptions, limited to the taxonomic level of family. For example, 97% of the myovirus OTUs could only be classified in the group "uncultured myoviridae", providing no insight into the potential hosts of these viruses, albeit strongly suggesting that thaw pond myoviruses are divergent from other myoviruses. Of the 3% of OTUs that were classified to the species level, most of them were grouped with myoviruses that infect humans or marine organisms that are unlikely to be present in a freshwater thaw pond. Similar to the myoviruses, most thaw pond chlorovirus OTUs (93%) could not be classified past the family level and were categorized as unknown members of the family *Phycodnaviridae*. In an effort to gain a better taxonomic resolution, we constructed phylogenetic trees with thaw pond chlorovirus OTUs and representative genomes from the genera of the family *Phycodnaviridae*. All of the OTUs from this study clustered most closely with the chlorovirus species *Acanthocystis turfacea* *Chlorella* virus 1, a virus that infects a *Chlorella* symbiont of a heliozoan. It is certainly plausible that sun animalcules were present in our samples (although the 18S sequencing does not support this), as freshwater is their preferred habitat. It is also possible that these viruses infect *Chlorella* species that are symbionts of other organisms such as the ciliates that were present in all the ponds. It is also possible, and perhaps more

likely, that the hosts of these viruses are the abundant free-living chlorophytes that comprised a significant percentage of the total eukaryotes.

In conclusion, myo- and chlorovirio plankton were detectable in palsa and lithalsa thaw ponds as well as a shallow rock-basin reference pond in subarctic Quebec. Although the thaw pond viral OTUs were related to viruses in the family *Myoviridae* and *Phycodnaviridae*, they appeared to be divergent from classified viruses within these groups. There was a distinct difference in viral community composition between different thaw ponds, suggesting that the permafrost landscape type selects the bacterial and eukaryotic host communities and dictates the relative importance of viral loss mechanisms, which in turn drives viral diversity in these microbial ecosystems.

2.7 Acknowledgments

We acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC), the Northern Scientific Training Program (NSTP), the Fond de Recherche du Québec – Nature et Technologies (FRQNT) and ArcticNET for their financial support. We are also grateful to Claude Tremblay who were the manager of the Whapmagoostui-Kuujuarapik CEN station, Vani Mohit, Bethany Deshpande and Alex Matveev for their help concerning field work.

2.7 Supporting information

Table S2.1 : Alpha-diversity statistics for samples grouped by organism, site and depth (S, surface ; B, bottom)

Organism	Lake	Depth	Chao1		Shannon	
			Mean	SD	Mean	SD
Myovirus	SAS	S	267,67 ± 81,25		6,67 ± 0,407	
		B	307,33 ± 86,4		6,66 ± 0,551	
	BGR	S	201,33 ± 13,58		5,02 ± 0,16	
		B	193,33 ± 13,65		5,38 ± 0,246	
	Olsha	S	265 ± 7,81		6,17 ± 0,094	
		B	250 ± 13		6,09 ± 0,104	
	Total	S	244,67 ± 52,62		5,95 ± 0,76	
		B	250,22 ± 53,57		6,04 ± 0,61	
Chlorovirus	SAS	S	19,67 ± 3,21		2,79 ± 0,42	
		B	10,67 ± 4,93		2,55 ± 0,72	
	BGR	S	24,67 ± 0,58		3,68 ± 0,09	
		B	25,67 ± 1,15		3,67 ± 0,09	
	Olsha	S	15,5 ± 4,95		2,82 ± 0,72	

		B	6,5 ± 6,36	1,44 ± 1,96
	Total	S	20,5 ± 4,63	3,13 ± 0,58
		B	15,25 ± 9,51	2,69 ± 1,25
Bacteria	SAS	S	576,67 ± 36,83	4,56 ± 0,41
		B	534 ± 64,21	4,85 ± 0,19
	BGR	S	213 ± 5,29	5,06 ± 0,04
		B	285 ± 10,15	5,04 ± 0,11
	Olsha	S	391,67 ± 20,6	6,25 ± 0,05
		B	426 ± 45,92	6,2 ± 0,19
Total	S	393,78 ± 158,91	5,29 ± 0,78	
		B	415 ± 115,23	5,36 ± 0,65
Eukaryotes	SAS	S	324,67 ± 54,6	5,58 ± 0,07
		B	283 ± 13,08	5,04 ± 0,21
	BGR	S	301,67 ± 21,01	5,31 ± 0,07
		B	260,33 ± 31,82	3,9 ± 0,61
	Olsha	S	425,67 ± 10,69	7,4 ± 0,08
		B	462 ± 52,05	7,48 ± 0,02
	Total	S	350,67 ± 64,4	6,09 ± 0,99
		B	335,11 ± 100,63	5,47 ± 1,62

At each site, n=3 for each depth. "Total" is the mean of the 3 sites (n=9). SD refers to standard deviation.

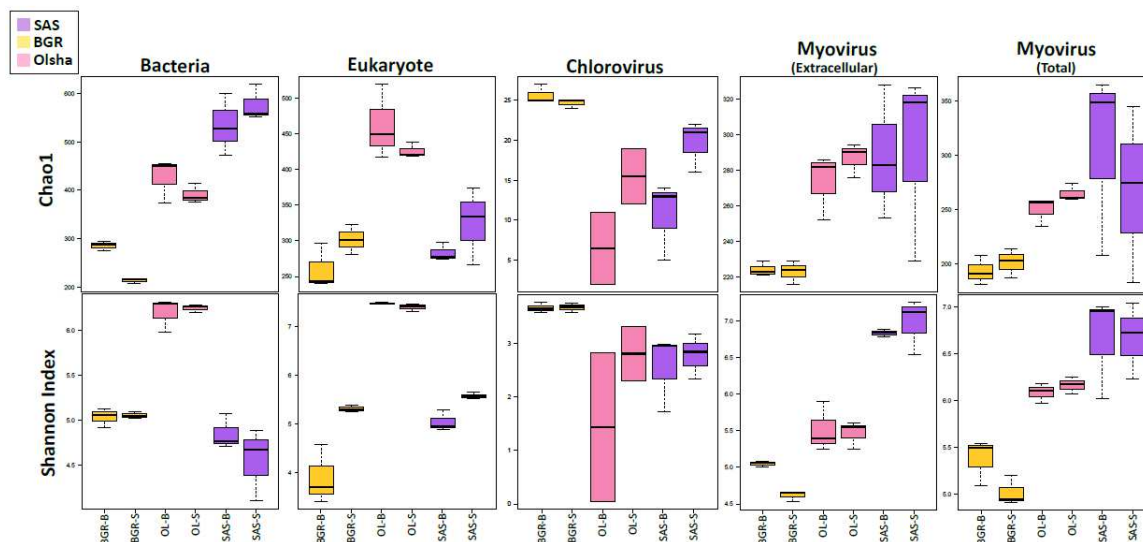


Figure S2.1: Box plots representing Chao1 richness and Shannon diversity of each sample, grouped by tested gene, according to the lake and the depth sampled. The solid black line is the median, the upper and lower limits of the box are the first and the third quartiles, and the whiskers extend to the highest and lowest datum.



Chapitre 3 : Isolement et séquençage de deux nouveaux cyanophages d'origine subarctique

Alice Lévesque^{1,2,4,5}, Catherine Marois^{2,5}, Antony T. Vincent^{2,5}, Simon J. Labrie^{2,4}, Sylvain Moineau^{2,4}, Warwick F. Vincent^{1,3} & Alexander I. Culley^{1,2,4,5}

1. Centre d'études nordiques (CEN), Université Laval, Québec, Canada
2. Département de biochimie, de microbiologie et de bio-Informatique, Université Laval, Québec, QC, Canada
3. Département de biologie, Université Laval, Québec, QC, Canada
4. Groupe de recherche en écologie buccale (GREB), Université Laval, Québec, Canada
5. Institut de biologie intégrative et des systèmes (IBIS), Université Laval, Québec, Canada

3.1 Mise en contexte

Ce troisième chapitre présente les résultats obtenus dans le cadre de mon second objectif de recherche qui visait à isoler et séquencer un nouveau cyanophage polaire. L'objectif global du projet de recherche étant d'étudier les virus demeurant dans les lacs de thermokarst subarctiques, l'isolement d'un cyanophage a pour but de mieux comprendre les interactions hôte-virus dans les milieux aquatiques nordiques et de vérifier si ces derniers ont des gènes uniques leur permettant de survivre dans des conditions extrêmes.

3.1.1 Les cyanobactéries

Les cyanobactéries (*Cyanobacteria*) constituent un embranchement au sein du domaine des procaryotes. Malgré leur apparence et fonctions écologiques similaires, les cyanobactéries ne sont pas des algues (eucaryotes) mais bien des bactéries. Elles sont dépourvues de noyau et possèdent l'ARNr 16S caractéristique des bactéries. Les cyanobactéries réagissent négativement au test de gram puisqu'elles sont enveloppées d'une paroi cellulaire composée d'une membrane externe recouvrant une couche de peptidoglycane. Sur le plan morphologique, elles sont parmi les groupes bactériens les plus diversifiés. En effet, certaines espèces ont une forme filamenteuse tandis que d'autres sont unicellulaires (Figure 3.1) (Sarma, 2012). Les cyanobactéries peuvent ainsi vivre autant sous forme de

tapis microbien que planctonique (seule ou en colonie), et ce, dans une grande variété d'habitats (sol, milieu aquatique, etc.). Elles se retrouvent également dans un large éventail d'environnements en pouvant croître notamment dans des conditions extrêmes (Antarctique, lac alcalin, sources chaudes, etc.) (Genuario et al., 2017; Ionescu et al., 2010; Jungblut et al., 2005).

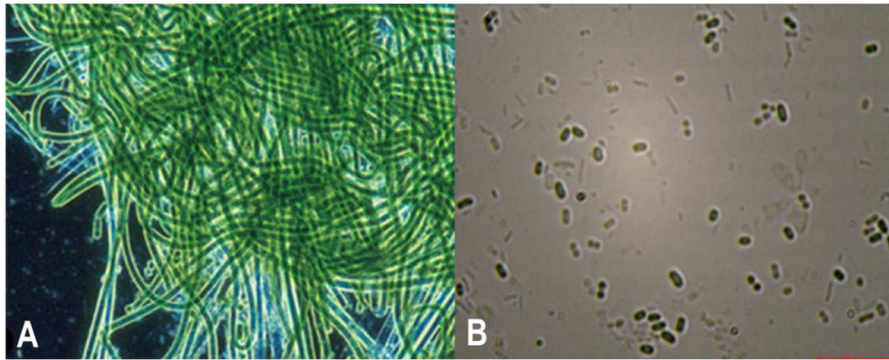


Figure 3.1: Les cyanobactéries peuvent posséder deux morphologies distinctes : filamenteuse (A) ou unicellulaire (B). Les photos en microscopie à fond clair présentent les genres *Lyngbya* sp. (A, (Paerl and Otten, 2013)) et *Cyanobium* sp. (B, (<http://bccm.belspo.be>, 2016)).

Les cyanobactéries pratiquent la photosynthèse oxygénique grâce à leurs thylakoïdes qui fixent le CO₂ et libèrent de l'oxygène (O₂) en retour (Soo et al., 2017). D'ailleurs, il est estimé que ce sont des tapis de cyanobactéries appelés stromatolithes qui ont majoritairement contribué à oxygéner l'atmosphère terrestre il y a de cela plus de 2,5 milliards d'années (Planavsky et al., 2014). Les cyanobactéries sont donc parmi les plus anciennes formes de vie sur Terre. D'autre part, certaines espèces sont également capables d'effectuer la photosynthèse anoxygénique en condition anaérobie en utilisant les sulfures comme donneurs d'électrons (Oren, 2015). D'autres espèces ont la capacité de fixer l'azote atmosphérique à l'aide de l'enzyme nitrogénase contenue dans une cellule spécialisée de la colonie appelée hétérocyste (Herrero et al., 2016). Les cyanobactéries filamenteuses ont la capacité de produire ce type de cellule lorsqu'elles sont cultivées dans un milieu sans azote combiné. Les hétérocystes, contrairement aux cellules végétatives, possèdent une paroi beaucoup plus épaisse car elle comprend une membrane supplémentaire composée de deux couches : l'une de glycolipides et l'autre de polysaccharides (Herrero et al., 2016). Dans ces cellules spécialisées, la concentration en oxygène est maintenue très basse en raison de la sensibilité élevée de la nitrogénase à l'O₂. En milieu océanique, *Prochlorococcus* est l'espèce d'autotrophes microbiens la plus abondante et celle-ci participe significativement à la productivité des océans (Partensky et al., 1999). Il est donc clair que

les cyanobactéries constituent une composante essentielle de tous les écosystèmes en raison de leur influence sur les cycles du carbone et de l'azote, en plus de générer une grande partie de l'oxygène atmosphérique.

En région polaire, les cyanobactéries sont parmi les taxons les plus abondants et écologiquement importants (Makhalanyane et al., 2015). Elles se retrouvent dans une grande variété d'habitats, autant dans le pergélisol que dans les cryoconites et les tapis microbiens au fond des lacs arctiques (Christner et al., 2003; Mohit et al., 2017). Des études réalisées en hautes latitudes ont détecté de 10^5 - 10^6 cellules/mL dans des lacs en région canadienne subarctique (Przytulska et al., 2016).

3.1.2 Les cyanophages

L'identité et les rôles écologiques des cyanophages ont été précédemment décrits dans la section 1.2.2.2. En quelques mots, les cyanophages influencent de manière significative les flux de carbone et de nutriments à travers les niveaux trophiques en infectant les cyanobactéries, constituants majeurs du phytoplancton situé à la base des écosystèmes aquatiques. Considérant le biais des bases de données pour les cyanophages isolés en eaux tempérés, il semble essentiel d'approfondir nos connaissances sur la biologie et la génomique des cyanophages provenant des milieux polaires. En effet, la découverte récente du cyanopodovirus SEIV-1 isolé du Haut-Arctique suggère que les cyanophages polaires peuvent être très divergents sur le plan génétique des autres cyanophages connus à ce jour (Chenard et al., 2015).

3.2 Matériel et Méthodes

Les différents sites d'échantillonnage et les protocoles d'échantillonnage et de filtration ont précédemment été expliqués en détails dans les sections 2.4.1 et 2.4.2 de ce présent mémoire. Voici maintenant les étapes spécifiques effectuées pour l'isolement des cyanophages.

3.2.1 Amplification des cyanophages

3.2.1.1 Cultures de cyanobactéries

Préalablement à la session d'échantillonnage, 24 cultures de cyanobactéries faisant partie de la Collection de Cyanobactéries Polaires (CEN, Université Laval) ont été sélectionnées sur la base de l'environnement d'où elles avaient été isolées à l'origine. Ces dernières provenaient de milieux arctiques et subarctiques variés présentant des similarités avec le site à l'étude (Toolik lake, Lac à l'eau claire, Île Bylot, etc.) (Tableau 3.1). Ainsi, avant de quitter pour la session d'échantillonnage, un mélange contenant 1 mL de chaque culture de cyanobactéries en phase exponentielle (24) a été préparé et conservé à 4°C et à la noirceur jusqu'à son utilisation au laboratoire de W-K. Au final, il y avait deux mélanges par lac : un pour les eaux de surface et un pour les eaux de fond.

Il est à noter que les cultures utilisées n'étaient pas axéniques, c'est-à-dire qu'elles contenaient d'autres bactéries hétérotrophes en plus des cyanobactéries. Le contenu de certaines cultures de la collection a été séquencé dans le cadre d'un précédent projet. Par contre, l'identité précise de chaque cyanobactérie n'a pas encore été établie. Toutes les cyanobactéries de la collection sont conservées en culture liquide dans environ 100 mL de milieu BG11, et sont transférées dans du nouveau milieu frais à tous les mois. L'incubateur est maintenu à une température de 11-12°C avec une intensité lumineuse continue de 33.212 $\mu\text{mol photon/m}^2\text{s}$.

Tableau 3.1 : Description des cultures de cyanobactéries sélectionnées

Numéro d'identification	Nom de la culture	Morphologie	Origine
1	CLA17*	Filamenteuse	Ward Hunt Island; neige
2	O-101*	Filamenteuse	Arctique (Toolik Lake); épilithique; étang
3	O-120*	Filamenteuse	Arctique (Toolik Lake); épilithique; étang
4	O-152*	Filamenteuse	Arctique (Ile Bylot); épilithique; étang
5	O-154*	Filamenteuse	Arctique (Ile Bylot); épilithique; étang
6	P-201*	Unicellulaire	Arctique (Toolik Lake) rivière
7	P-202*	Unicellulaire	Arctique (Toolik Lake) étang
8	P-211	Unicellulaire	Arctique (Ile Bylot); étang
9	P-212	Unicellulaire	Arctique (Ile Bylot); étang
10	P-213*	Unicellulaire	Arctique (Ile Bylot); étang
11	P-214*	Unicellulaire	Arctique (Ile Bylot); étang

12	P-215	Unicellulaire	Arctique (Ile Bylot); étang
13	O-201*	Filamenteuse	Subarctique (Poste de la Baleine); épilithique; lac
14	O-202*	Filamenteuse	Subarctique (Poste de la Baleine); épilithique; étang
15	O-203*	Filamenteuse	Subarctique (Poste de la Baleine); épilithique; étang
16	O-204*	Filamenteuse	Subarctique (Poste de la Baleine); épilithique; étang
17	O-210*	Filamenteuse	Subarctique (Lac à l'eau claire); épilithique; lac
18	O-303*	Filamenteuse	Subarctique (Lac à l'eau claire); épipsammique; lac
19	O-304*	Filamenteuse	Subarctique (Lac à l'eau claire); épipsammique; lac
20	O-310*	Filamenteuse	Subarctique (Lac à l'eau claire); épipsammique; lagune
21	O-401	Filamenteuse	Subarctique (Lac à l'eau claire); planctonique; lac
22	O-402*	Filamenteuse	Subarctique (Lac à l'eau claire); planctonique; lac
23	P-101*	Unicellulaire	Subarctique (Poste de la Baleine); étang
24	P-102a	Unicellulaire	Subarctique (Poste de la Baleine); étang

* : Cultures dont le contenu a été séquencé.

3.2.1.2 Inoculation avec les virus

À W-K, le mélange de cyanobactéries a été inoculé avec 8 mL de chaque triplicat d'eau de surface ou de fond préfiltrée sur 20 µm, 0,45 µm et 0,22 µm (voir section 2.4.2). En théorie, ce filtrat n'était censé contenir que les virus plus petits que 0,22 µm libres dans l'échantillon d'eau, les fractions cellulaire et particulaire ayant été soustraites grâce à la préfiltration. Le tout a été complété à 50 mL avec 2 mL de milieu BG11 afin de restimuler la croissance des cyanobactéries. Les cultures ont ainsi été conservées à 4°C et à la noirceur jusqu'au retour à Québec. Par contre, il est possible que cette période de transition ait causé la perte de plusieurs virus.

3.2.1.3 Rondes d'amplification

Ce premier mélange a été considéré comme la première ronde d'amplification puisque si un virus présent dans l'eau des lacs était en mesure d'infecter une des cyanobactéries contenues dans le mélange, celui-ci a été légèrement amplifié en se multipliant à l'aide de sa cellule hôte. Au laboratoire de Québec, cette première amplification a été filtrée à travers un filtre 0,45 µm (Sarstedt) pour obtenir le lysat 1. Les 24 cultures de cyanobactéries ont ensuite été inoculées individuellement dans une plaque 96 puits où, dans chaque puit, 67,5 µL de cyanobactéries en phase exponentielle était mélangées à 67,5 µL de milieu BG11 et 15 µL de lysat 1. Pour le contrôle négatif, les cultures ont été inoculées avec 15 µL d'une portion du lysat 1 ayant été chauffée à 85°C pendant 15 minutes. Cette

manœuvre avait pour but de désactiver les virus tout en conservant tout produit chimique pouvant nuire à la croissance de la cyanobactérie et qui aurait eu le même effet visuel qu'une infection virale. De plus, les tests ont été faits en duplicata. Toutes les plaques ont ensuite été incubées à 15°C avec lumière continue (22,14 $\mu\text{mol photons/m}^2\text{s}$) pour une période de 2 semaines. Les plaques ont été examinées à tous les 2-3 jours afin de détecter une anomalie dans la croissance des cultures.

Suite à cette période d'incubation, les cultures ont été filtrées à travers un filtre 0,45 μm pour obtenir le lysat 2. Les «lysats 2» ayant montré des différences de croissance entre les cultures contrôle et infectée ont été utilisés pour réaliser une troisième amplification dans un plus grand volume. Les puits d'une plaque 24 puits ont été chargés avec 900 μL de cyanobactéries mélangées à 900 μL de milieu BG11 et 200 μL de lysat 2. Les plaques ont été incubées dans les mêmes conditions que celles mentionnées précédemment. Ainsi, au bout de 7-10 jours toutes les cultures ont été filtrées sur un filtre 0,45 μm pour obtenir le lysat 3. Les étapes subséquentes ont été réalisées sur les trois échantillons ayant généré une lyse claire (L4K B3, L4K B23 et L4K S23).

3.2.2 Microscopie électronique

Les «lysats 3» des trois échantillons (L4K B3, L4K B23 et L4K S23) ayant produit une lyse claire dans les cultures infectées ont été observés au microscope électronique à transmission (TEM) afin de vérifier si l'absence de croissance dans les puits était bien due à une infection virale. Pour ce faire, une quatrième ronde d'amplification, identique à la troisième ronde, a été effectuée pour obtenir un plus grand volume de lysat (6 mL). Ce lysat a été filtré sur un filtre 0,45 μm puis 1 mL a été concentré à l'aide d'une colonne *Amicon Ultra 0.5 mL filters for Protein Purification/Concentration* (EMD Millipore) en centrifugeant 2 fois 500 μL pendant 20 minutes à 2000xg. Une troisième centrifugation a été réalisée dans les mêmes conditions pour effectuer un échange de tampon (500 μL d'acétate d'ammonium 0,1 M). Un volume final entre 25-45 μL a été récupéré pour chaque échantillon. Un volume de 10 μL a été mélangé à 10 μL d'uranyl acétate 2% sur une grille *Formvar/Carbon 200 Mesh* (Ted Pella), et le tout a été observé au microscope électronique à transmission JEOL 1230 de la Plateforme de Microscopie de l'IBIS (Université Laval, Québec). Plusieurs particules virales ont pu être observées de cette manière, et du même coup mesurées.

3.2.3 Isolement

Comme la microscopie électronique a confirmé la présence d'un titre élevé de virus dans les échantillons, il ne reste qu'à isoler et purifier une souche par échantillon en vue du séquençage. À cette fin, la technique des dilutions limites a été répétée à trois reprises en effectuant à chaque fois une série (8) de dilutions 1/10 (12 réplicats/dilution) afin de diluer au maximum les virus dans l'échantillon et d'en isoler une seule souche. Les cultures de la dilution où seulement une partie des réplicats présentait de la lyse ont été utilisées pour la seconde ronde de dilutions. En théorie, il n'y avait à l'origine qu'une seule souche de virus dans ces puits.

3.2.4 Extraction d'ADN et séquençage

Dans le but d'obtenir une quantité d'ADN suffisante pour le séquençage haut-débit, les virus ont été concentrés 40 fois à l'aide de polyéthylène glycol 8000 (PEG 8000). Un lysat de 10 mL de chaque virus a été produit, et ensuite 0,2922 g de NaCl (0,5 M) et 1 g de PEG 8000 (10%) préalablement stérilisés aux rayons ultra-violet ont été ajoutés au lysat pour effectuer la précipitation toute la nuit. Le lendemain, les lysats ont été centrifugés à 16 300x g pendant 15 minutes à 4°C et les culots ont été resuspendus dans 250 µL de tampon phage 1X (Tris-HCl 50 mM pH 7.5, NaCl 100 mM et MgSO₄ 8 mM). Les acides nucléiques totaux ont été extraits en suivant le protocole de la trousse *MasterPure Complete DNA and RNA Purification* (Epicentre) et ont été finalement resuspendus dans 65 µL de tampon TE (Tris-EDTA 1X pH 8.0, Integrated DNA Technologies). La concentration en ADN de chaque échantillon a été mesurée grâce au fluoromètre Qubit 3.0 (Invitrogen, ThermoFisher) et à la trousse associée *Molecular Probes dsDNA assay High Sensitivity* (Invitrogen).

Les bibliothèques ont été préparées à l'aide de la trousse *Next Ultra II* (DNA library prep kit for Illumina; NEB). La fragmentation s'est déroulée dans une machine *Covaris* par sonication. Les étapes de purification du protocole ont toutes été réalisées avec des billes magnétiques (Axygen). Le génome de chaque cyanophage a été séquençé par MiSeq (Illumina) à la Plateforme d'Analyses Génomiques de l'IBIS (Université Laval, Québec).

3.2.5 Analyses génomiques

Les séquences brutes ont été filtrées à l'aide de Trimmomatic (Bolger et al., 2014) dans le but de retirer les régions de mauvaise qualité et les adaptateurs Illumina™. Un échantillonnage aléatoire des lectures a été réalisé avec seqtk (Li, 2012) afin de diminuer la couverture lors de l'assemblage. Les génomes ont ensuite été assemblés *de novo* à l'aide du logiciel SPAdes (Bankevich et al., 2012) et visualisés à l'aide de Geneious v8.1.7 (Kearse et al., 2012). L'annotation des génomes a été réalisée avec le logiciel RAST en utilisant l'option «virus» (Aziz et al., 2008). Les deux génomes ont également été alignés au niveau des acides aminés pour vérifier leur similarité en utilisant d'abord BLASTp 2.2.28+ pour l'alignement, et GenePlotR intégré dans R pour la visualisation (Mercanti et al., 2016).

3.2.6 Analyses phylogénétiques

Les séquences en acides aminés du gène de la primase de B3 et B23 ont été alignées avec celles de 275 autres phages à l'aide du logiciel PROMALS3D (Pei et al., 2008). Un arbre phylogénétique «*maximum-likelihood*» a ensuite été construit en utilisant le logiciel IQ-Tree (Nguyen et al., 2015).

3.3 Résultats

3.3.1 Lyse des cultures de cyanobactéries

Lors des premiers événements de lyse dans le lysat 2, il était possible de voir des différences dans les teintes de verts et dans la texture de plusieurs cultures infectées (Figure 3.2). Par exemple, certaines cultures infectées semblaient moins étendues, et plus foncées que les cultures contrôles (Figure 3.2A). Au lysat 3, les signes de lyse étaient plus clairs dans les échantillons L4K B3, L4K B23 et L4K S23 (Figure 3.2B). Une lyse claire était observable dans les échantillons B23 et S23, tandis que des débris vert foncé demeuraient au fond des puits infectés avec B3. Par contre, les cultures contrôles #3 étaient légèrement turbides et vertes, tandis que les cultures infectées étaient claires. Les autres cultures repiquées ne présentant plus de différences significatives ont été laissées de côté.

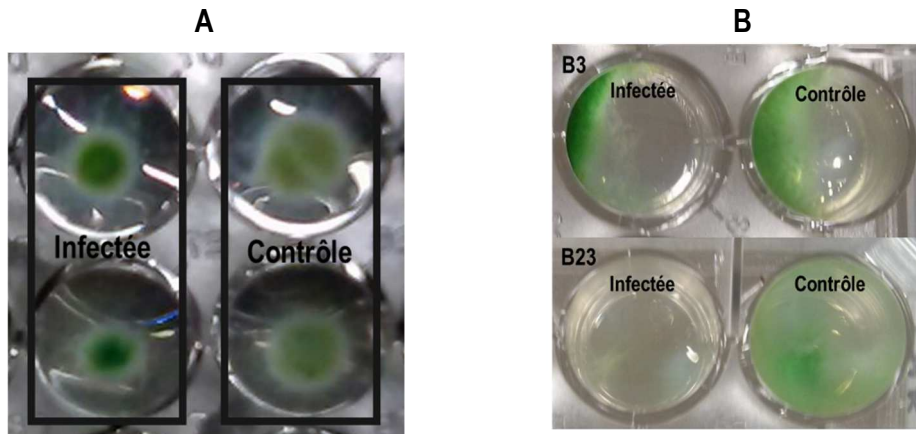


Figure 3.2: Différences observées entre les cultures contrôles et infectées lors de l'amplification des cyanophages. Exemple d'évènement de lyse obtenu dans le lysat 2 (A). Évènement de lyse obtenu dans le lysat 3 pour les échantillons B3 et B23 (B23 semblable à S23) (B).

3.3.2 Caractéristiques génomiques

L'assemblage des génomes a permis de découvrir que L4K S23 et L4K B23 possédaient un génome parfaitement similaires, suggérant qu'il s'agit du même phage (désigné B23 à partir de maintenant). Or, ce phage est différent de celui obtenu à partir de L4K B3 (désigné B3) puisque la taille de leur génome diffère d'environ 1,3 kb. Les génomes circulaires de B3 et B23 ont respectivement une longueur de 244 930 pb et 243 633 pb, et ont une identité similaire à 99,1% au niveau nucléotidique. Les deux génomes ont également un G+C% similaire, soient 35,36% pour B3 et 35,39% pour B23. La comparaison des deux génomes à la base de données du NCBI a montré que ces derniers ne correspondaient significativement à aucun virus recensé à ce jour. L'annotation des génomes a permis d'identifier la fonction potentielle de seulement 20% des séquences d'ADN codantes (CDS), tandis que 80% des CDSs ont été identifiés en tant que protéine hypothétique (Figure 3.3, Annexe Tableaux A.1 et A.2). Au total, RAST a identifié 370 CDSs chez B3, et 365 chez B23 et annoté la fonction pour seulement 75 de ces CDS pour chaque génome. Les deux génomes possèdent 18 CDSs codés sur le brin sens (-). Parmi les CDSs dont la fonction a pu être identifiée, on en retrouve deux particulièrement intéressants : deux copies d'une «*high-light inducible protein*» (*hliP*) et une «*phosphate starvation-inducible protein*» (*phoH*) tous deux reconnus comme AMGs (Figure 3.3). La phylogénie du gène de la primase a aussi permis de constater que B3 et B23 tendent à se regrouper avec d'autres phages de *Synechococcus* et de *Prochlorococcus* plutôt qu'avec des phages d'*Enterobacteriaceae* et d'*Aeromonas*, suggérant qu'ils sont plus rapprochés génétiquement des cyanophages, et donc qu'ils

seraient bel et bien des virus infectant des cyanobactéries (Figure 3.4). L'annotation a également pu identifier 21 ARN de transfert (ARNt) chez B3, et 19 chez B23.

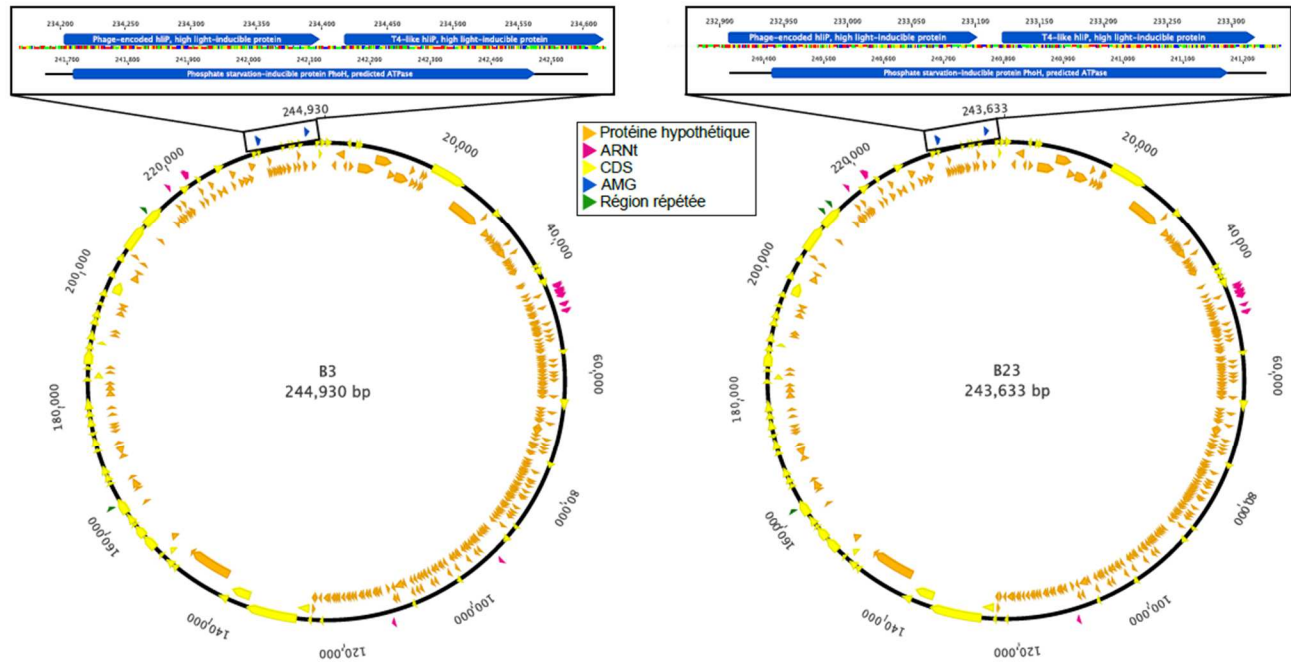


Figure 3.3 : Représentation du génome des cyanophages B3 et B23. Un agrandissement de la région des AMGs (bleu) permet de discerner deux copies du gène *hliP*. Les différentes couleurs correspondent aux protéines hypothétiques (orange), ARNt (rose), CDS avec fonction identifiée (jaune) et régions répétées (vert).

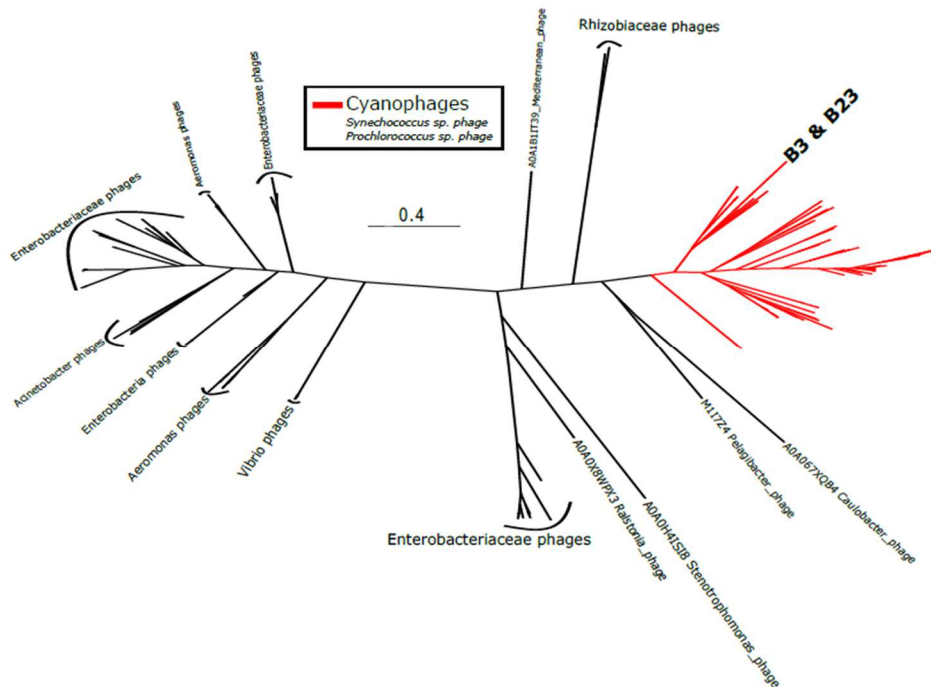


Figure 3.4 : Arbre phylogénétique sans racine basé sur l’alignement au niveau des acides aminés du gène codant pour la primase des cyanophages B3 et B23 avec ceux de 275 autres phages.

L’alignement des deux génomes au niveau des acides aminés a démontré qu’une région en particulier était divergente entre les deux génomes (Annexe Figure A.1). En effet, cette région située à 106 kb du génome de B3 possède une identité >90% lorsque comparée au génome de B23, contrairement au reste du génome dont l’identité est majoritairement supérieure à 99%. La différence réside dans le fait qu’on retrouve quatre CDSs dans le génome de B3 qui sont complètement absents de B23. Il existe également un CDS dont l’identité est de 95% entre les deux génomes, et celui-ci présente une homologie avec un antigène précurseur sécrété SsaA («*Secretory antigen precursor SsaA*»), dont le rôle précis n’est pas encore identifié mais l’annotation suggère un rôle potentiel dans la pathogénicité de certains *staphylocoques* (CDS314 et CDS309). Les différentes teintes de gris dans l’alignement illustrent que l’identité au niveau nucléotidique est également plus basse.

3.3.3 Morphologie des particules virales

L’observation au microscope électronique à transmission a permis de constater une morphologie typique des virus de la famille des *Myoviridae* dans chaque échantillon (Figure 3.5). En effet, les particules virales présentent une morphologie complexe composée d’une capsidie icosaédrique et

d'une longue queue contractile. La capacité de la queue à se contracter a été observée dans les échantillons de B23 seulement (Figure 3.5 C). Les fibres de queue se sont probablement détachées lors de la préparation des échantillons, c'est pourquoi elles ne sont pas visibles sur les photos. En moyenne, les capsides ont un diamètre de $107,94 \pm 3,18$ nm et les queues une longueur de $155,56 \pm 4,78$ nm pour le phage B3 (n=28), et $110,32 \pm 3,28$ nm et $147,82 \pm 2,76$ nm pour le phage B23 (n=28).

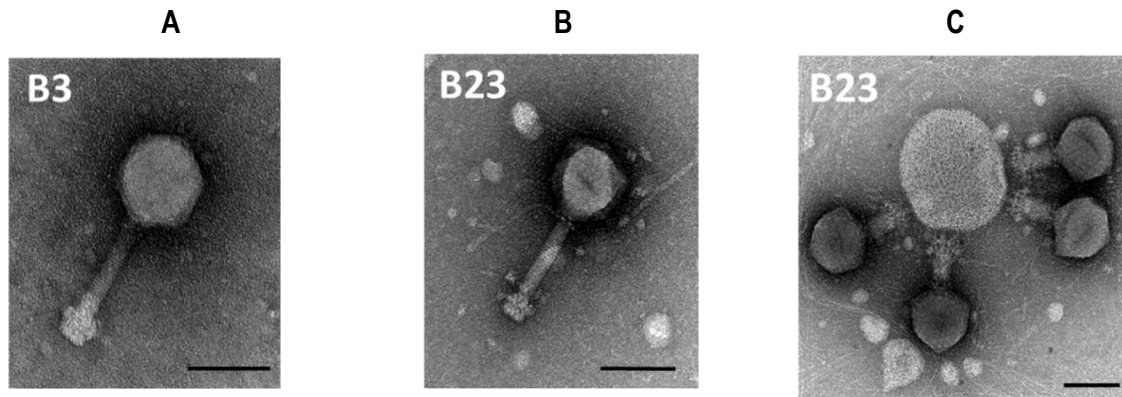


Figure 3.5 : Photos obtenues en TEM des deux cyanophages isolés. Elles illustrent B3 (A), B23 (B) et B23 sous sa forme contractée (C). La barre d'échelle représente 100 nm.

Chapitre 4 : Discussion générale et conclusion

Ce projet de recherche aura permis dans son ensemble d'apporter de nouvelles connaissances dans le domaine de l'écologie virale arctique, autant sur les aspects moléculaires que biologiques. Il sera question dans cette section de discuter des résultats obtenus lors de cette étude, et de compléter ce qui a déjà été mentionné dans le cadre du premier objectif de recherche. Une conclusion du projet suivie de ses perspectives compléteront ce quatrième et dernier chapitre.

4.1 Discussion

4.1.1 Objectif de recherche 1

Ces données obtenues sur les communautés virales dans les lacs de thermokarst subarctiques de la région de W-K viennent compléter celles récoltées par le passé sur les bactéries, les eucaryotes (phytoplancton et zooplancton) et les archées (Bégin and Vincent, 2017; Comte et al., 2016b; Crevecoeur et al., 2016; Przytulska et al., 2016). En combinant toutes ces études, il est maintenant plus facile d'avoir un portrait complet de la diversité biologique et des interactions qui ont lieu à l'intérieur de ces lacs. Il est probable que la productivité virale soit fortement favorisée dans ces lacs puisqu'en période estivale les lacs de fonte sont caractérisés par une productivité bactérienne très élevée jumelée à un taux de croissance bactérien de 4 à 7 fois plus élevé que dans d'autres types de lacs (Deshpande et al., 2016; Roiha et al., 2015). L'apport constant de carbone organique et inorganique provenant de la dégradation du pergélisol stimule sans cesse le métabolisme et la croissance des micro-organismes demeurant dans la colonne d'eau. Ce processus avantage du même coup la multiplication et la propagation des virus dans ces milieux en raison de l'augmentation du taux de contact entre les virus et leurs hôtes. L'étude de Deshpande et al. (2016) suggère notamment que la lyse virale participerait activement, de concert avec les brouteurs, au déclin des populations bactériennes mesuré lors d'expériences *in situ*. Ainsi, ceci expliquerait la présence de myovirus dans le milieu et confirmerait que ces derniers sont bel et bien actifs.

Les assemblages viraux obtenus dans le cadre de cette étude montrent les mêmes patrons de distributions que ceux observés dans de précédentes études réalisées dans les mêmes lacs sur les bactéries (Comte et al., 2016b; Crevecoeur et al., 2015). En effet, toutes ces études, incluant la nôtre,

ont démontré une différenciation des communautés selon le type de lac de fonte (SAS-2A, pale versus BGR-1, lithalse). Ce phénomène est dû en partie aux conditions physico-chimiques distinctes dans les deux types de lac (Deshpande et al., 2015). En effet, en plus de la concentration en DOC plus élevée à SAS 2A, ce site possède un pH plus acide, tandis que BGR-1 présente une conductivité plus élevée. Leur concentration en phosphore et en azote diffèrent également. Tous ces paramètres abiotiques sont reconnus pour influencer significativement la composition des populations microbiennes, et cela se reflète dans la composition des échantillons aux différents sites (β -diversité) (Langenheder et al., 2012; Peeters et al., 2012; Percent et al., 2008). De la même manière, les communautés virales varient en fonction de la distribution leur hôte cellulaire, c'est pourquoi les propriétés limnologiques distinctes de chaque lac entraînent indirectement la présence de virus appartenant à des phylotypes et écotypes différents.

L'utilisation d'un seuil de similarité de 99% et d'une longueur minimale de 400 pb lors du traitement des séquences de chlorovirus ont pu causer des pertes importantes de diversité. Ce seuil de similarité peut sembler très conservateur mais est déjà utilisé pour définir les OTUs viraux à partir d'autres paires d'amorces (Clasen et al., 2013; Needham et al., 2017). Or, après avoir testé différents seuils et longueurs minimales, ce sont ces deux paramètres qui impliquaient la plus basse proportion de séquences cellulaires non-désirées dans notre étude. Une grande quantité de séquences a tout de même dû être retirée du jeu de données pour ne conserver que 39 séquences par échantillons, ce qui est infime comparé aux jeux de données des myovirus (28 511), bactéries (27 396) et eucaryotes (27 221). En conséquence, il est possible que les résultats statistiques obtenus avec les OTUs de chlorovirus soient moins fort statistiquement que les trois autres groupes testés et qu'ils représentent moins bien la diversité virale réelle dans les lacs de fonte (Caporaso et al., 2011). De plus, en ne conservant que les séquences identifiées comme *Phycodnaviridae*, il est possible qu'un biais ait été créé envers les séquences connues, et que des chlorovirus encore inconnus ait pu être laissés de côté. Par exemple, 6 OTUs ont été identifiés comme «Virus» et 2 OTUs ont été classés dans la catégorie «virus à ADN db sans étape à ARN» sans taxonomie plus détaillée. Il existe donc une possibilité que parmi ces 8 OTUs, il aurait pu y avoir d'autres *Phycodnaviridae*.

4.1.2 Objectif de recherche 2

B3 et B23 constituent deux nouveaux membres du grand groupe des cyanomyovirus dont les représentants sont largement répandus à travers le monde autant en eau douce qu'en eau marine (Deng and Hayes, 2008; Sabehi et al., 2012). Il semble également que B3 et B23 (243-244 kb) possèdent parmi les plus longs génomes recensés à ce jour chez les cyanomyovirus qui généralement oscillent entre 170-200 kb, avec quelques exceptions allant jusqu'à 232 et 252 kb (Crummett et al., 2016). Le G+C% de chaque génome est plus bas que la moyenne, mais est semblable à celui de cyanomyovirus (P-SSM2 et P-SSM5) à gros génomes (252 kb) isolés de la Mer des Sargasses infectant *Prochlorococcus* sp. (Sullivan et al., 2005; Sullivan et al., 2010). Par contre, B3 et B23 possèdent beaucoup plus d'ARNt (19-21 contre un seul) et un nombre moindre d'AMGs (2 contre 26) que ces derniers phages de *Prochlorococcus*. D'ailleurs, le nombre d'AMGs est étonnamment très bas comparés aux autres cyanomyovirus qui en possèdent pour la plupart >10. Il est possible que l'annotation n'ait pas identifié la fonction de tous les gènes en raison d'un seuil d'homologie trop élevé utilisé dans RAST. De surcroît, le nombre de CDS prédit pour B3 et B23, soit 370 et 365, est plus élevé que tous les cyanomyovirus recensés dans l'étude de Crummett et al. (2016) qui possèdent entre 204-334 CDS. Ces résultats suggèrent une fois de plus à quel point les deux cyanophages isolés dans ce projet sont distincts des autres cyanomyovirus étudiés jusqu'à maintenant.

Les 2 AMGs (*hliP* et *phoH*) retrouvés dans chaque génome sont parmi les plus communément retrouvés chez les cyanomyovirus (Crummett et al., 2016). Concernant *hliP*, il est courant que ce gène soit présent en plusieurs copies dans le génome viral, avec certains phages tel que P-SSM2 possédant jusqu'à six copies de *hliP* (Crummett et al., 2016). Or, il est estimé que les cyanophages d'origine océanique (haute mer) auront tendance à posséder un plus grand nombre de copies de *hliP* comparés aux phages d'eaux côtières en raison du haut niveau de radiation à la surface des océans (Crummett et al., 2016). Ces protéines auraient comme fonction de dissiper l'énergie lumineuse en excès afin de protéger l'appareil photosynthétique de l'hôte en condition d'intense radiation solaire (He et al., 2001). Une étude a également précédemment suggéré que le nombre de copies d'*hliP* serait fortement influencé par une sélection environnementale (Lindell et al., 2004). Ainsi, le faible nombre de copies de *hliP* dans B3 et B23 pourrait être en raison de leur origine subarctique et lacustre, où la radiation solaire est moins intense durant l'année. Quant à *phoH*, une analyse des génomes recensés dans la GenBank a démontré que 40% des cyanophages marins possédait ce gène tandis que seulement 4%

des cyanophages d'origine non-marine portait ce gène (Goldsmith et al., 2011). Ainsi, B3 et B23 seraient parmi les rares cyanophages d'eau douce à posséder cet AMG. Une récente étude réalisée sur l'eau des rizières en Chine a également démontré une grande diversité du gène *phoH* en utilisant les mêmes amorces que Goldsmith et al. 2011 (Wang et al., 2016). Il serait donc intéressant d'investiguer davantage la présence du gène *phoH* dans les écosystèmes d'eau douce puisque sa diversité dans les milieux non-salés semble fortement sous-estimée dans les bases de données. La fonction précise de *phoH* (ATPase) n'a pas encore été établie, mais celui-ci ne serait pas impliqué dans le stress dû au manque de phosphate contrairement à ce qui a déjà été suggéré (communication personnelle, Lindell D., 2017).

Par contre, il est étonnant que le gène *psbA* n'ait pas été annoté dans aucun des deux génomes. En effet, il avait jusqu'à maintenant été répertorié dans tous les génomes de cyanomyovirus, à l'exception des deux isolats de S-CAM7 (0910CC49 et 0910SB42) qui représentent un clade distinct (Crummett et al., 2016; Sullivan et al., 2010). Ces derniers ont tous deux été isolés des eaux Pacifique californiennes. L'origine géographique n'influencerait donc pas la présence du gène *psbA* dans le génome phagique. Il y a probablement d'autres facteurs environnementaux et évolutifs qui ont conduit la majorité des cyanomyovirus à acquérir *psbA*. Ce gène codant pour la protéine D1 du photosystème II est considéré parmi les gènes les mieux conservés chez les cyanomyovirus, et fait partie des AMGs de cœur avec *mazG*, *phoH*, *cobS* et *hsp* (Sullivan et al., 2010). L'absence de ce gène suggère alors que les cyanophages B3 et B23 ont recours à une stratégie différente pour combler les besoins énergétiques, et qu'ils ne dépendent pas du maintien de la photosynthèse pour pouvoir se répliquer lors de l'infection virale. Il est à noter que *mazG*, *cobS* et *hsp* n'ont pas non plus été annotés dans les deux génomes.

L'identification d'AMGs dans les deux cyanophages suggère que des événements de transferts horizontaux de gènes se sont produits entre les virus et leur hôte par le passé. En effet, l'acquisition des AMGs par les virus à partir de leur hôte a été démontré dans plusieurs cas, avec les gènes *hli* et de photosynthèse notamment (Ignacio-Espinoza and Sullivan, 2012; Lindell et al., 2004; Sharon et al., 2009). Dans le cas de B3 et B23, les mécanismes par lesquels ils ont acquis ces gènes sortent du cadre de ce projet, mais il est tout même intéressant de constater que les virus arctiques (Chenard et

al., 2015) ont pu adopter des stratégies évolutives similaires à ceux retrouvés en milieux tempérés et océaniques (Sullivan et al., 2010).

La phylogénie de la primase indique clairement que B3 et B23 appartiennent au groupe des cyanophages. En effet, ils sont génétiquement plus rapprochés des virus infectant *Prochlorococcus sp.* et *Synechococcus sp.* que des virus infectant des bactéries du genre *Aeromonas sp.* ou de la famille des *Enterobacteriaceae*. Ainsi, malgré que les cultures de cyanobactéries n'étaient pas axéniques, il est possible d'affirmer que les cellules hôtes sont des cyanobactéries. Par contre, l'hôte n'a pas encore été identifié officiellement. Une première analyse morphologique a déterminé que les hôtes de B3 et B23 sont respectivement de type filamenteuse (#3) et unicellulaire (#23). Or, une analyse par séquençage MiSeq plus ou moins concluante (moins de 10X de couverture) suggère que les deux cultures seraient mixtes et composées de *Cyanobium* et de *Synechococcus*, deux espèces unicellulaires. Ce qui demeure curieux puisque la culture #3 produit bel et bien des filaments lorsque cultivée en milieu liquide et solide. Il reste donc à purifier les cultures et à séquencer les cyanobactéries hôtes. À ce jour, il n'existe aucun phage infectant *Cyanobium sp.* dans les bases de données, ce qui constituerait une première.

La faible proportion de CDS annotés est un problème communément rencontré lors de l'isolation de nouveaux cyanophages (Chenard et al., 2015; Coloma et al., 2017). Il est possible que certains CDS codent pour des protéines connues, mais dont la séquence est si divergente que le logiciel d'annotation est incapable de reconnaître l'homologie en raison du seuil d'identité par défaut trop élevé. Une solution serait de réaliser une seconde analyse en diminuant cette fois-ci le seuil d'identité afin de permettre l'annotation d'un plus grand nombre de CDS. Aussi, malgré qu'une «protéine majeure de la capsid» (CDS269, B3 et CDS264, B23) et qu'un «phage portal vertex of the head» (CDS261, B3 et CDS256, B23) aient été identifiées dans chaque génome, les gènes généralement conservés à travers les myovirus *gp20* et *gp23* n'ont pas été formellement annotés. En effet, ces deux gènes sont habituellement utilisés comme gènes marqueurs des virus de la famille des *Myoviridae* (Chow and Fuhrman, 2012; Sullivan et al., 2008). Leur absence peut donc soulever des questions sur la précision à laquelle les génomes ont été annotés, ou encore sur la possibilité que ces deux cyanophages présentent des structures distinctes des autres myovirus.

4.2 Conclusion

Ce projet de recherche avait pour but d'étudier la diversité des assemblages viraux dans les mares de thermokarst subarctiques situées au nord du Québec. Deux approches divisées en deux objectifs de recherche ont été utilisées dans le but de caractériser les virus dans ces écosystèmes : une approche moléculaire pour établir la diversité spécifique des myovirus et des chlorovirus, et une approche par culture afin d'isoler et d'étudier plus en détails la biologie d'un cyanophage originaire de ces mares.

Les résultats obtenus dans le cadre du premier objectif de recherche nous permettent de conclure que les mares de fonte subarctiques abritent des communautés virales hautement diverses, surtout du côté des myovirus. Les analyses bio-informatiques et statistiques ont révélé des patrons de distributions intéressants concernant les communautés virales, bactériennes et eucaryotes. En effet, il est clair que SAS (palse) et BGR (lithalse) possèdent des assemblages viraux et microbiens significativement distincts dû à un processus de filtration environnementale favorisant différents types de bactéries et de micro-eucaryotes (hôtes) dans les deux mares. De cette façon, la distribution des populations virales suit celle de leurs hôtes qui varie en fonction des conditions physico-chimiques retrouvées dans chaque mare. Ainsi, comme les palse et les lithalse ont des propriétés limnologiques distinctes, il est logique de retrouver des populations différentes à SAS et à BGR.

Le second objectif a permis, quant à lui, de découvrir deux nouveaux cyanophages appartenant à la famille des *Myoviridae* (B3 et B23) inconnus jusqu'à maintenant des bases de données. Les analyses génomiques ont démontré la présence de 2 AMGs (*phoH* et *hliP*) communément retrouvés à travers les cyanophages. Par contre, comme une grande proportion des génomes demeure inconnue et que rien ne leur correspond dans les bases de données, cela suggère que ces 2 cyanophages d'origine polaires seraient très divergents génétiquement des autres cyanomyovirus découverts par le passé. En effet, ils infectent possiblement *Cyanobium sp.* dont aucun phage n'a encore été séquencé. La découverte de B3 et B23 permettra d'en apprendre davantage sur la biologie des cyanophages, et surtout sur la dynamique des phages arctiques, étant donné le peu de données existantes sur l'écologie des virus polaires.

4.2.1 Perspectives

Il existe plusieurs possibilités concernant l'avenir de ce projet. D'abord, il serait capital d'échantillonner un plus grand nombre de lac par région afin de pouvoir comparer les communautés virales entre les régions (palse et lithalse) et établir des tendances statistiquement valables. Comme il a été établi dans de précédentes études, il existe une grande variabilité intra-régionale concernant les populations microbiennes ainsi que les propriétés physico-chimiques. Le profil limnologique de chaque lac devrait également être mesuré dans le but de pouvoir réaliser des analyses statistiques qui incluent les paramètres environnementaux. Ces dernières pourraient mettre en évidence des conditions environnementales influençant de manière significative la composition des populations microbiennes et virales. Des réactions PCR visant d'autres familles ou groupes de virus pourraient être effectuées sur les échantillons utilisés dans la présente étude pour tracer un portrait plus détaillé de la communauté virale. Il serait intéressant de tester des amorces ciblant les *Podoviridae*, les *Siphoviridae* ou encore les virus géants puisque ces derniers sont présents en abondance dans les écosystèmes aquatiques et nordiques. La diversité des virus à ARN pourrait également constituer un projet intéressant, compte tenu de leur importance écologique majeure. Enfin, la présence et la diversité des AMG dans les lacs de thermokarst pourraient également être évaluées en testant, par exemple, une paire d'amorce amplifiant le gène *phoH*. Bref, il va sans dire qu'il demeure essentiel d'effectuer davantage de recherches sur la diversité des communautés virales en milieu subarctique, étant donné le peu de donnée disponible à ce jour.

Du côté de B3 et B23, il serait intéressant de pouvoir les quantifier par qPCR ou cytométrie en flux à l'aide d'un fluorochrome spécifique dans les échantillons récoltés en 2015, mais aussi dans de futurs échantillons dans le but de vérifier leur abondance dans les lacs de la région de W-K. Des analyses génomiques et phylogénétiques supplémentaires sont également requises pour connaître avec certitude à quels clade et phylotype appartiennent B3 et B23. Il serait également important d'investiguer davantage la nature des CDS (75%) dont la fonction n'a pas pu être identifiée. Il faudrait aussi comparer les génomes de B3 et B23 avec des métaviromes obtenus à l'échelle mondiale dans le but de définir leur biogéographie. Une fois que les hôtes seront identifiés, il sera possible de caractériser le cycle de reproduction de chaque cyanophage (taille de la progéniture, temps de latence, etc.). Finalement, une étude de métatranscriptomique pourrait permettre d'évaluer l'activité et l'impact des AMG (*phoH* et

hliP) au cours de l'infection virale. Ainsi, il en reste encore beaucoup à découvrir du génome et de la biologie de ces deux cyanomyovirus.

Bibliographie

- Abergel, C., M. Legendre, and J.M. Claverie. 2015. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiology Reviews*. 39:779-796.
- ACIA. 2005. Arctic Climate Impact Assessment. Cambridge University Press, Cambridge.
- Ackermann, H.W. 1999. Tailed bacteriophages: The order *Caudovirales*. *Advances in Virus Research*, Vol 51. 51:135-201.
- Allard, M., and M.K. Seguin. 1987. The Holocene evolution of permafrost near the tree line, on the eastern coast of Hudson-bay (Northern Quebec). *Canadian Journal of Earth Sciences*. 24:2206-2222.
- Anesio, A.M., B. Mindl, J. Laybourn-Parry, A.J. Hodson, and B. Sattler. 2007. Viral dynamics in cryoconite holes on a high Arctic glacier (Svalbard). *Journal of Geophysical Research-Biogeosciences*. 112:10.
- Anisimov, O.A., N.I. Shiklomanov, and F.E. Nelson. 1997. Global warming and active-layer thickness: results from transient general circulation models. *Global and Planetary Change*. 15:61-77.
- Arlen-Pouliot, Y., and N. Bhiry. 2005. Palaeoecology of a palsa and a filled thermokarst pond in a permafrost peatland, subarctic Quebec, Canada. *Holocene*. 15:408-419.
- Azam, F., T. Fenchel, J.G. Field, J.S. Gray, L.A. Meyerreil, and F. Thingstad. 1983. The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*. 10:257-263.
- Aziz, R.K., D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko. 2008. The RAST server: Rapid annotations using subsystems technology. *BMC Genomics*. 9:15.
- Bankevich, A., S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Pribelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, and P.A. Pevzner. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. 19:455-477.
- Bellas, C.M., A.M. Anesio, J. Telling, M. Stibal, M. Tranter, and S. Davis. 2013. Viral impacts on bacterial communities in Arctic cryoconite. *Environmental Research Letters*. 8:9.
- Berger, S.A., D. Krompass, and A. Stamatakis. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic Biology*. 60:291-302.
- Bergh, O., K.Y. Borsheim, G. Bratbak, and M. Heldal. 1989. High abundance of viruses found in aquatic environments. *Nature*. 340:467-468.
- Bhiry, N., A. Delwaide, M. Allard, Y. Begin, L. Filion, M. Lavoie, C. Nozais, S. Payette, R. Pienitz, E. Saulnier-Talbot, and W.F. Vincent. 2011. Environmental change in the Great Whale River region, Hudson Bay: Five decades of multidisciplinary research by Centre d'etudes nordiques (CEN). *Ecoscience*. 18:182-203.
- Bidle, K.D., L. Haramaty, J.B.E. Ramos, and P. Falkowski. 2007. Viral activation and recruitment of metacaspases in the unicellular coccolithophore, *Emiliana huxleyi*. *Proceedings of the National Academy of Sciences of the United States of America*. 104:6049-6054.
- Birge, E.A. 2006. Transduction. In *Bacterial and Bacteriophage Genetics*. E.A. Birge, editor. Springer New York, New York, NY. 305-325.
- Boeckli, L., A. Brenning, S. Gruber, and J. Noetzli. 2012. Permafrost distribution in the European Alps: calculation and evaluation of an index map and summary statistics. *Cryosphere*. 6:807-820.
- Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114-2120.
- Boras, J.A., M.M. Sala, J.M. Arrieta, E.L. Sa, J. Felipe, S. Agusti, C.M. Duarte, and D. Vaque. 2010. Effect of ice melting on bacterial carbon fluxes channelled by viruses and protists in the Arctic Ocean. *Polar Biology*. 33:1695-1707.

- Borrel, G., D. Jezequel, C. Biderre-Petit, N. Morel-Desrosiers, J.P. Morel, P. Peyret, G. Fonty, and A.C. Lehours. 2011. Production and consumption of methane in freshwater lake ecosystems. *Research in Microbiology*. 162:832-847.
- Bouchard, F., P. Francus, R. Pienitz, I. Laurion, and S. Feyte. 2014. Subarctic thermokarst ponds: Investigating recent landscape evolution and sediment dynamics in thawed permafrost of northern Quebec (Canada). *Arctic Antarctic and Alpine Research*. 46:251-271.
- Bouchard, F., L.A. MacDonald, K.W. Turner, J.R. Thienpont, A.S. Medeiros, B.K. Biskaborn, J. Korosi, R.I. Hall, R. Pienitz, and B.B. Wolfe. 2016. Paleolimnology of thermokarst lakes: a window into permafrost landscape evolution. *Arctic Science*. 3:91-117.
- Breitbart, M. 2012. Marine Viruses: Truth or Dare. In *Annual Review of Marine Science*, Vol 4. Vol. 4. C.A. Carlson and S.J. Giovannoni, editors. Annual Reviews, Palo Alto. 425-448.
- Breitbart, M., L.R. Thompson, C.A. Suttle, and M.B. Sullivan. 2007. Exploring the vast diversity of marine viruses. *Oceanography*. 20:135-139.
- Breton, J., C. Vallieres, and I. Laurion. 2009. Limnological properties of permafrost thaw ponds in northeastern Canada. *Canadian Journal of Fisheries and Aquatic Sciences*. 66:1635-1648.
- Briee, C., D. Moreira, and P. Lopez-Garcia. 2007. Archaeal and bacterial community composition of sediment and plankton from a suboxic freshwater pond. *Research in Microbiology*. 158:213-227.
- Brown, C.M., J.E. Lawrence, and D.A. Campbell. 2006. Are phytoplankton population density maxima predictable through analysis of host and viral genomic DNA content? *Journal of the Marine Biological Association of the United Kingdom*. 86:491-498.
- Brown, J., O.J. Ferrians, J.A. Heginbottom, and E.S. Melnikov. 1998. revised February 2001. Circum-Arctic Map of Permafrost and Ground Ice Conditions. Boulder, CO: National Snow and Ice Data Center; Digital media.
- Brum, J.R., R.O. Schenck, and M.B. Sullivan. 2013. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *ISME Journal*. 7:1738-1751.
- Brussaard, C.P.D. 2004. Viral control of phytoplankton populations - a review. *Journal of Eukaryotic Microbiology*. 51:125-138.
- Brussaard, C.P.D., R.S. Kempers, A.J. Kop, R. Riegman, and M. Heldal. 1996. Virus-like particles in a summer bloom of *Emiliania huxleyi* in the North Sea. *Aquatic Microbial Ecology*. 10:105-113.
- Bubeck, J.A., and A.J.P. Pfitzner. 2005. Isolation and characterization of a new type of chlorovirus that infects an endosymbiotic *Chlorella* strain of the heliozoon *Acanthocystis turfacea*. *Journal of General Virology*. 86:2871-2877.
- Buchfink, B., C. Xie, and D.H. Huson. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*. 12:59-60.
- Buskey, E.J., and C.J. Hyatt. 1995. Effects of the Texas (USA) brown-tide algae on planktonic grazers. *Marine Ecology Progress Series*. 126:285-292.
- Bégin, P.N., and W.F. Vincent. 2017. Permafrost thaw lakes and ponds as habitats for abundant rotifer populations. *Arctic Science*. 3:354-377.
- Calmels, F., M. Allard, and G. Delisle. 2008. Development and decay of a lithalsa in Northern Quebec: A geomorphological history. *Geomorphology*. 97:287-299.
- Caporaso, J.G., J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*. 7:335-336.
- Caporaso, J.G., C.L. Lauber, W.A. Walters, D. Berg-Lyons, C.A. Lozupone, P.J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*. 108:4516-4522.
- Chappellaz, J., T. Blunier, S. Kints, A. Dallenbach, J.M. Barnola, J. Schwander, D. Raynaud, and B. Stauffer. 1997. Changes in the atmospheric CH₄ gradient between Greenland and Antarctica during the Holocene. *Journal of Geophysical Research-Atmospheres*. 102:15987-15997.

- Chen, F., and C.A. Suttle. 1995. Amplification of DNA-polymerase gene fragments from viruses infecting microalgae. *Applied and Environmental Microbiology*. 61:1274-1278.
- Chen, H., and P.C. Boutros. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *Bmc Bioinformatics*. 12:7.
- Chen, M., J.C. Rowland, C.J. Wilson, G.L. Altmann, and S.P. Brumby. 2014. Temporal and spatial pattern of thermokarst lake area changes at Yukon Flats, Alaska. *Hydrological Processes*. 28:837-852.
- Chenard, C., A.M. Chan, W.F. Vincent, and C.A. Suttle. 2015. Polar freshwater cyanophage S-EIV1 represents a new widespread evolutionary lineage of phages. *ISME Journal*. 9:2046-2058.
- Chenard, C., and C.A. Suttle. 2008. Phylogenetic diversity of sequences of cyanophage photosynthetic gene psbA in marine and freshwaters. *Applied and Environmental Microbiology*. 74:5317-5324.
- Chow, C.-E.T., and J.A. Fuhrman. 2012. Seasonality and monthly dynamics of marine myovirus communities. *Environmental Microbiology*. 14:2171-2183.
- Chow, C.E.T., and C.A. Suttle. 2015. Biogeography of viruses in the sea. *Annual Review of Virology*. 2:41-66.
- Christner, B.C., B.H. Kvitko, and J.N. Reeve. 2003. Molecular identification of Bacteria and Eukarya inhabiting an Antarctic cryoconite hole. *Extremophiles*. 7:177-183.
- Clasen, J.L., C.A. Hanson, Y. Ibrahim, C. Weihe, M.F. Marston, and J.B.H. Martiny. 2013. Diversity and temporal dynamics of Southern California coastal marine cyanophage isolates. *Aquatic Microbial Ecology*. 69:17-+.
- Clasen, J.L., and C.A. Suttle. 2009. Identification of freshwater *Phycodnaviridae* and their potential phytoplankton hosts, using DNA pol sequence fragments and a genetic-distance analysis. *Applied and Environmental Microbiology*. 75:991-997.
- Colangelo-Lillis, J., B.A. Wing, and L.G. Whyte. 2016. Low viral predation pressure in cold hypersaline Arctic sediments and limits on lytic replication. *Environmental Microbiology Reports*. 8:250-260.
- Coloma, S.E., A. Dienstbier, D.H. Bamford, K. Sivonen, E. Roine, and T. Hiltunen. 2017. Newly isolated Nodularia phage influences cyanobacterial community dynamics. *Environmental Microbiology*. 19:273-286.
- Comeau, A.M., and H.M. Krisch. 2005. War is peace - dispatches from the bacterial and phage killing fields - Commentary. *Current Opinion in Microbiology*. 8:488-494.
- Comeau, A.M., W.K.W. Li, J.E. Tremblay, E.C. Carmack, and C. Lovejoy. 2011. Arctic ocean microbial community structure before and after the 2007 record sea ice minimum. *Plos One*. 6:12.
- Comte, J., C. Lovejoy, S. Crevecoeur, and W.F. Vincent. 2016a. Co-occurrence patterns in aquatic bacterial communities across changing permafrost landscapes. *Biogeosciences*. 13:175-190.
- Comte, J., A. Monier, S. Crevecoeur, C. Lovejoy, and W.F. Vincent. 2016b. Microbial biogeography of permafrost thaw ponds across the changing northern landscape. *Ecography*. 39:609-618.
- Cram, J.A., A.E. Parada, and J.A. Fuhrman. 2016. Dilution reveals how viral lysis and grazing shape microbial communities. *Limnology and Oceanography*. 61:889-905.
- Crevecoeur, S., W.F. Vincent, J. Comte, and C. Lovejoy. 2015. Bacterial community structure across environmental gradients in permafrost thaw ponds: methanotroph-rich ecosystems. *Frontiers in Microbiology*. 6:15.
- Crevecoeur, S., W.F. Vincent, J. Comte, A. Matveev, and C. Lovejoy. 2017. Diversity and potential activity of methanotrophs in high methane-emitting permafrost thaw ponds. *Plos One*. 12:22.
- Crevecoeur, S., W.F. Vincent, and C. Lovejoy. 2016. Environmental selection of planktonic methanogens in permafrost thaw ponds. *Scientific Reports*. 6:10.
- Crummett, L.T., R.J. Puxty, C. Weihe, M.F. Marston, and J.B.H. Martiny. 2016. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology*. 499:219-229.
- Dammeyer, T., S.C. Bagby, M.B. Sullivan, S.W. Chisholm, and N. Frankenberg-Dinkel. 2008. Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Current Biology*. 18:442-448.
- Danovaro, R., A. Dell'Anno, C. Corinaldesi, E. Rastelli, R. Cavicchioli, M. Krupovic, R.T. Noble, T. Nunoura, and D. Prangishvili. 2016. Virus-mediated archaeal hecatomb in the deep seafloor. *Science Advances*. 2:9.

- de Carcer, D.A., C. Pedros-Alio, D.A. Pearce, and A. Alcamí. 2016. Composition and interactions among bacterial, microeukaryotic, and T4-like viral assemblages in lakes from both polar zones. *Frontiers in Microbiology*. 7:11.
- Dekel-Bird, N.P., S. Avrani, G. Sabehi, I. Pekarsky, M.F. Marston, S. Kirzner, and D. Lindell. 2013. Diversity and evolutionary relationships of T7-like podoviruses infecting marine cyanobacteria. *Environmental Microbiology*. 15:1476-1491.
- Deng, L., and P.K. Hayes. 2008. Evidence for cyanophages active against bloom-forming freshwater cyanobacteria. *Freshwater Biology*. 53:1240-1252.
- Deshpande, B.N., S. Crevecoeur, A. Matveev, and W.F. Vincent. 2016. Bacterial production in subarctic peatland lakes enriched by thawing permafrost. *Biogeosciences*. 13:4411-4427.
- Deshpande, B.N., S. MacIntyre, A. Matveev, and W.F. Vincent. 2015. Oxygen dynamics in permafrost thaw lakes: Anaerobic bioreactors in the Canadian subarctic. *Limnology and Oceanography*. 60:1656-1670.
- Dunigan, D.D., R.L. Cerny, A.T. Bauman, J.C. Roach, L.C. Lane, I.V. Agarkova, K. Wulser, G.M. Yanai-Balser, J.R. Gurnon, J.C. Vitek, B.J. Kronschnabel, A. Jeanniard, G. Blanc, C. Upton, G.A. Duncan, O.W. McClung, F.R. Ma, and J.L. Van Etten. 2012. Paramecium bursaria chlorella virus 1 proteome reveals novel architectural and regulatory features of a giant virus. *Journal of Virology*. 86:8821-8834.
- Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792-1797.
- Edgar, R.C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*. 10:996-+.
- Enav, H., Y. Mandel-Gutfreund, and O. Beja. 2014. Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. *Microbiome*. 2:11.
- Fenchel, T. 2008. The microbial loop-25 years later. *Journal of Experimental Marine Biology and Ecology*. 366:99-103.
- Fenchel, T., and B. Finlay. 2008. Oxygen and the spatial structure of microbial communities. *Biological Reviews*. 83:553-569.
- Filee, J., F. Tetart, C.A. Suttle, and H.M. Krisch. 2005. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proceedings of the National Academy of Sciences of the United States of America*. 102:12471-12476.
- Finke, J.F., D.M. Winget, A.M. Chan, and C.A. Suttle. 2017. Variation in the genetic repertoire of viruses infecting *Micromonas pusilla* reflects horizontal gene transfer and links to their environmental distribution. *Viruses-Basel*. 9:18.
- Firsova, A.D., A.E. Kuzmina, I.V. Tomberg, T.G. Potemkina, and Y.V. Likhoshway. 2008. Seasonal dynamics of chrysophyte stomatocyst formation in the plankton of Southern Baikal. *Biology Bulletin*. 35:507-514.
- Fitzgerald, L.A., M.V. Graves, X. Li, J. Hartigan, A.J.P. Pfitzner, E. Hoffart, and J.L. Van Etten. 2007. Sequence and annotation of the 288-kb ATCV-1 virus that infects an endosymbiotic chlorella strain of the heliozoon *Acanthocystis turfacea*. *Virology*. 362:350-361.
- Fortier, D., and M. Allard. 2004. Late Holocene syngenetic ice-wedge polygons development, Bylot Island, Canadian Arctic Archipelago. *Canadian Journal of Earth Sciences*. 41:997-1012.
- Frada, M., I. Probert, M.J. Allen, W.H. Wilson, and C. de Vargas. 2008. The "Cheshire Cat" escape strategy of the coccolithophore *Emiliana huxleyi* in response to viral infection. *Proceedings of the National Academy of Sciences of the United States of America*. 105:15944-15949.
- Fujimoto, A., S. Kondo, R. Nakao, Y. Tomaru, and K. Nagasaki. 2013. Co-occurrence of *Heterocapsa circularisquama* bloom and its lytic viruses in Lake Kamo, Japan, 2010. *Jarq-Japan Agricultural Research Quarterly*. 47:329-338.
- Gao, E.B., Y.H. Huang, and D.G. Ning. 2016. Metabolic genes within cyanophage genomes: Implications for diversity and evolution. *Genes*. 7:15.
- Genuario, D.B., A.P.D. Andreote, M. Vaz, and M.F. Fiore. 2017. Heterocyte-forming cyanobacteria from Brazilian saline-alkaline lakes. *Molecular Phylogenetics and Evolution*. 109:105-112.

- Godoi, R.H.M., K. Aerts, J. Harlay, R. Kaegi, C.U. Ro, L. Chou, and R. Van Grieken. 2009. Organic surface coating on Coccolithophores - *Emiliania huxleyi*: Its determination and implication in the marine carbon cycle. *Microchemical Journal*. 91:266-271.
- Goldsmith, D.B., G. Crosti, B. Dwivedi, L.D. McDaniel, A. Varsani, C.A. Suttle, M.G. Weinbauer, R.A. Sandaa, and M. Breitbart. 2011. Development of phoH as a novel signature gene for assessing marine phage diversity. *Applied and Environmental Microbiology*. 77:7730-7739.
- Grattan, L.M., S. Holobaugh, and J.G. Morris. 2016. Harmful algal blooms and public health. *Harmful Algae*. 57:2-8.
- Gross, M., I. Marianovsky, and G. Glaser. 2006. MazG -- a regulator of programmed cell death in *Escherichia coli*. *Mol Microbiol*. 59:590-601.
- Guixa-Boixereu, N., D. Vaque, J.M. Gasol, J. Sanchez-Camara, and C. Pedros-Alio. 2002. Viral distribution and activity in Antarctic waters. *Deep-Sea Research Part II-Topical Studies in Oceanography*. 49:827-845.
- Gurney, S.D. 2001. Aspects of the genesis, geomorphology and terminology of palsas: perennial cryogenic mounds. *Progress in Physical Geography*. 25:249-260.
- He, Q.F., N. Dolganov, O. Bjorkman, and A.R. Grossman. 2001. The high light-inducible polypeptides in *Synechocystis* PCC6803 - Expression and function in high light. *Journal of Biological Chemistry*. 276:306-314.
- Herrero, A., J. Stavans, and E. Flores. 2016. The multicellular nature of filamentous heterocyst-forming cyanobacteria. *FEMS Microbiology Reviews*. 40:831-854.
- Hessen, D.O. 1996. Competitive trade-off strategies in Arctic *Daphnia* linked to melanism and UV-B stress. *Polar Biology*. 16:573-579.
- <http://bccm.belspo.be>. 2016. ULC species details, Cyanobium sp. FW134.
- Hultman, J., M.P. Waldrop, R. Mackelprang, M.M. David, J. McFarland, S.J. Blazewicz, J. Harden, M.R. Turetsky, A.D. McGuire, M.B. Shah, N.C. VerBerkmoes, L.H. Lee, K. Mavrommatis, and J.K. Jansson. 2015. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature*. 521:208-+.
- Huson, D.H., S. Beier, I. Flade, A. Gorska, M. El-Hadidi, S. Mitra, H.J. Ruscheweyh, and R. Tappu. 2016. MEGAN community edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *Plos Computational Biology*. 12:12.
- Hynes, A.P., M. Shakya, R.G. Mercer, M.P. Grull, L. Bown, F. Davidson, E. Steffen, H. Matchem, M.E. Peach, T. Berger, K. Grebe, O. Zhaxybayeva, and A.S. Lang. 2016. Functional and evolutionary characterization of a gene transfer agent's multilocus "genome". *Molecular Biology and Evolution*. 33:2530-2543.
- Ignacio-Espinoza, J.C., and M.B. Sullivan. 2012. Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. *Environmental Microbiology*. 14:2113-2126.
- Ionescu, D., M. Hindiye, H. Malkawi, and A. Oren. 2010. Biogeography of thermophilic cyanobacteria: insights from the Zerka Ma'in hot springs (Jordan). *FEMS Microbiology Ecology*. 72:103-113.
- Jones, B.M., G. Grosse, C.D. Arp, M.C. Jones, K.M.W. Anthony, and V.E. Romanovsky. 2011. Modern thermokarst lake dynamics in the continuous permafrost zone, northern Seward Peninsula, Alaska. *Journal of Geophysical Research-Biogeosciences*. 116:13.
- Judd, K.E., B.C. Crump, and G.W. Kling. 2006. Variation in dissolved organic matter controls bacterial production and community composition. *Ecology*. 87:2068-2079.
- Jungblut, A.D., I. Hawes, D. Mountfort, B. Hitzfeld, D.R. Dietrich, B.P. Burns, and B.A. Neilan. 2005. Diversity within cyanobacterial mat communities in variable salinity meltwater ponds of McMurdo Ice Shelf, Antarctica. *Environmental Microbiology*. 7:519-529.
- Kang, M., D.D. Dunigan, and J.L. Van Etten. 2005. Chlorovirus: a genus of *Phycodnaviridae* that infects certain chlorella-like green algae. *Molecular Plant Pathology*. 6:213-224.
- Kawakami, H., and N. Kawakami. 1978. Behavior of a virus in a symbiotic system, *Paramecium bursaria* zoochlorella. *Journal of Protozoology*. 25:217-225.
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: An integrated

- and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 28:1647-1649.
- Keeling, P.J. 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Current Opinion in Genetics & Development*. 19:613-619.
- Kiikkila, O., A. Smolander, and L. Ukonmaanaho. 2014. Properties of dissolved organic matter in peatland: Implications for water quality after harvest. *Vadose Zone Journal*. 13:9.
- King, A.M.Q., E. Lefkowitz, M.J. Adams, and E.B. Carstens. 2011. Virus taxonomy classification and nomenclature of viruses : Ninth report of the international committee on taxonomy of viruses.
- Koven, C.D., B. Ringeval, P. Friedlingstein, P. Ciais, P. Cadule, D. Khvorostyanov, G. Krinner, and C. Tarnocai. 2011. Permafrost carbon-climate feedbacks accelerate global warming. *Proceedings of the National Academy of Sciences of the United States of America*. 108:14769-14774.
- Kritzberg, E.S., S. Langenheder, and E.S. Lindstrom. 2006. Influence of dissolved organic matter source on lake bacterioplankton structure and function - implications for seasonal dynamics of community composition. *FEMS Microbiology Ecology*. 56:406-417.
- La Scola, B., S. Audic, C. Robert, L. Jungang, X. de Lamballerie, M. Drancourt, R. Birtles, J.M. Claverie, and D. Raoult. 2003. A giant virus in amoebae. *Science*. 299:2033-2033.
- Landsberg, J.H. 2002. The effects of harmful algal blooms on aquatic organisms. *Reviews in Fisheries Science*. 10:113-390.
- Lang, A.S., O. Zhaxybayeva, and J.T. Beatty. 2012. Gene transfer agents: phage-like elements of genetic exchange. *Nature Reviews Microbiology*. 10:472-482.
- Langenheder, S., M. Berga, O. Ostman, and A.J. Szekely. 2012. Temporal variation of beta-diversity and assembly mechanisms in a bacterial metacommunity. *ISME Journal*. 6:1107-1114.
- Laurion, I., and N. Mladenov. 2013. Dissolved organic matter photolysis in Canadian arctic thaw ponds. *Environmental Research Letters*. 8:12.
- Laurion, I., W. Vincent, S. Macintyre, L. Retamal, C. Dupont, P. Francus, and R. Pienitz. 2010. Variability in greenhouse gas emissions from permafrost thaw ponds. *Limnology and Oceanography*. 55:115-133.
- Lawrence, J.E., A.M. Chan, and C.A. Suttle. 2001. A novel virus (HaNIV) causes lysis of the toxic bloom-forming alga *Heterosigma akashiwo* (Raphidophyceae). *Journal of Phycology*. 37:216-222.
- Legendre, M., J. Bartoli, L. Shmakova, S. Jeudy, K. Labadie, A. Adrait, M. Lescot, O. Poirot, L. Bertaux, C. Bruley, Y. Coute, E. Rivkina, C. Abergel, and J.M. Claverie. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences of the United States of America*. 111:4274-4279.
- Levasseur, A., M. Bekliz, E. Chabriere, P. Pontarotti, B. La Scola, and D. Raoult. 2016. MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. *Nature*. 531:249-+.
- Li, H. 2012. Seqtk, <https://github.com/lh3/seqtk>.
- Lindell, D., J.D. Jaffe, M.L. Coleman, M.E. Futschik, I.M. Axmann, T. Rector, G. Kettler, M.B. Sullivan, R. Steen, W.R. Hess, G.M. Church, and S.W. Chisholm. 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*. 449:83-86.
- Lindell, D., J.D. Jaffe, Z.I. Johnson, G.M. Church, and S.W. Chisholm. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*. 438:86-89.
- Lindell, D., M.B. Sullivan, Z.I. Johnson, A.C. Tolonen, F. Rohwer, and S.W. Chisholm. 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proceedings of the National Academy of Sciences of the United States of America*. 101:11013-11018.
- Lloyd, A.H., K. Yoshikawa, C.L. Fastie, L. Hinzman, and M. Fraver. 2003. Effects of permafrost degradation on woody vegetation at arctic treeline on the Seward Peninsula, Alaska. *Permafrost and Periglacial Processes*. 14:93-101.
- Lopez-Bueno, A., J. Tamames, D. Velazquez, A. Moya, A. Quesada, and A. Alcamí. 2009. High diversity of the viral community from an Antarctic Lake. *Science*. 326:858-861.
- Lovejoy, C., A. Comeau, and M. Thaler. 2016. Curated reference database of SSU rRNA for northern marine and freshwater communities of Archaea, Bacteria and microbial eukaryotes, v. 1.1 (2002-2008) Nordicana D23. DOI:10.5885/45409XD-45479A45199B45476BCC44110.

- Makhalanyane, T.P., A. Valverde, D. Velazquez, E. Gunnigle, M.W. Van Goethem, A. Quesada, and D.A. Cowan. 2015. Ecology and biogeochemistry of cyanobacteria in soils, permafrost, aquatic and cryptic polar habitats. *Biodiversity and Conservation*. 24:819-840.
- Manasyov, R.M., O.S. Pokrovsky, S.N. Kirpotin, and L.S. Shirokova. 2014. Thermokarst lake waters across the permafrost zones of western Siberia. *Cryosphere*. 8:1177-1193.
- Mann, N.H., M.R.J. Clokie, A. Millard, A. Cook, W.H. Wilson, P.J. Wheatley, A. Letarov, and H.M. Krisch. 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. *Journal of Bacteriology*. 187:3188-3200.
- Mann, N.H., A. Cook, A. Millard, S. Bailey, and M. Clokie. 2003. Marine ecosystems: Bacterial photosynthesis genes in a virus. *Nature*. 424:741-741.
- Margesin, R. 2009. Permafrost soils. Springer, Berlin.
- Maslov, S., and K. Sneppen. 2017. Population cycles and species diversity in dynamic Kill-the-Winner model of microbial ecosystems. *Scientific Reports*. 7:8.
- Matheus Carnevali, P.B., M. Rohrsen, M.R. Williams, A.B. Michaud, H. Adams, D. Berisford, G.D. Love, J.C. Priscu, O. Rassuchine, K.P. Hand, and A.E. Murray. 2015. Methane sources in arctic thermokarst lake sediments on the North Slope of Alaska. *Geobiology*. 13:181-197.
- Matteson, A.R., S.N. Loar, R.A. Bourbonniere, and S.W. Wilhelm. 2011. Molecular enumeration of an ecologically important cyanophage in a Laurentian Great Lake. *Applied and Environmental Microbiology*. 77:6772-6779.
- Matteson, A.R., J.M. Rowe, A.J. Ponsoero, T.M. Pimentel, P.W. Boyd, and S.W. Wilhelm. 2013. High abundances of cyanomyoviruses in marine ecosystems demonstrate ecological relevance. *FEMS Microbiology Ecology*. 84:223-234.
- Matveev, A., I. Laurion, B.N. Deshpande, N. Bhiry, and W.F. Vincent. 2016. High methane emissions from thermokarst lakes in subarctic peatlands. *Limnology and Oceanography*. 61:S150-S164.
- McDaniel, L.D., E. Young, J. Delaney, F. Ruhnau, K.B. Ritchie, and J.H. Paul. 2010. High frequency of horizontal gene transfer in the oceans. *Science*. 330:50-50.
- Meints, R.H., J.L. Vanetten, D. Kuczmarski, K. Lee, and B. Ang. 1981. Viral infection of the symbiotic chlorella-like algae present in *Hydra viridis*. *Virology*. 113:698-703.
- Mercanti, D.J., G.M. Rousseau, M.L. Capra, A. Quiberoni, D.M. Tremblay, S.J. Labrie, and S. Moineau. 2016. Genomic diversity of phages infecting probiotic strains of *Lactobacillus paracasei*. *Applied and Environmental Microbiology*. 82:95-105.
- Millard, A.D., K. Zwirgmaier, M.J. Downey, N.H. Mann, and D.J. Scanlan. 2009. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environmental Microbiology*. 11:2370-2387.
- Mohit, V., A. Culley, C. Lovejoy, F. Bouchard, and W.F. Vincent. 2017. Hidden biofilms in a far northern lake and implications for the changing Arctic. *npj Biofilms and Microbiomes*. 3:17.
- Mondav, R., B.J. Woodcroft, E.H. Kim, C.K. McCalley, S.B. Hodgkins, P.M. Crill, J. Chanton, G.B. Hurst, N.C. VerBerkmoes, S.R. Saleska, P. Hugenholtz, V.I. Rich, and G.W. Tyson. 2014. Discovery of a novel methanogen prevalent in thawing permafrost. *Nature Communications*. 5:7.
- Mueller, J.A., A.I. Culley, and G.F. Steward. 2014. Variables influencing extraction of nucleic acids from microbial plankton (Viruses, Bacteria, and Protists) collected on nanoporous aluminum oxide filters. *Applied and Environmental Microbiology*. 80:3930-3942.
- Needham, D.M., R. Sachdeva, and J.A. Fuhrman. 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME Journal*. 11:1614-1629.
- Negandhi, K., I. Laurion, and C. Lovejoy. 2016. Temperature effects on net greenhouse gas production and bacterial communities in arctic thaw ponds. *FEMS Microbiology Ecology*. 92:12.
- Nelson, F.E. 2003. (Un)frozen in time. *Science*. 299:1673-1675.

- Nguyen, L.T., H.A. Schmidt, A. von Haeseler, and B.Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*. 32:268-274.
- Noges, T., and K. Kangro. 2005. Primary production of phytoplankton in a strongly stratified temperate lake. *Hydrobiologia*. 547:105-122.
- Nuttall, M. 2005. Microbes. In *Encyclopedia of the Arctic*. Routledge, New York. 1287.
- Ochman, H., J.G. Lawrence, and E.A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 405:299-304.
- Oksanen, J., F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, and H. Wagner. 2016. Vegan: Community ecology package. R package version 2.3-5. <https://CRAN.R-project.org/package=vegan>.
- Olefeldt, D., S. Goswami, G. Grosse, D. Hayes, G. Hugelius, P. Kuhry, A.D. McGuire, V.E. Romanovsky, A.B.K. Sannel, E.A.G. Schuur, and M.R. Turetsky. 2016. Circumpolar distribution and carbon storage of thermokarst landscapes. *Nature Communications*. 7:11.
- Oren, A. 2015. Cyanobacteria in hypersaline environments: biodiversity and physiological properties. *Biodiversity and Conservation*. 24:781-798.
- Paasche, E. 2001. A review of the coccolithophorid *Emiliana huxleyi* (*Prymnesiophyceae*), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia*. 40:503-529.
- Paerl, H.W., and T.G. Otten. 2013. Harmful cyanobacterial blooms: Causes, consequences, and controls. *Microbial Ecology*. 65:995-1010.
- Pannard, A., D. Planas, and B.E. Beisner. 2015. Macrozooplankton and the persistence of the deep chlorophyll maximum in a stratified lake. *Freshwater Biology*. 60:1717-1733.
- Partensky, F., W.R. Hess, and D. Vaultot. 1999. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiology and Molecular Biology Reviews*. 63:106-+.
- Peeters, K., E. Verleyen, D.A. Hodgson, P. Convey, D. Ertz, W. Vyverman, and A. Willems. 2012. Heterotrophic bacterial diversity in aquatic microbial mat communities from Antarctica. *Polar Biology*. 35:543-554.
- Pei, J., B.H. Kim, and N.V. Grishin. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res*. 36:2295-2300.
- Percent, S.F., M.E. Frischer, P.A. Vescio, E.B. Duffy, V. Milano, M. McLellan, B.M. Stevens, C.W. Boylen, and S.A. Nierzwicki-Bauer. 2008. Bacterial community structure of acid-impacted lakes: What controls diversity? *Applied and Environmental Microbiology*. 74:1856-1868.
- Philippe, N., M. Legendre, G. Doutre, Y. Coute, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J.M. Claverie, and C. Abergel. 2013. Pandoraviruses: Amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science*. 341:281-286.
- Pienitz, R., P.T. Doran, and S.F. Lamoureux. 2008. Origin and geomorphology of lakes in the polar regions. In *Polar Lakes and Rivers: Limnology of Arctic and Antarctic Aquatic Ecosystems*. W.F. Vincent and J. Laybourn-Parry, editors. Oxford Scholarship Online. DOI: 10.1093/acprof:oso/9780199213887.9780199213001.9780199210001.
- Pierce, S.K., T.K. Maugel, M.E. Rumpho, J.J. Hanten, and W.L. Mondy. 1999. Annual viral expression in a sea slug population: Life cycle control and symbiotic chloroplast maintenance. *Biological Bulletin*. 197:1-6.
- Planavsky, N.J., D. Asael, A. Hofmann, C.T. Reinhard, S.V. Lalonde, A. Knudsen, X.L. Wang, F.O. Ossa, E. Pecoits, A.J.B. Smith, N.J. Beukes, A. Bekker, T.M. Johnson, K.O. Konhauser, T.W. Lyons, and O.J. Rouxel. 2014. Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. *Nature Geoscience*. 7:283-286.
- Polz, M.F., E.J. Alm, and W.P. Hanage. 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics*. 29:170-175.
- Poorvin, L., J.M. Rinta-Kanto, D.A. Hutchins, and S.W. Wilhelm. 2004. Viral release of iron and its bioavailability to marine plankton. *Limnology and Oceanography*. 49:1734-1741.

- Pruesse, E., C. Quast, K. Knittel, B.M. Fuchs, W.G. Ludwig, J. Peplies, and F.O. Glockner. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*. 35:7188-7196.
- Przytulska, A., M. Bartosiewicz, and W.F. Vincent. 2017. Increased risk of cyanobacterial blooms in northern high-latitude lakes through climate warming and phosphorus enrichment. *Freshwater Biology*.
- Przytulska, A., J. Comte, S. Crevecoeur, C. Lovejoy, I. Laurion, and W.F. Vincent. 2016. Phototrophic pigment diversity and picophytoplankton in permafrost thaw lakes. *Biogeosciences*. 13:13-26.
- Quebatte, M., M. Christen, A. Harms, J. Korner, B. Christen, and C. Dehio. 2017. Gene transfer agent promotes evolvability within the fittest subpopulation of a bacterial pathogen. *Cell Systems*. 4:611-+.
- R Core Team. 2016 R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Rachel, R., M. Bettstetter, B.P. Hedlund, M. Haring, A. Kessler, K.O. Stetter, and D. Prangishvili. 2002. Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Archives of Virology*. 147:2419-2429.
- Rautio, M., F. Dufresne, I. Laurion, S. Bonilla, W.F. Vincent, and K.S. Christoffersen. 2011. Shallow freshwater ecosystems of the circumpolar Arctic. *Ecoscience*. 18:204-222.
- Reisser, W., D.E. Burbank, R.H. Meints, B. Becker, and J.L. Vanetten. 1991. Viruses distinguish symbiotic *Chlorella spp* of *Paramecium Bursaria*. *Endocytobiosis and Cell Research*. 7:245-251.
- Rohwer, F., D. Prangishvili, and D. Lindell. 2009. Roles of viruses in the environment. *Environmental Microbiology*. 11:2771-2774.
- Rohwer, F., and R.V. Thurber. 2009. Viruses manipulate the marine environment. *Nature*. 459:207-212.
- Roiha, T., I. Laurion, and M. Rautio. 2015. Carbon dynamics in highly heterotrophic subarctic thaw ponds. *Biogeosciences*. 12:7223-7237.
- Roncalli, V., J.T. Turner, D. Kulis, D.M. Anderson, and P.H. Lenz. 2016. The effect of the toxic dinoflagellate *Alexandrium fundyense* on the fitness of the calanoid copepod *Calanus finmarchicus*. *Harmful Algae*. 51:56-66.
- Rossi, P.-G., I. Laurion, and C. Lovejoy. 2013. Distribution and identity of Bacteria in subarctic permafrost thaw ponds. *Aquatic Microbial Ecology*. 69:231-245.
- Roux, S., J.R. Brum, B.E. Dutilh, S. Sunagawa, M.B. Duhaime, A. Loy, B.T. Poulos, N. Solonenko, E. Lara, J. Poulain, S. Pesant, S. Kandels-Lewis, C. Dimier, M. Picheral, S. Searson, C. Cruaud, A. Alberti, C.M. Duarte, J.M. Gasol, D. Vaque, P. Bork, S.G. Acinas, P. Wincker, M.B. Sullivan, and C. Tara Oceans. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*. 537:689-+.
- Sabehi, G., L. Shaulov, D.H. Silver, I. Yanai, A. Harel, and D. Lindell. 2012. A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. *Proceedings of the National Academy of Sciences of the United States of America*. 109:2037-2042.
- Sanmartin, M., L. Jaroszewski, N.V. Raikhel, and E. Rojo. 2005. Caspases. Regulating death since the origin of life. *Plant Physiology*. 137:841-847.
- Sarma, T.A. 2012. Taxonomy, Phylogeny and Evolution. *In Handbook of Cyanobacteria*. CRC Press. 1-75.
- Sawstrom, C., J. Ask, and J. Karlsson. 2009. Viruses in subarctic lakes and their impact on benthic and pelagic bacteria. *FEMS Microbiology Ecology*. 70:471-482.
- Sawstrom, C., W. Graneli, J. Laybourn-Parry, and A.M. Anesio. 2007. High viral infection rates in Antarctic and Arctic bacterioplankton. *Environmental Microbiology*. 9:250-255.
- Sharon, I., A. Alperovitch, F. Rohwer, M. Haynes, F. Glaser, N. Atamna-Ismaeel, R.Y. Pinter, F. Partensky, E.V. Koonin, Y.I. Wolf, N. Nelson, and O. Beja. 2009. Photosystem I gene cassettes are present in marine virus genomes. *Nature*. 461:258-262.
- Sharon, I., S. Tzahor, S. Williamson, M. Shmoish, D. Man-Aharonovich, D.B. Rusch, S. Yooseph, G. Zeidner, S.S. Golden, S.R. Mackey, N. Adir, U. Weingart, D. Horn, J.C. Venter, Y. Mandel-Gutfreund, and O. Beja. 2007. Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME Journal*. 1:492-501.

- Shokralla, S., J.L. Spall, J.F. Gibson, and M. Hajibabaei. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*. 21:1794-1805.
- Short, S.M. 2012. The ecology of viruses that infect eukaryotic algae. *Environmental Microbiology*. 14:2253-2271.
- Short, S.M., O. Rusanova, and M.A. Staniewski. 2011. Novel phycodnavirus genes amplified from Canadian freshwater environments. *Aquatic Microbial Ecology*. 63:61-67.
- Skerratt, J.H., J.P. Bowman, G. Hallegraeff, S. James, and P.D. Nichols. 2002. Algicidal bacteria associated with blooms of a toxic dinoflagellate in a temperate Australian estuary. *Marine Ecology Progress Series*. 244:1-15.
- Smith, M.W., and D.W. Riseborough. 2002. Climate and the limits of permafrost: A zonal analysis. *Permafrost and Periglacial Processes*. 13:1-15.
- Soo, R.M., J. Hemp, D.H. Parks, W.W. Fischer, and P. Hugenholtz. 2017. On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science*. 355:1436-1439.
- Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312-1313.
- Stern, A., and R. Sorek. 2011. The phage-host arms-race: Shaping the evolution of microbes. *Bioessays*. 33:43-51.
- Sullivan, M.B., M.L. Coleman, V. Quinlivan, J.E. Rosenkrantz, A.S. DeFrancesco, G. Tan, R. Fu, J.A. Lee, J.B. Waterbury, J.P. Bielawski, and S.W. Chisholm. 2008. Portal protein diversity and phage ecology. *Environmental Microbiology*. 10:2810-2823.
- Sullivan, M.B., M.L. Coleman, P. Weigele, F. Rohwer, and S.W. Chisholm. 2005. Three Prochlorococcus cyanophage genomes: Signature features and ecological interpretations. *Plos Biology*. 3:790-806.
- Sullivan, M.B., K.H. Huang, J.C. Ignacio-Espinoza, A.M. Berlin, L. Kelly, P.R. Weigele, A.S. DeFrancesco, S.E. Kern, L.R. Thompson, S. Young, C. Yandava, R. Fu, B. Krastins, M. Chase, D. Sarracino, M.S. Osburne, M.R. Henn, and S.W. Chisholm. 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental Microbiology*. 12:3035-3056.
- Sullivan, M.B., D. Lindell, J.A. Lee, L.R. Thompson, J.P. Bielawski, and S.W. Chisholm. 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *Plos Biology*. 4:1344-1357.
- Sullivan, M.B., J.B. Waterbury, and S.W. Chisholm. 2003. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature*. 424:1047-1051.
- Suttle, C.A. 2005. Viruses in the sea. *Nature*. 437:356-361.
- Suttle, C.A. 2007. Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology*. 5:801-812.
- Suttle, C.A., and A.M. Chan. 1994. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Applied and Environmental Microbiology*. 60:3167-3174.
- Suttle, C.A., and C. Feng. 1992. Mechanisms and rates of decay of marine viruses in seawater. *Applied and Environmental Microbiology*. 58:3721-3729.
- Tarnocai, C., J.G. Canadell, E.A.G. Schuur, P. Kuhry, G. Mazhitova, and S. Zimov. 2009. Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochemical Cycles*. 23:11.
- Thingstad, T.F., and R. Lignell. 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology*. 13:19-27.
- Thompson, L.R., Q. Zeng, L. Kelly, K.H. Huang, A.U. Singer, J. Stubbe, and S.W. Chisholm. 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences of the United States of America*. 108:E757-E764.
- Traxler, M.F., S.M. Summers, H.T. Nguyen, V.M. Zacharia, G.A. Hightower, J.T. Smith, and T. Conway. 2008. The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Molecular Microbiology*. 68:1128-1148.
- Tyrrell, T., and A.H. Taylor. 1996. A modelling study of *Emiliania huxleyi* in the NE Atlantic. *Journal of Marine Systems*. 9:83-112.

- Van Etten, J.L., I. Agarkova, D.D. Dunigan, M. Tonetti, C. De Castro, and G.A. Duncan. 2017. Chloroviruses Have a Sweet Tooth. *Viruses-Basel*. 9:23.
- Van Etten, J.L., R.H. Meints, D. Kuczmarski, D.E. Burbank, and K. Lee. 1982. Viruses of symbiotic *Chlorella*-like algae isolated from *Paramecium Bursaria* and *Hydra Viridis*. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*. 79:3867-3871.
- van Huissteden, J., C. Berrittella, F.J.W. Parmentier, Y. Mi, T.C. Maximov, and A.J. Dolman. 2011. Methane emissions from permafrost thaw lakes limited by lake drainage. *Nature Climate Change*. 1:119-123.
- Van Valen, L. 1973. A new evolutionary law. *Evolutionary Theory*. 1:1-30.
- Walter, K.M., M.E. Edwards, G. Grosse, S.A. Zimov, and F.S. Chapin. 2007. Thermokarst lakes as a source of atmospheric CH₄ during the last deglaciation. *Science*. 318:633-636.
- Wang, M.N., X.Y. Ge, Y.Q. Wu, X.L. Yang, B. Tan, Y.J. Zhang, and Z.L. Shi. 2015. Genetic diversity and temporal dynamics of phytoplankton viruses in East Lake, China. *Virologica Sinica*. 30:290-300.
- Wang, X.Z., J.J. Liu, Z.H. Yu, J. Jin, X.B. Liu, and G.H. Wang. 2016. Novel groups and unique distribution of phage phoH genes in paddy waters in northeast China. *Scientific Reports*. 6:8.
- Watanabe, S., I. Laurion, K. Chokmani, R. Pienitz, and W.F. Vincent. 2011. Optical diversity of thaw ponds in discontinuous permafrost: A model system for water color analysis. *Journal of Geophysical Research-Biogeosciences*. 116:17.
- Weitz, J.S., and S.W. Wilhelm. 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 biology reports*. 4:17.
- Wells, L.E., and J.W. Deming. 2006. Modelled and measured dynamics of viruses in Arctic winter sea-ice brines. *Environmental Microbiology*. 8:1115-1121.
- Wik, M., R.K. Varner, K.W. Anthony, S. MacIntyre, and D. Bastviken. 2016. Climate-sensitive northern lakes and ponds are critical components of methane release. *Nature Geoscience*. 9:99-+.
- Wilhelm, S.W., and A.R. Matteson. 2008. Freshwater and marine viroplankton: a brief overview of commonalities and differences. *Freshwater Biology*. 53:1076-1089.
- Wilhelm, S.W., and C.A. Suttle. 1999. Viruses and Nutrient Cycles in the Sea - Viruses play critical roles in the structure and function of aquatic food webs. *Bioscience*. 49:781-788.
- Williamson, S.J., D.B. Rusch, S. Yooseph, A.L. Halpern, K.B. Heidelberg, J.I. Glass, C. Andrews-Pfannkoch, D. Fadrosh, C.S. Miller, G. Sutton, M. Frazier, and J.C. Venter. 2008. The Sorcerer II Global Ocean Sampling Expedition: Metagenomic characterization of viruses within aquatic microbial samples. *Plos One*. 3:14.
- Wilson, W.H., J.L. Van Etten, D.C. Schroeder, K. Nagasaki, C. Brussaard, G. Bratbak, and C. Suttle. 2011. Family - Phycodnaviridae. In *Virus Taxonomy : Ninth Report of the International Committee on Taxonomy of Viruses*. A.M.Q. King, E. Lefkowitz, M.J. Adams, and E.B. Carstens editors. Elsevier Inc. 249-262.
- Wu, D.D., and Y.P. Zhang. 2011. Eukaryotic origin of a metabolic pathway in virus by horizontal gene transfer. *Genomics*. 98:367-369.
- Wu, Q.B., T.J. Zhang, and Y.Z. Liu. 2010. Permafrost temperatures and thickness on the Qinghai-Tibet Plateau. *Global and Planetary Change*. 72:32-38.
- Yan, X.D., N.H. Olson, J.L. Van Etten, M. Bergoin, M.G. Rossmann, and T.S. Baker. 2000. Structure and assembly of large lipid-containing dsDNA viruses. *Nature Structural Biology*. 7:101-103.
- Yoshida, T., Y. Takashima, Y. Tomaru, Y. Shirai, Y. Takao, S. Hiroishi, and K. Nagasaki. 2006. Isolation and characterization of a cyanophage infecting the toxic cyanobacterium *Microcystis aeruginosa*. *Applied and Environmental Microbiology*. 72:1239-1247.
- Zeigler Allen, L., J.P. McCrow, K. Ininbergs, C.L. Dupont, J.H. Badger, J.M. Hoffman, M. Ekman, A.E. Allen, B. Bergman, and J.C. Venter. 2017. The Baltic sea virome: Diversity and transcriptional activity of DNA and RNA viruses. *mSystems*. 2.
- Zhang, T., J.A. Heginbottom, R.G. Barry, and J. Brown. 2000. Further statistics on the distribution of permafrost and ground ice in the Northern Hemisphere. *Polar Geography*. 24:126-131.

- Zhong, X., A.S.P. Ram, J. Colombet, and S. Jacquet. 2014. Variations in abundance, genome size, morphology, and functional role of the viroplankton in Lakes Annecy and Bourget over a 1-year period. *Microbial Ecology*. 67:66-82.
- Zhong, Y., F. Chen, S.W. Wilhelm, L. Poorvin, and R.E. Hodson. 2002. Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. *Applied and Environmental Microbiology*. 68:1576-1584.

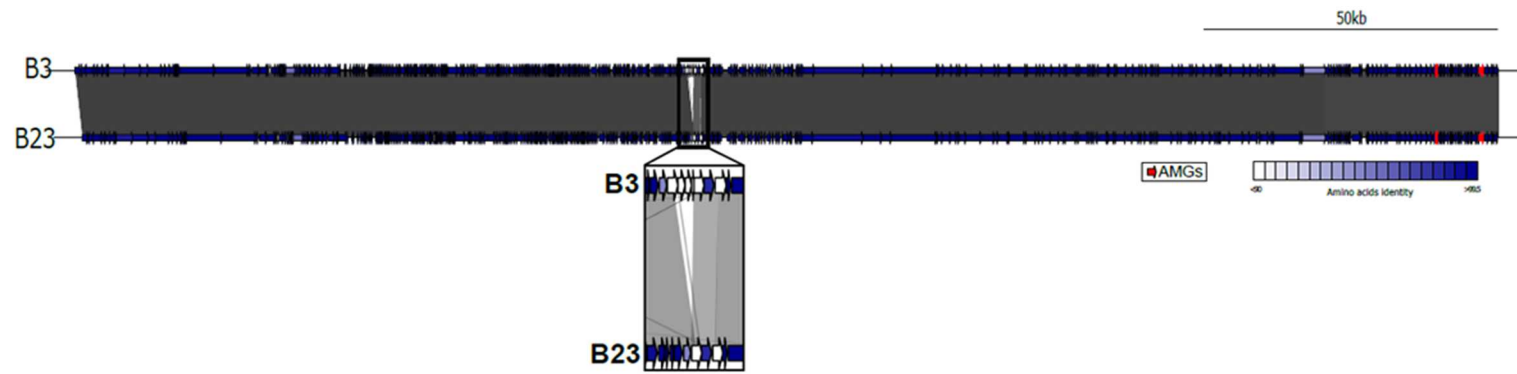


Figure A.1 : Alignement des génomes des cyanophages B3 et B23. Les différentes teintes de bleus réfèrent au niveau d'identité entre les séquences d'acides aminés, et le gradient de gris correspond au niveau d'identité des nucléotides. Les flèches rouges réfèrent aux AMGs.

***À noter que les CDSs annotés comme protéine hypothétique ont été retirés des tableaux.

Tableau A.1 : Annotation génomique du cyanophage B3

Type	Début	Fin	Brin	Fonction
CDS1	1	951	+	Single-stranded DNA-binding protein, phage-associated #T4-like phage gp32 #T4 GC0005
CDS5	4109	3669	-	Phage head completion protein
CDS7	4918	5640	+	Phage baseplate assembly chaperone (T4-like gp26)
CDS8	5652	5816	+	gp51
CDS20	18006	24023	+	Phage tail fiber protein
CDS23	30483	31715	+	helicase
CDS45	41539	42162	+	Translation initiation factor 2
CDS46	42316	43341	+	Nicotinamide-nucleotide adenylyltransferase, NadM family (EC 2.7.7.1) / ADP-ribose pyrophosphatase (EC 3.6.1.13)
CDS49	44103	45491	+	Nicotinamide phosphoribosyltransferase (EC 2.4.2.12)
tRNA	45857	45929	+	tRNA-Lys-TTT
tRNA	46062	46132	+	tRNA-Trp-CCA
tRNA	46277	46350	+	tRNA-Asp-GTC
CDS52	46505	46708	+	Hypothetical-Protein belonging to T4-LIKE GC: 783
tRNA	46768	46840	+	tRNA-Pseudo-GTT
tRNA	46869	46941	+	tRNA-Ile-GAT
tRNA	46964	47041	+	tRNA-Met-CAT
tRNA	47110	47182	+	tRNA-Ala-TGC
tRNA	47205	47276	+	tRNA-Pseudo-TTC
tRNA	47515	47601	+	tRNA-Ser-GCT
tRNA	47694	47785	+	tRNA-Pseudo-TGA
tRNA	49083	49154	+	tRNA-Met-CAT
tRNA	49156	49230	+	tRNA-His-GTG
tRNA	50074	50145	+	tRNA-Thr-TGT
tRNA	50283	50354	+	tRNA-Cys-GCA
CDS79	56071	56856	+	Ser/Thr protein phosphatase family protein
CDS106	64212	65837	+	Flagellar hook-length control protein FliK
CDS110	67118	67774	+	FIG01149310: hypothetical protein
CDS128	74856	75728	+	DNA ligase, phage-associated
CDS156	83298	83555	+	Hypothetical-Protein belonging to T4-LIKE GC: 780
CDS168	86809	87165	+	Pentapeptide repeat family protein
CDS173	88939	89859	+	Phage pentapeptide repeat family protein (ACLAME 560)
tRNA	92431	92514	+	tRNA-Pseudo-CAA
CDS196	99227	99958	+	CCA tRNA nucleotidyltransferase (EC 2.7.7.72)
CDS215	107633	108229	+	Pentapeptide repeat family protein
tRNA	112028	112111	+	tRNA-Leu-TAA
CDS241	123108	123644	+	HNH homing endonuclease
CDS246	125110	125502	+	Phage baseplate wedge subunit (T4-like gp25)

CDS247	125502	127400	+	Phage baseplate wedge subunit (T4-like gp6)
CDS248	127397	135934	+	Phage baseplate wedge initiator (T4-like gp7)
CDS249	135931	139176	+	Phage baseplate wedge initiator (T4-like gp7)
CDS250	139250	140830	+	Phage baseplate wedge subunit (T4-like gp8)
CDS253	148830	149648	+	T4-like phage head completion, neck hetero-dimeric protein #T4-like phage gp13 #T4 GC0116
CDS254	149652	151058	+	T4-like phage head completion, neck hetero-dimeric protein #T4-like phage gp14 #T4 GC0117
CDS255	151058	152059	+	Phage tail completion protein (T4-like gp15)
CDS256	152068	152520	+	Phage terminase, small subunit
CDS258	153663	156305	+	Phage terminase, large subunit
CDS259	156364	158550	+	Phage tail sheath monomer
CDS260	158587	159210	+	Phage tail tube protein
CDS261	159244	160917	+	Phage portal vertex of the head
CDS263	161163	163793	+	Phage tail fiber
RR	163053	163192	+	repeat region
CDS267	166326	166976	+	Phage prohead assembly (scaffolding) protein
CDS268	167097	168269	+	Phage prohead assembly (scaffolding) protein
CDS269	168323	169720	+	Phage major capsid protein of Caudovirales
CDS272	172077	172649	+	Phage tail completion protein
CDS274	173152	174618	+	Phge RNA-DNA + DNA-DNA helicase (EC 3.6.4.12) # Phage T4-like UvsW #T4 GC0138
CDS276	175401	175090	-	Phage protein
CDS279	176473	177987	+	phage T4-like gp55, RNA polymerase sigma factor for late transcription # T4 GC 140
CDS281	178135	179166	+	Phage recombination-related endonuclease Gp47
CDS283	179425	181140	+	Phage recombination-related endonuclease Gp46
CDS287	183845	184519	+	Sliding clamp DNA polymerase accessory protein, phage associated # T4-like phage gp45 #T4 GC0153
CDS288	184516	185451	+	Replication factor C small subunit / Phage DNA polymerase clamp loader subunit
CDS290	185847	186296	+	Phage endoribonulcease translational repressor of early genes, RegA #Phage T4-like RegA #T4 GC0168
CDS292	186626	189118	+	DNA polymerase (EC 2.7.7.7), phage-associated
CDS293	189147	190202	+	UvsX RecA-like protein
CDS294	190202	190762	+	2-Oxoglutarate/Fe(II)-dependent oxygenase (EC 1.14.11) #T4 GC0104
CDS295	190762	192129	+	DNA primase/helicase, phage-associated
CDS298	193247	194479	+	Cyanophage-encoded cytidyltransferase (EC 2.7.7.-) #T4 GC0190
CDS299	194484	195716	+	Phage protein (ACLAME 1119)
CDS301	196639	197025	+	Phage protein
CDS304	198110	199072	+	DNA primase subunit

CDS305	199069	201339	+	Ribonucleotide reductase of class Ia (aerobic), alpha subunit (EC 1.17.4.1)
CDS306	201350	202489	+	Ribonucleotide reductase of class Ia (aerobic), beta subunit (EC 1.17.4.1)
CDS309	204411	205673	+	2',3'-cyclic-nucleotide 2'-phosphodiesterase (EC 3.1.4.16) / 3'-nucleotidase (EC 3.1.3.6)
CDS312	206702	211195	+	Phage tail fiber protein
CDS314	211560	215252	+	Secretory antigen precursor SsaA
RR	213459	213598	+	repeat region
tRNA	218586	218660	+	tRNA-Pro-TGG
CDS326	219837	221312	+	Phage protein
tRNA	221705	221814	+	tRNA-Pseudo-GCA
tRNA	221885	221957	+	tRNA-Arg-TCT
tRNA	222000	222072	+	tRNA-Gly-TCC
tRNA	222110	222183	+	tRNA-Ile-TAT
CDS330	222744	223535	+	glycosyltransferase family 11
CDS335	226385	227758	+	Glucose-1-phosphate thymidyltransferase (EC 2.7.7.24)
CDS342	232759	233556	+	glycosyltransferase family 25
CDS343	233708	234187	+	Septum-associated rare lipoprotein A
CDS344	234204	234398	+	Phage-encoded high light-inducible protein, HliPP #T4 GC0267
CDS345	234418	234615	+	phage T4-like hliPP, high light inducible protein # T4 GC 267
CDS352	236816	237046	+	Hypothetical-Protein belonging to T4-LIKE GC: 61
CDS355	237918	238172	+	Glutaredoxin, NrdC #T4 GC0311
CDS358	238905	239480	+	T4-like phage protein, cyanophage only #T4-like GC0313
CDS360	239705	240031	+	S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50), prokaryotic class 1B
CDS363	240497	241249	+	Thymidylate synthase ThyX (EC 2.1.1.148)
CDS366	241711	242472	+	Phosphate starvation-inducible protein PhoH, predicted ATPase #T4 GC0322
CDS368	243327	243998	+	Phage T4-like exonuclease A, DexA #T4 GC0325
CDS369	243985	244257	+	T4-like late promoter transcriptional accessory protein
CDS370	244268	244870	+	Phage DNA helicase loader # T4-like phage gp59, loader of gp41 DNA helicase #T4 GC0003

Tableau A.2 : Annotation génomique du cyanophage B23

Type	Début	Fin	Brin	Fonction
CDS1	1	951	+	Single-stranded DNA-binding protein, phage-associated #T4-like phage gp32 #T4 GC0005
CDS5	4109	3669	-	Phage head completion protein
CDS7	4918	5640	+	Phage baseplate assembly chaperone (T4-like gp26)
CDS8	5652	5816	+	gp51
CDS20	18012	24029	+	Phage tail fiber protein
CDS23	30489	31721	+	helicase

CDS45	41542	42165	+	Translation initiation factor 2
CDS46	42319	43344	+	Nicotinamide-nucleotide adenylyltransferase, NadM family (EC 2.7.7.1) / ADP-ribose pyrophosphatase (EC 3.6.1.13)
CDS47	43357	43887	+	Phage protein (ACLAME 1218)
CDS49	44106	45494	+	Nicotinamide phosphoribosyltransferase (EC 2.4.2.12)
tRNA	45860	45932	+	tRNA-Lys-TTT
tRNA	46065	46135	+	tRNA-Pseudo-CCA
tRNA	46280	46353	+	tRNA-Asp-GTC
CDS52	46508	46711	+	Hypothetical-Protein belonging to T4-LIKE GC: 783
tRNA	46771	46843	+	tRNA-Pseudo-GTT
tRNA	46872	46944	+	tRNA-Ile-GAT
tRNA	46967	47044	+	tRNA-Met-CAT
tRNA	47113	47185	+	tRNA-Ala-TGC
tRNA	47208	47279	+	tRNA-Glu-TTC
tRNA	47518	47604	+	tRNA-Pseudo-GCT
tRNA	47697	47788	+	tRNA-Pseudo-TGA
tRNA	49086	49157	+	tRNA-Met-CAT
tRNA	50077	50148	+	tRNA-Thr-TGT
tRNA	50286	50357	+	tRNA-Cys-GCA
CDS79	56074	56859	+	Ser/Thr protein phosphatase family protein
CDS106	64215	65840	+	Flagellar hook-length control protein FliK
CDS128	74859	75731	+	DNA ligase, phage-associated
CDS156	83301	83558	+	Hypothetical-Protein belonging to T4-LIKE GC: 780
CDS168	86812	87168	+	Pentapeptide repeat family protein
CDS173	88942	89862	+	Phage pentapeptide repeat family protein (ACLAME 560)
CDS196	99230	99961	+	CCA tRNA nucleotidyltransferase (EC 2.7.7.72)
CDS211	106204	106740	+	Pentapeptide repeat family protein
tRNA	110539	110622	+	tRNA-Leu-TAA
CDS236	121619	122155	+	HNH homing endonuclease
CDS241	123621	124013	+	Phage baseplate wedge subunit (T4-like gp25)
CDS242	124013	125911	+	Phage baseplate wedge subunit (T4-like gp6)
CDS243	125908	134445	+	Phage baseplate wedge initiator (T4-like gp7)
CDS244	134442	137687	+	Phage baseplate wedge initiator (T4-like gp7)
CDS245	137761	139341	+	Phage baseplate wedge subunit (T4-like gp8)
CDS248	147341	148159	+	T4-like phage head completion, neck hetero-dimeric protein #T4-like phage gp13 #T4 GC0116
CDS249	148163	149569	+	T4-like phage head completion, neck hetero-dimeric protein #T4-like phage gp14 #T4 GC0117
CDS250	149569	150570	+	Phage tail completion protein (T4-like gp15)
CDS251	150579	151031	+	Phage terminase, small subunit
CDS253	152174	154816	+	Phage terminase, large subunit
CDS254	154875	157061	+	Phage tail sheath monomer
CDS255	157098	157721	+	Phage tail tube protein

CDS256	157755	159428	+	Phage portal vertex of the head
CDS258	159674	162304	+	Phage tail fiber
RR	161471	161703	+	repeat region
CDS262	164837	165487	+	Phage prohead assembly (scaffolding) protein
CDS263	165608	166780	+	Phage prohead assembly (scaffolding) protein
CDS264	166834	168231	+	Phage major capsid protein of Caudovirales
CDS267	170588	171160	+	Phage tail completion protein
CDS269	171663	173129	+	Phage RNA-DNA + DNA-DNA helicase (EC 3.6.4.12) # Phage T4-like UvsW #T4 GC0138
CDS271	173912	173601	-	Phage protein
CDS274	174984	176498	+	phage T4-like gp55, RNA polymerase sigma factor for late transcription # T4 GC 140
CDS276	176646	177677	+	Phage recombination-related endonuclease Gp47
CDS278	177936	179651	+	Phage recombination-related endonuclease Gp46
CDS282	182356	183030	+	Sliding clamp DNA polymerase accessory protein, phage associated # T4-like phage gp45 #T4 GC0153
CDS283	183027	183962	+	Replication factor C small subunit / Phage DNA polymerase clamp loader subunit
CDS285	184358	184807	+	Phage endoribonuclease translational repressor of early genes, RegA #Phage T4-like RegA #T4 GC0168
CDS287	185137	187629	+	DNA polymerase (EC 2.7.7.7), phage-associated
CDS288	187658	188713	+	UvsX RecA-like protein
CDS289	188713	189273	+	2-Oxoglutarate/Fe(II)-dependent oxygenase (EC 1.14.11) #T4 GC0104
CDS290	189273	190640	+	DNA primase/helicase, phage-associated
CDS293	191758	192990	+	Cyanophage-encoded cytidyltransferase (EC 2.7.7.-) #T4 GC0190
CDS294	192995	194227	+	Phage protein (ACLAME 1119)
CDS296	195150	195536	+	Phage protein
CDS299	196621	197583	+	DNA primase subunit
CDS300	197580	199850	+	Ribonucleotide reductase of class Ia (aerobic), alpha subunit (EC 1.17.4.1)
CDS301	199861	201000	+	Ribonucleotide reductase of class Ia (aerobic), beta subunit (EC 1.17.4.1)
CDS304	202922	204184	+	2',3'-cyclic-nucleotide 2'-phosphodiesterase (EC 3.1.4.16) / 3'-nucleotidase (EC 3.1.3.6)
CDS307	205213	209706	+	Phage tail fiber protein
CDS309	210071	213955	+	Secretory antigen precursor SsaA
RR	211970	212109	+	repeat region
RR	213455	213582	+	repeat region
tRNA	217289	217363	+	tRNA-Pro-TGG
CDS321	218540	220015	+	Phage protein
tRNA	220408	220517	+	tRNA-Undet-???

tRNA	220588	220660	+	tRNA-Arg-TCT
tRNA	220703	220775	+	tRNA-Gly-TCC
tRNA	220813	220886	+	tRNA-Ile-TAT
CDS325	221447	222238	+	glycosyltransferase family 11
CDS330	225088	226461	+	Glucose-1-phosphate thymidyltransferase (EC 2.7.7.24)
CDS337	231462	232259	+	glycosyltransferase family 25
CDS338	232411	232890	+	Septum-associated rare lipoprotein A
CDS339	232907	233101	+	Phage-encoded high light-inducible protein, HliPP #T4 GC0267
CDS340	233121	233318	+	phage T4-like hliPP, high light inducible protein # T4 GC 267
CDS347	235519	235749	+	Hypothetical-Protein belonging to T4-LIKE GC: 61
CDS350	236621	236875	+	Glutaredoxin, NrdC #T4 GC0311
CDS353	237608	238183	+	T4-like phage protein, cyanophage only #T4-like GC0313
CDS355	238408	238734	+	S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50), prokaryotic class 1B
CDS358	239200	239952	+	Thymidylate synthase ThyX (EC 2.1.1.148)
CDS361	240414	241175	+	Phosphate starvation-inducible protein PhoH, predicted ATPase #T4 GC0322
CDS363	242030	242701	+	Phage T4-like exonuclease A, DexA #T4 GC0325
CDS364	242688	242960	+	T4-like late promoter transcriptional accessory protein
CDS365	242971	243573	+	Phage DNA helicase loader # T4-like phage gp59, loader of gp41 DNA helicase #T4 GC0003
