
RECONSTITUTION DE L'HISTOIRE ÉVOLUTIVE DE DEUX CHAMPIGNONS LICHÉNISÉS LORS DES CHANGEMENTS CLIMATIQUES DU QUATERNAIRE GRÂCE À UNE APPROCHE DE GÉNOMIQUE DES POPULATIONS

Auteur : Balthazar, Alexis

Promoteur(s) : Magain, Nicolas; 9936

Faculté : Faculté des Sciences

Diplôme : Master en biologie des organismes et écologie, à finalité approfondie

Année académique : 2019-2020

URI/URL : <http://hdl.handle.net/2268.2/9842>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

RECONSTITUTION DE L'HISTOIRE ÉVOLUTIVE DE DEUX ESPÈCES DE CHAMPIGNONS LICHÉNISÉS LORS DES CHANGEMENTS CLIMATIQUES DU QUATERNAIRE GRÂCE À UNE APPROCHE DE GÉNOMIQUE DES POPULATIONS

BALTHAZAR ALEXIS

Sous la supervision des Drs. Nicolas Magain et Luc Cornet



*Mémoire de fin d'études présenté en vue de l'obtention du grade de Master en Biologie des
Organismes et Écologie*

Aout 2020



Faculté des Sciences
Département de Biologie, Écologie et Évolution
Unité de Biologie de l'Évolution et de la Conservation

**RECONSTITUTION DE L'HISTOIRE ÉVOLUTIVE DE DEUX ESPÈCES DE CHAMPIGNONS
LICHÉNISÉS LORS DES CHANGEMENTS CLIMATIQUES DU QUATERNAIRE GRÂCE À UNE
APPROCHE DE GÉNOMIQUE DES POPULATIONS**

— Année académique 2019-2020 —

BALTHAZAR ALEXIS

Membres du jury :

Professeur Patrick **DAUBY**
Professeur Jean-Christophe **PLUMIER**
Professeur Denis **BAURAIN**
Docteur Nicolas **MAGAIN**
Docteur Luc **CORNET**
Docteur Camille **TRUONG**
Docteur Alain **VANDERPOORTEN**

REMERCIEMENTS

Au terme de ce travail, je voulais tout particulièrement remercier mon promoteur, le Dr Nicolas Magain pour sa sympathie, ses conseils judicieux ainsi que pour son implication tout le long du mémoire. Je le remercie aussi pour m'avoir fait découvrir les États-Unis à travers une mission de récolte de plusieurs semaines. J'en profite pour remercier par la même occasion toutes ses connaissances, amis et famille qui m'ont accueilli pendant ce séjour.

Je tiens à remercier mon co-promoteur, le Dr Luc Cornet, pour son aide durant l'intégralité de la réalisation du pipeline informatique, ainsi que pour sa disponibilité de tous les instants.

Je remercie aussi Laurent Gohy pour m'avoir épaulé durant les périodes de laboratoire.

Je remercie le Professeur Denis Baurain pour m'avoir permis d'utiliser le cluster informatique Durandal afin de réaliser mes assemblages métagénomiques (financé par l'Université de Liège : 'Crédit de démarrage 2012' SFRD-12/04 et FRS-FNRS 'Crédit de recherche 2014' CDR J.0080.15.). Je remercie également le CÉCI pour m'avoir permis d'utiliser leurs clusters informatiques pour l'étape de tri des génomes (financés par le FRS-FNRS subvention n° 2.5020.11.).

Qu'aurait été ce travail sans le soutien de mes amis, en particulier Corentin et Marie. Ils ont été un véritable moteur de travail grâce à leur bonne ambiance.

Je ne pourrais terminer sans remercier Pauline, pour son soutien infaillible, ses nombreuses relectures et son Amour.

Pour finir, un énorme Merci à mes parents pour m'avoir permis d'entreprendre ces études, ainsi que pour le soutien apporté durant toutes ces années

Source photo de couverture :
Balthazar Alexis, *Peltigera appalachensis* ad. int.

RÉSUMÉ

Auteur : Alexis Balthazar

Date : août 2020

Promoteurs : Nicolas Magain, Luc Cornet

Titre : Reconstitution de l'histoire évolutive de deux espèces de champignons lichénisés lors des changements climatiques du Quaternaire grâce à une approche de génomique des populations.

Ce mémoire s'intéresse à deux espèces de champignons lichénisés, *Peltigera appalachensis* ad. int. et *Peltigera vitikainenii* ad. int. (Peltigerales, Lecanoromycetes), qui ont toutes les deux une répartition panboréale, mais qu'on retrouve également dans les montagnes tempérées d'Amérique du Nord pour la première, et au Japon pour la seconde. Notre hypothèse est que ces deux zones tempérées ont servi de refuge glaciaire pour l'ancêtre commun de ces deux espèces, que ces espèces se sont isolées l'une de l'autre pendant une période glaciaire, lorsque la zone boréale était sous glace, et n'ont recolonisé la zone boréale que récemment. L'objectif général de ce mémoire est de tester cette hypothèse via des méthodes de génétiques des populations sur des données génomiques. En effet, une recolonisation unique et récente de la zone boréale devrait conduire à une diversité génétique faible au sein de cette zone.

Plus particulièrement, le premier objectif a été de réaliser une mission sur le terrain dans la chaîne de montagnes des Appalaches et des Ozark aux États-Unis pour récolter les spécimens de *Peltigera appalachensis* ad. int. destinés au séquençage métagénomique.

Ensuite, nous avons sélectionné 57 thalles de *Peltigera* sur base de leurs séquences ITS et de leur origine géographique. Nous avons séquencé les métagénomomes de ces lichens avec la plateforme Illumina NovaSeq 6000. Les lichens étant des symbioses entre un champignon (dans notre cas, du genre *Peltigera*), un partenaire photosynthétique et une grande diversité de microorganismes qui forment son microbiome, nous avons séquencé les métagénomomes de ces thalles, c'est-à-dire le génome de tous les organismes qui le composent. Le second objectif de ce mémoire a été d'assembler ces métagénomomes et de réaliser un binning, c'est-à-dire de séparer les différentes séquences du métagénome selon l'organisme auquel elles appartiennent.

Enfin, le troisième objectif était de réaliser des analyses de génétique des populations sur la bin contenant le génome des *Peltigera*. Pour ce faire, nous avons fait du « SNP calling », nous avons détecté plusieurs milliers de sites du génome avec des variations dans la séquence d'ADN entre les différents individus (SNP, ou Single Nucleotide Polymorphism). Nous avons utilisé ces données SNP pour réaliser des analyses de génétique des populations. Parmi celles-ci, le clustering automatique et les analyses en composantes principales (ACP) séparent effectivement *P. appalachensis* et *P. vitikainenii* en deux groupes distincts, ce qui confirme qu'il s'agit bien de deux espèces bien isolées. De plus, ce dernier sépare également les spécimens boréaux et tempérés en ce qui concerne *P. appalachensis*. Les ACP groupent la plupart des échantillons boréaux de *P. vitikainenii*, qui sont génétiquement très proches, ensemble, à l'exception d'un échantillon d'Alaska qui représente sans doute un événement distinct de recolonisation de la zone boréale. La diversité génétique est plus élevée au sein des *P. appalachensis* tempérés en comparaison aux *P. appalachensis* boréaux. Cette même tendance est observée plus nettement encore concernant *P. vitikainenii*. L'hypothèse d'une recolonisation récente de la zone boréale à partir des potentielles zones refuges tempérées (Appalaches/Ozark et Japon, respectivement) est donc soutenue par nos données génomiques.

Lexique

ACP : Analyses en Composantes Principales

ADN : Acide DésoxyRibonucléique

CÉCI : Consortium des Équipements de Calcul Intensif

Bin : groupe de séquences partageant des caractères communs

Sub-bin : Sous-ensemble d'une bin

pb : paire(s) de bases

e-value : Dans un contexte de BLAST dans une base de données, la e-value (associée à un score d'alignement S) est le nombre d'alignements distincts, avec un score d'alignement équivalent ou supérieur à S, qui devraient se produire par chance pour notre recherche dans la base de données. Plus la e-value est proche de zéro, plus la « correspondance » entre notre séquence et celles renvoyées par la base de données sont significatives.

hits : Quand un alignement est trouvé entre une séquence en entrée et une base de données.

RefSeq : Reference Sequence Database. Cette collection de séquences est un sous-ensemble de NCBI, elle fournit un ensemble complet, intégré et non-redondant de séquences bien annotées.

LCA : « Last Common Ancestor ». Un algorithme LCA cherche à trouver le dernier ancêtre commun

NCBI : National Center for Biotechnology Information

BLAST : Basic Local Alignment Search tool

Fst : indice de fixation entre les populations

SNP : Single Nucleotide Polymorphism

NGS : Next Generation Sequencing

PCR : Polymerase Chain Reaction

ITS : Internal Transcribed Spacer

BAM : Binary Alignment/Map format ; C'est la version binaire du format SAM.

SAM : Sequence Alignment/Map format

Reads : Fragments de génome provenant d'une digestion partielle lors du séquençage.

Contig : Séquence résultant de l'assemblage de reads.

Scaffolds : Ensemble de contigs positionnés et orientés les uns par rapport aux autres sur base de leurs identité mais aussi grâce aux inserts des reads paired-end d'Illumina. Les scaffolds sont composés de séquences et de trous les séparant.

Splits : Les contigs de plus de 20.000 pb sont scindés pour être analysés par Anvio v.6.1. Ils sont alors nommés « splits ».

Sympatrique : Employé en écologie pour désigner deux espèces ou populations qui coexistent dans une même zone géographique

Ombilique : Point de fixation central de certains lichens foliacés.

Tomentum : Ensemble de poils plus ou moins courts, serrés, le plus souvent dressés.

Rhizine : Organe de fixation des thalles foliacés, simple ou ramifiée, formée d'un faisceau d'hyphes plus ou moins soudées et recouverte d'une gaine mucilagineuse facilitant l'adhésion au substrat.

Purine : Pellicule poudreuse ressemblant à du givre et qui recouvre certains thalles. Il s'agit soit de cristaux d'oxaloacétate, soit de cellules corticales mortes diffusant la lumière.

Thalle : partie végétative du lichen constituée des divers partenaires de la symbiose.

Symbiose : au sens anglosaxon — coexistence durable d'organismes différents, à relations variables

Apothécie : Ascome, le plus souvent sous forme de coupe arrondie, contenant, entre autre, l'hyménium.

Ascome : Structure issue de la reproduction sexuée du champignon, contenant les asques

Asque : Structure hyménial dans laquelle se forme les spores (souvent 8 par asque).

Foliacé : Thalle formé de lames plus ou moins lobées, facilement détachable du substrat auquel il est parfois fixé par des rhizines

Picoeucaryote : eucaryote de petite taille, souvent unicellulaire. Ce terme est souvent utilisé pour faire référence au plancton de très petite taille.

Pipeline : en informatique, il désigne un groupe de logiciel exécuté en série, de sorte que la sortie d'un logiciel serve d'entrée pour le suivant.

Binning : De l'anglais « mise en récipient », il désigne l'action de regrouper les données par classes. Dans le cas présent, de regrouper les séquences génétiques selon les organismes auxquelles ils appartiennent.

LISTES DES FIGURES ET TABLEAUX

FIGURE 1: REPRISE DE MIADLIKOWSKA ET AL. (2014). CLADOGRAMME SCHÉMATIQUE MONTRANT LES RELATIONS PHYLOGÉNÉTIQUE AU SEIN DES LECANOROMYCETES, AINSI QUE LES PHOTOBIONTES AVEC LESQUELS S'ASSOCIENT LES DIFFÉRENTS GROUPES (VOIR LÉGENDE À GAUCHE DE LA FIGURE).....	3
FIGURE 2 : PHOTO DE <i>PELTIGERA APPALACHENSIS</i> AD. INT., VERMONT, USA.....	4
FIGURE 3: ARBRE PHYLOGÉNÉTIQUE ADAPTÉ DE MAGAIN ET AL. (2017B) MONTRANT LES RELATIONS PHYLOGÉNÉTIQUES AU SEIN DE LA SECTION <i>POLYDACTYLON</i> . LES ESPÈCES IDENTIFIÉES COMME <i>P. NEOPOLYDACTYLA</i> AVANT LES ANALYSES MOLÉCULAIRES SONT REPRÉSENTÉES EN GRAS, ET LE CADRE GRIS MET EN ÉVIDENCE LE CLADE D'INTÉRÊT DE CE MÉMOIRE.	6
FIGURE 4: CARTE REPRÉSENTANT L'ÉTENDUE DES CALOTTES GLACIAIRES DURANT LE DERNIER MAXIMUM GLACIAIRE DANS LE MONDE, IL Y A ENVIRON 27 000 À 21 000 ANS. D'APRÈS LES DONNÉES D'EHLERS ET AL., 2011.....	7
FIGURE 5: MODIFICATION DES TEMPÉRATURES MOYENNES SELON LE MODÈLE GFDL POUR UN DOUBLEMENT DE LA CONCENTRATION EN CO ₂ . D'APRÈS GEOPHYSICAL FLUID DYNAMICS LABORATORY, PRINCETON UNIVERSITY, 2002.	8
FIGURE 6: CLADOGRAMME INFÉRÉ À PARTIR DE DONNÉES MOLÉCULAIRES PERMETTANT D'ESTIMER LES PÉRIODES DE DIVERGENCES DES DIFFÉRENTES LIGNÉES COMPOSANT LE GENRE <i>PELTIGERA</i> , ADAPTÉ DE LA DESCRIPTION DU PROJET SCIENTIFIQUE EN VUE DE L'OBTENTION DU CRÉDIT DE RECHERCHE F.R.S.-FNRS N°33695887 – LICHEN GENOMIC AND CLIMATE CHANGE.....	9
FIGURE 7: REPRÉSENTATION SCHÉMATIQUE DE LA PHYLOGÉNIE DE <i>P. APPALACHENSIS</i> ET <i>P. VITIKAINENII</i> . N. MAGAIN. IMAGE EXTRAITE DE LA DESCRIPTION DU PROJET SCIENTIFIQUE EN VUE DE L'OBTENTION DU CRÉDIT DE RECHERCHE F.R.S.-FNRS N°33695887 – LICHEN GENOMIC AND CLIMATE CHANGE.	10
FIGURE 8: CARTE EXPLICATIVE D'UN SCÉNARIO POSSIBLE, PERMETTANT D'EXPLIQUER L'ACTUELLE DISTRIBUTION GÉOGRAPHIQUE DES DIFFÉRENTS CLADES DE <i>PELTIGERA NEOPOLYDACTYLA</i> S.L. . N. MAGAIN. IMAGE EXTRAITE DE LA DESCRIPTION DU PROJET SCIENTIFIQUE EN VUE DE L'OBTENTION DU CRÉDIT DE RECHERCHE F.R.S.-FNRS N°33695887 – LICHEN GENOMIC AND CLIMATE CHANGE.	11
FIGURE 9: CARTOGRAPHIE DES DIFFÉRENTS LIEUX DE RÉCOLTE. LES LIEUX ÉCHANTILLONNÉS REPRÉSENTÉS EN VERT SONT CEUX OÙ <i>PELTIGERA APPALACHENSIS</i> AD. INT. A ÉTÉ TROUVÉ. LES LIEUX INDIQUÉS EN ROUGE SONT CEUX OÙ L'ESPÈCE N'A PAS ÉTÉ TROUVÉE.	13
TABLEAU 1 : LISTE DES ÉCHANTILLONS SÉQUENCÉS.....	15
FIGURE 10: VUE D'ENSEMBLE DES DIFFÉRENTES OPÉRATIONS RÉALISÉES AFIN D'ASSEMBLER LES READS.	16
FIGURE 11: VUE D'ENSEMBLE DES DIFFÉRENTES OPÉRATIONS RÉALISÉES POUR LE BINNING.....	17
FIGURE 12: INTERFACE DE VISUALISATION DES CONTIGS GRÂCE À <i>ANVIO</i> . DANS CET EXEMPLE, IL S'AGIT DE LA BIN N°6 GÉNÉRÉE PAR <i>MÉTABAT2</i> . CELLE-CI CONTENANT, ENTRE AUTRE, UN GRAND NOMBRE DE CONTIGS DU MYCOBIONTE DE L'ÉCHANTILLON <i>S18 - PELTIGERA VITIKAINENII</i> AD. INT.. L'ÉPAISSEUR DE LA BANDE CIRCULAIRE VERTE INFORME SUR LA PROPORTION DE GC DANS CHAQUE CONTIG. LES DEUX BANDES NOIRES AU-DESSUS CORRESPONDENT AU COVERAGE. LA BANDE ANNOTÉE « TAXONOMY » CONTIENT POUR CHACUN DES CONTIGS, L'AFFILIATION TAXONOMIQUE DE CELUI-CI (GÉNÉRÉ GRÂCE À <i>DIAMOND BLASTX</i> ET <i>PHYSETER</i>). LA LÉGENDE EN HAUT À DROITE INDIQUE LA TAXONOMIE DES CONTIGS (PHYLUM). LA BANDE CIRCULAIRE LA PLUS À L'EXTÉRIEUR INDIQUE LA SUB-DIVISION MANUELLE EFFECTUÉE SUR BASE DES DIFFÉRENTS PARAMÈTRES CITÉS CI-DESSUS.	20
FIGURE 13: VUE D'ENSEMBLE DES DIFFÉRENTES OPÉRATIONS RÉALISÉES POUR LE SNP CALLING.....	22
TABLEAU 2: TABLEAU RÉCAPITULATIF INCLUANT LA COMPLÉTUDE ET LA CONTAMINATION OBTENUES AVEC <i>BUSCO v 3.0.1</i> POUR LES 57 GÉNOMES DE <i>PELTIGERA</i> . LES LIGNES EN VERT REPRÉSENTENT LES ASSEMBLAGES GÉNOMIQUES QUI ONT ÉTÉ CONSERVÉS POUR LES ANALYSES DE GÉNÉTIQUE DES POPULATIONS. LES LIGNES ROUGES FONT RÉFÉRENCE AUX ASSEMBLAGES QUI N'ONT PAS ÉTÉ REPRIS SUR BASE DE LEUR FAIBLE COMPLÉTUDE (<40%). LES LIGNES JAUNES REPRÉSENTENT LES ASSEMBLAGES NON TRAITÉS CAR ILS NE CONCERNENT NI <i>PELTIGERA APPALACHENSIS</i> . NI <i>PELTIGERA VITIKAINENII</i>	23
FIGURE 14: ILLUSTRATION DU FORMAT VCF "VARIANT CALLING FORMAT". IMAGE ANNOTÉE PROVENANT DU SITE HTTP://DIGITHEADSLABNOTEBOOK.BLOGSPOT.COM/2013/01/VCF-VARIANT-CALL-FORMAT.HTML À PARTIR DE DANECEK ET AL., (2011).....	24
FIGURE 15: RÉSEAUX D'HAPLOTYPES ITS DE <i>P. APPALACHENSIS</i> . 9 RÉGIONS SONT REPRÉSENTÉES DANS (A), TANDIS QUE DANS (B), LE SUD (APPALACHIAN S) ET LE NORD (APPALACHIAN N) DES APPALACHES ONT ÉTÉ REGROUPÉS; ET LA NORVÈGE, LE QUÉBEC, LA RUSSIE, L'ALBERTA, L'ALASKA ET LA COLOMBIE BRITANNIQUE (BC) ONT ÉTÉ REGROUPÉS EN "BOREAL".....	27
FIGURE 16: RÉSEAUX D'HAPLOTYPES ITS DE <i>P. VITIKAINENII</i> . 10 RÉGIONS GÉOGRAPHIQUES SONT REPRÉSENTÉES DANS (A) TANDIS QUE JAPON ET EAST RUSSIA SONT REGROUPÉS; ET BC (COLOMBIE BRITANNIQUE), ESTONIA, FRANCE, NORWAY, UKRAINE, USA ET WEST RUSSIA SONT REGROUPÉS EN TANT QUE "OTHER" DANS (B).	28
TABLEAU 3 : TABLEAU REPRENANT DIFFÉRENTES STATISTIQUES CONCERNANT LES DONNÉES GÉNOMIQUES AVANT ET APRÈS L'ASSEMBLAGE. LES DONNÉES DANS LES COLONNES MARQUÉES D'UN ASTÉRISQUE NE TIENNENT COMPTE QUE DES CONTIGS AYANT UNE TAILLE SUPÉRIEURE OU ÉGALE À 500PB. PB= PAIRES DE BASES.	30

TABLEAU 4: TABLEAU RÉCAPITULATIF DES BINS ET SUB-BINS	32
FIGURE 17: VISUALISATION <i>ANVIO</i> DES 3 BINS GÉNÉRÉES PAR <i>METABAT 2</i> POUR L'ASSEMBLAGE <i>S67</i> . CES DERNIÈRES SONT SÉPARÉES EN SUB-BINS (BANDE EXTÉRIEURE) GRÂCE AUX INFORMATIONS COMBINÉES DU COVERAGE (BANDES NOIRES), DU CONTENU EN GC (BANDE VERTE) ET EN TENANT COMPTE DES AFFILIATIONS TAXONOMIQUES (BANDE "TAXONOMY"). A. BIN CONTENANT LA MAJORITÉ DES CONTIGS AFFILIÉS AUX ASCOMYCÈTES (2023). B. BIN CONTENANT LE RESTE DES CONTIGS ASCOMYCÈTES (944). C. BIN CONTENANT LES CONTIGS DE LA CYANOBACTÉRIE (551).	35
FIGURE 18 : CLUSTER DENDROGRAM DES 33 MYCOBIONTES DE <i>PELTIGERA</i> , RÉALISÉ SUR BASE DE LA DISTANCE DE PERVOSTI SUR LES DONNÉES SNP.	36
FIGURE 19: GRAPHIQUE DES PROBABILITÉS POSTÉRIEURS D'APPARTENANCE À UN GROUPE POUR LES 33 MYCOBIONTES DE <i>PELTIGERA</i> APRÈS ANALYSE DES DISTANCES DE PERVOSTI SUR LES DONNÉES SNP.	36
FIGURE 20: ÉVOLUTION DE LA VALEUR DU BIC (BAYESIAN INFORMATION CRITERION) OBTENU EN FONCTION DU NOMBRE DE CLUSTERS POUR LES 32 MYCOBIONTES DE <i>PELTIGERA</i> (APRÈS EXCLUSION DU <i>S44</i>).....	37
FIGURE 21: GRAPHIQUE DES PROBABILITÉS POSTÉRIEURS D'APPARTENANCE À UN GROUPE AVEC K=2 CLUSTERS, POUR LES 32 MYCOBIONTES DE <i>PELTIGERA</i> (APRÈS EXCLUSION DU <i>S44</i>) APRÈS ANALYSE DES DISTANCES DE PERVOSTI SUR LES DONNÉES SNP.....	37
FIGURE 22: GRAPHIQUE DES PROBABILITÉS POSTÉRIEURS D'APPARTENANCE À UN GROUPE AVEC K=5 CLUSTERS POUR LES 32 MYCOBIONTES DE <i>PELTIGERA</i> (APRÈS EXCLUSION DU <i>S44</i>) SUR BASE DES DONNÉES SNP.	38
FIGURE 23: LES FIGURES A ET B REPRÉSENTENT LE MÊME « CLUSTER DENDROGRAM » DES 32 MYCOBIONTES DE <i>PELTIGERA</i> , RÉALISÉ SUR BASE DE LA DISTANCE DE PERVOSTI SUR LES DONNÉES SNP. DANS LA FIGURE A, LE CODE COULEUR CORRESPOND AUX K=5 CLUSTERS CONSEILLÉS SUR BASE DU BIC (IDEM <i>FIGURE 22</i>). DANS LA FIGURE B, LES COULEURS CORRESPONDENT À L'ORIGINE GÉOGRAPHIQUE DES SPECIMENS: <i>P. APPALACHENSIS</i> TEMPÉRÉS EN ORANGE, <i>P. APPALACHENSIS</i> BORÉAUX EN BLEU, <i>P. VITIKAINENII</i> TEMPÉRÉS EN GRIS ET <i>P. VITIKAINENII</i> BORÉAUX EN VERT.....	39
FIGURE 24: ANALYSE EN COMPOSANTE PRINCIPALE (PCoA) SUR LES DONNÉES SNP POUR LES 32 MYCOBIONTES DE <i>PELTIGERA</i> AVEC <i>P. APPALACHENSIS</i> TEMPÉRÉS EN ORANGE, <i>P. APPALACHENSIS</i> BORÉAUX EN BLEU, <i>P. VITIKAINENII</i> TEMPÉRÉS EN GRIS ET <i>P. VITIKAINENII</i> BORÉAUX EN VERT. LA FIGURE A REPRÉSENTE LES DEUX PRINCIPAUX AXES DE LA PCoA, TANDIS QUE LA FIGURE B REPRÉSENTE L'AXE 1 ET L'AXE 3.....	40

TABLE DES MATIÈRES

INTRODUCTION	1
1. La symbiose lichénique	1
2. Le genre <i>Peltigera</i>	4
3. Les analyses des SNP (Single Nucleotide Polymorphism)	7
4. Changements climatiques du Quaternaire et spéciation	7
5. Les changements climatiques comme moteurs de la spéciation et la répartition de <i>P. appalachensis</i> et <i>P. vitikainenii</i>?	9
6. Objectifs du mémoire	12
MATÉRIELS ET MÉTHODES	13
1. Récolte des données	13
1.1 Échantillonnage.....	13
1.2 Extraction d'ADN.....	14
1.3 Sélection des échantillons pour le NGS	14
1.4 Séquençage	14
2. Traitement bioinformatique des données	16
2.1 Assemblage des métagénomés.....	16
2.1.1 Contrôle de la qualité	16
2.1.2 Suppression des adaptateurs d'illumina :.....	16
2.1.3 Assemblage des séquences via metaSPAdes :.....	17
2.2 Tri des génomes	17
2.2.1 Algorithmes de binning	17
2.2.2 Vérification du binning via Anvio.....	18
2.2.2.1 Affiliations taxonomiques des contigs.....	18
2.2.2.2 <i>Diamond BLASTX</i>	19
2.2.2.3 <i>Physeter</i>	19
2.2.2.4 Binning manuel via <i>Anvio</i>	19
2.2.3 Fusion des sub-bins.....	20
2.2.3.1 Affiliation taxonomique des sub-bins via Kraken.....	20
2.2.3.2 Affiliation taxonomique des sub-bins via <i>Diamond BLASTX</i> et <i>Physeter</i>	21
2.2.3.3 Regroupement des sub-bins	21
2.2.3.4 Complétude et contamination des assemblages génomiques finaux.....	21
3. Génétique des populations	22
3.1 SNP calling.....	22
3.1.1 Sélection des génomes	22
3.1.2 Alignement via BWA.....	24
3.1.3 SNP calling.....	24
3.2 Analyses de génétique des populations sur R.....	24
RÉSULTATS	26
1. Récolte des données	26
1.1 Mission de terrain et séquençage ITS	26
1.2 Réseaux d'haplotypes ITS	26
1.3 Séquençage	29
2.1 Tri des génomes	31
2.2 Complétude et contamination	35
3. Analyses de génétique des populations sur R	35
3.1 SNP calling.....	35

3.2 Clustering	36
3.3 Diversité génétique	41
DISCUSSION.....	42
1. Collecte des données	42
1.1 Mission de terrain.....	42
1.2 Cohérence des données génomiques avec les données ITS.....	42
2. Séquençage des génomes	42
2.1 Qualité du séquençage	42
3. Génétique des populations	43
3.1 Isolation des deux espèces <i>P. appalachensis</i> et <i>P. vitikainenii</i>	43
3.2 Structuration génétique des deux populations et hypothèse des refuges glaciaires.....	44
REFERENCES BIBLIOGRAPHIQUES	
ANNEXES	

INTRODUCTION

1. La symbiose lichénique

Un lichen est une association symbiotique stable entre un champignon hétérotrophe appelé mycobionte et un ou plusieurs partenaires photosynthétiques dénommés photobiontes (Nash et al. 2008). Les différents partenaires forment une structure morphologique unique qu'on appelle le thalle. La lichénisation est un mode de vie répandu au sein du règne des Fungi. Plus d'un cinquième des champignons connus sont lichénisés, soit plus de 19 000 espèces (Lücking et al. 2017). Les lichens ne forment pas un groupe monophylétique, ils regroupent des organismes partageant des similitudes dans leur mode de vie, à savoir obtenir des sucres produits par la photosynthèse de leur partenaire symbiotique unicellulaire (Lutzoni et al., 2001). Les données phylogénétiques modernes suggèrent que le phénomène de lichénisation est apparu indépendamment et à plusieurs reprises au cours de l'évolution (Lücking et al., 2017). 98 % des champignons lichénisés sont des Ascomycètes, mais il existe aussi des champignons lichénisés Basidiomycètes.

Généralement, le mycobionte s'associe avec une algue verte eucaryote comme photobionte. Cependant 8 % des lichens sont en symbiose avec des cyanobactéries. On parle alors de cyanolichens (Rikkinen 2002). La cyanobactérie peut soit être le seul photobionte du champignon (symbiose bipartite), soit être présente en complément d'une algue verte eucaryote (symbiose tripartite). Cette symbiose peut donc potentiellement impliquer 3 Règnes : Plantae, Fungi et Bacteria (Nash et al. 2008). La plupart des auteurs qualifient la relation symbiotique de mutualiste, impliquant donc des bénéfices réciproques. Cela étant, l'hypothèse d'un parasitisme contrôlé de la part du champignon n'est pas à exclure, car il semble que celui-ci tire plus de bénéfice de la symbiose que le photobionte. Un des arguments à l'encontre de l'hypothèse mutualiste est une croissance nettement plus rapide des photobiontes lorsqu'ils sont non-lichenisés (Ahmadjian 1993).

Le degré de lichénisation varie beaucoup selon les espèces, allant de cellules du mycobionte s'associant avec quelques cellules du photobionte sans organisation spatiale apparente, à une couche bien distincte de cellules du photobionte, qu'on appelle la couche photosynthétique, située sous le cortex du champignon lichénisé qui forme l'essentiel du thalle (Nash et al. 2008). Lorsqu'il est en symbiose avec une cyanobactérie, le mycobionte profite généralement non seulement des produits de la photosynthèse, mais aussi de la fixation d'azote atmosphérique par la cyanobactérie (Rai 2002).

La symbiose lichénique a permis aux différents partenaires d'occuper presque tous les habitats terrestres, depuis les tropiques jusqu'aux pôles (Nash 2008). Sans l'apparition de celle-ci, il est probable que ni les photobiontes ni les mycobiontes, seuls, n'auraient été aptes à la colonisation d'autant de niches diverses et variées. Certains lichens endolithes, se développent dans les micro-infractuosités présentes à la surface de roches en Antarctique (Friedmann 1982). On peut aussi en retrouver colonisant des sols nus dans des régions extrêmement arides (Evans and Johansen 1999 ; Belnap and Lange 2003). Dans ces niches extrêmes, les photobiontes et mycobiontes ne sont pas présents en dehors de ces associations symbiotiques. La symbiose serait donc la clé permettant de comprendre le succès évolutif de ces organismes au sein de niches écologiques extrêmes (Nash et al. 2008).

Un lichen ne se résume pas seulement à une symbiose bipartite ou tripartite entre un mycobionte et son/ses photobiontes. En effet, il est désormais établi que les lichens sont formés par des interactions entre de nombreux organismes, y compris un microbiome principalement composé de bactéries et de champignons se situant à l'intérieur du thalle, appelés endolichéniques (Aschenbrenner et al., 2016). Des études ont révélé chez différents genres de champignons lichénisés que les communautés bactériennes peuvent être spécifiques de leurs hôtes (Sierra et al., 2020). Par ailleurs, les compositions des communautés bactériennes de lichens apparentés sont généralement similaires (Ochman et al., 2010). Ces communautés se composent de différents phylum: parmi les plus fréquents, Actinobacteria, Bacteroidetes, Proteobacteria, Verrucomicrobia (Sierra et al. 2020).

Au sein des Ascomycètes, trois classes contiennent des champignons lichénisés: les Sordariomycètes, les Lecanoromycètes et les Eurotiomycètes (Nash et al., 2008). La classe des Lecanoromycètes contient environ $\frac{3}{4}$ des espèces de champignons lichénisés et regroupe 70% des genres (Lücking et al., 2017) (*Annexe 1*). Dans cette classe (*Figure 1*) figure l'ordre des Peltigerales qui contient actuellement plus de 1300 espèces réparties dans 67 genres (Lücking et al., 2017). Une des caractéristiques les plus remarquables des Peltigerales est que la plupart des espèces de champignons lichénisés de cet ordre le sont en symbiose avec une cyanobactérie, soit en symbiose bipartite, soit en symbiose tripartite (avec dans ce cas également une algue verte de la famille des Trebouxiaceae, Miadlikowska et al. 2004).

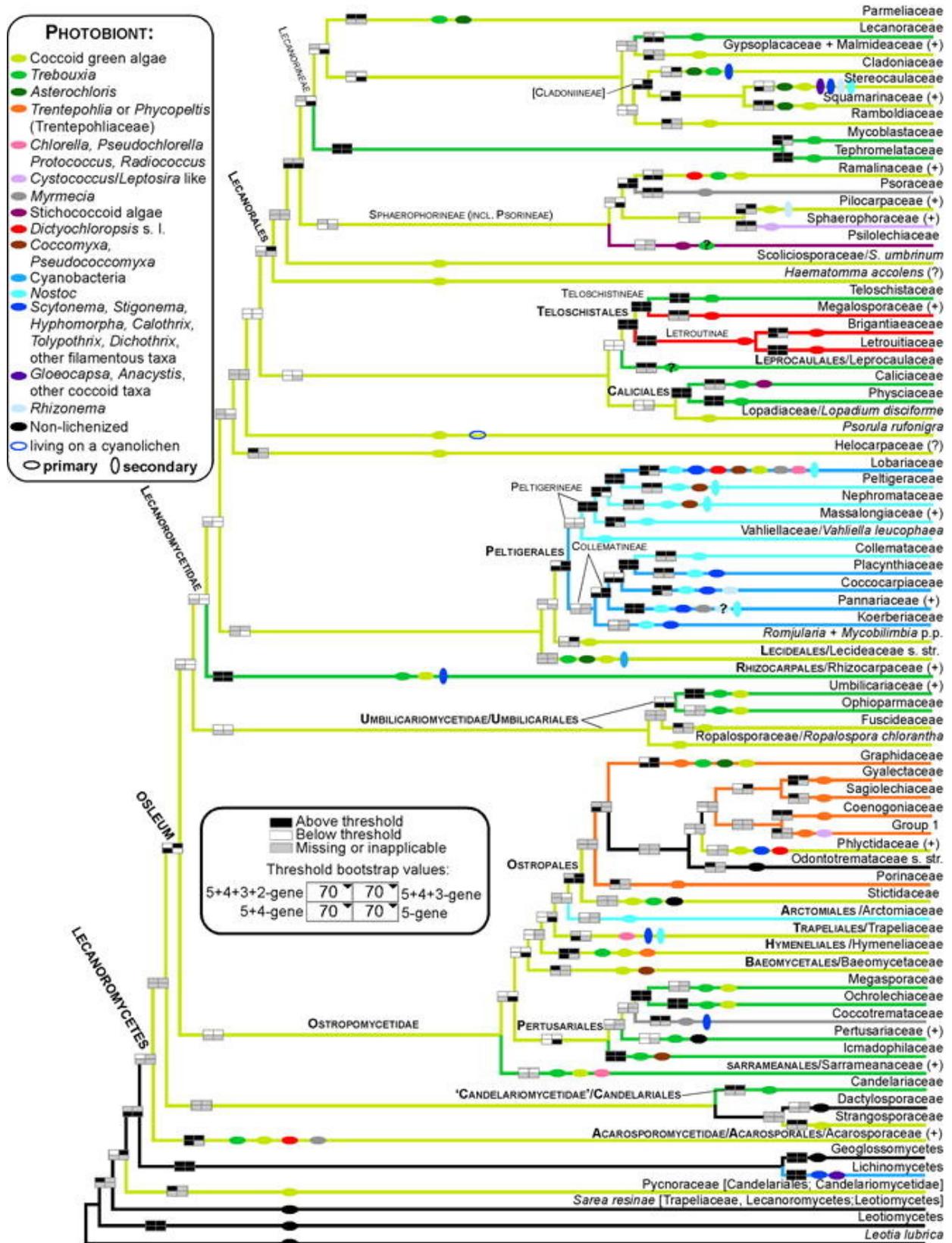


Figure 1: reprise de Miadlikowska et al. (2014). Cladogramme schématique montrant les relations phylogénétique au sein des Lecanoromycetes, ainsi que les photobiontes avec lesquels s'associent les différents groupes (voir légende à gauche de la figure).

2. Le genre *Peltigera*

Au sein du genre *Peltigera* (Lecanoromycetes, Peltigerales), toutes les espèces de champignons lichénisés sont en association avec une cyanobactérie du genre *Nostoc* (Vitikainen 1994). La plupart des espèces de *Peltigera* sont associées uniquement à *Nostoc* (symbiose bipartite) mais environ 10% des espèces de *Peltigera* sont également associés à une algue verte du genre *Coccomyxa* dans une association tripartite (Miadlikowska et al. 2004).

Le genre *Nostoc* est très diversifié, et paraphylétique dans sa circonscription actuelle (Otalora et al. 2010, Magain et al. 2017a). Les espèces qui le composent peuvent se présenter sous de multiples formes dans l'environnement. Elles peuvent vivre en colonie lamellaire ou ayant l'apparence de petites boules muqueuses pouvant aller jusqu'à plusieurs centimètres de diamètre. De nombreuses espèces de *Nostoc* sont également impliquées dans des symbioses stables avec divers végétaux et champignons, notamment des Anthocérotes, des Cycadales, des champignons lichénisés (O'Brien et al. 2005).

Les *Peltigera* sont des lichens foliacés, ce qui signifie que leur thalle se développe principalement en deux dimensions, comme la feuille d'un arbre (*Figure 2*; Nash 2008). Leur thalle peut atteindre jusqu'à 15 cm de diamètre. La face supérieure du thalle est de couleur verte pour les *Peltigera* en symbiose tripartite avec une algue verte du genre *Coccomyxa* et plutôt gris-brun aux reflets parfois bleuâtres pour les *Peltigera* en symbiose bipartite à cyanobactéries du genre *Nostoc* (*Figure 2*). La couleur est un paramètre pouvant varier considérablement suivant le degré d'hydratation du thalle et dépend de l'identité du *Nostoc* associé (Magain et al. 2017b). La face inférieure du thalle est munie de rhizines bien visibles, permettant de s'accrocher au substrat, et d'un réseau de veines plus ou moins marqué. Ces deux caractéristiques permettent d'identifier facilement le genre *Peltigera* (Vitikainen 1994). Les apothécies de *Peltigera* sont brun-noir à orange, le plus souvent allongée en forme de doigt, et située aux extrémités des lobes (Vitikainen 1994).



Figure 2 : Photo de *Peltigera appalachensis* ad. int., Vermont, USA.

Les thalles de *Peltigera* jouent un rôle très important au sein des écosystèmes arctiques et boréaux, notamment via la fixation d'azote atmosphérique par leurs symbiotes cyanobactériens (Henriksson et al., 1971), dans des environnements pauvres, où celui-ci est limitant.

Le genre *Peltigera* a été divisé en huit sections (Miadlikowska et al. 2000). Cinq sections, dont, la section *Polydactylon* (la deuxième plus riche en espèces), ne comprennent que des espèces formant des symbioses bipartites. Au sein de la section *Polydactylon*, l'épithète *Peltigera neopolydactyla s.l.* a été fréquemment utilisé pour identifier des espèces formant de grands thalles glabres à larges lobes dans l'hémisphère nord (Holtan-Hartwig 1993, Vitikainen 1994). La face supérieure du thalle est lisse, dépourvue de tomentum et de pruines, brun-noir ou gris sombre brillant avec des reflets bleus à l'état humide. Les extrémités du thalle ne portent jamais de phyllidies (propagules végétatives), mais possèdent régulièrement des apothécies toujours dressées, brun-rougeâtre à orangées. En ce qui concerne sa face inférieure, les veines sont très marquées, sombres et contrastent avec le fond blanc très clair. Les rhizines sont plutôt simples à la base et peuvent se ramifier légèrement au sommet. Elles sont longues de 7 à 10 mm. (Vitikainen, 1994).

Des analyses phylogénétiques sur 8 *loci* nucléaires menées sur plus de 250 thalles de lichens de la section *Polydactylon*, récoltés à travers le monde (Magain et al. 2017b) a montré que le nombre d'espèces de la section *Polydactylon* a été sous-évalué avant la vérification par données génétiques. Cette étude, mettant en lumière un grand nombre d'espèces cryptiques, propose de doubler le nombre d'espèces présentes dans cette section (Magain et al. 2017b). En particulier, cette étude a montré que les champignons lichénisés identifiés jusqu'alors comme *Peltigera neopolydactyla s.l.* représentaient un assemblage paraphylétique d'au moins dix espèces (Figure 3). Le terme « cryptique » est utilisé pour qualifier les espèces très difficiles, voire impossibles à distinguer d'un point de vue morphologique, mais qui représentent bien des lignées évolutives distinctes et isolées les unes des autres.

Ce mémoire s'intéresse à six espèces très proches phylogénétiquement, qui forment un groupe monophylétique (Figure 3; Magain et al. 2017a, 2017b). Cinq de ces six espèces sont similaires morphologiquement et étaient identifiées en tant que *Peltigera neopolydactyla s. l.* avant l'apparition des données moléculaires. Ces espèces n'ont pas encore été formellement décrites, mais une révision taxonomique est en cours (Magain et al. in prep.). Les cinq espèces en question sont *Peltigera appalachensis* ad. int., *Peltigera gayae* ad. int., *Peltigera vitikainenii* ad. int., *Peltigera mikado* ad. int. et *Peltigera asiatica* ad. int. (*Peltigera neopolydactyla* 1, *Peltigera neopolydactyla* 1b, *Peltigera neopolydactyla* 2a, *Peltigera neopolydactyla* 2b et *Peltigera neopolydactyla* 3 dans Magain et al. (2017a, 2017b), respectivement). Pour simplifier l'écriture ainsi que la compréhension, elles seront nommées *Peltigera appalachensis*, *Peltigera gayae*, *Peltigera vitikainenii*, *Peltigera mikado* et *Peltigera asiatica* dans le texte ci-après. La sixième, *Peltigera pacifica* (Vitikainen, 1985), est incluse dans cette étude car elle complète ce groupe monophylétique.



Figure 3: Arbre phylogénétique adapté de Magain et al. (2017b) montrant les relations phylogénétiques au sein de la section *Polydactylon*. Les espèces identifiées comme *P. neopolydactyla* avant les analyses moléculaires sont représentées en gras, et le cadre gris met en évidence le clade d'intérêt de ce mémoire.

3. Les analyses des SNP (Single Nucleotide Polymorphism)

Depuis quelques années, l'arrivée et la démocratisation des techniques de séquençage de nouvelle génération à haut débit (NGS), ont permis d'avoir accès à l'information contenue dans les génomes complets, à des coûts réduits (Schuster et al., 2008). De ce fait, les analyses de génétique des populations ne sont plus limitées à la comparaison de quelques dizaines de sites variables (comme les régions microsatellites, par exemple). Elles peuvent s'effectuer grâce à l'analyse de milliers de sites avec polymorphisme nucléotidique, plus communément appelés SNP (Single Nucleotide Polymorphism) sur l'entièreté du génome (Brumfield et al., 2003). Ce sont des variations (polymorphismes) d'une seule paire de bases entre différents génomes. Elles sont assez fréquentes chez la plupart des organismes et peuvent aussi bien se trouver dans les régions codantes (exons) que dans les régions non codantes des gènes (introns), ainsi que dans les régions intergéniques. Ces SNP permettent donc d'obtenir des milliers, voire des dizaines de milliers de points de comparaison entre nos génomes. Elles offrent de ce fait une meilleure résolution dans les analyses de distances génétiques entre populations et assurent une compréhension plus fine des mécanismes de spéciation (Brumfield et al., 2003).

4. Changements climatiques du Quaternaire et spéciation

La période géologique du Quaternaire s'étend de -2,59 Ma (Millions d'années) à nos jours. Elle couvre deux époques : Le Pléistocène et l'Holocène. Certains auteurs en ajoutent une troisième, l'Anthropocène (Lewis et al. 2015). Durant le Quaternaire a lieu une alternance des périodes glaciaires et interglaciaires. La dernière période glaciaire débute à la fin du Pléistocène, il y a environ 110 ka (milliers d'années). La fin de cette dernière, marquée par le début de la période interglaciaire dans laquelle nous nous trouvons actuellement, coïncide avec le début de l'Holocène il y a environ 10 ka.

Il y a 21 000 ans, en plein maximum glaciaire (Bartlein et al. 2010), la banquise s'étendait sur des latitudes beaucoup plus basses que ce que l'on observe actuellement (Figure 4). Ces changements climatiques sont à l'origine de nombreuses spéciations (Liu et al. 2013).

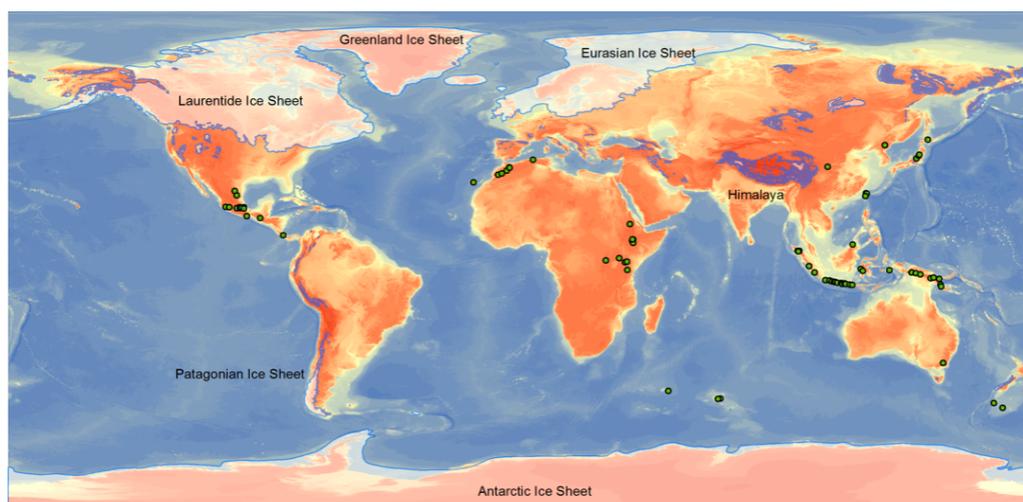


Figure 4: Carte représentant l'étendue des calottes glaciaires durant le dernier maximum glaciaire dans le monde, il y a environ 27 000 à 21 000 ans. D'après les données d'Ehlers et al., 2011.

Comprendre comment les organismes se sont adaptés aux événements climatiques passés, peut nous permettre d'affiner les modèles prédictifs concernant les changements climatiques du présent et du futur. Ceux-ci peuvent être à l'origine de spéciations, notamment via certains changements d'aires de distribution d'espèces (Excoffier et al., 2009). La compréhension de ces mécanismes de spéciation, pourrait permettre de mieux anticiper l'impact futur des changements climatiques en cours.

L'impact des changements climatiques en cours n'est pas uniforme sur terre. En effet, les hautes latitudes, telles que les zones arctiques et boréales seront plus rapidement et plus fortement impactées par l'augmentation de température (*Figure 5*).

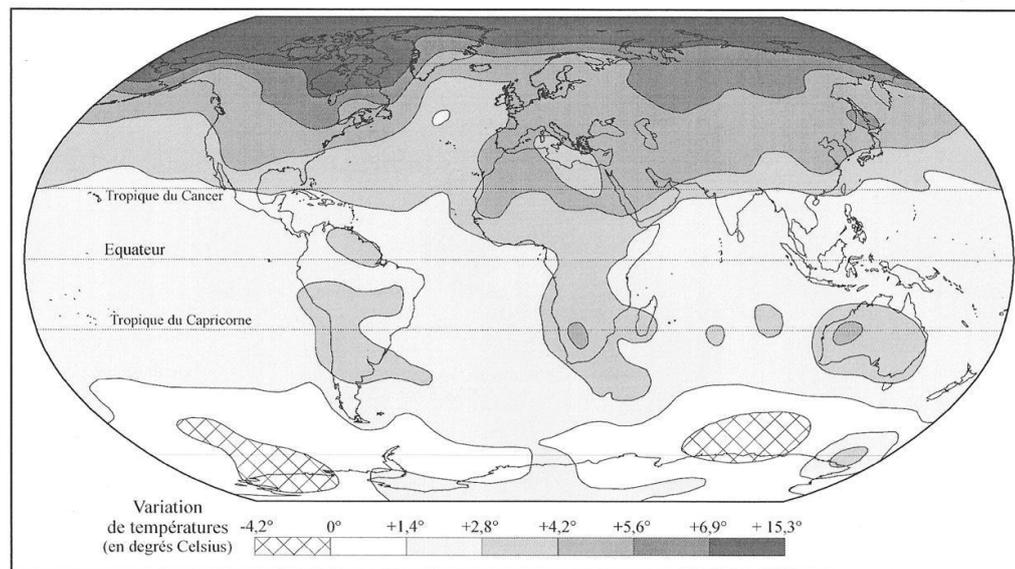


Figure 5: Modification des températures moyennes selon le modèle GFDL pour un doublement de la concentration en CO₂. d'après Geophysical Fluid Dynamics Laboratory, Princeton University, 2002.

Il est très important de pouvoir prédire la réponse de la biodiversité face aux changements climatiques. Les modèles prédictifs peuvent servir à alerter la communauté scientifique, mais aussi les décideurs politiques afin de réduire l'impact de des changements climatiques sur la biodiversité (Pereira et al. 2010; Parmesan et al., 2011). Le « Millenium Ecosystem Assessment » prédit des bouleversements importants pour 5-20% des écosystèmes terrestres (Sala et al., 2005). Ce dernier prévoit des perturbations plus conséquentes concernant les forêts de conifères tempérées, la toundra, les savanes ainsi que les forêts boréales. Ces écosystèmes pourraient dépasser leurs seuils de tolérance aux perturbations menant ainsi à des changements irréversibles de biomes (Leadley et al., 2010).

Les espèces peuvent s'adapter aux changements climatiques en modifiant leurs niches écologiques de trois différentes manières : en changeant leur phénologie, leurs aires de distribution ou leur physiologie (Bellard et al., 2011). Ces changements de niches spatiales ont déjà été observés chez plus de 1000 espèces (Parmesan, 2006). La limite des arbres devrait s'étendre sur les zones de hautes altitudes ou latitudes en provoquant un changement des communautés alpines et de la toundra (Alo et al., 2008). Les changements les plus observés et les mieux documentés sont ceux impactant la phénologie, notamment dans une méta-analyse portant sur de nombreux animaux et végétaux présentant des réponses aux changements climatiques (Root et al., 2003). Cette dernière

met en évidence le décalage temporel d'évènements phénologiques clés. Ils se produisent en moyenne 5.1 jours plus tôt, par tranche de 10 ans et ce depuis 50 ans. Ceci peut provoquer une asynchronisation dans les relations nécessaires au bon fonctionnement de communautés biologiques (interactions proie-prédateur et insecte-plante par exemple) (Parmesan, 2006).

Différents modèles prédictifs sont comparés et détaillés dans Bellard et al. (2011). Dans certains modèles, une extinctions de 39% à 43% des espèces endémiques est attendue. Ces chiffres se basent sur l'analyse de 25 hotspots de biodiversité (Malcolm et al., 2006). Une étude de Maclean & Wilson (2011) donne de nouvelles estimations sur la perte nette de biodiversité imputable aux changements climatiques d'ici 2100. Selon leurs estimations basées sur différentes méta-analyses multi-taxa à l'échelle du globe, 12,6% des plantes seraient amenées à disparaître. Il en va de même concernant 9,4% des invertébrés et 17,7% des vertébrés.

5. Les changements climatiques comme moteurs de la spéciation et la répartition de *P. appalachensis* et *P. vitikainenii*?

Ce travail va se concentrer sur l'histoire évolutive du clade comportant *Peltigera vitikainenii* et *Peltigera appalachensis*. Nous savons, grâce aux travaux de Magain et al. (2017a, b), que *P. appalachensis* et *P. vitikainenii* sont deux espèces très proches d'un point de vue phylogénétique (Figure 6) en plus d'être cryptiques du point de vue morphologique.

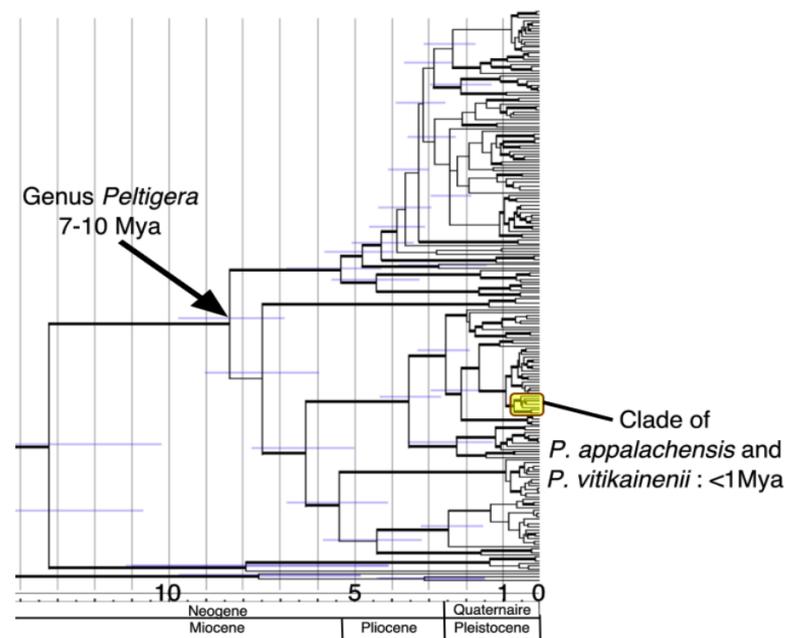


Figure 6: Cladogramme inféré à partir de données moléculaires permettant d'estimer les périodes de divergences des différentes lignées composant le genre *Peltigera.*, adapté de la description du projet scientifique en vue de l'obtention du Crédit de recherche F.R.S.-FNRS n°33695887 – Lichen genomic and climate change

Le clade dans lequel se trouvent *P. appalachensis* et *P. vitikainenii* est relativement récent, puisqu'il est apparu il y a moins d'un million d'années (Figure 6). Ceci laisse penser que ces deux espèces ont pu diverger lors des derniers changements climatiques du Quaternaire.

Ces deux espèces, parmi les plus communes identifiées comme “*P. neopolydactyla* s.l.” sont considérées par la plupart des spécialistes comme des espèces boréales (voir par exemple Holtan-Hartwig 1993, Vitikainen 1994). Les deux espèces coexistent en sympatrie dans une grande partie de la zone boréale, notamment au Canada, en Norvège et en Russie. *P. appalachensis* est également présent plus au sud en zone tempérée en Amérique du Nord, dans les Appalaches (d’Alabama jusqu’au Vermont) et les Ozark (Missouri et Arkansas) (*Figure 7*). Quant à *P. vitikainenii*, on le retrouve également au Japon ainsi que dans les régions océaniques de l’Est de la Russie (Sakhaline, Khabarovsk); *P. vitikainenii* est aussi présent en Europe dans les Alpes, les Vosges et les Carpathes. Or, sur base des données ITS (locus très variable très largement utilisé chez les champignons, notamment utilisé comme code-barre pour identifier les espèces; Schoch et al. 2012), il est intéressant de constater que la diversité génétique des deux espèces est très faible dans la région boréale (un haplotype ITS très majoritaire par espèce sur les trois continents), alors que la diversité génétique de *P. appalachensis* semble plus importante dans les zones tempérées d’Amérique du Nord, et celle de *P. vitikainenii* dans les zones tempérées et sub-boréales est-asiatique (*Figure 7*).

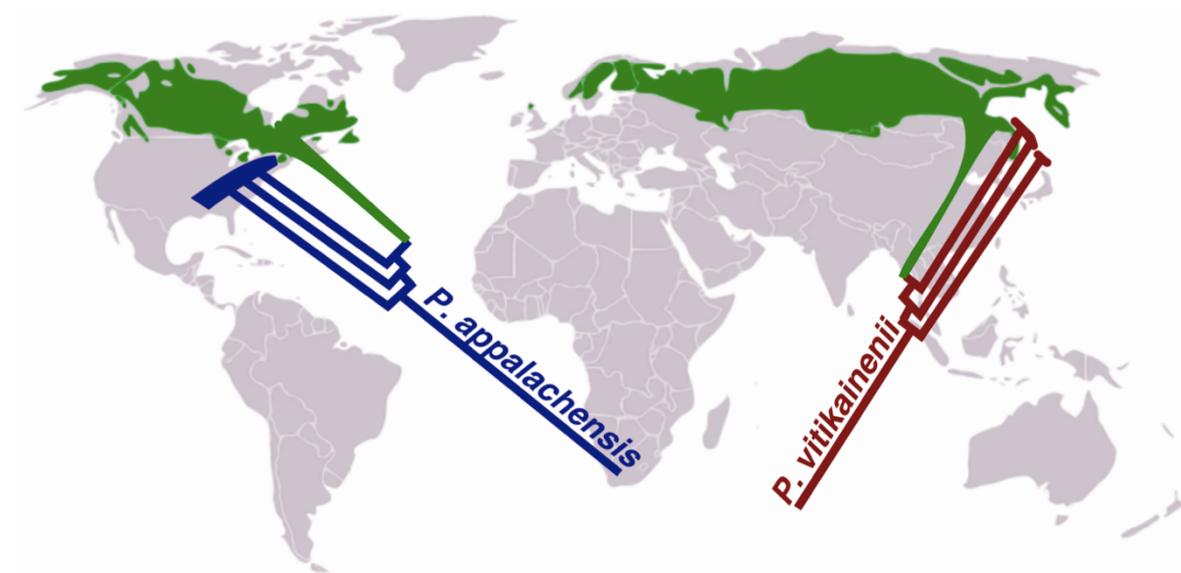


Figure 7: Représentation schématique de la phylogénie de *P. appalachensis* et *P. vitikainenii*. N. Magain. Image extraite de la description du projet scientifique en vue de l’obtention du Crédit de recherche F.R.S.-FNRS n°33695887 – Lichen genomic and climate change.

La situation pourrait s’expliquer par un scénario où l’ancêtre commun de ces deux espèces était une espèce avec une répartition panboréale (*Figure 8*). Ensuite, lors d’une période glaciaire, lorsque la zone boréale était sous glace, une population de l’espèce aurait pu trouver un refuge dans les régions tempérées d’Amérique du Nord, et une autre en Asie, la banquise de l’époque bloquant la possibilité de brassage génétique entre les sous-populations. L’isolement spatial de ces sous-populations pendant la dernière période glaciaire, pourrait donc être à l’origine de la divergence génétique observée chez ces deux espèces sœurs. Ces deux populations auraient divergé et seraient devenues deux espèces distinctes, avant de recoloniser chacune récemment la zone boréale à partir de leurs refuges, une fois que les conditions climatiques le permettaient à nouveau.

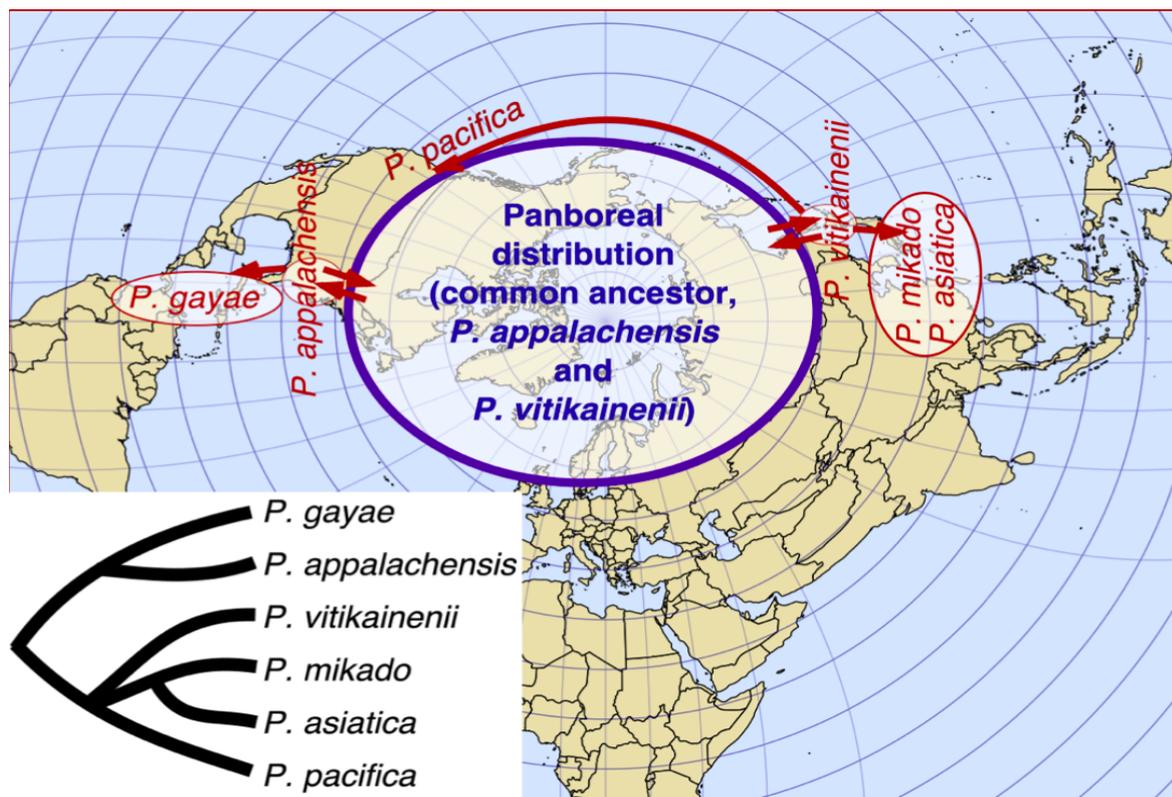


Figure 8: Carte explicative d'un scénario possible, permettant d'expliquer l'actuelle distribution géographique des différents clades de *Peltigera neopolydactyla* s.l. . N. Magain. Image extraite de la description du projet scientifique en vue de l'obtention du Crédit de recherche F.R.S.-FNRS n°33695887 – Lichen genomic and climate change.

Pour tester cette hypothèse, deux grands axes complémentaires ont été envisagés. Ces deux axes reposent sur l'obtention des génomes de représentants de ces espèces.

Le premier consistera en une approche phylogénomique, ou la reconstitution de l'histoire évolutive de ces deux espèces sur base de centaines de loci permettra à la fois de dater les différents événements (divergence des deux espèces, reconquête éventuelle de la zone boréale par les deux espèces) pour voir s'ils correspondent au tempo des changements climatiques, et de vérifier que l'origine de chaque espèce est bien la zone tempérée, si les thalles boréaux forment un clade « nésté » au sein de la diversité phylogénétique des lignées tempérées.

Le second axe, dont le présent travail est la première étape, consistera à vérifier que la diversité génétique est plus importante dans les zones tempérées que dans les zones boréales, ce qui suggérerait que les populations y sont établies depuis plus longtemps, et qu'un sous-ensemble de ces populations a ensuite migré vers les zones boréales. Cette approche de génomique des populations se fera via l'analyse de milliers de SNP (Single Nucleotide Polymorphism). Ces données permettront aussi de tester si les deux espèces sont bien isolées ou s'il existe encore un certain degré d'hybridation dans les zones où elles coexistent en sympatrie.

6. Objectifs du mémoire

1. Récolter des échantillons de *Peltigera appalachensis* dans l'ensemble de sa distribution tempérée connue, afin d'obtenir du matériel frais pour le séquençage de nouvelle génération de métagénomes de cette espèce.
2. Séquencer les metagénomes de 57 thalles de *P. appalachensis*, *P. vitikainenii* et leurs espèces proches, et établir un pipeline bioinformatique pour obtenir les séquences correspondant aux différents partenaires présents, en particulier le champignon *Peltigera* et la cyanobactérie *Nostoc*.
3. Vérifier grâce aux données génomiques que *P. appalachensis* et *P. vitikainenii* sont bien isolées du point de vue génétique, et représentent deux espèces distinctes.
4. Vérifier l'hypothèse que les zones tempérées, en Amérique du Nord et en Asie de l'Est pour *P. appalachensis* et *P. vitikainenii*, respectivement, ont servi de refuge à ces espèces pendant la période glaciaire, et que la recolonisation de la zone boréale est plus récente

MATÉRIELS ET MÉTHODES

1. Récolte des données

1.1 Échantillonnage

La majorité des spécimens de *Peltigera appalachensis* ont été récoltés dans les Appalaches lors d'une mission qui s'est déroulée en septembre et octobre 2019. L'intérêt principal de cette mission était de récolter du matériel frais, susceptible de donner de meilleurs résultats lors du séquençage. Les zones d'échantillonnage couvrent la majorité des états sur lesquels la chaîne de montage des Appalaches est présente, mais aussi le Missouri et l'Arkansas (montagnes des Ozark) (Figure 9). La détermination précise des différents lieux d'échantillonnage fut élaborée à l'aide des annotations des lieux de récoltes de *Peltigera neopolydactyla s.l.* présentes dans différents herbiers, notamment en utilisant la base de données du CNALH (Consortium of North American Lichen Herbaria, <http://lichenportal.org/cnalh/>). Les échantillons récoltés ont été placés dans des enveloppes en papier annotées. Celles-ci étaient mises à sécher à la fin de chaque journée de terrain pour éviter que de la moisissure ne se forme sur le thalle. Le séchage des échantillons s'effectue rapidement et simplement en laissant les enveloppes ouvertes dans un environnement sec. Les échantillons récoltés ont été envoyés à l'Université de Liège pour extraction d'ADN et séquençage.



Figure 9: Cartographie des différents lieux de récolte. Les lieux échantillonnés représentés en vert sont ceux où *Peltigera appalachensis* ad. int. a été trouvé. Les lieux indiqués en rouge sont ceux où l'espèce n'a pas été trouvée.

1.2 Extraction d'ADN

L'ADN de 48 thalles identifiés comme *Peltigera appalachensis* collectés lors de la mission de récolte décrite ci-dessus a été extrait en suivant le protocole de Cubero et al, 1999. Les spécimens choisis pour l'extraction d'ADN étaient bien préservés et ne présentaient pas de signe visible d'infection par des champignons parasites. Le locus ITS a été séquencé pour vérifier l'identité du mycobionte et pour que l'haplotype ITS guide le choix des échantillons pour le séquençage de nouvelle génération. Le locus ITS a été séquencé avec les amorces ITS1F (Gardes and Bruns, 1993) et ITS4 (White et al., 1990), en suivant le programme préconisé dans Magain et al. (2017a).

1.3 Sélection des échantillons pour le NGS

Des réseaux d'haplotypes ont été réalisés en utilisant l'ensemble des séquences ITS disponibles pour *P. appalachensis* (116 séquences, y compris celles récoltées dans la mission décrite plus haut) et *P. vitikainenii* (161 séquences), en utilisant la fonction haploNet du package R pegas (Paradis 2010), afin de comparer la diversité génétique présente au sein de ces espèces, et de sélectionner les haplotypes pour le séquençage NGS.

Sur les 57 métagénomés séquencés, 11 proviennent de la mission de récolte décrite ci-dessus. Les 46 autres spécimens sont issus de précédentes études de Magain et ses collègues (y compris une mission au Japon en 2019) ou sont des emprunts issus d'herbiers institutionnels (*Tableau 1*).

1.4 Séquençage

58 thalles ont été sélectionnés pour le séquençage de nouvelle génération. Pour *Peltigera appalachensis* et *P. vitikainenii*, l'haplotype ITS et l'origine géographique ont été utilisés comme critères de sélection. Pour les autres espèces, l'identité géographique des spécimens a été utilisée. Les 58 thalles choisis (voir *tableau 1*) représentaient:

-14 thalles de *P. appalachensis* représentant l'ensemble de sa répartition tempérée, de l'Arkansas au Vermont, récoltés lors de notre mission de terrain (zone du refuge glaciaire potentiel)

-10 thalles de *P. appalachensis* provenant de la zone boréale sur les trois continents

-12 thalles de *P. vitikainenii* provenant du Japon et de l'Est de la Russie (zone du refuge glaciaire potentiel)

-10 thalles de *P. vitikainenii* provenant du reste de sa distribution, dans les zones boréales des trois continents ainsi qu'une population des Vosges (France) et une des Alpes (Suisse).

-3 thalles pour chacune des espèces voisines suivantes: *P. asiatica*, *P. mikado*, *P. gayae*, *P. pacifica*

58 métagénomés ont été envoyés pour séquençage au Duke University Sequencing and Genomic Technologies Shared Resource (Durham, NC, USA). Le séquençage a été réalisé avec la plateforme de séquençage NovaSeq 6000 d'Illumina (États-Unis, San Diego, CA 92122). Avant le séquençage, des bibliothèques de reads paired-end de 151pb avec des tailles d'insert d'environ 500pb ont été produites pour les échantillons. La création de la bibliothèque a échoué pour un échantillon de *P. gayae*, et les 57 échantillons restants ont été séquencés.

Tableau 1 : Liste des échantillons séquencés.

N°	Espèces	Origine géographique	Année de collecte	Herbier	Collecteurs
S14	<i>Peltigera pacifica</i>	Canada:British Columbia	2012	UBC	T. Goward
S15	<i>Peltigera pacifica</i>	USA:Oregon	2005	OSC	B. McCune
S16	<i>Peltigera pacifica</i>	Canada:British Columbia	1994	H	O. Vitikainen
S17	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S18	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S19	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S20	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S21	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S22	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S23	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S24	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S25	<i>Peltigera vitikainenii</i>	Japan:Honshu	2017	TU	I. Juriado
S26	<i>Peltigera vitikainenii</i>	Japan:Honshu	2017	TU	i. Juriado
S27	<i>Peltigera vitikainenii</i>	Russia:Khabarovsk	2018	MHA	L Konoreva
S28	<i>Peltigera vitikainenii</i>	Russia:Sakhalin	2013	SAKH	S. Tchabanenko
S29	<i>Peltigera vitikainenii</i>	France: Vosges	2019	LG	N. Magain
S30	<i>Peltigera vitikainenii</i>	Suisse	1998	G	Mathias Vust
S31	<i>Peltigera vitikainenii</i>	Canada:British Columbia	inconnu	UBC	T. Goward
S32	<i>Peltigera vitikainenii</i>	Canada:British Columbia	inconnu	UBC	T. Goward
S33	<i>Peltigera vitikainenii</i>	USA:Alaska	2013	OSC	K. Spickerman
S34	<i>Peltigera vitikainenii</i>	Estonia	2017	TU	i. Juriado
S35	<i>Peltigera vitikainenii</i>	Norway	2011	LG	N. Magain
S36	<i>Peltigera vitikainenii</i>	Norway	2011	LG	N. Magain
S37	<i>Peltigera vitikainenii</i>	Norway	2011	LG	N. Magain
S38	<i>Peltigera vitikainenii</i>	Norway	2011	LG	N. Magain
S39	<i>Peltigera appalachensis</i>	USA:Vermont	2019	LG	Magain Balthazar
S40	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	LG	Magain Balthazar
S41	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	LG	Magain Balthazar
S42	<i>Peltigera appalachensis</i>	USA:Vermont	2019	LG	Magain Balthazar
S43	<i>Peltigera appalachensis</i>	USA: Alabama	2019	LG	Magain Balthazar
S44	<i>Peltigera appalachensis</i>	USA: Missouri	2019	LG	Magain Balthazar
S45	<i>Peltigera appalachensis</i>	USA: Connecticut	2019	LG	Goffinet
S46	<i>Peltigera appalachensis</i>	USA: Alabama	2019	LG	Magain Balthazar
S47	<i>Peltigera appalachensis</i>	USA: South Carolina	2019	LG	Magain Balthazar
S48	<i>Peltigera appalachensis</i>	USA: Arkansas	2019	LG	Magain Balthazar
S49	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	LG	Magain Balthazar
S50	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	LG	Magain Balthazar
S51	<i>Peltigera appalachensis</i>	Canada: Quebec	2011	DUKE	Miadlikowska et al.
S52	<i>Peltigera appalachensis</i>	Canada: Quebec	2011	DUKE	Miadlikowska et al.
S53	<i>Peltigera appalachensis</i>	Canada: Alberta	inconnu	PMAE	D. Haughland
S54	<i>Peltigera appalachensis</i>	Canada: Alberta	inconnu	PMAE	D. Haughland
S55	<i>Peltigera appalachensis</i>	Canada: British Columbia	inconnu	UBC	T. Goward
S56	<i>Peltigera appalachensis</i>	Russia: Sakha	2011	H	Teuvo Ahti
S57	<i>Peltigera appalachensis</i>	Russia: Komi	2019	TU	Inga Juriado
S58	<i>Peltigera appalachensis</i>	Norway	2011	LG	Magain
S59	<i>Peltigera appalachensis</i>	USA: Alaska	2006	CONN	Bernard Goffinet
S60	<i>Peltigera appalachensis</i>	USA: Alaska	2011	OSC	Bruce McCune
S61	<i>Peltigera asiatica</i>	Malaysia: Borneo	2018	DUKE	Magain et al.
S62	<i>Peltigera asiatica</i>	China: Yunnan	2003	DUKE	Miadlikowska
S63	<i>Peltigera asiatica</i>	Taiwan	2016	LG	Serussiaux
S64	<i>Peltigera mikado</i>	Russia: Khabarovsk	2013	DUKE	Miadlikowska
S65	<i>Peltigera mikado</i>	China: Yunnan	2010	CONN	Goffinet
S66	<i>Peltigera mikado</i>	Taiwan	2016	LG	Serussiaux
S67	<i>Peltigera gayae</i>	Peru	2011	Duke	Gaya
S68	<i>Peltigera gayae</i>	Puerto Rico	2018	F	Mercado-Diaz
S91	<i>Peltigera appalachensis</i>	USA:South Carolina	2019	LG	Magain Balthazar
S93	<i>Peltigera appalachensis</i>	USA:West Virginia	2019	LG	Magain Balthazar

2. Traitement bioinformatique des données

2.1 Assemblage des métagénomomes

Aperçu du pipeline d'assemblage :

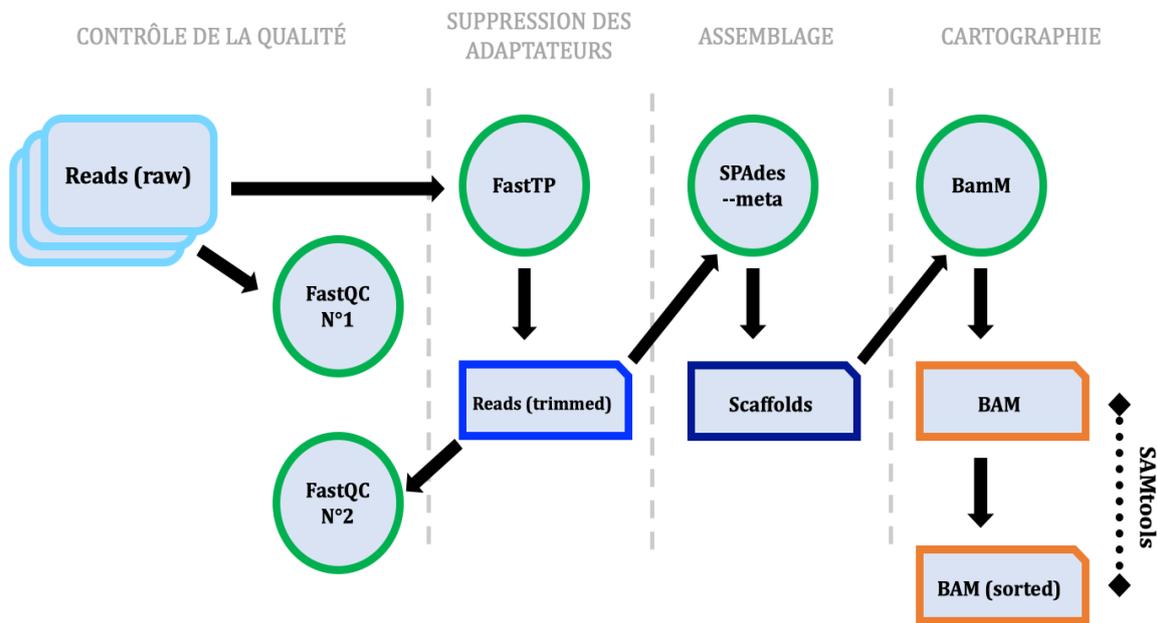


Figure 10: Vue d'ensemble des différentes opérations réalisées afin d'assembler les reads.

2.1.1 Contrôle de la qualité

Le séquençage produit deux fichiers au format FastQ contenant les séquences brutes, ou reads, dans les deux sens (forward et reverse, respectivement). La première étape du pipeline bioinformatique utilise *FastQC* v 0.11.9 (Brown et al., 2017) avec les réglages par défaut pour contrôler la qualité des séquences brutes des 57 individus au format FASTQ. Cet outil fournit des statistiques permettant d'avoir un rapide aperçu de la qualité des séquences brutes, notamment après un séquençage haut débit de type Illumina. Ainsi *FastQC* génère un rapport HTML qui permet de visualiser des scores de qualité résumés selon plusieurs paramètres tels que la longueur moyenne des séquences, la qualité des séquences par base, ou encore la présence/absence d'adaptateur. Ceci afin de s'assurer que la qualité du séquençage est suffisante.

2.1.2 Suppression des adaptateurs d'Illumina :

Après fractionnement de l'ADN, des adaptateurs sont ajoutés aux reads durant le processus de séquençage Illumina. Il est nécessaire de les retirer afin d'éviter qu'ils n'interfèrent dans les analyses faites en aval ou pendant l'assemblage des séquences. *FastTP* v. 0.19.6 (Shifu et al., 2018), avec les réglages par défaut, permet de cibler les séquences des adaptateurs et de les enlever. *FastTP* reconnaît automatiquement les adaptateurs utilisés dans la plupart des NGS « Next Generation Sequencing », notamment ceux d'Illumina NovaSeq 6000. Ensuite, *FastQC* v 0.11.9 est à nouveau utilisé sur les séquences pour s'assurer que les adaptateurs ont été correctement repérés et retirés efficacement par *FastTP*.

2.1.3 Assemblage des séquences via metaSPAdes :

L'assemblage permet de grouper les reads pour former des séquences plus longues, appelées contigs. Un assembleur métagénomique est utilisé. Il s'agit de *metaSPAdes* v.3.7.0 (Nurk et al., 2017), spécialement conçu pour procéder à l'assemblage de bibliothèques de séquences provenant d'organismes différents. Les reads, sont réalignés sur l'assemblage via *BamM* (<http://ecogenomics.github.io/BamM/>) avec la commande `-make`. Cette opération permet de convertir les données du séquençage en format BAM « Binary Alignment Map » (Li et al., 2009), que l'on peut traduire par format de cartographie d'alignement binaire. Le format BAM est une représentation binaire, compressée et sans perte, du format SAM « Sequence Alignment Map ». Ce dernier est un format texte conçu pour stocker des séquences biologiques alignées sur une séquence de référence

À l'aide de Samtools v.1.3.1 et de l'option `-sort`, les séquences sont automatiquement classées en fonction de la taille des contigs. Chaque scaffold se voit attribuer un identifiant, d'abord en fonction de leurs tags d'alignements, ensuite en fonction de leurs positions (<http://www.htslib.org/doc/samtools-sort.html>). Cette étape est nécessaire afin de procéder au tri des génomes.

2.2 Tri des génomes

Aperçu du pipeline de binning :

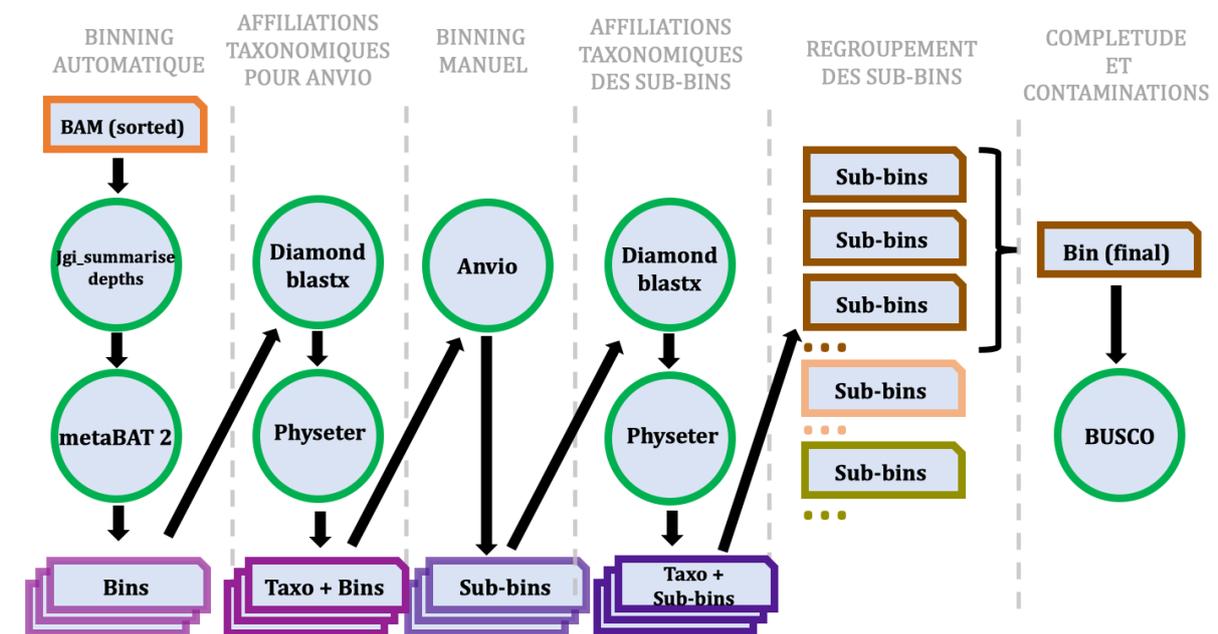


Figure 11: Vue d'ensemble des différentes opérations réalisées pour le binning.

2.2.1 Algorithmes de binning

Actuellement, il n'existe pas, dans la littérature, de pipeline bioinformatique permettant de faire le binning d'un métagénome composé à la fois d'un eucaryote prédominant et de bactéries. Toutefois, pour essayer de réaliser le binning au mieux, différents essais (méthodes et algorithmes) ont été réalisés. Le contrôle des contaminations a guidé la prise de décision concernant les différentes méthodes et algorithmes à adopter en vue de l'élaboration du pipeline de binning.

Deux algorithmes de tri automatiques ont été testés : *Maxbin v 2.0* (Wu et al., 2016) et *MetaBAT v.2.12.1* (Kang et al., 2019). Ils ont pour but de trier automatiquement les séquences de mêmes organismes pour les regrouper dans différents fichiers. Ils génèrent un bin par organisme détecté. Pour ce faire, ils se basent à la fois sur les différentes profondeurs de séquençage, mais aussi sur la fréquence variable en tétra-nucléotides (ou k-mer). Ces dernières diffèrent selon les organismes et permettent donc de procéder au tri automatique.

Les deux programmes ont été testés sur les métagénomés *S25*, *S64* et *S75*. *Maxbin* a moins bien fonctionné pour le tri automatique des 57 métagénomés. Il présentait nettement plus de contaminations que celui de *MetaBAT*. C'est donc ce dernier qui fut sélectionné pour procéder au binning automatique des assemblages.

MetaBAT nécessite, pour fonctionner, de calculer préalablement la profondeur de séquençage moyenne par contig. Ce calcul est réalisé avec *jgi_summarize_bam_contig_depths* (Kang et al., 2019) ainsi que l'option `-outputDepth`.

2.2.2 Vérification du binning via Anvio

Anvio v.6.1 (Eren et al., 2015) est un programme open-source qui permet la visualisation et l'analyse des données omics à l'aide de 3 informations (*Figure 12*): le coverage (symbolisé par les deux bandes noires), le contenu en GC (bande circulaire verte) ainsi que les affiliations taxonomiques des différents contigs. *Anvio v.6.1* a été utilisé afin de rendre compte, à l'aide d'une interface visuelle, de l'état de contamination des bins générées automatiquement par *MetaBAT v.2.12.1*.

2.2.2.1 Affiliations taxonomiques des contigs

Anvio nécessite pour son bon fonctionnement la connaissance des affiliations taxonomiques des contigs présents dans les bins. Il est nécessaire de faire correspondre une taxonomie à chacun des contigs présents dans les bins. Ceci a été réalisé grâce à la succession de deux programmes. D'une part *Diamond BLASTX* et d'autre part *Physeter*.

Il existe un avantage à utiliser *Diamond BLASTX* avec la banque de *Kraken 2* (Wood et al., 2019) par rapport à *Kraken 2* directement. En effet, *Kraken 2* affine la taxonomie grâce à la similitude de k-mers entre les séquences de nucléotides. Or, *Diamond BLASTX* le réalise en comparant les séquences avec une base de données de protéine et ces dernières sont beaucoup plus conservées que les nucléotides au cours de l'évolution. Dès lors, *Diamond BLASTX* permet de retrouver des similitudes pour des organismes, même s'ils sont distants de ceux présents dans la base de données (Boratyn et al., 2012). Ce qui est le cas pour ce travail puisque aucun génome complet de *Peltigera* n'est disponible dans RefSeq. Cependant, la fraction de séquences non classée est plus importante avec *Diamond BLASTX* qu'avec *Kraken*. Ceci est explicable par le fait que seules les régions codantes du génome sont utilisables (Cornet et al., 2018).

2.2.2.2 Diamond BLASTX

BLASTX est un algorithme heuristique visant à aligner la traduction d'une séquence d'ADN en protéine contre une base de données de protéines. Plus spécifiquement, c'est *Diamond BLASTX* v0.9.21 (Bushfink et al., 2015) qui est utilisé, comme préconisé dans Cornet et al. (2018). Ce programme est une variante algorithmique améliorée du *BLASTX*, le rendant plus rapide.

Chacune des bins des 57 métagénomés est « blastée » à l'aide de *Diamond BLASTX* v0.9.21 avec le seuil maximal de e-value fixé à 1e-10 et le nombre de hits maximums à k=30. Les bins sont « blastées » contre la banque de *Kraken 2*. Celle-ci contient la quasi-totalité des séquences de RefSeq.

Les algorithmes de *BLASTX* fonctionnent en comparant les contigs des bins, traduites selon les six cadres de lecture possibles, à une base de données protéique. En effet, puisqu'un codon est constitué de trois nucléotides successifs, il y a donc trois façons distinctes pour diviser une séquence en codons. Comme l'ADN est bicaténaire (composé de deux chaînes), cela double le nombre possible de cadres de lecture à prendre en compte. Le code génétique est redondant, il est donc fréquent que deux séquences différentes codent pour la même protéine. De plus, il y a seulement 4 nucléotides pour 20 acides aminés. Ces deux facteurs rendent le *BLAST* contre des protéines plus spécifique et précis que le *BLAST* contre des nucléotides. Ceci car les séquences sont plus susceptibles de varier que les protéines. Cependant, cette méthode comporte un désavantage, elle ne fonctionne que pour les séquences codantes.

2.2.2.3 Physeter

Physeter (Balthazar et al., in prep.) est un parseur semblable à *MEGAN* (Huson et al., 2007). Il se base sur un algorithme de type LCA «Last Common Ancestor». Ce programme prend en entrée la taxonomie de chaque contig provenant du *Diamond BLASTX* sur l'ensemble des bins des 57 métagénomés afin d'en déterminer le dernier ancêtre commun ou LCA. Ce dernier associe donc à chacun des contigs, une taxonomie. Les affiliations taxonomiques des contigs seront nécessaires à *Anvio* v6.1 ci-après.

2.2.2.4 Binning manuel via Anvio

Anvio v.6.1 permet non seulement de passer les bins en revue afin d'en estimer la contamination mais également de sub-diviser celles-ci, dans le but de retirer manuellement les contaminations observées. Ainsi les bins soupçonnées de ne pas contenir uniquement les séquences d'un seul et même organisme ont été divisées manuellement en sous-ensembles (appelés sub-bins). Ces sub-bins ont été divisées en essayant de maximiser l'homogénéité du coverage et du contenu en GC, le tout en tenant compte des affiliations taxonomiques quand celles-ci sont connues comme le montre la capture d'écran (*Figure 12*). Au total, 380 bins provenant des 57 métagénomés et générées automatiquement par *MetaBAT* ont été passées en revue grâce à *Anvio* et divisées en sub-bins lorsque nécessaire. Un fichier PDF reprenant l'ensemble des bins qui ont été retravaillées pour le binning est disponible en ligne (<https://github.com/balthazar-alexis/memoire>).

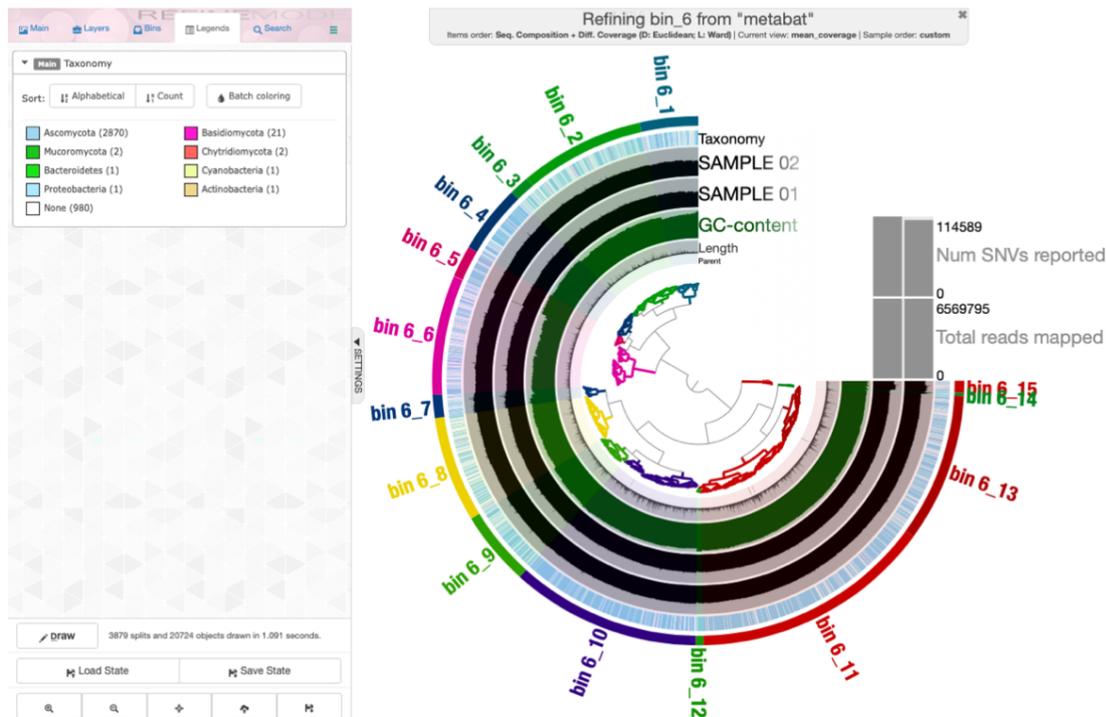


Figure 12: Interface de visualisation des contigs grâce à *Anvio*. Dans cet exemple, il s'agit de la bin n°6 générée par *MétabAT2*. Celle-ci contenant, entre autre, un grand nombre de contigs du mycobionte de l'échantillon *S18 - Peltigera vitikainenii* ad. int.. L'épaisseur de la bande circulaire verte informe sur la proportion de GC dans chaque contig. Les deux bandes noires au-dessus correspondent au coverage. La bande annotée « taxonomy » contient pour chacun des contigs, l'affiliation taxonomique de celui-ci (généré grâce à *Diamond BLASTX* et *Physeter*). La légende en haut à droite indique la taxonomie des contigs (Phylum). La bande circulaire la plus à l'extérieur indique la sub-division manuelle effectuée sur base des différents paramètres cités ci-dessus.

2.2.3 Fusion des sub-bins

Le but est de rassembler les sub-bins contenant les séquences génétiques d'un seul et même organisme. Dans le cadre de ce travail, seuls les génomes du mycobionte seront rassemblés pour former 57 assemblages ne contenant plus que les génomes du mycobionte *Peltigera*.

2.2.3.1 Affiliation taxonomique des sub-bins via Kraken.

Kraken v.2 est un programme, qui permet de donner une affiliation taxonomique aux séquences d'ADN métagénomiques (Wood et al., 2014). Pour les séquences métagénomiques analysées, celui-ci détermine le pourcentage de séquences affiliées aux différents rangs taxonomiques, depuis le Règne, jusqu'à l'espèce.

Après utilisation de *Kraken 2* sur les 4059 sub-bins générées via *Anvio*, la taxonomie de chacune des sub-bins n'était pas assez précise pour permettre de rassembler les sub-bins provenant uniquement du champignon principal de *Peltigera*. Les affiliations taxonomiques des sub-bins descendaient jusqu'à la sous-division des Pezizomycotina. Bien que *Peltigera* fasse partie de cette sous-division, cette classification restait trop vague pour permettre de différencier les sub-bins contenant les séquences du mycobionte de *Peltigera* et celles provenant d'éventuelles contaminations comme *Aspergillus*, faisant partie des contaminations fréquentes. Ce dernier fait

partie de la classe des Eurotiomycetes, alors que *Peltigera* est inclus dans la classe des Lecanoromycetes. Une méthode d'affiliation taxonomique des sub-bins allant minimum jusqu'à la classe était dès lors nécessaire pour éviter autant que possible de rassembler des sub-bins ne provenant pas du même organisme.

2.2.3.2 Affiliation taxonomique des sub-bins via *Diamond BLASTX* et *Physeter*.

L'utilisation de *Kraken 2* donnant des affiliations taxonomiques trop superficielles pour permettre le binning, il était nécessaire d'ajouter des génomes de Lecanoromycetes à la banque de *Kraken 2*. Cependant, l'ajout de nouvelles séquences à cette banque de données n'était pas réalisable pour des raisons de format de fichiers. Il a dès lors été nécessaire de créer une nouvelle banque de données. Cette dernière fut élaborée à partir de tous les génomes d'Ascomycètes répertoriés dans RefSeq ainsi que deux génomes de Lecanoromycetes présents dans NCBI.

Une taxonomie a été attribuée aux sub-bins grâce à la combinaison, précédemment décrite, de *Diamond BLASTX* suivie de *Physeter*, couplées cette fois avec la nouvelle banque de données contenant les Lecanoromycetes. Grâce à cette dernière, une classification plus fine, allant jusqu'à la classe a pu être établie.

2.2.3.3 Regroupement des sub-bins

Les sub-bins ayant reçu pour affiliation taxonomique Lecanoromycetes ont été groupées. Ceci est réalisé pour chacun des 57 métagénomes de départ. Cette opération a permis de rassembler toutes les sub-bins contenant les séquences du champignon principal de *Peltigera* dans un seul assemblage génomique final.

2.2.3.4 Complétude et contamination des assemblages génomiques finaux.

BUSCO v3.0.1 (Seppey et al., 2019) permet d'estimer la complétude d'un génome ainsi que le degré de contamination de celui-ci. Ce programme se base sur la reconnaissance de gènes orthologues extrêmement bien conservés au cours de l'évolution. Les gènes orthologues sont d'abord homologues, c'est-à-dire qu'ils descendent d'un même gène ancestral. De plus, leur divergence doit provenir d'une spéciation et non d'une duplication ancestrale, auquel cas ils seraient qualifiés de paralogues.

Pour fonctionner, *BUSCO* se base sur une série de gènes orthologues présents en une seule copie au sein du génome. Ils ont été sélectionnés car ils ont été extrêmement conservés au cours de l'évolution des lignées. Grâce à cela, le pourcentage de ces gènes orthologues retrouvés dans notre génome d'intérêt est directement proportionnel à la complétude de celui-ci. De plus, *BUSCO* peut témoigner de l'éventuelle contamination du génome d'intérêt. En effet, si celui-ci repère plus d'une copie d'un gène orthologue — censé n'être présent qu'en simple copie dans le génome — c'est que ce gène provient d'un autre organisme. Le nombre de gènes présents en double copie atteste donc du degré de contamination de notre génome.

Ascomycota odb9 (Seppey et al., 2019) est la base de données utilisée par *BUSCO* pour attester de la complétude et des contaminations de nos 57 génomes. Elle contient 1315 gènes orthologues conservés et communs à l'embranchement des Ascomycètes. *BUSCO* retourne alors le nombre de gènes présents en simple et double copie pour chacun de nos 57 génomes.

3. Génétique des populations

3.1 SNP calling

Aperçu du pipeline SNP :

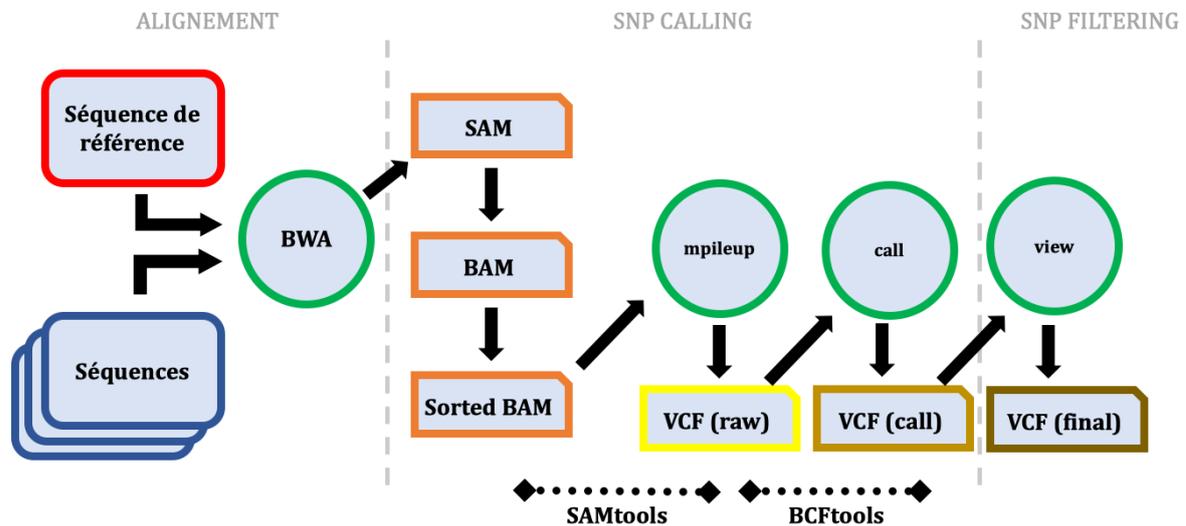


Figure 13: Vue d'ensemble des différentes opérations réalisées pour le SNP calling.

3.1.1 Sélection des génomes

La sélection des génomes pour les analyses de génomique des populations a été effectuée sur base de deux critères grâce aux résultats de BUSCO v 3.0.1 (Tableau 2):

1. La complétude : Nous l'avons fixée à 40% minimum. Cette valeur seuil a été retenue car elle reflète un compromis entre complétude minimum et nombre d'échantillons minimum. Les génomes ayant une valeur de complétude inférieure à la valeur seuil ci-dessus, ne seront pas sélectionnés. (Tableau 2, cases rouges)
2. Seuls les génomes de *Peltigera vitikainenii* et *Peltigera appalachensis* sont sélectionnés. Ceux de *Peltigera pacifica*, *Peltigera asiatica*, *Peltigera mikado* et *Peltigera gayae* (Tableau 2, cases jaunes) ne feront donc pas partie des analyses ultérieures dans le cadre de ce mémoire.

Parmi les 57 génomes, 33 ont été sélectionnés : 14 *Peltigera vitikainenii* et 19 *Peltigera appalachensis* (Tableau 2, cases vertes).

Tableau 2: Tableau récapitulatif incluant la complétude et la contamination obtenues avec *BUSCO v 3.0.1* pour les 57 génomes de *Peltigera*. Les lignes en vert représentent les assemblages génomiques qui ont été conservés pour les analyses de génétique des populations. Les lignes rouges font référence aux assemblages qui n'ont pas été repris sur base de leur faible complétude (<40%). Les lignes jaunes représentent les assemblages non traités car ils ne concernent ni *Peltigera appalachensis*, ni *Peltigera vitikainenii*.

N°	Espèce	Origine géographique	Date de collecte	Complétude	Contamination
S14	<i>Peltigera pacifica</i>	Canada:British Columbia	2012	71.90%	0.10%
S15	<i>Peltigera pacifica</i>	USA:Oregon	2005	82.40%	0.40%
S16	<i>Peltigera pacifica</i>	Canada:British Columbia	1994	20.70%	0.00%
S17	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	81.80%	0.10%
S18	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	79.40%	0.20%
S19	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	79.30%	0.20%
S20	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	84.40%	0.30%
S21	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	36.40%	0.00%
S22	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	65.70%	0.20%
S23	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	78.40%	0.20%
S24	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	80.10%	0.20%
S25	<i>Peltigera vitikainenii</i>	Japan:Honshu	2017	0.00%	0.00%
S26	<i>Peltigera vitikainenii</i>	Japan:Honshu	2017	11.90%	0.00%
S27	<i>Peltigera vitikainenii</i>	Russia:Khabarovsk	2018	18.10%	0.10%
S28	<i>Peltigera vitikainenii</i>	Russia:Sakhalin	2013	14.10%	0.00%
S29	<i>Peltigera vitikainenii</i>	France: Vosges	2019	85.60%	0.40%
S30	<i>Peltigera vitikainenii</i>	Suisse	1998	11.70%	0.00%
S31	<i>Peltigera vitikainenii</i>	Canada:British Columbia	inconnu	80.80%	0.20%
S32	<i>Peltigera vitikainenii</i>	Canada:British Columbia	inconnu	81.50%	0.30%
S33	<i>Peltigera vitikainenii</i>	USA:Alaska	2013	45.00%	0.10%
S34	<i>Peltigera vitikainenii</i>	Estonia	2017	2.40%	0.00%
S35	<i>Peltigera vitikainenii</i>	Norway	2011	51.60%	0.00%
S36	<i>Peltigera vitikainenii</i>	Norway	2011	23.80%	0.00%
S37	<i>Peltigera vitikainenii</i>	Norway	2011	56.60%	0.20%
S38	<i>Peltigera vitikainenii</i>	Norway	2011	66.20%	0.20%
S39	<i>Peltigera appalachensis</i>	USA:Vermont	2019	72.50%	0.20%
S40	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	79.10%	0.50%
S41	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	77.40%	0.40%
S42	<i>Peltigera appalachensis</i>	USA:Vermont	2019	67.70%	0.20%
S43	<i>Peltigera appalachensis</i>	USA: Alabama	2019	80.00%	0.20%
S44	<i>Peltigera appalachensis</i>	USA: Missouri	2019	83.30%	0.40%
S45	<i>Peltigera appalachensis</i>	USA: Connecticut	2019	68.50%	0.20%
S46	<i>Peltigera appalachensis</i>	USA: Alabama	2019	73.40%	0.40%
S47	<i>Peltigera appalachensis</i>	USA: South Carolina	2019	63.30%	0.00%
S48	<i>Peltigera appalachensis</i>	USA: Arkansas	2019	77.30%	0.20%
S49	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	58.80%	0.00%
S50	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	67.60%	0.10%
S51	<i>Peltigera appalachensis</i>	Canada: Quebec	2011	0.00%	0.00%
S52	<i>Peltigera appalachensis</i>	Canada: Quebec	2011	49.90%	0.20%
S53	<i>Peltigera appalachensis</i>	Canada: Alberta	inconnu	71.40%	0.20%
S54	<i>Peltigera appalachensis</i>	Canada: Alberta	inconnu	53.40%	0.20%
S55	<i>Peltigera appalachensis</i>	Canada: British Columbia	inconnu	82.60%	0.20%
S56	<i>Peltigera appalachensis</i>	Russia: Sakha	2011	3.10%	0.00%
S57	<i>Peltigera appalachensis</i>	Russia: Komi	2019	77.40%	0.20%
S58	<i>Peltigera appalachensis</i>	Norway	2011	3.20%	0.00%
S59	<i>Peltigera appalachensis</i>	USA: Alaska	2006	7.20%	0.00%
S60	<i>Peltigera appalachensis</i>	USA: Alaska	2011	10.70%	0.10%
S61	<i>Peltigera asiatica</i>	Malaysia: Borneo	2018	80.70%	0.20%
S62	<i>Peltigera asiatica</i>	China: Yunnan	2003	13.50%	0.00%
S63	<i>Peltigera asiatica</i>	Taiwan	2016	0.10%	0.00%
S64	<i>Peltigera mikado</i>	Russia:Khabarovsk	2013	74.60%	0.20%
S65	<i>Peltigera mikado</i>	China: Yunnan	2010	36.30%	0.20%
S66	<i>Peltigera mikado</i>	Taiwan	2016	18.90%	0.00%
S67	<i>Peltigera gayae</i>	Peru	2011	78.10%	0.20%
S68	<i>Peltigera gayae</i>	Puerto Rico	2018	48.70%	0.20%
S91	<i>Peltigera appalachensis</i>	USA:South Carolina	2019	66.50%	0.10%
S93	<i>Peltigera appalachensis</i>	USA:West Virginia	2019	78.90%	0.10%

3.1.2 Alignement via BWA

BWA « Burrows-Wheeler Alignment Tool », est un package qui permet le mapping de séquences proches entre elles (Li et al., 2009). Il nécessite un génome de référence contre lequel effectuer le mapping. Comme aucun génome de *Peltigera* n'est disponible à l'heure actuelle, l'assemblage *S15* (*P. pacifica*) est choisi pour jouer le rôle de séquence de référence sur laquelle viennent s'aligner les autres. La séquence de référence est tout d'abord indexée grâce à la commande *-index*, comprise dans le package *BWA*. Ensuite, à l'aide de la commande *-aln*, les différentes séquences sont alignées contre la séquence de référence. Enfin, c'est grâce à l'option *-sampe* qu'un alignement au format SAM est généré.

3.1.3 SNP calling

SAMtools, à l'aide de la commande *view*, permet de transformer l'alignement du format SAM au format BAM. Ensuite, la commande *--sort* ordonne le fichier BAM. La commande *--faidx* indexe une nouvelle fois notre génome de référence. La commande *--mpileup* permet ensuite le réarrangement des données génomique en un tableau, contenant les informations de positions base par base, préparant ainsi le SNP calling.

BCFtools, commande *--call*, permet d'appliquer les options choisies et de faire le SNP calling. C'est ensuite grâce à la commande *--view* de *BCFtools*, que seuls les sites ne comportant aucune donnée manquante sont gardés. C'est alors que le format VCF « Variant Calling Format » final est généré (Figure 14).

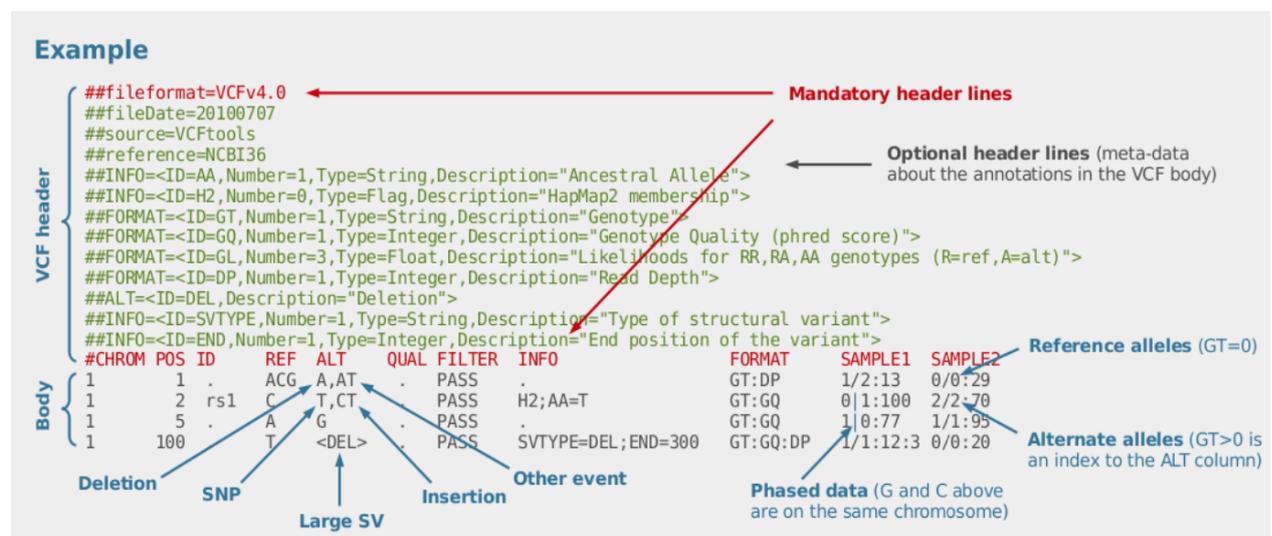


Figure 14: Illustration du format VCF "Variant Calling Format". Image annotée provenant du site <http://digitheadslabnotebook.blogspot.com/2013/01/vcf-variant-call-format.html> à partir de Danecek et al., (2011).

3.2 Analyses de génétique des populations sur R

Les analyses de génétique des populations à partir des fichiers VCF ont été réalisées avec R v 4.0.2 (R Core Team, 2020), via l'interface *R studio v 1.1.463* (RStudio Team, 2020).

Le package *pegas* (Paradis, 2010) a permis la lecture et la manipulation des données au format VCF. Ensuite, c'est grâce au package *adegenet* (Jombart, 2008) que le clustering hiérarchique des 33 génomes a été réalisé. Un dendrogramme a été élaboré sur base des distances génétiques entre chacun des 33 individus deux par deux, estimées grâce à la méthode de Pervosti (Pervosti et al., 1975). Le package *adegenet* a également permis le calcul des probabilités postérieures d'appartenance à un groupe d'après une DAPC « Discriminant Analysis of Principal Components » sur les 33 individus, comme préconisé dans Paradis et al. (2017). Avec nombre de clusters $K=5$, comme suggéré sur base du BIC « Bayesian Information Criterion ». Le découpage en $K=2$ clusters a également été réalisé avec pour objectif de vérifier la séparation de *P. vitikainenii* et *P. appalachensis* en deux cluster distinct.

Ensuite, une analyse en composantes principales (Principal Component Analysis, PCoA) a été réalisée sur les données SNP grâce à la fonction *glPCA* du package *adegenet*.

Finalement, les échantillons ont été divisés en quatre populations (*P. appalachensis* de la zone tempérée, *P. appalachensis* de la zone boréale, *P. vitikainenii* de la zone Japon-Est Russie, et autres *P. vitikainenii*) et des mesures de diversité génétique ont été réalisées pour chaque population grâce à la fonction *get.diversity* du package *PopGenome* (Pfeifer et al. 2014).

RÉSULTATS

1. Récolte des données

1.1 Mission de terrain et séquençage ITS

Peltigera appalachensis a été récolté avec succès dans 18 localités réparties dans 8 états américains (Vermont, Connecticut, Virginie Occidentale, Caroline du Nord, Caroline du Sud, Alabama, Arkansas, Missouri). L'espèce n'a pas été récoltée dans deux états pour lesquels nous n'avons échantillonné qu'une seule localité: la Pennsylvanie et le Kentucky. Deux autres espèces de la section Polydactylon ont été récoltées lors de la mission: *Peltigera polydactylon* ssp. *udeghe* (Vermont) et *P. occidentalis* (Caroline du Nord). Il s'agit de la première confirmation de la présence de *Peltigera polydactylon* ssp. *udeghe* au Vermont.

L'identité de *P. appalachensis* a été confirmée grâce au séquençage ITS de 48 thalles répartis dans ces 18 localités.

1.2 Réseaux d'haplotypes ITS.

Sur base des données ITS, 12 haplotypes ITS ont été détectés pour *P. appalachensis*. Parmi ceux-ci, 8 proviennent exclusivement de la zone tempérée nord-américaine, et 4 de l'ensemble de la zone boréale (*Figure 15*). Deux haplotypes sont partagés entre les Appalaches et les Ozark. Il ne semble donc pas y avoir de structuration génétique entre ces deux zones, ni entre la partie plus au nord des Appalaches (Vermont, Connecticut, Virginie occidentale) et la partie plus au sud (Caroline du Nord, Caroline du Sud, Alabama). Aucun haplotype n'est partagé entre la zone tempérée et la zone boréale. Au sein de la zone boréale, il ne semble pas y avoir de structuration géographique, chaque haplotype ayant été trouvé sur au moins deux continents.

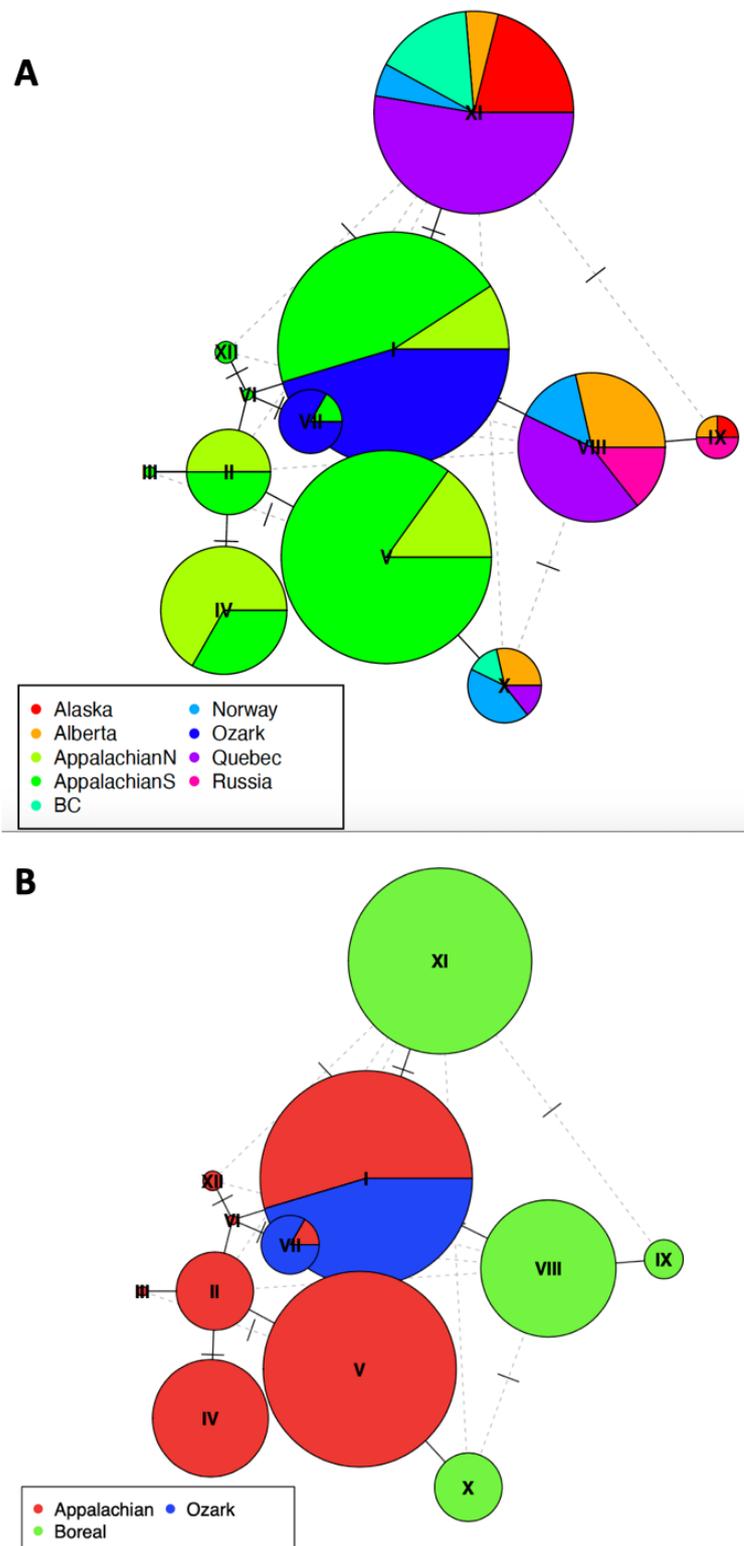


Figure 15: Réseaux d’haplotypes ITS de *P. appalachensis*. 9 régions sont représentées dans (A), tandis que dans (B), le sud (AppalachianS) et le nord (AppalachianN) des Appalaches ont été regroupés; et la Norvège, le Québec, la Russie, l’Alberta, l’Alaska et la Colombie Britannique (BC) ont été regroupés en “Boreal”.

16 haplotypes ont été détectés chez *P. vitikainenii* (Figure 16), mais 7 de ces haplotypes n’ont été détectés qu’une seule fois. 12 haplotypes n’ont été détecté que dans l’est asiatique (Japon et Est de la Russie), alors que la majorité des échantillons hors de cette zone partagent le même haplotype, aussi détecté au Japon (haplotype IV). Un échantillon d’Alaska est identique à un échantillon fréquemment trouvé au Japon et dans l’est de la Russie (haplotype I). Enfin, un haplotype (II) a été détecté au Montana (USA) et en Colombie Britannique (Canada).

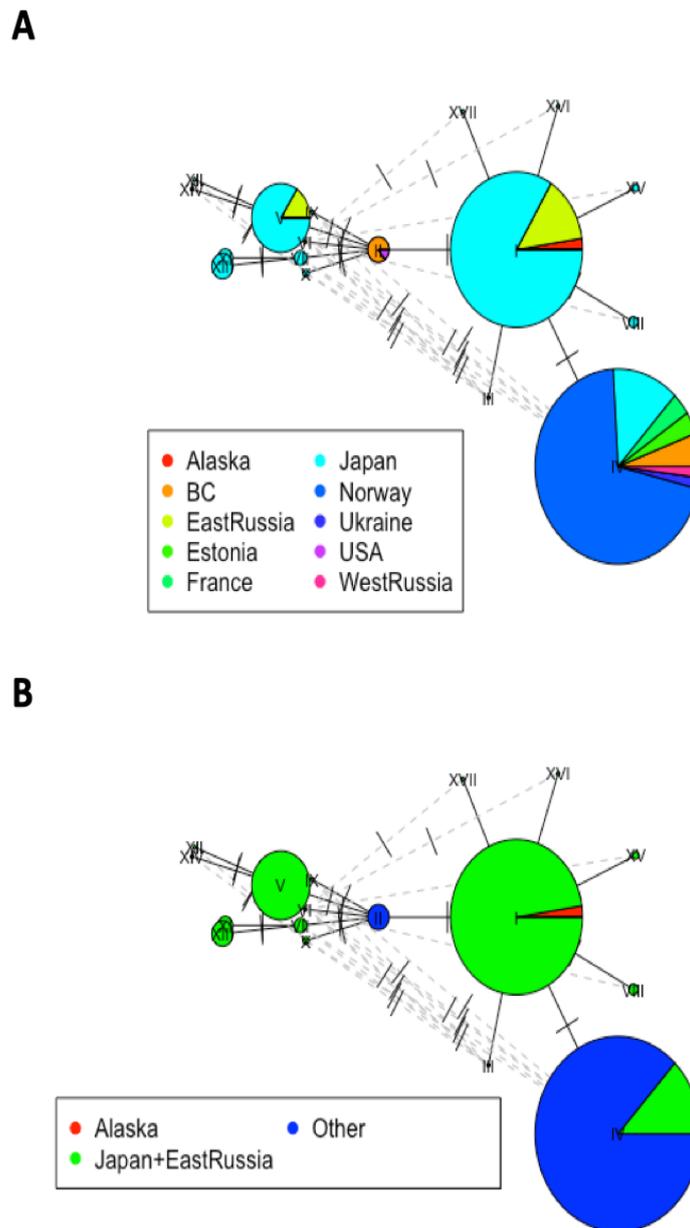


Figure 16: Réseaux d’haplotypes ITS de *P. vitikainenii*. 10 régions géographiques sont représentées dans (A) tandis que Japon et East Russia sont regroupés; et BC (Colombie Britannique), Estonia, France, Norway, Ukraine, USA et West Russia sont regroupés en tant que “other” dans (B).

1.3 Séquençage

Le séquençage NovaSeq 6000 de la plateforme Illumina (États-Unis, San Diego, CA 92122) a produit des reads paired-end de 151pb avec des tailles d'inserts de ~500pb. L'ensemble des reads présentaient un excellent score de qualité par base. Et aucune séquence n'a été signalée comme étant de mauvaise qualité par *FastQC*.

Le séquençage de chacun des 57 thalles de lichen a produit, en moyenne, 18.5 millions de reads ce qui représente 2.8 milliards de paires de bases (*Tableau 3*). Ces valeurs sont relativement proches les unes des autres. Si on ne considère pas *Peltigera gayae* S69 pour lequel la préparation de la librairie a échoué, il n'y a pas d'échantillons pour lesquelles le séquençage a produit beaucoup moins de données.

La N50 est une mesure commune pour évaluer la continuité d'un assemblage. La valeur donnée par celle-ci représente la taille du scaffold ou du contig tel que 50% des bases de l'assemblage sont comprises dans des scaffolds de tailles supérieures ou égales à cette taille. Elle est de 3338 en moyenne (*Tableau 3*). Cependant, elle varie fortement selon les assemblages. Les gammes de valeurs se situent entre 974 pour l'assemblage *S34* et 13088 pour le *S67*.

Tableau 3 : Tableau reprenant différentes statistiques concernant les données génomiques avant et après l'assemblage. Les données dans les colonnes marquées d'un astérisque ne tiennent compte que des contigs ayant une taille supérieure ou égale à 500pb. pb= paires de bases.

N°	AVANT ASSEMBLAGE			APRÈS ASSEMBLAGE					
	Reads	Taille (pb)	contigs (>= 0 pb)	Taille (pb) (>= 0 pb)	contigs (>=500pb)	Le plus long contig*	Taille (pb)*	N50*	L50*
S14	17292950	2611235450	538366	266553095	80139	319522	152536656	3312	10969
S15	13392292	2022236092	340127	163549623	50877	100488	101443537	3313	7577
S16	17855316	2696152716	520539	226957661	89726	239653	119896749	1572	19960
S17	18379544	2775311144	333593	177080287	49072	148320	112335949	4374	6208
S18	18286212	2761218012	648262	269216678	63314	231995	127870686	3654	7971
S19	18258062	2756967362	472188	211941258	54968	205652	112331956	3608	7329
S20	20053588	3028091788	236512	159022908	35926	410277	116480101	7197	4012
S21	16238260	2451977260	695768	258380469	79220	154926	103519543	1610	16143
S22	16843702	2543399002	697777	275235732	75088	380409	123656677	2256	12412
S23	24113088	3641076288	980858	391380840	93817	255496	163862990	3496	9383
S24	17285286	2610078186	464958	219986688	62268	205861	121714025	3842	7269
S25	16320550	2464403050	544523	190560952	61073	155086	64027672	997	13161
S26	18508104	2794723704	576512	252141458	98497	197813	129565730	1511	18669
S27	17878802	2699699102	894193	321979135	99395	167291	119034095	1338	22159
S28	18242706	2754648606	461332	192835534	85655	198067	97161283	1219	20347
S29	21709148	3278081348	570819	259931112	66269	237686	136483171	3728	8269
S30	17949536	2710379936	345717	174656533	76468	172077	107591650	1697	15392
S31	19245346	2906047246	360640	198315319	65345	183490	126632210	3427	8403
S32	23732828	3583657028	490343	243401165	69912	189078	139018845	3878	8194
S33	20118348	3037870548	443694	179805801	58463	160710	88967238	1943	11285
S34	21133130	3191102630	1028015	329683702	89298	249641	87568591	974	23823
S35	16557242	2500143542	397427	203420144	74308	237711	124782826	2367	12916
S36	22680280	3424722280	374034	177866001	65849	60382	104454099	2186	12928
S37	17280136	2609300536	388754	198050885	57822	172289	118877265	3239	8618
S38	20423506	3083949406	485495	261268486	87860	171782	161226816	2849	13200
S39	20347094	3072411194	602874	257676035	53431	188015	124812772	4519	6829
S40	21327242	3220413542	441875	235803453	48454	259470	138351449	9449	3424
S41	17618302	2660363602	215015	150998488	26070	106532	111209576	11467	2645
S42	17717862	2675397162	638046	270672677	78382	114849	130978259	2388	13706
S43	19256216	2907688616	710661	283939852	55491	189649	122733763	4150	7260
S44	20333622	3070376922	637474	304349280	66329	277730	154420434	5382	6356
S45	17472868	2638403068	518848	223267866	58254	179357	110145720	2757	9962
S46	19635110	2964901610	688323	280197418	69366	227855	126800611	2851	10632
S47	17732902	2677668202	605332	258200687	77311	101906	125257152	2390	12933
S48	16699298	2521593998	446803	220014555	38032	238004	120736603	6746	4533
S49	21505688	3247358888	960630	345708661	74907	141858	124525089	2478	12201
S50	21267494	3211391594	966945	322441592	59832	30239	100952501	2465	11269
S51	16354916	2469592316	621143	253310597	112045	129301	116769939	1057	28098
S52	18077768	2729742968	588149	224381883	58994	201506	96602578	2239	10844
S53	15187740	2293348740	329696	157683812	46165	173586	94139316	3096	7448
S54	11357378	1714964078	378159	171180955	68173	165334	92357260	1676	12458
S55	16119350	2434021850	363480	181178969	56598	214269	109119286	3352	7334
S56	19520554	2947603654	772532	261088260	84309	119586	83739830	995	22568
S57	18677768	2820342968	437865	228939583	67859	192448	139511209	3866	8196
S58	23474956	3544718356	979951	306219337	76830	148174	79708878	1064	19796
S59	19183442	2896699742	697096	283473318	107559	171321	130368449	1346	22353
S60	18946902	2860982202	703994	277353574	114474	226794	123330842	1140	29229
S61	18805816	2839678216	429358	244554236	49789	195402	150004074	9902	3716
S62	17895574	2702231674	519303	216410361	90239	289224	112007692	1405	21638
S63	14841090	2241004590	395930	154672457	73442	94453	74521131	1007	19634
S64	18450386	2786008286	379544	189284698	57679	178127	111957955	3344	7778
S65	21500132	3246519932	612803	244808810	69856	239754	108187443	2088	13119
S66	16363616	2470906016	754200	247349878	63683	127738	80212274	1407	14653
S67	15480812	2337602612	321982	186184038	30542	143912	112842708	13088	2146
S68	16912226	2553746126	556617	235637677	79738	139146	116662718	1886	15760
S91	17766408	2682727608	617112	288832889	93387	121640	153443339	2535	14049
S93	18588672	2806889472	463477	227213275	47047	150315	128169986	7150	4355
MOYENNE	18494687	2792697756	555187.1	237478081	69208.7	187424.4912	116940723	3338.1	12167
ÉCART-TYPE	2424510.4	366101063	193969.9	52763386.9	19296.72	68732.91421	21443683.4	2598.8	6427

2. Traitement bioinformatique des données

2.1 Tri des génomes

MetaBAT 2 a généré 6.6 bins en moyenne pour chacun des 57 métagénomes, pour un total de 380 bins. *MétabAT* a employé entre 2% (S25 ; S58) et 9% (S30) des contigs contenu dans l'assemblage métagénomique généré par metaSPAdes. C'est pour l'assemblage S67 que l'on observe le moins de bins métabat avec seulement 3 bins. À l'inverse, le maximum est atteint pour l'assemblage S38 ou l'on obtient 16 bins.

Les 380 bins générées par *MetaBAT 2* ont été scindées en 4059 sub-bins grâce à *Anvio* (Tableau 4). En moyenne, il a fallu rassembler 22 sub-bins pour former chacun des assemblages contenant le champignon principal de *Peltigera*. Une fois rassemblés, ces assemblages contiennent en moyenne 40% des nucléotides qui étaient présents sur l'ensemble des bins de *MetaBAT*.

Tableau 4: Tableau récapitulatif des bins et sub-bins

N°	Pourcentage de contigs de départ utilisés par metabat	Bins	sub-bins	Sub-bins rassemblées (Mycobionte <i>Peltigera</i>)	Somme des nucléotides sur l'ensemble des bins	Somme des nucléotides des sub-bins rassemblées (Mycobionte <i>Peltigera</i>)
S14	7	8	65	11	78595985	27580030
S15	7	4	69	13	56315212	28186313
S16	7	7	157	47	34075666	9784927
S17	7	9	90	37	70100166	33297872
S18	4	10	81	18	74611916	30170518
S19	5	5	58	14	64932604	30941621
S20	8	7	125	34	84424783	33137329
S21	4	4	131	55	32463793	15819993
S22	5	7	138	39	55888776	25870055
S23	3	9	107	19	88114190	27559233
S24	5	6	83	31	70378194	29315808
S25	2	7	42	1	15209519	90384
S26	6	11	91	18	40313893	6319166
S27	4	8	94	20	29822048	8612882
S28	6	4	49	20	11347340	7383682
S29	5	9	87	23	78843353	33784507
S30	9	10	69	21	32782576	9037944
S31	8	7	80	21	69795671	31977773
S32	6	10	94	18	80266733	32491868
S33	6	5	46	22	32462444	18868191
S34	2	5	31	13	11352248	1652889
S35	9	8	67	22	55610756	24538409
S36	8	6	92	30	42830126	14506175
S37	8	8	86	19	67358068	27629903
S38	8	16	140	24	83348494	29427466
S39	4	7	78	29	72726354	28805745
S40	4	9	38	14	98473712	29727790
S41	6	5	45	11	84464456	31352881
S42	6	7	99	24	61124663	26652381
S43	3	4	72	23	77950825	33161823
S44	4	11	97	37	102671883	38605330
S45	6	6	63	28	56645979	26237226
S46	4	7	82	23	65608966	29045482
S47	5	4	49	21	57490493	24373701
S48	4	7	61	15	89996126	29105405
S49	3	6	72	34	59355038	27886237
S50	3	4	40	13	48714787	26167063
S51	5	8	71	3	20516184	92941
S52	5	5	37	15	42292999	20318344
S53	7	5	54	21	53065285	27765574
S54	7	4	45	13	33207815	19930104
S55	7	5	51	13	59768222	29893325
S56	2	4	25	9	9697797	1274894
S57	7	8	82	15	81886744	29483471
S58	2	4	33	11	11469958	2211666
S59	5	8	67	9	26241622	3551639
S60	5	7	59	13	20470896	5336343
S61	5	5	49	18	110768909	32395718
S62	7	6	45	14	24834982	8299222
S63	5	4	28	28	10148068	293684
S64	7	5	51	17	61146642	31932362
S65	5	4	54	28	34892508	18197301
S66	3	5	61	27	18499892	7810103
S67	4	3	43	43	88777671	33169765
S68	6	7	76	76	42674089	22164217
S91	6	8	84	15	75357193	23590872
S93	4	8	76	12	88753265	29626544
SOMME	305	380	4059	1262	3126103595	1236444091
MOYENNE	5.35	6.67	71.21	22.14	55823278.48	21692001.6

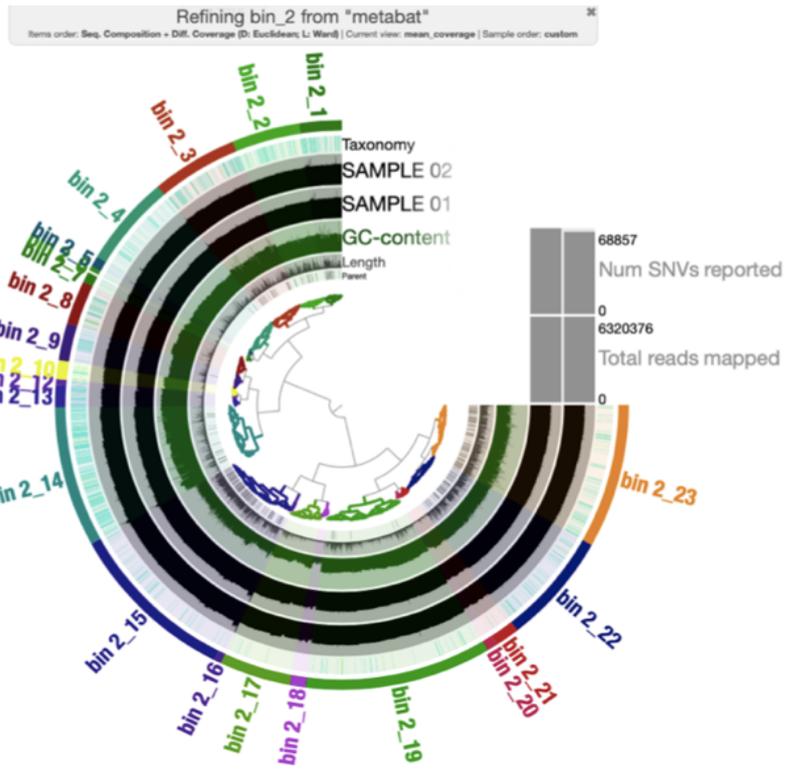
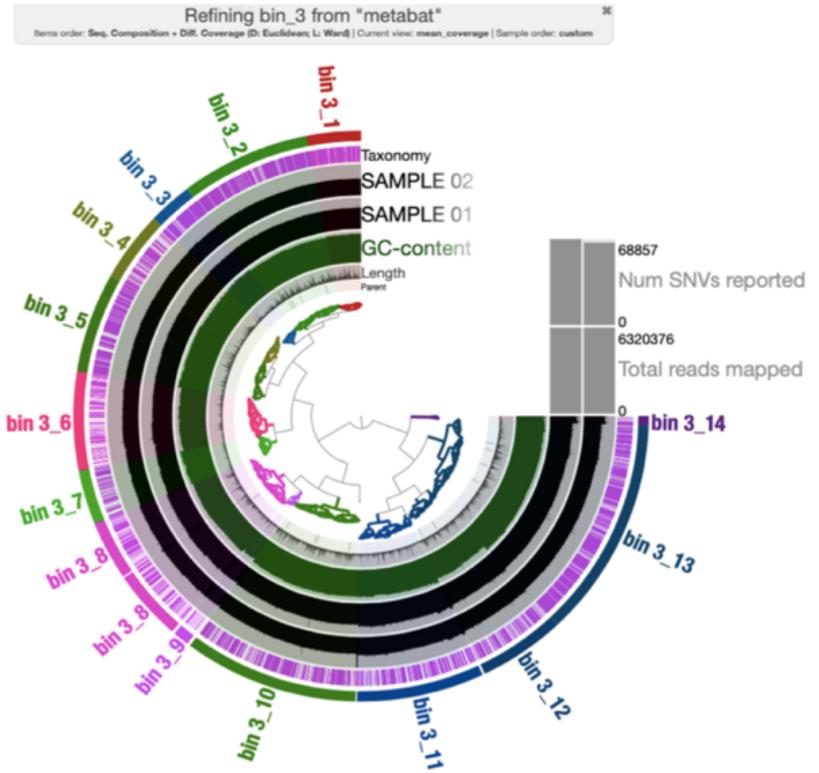
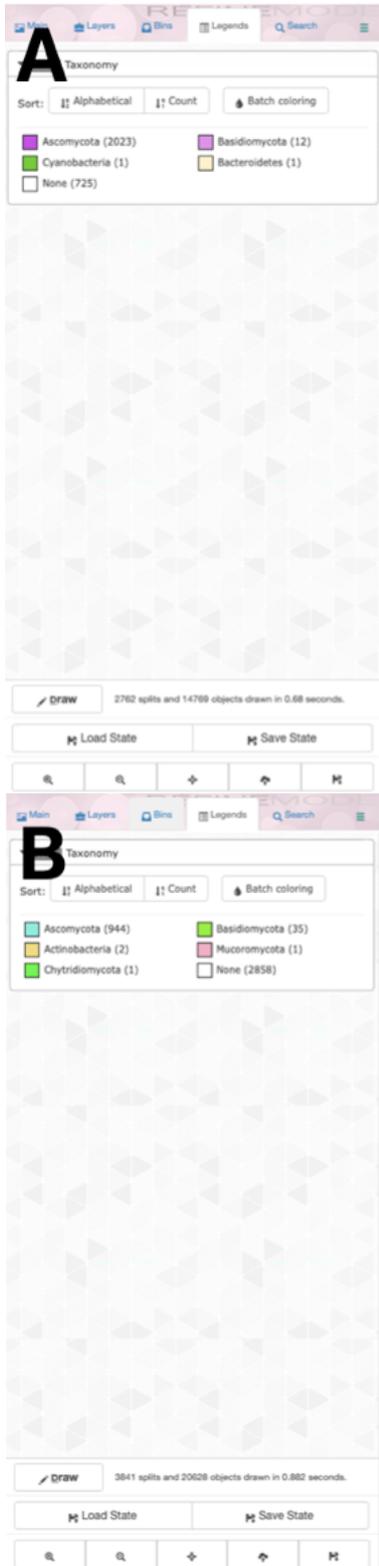
Le détail des 380 bins est disponible à l'adresse <https://github.com/balthazar-alexis/memoire>, ainsi que la sélection des sub-bins à regrouper pour former l'assemblage final (parmi les 4059 sub-bins). Seules les bins de l'assemblage *S67* seront détaillées dans cette section. Ce dernier est assez représentatif des regroupements effectués automatiquement par *MétabAT 2*. Les 3 bins du *S67* rendent assez bien compte du degré général des contaminations que présentent en moyenne les bins Ascomycète et Cyanobactéries des 57 assemblages metagénomiques.

Comme c'est le cas pour la majorité des 57 assemblages, on observe plus d'un bin contenant des contigs d'Ascomycètes, deux dans le cas du *S67* (*Figure 17a et 17b*), ainsi qu'un bin de Cyanobactérie (*Figure 17c*)

La première bin concernant le nombre de contigs d'Ascomycètes (*Figure 17a*) possède un coverage ainsi qu'un contenu en CG relativement homogènes. Le découpage en sub-bins a été jugé nécessaire par endroits afin d'éviter au maximum la présence de contamination dans l'assemblage final. Seul la sub-bin 3_14 a été écartée de l'assemblage final, les 13 autres ont été rassemblées.

Concernant la deuxième bin Ascomycètes (*Figure 17b*), le coverage ainsi que le contenu en GC ne sont absolument pas homogènes sur l'ensemble de la bin avec d'importantes variations ainsi que quelques pics par endroit. La bin est légèrement contaminée (34 contigs identifiés comme appartenant aux Basidiomycètes) et est composée de 3/4 de contigs ne possédant pas d'affiliation taxonomique. Une division manuelle en sub-bins afin d'éviter les contamination était nécessaire puisque seuls les sub_bins 2_1, 2_5, 2_13, 2_14 et 2_16 ont été conservés pour faire partie de l'assemblage final.

Le contraste se fait ressentir au niveau de la 3ème bin (*Figure 17c*), contenant les séquences du photobionte *Nostoc*. Le contenu en GC ainsi que le coverage est pratiquement homogène sur toute la bin. Plus de 99% des contigs de cette bin sont affiliés en tant que Cyanobacteria. À noter toutefois que celle-ci a été découpée en sub-bins par précaution, bien que le binning automatique semble ne pas contenir beaucoup de contamination.



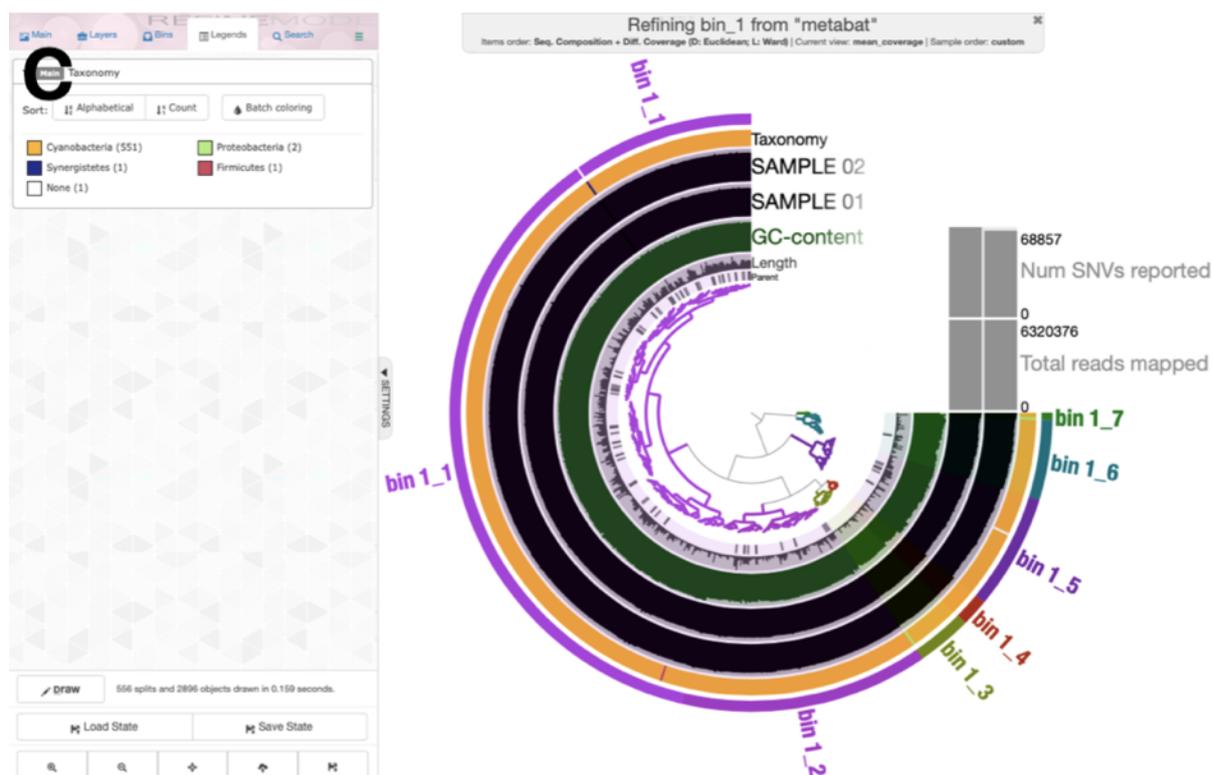


Figure 17: Visualisation *Anvio* des 3 bins générées par *MetaBAT 2* pour l'assemblage *S67*. Ces dernières sont séparées en sub-bins (bande extérieure) grâce aux informations combinées du coverage (Bandes noires), du contenu en GC (Bande verte) et en tenant compte des affiliations taxonomiques (bande "taxonomy"). A. Bin contenant la majorité des contigs affiliés aux Ascomycètes (2023). B. Bin contenant le reste des contigs Ascomycètes (944). C. Bin contenant les contigs de la cyanobactérie (551).

2.2 Complétude et contamination

Les 57 assemblages génomiques triés pour ne garder que des séquences du mycobionte *Peltigera* ont des gammes de valeurs très diverses concernant les complétudes de *BUSCO*. Elles sont comprises entre 0% pour les assemblages génomiques *S25* ainsi que *S51* et montent jusqu'à 85,6% pour l'assemblage le plus complet (*S29*). 25 assemblages génomiques ont une complétude relativement élevée, comprise entre 70% et 85%. Pour les 57 génomes, le seuil de contamination, estimé à partir des gènes retrouvés en plus d'une copie, ne subit pas de grands écarts de variations entre les différents génomes et se situe entre 0% et 0,5%. La moyenne des contamination est de 0,15%.

3. Analyses de génétique des populations sur R

3.1 SNP calling

Mon analyse de SNP calling a détecté 180 516 SNPs présents sur l'ensemble des 33 assemblages de *Peltigera*. Ce nombre chute à 26 561 SNPs en ne gardant que les sites pour lesquels il n'y a aucune donnée manquante au sein des 33 génomes.

3.2 Clustering

Le clustering des 33 *Peltigera* basé sur la distance de Pervosti montre que le *S44* est fortement distant des autres individus (Figure 18). La même observation est faite sur base des probabilités postérieures d'appartenance à un groupe pour K=3 populations (Figure 19). Elle place tous les *Peltigera appalachensis* dans un groupe (en bleu sur la Figure 19) et tous les *Peltigera vitikainenii* dans l'autre (en rouge) mais toujours à l'exception du *S44* (en jaune). Il est probable que l'échantillon *S44* ait été contaminé lors de l'extraction d'ADN ou du séquençage. L'échantillon *S44* a été exclu et les analyses ont ensuite été répétées sans lui.

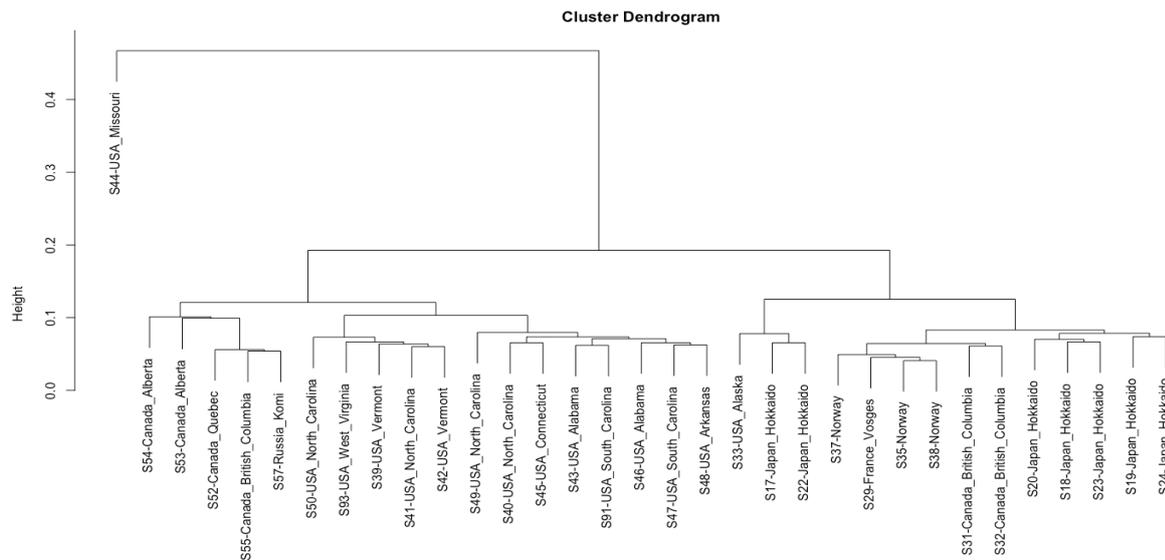


Figure 18 : Cluster Dendrogram des 33 mycobiontes de *Peltigera*, réalisé sur base de la distance de Pervosti sur les données SNP.

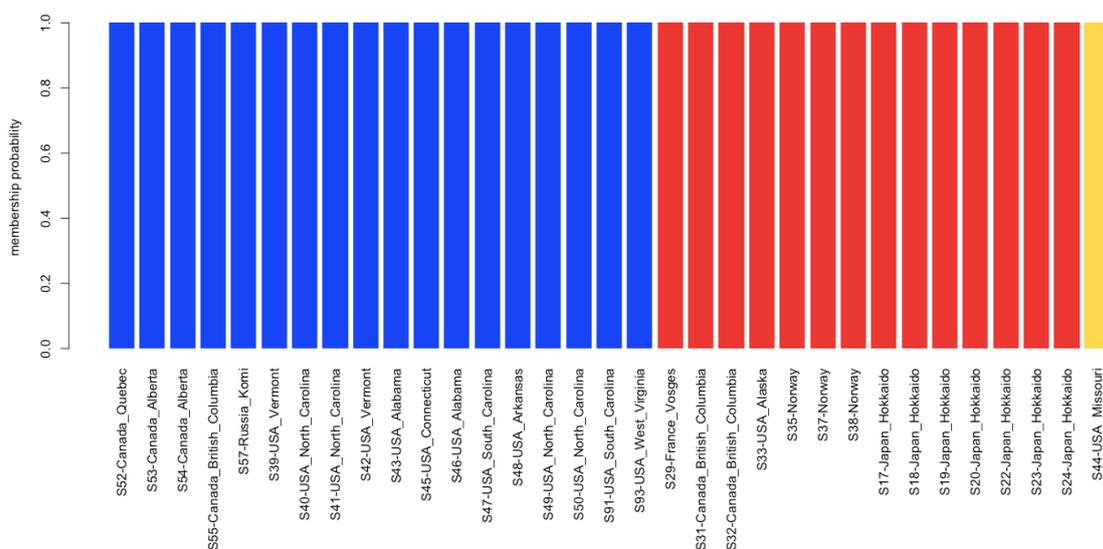


Figure 19: Graphique des probabilités postérieures d'appartenance à un groupe pour les 33 mycobiontes de *Peltigera* après analyse des distances de Pervosti sur les données SNP.

Après exclusion de l'échantillon S44, le BIC est optimal pour 5 clusters (Figure 20). Cependant, lorsque l'on observe les analyses pour k=2, on obtient bien une séparation entre les *P. appalachensis* (en bleu sur la Figure 21) et les *P. vitikainenii* (en rouge).

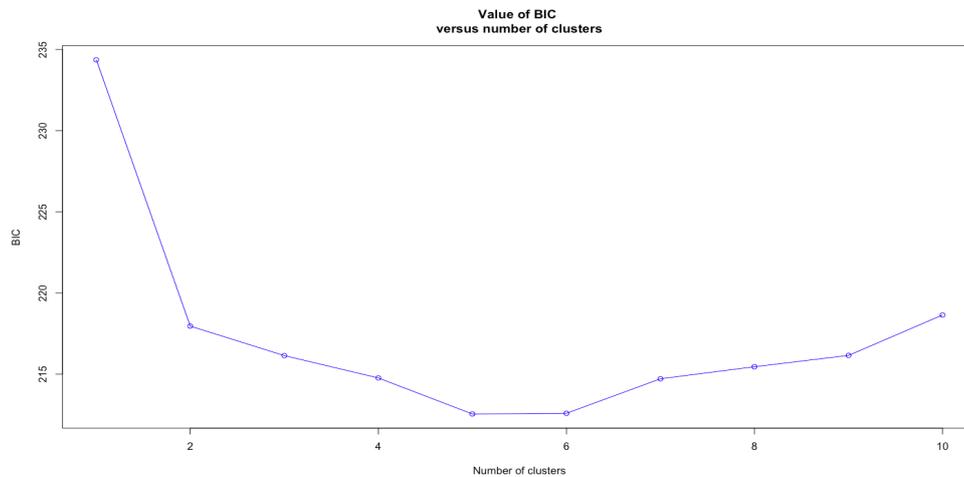


Figure 20: Evolution de la valeur du BIC (Bayesian Information Criterion) obtenu en fonction du nombre de clusters pour les 32 mycobiontes de *Peltigera* (après exclusion du S44).

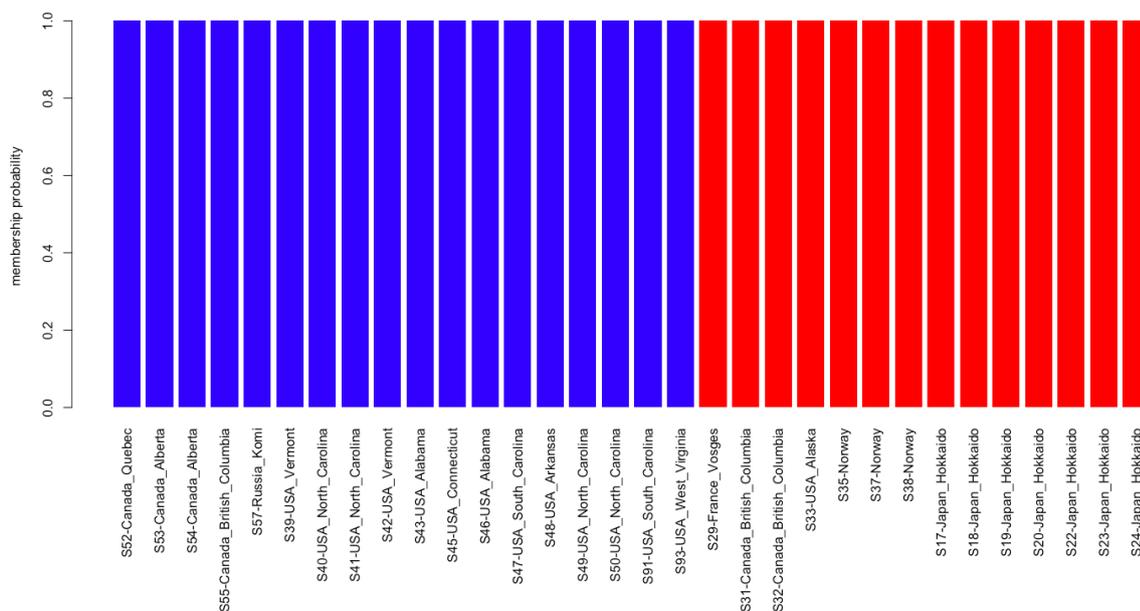


Figure 21: Graphique des probabilités postérieurs d'appartenance à un groupe avec k=2 clusters, pour les 32 mycobiontes de *Peltigera* (après exclusion du S44) après analyse des distances de Pervosti sur les données SNP.

Ensuite, pour k=5, on obtient une population correspondant au *P. appalachensis* de la zone boréale (en vert sur la Figure 22), et les *P. appalachensis* de la zone tempérée divisé en deux populations (en rose et en rouge). Ensuite, les *P. vitikainenii* sont divisés en deux populations, une avec deux échantillons du Japon et un d'Alaska (en bleu) et l'autre avec le reste des *P. vitikainenii* (en jaune).

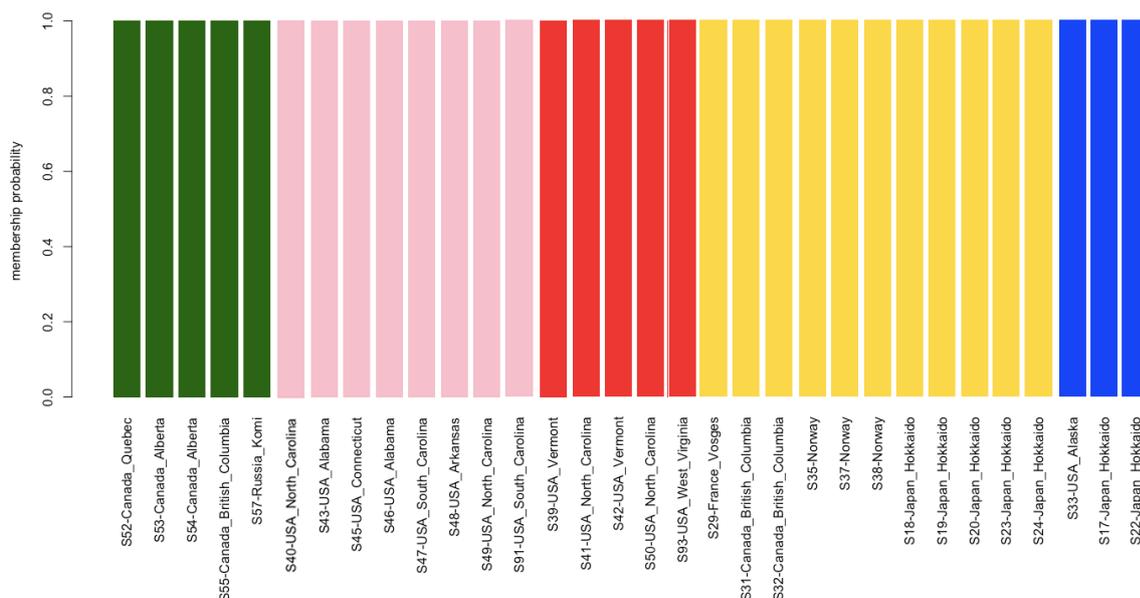


Figure 22: Graphique des probabilités postérieurs d'appartenance à un groupe avec k=5 clusters pour les 32 mycobiontes de *Peltigera* (après exclusion du *S44*) sur base des données SNP.

En visualisant le « cluster dendrogram » (*Figure 23*), on s’aperçoit que pour *P. vitikainenii*, tous les spécimens boréaux se groupent ensemble, à l’exception d’un spécimen d’Alaska. Ce groupe des boréaux apparaît « nesté » au sein de la diversité des tempérés, ce qui tendrait à confirmer l’hypothèse d’une colonisation de la zone boréale à partir du refuge tempéré.

Dans le « cluster dendrogram », pour *P. appalachensis*, tous les échantillons boréaux se groupent ensemble, et tous les tempérés ensemble. Les échantillons boréaux n'apparaissent pas « nestés » par rapport aux tempérés.

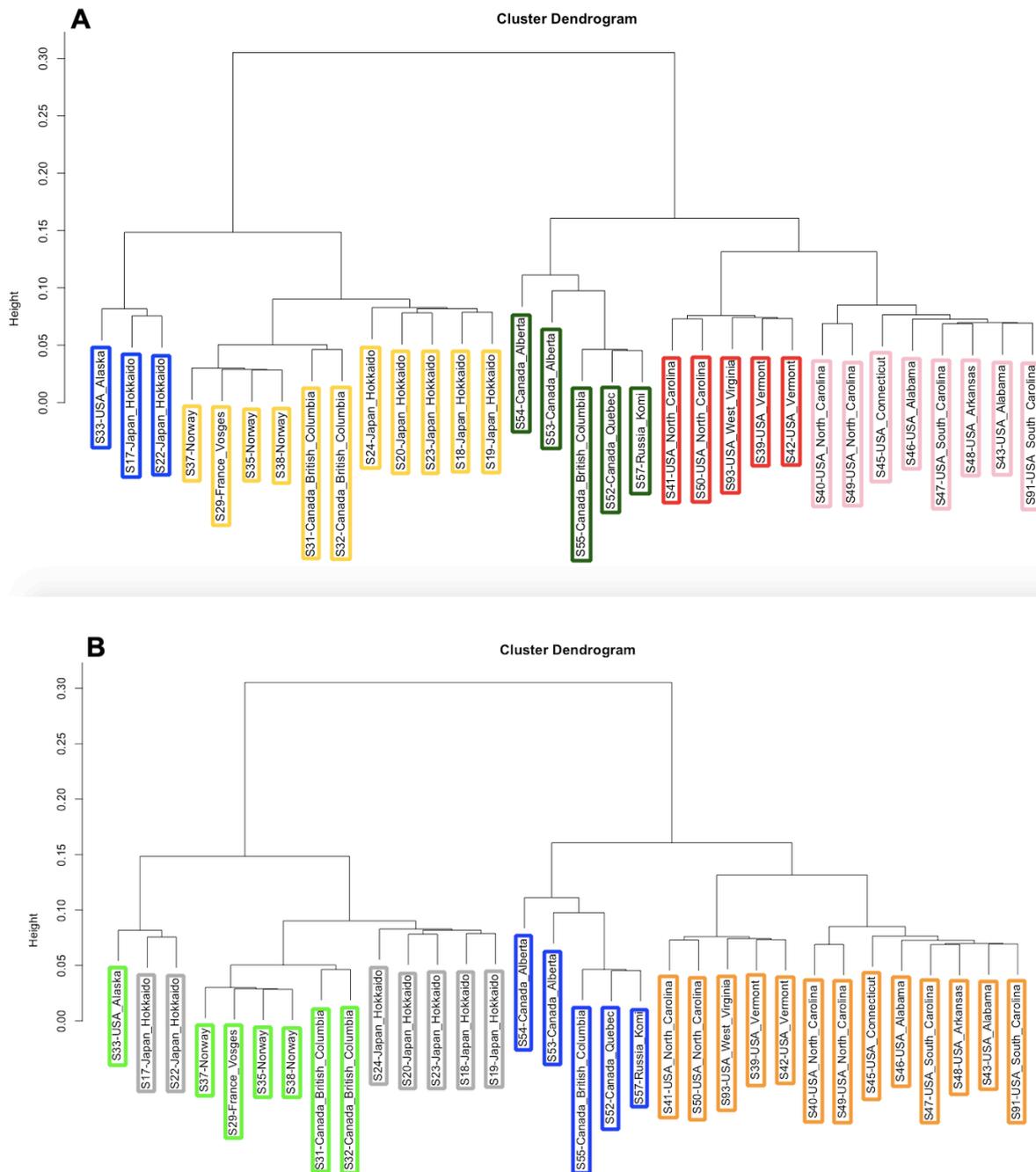


Figure 23: Les figures A et B représentent le même « Cluster Dendrogram » des 32 mycobiontes de *Peltigera*, réalisé sur base de la distance de Pervosti sur les données SNP. Dans la figure A, le code couleur correspond aux k=5 clusters conseillés sur base du BIC (idem *Figure 22*). Dans la figure B, les couleurs correspondent à l'origine géographique des spécimens : *P. appalachensis* tempérés en orange, *P. appalachensis* boréaux en bleu, *P. vitikainenii* tempérés en gris et *P. vitikainenii* boréaux en vert.

En ce qui concerne la PCA sur les données SNP (Figure 24), à nouveau les échantillons de *P. appalachensis* (orange et bleu) et de *P. vitikainenii* (vert et gris) apparaissent bien séparés, au niveau du 1er axe de la PCA.

Au niveau de *P. appalachensis*, le deuxième axe sépare clairement les échantillons tempérés (orange) des échantillons boréaux (bleu). Chacune de ces populations semblent ensuite se séparer en deux clusters. Parmi les boréaux, les deux génomes d'Alberta (Canada) semblent occuper une position intermédiaire entre le reste des échantillons boréaux (Québec, Colombie Britannique et Russie). Concernant les échantillons tempérés, il ne semble pas y avoir de différenciation géographique claire entre les échantillons des Appalaches, avec un recouvrement entre les deux groupes (du Vermont à la Caroline du Nord dans l'un, du Connecticut à l'Alabama dans l'autre, la Caroline du Nord représentée dans les deux clusters). Les échantillons des Ozark sont dans le groupe du haut.

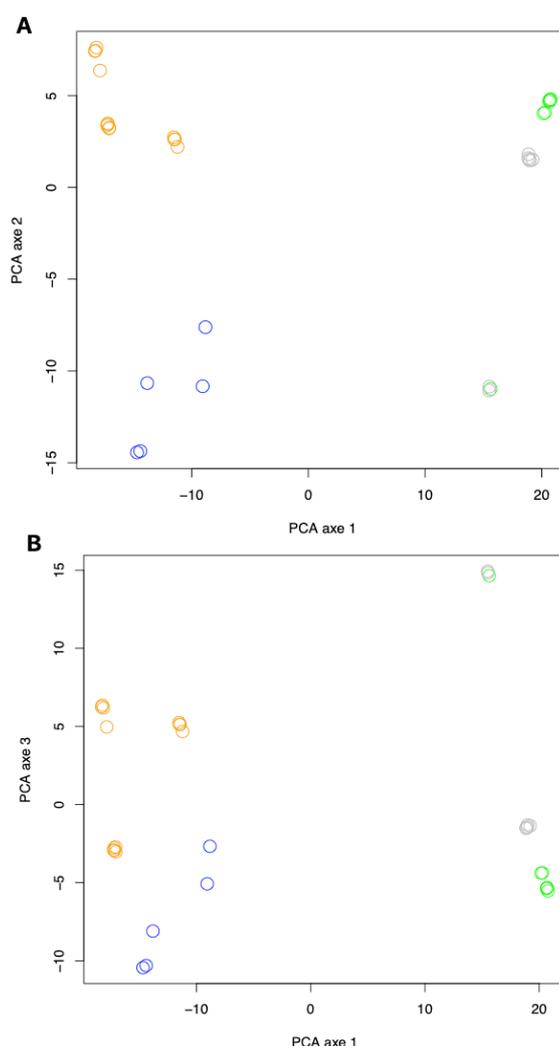


Figure 24: Analyse en composante principale (PCoA) sur les données SNP pour les 32 mycobiontes de *Peltigera* avec *P. appalachensis* tempérés en orange, *P. appalachensis* boréaux en bleu, *P. vitikainenii* tempérés en gris et *P. vitikainenii* boréaux en vert. La figure A représente les deux principaux axes de la PCoA, tandis que la figure B représente l'axe 1 et l'axe 3.

En ce qui concerne les *P. vitikainenii*, trois échantillons (deux du Japon et un d'Alaska) se regroupent loin des autres (en bas à droite dans la figure 24A, en haut à droite dans la figure 24B). En ce qui concerne les autres, on voit tous les boréaux (vert) se séparer des tempérés (gris) (en haut à droite de la figure 24A, en bas à droite de la figure 24B). Cela montre une différenciation génétique en deux groupes au sein de *P. vitikainenii*, ces 2 groupes étant présents au Japon. Cela suggère aussi que l'échantillon d'Alaska correspond sans doute à une autre dispersion vers la zone boréale, à travers le pacifique, tandis que le reste des échantillons boréaux (France, Norvège, Colombie Britannique).

3.3 Diversité génétique

En ce qui concerne les valeurs de diversité génétique à l'intérieur des populations, chez *P. appalachensis*, la diversité génétique est légèrement plus faible chez les échantillons boréaux (2202.4) par rapport aux tempérés (2626.4).

Chez les *P. vitikainenii*, la diversité nucléotidique des échantillons du Japon et de l'Est de la Russie (2859.4) est nettement supérieure à la diversité des échantillons hors de cette zone (1806.2), et cette valeur tombe à 1029.1 si on ne considère pas l'échantillon d'Alaska qui ne se groupe pas avec le reste des boréaux (*Figure 18*).

L'hypothèse de la recolonisation récente de la zone boréale par les *P. vitikainenii* se trouve donc fortement renforcée ici, avec une diversité génétique très faible des échantillons boréaux, de la Colombie Britannique jusqu'en Norvège, alors que la diversité génétique est beaucoup plus importante à l'intérieur du Japon.

En ce qui concerne les *P. appalachensis*, une tendance se dégage aussi dans ce sens, mais elle est plus difficile à interpréter, et les données phylogénomiques seront nécessaires pour approfondir ces scénarios biogéographiques.

DISCUSSION

1. Collecte des données

1.1 Mission de terrain

Afin de maximiser les chances de récolter *P. appalachensis* tout en minimisant la durée de la mission de récolte, la stratégie choisie fut d'échantillonner des localités où l'espèce avait déjà été récoltée, soit en utilisant la base de données du CNALH (Consortium of North American Lichen Herbaria), soit les informations présentes sur l'étiquette des échantillons empruntés à des herbiers. Cette stratégie s'est avérée très efficace. Sur 20 localités échantillonnées, nous avons récolté *P. appalachensis* dans 16 de celles-ci. Ceci souligne l'intérêt de ce type de base de données d'herbiers pour la recherche actuelle, et non uniquement pour une valeur patrimoniale. Grâce à cette stratégie, nous avons pu échantillonner une zone de plus de 2500 km en moins de trois semaines, alors qu'il aurait fallu plusieurs mois pour échantillonner toute cette zone et trouver des échantillons de *P. appalachensis* en suffisance si nous n'avions pas disposé des informations de localités préalables fournies par ces herbaria.

1.2 Cohérence des données génomiques avec les données ITS

Depuis plusieurs années, le locus ITS est utilisé comme code-barre pour le Royaume des Fungi (Schoch et al. 2012), et on l'utilise donc pour identifier l'espèce à laquelle appartiennent les champignons lichénisés, mais aussi pour la phylogénie de nombreux groupes, qui est parfois reconstruite sur base de ce seul locus (voir par exemple Moncada et al. 2013) ou pour des analyses de génétique des populations (Domaschke et al. 2013). Il est donc important que ce locus reflète le plus précisément possible l'histoire évolutive du reste du génome, pour que l'information de ce locus (son "gene tree") reflète autant que possible l'histoire évolutive des organismes (son "species tree").

Les données génomiques que nous avons produites semblent confirmer les patterns observés via l'ITS. En effet, sur base des haplotypes ITS, nous avons détecté une diversité génétique plus importante dans les zones de montagnes tempérées d'Amérique du Nord pour *P. appalachensis*, et au Japon et dans l'Est de la Russie pour *P. vitikainenii*, et les données génomiques le confirment. Ceci tendrait donc à renforcer l'utilité du locus ITS comme code-barre pour les *Peltigera* et pour les champignons en général.

2. Séquençage des génomes

2.1 Qualité du séquençage

Nous n'avons pu utiliser que 33 génomes de *Peltigera* pour nos analyses de génétique des populations, sur les 57 métagénomés séquencés. Ceci a, évidemment, rendu nos résultats plus difficiles à interpréter. L'élimination de ces 24 génomes était dû à leur faible complétude (moins de 40% des gènes BUSCO présents dans la bin).

On remarque que l'âge de l'échantillon a fortement influencé la qualité de l'assemblage et du binning. En effet, le séquençage de la majorité des échantillons récoltés en 2019 a très bien fonctionné, tandis que les échantillons récoltés il y a plus de 5 ans ont donné de moins bons résultats (*Tableau 2*). Il faut noter qu'en général, en se basant sur l'étude des assemblages avant binning, les gènes manquants ont bien été séquencés et sont présents dans les séquences brutes. Mais en général, ces génomes ont un « coverage » (une profondeur de séquençage) plus faible, et c'est lors de l'étape de binning que les séquences correspondant au *Peltigera* n'ont pas été correctement ajoutées à la bin (ou dans les cas extrêmes, aucune bin de *Peltigera* n'a été créée). La profondeur de séquençage faible de certains génomes a donc empêché les programmes de binning de fonctionner de manière optimale.

A l'avenir, il sera donc important de viser une profondeur de séquençage plus importante, pour obtenir des assemblages métagénomiques de meilleure qualité. Une possibilité à explorer dans le futur sera de re-séquencer certains génomes avec une profondeur de séquençage plus importante, pour améliorer la qualité des données et la performance du pipeline bioinformatique.

Une autre étape qui permettrait d'améliorer les assemblages, ainsi que le SNP calling, serait de disposer d'un génome de référence pour *Peltigera*. Comme il n'a pour l'instant pas été possible de cultiver le champignon *Peltigera* en conditions axéniques, une possibilité serait d'utiliser une méthode de séquençage qui génère de plus longs reads, comme PacBio ou Nanopore, à la place d'Illumina, pour faciliter l'assemblage d'un génome de référence.

Il sera nécessaire de ré-analyser les données en améliorant la qualité du séquençage ou la qualité du binning pour pouvoir utiliser les 24 génomes que nous avons dû éliminer des présentes analyses car la complétude des bins était insuffisante.

Par ailleurs, il se peut que *BUSCO* sous-évalue légèrement la complétude de nos génomes. En effet, les gènes de la base de donnée *ascomycota_oddb9* sont communs à tous les Ascomycètes et très conservés. Cependant cette dernière contient uniquement des champignons non-lichénisés. Or, le mycobionte de *Peltigera* vit en symbiose obligatoire. Il n'est donc pas à exclure que ce dernier ait une moins bonne conservation de ces gènes orthologues au regard de son mode de vie symbiotique.

Afin d'améliorer le binning automatique de *MetaBAT 2* pour qu'il fournisse de meilleurs résultats sur les lichens, il serait judicieux d'entraîner cet algorithme de machine learning sur des sets de données contenant à la fois des bactéries et des eucaryotes car jusqu'à présent il n'a été entraîné à différencier que des sets de données bactériens ou de pico-eucaryotes (communautés planctoniques).

Dans un contexte où les études scientifiques se veulent reproductibles au maximum, il serait souhaitable, à l'avenir, de communiquer un script Nextflow (<https://www.nextflow.io>) contenant l'entièreté du pipeline bio-informatique. Cela permettrait de palier aux difficultés inhérentes à la perpétuelle évolution des systèmes d'exploitations ainsi que celle des versions des programmes utilisés.

3. Génétique des populations

3.1 Isolation des deux espèces *P. appalachensis* et *P. vitikainenii*

Des études précédentes, basées sur le séquençage de 8 loci et des méthodes de délimitation des espèces, avaient suggéré que *P. appalachensis* et *P. vitikainenii* étaient deux espèces bien distinctes

(Magain et al. 2017b). Cependant, ces deux espèces étant très proches phylogénétiquement, il était intéressant d'utiliser les données venant du génome entier, plutôt que de 8 loci seulement, pour voir si les deux espèces apparaissent bien isolées ou si on pouvait détecter des traces d'hybridation et de "gene flow" entre les deux espèces, en particulier dans les zones boréales, où les deux espèces sont présentes en sympatrie.

Nos résultats préliminaires, notamment le clustering et la PCoA, ne montrent pas de trace d'hybridation. Aucun échantillon boréal n'apparaît dans une position intermédiaire entre les *P. appalachensis* et les *P. vitikainenii*. L'hypothèse que ces deux espèces sont à présent bien isolées s'en trouve donc renforcée. Cependant, des analyses visant spécifiquement à tester l'isolation entre les deux espèces, tels que des tests avec le programme *IMA2* (Hey 2010) seront nécessaires pour confirmer cette hypothèse.

3.2 Structuration génétique des deux populations et hypothèse des refuges glaciaires.

Pour les deux espèces, la diversité génétique est plus importante dans la zone pour laquelle nous avons fait l'hypothèse du refuge glaciaire que dans l'ensemble de la zone boréale, beaucoup plus vaste. L'hypothèse des refuges glaciaires est donc plutôt soutenue par nos données.

Cependant, les données sont aussi plus complexes qu'attendues.

En ce qui concerne *P. appalachensis*, la diversité génétique est plus faible dans la zone boréale, mais peut-être moins significativement qu'attendu. De plus, les échantillons boréaux se divisent en deux groupes, avec les échantillons d'Alberta à part du reste (Figure 24B). Il sera intéressant de séquencer davantage de spécimens du Canada pour voir si ces deux groupes sont structurés géographiquement ou si on peut les retrouver tous les deux dans les mêmes régions. Certains échantillons de la zone tempérée apparaissent assez proches des échantillons boréaux (Figure 24B), il sera intéressant de tester si c'est un représentant de cette population qui est l'ancêtre de la population boréale.

Dans le cas de *P. vitikainenii*, l'espèce semble se diviser en deux clusters génétiques. Les deux clusters sont présents au Japon, et un échantillon d'Alaska appartient au premier groupe, tandis que le reste des échantillons boréaux se regroupent dans le second. Les échantillons boréaux (hors spécimen d'Alaska) se regroupent ensemble, et apparaissent très proches, ce qui soutient l'hypothèse d'une recolonisation récente de la zone boréale. Mais l'échantillon d'Alaska représente sans doute un événement de recolonisation de la zone boréale distinct. Davantage de séquençage dans le Nord-Ouest américain et le Nord-Est asiatique (Kamchatka, par exemple) permettrait d'affiner ces hypothèses.

Dans l'étude de délimitation des espèces (Magain et al. 2017b), les échantillons correspondant à *P. vitikainenii* dans le présent travail avaient été séparés en deux clades, représentant deux espèces potentielles, appelés dans cette étude *P. neopolydactyla* 2a et *P. neopolycatyla* 2c. Il serait intéressant de tester si les deux groupes qui apparaissent bien distincts dans nos analyses (Figure 24) correspondent à ces deux lignées dans Magain et al. (2017b). Il sera également intéressant de tester si ces deux lignées sont suffisamment divergentes au niveau du génome entier pour justifier d'un statut taxonomique distinct.

Dans les deux cas, les analyses phylogénomiques prévues, sur base de plusieurs centaines de loci, permettront de reconstituer dans quel ordre et à quel rythme ces différents événements se sont

produits. Nous pourrions donc tester si c'est bien d'ancêtres présents dans les zones tempérées que les spécimens boréaux dérivent. Même si les données de génomique des populations vont dans le sens de nos hypothèses, c'est en complément avec les données phylogénomiques qu'elles prendront tout leur sens, et que nous pourrions valider ou non le scénario des refuges glaciaires.

Il sera également intéressant de répéter toutes ces analyses sur le partenaire photosynthétique cyanobactérien *Nostoc*, pour voir si les deux partenaires ont eu une histoire évolutive similaire, ou si les événements ayant mené à la situation actuelle des deux partenaires sont très différents

REFERENCES BIBLIOGRAPHIQUES

- Ahmadjian, V. (1993). *The Lichen Photobiont: What Can It Tell Us about Lichen Systematics?* *The Bryologist*, 96(3), 310-313.
- Allan, R. P., Ramaswamy, V., & Slingo, A. (2002). *Diagnostic analysis of atmospheric moisture and clear-sky radiative feedback in the Hadley Centre and Geophysical Fluid Dynamics Laboratory (GFDL) climate models.* *Journal of Geophysical Research: Atmospheres*, 107(D17), ACL-4
- Alo, C.A. & Wang, G.L. (2008). *Potential future changes of the terrestrial eco- system based on climate projections by eight general circulation models.* *J. Geophys. Res. Biogeosci.*, 113, 16.
- Aschenbrenner, I. A., Cernava, T., Berg, G., & Grube, M. (2016). *Understanding microbial multi-species symbioses.* *Frontiers in Microbiology*, 7, 180.
- Balthazar V, Baurain D, Cornet L. *An analysis of the contamination state of RefSeq.* 2020. *In prep*
- Bartlein, P. J., Harrison, S. P., Brewer, S., Connor, S., Davis, B. A. S., Gajewski, K., ... & Prentice, I. C. (2011). *Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis.* *Climate Dynamics*, 37(3), 775-802.
- Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). *Impacts of climate change on the future of biodiversity.* *Ecology letters*, 15(4), 365-377.
- Belnap, J. (2003). *The world at your feet: desert biological soil crusts.* *Frontiers in Ecology and the Environment*, 1(4), 181-189.
- Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). *Domain enhanced lookup time accelerated BLAST.* *Biology direct*, 7(1), 12.
- Brown, J., Pirrung, M., & McCue, L. A. (2017). *FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool.* *Bioinformatics*, 33(19), 3137-3139.
- Brumfield, R. T., Beerli, P., Nickerson, D. A., & Edwards, S. V. (2003). *The utility of single nucleotide polymorphisms in inferences of population history.* *Trends in Ecology & Evolution*, 18(5), 249-256.
- Buchfink, B., Xie, C., & Huson, D. H. (2015). *Fast and sensitive protein alignment using DIAMOND.* *Nature methods*, 12(1), 59-60.
- Cornet, L., Bertrand, A. R., Hanikenne, M., Javaux, E. J., Wilmotte, A., & Baurain, D. (2018a). *Metagenomic assembly of new (sub) polar Cyanobacteria and their associated microbiome from non-axenic cultures.* *Microbial genomics*, 4(9).

- Cornet, L., Meunier, L., Van Vlierberghe, M., Léonard, R. R., Durieu, B., Lara, Y., ... & Wilmotte, A. (2018b). Consensus assessment of the contamination level of publicly available cyanobacterial genomes. *PloS one*, 13(7), e0200323.
- Cubero, O. F., Crespo, A. N. A., Fatehi, J., & Bridge, P. D. (1999). DNA extraction and PCR amplification method suitable for fresh, herbarium-stored, lichenized, and other fungi. *Plant Systematics and Evolution*, 216(3-4), 243-249.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158.
- Dodds, W. K., Gudder, D. A. and Mollenhauer, D. (1995), *The Ecology of Nostoc*. *Journal of Phycology*, 31 : 2-18. DOI : 10.1111/j.0022-3646.1995.00002.x
- Domaschke, S., Fernández-Mendoza, F., A. García, M., Martín, M., & Printzen, C. (2012). Low genetic diversity in Antarctic populations of the lichen-forming ascomycete *Cetraria aculeata* and its photobiont. *Polar Research*, 31(1), 17353
- Ehlers, J., & Gibbard, P. (2011). Quaternary glaciation. *Encyclopedia of snow, ice and glaciers*, 873-882.
- Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319
- Evans, R. D., & Johansen, J. R. (1999). Microbiotic crusts and ecosystem processes. *Critical reviews in plant sciences*, 18(2), 183-225.
- Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481-501
- Friedmann, E. I. (1982). Endolithic microorganisms in the Antarctic cold desert. *Science*, 215(4536), 1045-1053.
- Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular ecology*, 2(2), 113-118.
- Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, 1000 Genome Project Data Processing Subgroup, *The Sequence Alignment/Map format and SAMtools*, *Bioinformatics*, Volume 25, Issue 16, 15 August 2009, Pages 2078-2079,
- Henriksson, E., & Simu, B. (1971). Nitrogen fixation by lichens. *Oikos*, 119-121.
- Hey, J. (2010). Isolation with migration models for more than two populations. *Molecular biology and evolution*, 27(4), 905-920.
- Holtan-Hartwig, J., & Holtan-Hartwig, J. (1993). The Lichen Genus 'Peltigera', Exclusive of the 'P. Canina' Group, in Norway. *Botanical Garden and Museum-University of Oslo*.

- Hua, X., & Wiens, J. J. (2013). How does climate influence speciation?. *The American Naturalist*, 182(1), 1-12.
- Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome research*, 17(3), 377-386.
- Jombart, T. (2008) *ade4*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405. doi:10.1093/bioinformatics/btn129
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359.
- Lawrey JD (2009) Chemical defense in lichen symbioses. In: White JF Jr, Torres MS (eds) *Defensive Mutualism in Microbial Symbiosis*. CRC Press, Boca Raton, FL, USA.
- Leadley, P., Pereira, H.M., Alkemade, R., Fernandez-Manjarres, J.F., Proenca, V., Scharlemann, J.P.W. et al. (2010). Biodiversity scenarios: projections of 21st century change in biodiversity and associated ecosystem services. In: *Secretariat of the Convention on Biological Diversity (ed. Diversity SotCoB)*. Published by the Secretariat of the Convention on Biological Diversity, Montreal, p. 1–132. Technical Series no. 50.
- Lewis, S., Maslin, M. Defining the Anthropocene. *Nature* 519, 171–180 (2015). <https://doi.org/10.1038/nature14258>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- Liu, J., Möller, M., Provan, J., Gao, L. M., Poudel, R. C., & Li, D. Z. (2013). Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist*, 199(4), 1093-1108.
- Lutzoni, François, Frank Kauff, Cydon J Cox, David McLaughlin, Gail Celio, Bryn Dentinger, Mahajabeen Padamsee, et al. "Assembling the Fungal Tree of Life: Progress, Classification, and Evolution of Subcellular Traits." *American journal of botany*. 91, no. 10 (October 2004): 1446–1480.
- Lutzoni, F., Pagel, M., & Reeb, V. (2001). Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*, 411(6840), 937-940.
- Maclean, I.M.D. & Wilson, R.J. (2011). Recent ecological responses to climate change support predictions of high extinction risk. *Proc. Natl Acad. Sci. USA*, 108, 12337–12342.

- Magain, N M, J M Iadlikowska, B G Offinet, E S Érusiaux, and F L Uetzoni. 2017. "Macroevolution of Specificity in Cyanolichens of the Genus *Peltigera* Section *Polydactylon* (Lecanoromycetes, Ascomycota)" 66 (1): 74–99.
- Malcolm, J.R., Liu, C.R., Neilson, R.P., Hansen, L. & Hannah, L. (2006). Global warming and extinctions of endemic species from biodiversity hotspots. *Conserv. Biol.*, 20, 538–548.
- Miadlikowska, Jolanta, Frank Kauff, Filip Högnabba, Jeffrey C Oliver, Katalin Molnár, Emily Fraker, Ester Gaya, et al. 2014. "A Multigene Phylogenetic Synthesis for the Class Lecanoromycetes (Ascomycota): 1307 Fungi Representing 1139 Infrageneric Taxa, 317 Genera and 66 Families." *Molecular Phylogenetics and Evolution* 79: 132–68.
- Miadlikowska, J., & Lutzoni, F. (2004). Phylogenetic classification of peltigeralean fungi (Peltigerales, Ascomycota) based on ribosomal RNA small and large subunits. *American Journal of Botany*, 91(3), 449-464
- Miadlikowska, J., & Lutzoni, F. (2000). Phylogenetic revision of the genus *Peltigera* (lichen-forming Ascomycota) based on morphological, chemical, and large subunit nuclear ribosomal DNA data. *International Journal of Plant Sciences*, 161(6), 925-958.
- Miadlikowska, J., & Lutzoni, F. (2004). Phylogenetic classification of peltigeralean fungi (Peltigerales, Ascomycota) based on ribosomal RNA small and large subunits. *American Journal of Botany*, 91(3), 449-464.
- Moncada, B., Lücking, R., & Suárez, A. (2014). Molecular phylogeny of the genus *Sticta* (lichenized Ascomycota: Lobariaceae) in Colombia. *Fungal Diversity*, 64(1), 205-231
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome research*, 27(5), 824-834.
- O'Brien, H. E., Miadlikowska, J., & Lutzoni, F. (2005). Assessing host specialization in symbiotic cyanobacteria associated with four closely related species of the lichen fungus *Peltigera*. *European journal of phycology*, 40(4), 363-378.
- Ochman, H., Worobey, M., Kuo, C.-H., Ndjango, J.-B. N., Peeters, M., Hahn, B. H., et al. (2010). Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol.* 8:e1000546. doi: 10.1371/journal.pbio.1000546
- Otálora, M. A., Martínez, I., O'Brien, H., Molina, M. C., Aragón, G., & Lutzoni, F. (2010). Multiple origins of high reciprocal symbiotic specificity at an intercontinental spatial scale among gelatinous lichens (Collemataceae, Lecanoromycetes). *Molecular Phylogenetics and Evolution*, 56(3), 1089-1095.
- Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26: 419-420.
- Paradis, E., Gosselin, T., Goudet, J., Jombart, T., & Schliep, K. (2017). Linking genomics and population genetics with R. *Molecular ecology resources*, 17(1), 54-66.

- Parmesan, C. (2006). *Ecological and evolutionary responses to recent climate change*. *Ecol. Evol.*, 37, 637–669.
- Parmesan, C., Duarte, C.M., Poloczanska, E., Richardson, A.J. & Singer, M.C. (2011). *Overstretching attribution*. *Nat. Clim. Change*, 1, 2–4.
- Pereira, H.M., Leadley, P.W., Proenca, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarres, J.F. et al. (2010). *Scenarios for global biodiversity in the 21st century*. *Science*, 330, 1496–1501.
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). *PopGenome: an efficient Swiss army knife for population genomic analyses in R*. *Molecular biology and evolution*, 31(7), 1929-1936.
- Prevosti, A., Ocana, J., & Alonso, G. (1975). *Distances between populations of Drosophila subobscura, based on chromosome arrangement frequencies*. *Theoretical and Applied Genetics*, 45(6), 231-241.
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rai, A. N. (2002). *Cyanolichens: nitrogen metabolism*. In *Cyanobacteria in symbiosis* (pp. 97-115). Springer, Dordrecht.
- RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- Rikkinen, Jouko & Oksanen, Ilona & Lohtander, Katileena. (2002). *Rikkinen J, Oksanen J, Lohtander K. 2002. Lichen guilds share related cyanobacterial symbionts*. *Science* 297: Science (New York, N.Y.). 297. 357.
- Robert Lücking, Brendan P. Hodkinson, Steven D. Leavitt "The 2016 classification of lichenized fungi in the Ascomycota and Basidiomycota – Approaching one thousand genera," *The Bryologist*, 119(4), 361-416, (3 January 2017).
- Root, B.A., Price, J.T. & Hall, K. (2003). *Fingerprints of global warming on wild animals and plants*. *Nature*, 421, 47–60.
- Sala, O.E., van Vuuren, D., Pereira, H.M., Lodge, D., Alder, J., Cumming, G. et al. (2005). *Chap 10: biodiversity across Scenarios*. In: *Millenium Ecosystem Assesment* (ed. Island press, NY). *Millenium Ecosystem Assesment, Volume 2: Scenarios assessment*, pp. 375–408.
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., ... & Fungal Barcoding Consortium. (2012). *Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi*. *Proceedings of the National Academy of Sciences*, 109(16), 6241-6246.
- Schuster, S. C. (2008). *Next-generation sequencing transforms today's biology*. *Nature methods*, 5(1), 16-18.

Seppy, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. In *Gene Prediction* (pp. 227-245). Humana, New York, NY.

Shifu Chen, Yanqing Zhou, Yaru Chen, Jia Gu; *fastp: an ultra-fast all-in-one FASTQ preprocessor*, *Bioinformatics*, Volume 34, Issue 17, 1 September 2018, Pages i884–i890,
Nash, III, T. (Ed.). (2008). *Lichen Biology* (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511790478

Sierra, M. A., Danko, D. C., Sandoval, T. A., Pishchany, G., Moncada, B., Kolter, R., ... & Zambrano, M. M. (2020). The microbiomes of seven lichen genera reveal host specificity, a reduced core community and potential as source of antimicrobials. *Frontiers in microbiology*, 11, 398.

Tehler, Anders, Damon P Little, and James S Farris. 2003. "The Full-Length Phylogenetic Tree from 1551 Ribosomal Sequences of Chitinous Fungi, Fungi* *The Essence of the Paper Was Presented at the 7th International Mycological Congress IMC7 in Oslo on 12 August 2002." *Mycological Research* 107 (8): 901–16.

VITIKAINEN, ORVO. "Three New Species of *Peltigera* (lichenized Ascomycetes)." *Annales Botanici Fennici* 22, no. 4 (1985): 291-98. Accessed July 19, 2020.

Waterhouse RM, Zdobnov EM, Kriventseva EV (2011) Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. *Genome Biol Evol* 3:75–86. <https://doi.org/10.1093/gbe/evq083>

White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) *PCR protocols: a guide to methods and applications*. Academic Press, Inc., New York, pp 315–322

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology*, 15(3), 1-12.

Wu, Y. W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, 32(4), 605-607.

ANNEXES

Annexe 1 : Les cladogrammes 3A et 3B représentent des ordres actuellement acceptés au sein des Ascomycètes. 3A, ordres sans Eurotiomycetes ni Lecanoromycetes. 3B, ordres Eurotiomycetes et Lecanoromycetes. Le dernier cladogramme correspond aux Basidiomycètes. Les clades presque entièrement lichénisés ou entièrement lichénisés sont en vert. Et les clades partiellement lichénisés sont en bleu. Figures extraites de Lücking et al. (2017).

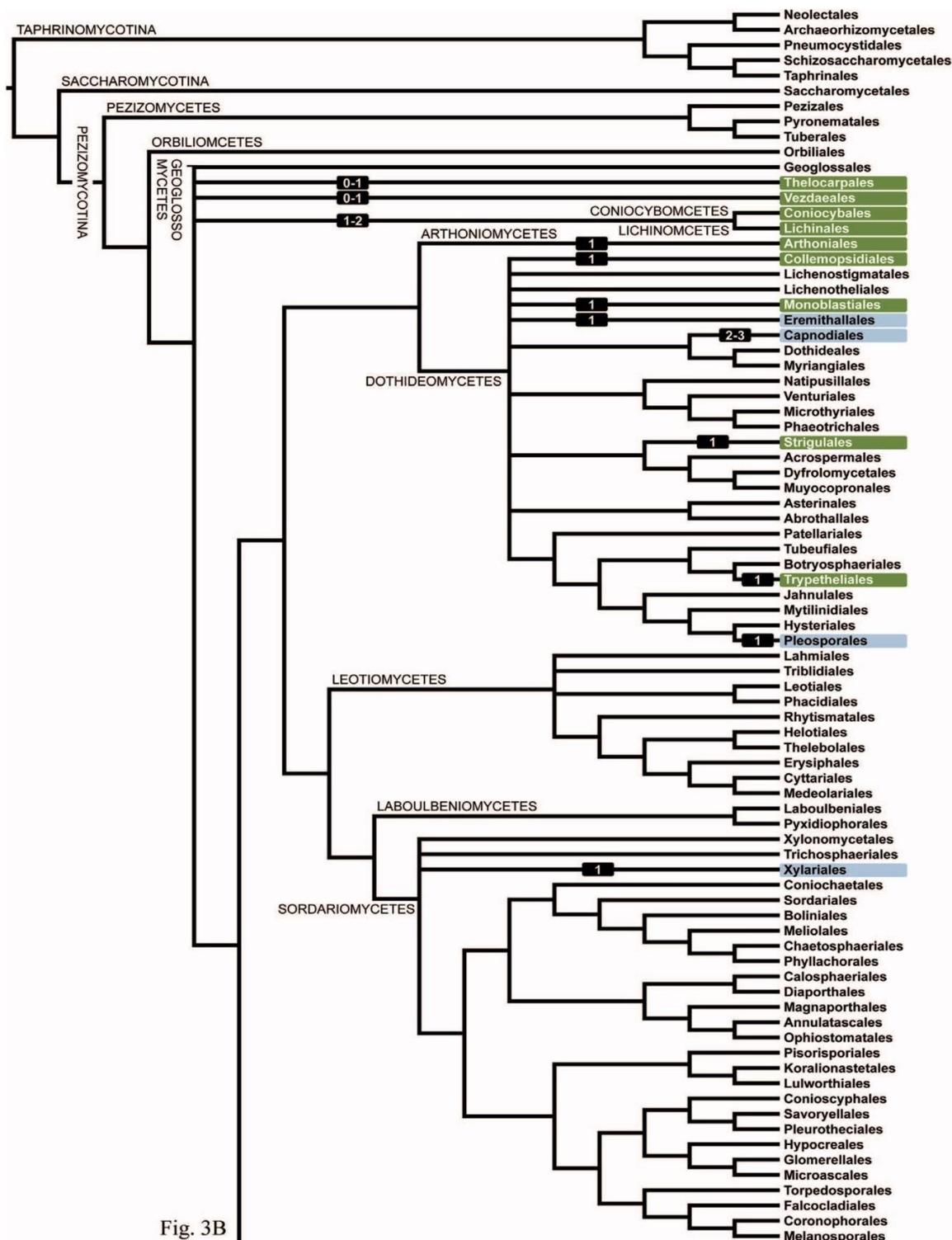
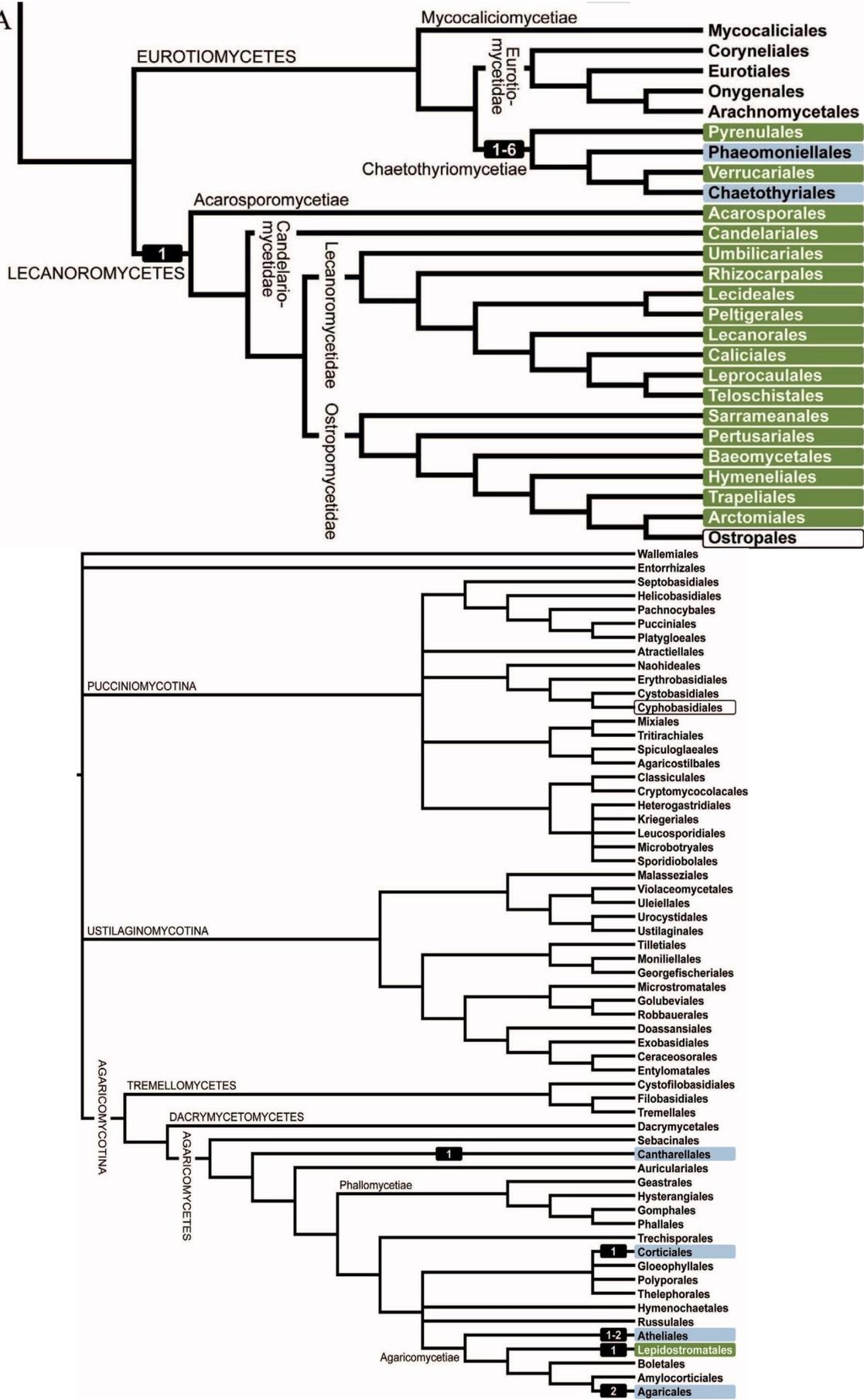


Fig. 3A



Annexe 2 : Tableau récapitulatif détaillant le numéro, l'origine géographique, la date de collecte, l'herbier ainsi que les collecteurs des lichens séquencés via NGS.

N°	Espèces	Origine géographique	Année de collecte	Herbier	Collecteurs
S14	<i>Peltigera pacifica</i>	Canada:British Columbia	2012	UBC	T. Goward
S15	<i>Peltigera pacifica</i>	USA:Oregon	2005	OSC	B. McCune
S16	<i>Peltigera pacifica</i>	Canada:British Columbia	1994	H	O. Vitikainen
S17	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S18	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S19	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S20	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S21	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S22	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S23	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S24	<i>Peltigera vitikainenii</i>	Japan:Hokkaido	2019	LG	Magain et al.
S25	<i>Peltigera vitikainenii</i>	Japan:Honshu	2017	TU	I. Juriado
S26	<i>Peltigera vitikainenii</i>	Japan:Honshu	2017	TU	i. Juriado
S27	<i>Peltigera vitikainenii</i>	Russia:Khabarovsk	2018	MHA	L Konoreva
S28	<i>Peltigera vitikainenii</i>	Russia:Sakhalin	2013	SAKH	S. Tchabanenko
S29	<i>Peltigera vitikainenii</i>	France: Vosges	2019	LG	N. Magain
S30	<i>Peltigera vitikainenii</i>	Suisse	1998	G	Mathias Vust
S31	<i>Peltigera vitikainenii</i>	Canada:British Columbia	inconnu	UBC	T. Goward
S32	<i>Peltigera vitikainenii</i>	Canada:British Columbia	inconnu	UBC	T. Goward
S33	<i>Peltigera vitikainenii</i>	USA:Alaska	2013	OSC	K. Spickerman
S34	<i>Peltigera vitikainenii</i>	Estonia	2017	TU	i. Juriado
S35	<i>Peltigera vitikainenii</i>	Norway	2011	LG	N. Magain
S36	<i>Peltigera vitikainenii</i>	Norway	2011	LG	N. Magain
S37	<i>Peltigera vitikainenii</i>	Norway	2011	LG	N. Magain
S38	<i>Peltigera vitikainenii</i>	Norway	2011	LG	N. Magain
S39	<i>Peltigera appalachensis</i>	USA:Vermont	2019	LG	Magain Balthazar
S40	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	LG	Magain Balthazar
S41	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	LG	Magain Balthazar
S42	<i>Peltigera appalachensis</i>	USA:Vermont	2019	LG	Magain Balthazar
S43	<i>Peltigera appalachensis</i>	USA: Alabama	2019	LG	Magain Balthazar
S44	<i>Peltigera appalachensis</i>	USA: Missouri	2019	LG	Magain Balthazar
S45	<i>Peltigera appalachensis</i>	USA: Connecticut	2019	LG	Goffinet
S46	<i>Peltigera appalachensis</i>	USA: Alabama	2019	LG	Magain Balthazar
S47	<i>Peltigera appalachensis</i>	USA: South Carolina	2019	LG	Magain Balthazar
S48	<i>Peltigera appalachensis</i>	USA: Arkansas	2019	LG	Magain Balthazar
S49	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	LG	Magain Balthazar
S50	<i>Peltigera appalachensis</i>	USA: North Carolina	2019	LG	Magain Balthazar
S51	<i>Peltigera appalachensis</i>	Canada: Quebec	2011	DUKE	Miadlikowska et al.
S52	<i>Peltigera appalachensis</i>	Canada: Quebec	2011	DUKE	Miadlikowska et al.
S53	<i>Peltigera appalachensis</i>	Canada: Alberta	inconnu	PMAE	D. Haughland
S54	<i>Peltigera appalachensis</i>	Canada: Alberta	inconnu	PMAE	D. Haughland
S55	<i>Peltigera appalachensis</i>	Canada: British Columbia	inconnu	UBC	T. Goward
S56	<i>Peltigera appalachensis</i>	Russia: Sakha	2011	H	Teuvo Ahti
S57	<i>Peltigera appalachensis</i>	Russia: Komi	2019	TU	Inga Juriado
S58	<i>Peltigera appalachensis</i>	Norway	2011	LG	Magain
S59	<i>Peltigera appalachensis</i>	USA: Alaska	2006	CONN	Bernard Goffinet
S60	<i>Peltigera appalachensis</i>	USA: Alaska	2011	OSC	Bruce McCune
S61	<i>Peltigera asiatica</i>	Malaysia: Borneo	2018	DUKE	Magain et al.
S62	<i>Peltigera asiatica</i>	China: Yunnan	2003	DUKE	Miadlikowska
S63	<i>Peltigera asiatica</i>	Taiwan	2016	LG	Serussiaux
S64	<i>Peltigera mikado</i>	Russia: Khabarovsk	2013	DUKE	Miadlikowska
S65	<i>Peltigera mikado</i>	China: Yunnan	2010	CONN	Goffinet
S66	<i>Peltigera mikado</i>	Taiwan	2016	LG	Serussiaux
S67	<i>Peltigera gayae</i>	Peru	2011	Duke	Gaya
S68	<i>Peltigera gayae</i>	Puerto Rico	2018	F	Mercado-Diaz
S91	<i>Peltigera appalachensis</i>	USA:South Carolina	2019	LG	Magain Balthazar
S93	<i>Peltigera appalachensis</i>	USA: West Virginia	2019	LG	Magain Balthazar

Les Annexes concernant le binning sont disponibles sur internet, via l'adresse <https://github.com/balthazar-alexis/memoire>. Elles reprennent un visuel *Anvio* de chacune des 380 bins de *MetaBAT 2* avec les sub-bins, ainsi qu'un tableau récapitulatif du regroupements de 4059 sub-bins effectués pour composer l'assemblage final.