

TITRE

Maxime Jaunatre, Master 1 BEE Grenoble

Mai 2019



Résumé

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Table des matières

1	Introduction	2
2	Matériels et méthodes	2
2.1	Échantillonnage	2
2.2	Bioinformatique	3
2.3	Génétique des populations	3
3	Résultats	4
3.1	Tri du jeu de données	4
3.2	Clade Hirsuta	4
3.3	Complexe ouest alpin	5
4	Discussion	5
5	Bibliographie	7
6	Ressources	7

1 Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

2 Matériels et méthodes

2.1 Échantillonnage

Les données génétiques utilisés lors de cette étude ont été produite dans le cadre de l'étude précédente de Boucher *et al.* [2016]. L'identification taxonomique des individus a été réalisé sur le terrain.

Les SNPs sont issus de séquences obtenus par séquençage via sondes RAD. Cette technique permet de génotyper le long du génome malgré des mutations sur les sites de restrictions. En effet les enzymes de restrictions sont trop sensibles à la mutation d'un nucléotide, tandis que les sondes ARN peuvent s'hybrider sur des sites plus nombreux sur le génome. La nécessité de capturer des sites malgré une faible variation provient du niveau interspécifique de l'étude, qui pose l'hypothèse que les mutations peuvent se placer sur les sites de restrictions et ainsi limiter leur capture par séquençage ddRAD.

2.2 Bioinformatique

Le jeu de données initial contient 90 individus pour les espèces échantillonnées initialement au sein de la section Auricula. Pour chacun de ces individus, l'information consiste en une séquence de SNPs appelés par Freebayes. Les analyses ont été portées sur deux jeux de données différents car extrait de Freebayes sous des seuils différents.

A partir des deux jeux de données initial, un pipeline est établi pour générer divers ensembles de données. Cette automatisation a permis entre autre de pouvoir évaluer l'effet des seuils posés au fur et à mesure de l'analyse. Les fonctions sont rassemblées en un package R hébergé sur Github (section 6) Dans un premier temps, le fichier initial est traité sous Rstudio [Team, 2017], avec la fonction `subset_reorder`, qui permet de reconstruire le fichier en ne gardant que les individus souhaités dans l'ordre indiqué. La fonction suivante, `rare`, permet de trier les allèles considérés comme étant présents dans un trop faible pourcentage des individus. Ces allèles rares sont écartés du jeu de données et les loci pour l'individu présentant ce loci sont considérés comme une donnée manquante. Suite aux deux tris précédents, il y a donc des loci pour lesquels tout les individus portent la même information. Afin de ne garder que les loci polymorphiques, une fonction `clean` est donc appelée à la fin de la fonction `rare`, afin de supprimer les loci monomorphiques. Enfin une dernière fonction de tri est appelée pour supprimer les loci puis individus présentant trop de données manquantes.

Afin de pouvoir utiliser les fichiers triés sous divers formats, la sous-sélection d'individus et de loci est enregistrée sous format `vcf`, au moyen d'une fonction appelant un script `bash`.

Chaque seuil est choisi après vérification que la décision n'impacte que peu les résultats de diversité génétique. Le nombre d'individus étant faible, il faut optimiser le nombre de SNP par individus car cela permet de limiter la perte d'information et d'atteindre les mêmes résultats qu'attendus avec un échantillon plus vaste. [Nazareno *et al.* , 2017]

2.3 Génétique des populations

Les analyses de génétique des populations ont été réalisées sous Rstudio.[Team, 2017]

3 Résultats

3.1 Tri du jeu de données

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

3.2 Clade Hirsuta

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.3 Complexe ouest alpin

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4 Discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipis-

cing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Remerciements Je tiens enfin à remercier sincèrement les personnes qui m’ont aidé durant ce stage. Tout d’abord Florian Boucher qui, en qualité de maître de stage, m’a accueilli au sein du LECA parmi l’équipe **EQUIPE???** et confié ses données. Ce stage m’a permis de me confronter à un sujet d’étude passionnant et je tiens donc à lui témoigner ma plus grande reconnaissance, en particulier pour sa patience, sa disponibilité et ses explications. Je remercie également Stéphanie Sherpa qui m’aura guidé durant mes recherches en génétique des populations.

Je n’oublie pas les autres membres du LECA, qui auront contribué à rendre mon stage agréable par leur attitude chaleureuse et bienveillante à mon égard. Je remercie enfin Caroline Kebaili, **MEH?** et **MEH?**, stagiaires de M2, avec qui j’ai partagé la salle 308 durant ces deux mois et qui ont toujours répondu présent à la moindre de mes questions.

5 Bibliographie

- Boucher, F. C., Casazza, G., Szövényi, P., & Conti, E. 2016. Sequence capture using RAD probes clarifies phylogenetic relationships and species boundaries in *Primula* sect. *Auricula*. *Molecular Phylogenetics and Evolution*, **104**, 60–72.
- Nazareno, Alison G., Bemmels, Jordan B., Dick, Christopher W., & Lohmann, Lúcia G. 2017. Minimum sample sizes for population genomics : an empirical study from an Amazonian plant species. *Molecular Ecology Resources*, **17**(6), 1136–1147.
- Team, R. 2017. R : A language and environment for statistical computing (Version 3.4. 2)[Computer software]. *Vienna, Austria : R Foundation for Statistical Computing*.

6 Ressources

<https://github.com/gowachin/Pedemontana>