

Délimitation d'espèces au sein du complexe de plantes des Alpes, *Primula pedemontana* s.l.

Maxime Jaunatre, Master 1 BEE Grenoble

1 Avril - 31 Mai 2019 — Soutenance : juin 2019

Enseignant référent :

Éric Coissac

Tuteur de stage :

Florian Boucher

Équipe :

DivAdapt



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.

Les nom de genre, section, sous-section, locus et loci en italiques, c'est du latin.

abstract

Todo list

| | |
|---|----|
| abstract | 2 |
| Les nom de genre, section, sous-section, locus et loci en italiques, c'est du latin. . | 2 |
| Pas clair : pourquoi c'est facilement observable ? Parce que c'est des patch d'habitat discrets ? | 1 |
| intro propre | 1 |
| intro trop détail | 1 |
| intro review | 2 |
| je ne sais pas si c'est l'approche standard, mais ça permet de garder l'information sur les autres individus. Au pire le <i>locus</i> devient monomorphique. Et au pire si c'est une erreur, il se peut que le <i>locus</i> ai un phred faible, et j'ai un seuil a plus tard pour ça. | 4 |
| hum on ne garde pas la notion de suite ? Parce que les deux tris sont consécutifs et non simultanés...peut-être trop de la cuisine ? | 5 |
| ça vaut peut être le coup de m'attarder sur comment il mesure ça. Si j'explique pas là et que j'ai a l'oral, je sais pas encore comment l'expliquer | 7 |
| proposer une explication ? peut etre parce qu'il y a une sous structure ? quid du Fis ;0 | 8 |
| Figure : Topologie | 8 |
| Figure : cluster | 8 |
| Figure : ACP | 8 |
| result Erythro | 8 |
| Figure : structure | 9 |
| result hirsuta | 9 |
| Figure : cline | 10 |
| result admixture | 10 |
| emo discuss | 11 |
| discussion complete | 12 |
| Marque page de fin de doc = 12!! | 12 |

Table des matières

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 2 | Matériels et méthodes | 3 |
| 2.1 | Cas d'étude | 3 |
| 2.2 | Échantillonnage | 3 |
| 2.3 | Bioinformatique | 4 |
| 2.4 | Génétique des populations | 5 |
| 2.5 | Hypothèse d'admixture | 6 |
| 3 | Résultats | 7 |
| 3.1 | Tri du jeu de données | 7 |
| 3.2 | Sous-section <i>Érythrodrosum</i> | 7 |
| 3.3 | Clade 'Hirsuta' | 8 |
| 3.4 | Admixture | 9 |
| 4 | Discussion | 11 |
| 5 | Bibliographie | 13 |
| 6 | Ressources | 14 |

1 Introduction

Par la diversité des milieux qu'elle abrite, la chaîne des Alpes est une zone géographique propice à une grande biodiversité. Cet ensemble de massifs et de vallées constitue donc un laboratoire parfait pour étudier l'évolution des organismes dont les aires de répartition semblent aisément observables.

Pas clair : pourquoi c'est facilement observable ? Parce que c'est des patch d'habitat discrets ?

Cependant, l'histoire de cette orogénèse et les différents cycles glaciaires qui ont joué sur l'Holarctique complexifient énormément ces histoires biologiques. Ainsi, l'isolement géographique propice au processus de spéciation, ou encore l'adaptation à de nouvelles niches écologiques se sont trouvés modifiés de nombreuses fois par ces facteurs abiotiques. Ces nombreuses modifications ont abouti à un état actuel de biodiversité où il est difficile de séparer ces ensembles de mécanismes, et conclure à une histoire biogéographique simple. En conséquence, on observe nombre d'espèces cryptiques, ou d'entités taxonomiques définies sur des critères évoluant rapidement dans la littérature.

Il faudrait réorganiser cette section. Introduire les espèces cryptiques est une très bonne idée, mais il faudrait expliquer pourquoi on en a pas mal dans la flore des Alpes. Pour l'instant on ne voit pas trop le lien.

Dans le cadre de cette étude, il a donc été proposé d'approfondir l'étude de la structure génétique d'un groupe de plantes d'altitude : *Primula* sect. *Auricula* Scott subsect. *Erythrodorum* Pax. Lors de récentes études phylogénétiques au niveau de la section *Auricula* (Boucher *et al.* , 2016, Zhang & Kadereit, 2004, Zhang *et al.* , 2004), cette sous-section a été restructurée en s'appuyant sur des données génétiques nucléaires. Cependant la dernière étude de Boucher *et al.* (2016) basée sur une nouvelle méthode de séquençage haut débit (hyRAD) suggère qu'un ensemble d'espèces (*P. apennina*, *P. cottia*, *P. pedemontana*) soit regroupé en une seule entité taxonomique, nommée ci-dessous *P. pedemontana* s.l.. Cette proposition vient également soulever l'hypothèse d'une nouvelle sous-espèce à cette entité dans le massif des Écrins, qui pourrait être un hybride entre *P. hirsuta* et *P. pedemontana*. Cette population des Écrins a longtemps été une difficulté taxonomique pour les botanistes (<http://www.ecrins-parcnational.fr/actualite/lenigme-de-la-primevere-du-valgaudemar>) et les dernières analyses ne la considèrent pas comme une espèce propre.

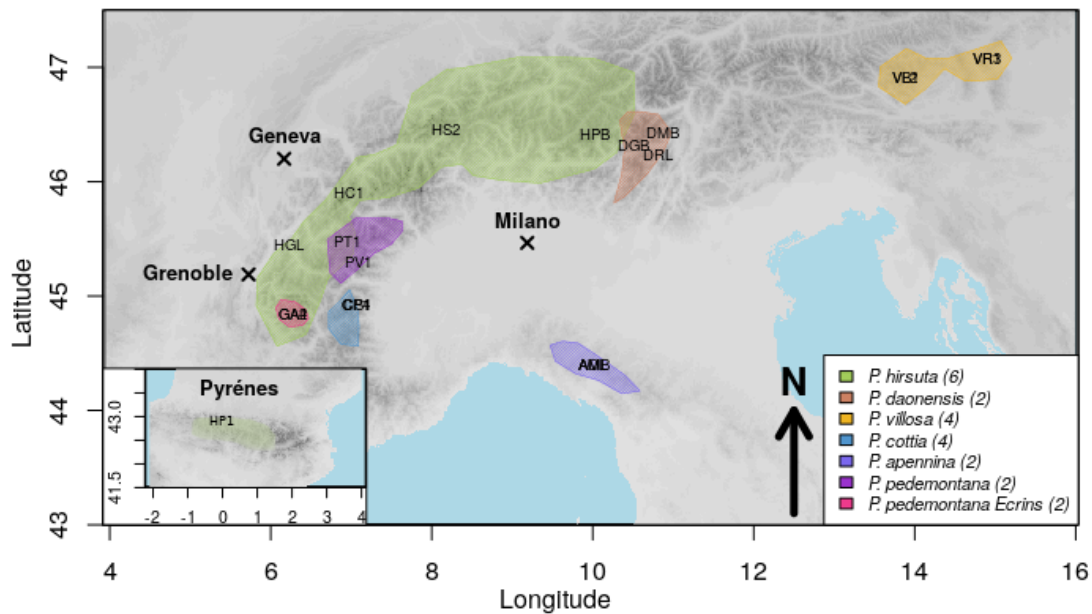


FIGURE 1 – **Carte de répartition de *Primula* sect. *Auricula* Scott subsect. *Erythrodrosom* Pax.** Espèces composant la sous-section, avec entre parenthèse le nombre d'échantillons de cette étude. Les aires de répartitions sont extrapolées depuis les observations de l'INPN et du CBNA. Fond de carte : Température moyenne annuelle extraite de WorldClim. Résolution 30s (1km²)

C'est trop de détails pour l'intro. il faut garder les détails taxonomiques/systématiques dans le premier paragraphe des méthodes. Ici ce qui compte à mon avis c'est de dire qu'on a un groupe d'espèces cryptiques et un taxon qu'on soupçonne d'être d'origine hybride. Pour être carré il faudra donner les noms d'auteurs des espèces que tu mentionnes, que tu trouveras dans la dernière révision taxonomique de Zhang et Kadereit.

De manière générale, je trouve cette intro un peu courte. Il faudrait essayer d'introduire les grandes questions de biogéographie et de biologie évolutive auxquelles ce sujet se rapporte. Du même coup ça permettrait d'expliquer pourquoi on s'intéresse à ce taxon en particulier. La section *Auricula* n'est pas présentée non plus, alors que c'est le groupe de plantes le plus riche dans le système Alpin Européen (cf. papier d'Ozenda joint).

Les buts de cette étude sont donc : (i) étudier le clade 'Hirsuta' à l'aide d'outils de génétique des populations ; (ii) clarifier la structure de *P. pedemontana* s.l. ; (iii) étudier l'hypothèse d'hybridation/introgression du taxon des Écrins.

intro
re-
view

2 Matériels et méthodes

2.1 Cas d'étude

La section *Auricula* Pax (1905) du genre *Primula* comprend 21 espèces et est considérée comme endémique du système Alpin (Ozenda, 1995). Au sein de cette section, la sous-section *Erythrodrosom* englobe 7 espèces : *P. villosa* ; *P. daonensis* ; *P. hirsuta* ; *P. apennina* ; *P. cottia* et *P. pedemontana*. L'ensemble de ces espèces sont hexaploïdes avec un nombre de chromosome dérivé de $x = 11$ (du fait de l'ancienneté de la polyploïdie, elles sont considérées ici comme diploïdes). Leur niche écologique est située entre 300 et 3050 mètres d'altitude, avec un substrat rocheux acide (crevasses, éboulis et prairies alpines) (Zhang & Kadereit, 2004). La reproduction est sexuée du fait d'une hétérostylie stricte, et a lieu au début du printemps après la fonte de la neige.

Leur répartition est soit très étendue (*P. hirsuta*), soit restreinte à des massifs spécifique et donc disjointe (Figure 1). C'est le cas pour *P. apennina* ; *P. cottia* et *P. pedemontana*.

Concernant la population de *P. pedemontana* des Écrins, la détermination est difficile à faire, car les individus présentent diverses morphologies, dont une morphologie intermédiaire : une hampe proche de *P. hirsuta* mais une pilosité semblable à *P. pedemontana*. La morphologie foliaire est également variée au sein de cette population avec un bord dentelé selon les individus.

2.2 Échantillonnage

Les données génétiques utilisées lors de cette étude ont été produites dans le cadre de l'étude précédente de Boucher *et al.* (2016). Il s'agit d'un jeu de données composé de 90 individus de toutes les espèces composant *Primula* section *Auricula*, collectés entre avril et septembre 2014. L'identification taxonomique des individus a été réalisée sur le terrain, mais un individu a dû être réattribué après analyse génétique au taxon *P. hirsuta*.

Les individus avaient été séquencés dans le but d'une étude phylogénétique, avec le besoin de récolter de l'information sur l'ensemble du génome, ce qui permet d'observer une histoire phylogénétique non restreinte à un marqueur. A ces fins, l'ADN avait été séquencé par hyRAD (Suchan *et al.* , 2016) car cette technique permet de génotyper le long du génome malgré des mutations sur les sites de restrictions requis par les enzymes utilisées en ddRAD. 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 simple séquençage ddRAD.

Les séquences obtenues par séquençage haut débit avaient été alignées sur un génome de référence partiel de *Primula veris* par le software Bowtie2. Le jeu de données utilisé ('m13_-q20_mincov10') est issu de divers filtres sur l'alignement, avec un score minimal de qualité (Phred) requis de 13 et une couverture minimale de 10 lectures par *locus*. A partir de ces séquences, les SNPs ont été isolés par Freebayes, avec un support de lecture de 30% minimum par allèle, et l'ensemble est exporté sous format VCF.

2.3 Bioinformatique

A partir du jeu de donnée initial, un pipeline est établi dans le langage R (R.Team, 2017) pour générer divers ensembles de données, spécifiques à chaque analyse. Les fonctions sont rassemblées en un package R hébergé sur Github (lien web 1). Dans un premier temps, le fichier initial est traité avec la fonction `subset_reorder`, qui permet de reconstruire le fichier en ne gardant que les individus souhaité dans l'ordre indiqué. Au maximum, les individus conservés à partir du jeu de donnée initial sont les individus de la sous-section *Erythrodrosom* avec : 2 individus de *P. apennina*, 3 individus de *P. cottia*, 2 individus de *P. pedemontana*, 2 individus de *P. pedemontana* des Écrins, 6 individus de *P. hirsuta*, 4 individus de *P. villosa*, 2 individus de *P. daonensis*. Un individu de *P. apennina* n'a pas été conservé comme dans l'étude de Boucher *et al.* (2016), car présentant trop de données manquantes. Toutes les informations concernant les individus de l'analyse sont présentés en annexe 1 et leur répartition est indiquée en Figure 1.

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 le locus pour l'individu présentant cet allèle rare est considéré comme une donnée manquante.

je ne sais pas si c'est l'approche standard, mais ça permet de garder l'information sur les autres individus. Au pire le *locus* devient monomorphique. Et au pire si c'est une erreur, il se peut que le *locus* ai un phred faible, et j'ai un seuil a plus tard pour ça.

Cette étape permet également de supprimer les SNPs qui ont strictement plus d'un variant, qui sont reconnus comme des artefacts des algorithmes utilisés pour appeler les SNPs. 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* polymorphes, nous utilisons ensuite la fonction `clean`.

Les *loci* sont aussi filtrés sur le score de qualité (QUAL) indiqué dans le vcf, qui correspond à la confiance dans l'assignation de l'allèle variant. Une fonction nommée `tri` est appelée pour supprimer les *loci* puis et les individus présentant trop de données manquantes selon les seuils posés.

hum on ne garde pas la notion de suite ? Parce que les deux tris sont consécutifs et non simultanés...peut-être trop de la cuisine ?

Enfin, afin de limiter le déséquilibre de liaison, les sites sont filtrés selon leurs positions sur les contigs du génome de référence : si deux SNPs sont situées à une distance inférieure à n paires de bases, seul le premier est conservé.

Afin de pouvoir utiliser les fichiers triés sous divers format, la sous-sélection d'individus et de *loci* est enregistrée sous quatre formats : `.vcf`, `.str`, `.geno`, `.snp`. La transformation d'un format en un autre se fait respectivement au moyen d'un script bash, par l'utilisation du software PGDSpider (Lischer & Excoffier, 2012), grâce au package LEA (Frichot & François, 2015) et avec un script R. La production de ces fichiers est inscrite dans le pipeline par la fonction `files`.

Toutes ces fonctions sont indépendantes mais peuvent être appelées dans le bon ordre au moyen de la fonction `dataset`, qui prend en entrée un vecteur de noms d'individus, un csv contenant l'assignation de chaque individu à une population ou une lignée, le fichier vcf original, le nom des fichiers de sortie et les seuils précisés au-dessus.

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 SNPs 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).

2.4 Génétique des populations

Les analyses de génétique des populations ont été réalisés sous Rstudio (R.Team, 2017) avec différents packages.

Dans un premier temps, des statistiques F ont été calculées avec le package `adegenet` (Jombart & Ahmed, 2011). Ces statistiques permettent de se faire une première idée du jeu de données au niveau de la sous-section *Erythrodrosum*, où le F_{st} permet d'apprécier la présence d'une structure génétique entre les différentes populations. En parallèle, un clustering de nos individus a été réalisé afin d'essayer d'attribuer des groupes sans a-priori de populations. Ce clustering fait par `adegenet` trouve les groupes pour lesquels la variance intergroupes est maxi-

male quand la variance intragroupes est minimale. En plus de ce clustering, une analyse par composante principale discriminante a été réalisée pour visualiser la répartition des individus au sein des groupes.

Afin d'observer plus spécifiquement le clade 'Hirsuta', sa structure a ensuite été étudiée par le package LEA (Frichot & François, 2015). Cette structuration est établie par un algorithme sNMF, et permet d'estimer des coefficients d'admixture pour tout les individus. Pour cette étude, la structuration a été modélisée pour des structures de K populations allant de 2 à 5, avec 20 simulations par K et un alpha de 10, en ne retenant que la meilleure simulation basée sur le critère de "cross-entropy". Les structures avec $K > 5$ ne sont pas étudiées, car le clade 'Hirsuta' ne contient que 5 populations et que ces populations sont échantillonnées avec peu d'individus. Un K plus grand reviendrait à structurer ces populations en assignant un individu par population.

2.5 Hypothèse d'admixture

Afin de tester l'hypothèse d'admixture pour le taxon des Écrins, un test "ABBA-BABA" a été réalisé. Ce test d'admixture développé par Durand *et al.* (2011) propose une statistique (D) basée sur quatre lignées partageant un ancêtre commun, selon la fréquence de SNPs observés avec un motif particulier. Considérant que les *loci* sont sous une évolution neutre et sans déséquilibre de liaison, on attend deux configurations différentes pour une même topologie. Cette topologie [(((P1,P2),P3),O)] propose ~~que "P1" et "P2" coalescent avant un autre événement de coalescence avec "P3", puis avec "O" l'outgroup~~ une divergence basale entre l'outgroup "O" et les autres taxons, suivie de la divergence de "P3" et finalement une divergence entre "P1" et "P2".

Sur cette topologie on attend deux allèles présents en fin de branche avec "A" l'allèle ancestral et "B" l'allèle alternatif porté par l'outgroup. Ce test ne s'intéresse qu'à deux cas : "ABBA" et "BABA", ce qui correspond à une minorité de *loci*. Sous un modèle neutre, on attend des proportions équilibrées de *loci* portant ces deux configurations. L'hypothèse alternative est qu'un déséquilibre de ces proportions peut être induit par deux cas : une topologie autre ou alors une introgression de "P3" avec "P1" ou "P2". Une introgression entre "P3" et "P1" verrait donc une plus grande proportion de *loci* à la configuration "BABA" et une valeur négative du "D"

Il a été décidé de calculer D à partir des fréquences alléliques, comme décrit dessous. P_{ij} correspond à l'allèle alternatif.

$$D(P1, P2, P3, 0) = \frac{\sum_{i=1}^n (1 - P_{i1})P_{i2}P_{i3}(1 - P_{i4}) - P_{i1}(1 - P_{i2})P_{i3}(1 - P_{i4})}{\sum_{i=1}^n (1 - P_{i1})P_{i2}P_{i3}(1 - P_{i4}) + P_{i1}(1 - P_{i2})P_{i3}(1 - P_{i4})}$$

Un intervalle de confiance sur cette statistique est obtenu par bootstrapping des sites. Ce bootstrapping a été réalisé 1000 fois par échantillonnage aléatoire avec remise des *loci*. Il est important de souligner que ce test ne permet en aucun cas de proposer un sens d'introgression, ni son intensité.

Une autre approche de l'admixture entre les taxons du clade 'Hirsuta' a été réalisée par le package introgress (Gompert & Alex Buerkle, 2010), utilisé pour mesurer l'index h, index d'hybridation, entre *P. pedemontana s.l.* et *P. hirsuta*. Ce package requiert l'import des deux taxons parents présumés et des "hybrides" afin de mesurer un coefficient d'admixture (ou h-index). ~~Ce coefficient basé sur une fonction de vraisemblance (Buerkle, 2005), où qui correspond à~~ la proportion de chaque génome parental ~~est mesurée dans (Buerkle, 2005) chez~~ l'individu hybride.

ça vaut peut être le coup de m'attarder sur comment il mesure ça. Si j'explique pas là et que j'ai à l'oral, je sais pas encore comment l'expliquer

Les marqueurs codominants utilisés ici n'étant pas tous fixés chez les parents, le calcul du h-index se basera ici sur les fréquences alléliques.

3 Résultats

3.1 Tri du jeu de données

Le jeu de donnée initial contient 175799 *loci*. Pour le jeu de donnée utilisé au niveau de la sous-section *Erythrodrosum*, 2078 *loci* ont été conservés pour les analyses, avec les seuils suivants : allèle rare 5%, Phred de 20, une distance de 10kb minimum entre deux *loci* et 5% des individus sans information par *locus*. Pour le jeu de donnée sur le clade 'Hirsuta', 1851 *loci* ont été conservés pour les analyses, avec les mêmes seuils.

Le changement de taille du jeu de donnée est dû à un nombre plus petit d'individus dans le second cas, qui rend des *loci* monomorphes. Le seuil le plus strict est le seuil sur les données manquantes par *locus*, qui ne conserve que respectivement 7.05% et 7.07% du jeu de donnée.

3.2 Sous-section *Érythrodrosum*

Les analyses de génétique des populations ont donc été faites sur **beaucoup** de *loci*. Dans un premier temps, le F_{st} général de la sous-section *Erythrodrosum* donne une valeur de 0.15, ce qui confirme la présence d'une structure génétique. Cette structure était attendue étant donné qu'il s'agit de plusieurs espèces à l'échelle des Alpes, mais pour autant il est important de remarquer que la valeur reste faible (> 0.5).

proposer une explication ? peut être parce qu'il y a une sous structure ? quid du F_{is} ;0

Les F_{st} par paire entre populations sont également tous supérieurs à zéro (valeurs comprises entre **tant et tant**), et la topologie par neighbor-joining rejoint celle décrite précédemment (Figure 2).

Cette structure est par ailleurs confirmée par le clustering réalisé par adegenet sans a-priori de populations. Ainsi, le nombre K de cluster avec le BIC le plus faible est atteint pour $K=3$. Pour cette valeur de K, les groupes sont corrélés à la géographie, avec un clade 'est-alpin' composé de *P. daonensis* et *P. villosa*, un clade composé de l'espèce *P. hirsuta* et enfin *P. pedemontana s.l.*.

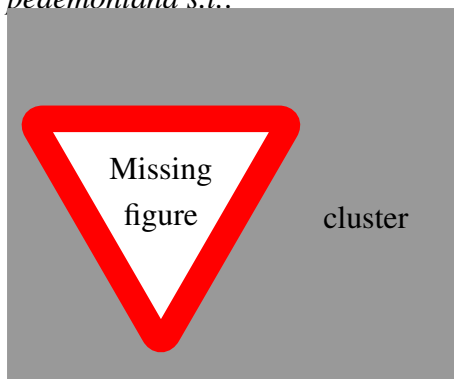


FIGURE 3 – **cluster** cluster

Ce clustering est plus facilement visualisable sous la forme d'une analyse en composante principale comme dans la figure 4. Les deux premières composantes montrent ce clustering avec quelques précisions de plus. Ainsi, *P. pedemontana s.l.* semble bien plus robuste que le clade 'est-alpin', avec des individus plus proches. Cependant, la population des Écrins n'est pas aussi groupée que le reste des populations composant *P. pedemontana s.l.*, et s'en éloigne vers *P. hirsuta*.

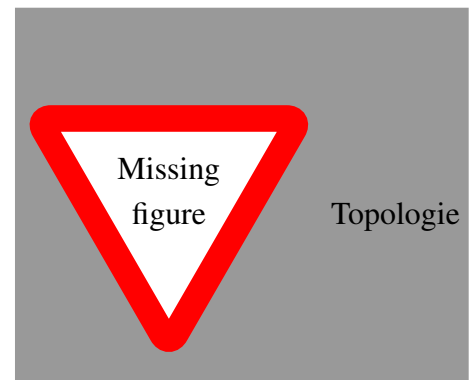


FIGURE 2 – **Topologie** faite avec les F_{st}

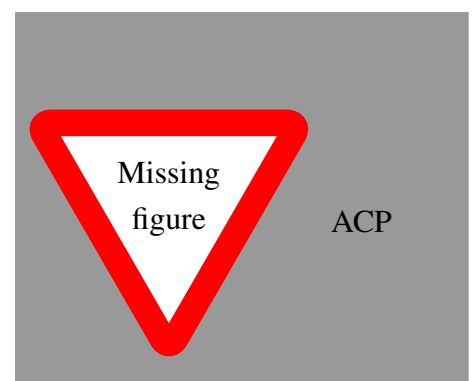


FIGURE 4 – **ACP** ACP

3.3 Clade ‘Hirsuta’

La structuration du clade ‘Hirsuta’ pour différents K permet de voir différentes informations (figure 5. Pour K=2, une séparation est déjà nette entre deux clades, conformément aux résultats précédents, entre *P. pedemontana s.l.* et *P. hirsuta*. Cette séparation présente néanmoins une légère trace d’admixture entre les individus des Écrins et *P. hirsuta*. L’individu de *P. hirsuta* présentant un peu d’admixture a été échantillonné en Belledonne, un massif proche des Écrins (figure 1).

Pour K=3, les individus des Écrins sont isolés et une admixture entre ces individus et les autres individus de *P. pedemontana s.l.*. Cette structure retrouvée ici reflète ce qui était observé sur l’ACP, où les individus des Écrins se regroupaient avec *P. pedemontana s.l.* tout en étant excentré.

Pour K=4, *P. pedemontana s.l.* se retrouve éclaté avec les différentes populations échantillonnées dans les divers massifs. Cependant, les deux espèces *P. apennina* et *P. cottia* sont toujours regroupées, même si cet ensemble n’est pas robuste. En effet, pour K=5, c’est *P. hirsuta* qui se retrouve scindé en deux avec d’un côté l’individu des Pyrénées et de l’autre l’ensemble des individus. Ici la structure de *P. pedemontana s.l.* est plus ambiguë, même si les individus des Écrins sont toujours isolés.

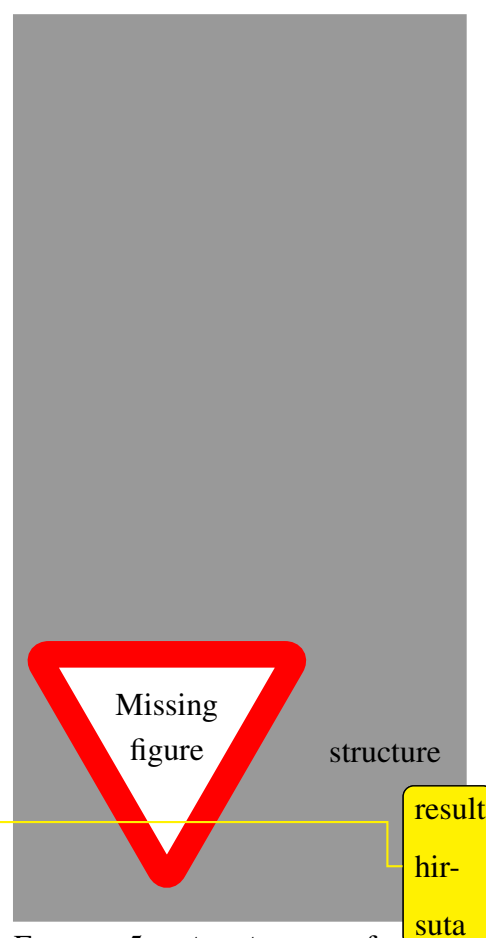


FIGURE 5 – **structure** snmf pour k de 2 a 5

3.4 Admixture

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.

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.

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

| Header-1 | Header-1 | Header-1 |
|----------|----------|----------|
| 2 | 3 | 5 |
| 2 | 3 | 5 |
| 2 | 3 | 5 |

TABLE 1 – ABBA-BABA

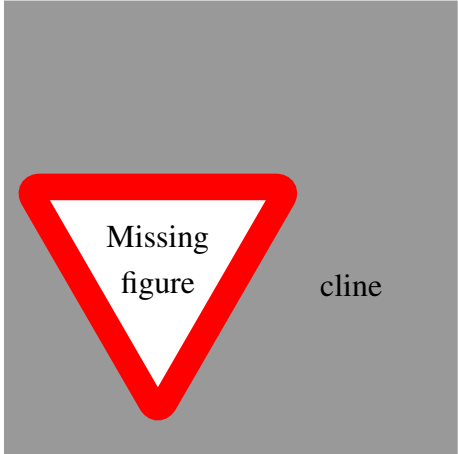


FIGURE 6 – **cline** hirsuta pedemontana s.l.

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.

result
ad-
mix-
ture

4 Discussion

C'est assez violent de commencer directement la discussion comme ça!

Tu présentes les choses de manière beaucoup trop négative. S'il s'agissait de ta thèse c'est sûr qu'on pourrait te critiquer pour ce choix d'échantillonnage, mais là la situation est bien différente, comme je te l'avais expliqué clairement quand tu m'as demandé ce stage. Pas besoin de te tirer une balle dans le pied d'entrée, il faut simplement expliquer que les données que tu as utilisées ont été collectées dans un autre but (phylogénétique et à plus large échelle) et qu'elles ne sont donc pas très adaptées à des analyses de génét des pops. Ensuite, il faut dire que justement, tu as fait une étude préliminaire qui permet d'aiguiller l'échantillonnage à venir. L'important c'est que tu as réussi à tirer de l'info pertinente de tout ça et que ça nous éclaire un peu plus!

Ça peut paraître une simple question d'écriture, mais c'est très important pour deux raisons : (1) c'est la vérité et (2) tu décourages tout de suite toute attaque sur l'adéquation des données : ce n'est pas ton problème mais le mien.

De manière générale il ne faut jamais commencer une discussion par les limitations, tu perds le lecteur direct qui ne croira pas du tout à tout ce que tu vas raconter par la suite. Au contraire il faut résumer les résultats principaux

Une grande limitation de cet étude provient de l'échantillonnage des populations. Ainsi, si l'échantillonnage actuel est suffisant pour les précédentes analyses phylogénétiques, la génétique des populations requiert davantage d'informations par populations. La plupart des analyses de génétiques des populations sont basées sur les variations de fréquences alléliques au sein des populations. La faible taille d'échantillon vient donc biaiser fortement ce prérequis, car plus la taille de l'échantillon est faible, moins l'estimation de la fréquence allélique représente la fréquence réelle au sein de la population. De fait, les potentielles structures génétiques des espèces étudiées ne peuvent être étudiées avec précision. Or, dans le cadre de cette étude,

différentes structures génétiques sont imbriquées pour les populations étudiées. Par exemple la population de *P. apennina* échantillonnée ici présente une structure génétique à l'échelle du massif des Apennins (Crema *et al.* , 2009), mais représente aussi une partie de la structure des populations d'Alpes de l'ouest. Enfin, cette population doit représenter une entité structurale pour la sous-section Erythrodrosum à l'échelle de l'arc Alpin. Si l'on ne considère que deux individus séquencés pour cette population, il est donc impossible d'avancer des analyses robustes pour les échelles les plus fines de structure. L'étude est donc de fait limitée à la plus grande échelle possible, sans possibilité de résolution à l'échelle des Alpes de l'ouest. Il est d'ailleurs fort probable que ce soit ce manque de résolution qui ait pénalisé les analyses par inférence bayésienne.

Un meilleur échantillonnage permettrait d'avoir suffisamment d'information génétique pour observer plus finement les statistiques F des populations. Plus d'échantillons permettrait également de pouvoir utiliser pleinement DIYABC et inférer une taille de population ou alors des paramètres historiques pour dater les événements biologiques. Enfin, un plus grand échantillon pourrait résoudre l'introgression entre *P. hirsuta* et le taxon des Écrins, afin de déterminer le sens de cette introgression et quelle quantité du génome a été modifié par cet événement, à l'aide d'un test du même type que ABBA-BABA mais avec P3 scindé en deux populations (Eaton *et al.* , 2015).

Autre aspect du rééchantillonnage, la délimitation des populations au niveau géographique est un prérequis à un bon échantillonnage génétique. Dans cette étude, les individus sont échantillonnés sur des distances parfois trop importantes (*P. pedemontana*) ou trop courtes (*P. pedemontana* des Écrins ou alors *P. apennina*).

Marque page de fin de doc = 12!!

discussion
com-
plete

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.
- Buerkle, C. Alex. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, **5**(3), 684–687.
- Crema, Silvia, Cristofolini, Giovanni, Rossi, Martina, & Conte, Lucia. 2009. High genetic diversity detected in the endemic *Primula apennina* Widmer (Primulaceae) using ISSR fingerprinting. *Plant Systematics and Evolution*, **280**(1-2), 29–36.
- Durand, Eric Y., Patterson, Nick, Reich, David, & Slatkin, Montgomery. 2011. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**(8), 2239–2252.
- Eaton, Deren A.R., Hipp, Andrew L., González-Rodríguez, Antonio, & Cavender-Bares, Jeanine. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, **69**(10), 2587–2601.
- Frichot, Eric, & François, Olivier. 2015. LEA : An Rpackage for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**(8), 925–929.
- Gompert, Zachariah, & Alex Buerkle, C. 2010. Introgress : A software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**(2), 378–384.
- Jombart, Thibaut, & Ahmed, Ismaïl. 2011. adegenet 1.3-1 : New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**(21), 3070–3071.
- Lischer, H. E L, & Excoffier, L. 2012. PGDSpider : An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**(2), 298–299.
- 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.
- Ozenda, Paul. 1995. L'endémisme au niveau de l'ensemble du Système alpin. *Acta Botanica Gallica*, **142**(7), 753–762.

- R.Team. 2017. R : A language and environment for statistical computing (Version 3.4. 2)[Computer software]. *Vienna, Austria : R Foundation for Statistical Computing*.
- Suchan, Tomasz, Pitteloud, Camille, Gerasimova, Nadezhda S., Kostikova, Anna, Schmid, Sarah, Arrigo, Nils, Pajkovic, Mila, Ronikier, Michal, & Alvarez, Nadir. 2016. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS ONE*, **11**(3), 1–22.
- Zhang, Li Bing, & Kadereit, Joachim W. 2004. Classification of *Primula* sect. *Auricula* (Primulaceae) based on two molecular data sets (ITS, AFLPs), morphology and geographical distribution. *Botanical Journal of the Linnean Society*, **146**(1), 1–26.
- Zhang, Li-Bing, Comes, Hans Peter, & Kadereit, Joachim W. 2004. The Temporal Course of Quaternary Diversification in the European High Mountain Endemic *Primula* sect. *Auricula* (Primulaceae) . *International Journal of Plant Sciences*, **165**(1), 191–207.

6 Ressources

Web 1 – <https://github.com/gowachin/Pedemontana>

| Species | Locality | Code | Morph | Collector | Date | Longitude | Latitude | Altitude | Reads raw | Reads trimmed | Voucher |
|-----------------------------|--------------------------------|------|--------------|-------------------------------|----------|-----------|----------|----------|-----------|---------------|---------|
| P. apennina* | Sella del Marmagna, Italy | AMB | Short-styled | F. Boucher/L. Gallien | 30/05/14 | 10.00575 | 44.3978 | 1610 | 6885928 | 6486849 | Photo |
| P. apennina | Monte Marmagna, Italy | AML | Long-styled | F. Boucher/L. Gallien | 30/05/14 | 9.99731 | 44.39672 | 1825 | 1856867 | 1663377 | Photo |
| P. apennina | Monte Orsaro, Italy | AOL | Long-styled | F. Boucher/L. Gallien | 30/05/14 | 9.99666 | 44.39883 | 1818 | 3494081 | 3230296 | Photo |
| P. cottia | Below locus classicus, Italy | CS1 | NA | F. Boucher | 23/07/14 | 7.0716 | 44.9271 | 1159 | 5127416 | 4814386 | Photo |
| P. cottia | Prali, locus classicus, Italy | CP1 | NA | F. Boucher | 23/07/14 | 7.06583 | 44.9186 | 1407 | 3160322 | 2941542 | Photo |
| P. cottia | Prali, locus classicus, Italy | CP4 | NA | F. Boucher | 23/07/14 | 7.06583 | 44.9186 | 1407 | 3482252 | 3201012 | Photo |
| P. daonensis | Passo di Gavia, Italy | DGB | Short-styled | F. Boucher/L. Gallien | 27/05/14 | 10.49701 | 46.31843 | 2219 | 6095146 | 5757485 | Photo |
| P. daonensis | Ritorto, Italy | DRL | Long-styled | F. Boucher/L. Gallien | 27/05/14 | 10.80429 | 46.23149 | 2083 | 4607717 | 4299840 | Photo |
| P. hirsuta | Malga Bordonolona, Italy | DMB | Short-styled | F. Boucher | 09/06/14 | 10.87383 | 46.43412 | 2214 | 6073360 | 5722516 | Photo |
| P. hirsuta | Refuge du Couvercle, France | HC1 | NA | C. Dentant | 15/07/14 | 6.9656 | 45.9103 | 2649 | 2639620 | 2384576 | NA |
| P. hirsuta | Grand Chat, France | HGL | Long-styled | F. Boucher/L. Gallien | 18/05/14 | 6.2147 | 45.4467 | 1986 | 4583323 | 4270118 | Photo |
| P. hirsuta | Steibensee, Switzerland | HS2 | NA | F. Boucher | 07/09/14 | 8.17 | 46.45 | 2414 | 2891228 | 2643122 | Photo |
| P. hirsuta | Pic du Midi d'Ossau, France | HP1 | NA | C. Roquet | 15/08/14 | -0.4381 | 42.8431 | 2739 | 1881282 | 1772502 | NA |
| P. hirsuta | Passo del Bernina, Switzerland | HPB | Short-styled | F. Boucher | 09/06/14 | 10.02717 | 46.41069 | 2328 | 6566463 | 6195227 | Photo |
| P. pedemontana | Barrage de Tignes, France | PT1 | NA | F. Boucher | 27/07/14 | 6.94633 | 45.4805 | 1836 | 6515454 | 6086010 | YES |
| P. pedemontana | Vallon d'Avérole, France | PV1 | NA | F. Boucher | 27/07/14 | 7.08707 | 45.29356 | 2144 | 6480484 | 6100355 | YES |
| P. sp. Lauzon Valley | Lauzon Valley, France | GA2 | NA | P. Salomé/R. Bonet/F. Boucher | 25/07/14 | 6.2784 | 44.8418 | 1732 | 4150458 | 3873470 | YES |
| P. sp. Lauzon Valley | Lauzon Valley, France | GA4 | NA | P. Salomé/R. Bonet/F. Boucher | 25/07/14 | 6.2773 | 44.8366 | 1899 | 4796528 | 4489119 | YES |
| P. villosa ssp. irringardis | Rappolt Kogel, Austria | VR3 | Short-styled | F. Boucher | 07/06/14 | 14.88541 | 47.08313 | 1871 | 4722814 | 4234314 | Photo |
| P. villosa ssp. irringardis | Rappolt Kogel, Austria | VR1 | Long-styled | F. Boucher | 07/06/14 | 14.88541 | 47.08313 | 1871 | 4789459 | 4283316 | Photo |
| P. villosa ssp. villosa | Turracher Hohe, Austria | VL2 | Long-styled | F. Boucher | 07/06/14 | 13.87581 | 46.91273 | 1801 | 3227112 | 2776420 | Photo |
| P. villosa ssp. villosa | Turracher Hohe, Austria | VB1 | Short-styled | F. Boucher | 07/06/14 | 13.87581 | 46.91273 | 1801 | 3004397 | 2708593 | Photo |

Annexe 1 – Individus séquencés pour cette étude lorem ipsum