

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

La dynamique des Alpes au Quaternaire a permis l'apparition d'une flore très diverse. Cependant, cette diversité porte les traces de multiples événements génétiques difficiles à différencier. Cette étude propose de réanalyser un jeu de données phylogénétique sur la sous-section *Erythrodrosom*, située au sein d'un large groupe de fleur de montagnes endémiques du système alpin : la section *Auricula* du genre *Primula*. Pour cela, les *loci* produits par séquençage à haut débit avec capture par sondes ARN, et alignés sur un génome de référence, sont re-triés pour obtenir un jeu de donnée analysable par les outils de génétique des populations. Ce nouveau jeu de donnée permet de valider les résultats précédemment établis sur la structure génétique de *P. pedemontana*, proposant une révision taxonomique, ainsi que la confirmation d'un flux de gène entre cette espèce et *P. hirsuta* dans le massif des Écrins. Finalement, cette étude préliminaire recentre les futures campagnes d'échantillonnage dans les massifs des Écrins et de la Vanoise, dans le but de clarifier l'impact des cycles de glaciations sur la génétique des fleurs de montagnes.

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>Erythrodrosum</i>	8
3.3	Clade 'Hirsuta'	9
3.4	Détection de flux de gènes	10
4	Discussion	10
4.1	Différentes structures génétiques	11
4.2	Flux de gènes et glaciations	11
4.3	Conclusion	12
5	Bibliographie	13
6	Ressources	15

1 Introduction

Par la diversité des milieux qu'elle abrite, la chaîne des Alpes représente une zone géographique avec une grande diversité floristique, dont une grande richesse endémique (Ozenda, 1995). Cependant, cette biodiversité possède une histoire complexe liée à l'histoire de cette orogénèse et les cycles glaciaires qui ont joué sur l'Holarctique au quaternaire. Ainsi, étudier l'évolution des végétaux présents dans un environnement très dynamique s'avère compliqué. En effet, si la spéciation peut-être facilitée par des répartitions allopatriques dues à l'extension glaciaire ou une spécialisation à différentes niches écologiques apparaissant au gré des cycles géoclimatiques, des contacts secondaires avec le retrait glaciaires sont propices aux flux de gènes et hybridations.

Ces différentes actions de la dynamique du paysage sur la structure génétique des espèces alpines ont donc parfois aboutit à des espèces cryptiques, dont le statut taxonomique est difficile à déterminer d'après les critères morphologiques. Traitant une plus grande quantité d'information autrefois inaccessible, la génétique a permis d'observer des phénomènes biologiques sur des critères plus fins et surtout préciser l'histoire évolutive de ces espèces. Ces nouvelles informations ont donc remis en question de nombreuses classifications. Cela a aboutit à une plus grande diversité taxonomique, actuellement retravaillée avec l'arrivée des techniques de génomique.

C'est notamment le cas pour la section *Auricula* du genre *Primula*. Parmi les groupes de plantes les plus riches du système Alpin Européen (i.e., l'ensemble des montagnes d'Europe (Ozenda, 1995) , cette section regroupe pas moins de 26 espèces (Zhang & Kadereit, 2004). Cette section s'étant fortement diversifiée durant les 5 derniers millions d'année (Boucher *et al.* , 2016b, Zhang *et al.* , 2004), elle ne permet pas une grande résolution phylogénétique à cause du peu de marqueurs différenciés. De plus, la présence de nombreux hybrides possibles remet en cause les rangs d'espèces proposés sur divers taxons (Kadereit *et al.* , 2011). C'est par exemple le cas dans la sous-section *Erythrodrosom* Pax, où une population d'une vallée du massif des Écrins présente une morphologie intermédiaire entre *P. pedemontana* et *P. hirsuta*.

L'origine de ce taxon reste inconnue et sa distribution géographique jouxte celles de différentes espèces proches (Figure 1). De plus, une révision taxonomique de ces espèces voisines a récemment été proposée, les regroupant sous une seule entité taxonomique au rang d'espèce, nommée ci-dessous *P. pedemontana* s.l. (Boucher *et al.* , 2016a).

La faible résolution phylogénétique obtenue jusqu'alors dans ce groupe étant probablement

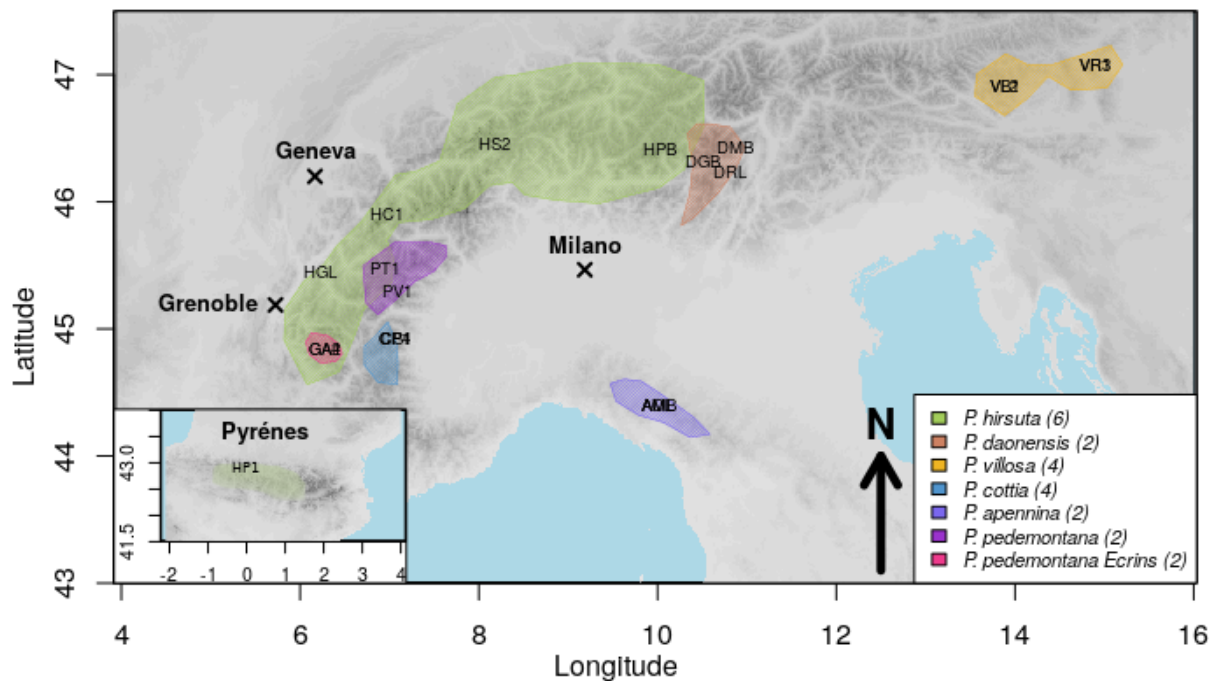


FIGURE 1 – Carte de répartition de *Primula* sect. *Auricula* Duby subsect. *Erythrodrosom* Pax. Espèces composant la sous-section, avec entre parenthèse le nombre d'échantillons utilisés dans cette étude. Les aires de répartition sont extrapolées depuis les observations de l'INPN et du CBNA. Les codes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

due à son origine récente (Boucher *et al.*, 2016b), la génétique des populations pourrait permettre d'éclairer le statut taxonomique de ces primevères. En effet, même si les données génétiques utilisées par Boucher *et al.* (2016a) ont été échantillonnées pour résoudre la phylogénie d'un groupe plus large, cette information sur tout le génome pourrait permettre d'estimer la structure génétique et les flux de gènes entre les différents massifs alpins. Ces premières estimations sont également nécessaires pour orienter de futures campagnes d'échantillonnage et approfondir la connaissance de cette histoire biogéographique façonnée par les cycles glaciaires.

Les objectifs de cette étude sont donc : (i) d'étudier la sous-section *Erythrodrosom* à l'aide d'outils de génétique des populations ; (ii) de clarifier la structure taxonomique de *P. pedemontana* s.l. ; (iii) de tester l'hypothèse d'admixture du taxon des Écrins.

2 Matériels et méthodes

2.1 Cas d'étude

La section *Auricula* Duby du genre *Primula* comprend 26 espèces et est endémique du système Alpin européen (Ozenda, 1995). Au sein de cette section, la sous-section *Erythrodosum* englobe 7 espèces : *P. villosa*; *P. daonensis*; *P. hirsuta*; *P. apennina*; *P. cottia* et *P. pedemontana*. L'ensemble de ces espèces est ancestralement hexaploïde mais du fait de l'ancienneté de cet événement de polyploïdisation, elles sont considérées ici comme diploïdes avec un nombre de chromosomes de $x = 33$. Leur niche écologique est située entre 1500 et 2500 mètres d'altitude (étages subalpins et alpins), et toutes sont spécialistes des substrats rocheux acides (crevasses, éboulis et prairies alpines) (Zhang & Kadereit, 2004). La reproduction est allogame 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 spécifiques 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, 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 (<http://www.ecrins-parcnational.fr/actualite/lenigme-de-la-primevere-du-valgaudemar>).

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.* (2016a). 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 au taxon *P. hirsuta* après analyse génétique.

Les individus avaient été génotypés dans le cadre d'une étude phylogénétique, avec le besoin de récolter de l'information sur l'ensemble du génome pour estimer une histoire phylogénétique non restreinte à un seul marqueur. A ces fins, l'ADN avait été génotypé par la méthode 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 RAD-seq (Davey &

Blaxter, 2010). En effet les enzymes de restrictions utilisées en RAD-seq sont sensibles à la mutation d'un nucléotide, tandis que les sondes ARN utilisés pour la capture de loci dans la méthode hyRAD peuvent s'hybrider sur ces sites de restriction RAD visés, même mutés. La nécessité de capturer des sites malgré une variance génétique 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 le génome de référence de *Primula veris* par le logiciel 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 identifiées en utilisant Freebayes, avec un support de lecture de 30% minimum par allèle, et ont été exportés sous format VCF.

2.3 Bioinformatique

A partir du jeu de données 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 souhaités dans l'ordre indiqué. Au maximum, les individus conservés à partir du jeu de données 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.* (2016a), 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.

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. 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 les individus présentant trop de données manquantes selon les seuils posés. Enfin, afin de limiter le déséquilibre de liaison, les sites sont filtrés selon leur position 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 tableau 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 (22 au total), 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 réalisées sous R 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 *Erythrodrosom*, 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*. Ce clustering fait par `adegenet` trouve les groupes pour lesquels la variance intergroupe est maximale quand la variance intragroupe 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 l'ensemble d'espèces de l'ouest des Alpes (*P. hirsuta*; *P. apennina*; *P. cottia*; *P. pedemontana* et *P. pedemontana des Écrins*, nommé ci-dessous 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 un nombre de groupes K 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 taxons nommés, chacun n'étant représenté que par quelques individus seulement dans cet échantillonnage. Un K plus grand reviendrait à structurer ces lignées en assignant un individu par groupe.

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 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" (Figure 2).

Sur cette topologie on attend deux allèles présents en fin de branche avec "A" l'allèle ancestral et "B" l'allèle alternatif porté par certains individus de l'ingroup. 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 une introgression de "P3" avec "P1" ou "P2". Une introgression entre "P3" et "P2" verrait donc une plus grande proportion de *loci* à la configuration "ABBA" et une valeur positive du "D".

Il a été décidé de calculer D à partir des fréquences alléliques, comme décrit dessous. P_{ij} correspond à la fréquence de 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 SNPs. 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,

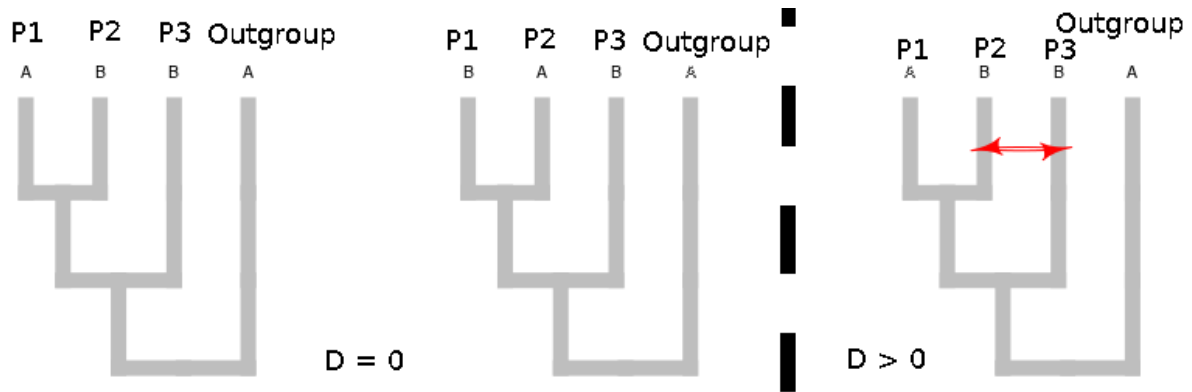


FIGURE 2 – **Topologies des *loci* inclus dans le test "ABBA-BABA"**. Les deux topologies de gauche représentent les configurations en équilibre sous un modèle neutre ($D=0$). La topologie de droite explicite un cas d'admixture entre "P1" et "P3", ce qui induit une plus grande proportion de la configuration schématisée (ABBA).

ni son intensité. Afin de vérifier l'absence de faux positifs, ce test sera effectué avec des taxons pour lesquels aucune hypothèse d'admixture n'existe (*P. apennina* et *P. cottia*). De plus, pour ne pas biaiser l'analyse avec l'*a priori* que le taxon des Écrins est issue d'une lignée de *P. pedemontana*, l'admixture sera testée entre tous les taxons de *P. pedemontana* s.l. et *P. hirsuta*. Ce besoin de considérer toutes les entités taxonomiques présentes vient du fait que pour identifier la source de l'introggression, il faut considérer toutes les lignées dérivées de l'ancêtre commun (Eaton *et al.*, 2015).

3 Résultats

3.1 Tri du jeu de données

Le jeu de données initial contient 175 799 *loci*. Pour le jeu de données utilisé au niveau de la sous-section *Erythrodromus*, 2 078 *loci* ont été conservés pour les analyses, avec les seuils suivants : fréquence des allèles rares $> 5\%$, Phred > 20 , une distance de 10kb minimum entre deux *loci* et 5% des individus sans information maximum par *locus*. Pour le jeu de données sur le clade 'Hirsuta', 1 851 *loci* ont été conservés pour les analyses, avec les mêmes seuils.

Le changement de taille du jeu de données 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 données.

3.2 Sous-section *Erythrodrosum*

Dans un premier temps, le F_{st} général entre les différents taxons de la sous-section *Erythrodrosum* donne une valeur de 0.15.

Les F_{st} entre paires d'espèces sont également tous supérieurs à zéro (valeurs comprises entre 0.093 et 0.214), et les relations phylogénétiques entre espèces obtenues par neighbor-joining à partir de ces valeurs de F_{st} rejoignent celles décrites précédemment (Figure 3). Cette structure est par ailleurs confirmée par le clustering réalisé par adegenet sans *a priori* de populations. Ainsi, le nombre K de clusters avec le critère d'information bayésien le plus faible est atteint pour K=3. La valeur de BIC la plus proche est celle de K=2, avec une différence de 0.41 (K=3 119.90; K=2 120.31). La valeur de K=4 présente également un BIC très proche (K=4 120.83). Pour la valeur de K=3, 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.*.

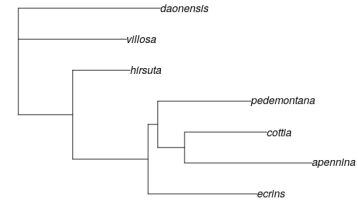


FIGURE 3 – **Topologie d'*Erythrodrosum* par Neighbor-joining**, réalisé à partir de la matrice de distances F_{st} par paires de populations.

Ce clustering est plus facilement visualisable sous la forme d'une analyse en composante principale discriminante comme dans la figure ???. Les deux premières composantes montrent ce clustering avec quelques précisions de plus. Ainsi, les individus de *P. pedemontana s.l.* sont plus proches que ceux du clade 'est-alpin'. 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*.

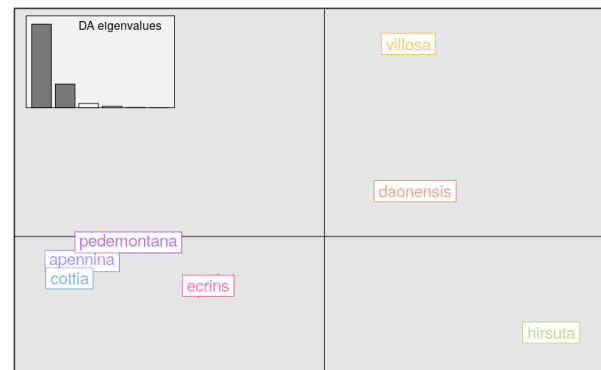


FIGURE 4 – **Analyse en composante principale discriminante de la sous-section *Erythrodrosum* Pax.** Les groupes *a priori* sont les populations échantillonnées.

3.3 Clade ‘Hirsuta’

La structuration du clade ‘Hirsuta’ pour différentes valeurs de 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 la DAPC, 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 retrouvé pour K=5. C’est alors *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 donc plus ambiguë, même si les individus des Écrins sont toujours isolés. Il est donc remarquable qu’au delà de K=3, aucune structure regroupant durablement des individus est observable.

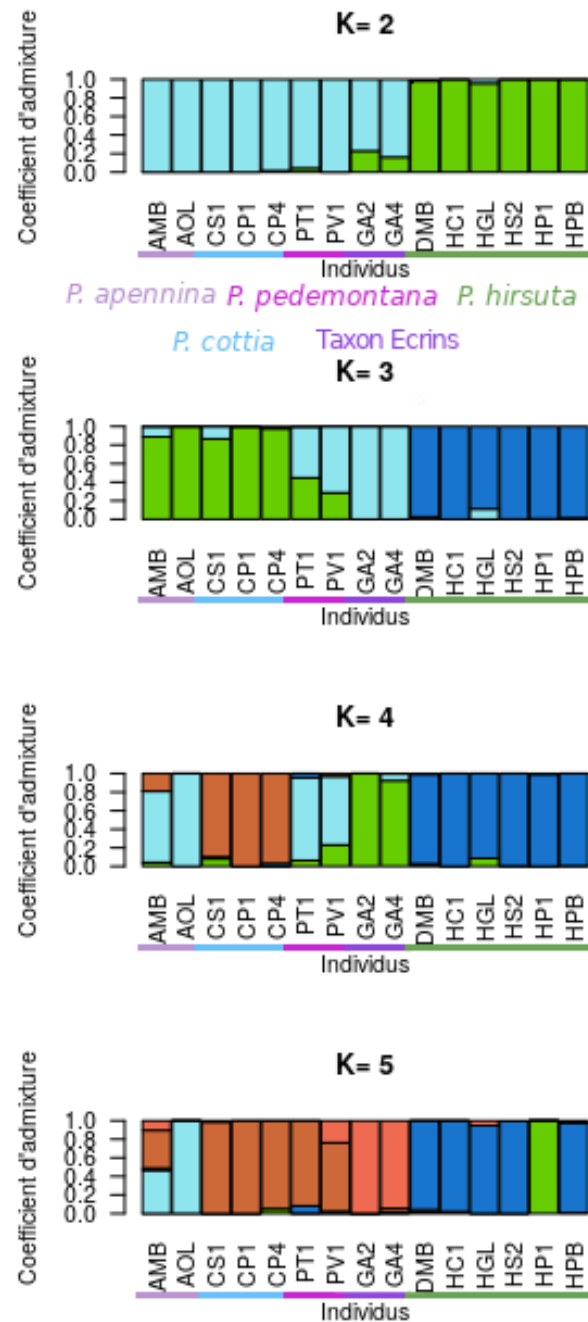


FIGURE 5 – **Structuration du clade ‘Hirsuta’ par sNMF.** Les K choisis vont de 2 à 5, avec 20 répétition par K et choix du meilleur run sur critère de cross-entropy.

3.4 Détection de flux de gènes

Le test d'admixture entre les différentes populations de *P. pedemontana s.l.* et *P. hirsuta*, avec *P. daonensis* en outgroup confirment l'admixture entre le taxon des Écrins et *P. hirsuta*. En effet, le D estimé est supérieur à 0 avec une moyenne de 0.102. A contrario, le test ne permet pas d'estimer un D différent de 0 pour les autres populations, ce qui reflète les observations précédentes en sNMF.

Cependant, il est important de noter que ce test ne permet pas d'estimer quelle population de *P. pedemontana s.l.* s'est admixté avec *P. hirsuta*, vu que le résultat est positif quelque soit la population sélectionnée en P1.

P1	P2	P3	Outgroup	D	p-valeur
<i>P. pedemontana</i>	<i>P. apennina</i>	<i>P. hirsuta</i>	<i>P. daonensis</i>	-0.027	0.212
<i>P. pedemontana</i>	<i>P. cottia</i>	<i>P. hirsuta</i>	<i>P. daonensis</i>	-0.033	0.0772
<i>P. pedemontana</i>	Taxon des Écrins	<i>P. hirsuta</i>	<i>P. daonensis</i>	0.082	5.761.10⁻⁵
<i>P. apennina</i>	Taxon des Écrins	<i>P. hirsuta</i>	<i>P. daonensis</i>	0.109	6.465.10⁻⁷
<i>P. cottia</i>	Taxon des Écrins	<i>P. hirsuta</i>	<i>P. daonensis</i>	0.116	1.237.10⁻⁸

TABLE 1 – **Test d'admixture par ABBA-BABA.** La p-valeur est estimée à partir d'un bootstrap sur les *loci* et recalcul de la valeur D. Un D supérieur à 0 indique une admixture entre P2 et P3.

4 Discussion

En réanalysant le jeu de donnée d'une précédente étude de phylogénétique, il été ici possible de conforter des résultats et d'en apporter de nouveaux concernant la génétique des populations étudiées. Ce passage de la phylogénétique à la génétique des populations sur un même jeu de donnée est permis grâce aux marqueurs utilisés et a autorisé l'analyse des données sous des angles différents. A la différence des marqueurs couramment utilisés en génétique des populations (comme les microsatellites), les SNPs peuvent être génotypés en grand nombre. Ce nombre important de *loci* informatifs permet d'estimer des paramètres génétiques qui auparavant demandaient un très grand nombre d'individus, tout en permettant une forte résolution phylogénétique. Cette complétion entre les deux analyses apporte donc de plus grandes capacités à résoudre des histoires complexes telles que celle de la sous-section *Erythrodrosom*.

4.1 Différentes structures génétiques

En effet, bien que le traitement des données initiales diffère de la précédente étude (Boucher *et al.*, 2016a), par un fort tri sur les données manquantes, les analyses de structurations de la sous-section *Erythrodrosum* confirment la présence de plusieurs groupes génétiques. Ces derniers sont associés à des espèces aux aires de répartitions bien localisées dans les Alpes. D'un côté un clade 'est-alpin' composé des espèces *P. daonensis* et *P. villosa*, et de l'autre un clade 'Hirsuta' composé de l'espèce homonyme et de *P. pedemontana s.l.*. Cette structure globale de la sous-section était attendue, étant donné qu'il s'agit de plusieurs espèces, mais pour autant il est important de remarquer que la valeur du F_{st} reste faible. Cela peut être dû à plusieurs phénomènes. Dans un premier temps, la radiation récente de la section *Auricula* (Boucher *et al.*, 2016b) a pu limiter la dérive génétique sur les massifs isolés. Une autre hypothèse pourrait être la présence de structures plus fines à celle étudiée comme c'est le cas dans le massif des Apennins (Crema *et al.*, 2009).

Cette dernière hypothèse est cependant contradictoire avec la structure observée ici pour *P. pedemontana s.l.*. Ce dernier taxon ne semble pas présenter une structure très marquée malgré les répartitions allopatriques des lignées la composant. Cette faible structuration pourrait également être causée par l'échelle spatiale et génétique de l'étude. En effet, il ne faut pas oublier que cette structure a été étudiée à l'échelle des Alpes et que par conséquent, la diversité génétique entre les massifs de la Vanoise et des Écrins (inter *P. pedemontana s.l.*) est forcément beaucoup moins importante qu'entre la Vanoise et l'est des Alpes (*P. daonensis*). Néanmoins, ce résultat rejoint le besoin de révision taxonomique de ce complexe d'espèce, avec une étude approfondie des relations génétiques entre lignées.

4.2 Flux de gènes et glaciations

Autre aspect de *P. pedemontana s.l.*, la lignée de *P. pedemontana* des Écrins présente un signal d'admixture avec *P. hirsuta*, ce qui permet d'affirmer l'existence d'un flux de gène (passé ou actuel) entre ces deux espèces. Cette information est par ailleurs cohérente avec les observations de terrains et explique en partie pourquoi les précédentes études ne peuvent l'assigner comme une espèce à part entière ou comme un hybride. Cependant, ce flux de gène reste difficile à étudier avec ce jeu de donnée, par le manque d'individus génotypés et par une méconnaissance des aires de répartitions dans le massif des Écrins.

Pour le manque d'individus, il serait en effet intéressant d'échantillonner davantage de *P.*

hirsuta à proximité des Écrins ainsi que *P. pedemontana* en Vanoise. Ce ré-échantillonnage permettrait de mettre en place les analyses proposés par Eaton *et al.* (2015) pour étudier le sens du flux de gène ainsi que la proportion du génome impacté. Mais ce ré-échantillonnage ne peut se faire sans une meilleure connaissance des populations des Écrins. Ainsi, comme on le remarque sur la figure 1, la limite de l'aire de distribution de *P. hirsuta* chevauche celle du taxon des Écrins (situé dans la vallée du Lauzon). Or cette répartition peut-être faussée par la difficulté de distinguer les deux lignées, à cause de ces morphologies intermédiaires.

Un plus grand échantillon permettrait enfin de proposer une datation pour les différentes séparations entre les lignées réparties sur ces différents massifs. Cette datation pourrait être obtenue avec une approche par approximate Bayesian computation (ABC) sur le logiciel DIYABC (Cornuet *et al.* , 2014). Ces datations permettraient d'évaluer l'impact que les cycles glaciaires du quaternaire ont eu sur ces plantes d'altitude et comment l'évolution de leurs niches écologique a impacté leur génétique. De telles connaissances ainsi qu'une meilleure appréciation des tailles effectives de populations guiderait plus sûrement les politiques de protection du milieu alpin face aux changements climatiques à venir. Ces éléments sont également importants pour essayer de comprendre l'avenir de cette lignée et si de futurs contacts secondaires avec les massifs voisins pourraient redistribuer des gènes entre ces différentes 'espèces'.

4.3 Conclusion

Finalement, la sous-section *Erythrodrosom* présente une structure génétique particulière, façonnée l'histoire des Alpes et son relief favorable aux répartitions allopatriques. Cette histoire aura notamment isolé les lignées sur des massifs éloignés tout en mettant en place d'autres flux de gènes, questionnant ainsi le statut d'espèce pour ces plantes d'altitudes.

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 DivAdapt 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, Théophile Tesseraud et Camille Puverel, 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. 2016a. Sequence capture using RAD probes clarifies phylogenetic relationships and species boundaries in *Primula* sect. *Auricula*. *Molecular Phylogenetics and Evolution*, **104**, 60–72.
- Boucher, Florian C., Zimmermann, Niklaus E., & Conti, Elena. 2016b. Allopatric speciation with little niche divergence is common among alpine Primulaceae. *Journal of Biogeography*, **43**(3), 591–602.
- Cornuet, Jean Marie, Pudlo, Pierre, Veyssier, Julien, Dehne-Garcia, Alexandre, Gautier, Matthieu, Leblois, Raphaël, Marin, Jean Michel, & Estoup, Arnaud. 2014. DIYABC v2.0 : A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**(8), 1187–1189.
- 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.
- Davey, John L., & Blaxter, Mark W. 2010. RADseq : Next-generation population genetics. *Briefings in Functional Genomics*, **9**(5-6), 416–423.
- 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.
- Jombart, Thibaut, & Ahmed, Ismail. 2011. adegenet 1.3-1 : New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**(21), 3070–3071.
- Kadereit, J. W., Goldner, H., Holstein, N., Schorr, G., & Zhang, L. B. 2011. The stability of Quaternary speciation : A case study in *Primula* sect. *Auricula*. *Alpine Botany*, **121**(1), 23–35.
- 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 Bordolona, 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