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 :	**Tuteur de stage :**	**Équipe :**
Éric Coissac	Florian Boucher	DivAdapt

Primula pedemontana Crédit photo : Florian Boucher

Résumé

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer 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.

abstract

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

Todo list

Table des matières

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 orogenèse et les cycles glaciaires qui ont joué sur l'Holarctique au quaternaire.

Si tu veux dire que l'environnement est très dynamique, dis-le.

Ainsi, étudier l'évolution des végétaux présents dans ces vallées et massifs s'avère compliqué, malgré leurs spécialisations dans des niches précises et fragiles.

Est-ce que tu peux expliquer en quoi les niches des espèces rend leur étude compliquée ?

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 la définition de l'espèce au sens biologique et de nombreuses classifications. Cela a aboutit à une plus grande diversité taxonomique, actuellement retravaillée avec l'arrivée des techniques de génomique.

Cette fin de paragraphe est pas super claire. La présence d'espèces cryptiques ne nie pas forcément le concept d'espèce biologique, pense aux polyploïdes de l'arctique par exemple.

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 (Zhang *et al.* , 2004)

Tu peux citer Boucher et al. 2016 J. of Biogeography aussi ici...

, 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 *Erythrodrosum* 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.* , 2016).

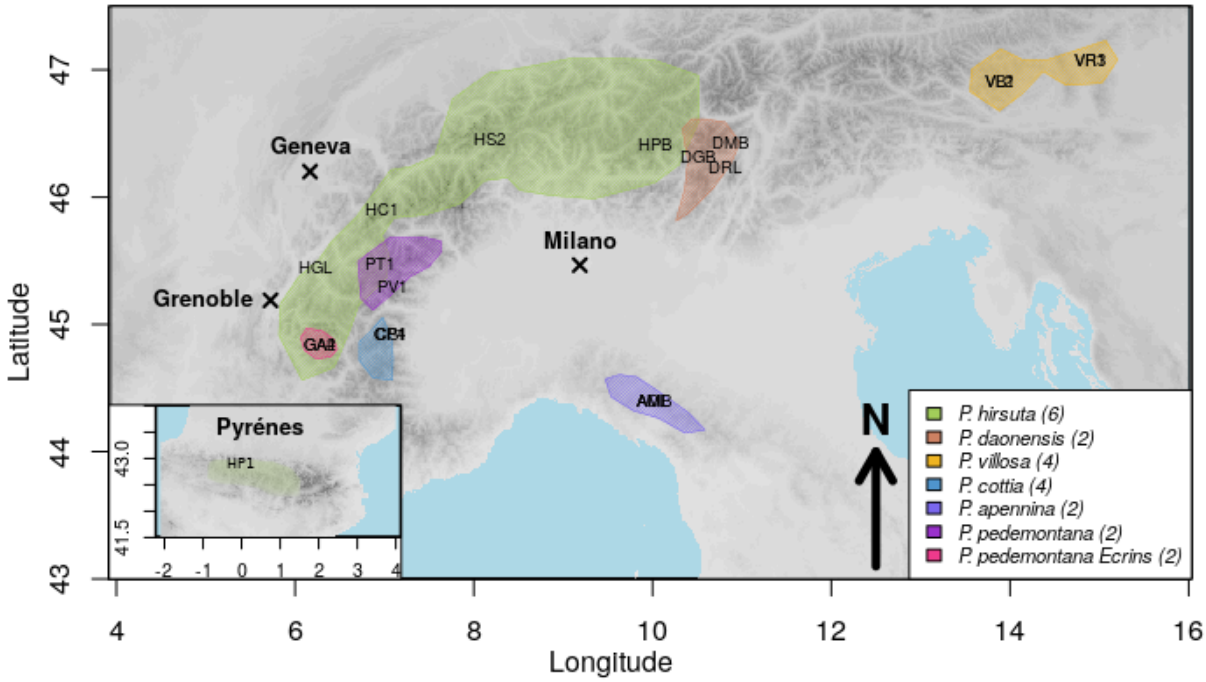


FIGURE 1 – **Carte de répartition de *Primula* sect. *Auricula* Duby subsect. *Erythrodrosum* Pax.** Espèces composant la sous-section, avec entre parenthèse le nombre d'échantillons utilisés dans cette étude. Les aires de répartitions sont extrapolées depuis les observations de l'INPN et du CBNA. Les codes d'individus sur la carte reflètent le lieu d'échantillonnage. Fond de carte : Température moyenne annuelle extraite de WorldClim. Résolution 30s (1km^2).

La faible résolution phylogénétique obtenue jusqu'alors dans ce groupe étant probablement due à ~~des phénomènes trop récents~~son origine récente (CITATION ?), 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.* (2016) 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 gla-

ciaires.

Les objectifs de cette étude sont donc : (i) d'étudier la sous-section *Erythrodrosum* 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 *Erythrodrosum* englobe 7 espèces : *P. villosa ; P. daonensis ; P. hirsuta ; P. apennina ; P. cottia et P. pedemontana*. L'ensemble de ces espèces sont ancestralement hexaploïdes mais du fait de l'ancienneté de cet évènement de polyploïdisation, elles sont considérés 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.* (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 du

être réattribué après analyse génétique au taxon *P. hirsuta*.

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 **(référence nécessaire ici)** . En effet les enzymes de restrictionsutilisées 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 des sites de restriction, même mutés. 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.

ref

C'est beaucoup mieux mais il manque encore deux points importants dans ce paragraphe (j'ai déjà bien ré-écrit) : 1) tu ne dis pas clairement qu'on fait de la capture hybride de loci RAD avec des sondes ARN, on le comprend mais ce n'est pas dit directement. 2) tu dis qu'on a peu de variation, mais ceci-dit que les sites de restriction sont mutés, ça paraît un peu contradictoire comme ça.

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.

Si tu manques de place tu peux réduire toute cette section à quelques lignes seulement : tu est parti d'un jeu de données de SNPs pour 90 individus, obtenus après génotypage de type hyRAD et alignement sur un génome de réf.

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 souhaité dans l'ordre indiqué. Au maximum, les individus conservés à partir du jeu de donnée initial sont les individus de la sous-section *Erythrodrosum* 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. 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 a 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 réalisés sous ~~Rstudio (R.Team, 2017)~~ R avec différents packages.

Rstudio c'est juste une interface et en plus ça t'économise une citation et ça te permet de citer un peu de biologie à la place ;)

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*. Ce clustering fait par adegenet trouve les groupes pour lesquels la variance intergroupes est maximale 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 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".

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'a 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 "P1" verrait donc une plus grande proportion de *loci* à la configuration "BABA" et une valeur négative du "D"

je pense qu'il faut absolument une figure avec un schéma ABBA-BABA ici, on se perd carrement en lisant les explications (dans le papier de Durand aussi, rassure-toi) et avec un figure ça sera hyper clair.

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 bootstraping des SNPs. Ce bootstraping 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.

3 Résultats

3.1 Tri du jeu de données

Le jeu de donnée initial contient 175 799 *loci*. Pour le jeu de donnée utilisé au niveau de la sous-section *Erythrodrosum*, 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 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 du 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 *Erythrodrosum*

Dans un premier temps, le F_{st} général entre les différents taxons de la sous-section *Erythrodrosum* donne une valeur de 0.15, ce qui appuie l'hypothèse la présence d'une structure génétique. Cette structure était attendue étant donné qu'il s'agit de plusieurs espèces, mais pour autant il est important de remarquer que la valeur reste faible.

proposer une explication ? peut etre parce qu'il y a une sous structure ? quid du Fis <0

les filtres un peu drastiques sur la quantité de données manquantes virent le gros de la variabilité ? A toi de voir si tu veux en parler ou si tu gardes ça pour l'oral. De toute façon, si tu donnes une explication c'est en Discussion, pas dans les résultats qui doivent être factuels. D'ailleurs je mettrais en Discussion la portion ci-dessus : 'ce qui appuie l'hypothèse la présence d'une structure génétique. Cette structure était attendue étant donné qu'il s'agit de plusieurs espèces, mais pour autant il est important de remarquer que la valeur reste faible.'

Les F_{st} entre paires d'espèces sont également tous supérieurs à zero (valeurs comprises entre 0.093 et 0.214), et les relations phylogénétiques entre espèces obtenues par neigbhor-joining à partir de ces valeurs de F_{st} rejoignent celles décrites 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

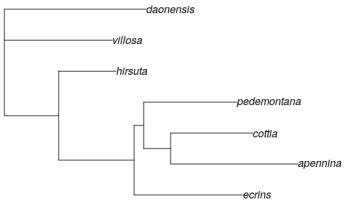


FIGURE 2 – **Topologie d'*Erythrodrosum* par Neighbor-joining**, réalisé à partir de la matrice de distances F_{st} par paires de populations.

clusters avec le critère d'information bayésien le plus faible ~~(BIC = 119.90)~~est atteint pour K=3.

L'info intéressante à donner c'est la différence de BIC, le BIC du meilleur modèle seulement ne donne aucune info intéressante.

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.*.

K	BIC
1	121.8305
2	120.3166
3	**119.9053**
4	120.8310
5	122.0690

Ce clustering est plus facilement visualisable sous la forme d'une analyse en composante principale comme dans la figure **??**. 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*.

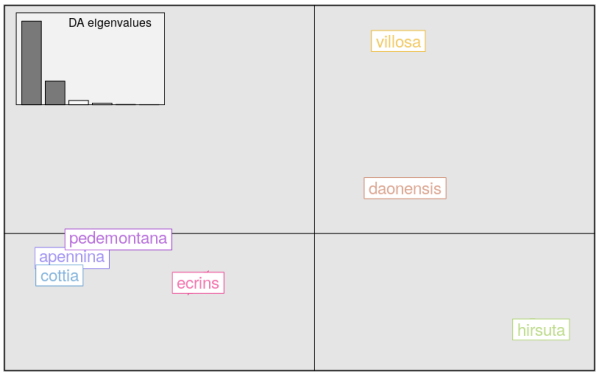


FIGURE 3 – **Analyse en composante principale discriminante de la sous-section *Erythrodrosum* Pax.** Les groupes à priori sont les populations échantillonnées.

3.3 Clade 'Hirsuta'

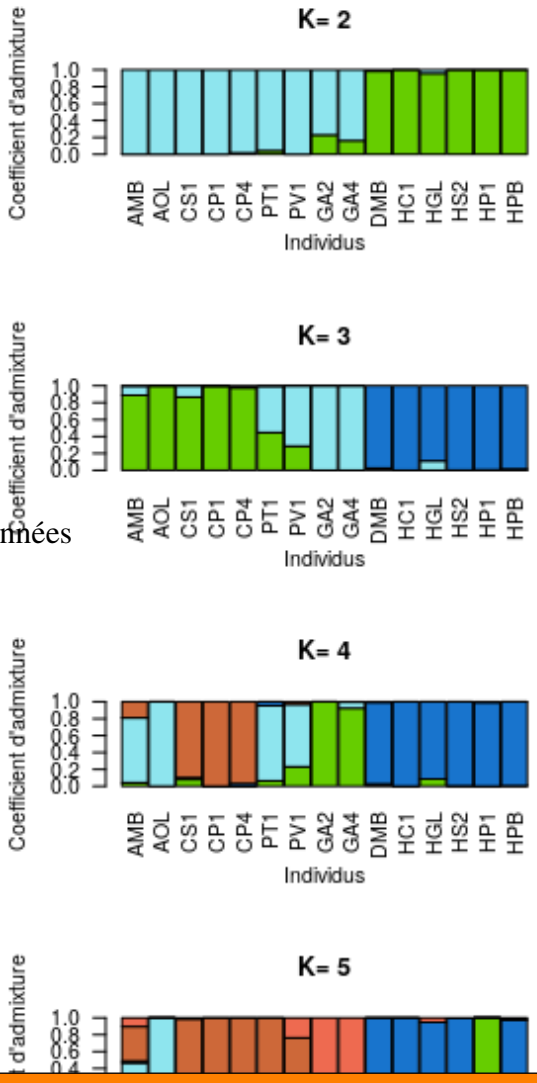
La structuration du clade 'Hirsuta' pour différentes valeurs de K permet de voir différentes informations (figure 4. Pour K=2, une séparation est déjà nette entre deux clades, conformément aux résultats précédents, entre *P. pedemontana s.l.* et *P. hirsuta*. Cette séparation présente néanmoins une légère 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).

Ici il faudrait entourer montrer les différentes espèces sur ces graphes, le relecteur ne connait pas mes codes d'individus...

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 étais 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 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 ambigue, même si les individus des Écrins sont toujours isolés.



Au final c'est quoi la meilleure valeur de K ? Tu perds pas mal de place à décrire ces différents graphes je trouve... Clairement l'intérêt de K=5 me paraît très limité. Je trouve également que tu emploies le terme 'robuste' un peu à la légère, je ne vois pas trop ce que tu veux dire, que le résultat ne te plaît pas ?

FIGURE 4 – **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 Admixture

Le test d'admixture entre les différentes populations de *P. pedemontana s.l.* et *P. hirsuta*, avec *P. daonensis* en outgroup confirment

P1	P2	P3	Outgroup	D	p-valeur
P. pedemontana	*P. apennina*	*P. hirsuta*	*P. daonensis*	-0.027	0.212
P. pedemontana	*P. cottia*	*P. hirsuta*	*P. daonensis*	-0.033	0.0772
P. pedemontana	Taxon des Écrins	*P. hirsuta*	*P. daonensis*	**0.082**	**$5.761.10^{-5}$**
P. apennina	Taxon des Écrins	*P. hirsuta*	*P. daonensis*	**0.109**	**$6.465.10^{-7}$**
P. cottia	Taxon des Écrins	*P. hirsuta*	*P. daonensis*	**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.

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. Ce résultat permet par contre de proposer deux hypothèses : la proximité génétique entre ces trois espèces ou alors une admixture passée avant séparation de ces populations sur des massifs isolés.

totale réécriture de la partie introgress, article a creuser davantage

Pas sur que mettre introgress soit nécessaire. Cette section admixture est très bien. Par contre, as-tu besoin de présenter les résultats 'contrôle', i.e. sans le taxon des Ecrins ? Si tu choisis de le faire alors il faut mieux expliquer pourquoi (en gros tu testes la méthode de Durand pour des faux positifs). Aussi, il faudrait expliquer dans les méthodes pourquoi tu teste les 3 P1 différents. Enfin, il faut vraiment une figure ABBA BABA, avec les taxons que tu as choisis, ça illustrera à la fois la méthode dans le mat et met et les résultats ici (tu pourrais par exemple mettre la flèche entre hirsuta et ecrins en plus gros que celle qui montre la situation BABA).

4 Discussion

discussion complete

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.

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.

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, Ismaïl. 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. irmingardis	Rappolt Kogel, Austria	VR3	Short-styled	F. Boucher	07/06/14	14.88541	47.08313	1871	4722814	4234314	Photo
P. villosa ssp. irmingardis	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