

Master 2e année Biodiversité Ecologie Evolution
Parcours Dynamique et modélisation de la biodiversité
(DynaMo)

Title : An example report in latex for my master
degree in Grenoble.

Maxime Jaunatre ¹

23 novembre 2020 | Soutenance : mardi 2 Juin 2020

Encadrant :

François Munoz

Laboratoire :

LECA
CIRAD, 1A A51/PS2, Boulevard de la Lironde,
34398 Montpellier cedex 5

Équipe :

Bouclés



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.

Check: this!

Abstract

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.

Check: this!

note that this is an example of abstract in 2 language

Table des matières

note that this page style is 'empty' and have no page number for example

Todo list

note that this todo list will disappear if you set notes on false or comment it

This document provide all the command I usually use for my reports. Gray notes are here to explain some stuff, while red ones are here to be a last line limit when I wrote my report too close to the deadline and misread colours. I defined a new todonotes juste for this, to help me manage my work.

Check: imp command

1 Basic commands

Here is some memory about basic latex commands. Like putting some text in *italic* for species names. Or **bold** for important parts. This is completed with a possibility to highlight with the **hl** command. The colour is defined at the very beginning of the document, but another command allow to change with **hlc[red]word**. Every colour is possible and you can define them by the xcolor package.

Below is an exemple of a list

- Item A.
- Item B.
 - 1. sub part 1
 - 2. sub part 2
- Item C

2 To do notes

To help in long work redaction, the todonotes package is loaded by default, allowing to set inline or marge todonotes . Authors can be specified and the list of to do is loaded at the beginning of the document. Everything is explain in more detail inside this document <http://tug.ctan.org/macros/latex/contrib/todonotes/todonotes.pdf>

inline

Gowachin: specifying the author of the todo comment

3 Maths

Some packages are also loaded to allow maths equations to be displayed.

$$Y \in B; B = a, c, g, t \quad (1)$$

$$\begin{cases} prey_{t+1} = prey_t + prey_t \cdot r \cdot (1 - prey_t/K) + int_{pred/prey} \cdot pred_t \\ pred_{t+1} = pred_t + pred_t \cdot r \cdot (1 - pred_t/K) + int_{prey/pred} \cdot prey_t \end{cases} \quad (2)$$

with : r the growth rate and K the capacity.

4 Captions

Multiple captions are possible, like tables, figures and boxes, as display below. To show all the possibilities, there is **Lorem ipsum** text around.

 Lorem ipsum dolor sit amet, consectetuer
 adipiscing elit. Ut purus elit, vestibulum
 ut, placerat ac, adipiscing vitae, felis. Cu-
 rabitur dictum gravida mauris. Nam arcu
 libero, nonummy eget, consectetuer id,
 vulputate a, magna. Donec vehicula augue
 eu neque. Pellentesque habitant morbi tris-
 tique senectus et netus et malesuada fames
 ac turpis egestas. Mauris ut leo. Cras vi-
 verra metus rhoncus sem. Nulla et lectus
 vestibulum urna fringilla ultrices. Phasel-
 lus eu tellus sit amet tortor gravida place-
 rat. 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

Glossary

Phylogeny: Object resuming evolutionary relation between living beings by illustrating distance between them. Mostly represented as a phylogenetic tree.

Community: An assemblage of populations from different species, defined by a time and geographic scale restriction. This assemblage can also be represented by interactions between these populations.

varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

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

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

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.

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

Boxe 1: More complex null model, Pigot et al. (2015)

Context : To study the assembly processes of a community, ecologist compare metrics (in this example the Mean Nearest Taxon Distance) with random draw species in the community. This null model is easily applicable but build assumption issues. For example, these null models don't take into account the historical aspect of community assembly.

Results : Authors found that their null models produce overdispersed community, by just taking into account historical aspect of colonisation and extinction in contrast with just random draw of species. Empirical data shown significantly overdispersed species in the community but fail to reject this null model.

Therefore, overdispersion could just arise from historical processes, independent from traits or biotic interactions. This new null model invite us to rethink more precedent result where overdispersion had been found.

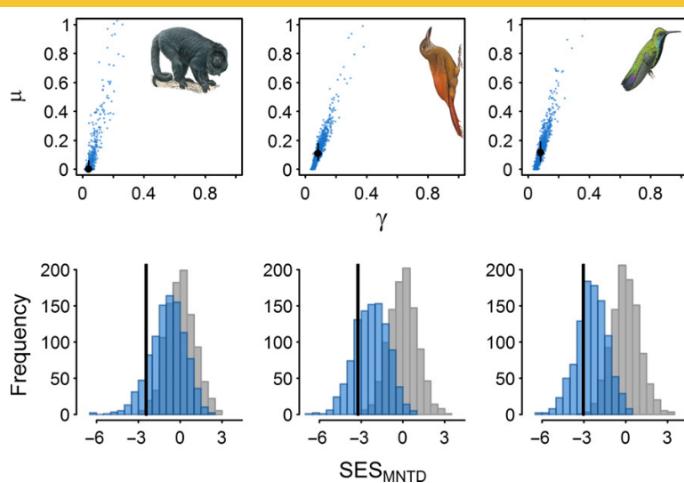


Figure 5 The dynamics and phylogenetic structure of community assembly for New World primates (left), ovenbirds (middle) and hummingbirds (right) in Manu. Maximum likelihood rate estimates (black points) of colonization (γ) and local extinction (μ) with uncertainty in estimates (black lines 75% CI) quantified from 1000 bootstrap replicates (blue points). Community phylogenetic structure according to the Mean Nearest Taxon Distance (SES_{MNTD}) showing the observed patterns (black lines), the expectation under the random-draw null model (grey) and the expectation under DAMOCLES given estimated rates (from 1000 replicate simulations).

urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

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

4.1 multicols

Captions are harder to set in twocolumn environnement, but here is an example. I don't really controle figure in two-columns docs to put them in a reproducible document, but I think time will help with this.

 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.

 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 li-

 gula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

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

5 Import other documents

This is a subfile, that act exactly like the main file in term of compilation. It means it can be compiled on its own

6 Adding R code

As R (?) is commonly used in this master, there is the possibility to paste directly R code and input it inside the document. It is important to note that long lines are broken and it is signaled with line numbers on the left part of the script.

```
1 # this is a random function taken out from a master project. code won't run
2 hudson_lotka <- beepodyna(label = "hudson_lotka", community = hudson,
3   interactions = hudson_int,
4   functions = c(lotka_prey, lotka_pred), verbose = TRUE)
5
5 jpeg("figure/lotka.jpg", width = 802, height = 445)
6 plot(simulate_n_pop_dynamic(hudson_lotka, 10)$community, col = c("tan3",
7   slateblue3),
7   pch = c(1, 2), xlim = c(1910, 1930), ylim = c(0, 100))
8 legend("bottomright", c("hare", "lynx"), col = c("tan3", "slateblue3"),
8   pch = c(1, 2))
9 graphics.off()
```

Next is my first master year report, to show examples with figures and all other stuff.

7 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 (?). 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. 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 distributions allopatriques dues à l’extension glaciaire ou par une spécialisation sur différentes niches écologiques apparaissant au gré des cycles géoclimatiques, des contacts secondaires avec le retrait glaciaire sont propices aux flux de gènes et hybridations.

Ces différentes actions de la dynamique du paysage sur la structure génétique des espèces alpines ont 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*. Regroupant pas moins de 26 espèces (?), c’est le groupe de plantes le plus riche du système Alpin européen (i.e., l’ensemble des montagnes d’Europe (?)). Cette section s’étant fortement diversifiée durant les 5 derniers millions d’années (??), 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 (?). C’est par exemple le cas dans la sous-section *Erythrodosum*, 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 ??). 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.* (?).

La faible résolution phylogénétique obtenue jusqu’alors dans ce groupe étant probablement due à son origine récente (?) , la génétique des populations pourrait permettre d’éclairer le statut

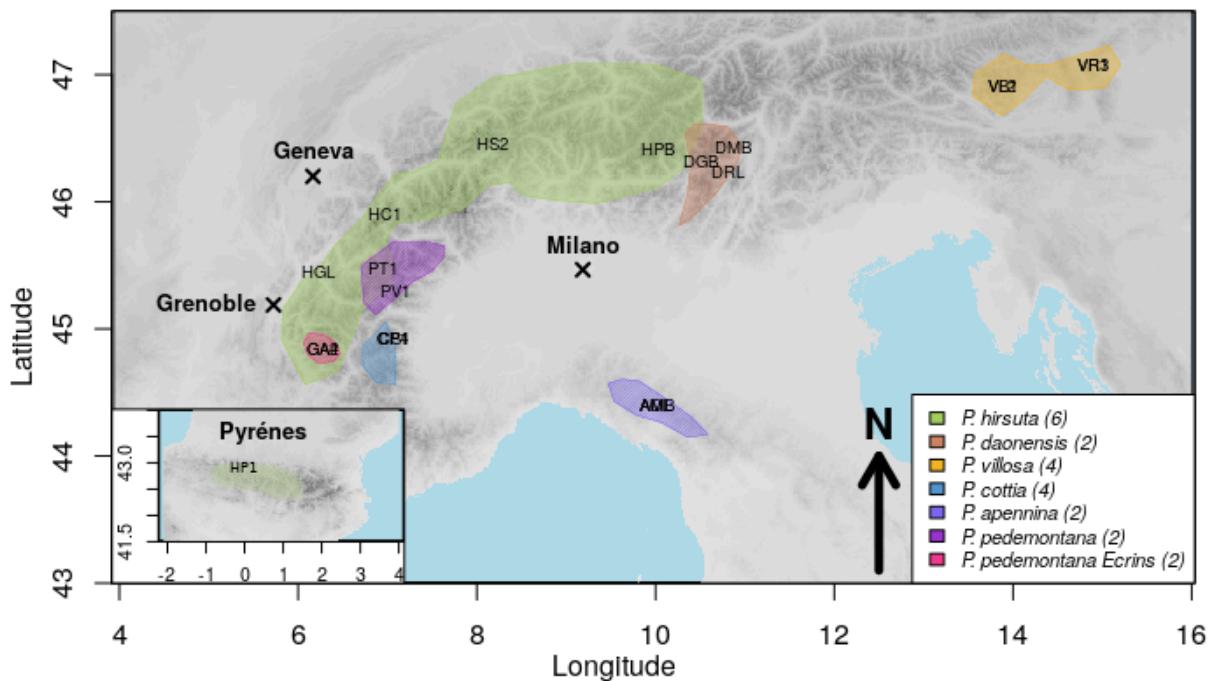


FIGURE 1 – Carte de répartition de *Primula* sect. *Auricula* Duby subsect. *Erythrodrosum* Pax. Espèces composant la sous-section, avec entre parenthèse le nombre d'échantillons utilisés dans cette étude. Les aires de répartitions sont extrapolées depuis les observations de l'INPN et du CBNA. Les codes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

taxonomique de ces primevères. En effet, même si les données génétiques utilisées par ? 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 populations des 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 *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.

8 Matériels et méthodes

8.1 Cas d'étude

La section *Auricula* Duby du genre *Primula* comprend 26 espèces et est endémique du système Alpin européen (?). Au sein de cette section, la sous-section *Erythrodrosum* Pax englobe 7 espèces : *P. villosa* Wulfen ; *P. daonensis* Leyb. ; *P. hirsuta* All. ; *P. apennina* Widmer ; *P. cottia* Widmer et *P. pedemontana* Thom. ex. Gaudin. L'ensemble de ces espèces est ancestralement hexaploïde mais du fait de l'ancienneté de cet évènement de polyploidisation, 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 (fissures de rochers, éboulis et prairies alpines) (?). 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 ??). 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 plus ou moins denté selon les individus (lien web ??).

8.2 Échantillonnage

Les données génétiques utilisées lors de cette étude ont été produites dans le cadre d'une étude précédente (?). 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 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. À ces fins, l'ADN avait été génotypé par la méthode hyRAD (?) 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 (?). En effet les enzymes de restrictions utilisées en RAD-seq sont sensibles à la mutation d'un nucléotide, tandis que les sondes ARN utilisées 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* L. 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.

8.3 Bioinformatique

A partir du jeu de données initial, un pipeline est établi dans le langage R (?) pour générer divers ensembles de données, spécifiques à chaque analyse. Les fonctions sont rassemblées en un package R hébergé sur Github (lien web ??). 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 donnée initial sont les individus de la sous-section *Erythrodosum* 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 ?, car présentant trop de données manquantes. Toutes les informations concernant les individus de l'analyse sont présentées en annexe ?? et leur répartition est indiquée en Figure ??.

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.

Si
be-
soin
de
place
du
peux
enle-
ver
'par
le
logi-
ciel
Bow-
tie2'

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 formats, la sous-sélection d'individus et de *loci* est enregistrée sous quatre formats (.vcf, .str, .geno, .snp) grâce à la fonction `files`. La transformation d'un format en un autre se fait respectivement au moyen d'un script bash, par l'utilisation du software PGDSpider (?), grâce au package LEA (?) et avec un script R.

Toutes ces fonctions sont indépendantes mais peuvent être appelées dans l'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 (?).

8.4 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` (?). 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 intergroupe est maximale quand la variance intragroupe est minimale.

Si
be-
soin
de
place
du
peux
enle-
ver
'La
trans-
for-
ma-
tion
d'un
for-
mat
en
un
autre
se
fait
res-

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 (?). 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.

8.5 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 ? propose une statistique (D) basée sur quatre lignées partageant un ancêtre commun, selon la fréquence de SNPs observés avec un motif particulier sur une topologie de référence. Dans cette topologie, nous avons d'abord une divergence entre l'outgroup "O" et les autres taxons, suivie de la divergence de "P3" et finalement une divergence entre "P1" et "P2" (Figure ??).

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" (Figure ??).

Il a été décidé de calculer D à partir des fréquences alléliques, comme décrit dans l'équation ci-dessous où P_{ij} correspond à la fréquence de l'allèle alternatif (B).

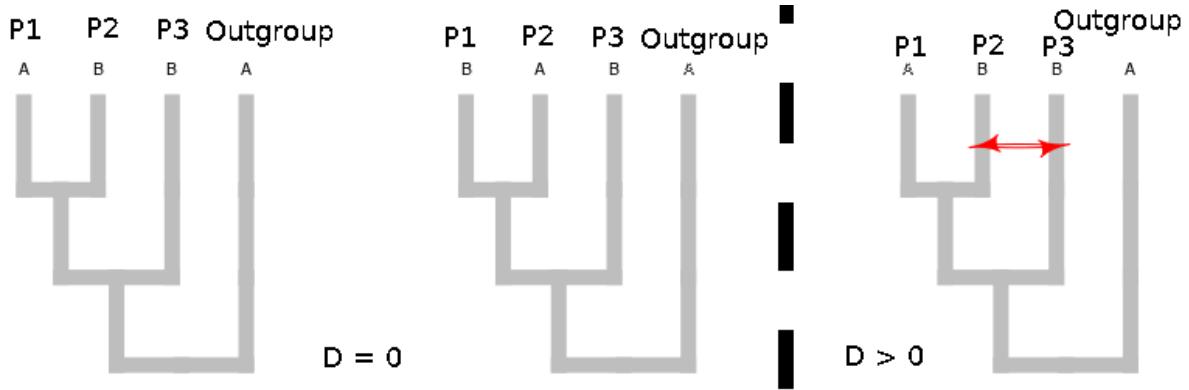


FIGURE 2 – Configurations d’allèles utilisées dans le test "ABBA-BABA". Les deux schémas de gauche représentent les deux configurations d’allèles (respectivement "ABBA" et "BABA") en équilibre sous un modèle neutre ($D=0$). Le schéma de droite montre un cas d’admixture entre "P1" et "P3", ce qui induit une plus grande proportion de la configuration schématisée (ABBA).

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

Un intervalle de confiance sur cette statistique a été obtenu par bootstrapping des SNPs, 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. Afin de vérifier l’absence de faux positifs, ce test a été 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 issu d’une lignée de *P. pedemontana*, l’admixture a été testé entre tout 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’introgression, il faut considérer toutes les lignées dérivées de l’ancêtre commun (?). Pour toutes les topologies différentes (voir ?? dans la partie résultat), *P. daonensis* a été considéré comme l’outgroup, et *P. hirsuta* comme P3. Ces choix découlent des résultats précédemment établis concernant la phylogénie de la sous-section *Erythrodrosum* (?).

9 Résultats

9.1 Tri du jeu de données

Le jeu de données initial contient 175 799 SNPs. Pour le jeu de données 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’, 1 851 *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.

9.2 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 zero (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 ??). 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.*.

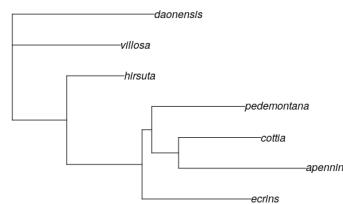


FIGURE 3 – Topologie de la sous-section *Erythrodrosum* obtenue par Neighbor-joining, réalisé à partir de la matrice de distances F_{st} par paires de populations.

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*.

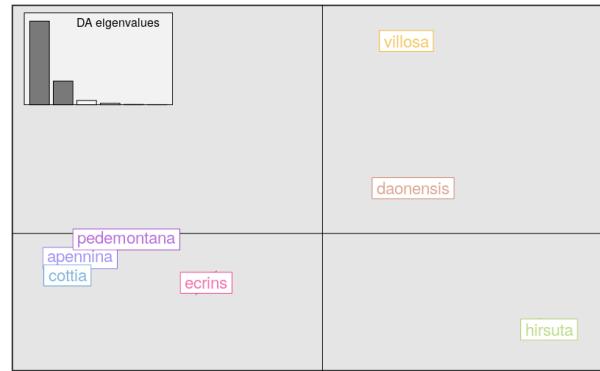


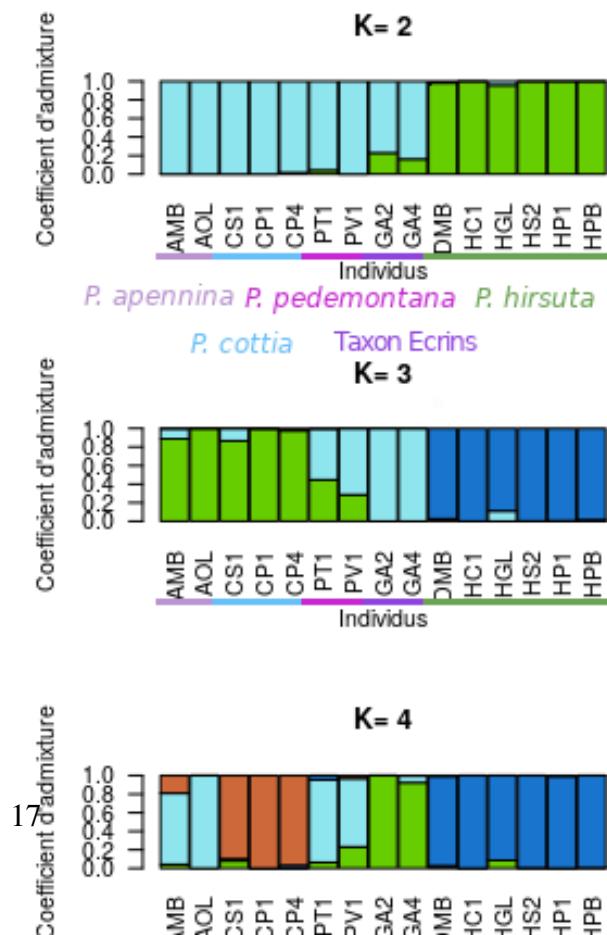
FIGURE 4 – Analyse en composante principale discriminante de la sous-section *Erythrodossum* Pax. Les groupes à priori sont les populations échantillonnées.

9.3 Clade 'Hirsuta'

La structuration du clade 'Hirsuta' pour différentes valeurs de K permet de voir différentes informations (figure ??). 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 échantillé en Belledonne, un massif proche des Écrins (figure ??).

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. apen-*



nina 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.

9.4 Hypothèse d'admixture

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*. Un D supérieur à 0 indique une admixture entre P2 et P3.

10 Discussion

En réanalysant le jeu de données d'une précédente étude phylogénétique, il a été ici possible de conforter l'histoire évolutive de la sous-section *Erythrodrosum* (*Primula* section *Auricula*) et sa structure génétique au niveau des Alpes. Cette étude a également validé l'hypothèse d'admixture pour un taxon présent dans le massif des Écrins, confirmant de fait le besoin de révision taxonomique pour l'espèce *P. pedemontana*. 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. À 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 complémentation entre les deux analyses apporte donc de plus grandes capacités à résoudre des histoires complexes telles que celle de la sous-section *Erythrodrosum*.

10.1 Structure génétique alpine

En effet, bien que le traitement des données initiales diffère de la précédente étude (?), 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 et que les résultats de phylogénie proposaient une structure à cette échelle. Pour autant il est important de remarquer que la valeur du F_{st} reste faible. Cela peut être du à la radiation récente de la sous-section *Erythrodrosum* (?), qui a pu limiter l'effet de la dérive génétique sur ces massifs isolés. Cette radiation récente permet également d'expliquer le peu de variations morphologiques que présentent les espèces de cette sous-section. C'est d'ailleurs ces fortes ressemblances qui ont longtemps modifiés la taxonomie de ce clade. Une autre hypothèse pour cette faible valeur de F_{st} serait la présence de flux de gènes, homogénéisant les pools d'allèles et luttant contre une dérive génétique.

Autre résultat de cette étude, l’absence de structure génétique très marquée pour *P. pedemontana s.l.* ajoute un argument en faveur de la révision taxonomique de ce groupe d’espèces. 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 (intra-*P. pedemontana s.l.*) est forcément beaucoup moins importante qu’entre la Vanoise et l’est des Alpes (*P. daonensis*). Néanmoins, ce résultat est identique à celui de la délimitation d’espèces (?) et soulève la question de la délimitation d’espèces au sein de cette sous-section. Cette question est d’autant plus délicate que ces espèces présentent naturellement des hybrides viables (?).

10.2 Flux de gènes et glaciations

Les résultats de cette étude confirment que la lignée de *P. pedemontana* du massif 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 données, par le manque d’individus génotypés et par une zone de contact secondaire (passée ou actuelle). Un ré-échantillonnage plus précis des massifs de Vanoise et des Écrins permettrait d’ailleurs de clarifier le statut de cette zone de contact. En effet, l’observation de divers degrés d’admixture affirmerait que le flux de gène est actuel, avec la présence de différentes générations d’hybrides.

Une plus fine compréhension de ce cas d’admixture aiderait à la compréhension de l’histoire évolutive de la flore alpine européenne. En effet, cet événement d’admixture ne représente pas un cas isolé et un certain nombre de cas d’hybrides sont observables entre les différentes espèces du genre *Primula* (?????). Cette redondance de cas pourrait trouver son origine dans l’histoire géoclimatique du système alpin, où les cycles glaciaires ont joué un rôle important dans la délimitation géographique des niches écologiques. Ainsi, avec le développement des glaciers alpins, de nombreuses populations se sont vues restreintes dans des refuges (nunataks) (?), et le retrait glaciaire a remis en contact ces populations. Ces contacts secondaires intra-spécifiques ont entraînés un brassage génétique cyclique, aboutissant à une perte d’allèle rares et donc de diversité (?). Ce processus limitant la structurations des différentes espèces a pu jouer un rôle

important dans la perméabilité des barrières reproductives inter-spécifiques.

Enfin, bien que la spéciation allopatrique soit la plus commune pour la flore des Alpes (?), la présence de flux de gènes et d'hybrides suggère que ce mode de spéciation requiert un isolement géographique bien plus long (plus de 3 Millions d'années) pour que la dérive génétique puisse effectivement aboutir à deux espèces sœurs. En l'absence de long isolement, ce mode de spéciation serait remplacé par la spéciation hybride, qui demande cependant de reconsidérer la notion d'espèce biologique. La prise en compte de ce mode de spéciation en biologie évolutive questionne finalement notre capacité à comprendre l'histoire de l'évolution, puisque des hybrides provenant de lignées aujourd'hui éteintes ont pu fausser la description d'événements biologiques passés.

10.3 Conclusion

Finalement, la sous-section *Erythrodrosum* présente une structure génétique particulière, façonnée par l'histoire du système alpin européen 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.

Check: change bibliography model

Web 1 – <https://github.com>

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	Ritorio, 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. D’après les informations de ?.