

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

Maxime Jaunatre, Master 1 BEE Grenoble

1 Avril - 31 Mai 2019 — Soutenance : juin 2019

Enseignant référent :

Éric Coissac

Tuteur de stage :

Florian Boucher

Équipe :

DivAdapt



Résumé

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

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

Table des matières

| | | |
|----------|-------------------------------------|-----------|
| 1 | Introduction | 2 |
| 2 | Matériels et méthodes | 3 |
| 2.1 | Échantillonnage | 3 |
| 2.2 | Bioinformatique | 3 |
| 2.3 | Génétique des populations | 5 |
| 2.4 | Inférences bayésiennes | 5 |
| 2.5 | Admixture -ABBA-BABA- | 6 |
| 3 | Résultats | 7 |
| 3.1 | Tri du jeu de données | 7 |
| 3.2 | Clade Hirsuta | 7 |
| 3.3 | Complexe ouest alpin | 7 |
| 4 | Discussion | 8 |
| 5 | Bibliographie | 9 |
| 6 | Ressources | 10 |

1 Introduction

Par la diversité des milieux qu'elle abrite, la chaîne des Alpes est une zone géographique propice à une grande biodiversité. Cet ensemble de massifs et de vallées constitue donc un laboratoire parfait pour étudier l'évolution des organismes dont les aires de répartition semblent aisément observables. Cependant, l'histoire de cette orogénèse et les différents cycles glaciaires qui ont joué sur l'Holarctique complexifient énormément ces histoires biologiques. Ainsi, l'isolement géographique propice aux spéciations, ou encore l'adaptation à de nouvelles niches écologiques se sont trouvés modifiés de nombreuses fois par ces facteurs abiotiques. Ces nombreuses modifications ont abouti à un état actuel de biodiversité où il est difficile de séparer ces ensembles de mécanismes, et conclure à une histoire biogéographique simple. En conséquence, on observe nombre d'espèces cryptiques, ou d'entités taxonomiques définies sur des critères évoluant rapidement dans la littérature.

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

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

To
check

2 Matériels et méthodes

2.1 Échantillonnage

Les données génétiques utilisés lors de cette étude ont été produite dans le cadre de l'étude précédente de Boucher *et al.* (2016). Il s'agit d'un jeu de donnée composé de 90 individus des espèces composant la section *Auricula*, au sein du genre *Primula*. Ces individus ont été prélevés à travers les Alpes entre avril et septembre 2014. L'identification taxonomique des individus a été réalisé sur le terrain, mais un individu a du être réattribué après analyse génétique au taxon *P. hirsuta*.

Les SNPs sont issus de séquençage haut débit via hyRAD (Suchan *et al.* , 2016). Le génome de référence proviens de *Primula veris*. Cette technique permet de génotyper le long du génome malgré des mutations sur les sites de restrictions. En effet les enzymes de restrictions sont trop sensibles à la mutation d'un nucléotide, tandis que les sondes ARN peuvent s'hybrider sur des sites plus nombreux sur le génome. La nécessité de capturer des sites malgré une faible variation provient du niveau interspécifique de l'étude, qui pose l'hypothèse que les mutations peuvent se placer sur les sites de restrictions et ainsi limiter leur capture par simple séquençage ddRAD.

2.2 Bioinformatique

Pour chacun des individus, l'information consiste en une séquence de SNPs appellés par Freebayes, exporté sous format VCF. Les analyses ont été porté sur deux jeux de données car filtrés sous des seuils différents. Le premier jeu de donnée ('m30_-q20_mincov20') est issus de filtres très strict, avec un score de qualité (Phred) requis de 30 et une couverture minimale de 20 lectures par site. Afin de ne pas biaiser l'analyse par des seuils favorisant les régions conservées, le second jeu de données ('m13_-q20_mincov10') est quant à lui produit avec un Phred minimal de 13 et une couverture de 10 lectures. A partir de ces séquences, les SNPs ont été isolés par Freebayes, avec un score de Phred minimal de 20 et un support de lecture de 30% minimum par allèle.

A partir des deux jeux de donnée initiaux, un pipeline est établis pour générer divers ensembles de données. Cette automatisation a permis de générer les jeux 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 sous Rstudio (R.Team, 2017), avec la fonction `subset_reorder`, qui permet de reconstruire le fichier en ne gardant que les indi-

vidus souhaité dans l'ordre indiqué. Au maximum, les individus conservés à partir des jeux de données initiaux sont les individus des espèces suivantes : 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.

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 donnée et le loci pour l'individu présentant cet allèle rare est considéré comme une donnée manquante. Cette étape permet également de supprimer les lectures avec de multiples allèles variants qui sont reconnus comme des artefacts des algorithmes utilisés pour appeler les SNP. Suite aux deux tris précédents, il y a donc des loci pour lesquels tout les individus portent la même information. Afin de ne garder que les loci polymorphiques, une fonction `clean` est donc appelée à la fin de la fonction `rare`, pour supprimer les locis monomorphiques. Les loci sont aussi filtré sur le score de qualité (QUAL) indiqué dans le `vcf`, qui correspond à la confiance dans l'assignation de l'allèle variant. Une fonction de `tri` est appelé pour supprimer les loci puis individus présentant trop de données manquantes selon les seuils posés. Enfin, afin de limiter le déséquilibre de liaisons, les sites sont filtrés selon leurs positions sur les contigs, où une distance minimale `n` doit être prise en compte entre deux sites d'un même contig pour que le second site soit 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), le package LEA (Frichot & François, 2015) et un script R. La production de ces fichiers est inscrite dans le pipeline par la fonction `files`.

L'ensemble de ces fonctions sont indépendantes mais peuvent être appelées dans le bon ordre au moyen de la fonction `dataset`, qui prend en entrée un vecteur de noms d'individus, un `csv` contenant les assignations aux populations, le fichier `vcf` original, le nom des fichiers de sortie et les seuils précisés au-dessus.

Chaque seuil est choisis après vérification que la décision n'impacte que peu les résultats de diversité génétique. Le nombre d'individus étant faible, il faut optimiser le nombre de SNPs par individus car cela permet de limiter la perte d'information et d'atteindre les mêmes résultats

qu'attendus avec un échantillon plus vaste. (Nazareno *et al.* , 2017)

2.3 Génétique des populations

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

Dans un premier temps, des statistiques F sont calculées avec le package adegenet (Jombart & Ahmed, 2011). Ces statistiques permettent de se faire une première idée du jeu de donnée. Ce package permet également de réaliser un analyse en composante principale discriminante. Cependant un clustering de nos individus a été réalisé avant, afin d'essayer d'attribuer des groupes sans a-priori de populations. Ce clustering fait par adegenet trouve les groupes pour lesquels la variance intergroupes est maximale quand la variance intragroupes est minimale. La DAPC a donc été réalisée avec ces groupes et sans afin d'observer les différences.

La structure des populations de nos taxons a ensuite été étudié par le package LEA (Frichot & François, 2015). Ce package permet une implémentation dans R des analyses auparavant menées par le logiciel STRUCTURE. Pour cette étude, la structuration a été modélisée pour des structure de K populations allant de 1 à 15, avec 20 simulations par K, en ne retenant que la meilleurs simulation basée sur le critère de "cross-entropy". Une dernière mesure de génétique des populations a été réalisé pour essayer d'estimer l'expansion ou la contraction récente des populations. Ce D de Tajima est mesuré à l'aide du package pegas (Paradis, 2010).

Concernant l'hypothèse d'hybridation pour le taxon des Écrins, le package introgress (Gompert & Alex Buerkle, 2010) a été utilisé pour mesurer l'index h, index d'hybridation, entre deux taxons. Ce package requiert l'import des deux taxons parents présumés et des hybrides afin de mesurer un coefficient d'admixture (ou h-index). Ce coefficient basé sur une fonction de vraisemblance (Buerkle, 2005), où la proportion de chaque génome parental est mesurée dans l'individu hybride. Les marqueurs codominants utilisés ici n'étant pas fixés, le calcul du h-index se basera ici sur les fréquences alléliques.

To
do

2.4 Inférences bayésiennes

Afin de caractériser l'admixture probable entre le taxon des Écrins et *P. hirsuta*, une approche par approximate Bayesian computation (ABC) a été réalisée sur le logiciel DIYABC (Cornuet *et al.* , 2014). Ce logiciel permet de simuler des jeux de données selon divers scénarios, en échantillonnant des paramètres entre des priors définis. Les scénarios sont ensuite classés

selon les probabilités a posteriori d'observer notre jeu de donnée initial selon les scénarios proposés. Seuls quelques scénarios ont été étudiés ici, en prenant en compte le fait qu'ils sont à chaque fois supportés par peu d'individus. Les priors sont également proposés dans un grand intervalle et selon une distribution uniforme, les temps de divergence entre populations et taille de population n'ayant pas été étudiés sur le terrain.

Pour les topologies proposés en scénarios, toutes les topologies "classiques" avec aucun événement d'hybridation sont essayées (5 cas possibles pour les 4 taxons *hirsuta*, *pedemontana*, *Écrins*, *cottia-apennina*). Des topologies avec des polytomies regroupant plusieurs populations ont également été proposé. Enfin des hypothèses d'hybridation de *pedemontana* et *hirsuta* sont étudiées pour l'origine du taxon des Écrins.

To
do

2.5 Admixture -ABBA-BABA-

Une autre approche de l'admixture entre plusieurs taxons du complexe de populations étudié est réalisé par le test "ABBA-BABA". 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. Cette méthode étant initialement pensée pour des séquences haploïdes. De fait la plupart des algorithmes présentés écartent les sites avec une ambiguïté (code IUPAC) ou alors résolvent l'ambiguïté par un tirage aléatoire entre deux bases. Il existe une méthode pour prendre en compte les sites hétérozygotes à partir des fréquences alléliques, comme présenté dans Durand *et al.* (2011), mais la faible taille de populations biaiserait les résultats. Il est donc plus judicieux d'écarter les sites hétérozygotes, rendant le test plus conservateur. Considérant que les locis sont sous une évolution neutre et sans déséquilibre de liaison, on attend deux configurations différentes pour une même topologie. Cette topologie [(((P1,P2),P3),O)] propose que "P1" et "P2" coalescent avant un autre événement de coalescence avec "P3", puis avec "O" l'outgroup. Sur cette topologie on attend deux allèles présents en fin de branche avec "A" l'allèle ancestral et "B" l'allèle alternatif porté par l'outgroup. Ce test ne s'intéresse qu'à deux cas : "ABBA" et "BABA". Sous un modèle neutre, on attend des proportions équilibrées de sites portant ces deux configurations. L'hypothèse alternative est qu'un déséquilibre de ces proportions peut être induit par deux cas : une topologie autre ou alors une introgression de "P3" avec "P1" ou "P2". Une introgression entre "P3" et "P1" verrait donc une plus grande proportion de locis à la configuration "BABA". Cela aboutit à une valeur négative du "D", comme explicité par l'équation suivante :

$$D(P1, P2, P3, 0) = \frac{\sum_{i=1}^n C_{ABBA}(i) - C_{BABA}(i)}{\sum_{i=1}^n C_{ABBA}(i) + C_{BABA}(i)}$$

Il est important de souligner que ce test ne permet en aucun cas de proposer un sens d'introgession, ni son intensité.

Devant le nombre réduit de sites informatifs, il a été décidé de calculer D à partir des fréquences alléliques, comme décrit dessous. P_{ij} correspond à l'allèle alternatif.

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

To
check

3 Résultats

3.1 Tri du jeu de données

To
do

3.2 Clade Hirsuta

To
do

3.3 Complexe ouest alpin

Pour les analyses sur DIYABC, aucun scénario proposé ne permet de simuler un jeu de donnée proche des données réelles. Le scénario observé le plus proche des données réelles était celui d'un polytomie où coalescent tout les taxons assigné à l'espèce *P. pedemontana*), puis un événement plus ancien de coalescence avec *P. hirsuta*. Ne pouvant conclure si cette polytomie est réelle ou alors un artefact issu d'un manque de résolution (soft poolytomie), les analyses n'ont pas été plus loin avec cette outil et aucun résultat ne peut être présenté.

To
do

4 Discussion

Une grande limitation de cet étude provient de l'échantillonnage des populations. Ainsi, si l'échantillonnage actuel est suffisant pour les précédentes analyses phylogénétiques, la génétique des populations requiert davantage d'informations par populations. La plupart des analyses de génétiques des populations sont basées sur les variations de fréquences alléliques au sein des populations. La faible taille d'échantillon vient donc biaiser fortement ce prérequis, car plus la taille de l'échantillon est faible, moins l'estimation de la fréquence allélique représente la fréquence réelle au sein de la population. De fait, les potentielles structures génétiques des espèces étudiées ne peuvent être étudiées avec précision. Or, dans le cadre de cette étude, différentes structures génétiques sont superposées pour les populations étudiées. Par exemple la population de *P. apennina* échantillonnée ici présente une structure génétique à l'échelle du massif des Apennins (Crema *et al.*, 2009), mais représente aussi une partie de la structure des populations d'Alpes de l'ouest. Enfin, cette population doit représenter une entité structurelle pour la sous-section *Erythrodrosum* à l'échelle de l'arc Alpin. Si l'on ne considère que deux individus séquencés pour cette population, il est donc impossible d'avancer des analyses robustes pour les échelles les plus fines de structure. L'étude est donc de fait limitée à la plus grande échelle possible, sans possibilité de résolution à l'échelle des Alpes de l'ouest. Il est d'ailleurs fort probable que ce soit ce manque de résolution qui a pénalisé les analyses par inférence bayésienne.

To
do

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

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

5 Bibliographie

- Boucher, F. C., Casazza, G., Szövényi, P., & Conti, E. 2016. Sequence capture using RAD probes clarifies phylogenetic relationships and species boundaries in *Primula* sect. *Auricula*. *Molecular Phylogenetics and Evolution*, **104**, 60–72.
- Buerkle, C. Alex. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, **5**(3), 684–687.
- Cornuet, Jean Marie, Pudlo, Pierre, Veyssier, Julien, Dehne-Garcia, Alexandre, Gautier, Mathieu, 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.
- Durand, Eric Y., Patterson, Nick, Reich, David, & Slatkin, Montgomery. 2011. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**(8), 2239–2252.
- Eaton, Deren A.R., Hipp, Andrew L., González-Rodríguez, Antonio, & Cavender-Bares, Jeanine. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, **69**(10), 2587–2601.
- Frichot, Eric, & François, Olivier. 2015. LEA : An Rpackage for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**(8), 925–929.
- Gompert, Zachariah, & Alex Buerkle, C. 2010. Introgress : A software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**(2), 378–384.
- Jombart, Thibaut, & Ahmed, Ismaïl. 2011. adegenet 1.3-1 : New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**(21), 3070–3071.
- Lischer, H. E L, & Excoffier, L. 2012. PGDSpider : An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**(2), 298–299.

- Nazareno, Alison G., Bemmels, Jordan B., Dick, Christopher W., & Lohmann, Lúcia G. 2017. Minimum sample sizes for population genomics : an empirical study from an Amazonian plant species. *Molecular Ecology Resources*, **17**(6), 1136–1147.
- Paradis, Emmanuel. 2010. Pegas : An R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**(3), 419–420.
- 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