



"Analyse de la connectivité des populations du Cuivré de la bistorte
dans un réseau de zones humides pyrénéen"
**Rapport scientifique concernant l'étude génétique de populations du Cuivré de
la bistorte *Lycaena helle***

Etude réalisée pour l'Association des Naturalistes de l'Ariège et le Parc Naturel des
Pyrénées catalanes par
Lise-Marie Pigneur et Johan Michaux
GeCoLAB
Laboratoire de Génétique de la Conservation
Université de Liège

Contexte

Une analyse de la connectivité des populations du cuivré de la bistorte (*Lycaena helle*) a été entreprise dans un réseau de zones humides pyrénéen. A cette fin, l'Association des Naturalistes de l'Ariège et le Parc Naturel des Pyrénées catalanes ont collecté des échantillons dans 4 sites de la région pyrénéenne. Ces échantillons ont été confiés au Laboratoire de Génétique de la Conservation (Université de Liège) afin de les analyses génétiquement.

Echantillonnage

En vue d'une analyse de structure et de diversité génétique, l'Association des Naturalistes de l'Ariège et le Parc Naturel des Pyrénées catalanes ont prélevé 1 patte sur 80 individus mâles de Cuivré de la bistorte dans 4 sites; les Mouillères de Noubals (N), de Balbonne (B), de Puyvalador (Puy) et de Porté-Puymorens (Po) (Tableau 1 ; Figure 1).

Cette opération a été autorisée sous couvert d'une autorisation préfectorale, s'agissant d'une espèce protégée - Arrêté préfectoral n°2018-s-17 du 28 mai 2018 portant sur l'autorisation de capture temporaire d'un papillon protégé (Direction Régionale de l'Environnement, de l'Aménagement et du Logement - Direction Ecologie - Division Biodiversité).

Les pattes ont été conservées dans l'éthanol absolu jusqu'à analyse.

Tableau 1. Sites de collecte et nombres d'échantillons collectés de Cuivré de la bistorte en France.

Site	Coordonnées géographiques (Lat./Long.)	Nombre d'individus
Mouillères de Noubals (N)	42.718110 ° / 2.053086 °	20
Mouillères de Balbonne (B)	42.685773 ° / 1.990902 °	17
Mouillères de Puyvalador (Puy)	42.646863 ° / 2.128322 °	23
Mouillères de Porté-Puymorens (Po)	42.546394 ° / 1.842973 °	20

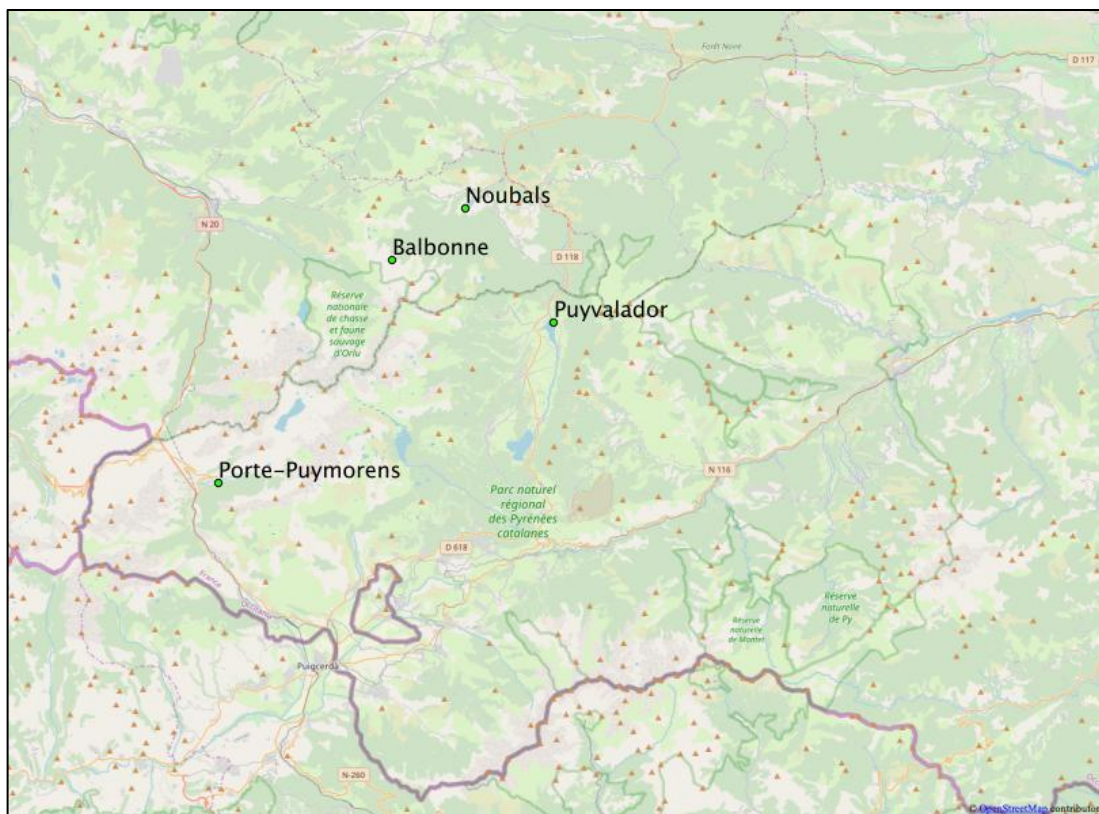


Figure 1. Localisation des 4 sites de collecte de Cuivré de la bistorte.

Matériel et Méthodes

Extraction de l'ADN et amplification

L'ADN de chaque patte collectée a été extrait à l'aide du Qiamp DNA microkit (Qiagen) selon les recommandations du fabricant.

Suivant les travaux de Habel et al. (2008) et Finger et al. (2009) nous avons développé nos analyses sur base des 5 marqueurs microsatellites polymorphes disponibles pour l'espèce : LheE12, LheB06, Lhe03, Lhe14, LheF12.

Les analyses ont été réalisées en utilisant un seul mix multiplex, permettant d'amplifier les 5 marqueurs microsatellites en un seul passage en séquenceur automatique. Pour cela, l'amorce *forward* de chaque microsatellite a été marquée par un fluorochrome

particulier. Le type de fluorochrome choisi pour chaque microsatellite a été basé sur la taille de fragments de chacun de ceux-ci (voir Habel et al. 2008 et Finger et al. 2009).

L'amplification du mix multiplex a été réalisée grâce au Qiagen Multiplex PCR kit (Qiagen) selon le protocole du fabricant, dans un volume final de 10 µl.

Génotypage et analyse des données

Les microsatellites amplifiés et marqués par fluorochromes ont été génotypés au moyen d'un séquenceur automatique multicapillaire *Applied Biosystems 3730xl Genetic Analyser*. Les résultats du génotypage sont informatisés sous forme d'électrophérogrammes qui permettent de mesurer la taille en paires de bases des différents allèles des microsatellites. La taille des allèles dépend du nombre de répétitions qu'ils contiennent. La taille de ces allèles pour chaque microsatellite et par individus a été identifiée à l'aide du programme GENEMAPPER 4.0 (Applied Biosystems).

Nous avons systématiquement répété les génotypages 2 fois pour chaque échantillon afin de garantir la qualité des données obtenues.

Analyses statistiques

Une fois les typages microsatellites effectués, la matrice de données a été analysée selon les approches statistiques suivantes :

- pour identifier la structure populationnelle globale sur l'ensemble des échantillons analysés, nous avons utilisé le logiciel STRUCTURE 2.3.4 (Pritchard *et al.*, 2000). Ce logiciel permet de déterminer s'il existe une structure populationnelle quelconque et d'identifier le nombre de populations (K) en assignant chaque individu à un des clusters sur base de son génotype (méthode du *clustering* bayésien). Aucune affiliation géographique *a priori* n'est fournie au logiciel. Il est également supposé que les fréquences alléliques sont corrélées entre les populations sur base du modèle d'admixture. Une série de « runs » indépendants a permis de tester des nombres de clusters K allant de 1 à 10 populations et en utilisant un « burn-in » de 100 000 itérations suivi d'une seconde analyse de 2 000 000 d'itérations (MCMC). L'analyse pour chaque K a été répétée 10 fois afin de vérifier la robustesse du résultat. Après correction selon Evanno, la meilleure valeur de K a été sélectionnée et les probabilités d'appartenance des individus aux différents clusters (pour chaque valeur de K) ont été synthétisées et représentées graphiquement via la plateforme CLUMPAK (Kopelman *et al.* 2015).
- la richesse allélique (A_r), les hétérozygoties observées (H_o) et attendues (H_e) et les indices de fixation (F_{st}) et de consanguinité (F_{is}) ont été calculés via le package *diveRsity* (Keenan et al. 2013) dans R (R Core Team, 2013). Les analyses ont été effectuées avec 1000 répétitions de bootstrap.
- une analyse discriminante des composantes principales (DAPC) a été réalisée afin d'établir le nombre de groupes génétiques via un autre algorithme (k-means) et de représenter ces groupes graphiquement. L'analyse a été réalisée dans R (R Core Team, 2013) via le package *adegenet* (Jombart 2008).

Résultats

Structure génétique

Concernant les résultats de l'analyse réalisée avec le logiciel STRUCTURE, la plus haute valeur de ΔK a été obtenue pour $K=2$, suggérant la présence de 2 clusters génétiques (Fig 2). Ces clusters semblent bien différenciés ($D_{Jost} = 0,31$).

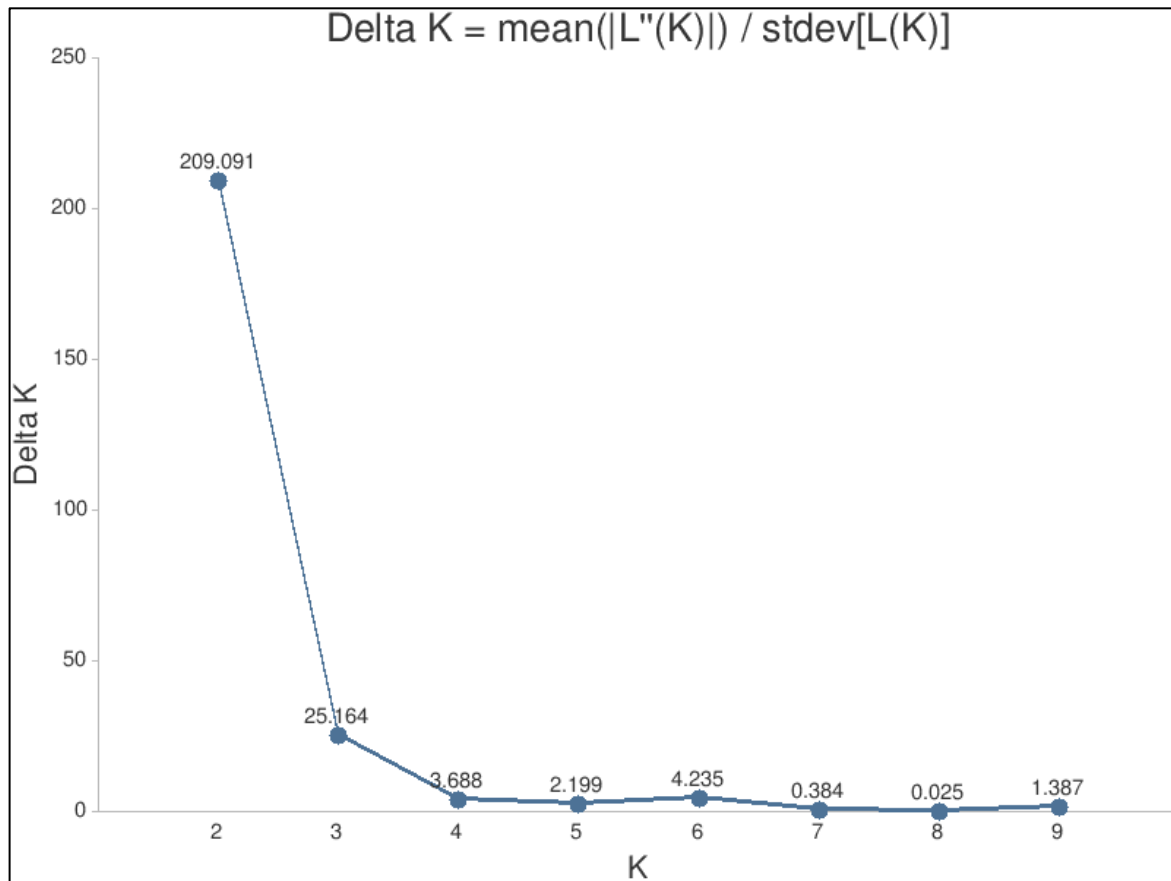


Figure 2. Valeurs de ΔK selon les nombres de clusters analysés.

Au total, 23 échantillons sont bien assignés au cluster 1 (coefficient d'appartenance $q > 0,90$) et 30 au cluster 2. Le reste des échantillons présente un patron génétique intermédiaire entre ceux-ci.

On note la présence des 2 clusters génétiques dans 3 des 4 sites étudiés, tandis que le site de Porté-Puymorens présente une grande homogénéité avec uniquement des individus attribués au cluster 2 (Figure 3). La même tendance est observée avec les résultats de la DAPC.

A l'exception de ce site, on trouve également des individus avec un patron génétique mixte (non attribués fortement à un des 2 clusters).

Il convient cependant de rester prudent quant à l'interprétation de ces résultats car certains individus présentent des données manquantes pour certains marqueurs microsatellites. Or, étant donné que nous ne disposons que de 5 de ces marqueurs, l'absence de données pour certains d'entre eux peut directement impacter la qualité des résultats.

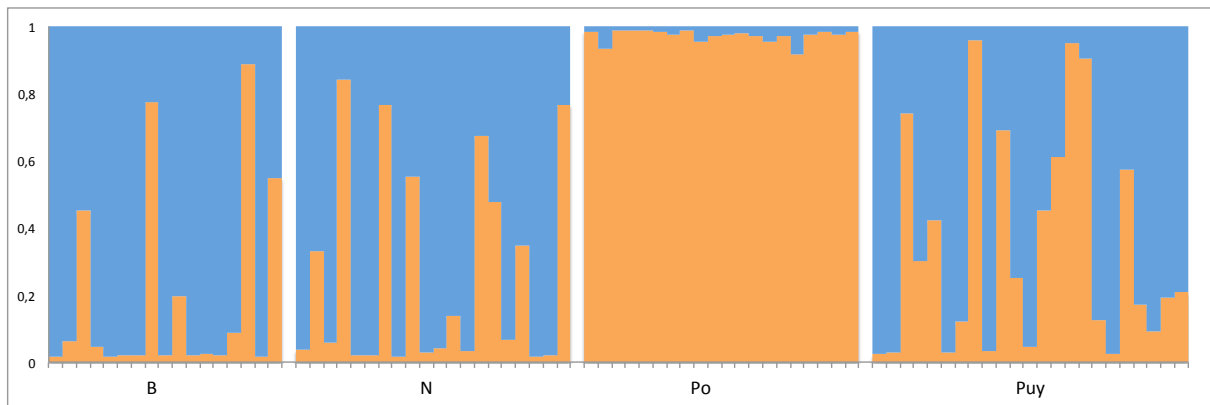


Figure 3. Résultats de l'analyse STRUCTURE pour $K=2$. Chaque bâtonnet représente la probabilité d'un individu d'appartenir à chacun des clusters génétiques. En bleu : cluster 1; en orange : cluster 2.

L'algorithme de k-means utilisé via la fonction *find.clusters* indique une valeur optimale du nombre de clusters de 2 (fig. 4) même si les valeurs de 3 à 7 restent intéressantes à explorer également. On voit à la figure 5, la répartition des individus selon une distribution bimodale. La répartition des individus des 4 sites étudiés au sein des 2 clusters définis est présentée au tableau 2.

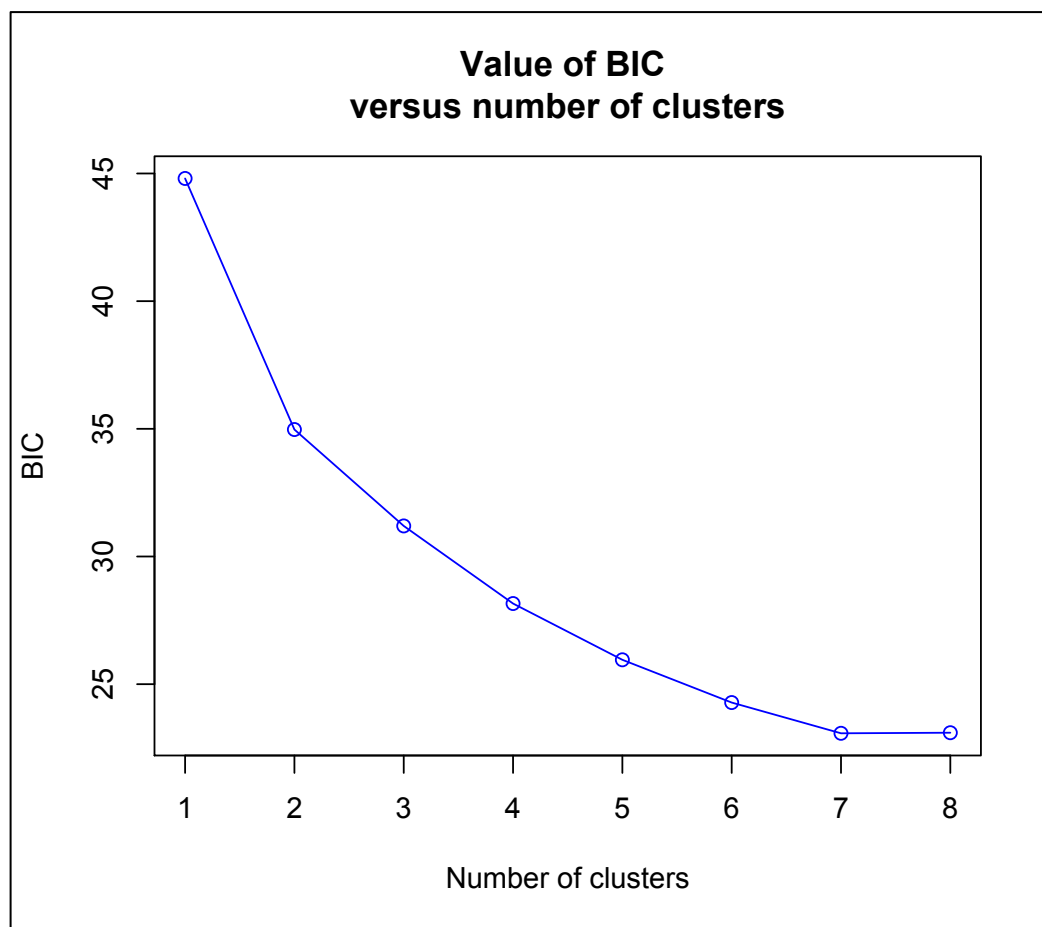


Figure 4. Critère d'information bayésien servant à établir le nombre de clusters au sein du jeu de données d'après un algorithme de k-means.

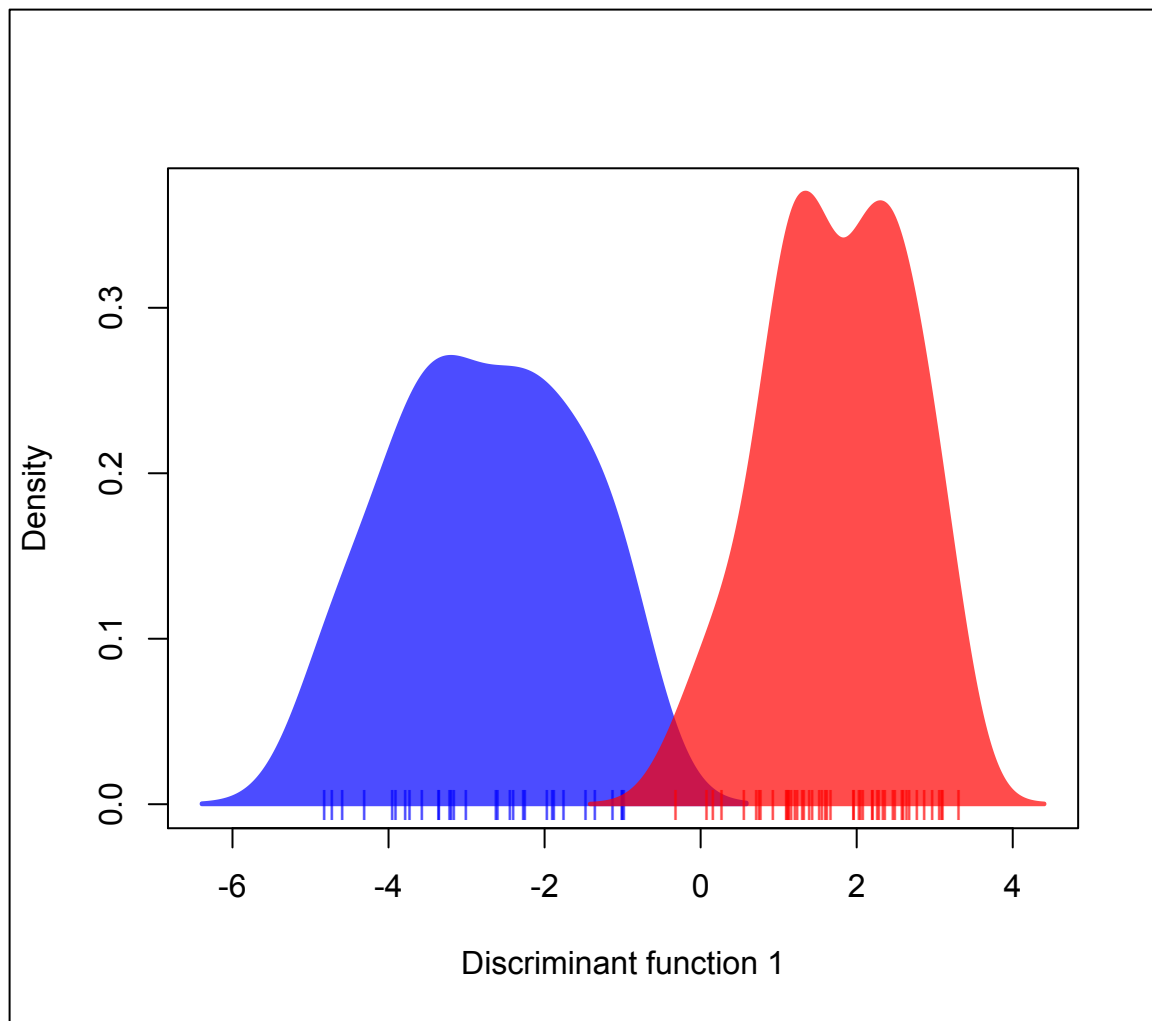


Figure 5. Résultat de la DAPC pour un nombre de clusters égal à 2.

Tableau 2. Répartition des individus des 4 sites dans les 2 clusters définis par la DAPC.

Site	Cluster 1	Cluster 2
B	12	5
N	11	9
Po	0	20
Puy	8	15

Si on répète l'analyse en choisissant un nombre de clusters égal à trois, on obtient la répartition présentée à la figure 6 (voir aussi tableau 3) et on note à nouveau que tous les individus du site Po se regroupent dans un seul et même cluster.

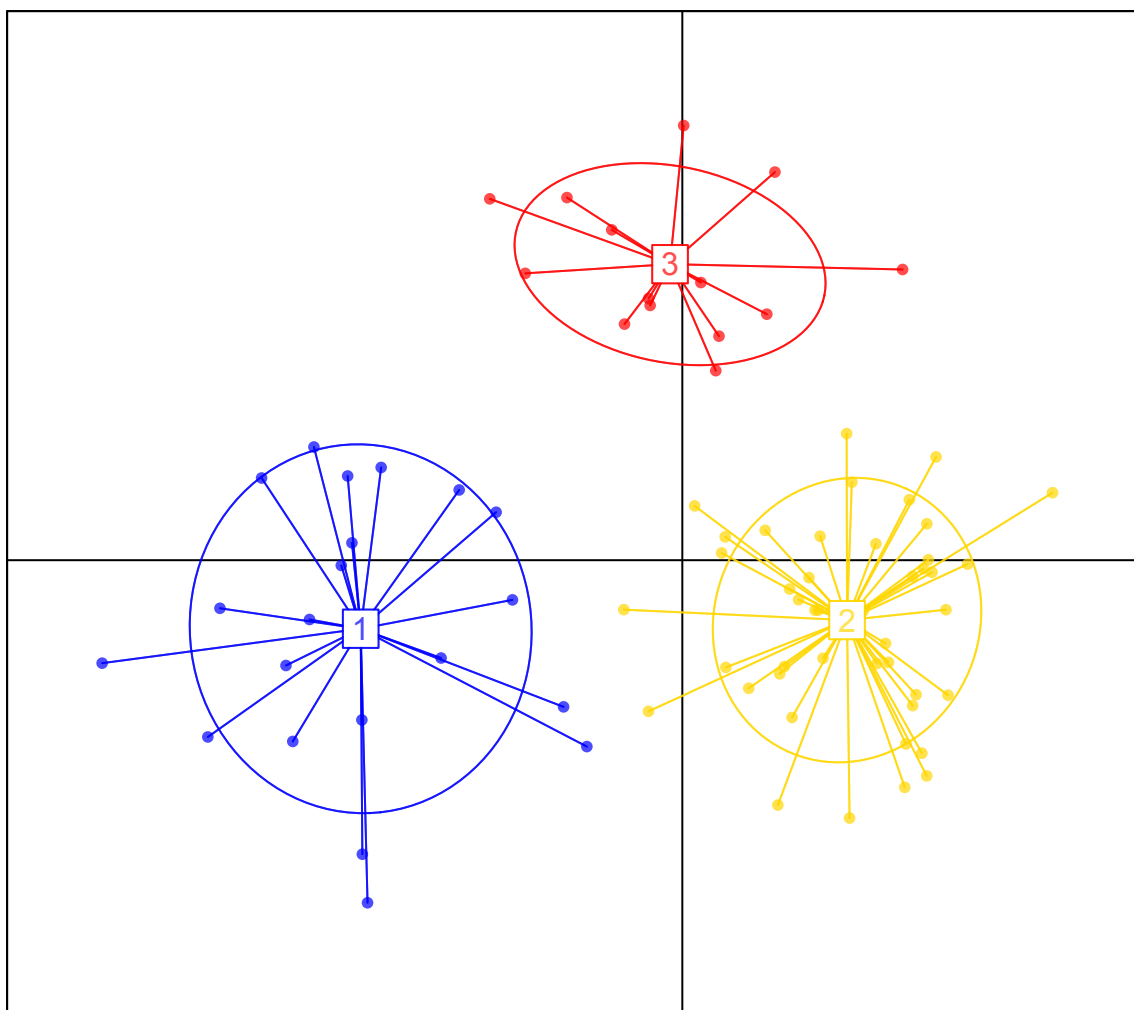


Figure 6. Résultat de la DAPC basée sur 3 clusters définis par l'algorithme de k-means.

Tableau 3. Répartition des individus des 4 sites dans les 3 clusters définis pas la DAPC.

Site	Cluster 1	Cluster 2	Cluster 3
B	14	1	2
N	4	9	7
Po	0	20	0
Puy	4	14	5

Enfin, si on réalise la DAPC sur base des sites, et non pas sur base de clusters obtenus par l'algorithme de k-means, on observe la répartition présentée à la figure 7. On voit à nouveau que les individus du site Po semblent former un groupe à part en comparaison des 3 autres sites.

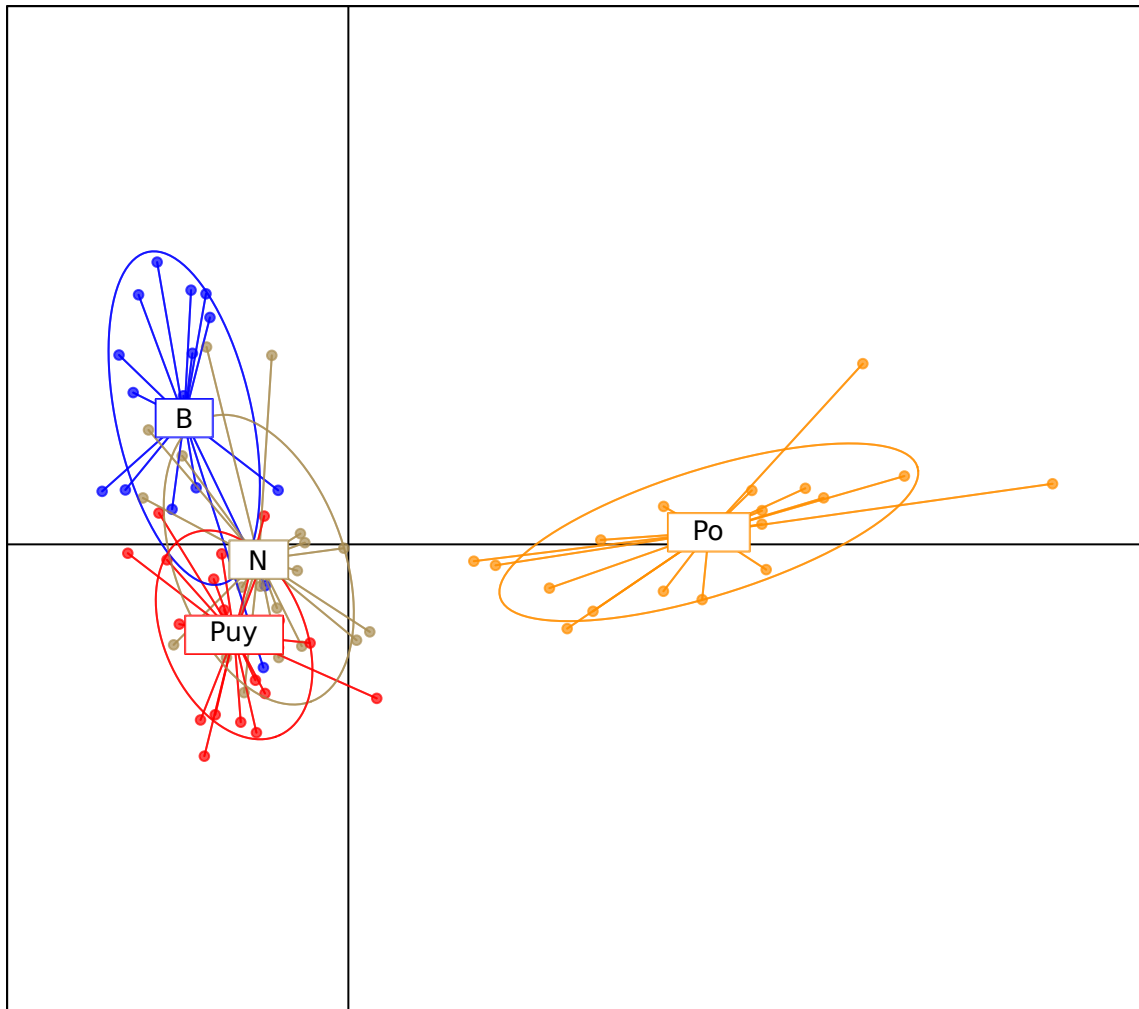


Figure 7. Résultat de la DAPC basée sur les 4 groupes définis par les sites de collecte.

Diversité génétique et flux géniques

- Au sein des sites

Les paramètres de diversité génétique (Tableau 4) montrent une richesse allélique relativement faible au sein des sites (moyenne = 3.4) en comparaison des valeurs mentionnées par Finger et al. (2009) pour la Grande Région (Allemagne, Luxembourg, Belgique). La valeur la plus élevée est trouvée au sein du site Po, qui semble pourtant le plus isolé génétiquement. Cette richesse allélique élevée pourrait s'expliquer par le fait que les animaux du site Po constitueraient une populations relique survivante d'une population ancienne beaucoup plus vaste et caractérisée par une grande diversité génétique. Ce signal de plus grande diversité génétique subsisterait ainsi dans cette population, mais avec néanmoins un risque important de la voir s'éroder peu à peu avec le temps.

Le coefficient de consanguinité (F_{is}) le plus élevé est observé pour la population de Po, ce qui suggère un risque de dépression de consanguinité pour cette population, suite à son isolement. Néanmoins les différences observées entre sites ne sont pas significatives eu égard aux intervalles de confiance (Tableau 4).

Tableau 4 : Paramètres de diversité génétique des populations analysées : coefficient de consanguinité (Fis), richesse allélique moyenne (Ar), hétérozygoties attendue (He) et observée (Ho), test d'équilibre d'Hardy-Weinberg (HWE). N=nombre d'individus analysés.

	N	Ar	Ho	He	HWE	Fis
B	17	3.01	0.43	0.53	0	0.18 [-0.06-0.40]
N	20	2.98	0.54	0.56	0.6382	0.02 [-0.14-0.17]
Po	20	4.53	0.5	0.63	0.0044	0.21 [0.06-0.35]
Puy	23	3.18	0.55	0.54	0.0481	-0.01 [-0.15-0.13]

Les valeurs des indices de fixation (Fst et D) montrent que les populations des sites B, N et Puy sont faiblement isolées génétiquement les unes des autres, tandis que Po apparaît le plus isolé, en particulier par rapport au site B et au site Puy selon la valeur de l'indice D (Tableau 5). Notons que Po est le site le plus éloigné géographiquement des 3 autres.

Tableau 5. Indices de fixation Fst (au-dessus de la diagonale) et D_Jost (sous la diagonale)

	B	N	Po	Puy
B	x	0.08	0.23	0.10
N	0.02	x	0.15	0.04
Po	0.37	0.21	x	0.13
Puy	0.45	0.03	0.16	x

- Au sein des 2 clusters

Le même type d'analyses réalisé en classant les individus dans les 2 clusters génétiques (donc en ne prenant que les individus bien assignés à ceux-ci; $q > 0.9$) montrent ces clusters sont bien différenciés génétiquement ($D = 0.31$). La richesse allélique moyenne du cluster 2 est plus élevée mais le coefficient de consanguinité l'est aussi (Tableaux 6 et 7).

Tableau 6. Indices de fixation Fst (au-dessus de la diagonale) et D_Jost (sous la diagonale)

	Cluster 1	Cluster 2
Cluster 1	x	0.27 [0.24-0.31]
Cluster 2	0.31 [0.22-0.42]	x

Tableau 7. Paramètres de diversité génétique des 2 clusters génétiques: coefficient de consanguinité (Fis), richesse allélique moyenne (Ar), hétérozygoties attendue (He) et observée (Ho), test d'équilibre d'Hardy-Weinberg (HWE). N=nombre d'individus analysés.

	N	Ar	Ho	He	HWE	Fis
Cluster 1	30	2.9	0.47	0.48	0.4326	0.0019 [-0.118 - 0.1246]
Cluster 2	23	4.99	0.52	0.63	3e-04	0.1676 [0.0237 - 0.2921]

Conclusions

Une analyse de la connectivité des populations du cuivré de la bistorte (*Lycaena helle*) a été entreprise dans un réseau de zones humides pyrénéen. Elle a permis d'obtenir une série d'informations inédites sur la structure et la diversité génétiques des populations de ces sites.

En conclusion, il existe une certaine fragmentation génétique entre les sites situés plus au nord et celui le plus au sud (Porté-Puymorens). La différenciation de la population de Porté-Puymorens par rapport aux autres est plus grande qu'entre les sites de Balbonne, Puyvalador et Noubals.

Les trois sites du nord renferment deux clusters génétiques tandis qu'on ne retrouve que le cluster 2 sein de la population de Porté-Puymorens. La différenciation entre les deux clusters est importante.

Assurer ou restaurer la connectivité entre site apparaît donc importante pour le Cuivré de la Bistorte, en particulier entre les sites du nord et celui de Porté-Puymorens. De plus, malgré un certainement isolement génétique de la population de Porté-Puymorens, sa richesse allélique plus élevée renforce l'intérêt d'augmenter la connectivité entre sites afin de permettre la transmission de diversité génétique entre ces différentes populations. L'intérêt d'une telle connectivité semble d'autant plus important pour le site de Porté-Puymorens que les résultats obtenus tendent à montrer un risque accru de dépression de consanguinité et une augmentation de marqueurs génétiques homozygotes pour ce site.

Enfin, la présente étude a été uniquement développée sur base des 5 marqueurs microsatellites disponibles pour l'espèce et utilisés avec succès dans de précédentes études (Finger et al. 2009 ; Habel et al. 2008). Une étude sur un plus grand nombre d'échantillons et basée sur un plus grand nombre de microsatellites (minimum 10) ou d'autres marqueurs (SNPs, etc) permettrait une connaissance encore plus fine de ces populations.

Bibliographie

- Finger A., Schmitt T., Zachos F.E., Meyer M., Assmann T., Habel J.C. 2009. The genetic status of the Violet Copper *Lycaena helle* – a relict of the cold past in times of global warming. *Ecography* 32: 382–390.
- Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 24:1403-1405.
- Habel J.C., Finger A., Meyer M., Schmitt T., Assmann T. 2008. Polymorphic microsatellite loci in the endangered butterfly *Lycaena helle* (Lepidoptera: Lycaenidae). *European Journal of Entomology* 105: 361–362.
- Keenan K., McGinnity P., Cross T.F., Crozier W.W., Prodöhl P.A. 2013. *diveRsity*: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4, 782–788.

- Kopelman N.M., Mayzel J., Jakobsson M., Rosenberg N.A., Mayrose I. 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Res*, 15(5): 1179-1191.
- Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype data. *Genetics*, **155**, 945-959.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.