

EXPERTISE SUR DES ANALYSES GÉNÉTIQUES DE GRANDS CANIDÉS DU PARC NATIONAL DU MONT-TREMBLANT ET SA PÉRIPHÉRIE

Rapport final

Glenn YANNIC

LABORATOIRE D'ÉCOLOGIE ALPINE - UMR CNRS 5553 UNIVERSITÉ SAVOIE
MONT BLANC, Le Bourget-du-Lac, France

Mars 2019



Résumé

La Société des Établissements de Plein Air du Québec (SEPAQ) et la Direction de l'Expertise sur la Faune terrestre, l'Herpétofaune et l'Avifaune du Ministère des Forêts, de la Faune et des Parcs (MFFP) ont mandaté deux laboratoires indépendants pour identifier génétiquement 21 Grands Canidés. Les résultats des deux laboratoires divergent sur l'identification génétique de plusieurs spécimens. Les raisons pouvant expliquer ces différences sont : i) la nature des espèces et populations de référence utilisées (3 espèces pour le **Premier Laboratoire** et 5 pour le **Second Laboratoire**); ii) l'identification et le degré de « pureté » des individus de référence; iii) l'origine géographique des individus de référence (essentiellement d'origine Québécoise pour le Premier Laboratoire et a priori très majoritairement Ontarienne pour le Second Laboratoire); et iv) des paramètres utilisées lors de l'identification des individus avec le logiciel STRUCTURE. Ce rapport détaille les données utilisées et les résultats obtenus par les deux Laboratoires, en soulignant les différences d'approche. Des pistes d'améliorations sont finalement avancées. Ce rapport ne permet cependant pas de conclure sur l'appartenance génétique des individus du Mont-Tremblant et de ses environs soumis pour identification aux deux Laboratoires.

1. Introduction

La Société des Établissements de Plein Air du Québec (SEPAQ) et la Direction de l'Expertise sur la Faune terrestre, l'Herpétofaune et l'Avifaune du Ministère des Forêts, de la Faune et des Parcs (MFFP) ont mandaté deux laboratoires indépendants pour identifier génétiquement 21 Grands Canidés du Mont-Tremblant et de sa région, identifiés visuellement comme étant 15 loups et 6 coyotes. Le laboratoire de la Direction de la Gestion de la Faune de l'Estrie, de Montréal, de la Montérégie et de Laval, secteur des opérations régionales du MFFP (dénommé ci-après : le **Premier Laboratoire**) a sous la supervision de Nathalie Tessier réalisé la première analyse. Le laboratoire du Natural Resources DNA Profiling and Forensic Centre (dénommé ci-après : la **Second Laboratoire**) a réalisé, sous la supervision de Brad Seyler, la seconde analyse.

Les deux laboratoires ont rendu chacun un rapport qui produisent des résultats contradictoires. Sur mandat de la SEPAQ, l'objet du présent rapport est de :

- i) réaliser une analyse critique des deux méthodologies employées (incluant le paramétrage du logiciel d'assignement de l'identité spécifique des échantillons) ainsi qu'évaluer la fiabilité des résultats reçus et la puissance des analyses réalisées.
- ii) tenter d'identifier les raisons de la divergence des résultats des 2 laboratoires.
- iii) proposer des pistes d'améliorations en vue d'une meilleure identification des individus génétiques inconnus.

2. Méthodologie

Les deux laboratoires ont reçu les 21 mêmes échantillons du Mont-Tremblant et de sa région qu'ils ont confrontés à leur propre base de données de référence génétique. Seuls ces 21 échantillons sont *a priori* communs aux deux études. Dans les deux cas, les laboratoires ont effectués des tests d'assignation génétique multilocus à l'aide du logiciel STRUCTURE (Pritchard *et al.*, 2000), logiciel classiquement utilisé pour pareille étude. Le présent rapport discutera essentiellement des résultats obtenus par les deux Laboratoires concernant les 21 échantillons Mont-Tremblant seulement, bien que les deux rapports mentionnent les résultats d'échantillons complémentaires

fournis par la SEPAQ et le MFFP, et ayant faits pour certains l'objet de rapports antérieurs (Rutledge and White 2013; Rogic *et al.*, 2014; Rutledge and White 2014).

2.1. Préambule

En préambule, il convient de revenir sur la « philosophie » du logiciel STRUCTURE. Ce logiciel basé sur une méthode Bayésienne vise à estimer le nombre le plus probable de populations qui contribuent à un échantillon d'individus, attribuer des individus à leurs populations sources et déduire des mélanges (hybridations / admixture) et les migrations entre populations inférées, sur la base de fréquences alléliques estimées à partir de génotypes individuels multilocus. Chaque individu se verra alors attribuer une probabilité d'appartenance à une population, noté q . Cette valeur est comprise entre 0 et 1, et seront considérés comme « pure » donc pleinement assignés les individus qui ont une valeur de $q_{individu} > 0.80$ (e.g., Rutledge *et al.*, 2010), voire $q_{individu} > 0.90$ pour un assignement plus strict (Manel *et al.*, 2002). Les individus qui ont une valeur $q_{individu} < 0.80$ pour toutes les populations seront considérés « admixés » : c'est à dire assignés à aucune des populations présentes dans la base de données de référence avec une probabilité suffisamment forte. Il peut s'agir d'hybride entre populations.

La puissance du logiciel STRUCTURE sera donc grandement affectée par 1) la représentativité équitable des espèces/populations de référence dans la base de données; 2) le nombre de marqueurs génétiques utilisés pour discriminer les individus; et 3) le niveau d'hybridation entre les échantillons de référence (François and Durand 2010; Wang 2016).

Le nombre de populations inférées va dépendre de la représentativité et de l'homogénéité des individus appartenant possiblement à chaque population dans la base de données (Kalinowski 2011; Neophytou 2014; Puechmaille 2016; Wang 2016). Une espèce ou une population sous-représentée ne sera pas détectée comme entité distincte, et ses individus seront assignés à l'entité la plus proche génétiquement ou dans certains cas considérés comme issu d'hybridation entre les populations les plus proches génétiquement.

Par ailleurs, le logiciel STRUCTURE est basé sur un modèle génétique Bayesian, qui requiert de fixer un certain nombre de priors en début d'analyse (connaissances a

priori que nous avons sur les données génétiques). Les différences de paramétrage conduisent à l'existence d'une vingtaine de modèles génétiques différents dans le logiciel STRUCTURE (Pritchard *et al.*, 2000; Falush *et al.*, 2003, 2007; Hubisz *et al.*, 2009). La paramétrage du logiciel va donc avoir une influence directe sur les inférences et résultats obtenus (Porras-Hurtado *et al.*, 2013). Il est donc aisé d'obtenir des résultats différents selon les priors établis (Porras-Hurtado *et al.*, 2013; Wang 2016). Ces paramètres doivent être clairement mentionnés afin de pouvoir interpréter ou reproduire des résultats (Gilbert *et al.*, 2012). La **Figure 1** illustre ce propos, où à partir de 34 analyses publiées dans la littérature, les auteurs n'ont pu reproduire que 16 (47%) des résultats obtenus avec STRUCTURE (Gilbert *et al.*, 2012).

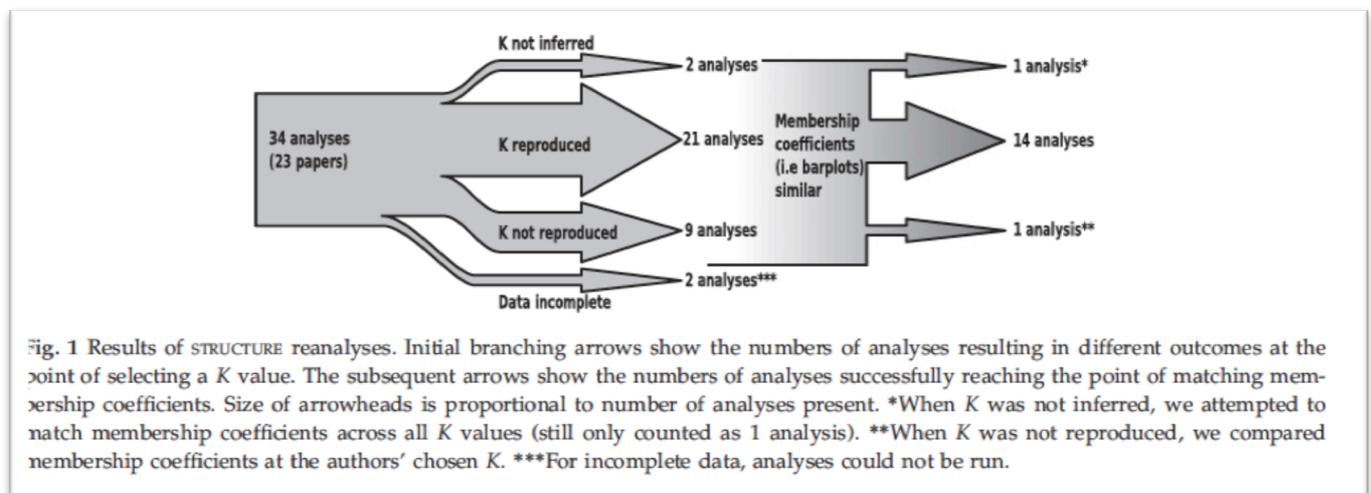


Figure 1. De la difficulté de reproduire les résultats obtenus avec le logiciel STRUCTURE. Figure tirée de Gilbert *et al.*, (2012)

3. Résultats du Premier laboratoire (sous la supervision de Nathalie Tessier)

3.1. Base de données de référence

La base de données de référence utilisée pour réaliser les analyses d'assignation génétique a été remise le 26 Janvier 2018 par courriel, copie à Hugues Tennier, par le Premier Laboratoire. Ce fichier contient une matrice des génotypes multilocus de Grands Canidés, génotypés avec 11 marqueurs microsatellites (PEZ15, PEZ19, FH2001, FH2422, Canis225, Canis109, Canis377, Canis20, Canis250, Canis204 et Canis172).

Selon l'identification du rapport remis par Nathalie Tessier (2017), la base de données utilisé pour les analyses se composent de 6 regroupements définis a priori dont la composition est décrite dans la **Table 1**.

Table 1. Description de la base de données établis par le Premier Laboratoire pour identifier les échantillons inconnus du Mont-Tremblant et sa région.

Acronyme	Description	N
#1 <i>C. latrans</i>	Coyotes (<i>Canis latrans</i>), dont 3 individus provenant de l'Ontario	70
#2 <i>C. lupus</i>	Loups gris (<i>Canis lupus</i>) de l' Ontario	8
	Loups gris (<i>Canis lupus</i>) du Québec	44
#3 <i>C. lycaon</i>	Loups de l'Est (<i>Canis lycaon</i>) de l'Ontario	12
	Loups de l'Est (<i>Canis lycaon</i>) du Québec	62
Inconnus	Échantillons du Parc national du Mont-Tremblant et de ses environs *	40
TOTAL		236

* Cette base de données inclue des échantillons issus d'une précédente étude (Rogic *et al.*, 2014).

3.2 Paramètres utilisés lors des analyses STRUCTURE

Le logiciel STRUCTURE a été utilisé selon différentes hypothèses de clustering incluant les modèles « no admixture » et « admixture » avec une valeur alpha (i.e., le niveau relatif d'admixture entre populations) de 1 et de 0,33 ($K/3$) en suivant les recommandations de Wang (2016). Le nombre d'entités génétiques présent dans le jeu de données a été testé en considérant $K=1$ à $K=8$ populations. Dix réplicats ont été calculés pour chacune des valeurs de K , sans qu'il ne soit indiqué comment les résultats issus des différents runs ont été traités par la suite. En testant différente valeur de K , le laboratoire n'a pas d'a priori sur le nombre de populations présentes dans le jeu de données. Le nombre de populations le plus vraisemblable a été inféré grâce à la statistique ΔK (Evanno *et al.*, 2005).

3.3 Interprétations des résultats

Le nombre de populations a été détecté sur la seule base du ΔK (Evanno *et al.*, 2005). Cette métrique permet de détecter le plus haut niveau hiérarchique de structure génétique présent dans un jeu de données (Evanno *et al.*, 2005), mais elle ne se suffit pas pour détecter le nombre le plus vraisemblable de populations lorsque de la sous-structure existe. Il est donc vivement recommandé (e.g., Wang 2016) d'inférer le nombre de populations présentes dans le jeu de données en combinant l'approche classique basée sur la valeur moyenne $\ln(K)$ décrite par Pritchard *et al.*, (2000) dans son papier original, avec le ΔK de Evanno *et al.*, (2005) – où d'utiliser une approche hiérarchique telle que proposée par Evanno *et al.*, (2005).

Les résultats bruts présentés permettent d'assigner au mieux 42% des individus à l'une ou l'autre des espèces considérées (**Table 2**). Les 3 modèles utilisés bien que non contradictoires - aucun individu assigné avec $q_{individu} > 0.80$ à une espèce dans une analyse n'est assigné à $q_{individu} > 0.80$ dans une autre - indiquent des différences dans l'assignation des individus, notamment pour le modèle « admixture alpha=1 ». Ces résultats démontrent que des variations dans la programmation du logiciel STRUCTURE conduisent à des résultats différents, pour le même individu génotypé dans le même laboratoire et confronté à la même base de données de référence.

Table 2. Résumé des résultats obtenus par le Premier Laboratoire. Les numéros de Tables correspondent aux Tables présentées par Tessier (2017). Table 2: modèle "no Admixture", Table 3: modèle "Admixture" alpha=1 et; Table 4: modèle "Admixture" alpha=0.33.

Sample name	ID	Table 2			Table 3				Table 4		
		#1 <i>C. latrans</i>	#2 <i>C. lupus</i>	#3 <i>C. lycaon</i>	#3 <i>C. latrans</i>	#2 <i>C. lupus</i>	#3 <i>C. lycaon</i>	#4	#1 <i>C. latrans</i>	#2 <i>C. lupus</i>	#3 <i>C. lycaon</i>
MT1-2016 (Jumeau)	CMT_01	0,68	0,00	0,32	0,29	0,02	0,16	0,53	0,65	0,05	0,30
MT2-2016 (Jumelle)	CMT_02	0,93	0,00	0,07	0,24	0,01	0,03	0,72	0,80	0,06	0,14
MT3-2016 (Boulé)	CMT_03	0,89	0,00	0,11	0,59	0,01	0,09	0,31	0,81	0,02	0,17
MT4-2016 (Cachée)	CMT_04	0,69	0,00	0,31	0,40	0,02	0,28	0,31	0,59	0,03	0,39
MT5-2016 (Foxy)	CMT_05	0,61	0,00	0,39	0,19	0,02	0,76	0,04	0,50	0,05	0,45
MT6-2016 (Galuzot)	CMT_06	0,10	0,31	0,59	0,08	0,09	0,02	0,81	0,29	0,51	0,20
MT7-2016 (Adriane)	CMT_07	0,00	0,92	0,08	0,02	0,37	0,05	0,56	0,03	0,87	0,10
MT8-2016 (Mattawinie)	CMT_08	0,60	0,28	0,12	0,45	0,24	0,10	0,20	0,48	0,41	0,11
MT9-2016 (Caravelle)	CMT_09	0,00	0,63	0,37	0,03	0,46	0,41	0,09	0,04	0,59	0,36
MT10-2016 (Libeau)	CMT_10	0,35	0,00	0,65	0,15	0,02	0,76	0,08	0,32	0,04	0,65
MT11-2016 (Lucée)	CMT_11	0,44	0,00	0,56	0,10	0,02	0,87	0,02	0,36	0,04	0,60
MT12-2016 (Cazalie)	CMT_12	0,36	0,01	0,63	0,18	0,08	0,68	0,06	0,30	0,11	0,59
MT13-2016 (Kendra)	CMT_13	0,01	0,79	0,20	0,08	0,21	0,04	0,67	0,16	0,71	0,13
MT14-2016 (Fourchu)	CMT_14	0,91	0,00	0,09	0,40	0,02	0,03	0,55	0,81	0,04	0,14
MT15-2016	CMT_15	0,95	0,00	0,05	0,62	0,03	0,07	0,29	0,81	0,04	0,15
MT16-2016	CMT_16	0,97	0,00	0,03	0,78	0,02	0,08	0,13	0,86	0,02	0,11
MT17-2016 (Crowfoot)	CMT_17	0,80	0,00	0,20	0,48	0,01	0,25	0,25	0,69	0,03	0,28
MT18-2016 (Comox)	CMT_18	1,00	0,00	0,00	0,89	0,02	0,03	0,06	0,94	0,02	0,04
MT93-2016 (Colombo)	CMT_19	0,68	0,00	0,32	0,43	0,01	0,27	0,29	0,60	0,02	0,38
MT20-2016 (Coeur-de-loup)	CMT_20	0,47	0,00	0,53	0,12	0,01	0,08	0,79	0,39	0,03	0,58
MT21-2016 (Lusignan)	CMT_21	1,00	0,00	0,00	0,92	0,01	0,02	0,05	0,96	0,01	0,03

3.4 Analyses complémentaires

3.4.1 Méthodologie

Sur la base des données brutes remises par Nathalie Tessier, des analyses complémentaires ont été effectuées:

- 1) avec le logiciel STRUCTURE 2.3.4 (Pritchard *et al.*, 2000; Falush *et al.*, 2003) afin de reproduire les résultats obtenus par Tessier (2017). C'est à dire en utilisant les modèles « no admixture », « admixture » avec ALPHA=1 et « admixture » avec ALPHA=0.33 (1/nombre de populations potentielles ; 3 espèces ici). Pour toutes les analyses, 100 000 itérations pour la période de rodage et 500 000 d'itérations de chaînes de Markov ont été utilisées (Wang 2016) et le modèle avec fréquences alléliques corrélées employé. Afin de déterminer le nombre d'itérations nécessaires, nous nous sommes assurés de la convergence des valeurs des statistiques synthétiques, comme recommandé par Pritchard *et al.*, (2007). Les autres paramètres par défaut ont été utilisés. Dix runs pour chaque valeur de K=1 à K=8 ont été calculés et les résultats analysés avec les plateformes STRUCTURE HARVESTER (Earl and vonHoldt 2012) et CLUMPAK (Kopelman *et al.*, 2015).
- 2) avec le logiciel R (R Development Core Team 2017), afin d'évaluer la capacité du jeu de données à assigner correctement un individu à sa population d'origine. Pour ce faire, une Analyse Discriminante des Composantes Principales (*Discriminant Analysis of Principal Components ; DAPC*) a été réalisée en utilisant la librairie « *adegenet* » (Jombart 2008). Cette approche multivariée peut être considérée comme un approche méthodologique alternative aux méthodes Bayésiennes implémentées dans le logiciel STRUCTURE.

3.4.2 Résultats et Discussion des analyses complémentaires

La base de données reçue diffère quelque peu de celle présentée dans le rapport remis par le Premier Laboratoire (Tessier 2017) et est détaillée dans la **Table 3**.

Table 3. Description de la base de données établis par le Premier Laboratoire pour identifier les échantillons inconnus du Mont-Tremblant et de ses environs.

Acronyme	Description	N
#1 <i>C. latrans</i>	Coyotes (<i>Canis latrans</i>)	55
#2 <i>C. lupus</i>	Loups gris (<i>Canis lupus</i>) de l' Ontario	8
	Loups gris (<i>Canis lupus</i>) du Québec	43
#3 <i>C. lycaon</i>	Loups de l'Est (<i>Canis lycaon</i>) de l'Ontario	12
	Loups de l'Est (<i>Canis lycaon</i>) du Québec	57
Inconnus	Échantillons du Parc national du Mont-Tremblant et de ses environs (n=40) * + 23 autres individus non-identifiés	73
TOTAL		248

* Cette base de données inclue des échantillons issus d'une précédente étude (Rogic *et al.*, 2014).

Basé sur le ΔK Evanno *et al.*, (2005), les 3 modèles utilisés convergent pour détecter 2 clusters génétiques principaux dans le jeu de données, soit le plus haut niveau hiérarchique présent, qui différencie selon toute vraisemblance les Coyotes et les Loups (**Figure 2**). Ce résultat est biologiquement attendu, mais est également l'un des faiblesses du ΔK (Janes *et al.*, 2017). 54% des 822 études analysées par ces auteurs identifient $K=2$ (Janes *et al.*, 2017). Cependant, l'inspection des résultats obtenus ici et basé sur le ΔK indique un second pic pour $K=5$. $K=5$ est également la partition la plus vraisemblable obtenue en utilisant le critère $\ln(K)$ moyen (**Figure 2**). Selon les analyses effectuées, il n'y a pas d'indication permettant de conclure à la présence de 3 groupes génétiques distincts dans la base de données (**Figure 2**).

Ces analyses indiquent un niveau d' admixture (hybridation) élevé entre individus, et notamment entre individus de la base de référence (**Figure 3**; voir également **Appendice 1**). Le faible pourcentage d'individus identifiés comme « pure » dans la base de référence va très certainement influencer la capacité de la matrice à assigner un génotype inconnu à l'une ou l'autre des espèces.

Finalement, les résultats obtenus avec ces analyses complémentaires, bien qu'ayant été générés *a priori* selon les mêmes paramètres, diffèrent de ceux obtenus initialement par le Premier Laboratoire (**Table 2** vs **Table 4**). Il doit être noté cependant que les probabilités d'appartenance q_{individu} sont corrélées entre les deux études : la plus haute valeur q_{individu} obtenue pour un individu - quelle soit > 0.80 ou pas - est toujours obtenu pour la même population source.

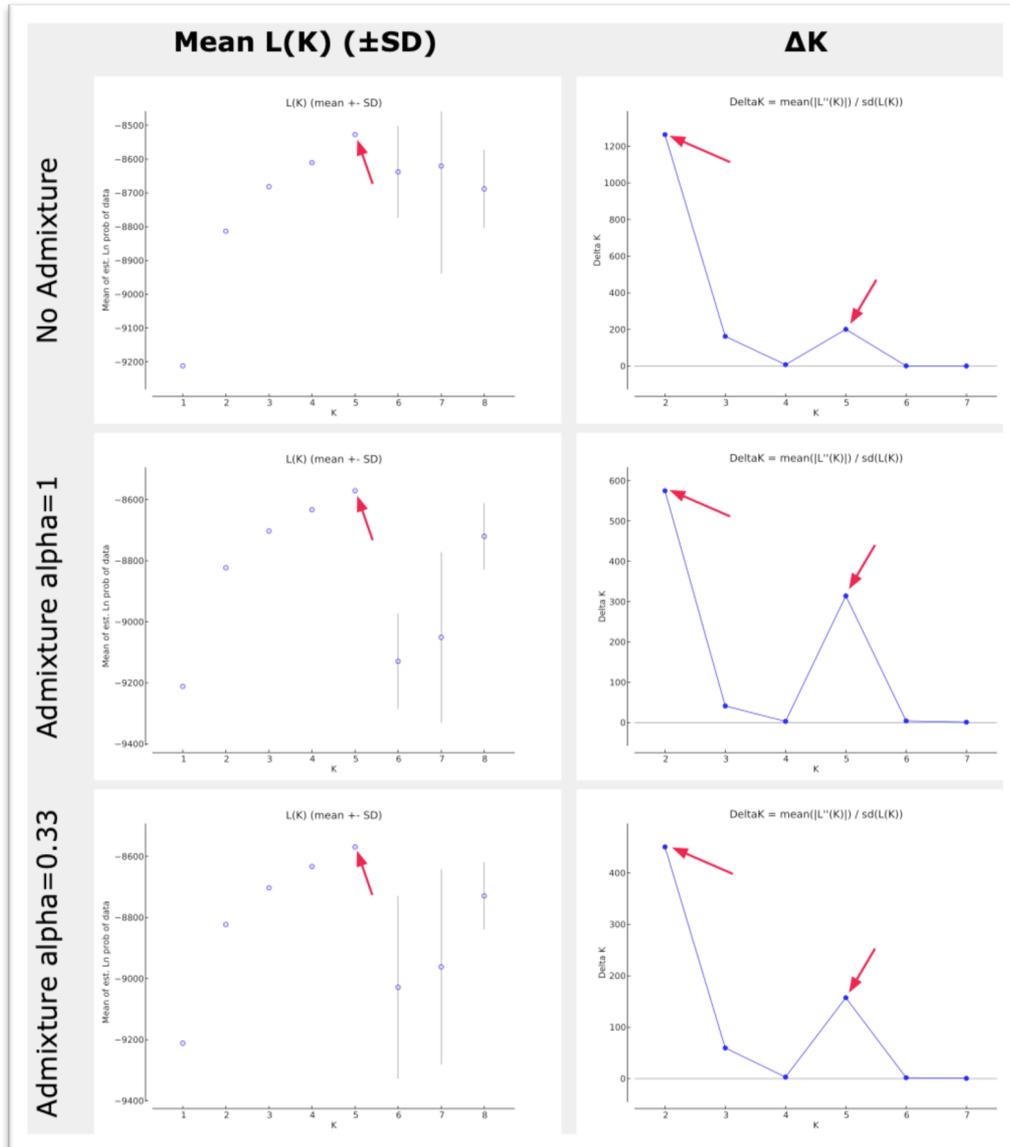


Figure 2. Nombre de populations génétiques détectées par STRUCTURE selon les 3 modèles. Les flèches rouges indiquent le nombre le plus probable de regroupements génétiques.

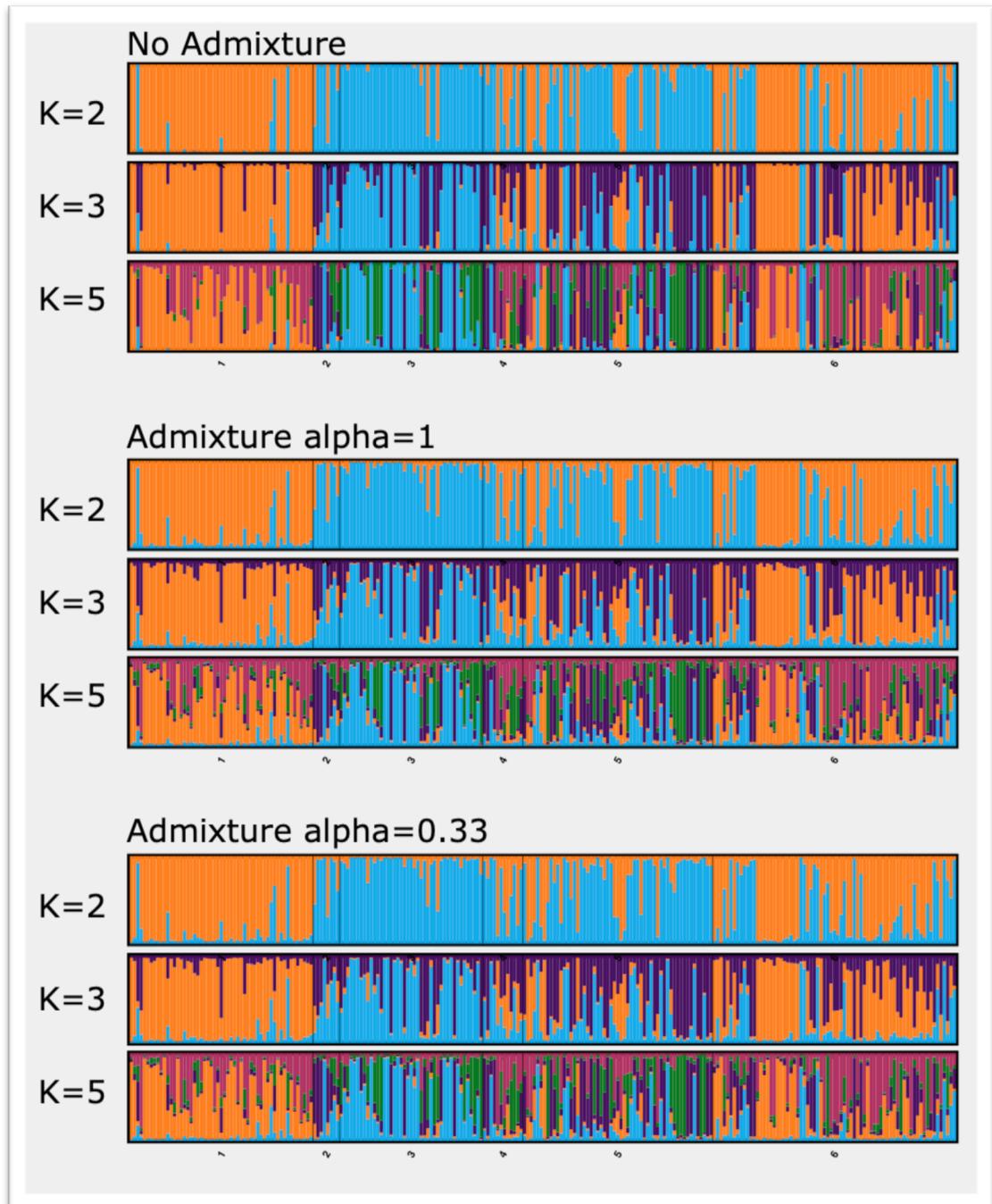


Figure 3. Valeurs d'assignation des Grands Canidés remis par le **Premier Laboratoire** pour des valeurs de $K=2$, $K=3$ et $K=5$. #1 Coyotes (*Canis latrans*); #2 Loups gris (*Canis lupus*) de l'Ontario; #3 Loups gris (*Canis lupus*) du Québec; #4 Loups de l'Est (*Canis lycaon*) de l'Ontario; #5 Loups de l'Est (*Canis lycaon*) du Québec; et finalement #6 Grands Canidés inconnus incluant ceux du Mont-Tremblant et de ses environs.

Table 4. Résumé des résultats obtenus par les analyses complémentaires, pour K=3 afin de comparaison avec les résultats obtenus par le **Premier Laboratoire**

Sample name	ID	No Admixture			Admixture alpha=1			Admixture alpha=0.33		
		# 1 <i>C. latrans</i>	# 2 <i>C. lupus</i>	# 3 <i>C. lycaon</i>	# 1 <i>C. latrans</i>	# 2 <i>C. lupus</i>	# 3 <i>C. lycaon</i>	# 1 <i>C. latrans</i>	# 2 <i>C. lupus</i>	# 3 <i>C. lycaon</i>
MT1-2016 (Jumeau)	CMT_01	0,61	0,04	0,35	0,538	0,142	0,319	0,537	0,142	0,321
MT2-2016 (Jumelle)	CMT_02	0,95	0,01	0,04	0,739	0,140	0,121	0,740	0,139	0,121
MT3-2016 (Boulé)	CMT_03	0,79	0,00	0,21	0,704	0,043	0,253	0,705	0,043	0,252
MT4-2016 (Cachée)	CMT_04	0,55	0,01	0,45	0,534	0,050	0,416	0,532	0,050	0,418
MT5-2016 (Foxy)	CMT_05	0,55	0,01	0,44	0,514	0,069	0,417	0,513	0,069	0,418
MT6-2016 (Galuzot)	CMT_06	0,03	0,82	0,16	0,278	0,623	0,099	0,280	0,621	0,099
MT7-2016 (Adrienne)	CMT_07	0,00	0,99	0,01	0,036	0,893	0,071	0,036	0,894	0,070
MT8-2016 (Mattawinie)	CMT_08	0,37	0,57	0,06	0,356	0,524	0,120	0,355	0,525	0,121
MT9-2016 (Caravelle)	CMT_09	0,00	0,63	0,36	0,044	0,613	0,343	0,044	0,611	0,345
MT10-2016 (Libeau)	CMT_10	0,18	0,00	0,82	0,307	0,064	0,629	0,303	0,064	0,633
MT11-2016 (Lucée)	CMT_11	0,18	0,00	0,82	0,327	0,048	0,625	0,325	0,048	0,627
MT12-2016 (Cazalie)	CMT_12	0,14	0,02	0,84	0,243	0,144	0,613	0,240	0,143	0,617
MT13-2016 (Kendra)	CMT_13	0,01	0,88	0,11	0,213	0,702	0,085	0,212	0,702	0,086
MT14-2016 (Fourchu)	CMT_14	0,91	0,00	0,08	0,755	0,080	0,165	0,755	0,079	0,165
MT15-2016	CMT_15	0,95	0,00	0,05	0,756	0,064	0,180	0,755	0,064	0,181
MT16-2016	CMT_16	0,96	0,00	0,04	0,793	0,041	0,166	0,793	0,041	0,166
MT17-2016 (Crowfoot)	CMT_17	0,66	0,00	0,34	0,596	0,067	0,336	0,597	0,067	0,336
MT18-2016 (Comox)	CMT_18	1,00	0,00	0,00	0,912	0,032	0,056	0,912	0,032	0,056
MT93-2016 (Colombo)	CMT_19	0,40	0,00	0,60	0,527	0,041	0,432	0,528	0,041	0,431
MT20-2016 (Coeur-de-loup)	CMT_20	0,51	0,03	0,47	0,421	0,075	0,504	0,421	0,075	0,504
MT21-2016 (Lusignan)	CMT_21	1,00	0,00	0,00	0,925	0,021	0,054	0,925	0,021	0,054

L'analyse discriminante (DAPC) effectuée sur les données de référence (5 populations) montre par ailleurs un large chevauchement génétique entre populations, en particulier entre les populations de Loups; les Loups gris du Québec, les Loups gris de l'Ontario et les Loups de L'Est du Québec notamment (**Figure 4**). Le groupe de Loups de l'Est du Québec semble par ailleurs peu homogène, avec la présence de 2 pics de densité, l'un centré sur celui du Loup de l'Est de l'Ontario, l'autre sur celui des Loups Gris. Les tests de validation-croisée - effectués avec un

premier sous-échantillonnage d'apprentissage correspondant à 90% des individus et un second sous-échantillonnage de test correspondant aux 10% restant - indiquent que la matrice génétique de référence permet d'assigner au mieux ~68% des individus à leur population d'origine.

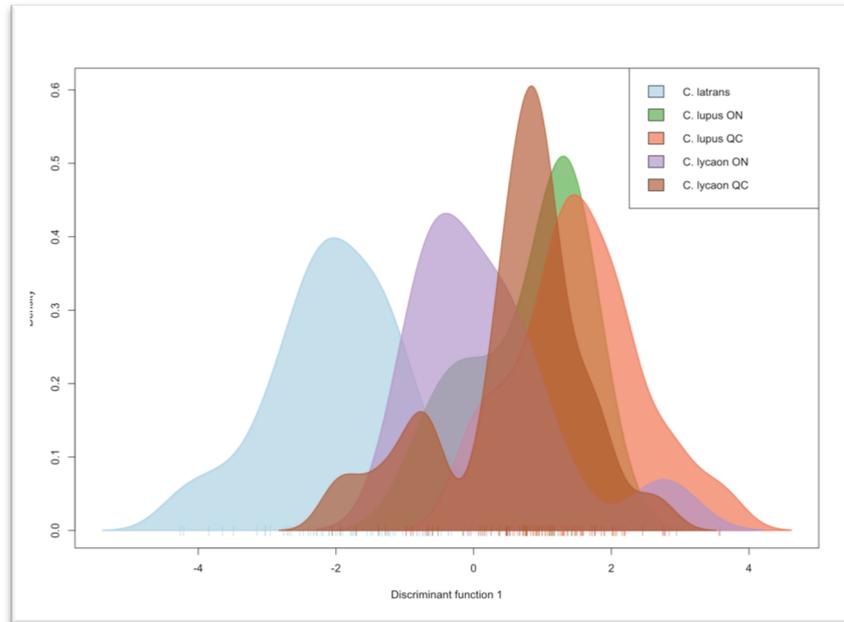


Figure 4. Diagramme de densité des individus obtenu grâce à la DAPC. Cette analyse indique un fort chevauchement génétique entre toutes les populations excepté *C. latrans*. En Bleu : Coyote (*C. latrans*); en Vert: Loup gris (*C. lupus*) de l'Ontario; en Rouge: Loup gris (*C. lupus*) du Québec; en Violet : Loup de l'Est (*C. lycaon*) de l'Ontario; en Marron: Loup de l'Est (*C. lycaon*) du Québec.

4. Résultats du Second Laboratoire (Laboratoire de Trent)

4.1. Base de données de référence

Les données issues des analyses effectuées par le **Second Laboratoire** sont parvenues le 25 janvier 2018, transmises par Hugues Tennier via WeTransfer. Ces fichiers correspondent aux résultats d'analyses effectuées avec le logiciel STRUCTURE selon différents paramètres, mais ne contenaient pas les données brutes utilisées pour générer les résultats.

Les génotypes ont été obtenus avec 12 marqueurs microsatellite (cxx225, cxx200, cxx123, cxx377, cxx250, cxx204, cxx172, cxx109, cxx253, cxx442, cxx410, et cxx147).

Un premier rapport a été remis par le Second Laboratoire portant sur l'identification des individus du Mont-Tremblant sur la base d'une matrice génétique composées de 4 populations de référence (Table 5; soit K=4) (Nguyen Nguyen *et al.*, 2017). Un second rapport a été remis en complément du rapport initial. Cette fois-ci l'identification des individus du Mont-Tremblant a été réalisée grâce à une base d'une matrice génétique composées de 5 populations de référence (Table 6; soit K=5) (Nguyen Nguyen *et al.*, 2018). Les deux bases de référence K=4 et K=5 diffèrent par l'ajout de Loup gris des Territoires du Nord Ouest et par les effectifs des différentes populations de référence (Tables 5 et 6). Dans les deux cas (K=4 et K=5), les effectifs des populations de référence sont assez débalancées (du simple au double pour K=4 et du simple au triple pour K=5).

Table 5. Description de la base de données utilisée dans le rapport initial, incluant 4 populations de référence (K=4).

Acronyme	Description	N
EC	Eastern Coyotes from southern Ontario	174
EW	Eastern Wolves from Algonquin Park	84
GLW	Great Lakes-Boreal Wolves from northern Ontario	174
DOG	Dogs	75
TOTAL		507

Table 6. Description de la base de données utilisée dans la mise à jour du rapport initial, incluant cette fois-ci 5 taxons de référence (K=5).

Acronyme	Description	N
EC	Eastern Coyotes from southern Ontario	174
EW	Eastern Wolves from Algonquin Park	84
GLW	Great Lakes-Boreal Wolves from northern Ontario	174
NWT	Grey Wolves from Northwest Territories	55
DOG	Dogs	75
TOTAL		562

4.2. Paramètres utilisés lors des analyses STRUCTURE

Le logiciel Structure a été utilisé ici selon un protocole détaillé dans Rutledge *et al.*, (2010), c'est à dire sans information *a priori* sur l'origine des individus, selon un modèle « admixture », avec les fréquences alléliques corrélées et alpha=1 (Falush *et al.*, 2003), avec 1,000,000 d'itérations MCMC précédées de 100,000 itérations de rodage. La procédure utilisée ici est différente de celle utilisée par le Premier

Laboratoire car un a priori fort est porté sur le nombre de populations dans le jeu de données de référence, soit $K=4$ dans la première analyse (Nguyen Nguyen *et al.*, 2017) et $K=5$ dans la seconde analyse qui incluait des Loups gris des Territoires du Nord-Ouest (Nguyen Nguyen *et al.*, 2018). Cinq réplicats pour chaque valeur de K ont été effectués et les résultats moyennés avec le logiciel CLUMPP (Jakobsson and Rosenberg 2007).

4.3. Interprétations des résultats

Selon les dossiers fournis par le laboratoire, les données ont été analysées avec le modèle «admixture» avec fréquences alléliques corrélées mais également avec le modèle «admixture» fréquences alléliques indépendantes (non-corrélées). Le modèle des fréquences alléliques corrélées a une plus grande puissance pour détecter des populations qui sont faiblement différenciées. Néanmoins, les résultats obtenus avec les deux modèles sont très fortement corrélés ici (corrélations de Pearson > 0.99). Au final, dans le rapport initial ($K=4$), ceux sont les résultats issus du modèle «admixture» avec fréquences alléliques corrélées qui sont présentés, alors que dans la mise à jour de ce rapport ($K=5$), les résultats issus d'un modèle «admixture» avec fréquences alléliques non-corrélées ont été retranscrits.

Une lecture initiale du rapport remis par le **Second Laboratoire** ($K=4$), puis sa mise à jour ($K=5$) a montré d'importantes contradictions dans l'assignements des individus du Mont-Tremblant à identifier. Aucune des identifications préalablement établies pour $K=4$ n'est confirmée avec $K=5$ populations !

Cependant, la consultation des fichiers de résultats fournis par le **Second Laboratoire** a permis de mettre en évidence des erreurs dans la retranscription des scores d'assignation. La **Table 7** présente les résultats corrigés et montrent une parfaite similitude entre les résultats obtenus avec les deux bases de référence ($K=4$ et $K=5$).

Pour les deux séries d'analyses ($K=4$ et $K=5$), peu d'admixture entre individus appartenant à des populations différentes de la base données n'est observée (Appendice 2; tous les $q_{\text{population}} > 0.85$), et la très grande majorité des individus de référence sont assignés à une population avec une probabilité $q_{\text{individu}} > 0.80$ (97.8% pour $K=4$ et 89.32% pour $K=5$) et $q_{\text{individu}} > 0.90$ (94.9% pour $K=4$ et 81.7% pour

K=5). Le plus grand taux d'admixture est observé au sein du groupe de référence indiqué comme étant du Loup des Grands-Lacs (Great-Lakes Wolf; Appendice 2). Néanmoins, le pourcentage d'individus du Mont-Tremblant et de ses environs assignés à l'une des espèces demeure extrêmement faible; seulement 5 sur 21 (24%) individus ont été identifiés (**Table 7**).

5. Interprétation générale des deux rapports reçus

5.1. Des différences dans les microsatellites utilisés

Les deux laboratoires ont utilisé des sets de marqueurs microsatellites différents avec un chevauchement partiel (6 loci : cxx109, cxx172, cxx204, cxx225, cxx250, cxx377 sont communs aux 2 études). Le nombre de marqueurs employés reste sensiblement le même (n=11 pour le **Premier Laboratoire** et n=12 pour le **Second Laboratoire**). Ce facteur ne devrait avoir *a priori* que peu d'effet sur l'assignation individuelle et ne devrait pas expliquer les divergences observées.

5.2. Des différences dans les bases de données de référence

L'utilisation de populations de référence différentes, aux effectifs souvent débalancés et dont le degré de « pureté » varie grandement a plus de chance de produire des résultats contradictoires.

Le **Premier Laboratoire** base sont analyses sur la présence de 3 espèces : le Loup gris (*C. lupus*), Le loup de L'Est (*C. Lycaon*) et le Coyote (*C. latrans*). Le **Second Laboratoire** utilise des bases de référence qui contiennent jusqu' à 5 espèces : le Coyote de l'Est (soit *C. latrans* x *C. lycaon*), le Loup de L'Est (*C. lycaon*), Great-Lakes Wolf (ou Loup boréal; soit *C. lupus* x *C. lycaon*), et le Loup gris (*C. lupus*). Il s'avère donc que les deux laboratoires n'ont pas utilisé le même set d'espèces pour identifier les individus du Mont-Tremblant. Le Gouvernement du Québec reconnaissant actuellement quatre regroupements génétiques distincts parmi les Canidés sauvages au Québec, le Loup gris (*C. lupus*), Le loup de L'Est (*C. Lycaon*), le Loup Boréal (*C. lupus* x *C. lycaon*) et le coyote de l'Est (*C. lycaon* x *C. latrans*), des précisions sur la nature même des espèces incluses dans les bases de référence, surtout pour celle utilisée par le Premier Laboratoire est indispensable pour s'assurer que les analyses comparent bien les mêmes choses.

Table 7. Résumé des résultats obtenus par le **Second Laboratoire**. Version corrigée en consultant les résultats bruts fournis par le laboratoire. EC : Eastern Coyote; EW : Eastern Wolf; GLW : Great-Lake Wolf; NWT : Wolf from North-West Territories; et DOG : Domestic Dog.

Sample Name	ID	K=4				K=5				
		EC	EW	GLW	DOG	EC	EW	GLW	NWT	DOG
MT1-2016 (Jumeau)	CMT_01	0,04	0,91	0,03	0,02	0,05	0,87	0,04	0,03	0,02
MT2-2016 (Jumelle)	CMT_02	0,05	0,76	0,18	0,01	0,05	0,64	0,27	0,03	0,01
MT3-2016 (Boulé)	CMT_03	0,08	0,64	0,25	0,03	0,07	0,50	0,38	0,01	0,03
MT4-2016 (Cachée)	CMT_04	0,45	0,10	0,40	0,05	0,41	0,14	0,38	0,01	0,06
MT5-2016 (Foxy)	CMT_05	0,66	0,32	0,01	0,01	0,64	0,33	0,01	0,01	0,01
MT6-2016 (Galuzot)	CMT_06	0,18	0,54	0,20	0,08	0,13	0,30	0,05	0,48	0,04
MT7-2016 (Adrienne)	CMT_07	0,01	0,79	0,19	0,01	0,01	0,56	0,41	0,02	0,01
MT8-2016 (Mattawinie)	CMT_08	0,46	0,13	0,36	0,05	0,47	0,15	0,22	0,12	0,04
MT9-2016 (Caravelle)	CMT_09	0,05	0,21	0,73	0,02	0,04	0,30	0,12	0,53	0,01
MT10-2016 (Libeau)	CMT_10	0,05	0,87	0,02	0,06	0,03	0,93	0,02	0,01	0,02
MT11-2016 (Lucée)	CMT_11	0,74	0,19	0,06	0,00	0,77	0,12	0,10	0,00	0,01
MT12-2016 (Cazalie)	CMT_12	0,61	0,13	0,25	0,01	0,65	0,11	0,20	0,04	0,01
MT13-2016 (Kendra)	CMT_13	0,24	0,10	0,64	0,02	0,15	0,19	0,05	0,61	0,01
MT14-2016 (Fourchu)	CMT_14	0,04	0,80	0,16	0,01	0,04	0,71	0,19	0,05	0,01
MT15-2016	CMT_15	0,58	0,19	0,03	0,20	0,58	0,21	0,03	0,01	0,18
MT16-2016	CMT_16	0,52	0,37	0,02	0,09	0,53	0,36	0,02	0,02	0,07
MT17-2016 (Crowfoot)	CMT_17	0,04	0,72	0,23	0,01	0,04	0,55	0,39	0,01	0,01
MT18-2016 (Comox)	CMT_18	0,97	0,02	0,00	0,01	0,97	0,02	0,00	0,00	0,01
MT93-2016 (Colombo)	CMT_19	0,05	0,74	0,21	0,00	0,06	0,49	0,44	0,01	0,00
MT20-2016 (Coeur-de-loup)	CMT_20	0,04	0,81	0,15	0,01	0,05	0,72	0,17	0,05	0,01
MT21-2016 (Lusignan)	CMT_21	0,97	0,02	0,01	0,00	0,97	0,02	0,01	0,01	0,00

Comme il a été mentionné par Tessier (2017), l'identification de nombreux spécimens constituant cette banque de génotypes de référence est incertaine, et il se pourrait

qu'elle inclut également du Loup Boréal. Ce dernier point reste à confirmer, mais la DAPC effectuée sur le jeu de données semble indiquer la présence de deux pools génétiques distincts au sein des Loups de l'Est du Québec. La nature même des échantillons constituant la banque de référence va de fait influencer les résultats d'assignation.

Les analyses effectuées par les deux Laboratoires semblent indiquer que les barrières à la reproduction ne sont pas étanches entre les différentes espèces de Grands Canidés présents dans l'aire d'étude. Dans ce cas, l'identification d'individus inconnus doit autant que possible se faire sur la base de pools génétiques « pures » et locaux, afin d'éviter que les effets de la dérive génétique ne viennent biaiser l'assignation des individus. Le plus faible pourcentage d'identification obtenu par le Second Laboratoire pourrait par ailleurs s'expliquer par l'absence d'individus génétiquement proches à ceux du Mont-Tremblant dans la base génétique de référence utilisée (la majorité des individus de cette base semblant venir de l'Ontario, voire plus loin encore pour les Loups gris des Territoires du Nord-Ouest); À cette échelle spatiale, l'effet de la dérive sera plus fort que celui la migration sur la différenciation génétique entre populations.

5.3. Des différences dans l'utilisation du logiciel STRUCTURE

Les paramètres décrits par le **Premier Laboratoire** (Tessier 2017) n'ont pas permis de reproduire exactement les résultats obtenus. Cela peut provenir des légères différences observées dans la base de données remise et celle initialement utilisée (Tessier 2017). Considérant que les deux bases sont majoritairement identiques et que les mêmes paramètres ont été utilisés avec le logiciel STRUCTURE, ces résultats indiquent que de micro-variation de fréquences alléliques liées à la présence de certains génotypes peut modifier le résultat des analyses effectuées avec STRUCTURE. Si cela est le cas, il est peu étonnant que les résultats obtenus par deux Laboratoires utilisant des bases de données a priori totalement différentes conduisent à des résultats divergents.

Les effectifs utilisés par le **Premier Laboratoire** sont relativement équilibrés entre les différentes espèces. Considérant que d'une part les Loups de l'Est du Québec et de l'Ontario et que d'autre par les Loups de Gris du Québec et de l'Ontario forment des populations génétiques homogènes, il était *a priori* pertinent de considérer l'existence de 3 populations dans le jeu de données. Les analyses complémentaires effectuées

apportent un portrait différent avec la présence plus vraisemblable de 5 groupes génétiques, avec sans doute un très fort chevauchement entre populations. La nature de ces groupes reste à définir. En revanche, les effectifs utilisés par le Second Laboratoire sont très hétérogènes et il a été démontré théoriquement que cela pouvait affecter les résultats d'assignation (Kalinowski 2011; Neophytou 2014; Puechmaille 2016; Wang 2016), mais également empiriquement sur les Canidés du Québec (Rutledge and White 2013, 2014). Dans ce rapport il a été montré que dépendamment de la base de référence utilisée, c'est à dire avec des effectifs équilibrés (New Analysis) ou non-équilibrés (Previous Analysis), les résultats variaient grandement pour des données générées dans le même laboratoire. Des analyses complémentaires pourraient être conduites en utilisant des bases de référence aux effectifs standardisés.

Finalement, les deux Laboratoires ont utilisé des approches différentes pour analyser leurs données avec le logiciel STRUCTURE (**Table 8**). Bien qu'il ait été démontré que des paramétrages différents du logiciel STRUCTURE pouvaient aboutir à des résultats différents (Porras-Hurtado *et al.*, 2013), il ne peut être affirmé ici que cela était la cause des contradictions observées entre les deux Laboratoires; alors que des différences de paramétrages conduisent à des différences d'assignations pour des analyses effectuées sur possiblement le même jeu de données (voir **Tables 2 et 4**).

Table 8. Résumé des paramètres utilisés avec le logiciel STRUCTURE par les deux laboratoires pour réaliser l'identification des individus.

	Burnin Period	MCMC répétitions	Ancestry	Alpha	Allele Frequency	Valeur de K	Nbre de répétitions	Critère pour déterminer K
Premier Laboratoire	100000	1100000	No Admixture	-	Correlated	K=1 à K=8	10	ΔK
	100000	1100000	Admixture	1	Correlated	K=1 à K=8	10	ΔK
	100000	1100000	Admixture	0.33	Correlated	K=1 à K=8	10	ΔK
Second Laboratoire	100000	1000000	Admixture	1	Correlated*	K=4	5	fixé <i>a priori</i>
	100000	1000000	Admixture	1	Uncorrelated	K=4	5	fixé <i>a priori</i>
	100000	1000000	Admixture	1	Correlated	K=5	5	fixé <i>a priori</i>
	100000	1000000	Admixture	1	Uncorrelated*	K=5	5	fixé <i>a priori</i>

* : dénotent les résultats présentés dans le rapport remis par le Second Laboratoire.

6. Conclusion

Les différences d'identification des Grands Canidés Québécois obtenus par les deux Laboratoires peuvent donc provenir de plusieurs facteurs non mutuellement exclusifs:

- De la nature et de l'identification des espèces et populations de référence utilisées (3 espèces pour le **Premier Laboratoire** et 5 espèces pour le **Second Laboratoire**);
- Du degré de « pureté » des individus utilisés comme référence;
- De l'origine géographique des individus appartenant aux différentes espèces utilisées pour identifier les canidés Québécois; soit essentiellement d'origine Québécoise pour le **Premier Laboratoire** et a priori uniquement Ontariennes pour le **Second Laboratoire**, à l'exception des échantillons de Loups gris des Territoires du Nord-Ouest. Les différences de succès d'identification observés entre les deux laboratoires pourraient refléter cette hypothèse, qui impliquerait que la diversité allélique des échantillons soumis pour identification n'est pas représentée dans la base de référence du Second Laboratoire.
- Et finalement, des paramètres d'assignation utilisées lors de l'identification des individus avec le logiciel STRUCTURE.

La contribution relative de ces différents facteurs a partiellement été explorée ici, notamment pour ce qui concerne les données mises à disposition par le **Premier Laboratoire**.

Or, globalement, les résultats obtenus ici ne permettent d'assigner qu'un faible pourcentage des individus à une espèce (~45% pour le **Premier Laboratoire**, ~25% pour le **Second Laboratoire**) et un très grand nombre d'individus demeure considéré hybride. Ce faible pourcentage est toute fois à mettre en relation avec les 27.6 % d'hybrides détectés sur l'ensemble du Québec (Mainguy *et al.*, 2017), soit un peu plus d'un quart des Grands Canidés présent sur le territoire Québécois. Ce résultat peut à la fois refléter un très haut niveau d'hybridation entre les espèces en présence et une incapacité des bases de référence à trancher sur l'origine des individus à identifier, si les individus constituant la base de référence sont tout simplement trop faiblement différenciés (sur la base de l'information génétique à disposition).

Si un tel taux d'hybridation entre toutes les espèces de Grands Canidés existe au Québec, il convient de s'interroger sur la notion même d'espèce (Coyne and Orr 1998, 2004) en présence d'un tel melting-pot génétique au sein des canidés analysés; considérant par ailleurs que les espèces décrites sont elles-mêmes issues d'hybridations inter-spécifiques plus ou moins récentes, *e.g.*, le Loup Boréal (*C. lupus* x *C. lycaon*) et le coyote de l'Est (*C. lycaon* x *C. latrans*), et dont le statut fait toujours débat dans la littérature scientifique (Rutledge *et al.*, 2010; vonHoldt *et al.*, 2011; Rutledge *et al.*, 2012; vonHoldt *et al.*, 2016; Hohenlohe *et al.*, 2017). Cela mène également à réfléchir à la place des individus et populations hybrides dans un contexte de conservation et de gestion, et à ne pas considérer les individus uniquement sur la base de leur statut génétique mais également d'intégrer leur rôle écologique et fonctionnel dans les écosystèmes (vonHoldt *et al.*, in press).

7. recommandations

Les résultats présentés se basent sur un nombre globalement limité de marqueurs utilisés pour créer les génotypes multilocus dans un contexte les flux de gènes interspécifiques semblent très important. Bien que le Second Laboratoire complète ses analyses par l'utilisation de marqueurs à transmission uniparentale (mtDNA et Chromosome Y), ces résultats montrent plusieurs incongruences d'assignation spécifique entre marqueurs, qui ne permettent pas de trancher davantage sur l'origine des spécimens. En conséquence, l'utilisation d'un plus grand nombre de marqueurs est hautement préconisée (par exemple de type SNPs; moins variables que les microsatellites, mais dont les techniques de séquençage actuelle permettent d'obtenir de 10K à 100K SNPs à un coût de revient bien moindre par unité que des microsatellites). Des puces en particulier existent pour les canidés et permettent d'inférer l'histoire évolutive, les flux de gènes et l'hybridation entre Grands Canidés avec une certaine précision (vonHoldt *et al.*, 2011; Schweizer *et al.*, 2016).

Actuellement, la distinction entre les différents types de Canidés se fait sur des variations parfois infimes de fréquences alléliques à un nombre somme toute limité de marqueurs (<20 loci). Ce faible nombre de loci ne peut capturer la complexité du génome des espèces, qui par ailleurs a connus de nombreux événements d'hybridation et d'introgession, passés ou récents (vonHoldt *et al.*, 2016; Hohenlohe *et al.*, 2017; Mainguy *et al.*, 2017). Dans ce contexte, l'utilisation ou le développement d'une puce (*e.g.*, 50K SNP) permettrait d'identifier avec davantage de

précisions les groupes génétiques distincts et leurs hybrides au Québec et serait sans aucun doute d'une aide précieuse pour la gestion et la conservation des Grands Canidés au Québec.

8. Bibliographie

- Coyne, J. A., Orr, H. A. (1998). The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London B* **353**, 287-305.
- Coyne, J. A., Orr, H. A. (2004). *Speciation* Sinauer Associates, Sunderland, Mass.
- Earl, D. A., vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.
- Evanno, G., Regnaut, S., Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
- Falush, D., Stephens, M., Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.
- Falush, D., Stephens, M., Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574-578
- François, O., Durand, E. (2010). Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources* **10**, 773-784.
- Gilbert, K. J., Andrew, R. L., Bock, D. G., *et al.* (2012). Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Molecular Ecology* **21**, 4925-4930.
- Hohenlohe, P. A., Rutledge, L. Y., Waits, L. P., *et al.* (2017). Comment on "Whole-genome sequence analysis shows two endemic species of North American wolf are admixtures of the coyote and gray wolf". *Science Advances* **3**, e1602250.
- Hubisz, M., Falush, D., Stephens, M., Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.

- Jakobsson, M., Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program with label switching and multimodality in analysis of population structure. *Bioinformatics* **3**, 1801-1896.
- Janes, J. K., Miller, J. M., Dupuis, J. R., *et al.* (2017). The K = 2 conundrum. *Molecular Ecology* **26**, 3594-3602.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.
- Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity* **106**, 625-632.
- 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. *Molecular Ecology Resources* **15**, 1179-1191.
- Mainguy, J., Hénault, M., Jolicoeur, H., Dalpé-Charron, E. (2017) Identification génétique et répartition spatiale des grands canidés sauvages au Québec, p. 82. Ministère des Forêts, de la Faune et des Parcs, Direction de l'expertise sur la faune terrestre, l'herpétofaune et l'avifaune et Direction de la gestion de la faune de Lanaudière et des Laurentides.
- Manel, S., Berthier, P., Luikart, G. (2002). Detecting Wildlife Poaching: Identifying the Origin of Individuals with Bayesian Assignment Tests and Multilocus Genotypes. *Conservation Biology* **16**, 650-659.
- Neophytou, C. (2014). Bayesian clustering analyses for genetic assignment and study of hybridization in oaks: effects of asymmetric phylogenies and asymmetric sampling schemes. *Tree Genetics & Genomes* **10**, 273-285.
- Nguyen Nguyen, T. X., Harden, M., White, B. N. (2017) Report on the DNA Analysis of Samples Submitted by Hugues Tennier - DNA Laboratory Case File Tennier1017. Natural Resources DNA Profiling & Forensic Centre (NRDPFC). Trent University, Peterborough, Canada, p. 7.

- Nguyen Nguyen, T. X., Harden, M., White, B. N. (2018) Update to Report on the DNA Analysis of Samples Submitted by Hugues Tennier - DNA Laboratory Case File Tennier1017. Natural Resources DNA Profiling & Forensic Centre (NRDPFC). Trent University, Peterborough, Canada, p. 8.
- Porras-Hurtado, L., Ruiz, Y., Santos, C., *et al.* (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics* **4**.
- Pritchard, J. K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Pritchard, J. K., Wen, W., Falush, D. (2007). Documentation for STRUCTURE software: Version 2.2, Available from <http://pritch.bsd.uchicago.edu/software>.
- Puechmaille, S. (2016). The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: sub-sampling and new estimators alleviate the problem. . *Molecular Ecology Resources* **16**, 608-627.
- R Development Core Team (2017) R version 3.3.3: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rogic, A., Tessier, N., Lapointe, F.-J. (2014) Identification des canidés du Parc national du Mont-Tremblant et de sa périphérie à l'aide de marqueurs microsatellites. Université de Montréal, Ministère des Forêts, de la Faune et des Parcs. Rapport présenté au responsable du service de la conservation et de l'éducation du Parc du Mont-Tremblant. Montréal, p. 30.
- Rutledge, L. Y., Garroway, C. J., Loveless, K. M., Patterson, B. R. (2010). Genetic differentiation of eastern wolves in Algonquin Park despite bridging gene flow between coyotes and grey wolves. *Heredity* **105**, 520–531.
- Rutledge, L. Y., Wilson, P. J., Klütsch, C. F. C., Patterson, B. R., White, B. N. (2012). Conservation genomics in perspective: A holistic approach to understanding Canis evolution in North America. *Biological Conservation* **155**, 186-192.
- Rutledge, L. Y., White, B. N. (2013) Genetic Characterization of Canis in Quebec- Analyse génétique des échantillons de chair provenant de canidés Québec.

Natural Resources DNA Profiling & Forensic Centre (NRDPFC). Trent University, Peterborough, Canada, p. 10.

Rutledge, L. Y., White, B. N. (2014) ADDENDUM: Genetic Characterization of Canis in Quebec-Analyse génétique des échantillons de chair provenant de canidés Québec. Natural Resources DNA Profiling & Forensic Centre (NRDPFC). Trent University, Peterborough, Canada, p. 10.

Schweizer, R. M., Vonholdt, B. M., Harrigan, R., *et al.* (2016). Genetic subdivision and candidate genes under selection in North American grey wolves. *Molecular Ecology* **25**, 380-402.

Tessier, N. (2017) Identification des canidés du Parc national du Mont-Tremblant et de sa périphérie à l'aide de marqueurs microsatellites (2014-2016). , p. 23. Ministère des Forêts, de la Faune et des Parcs, Direction de la gestion de la faune de l'Estrie, de Montréal, de la Montérégie et de Laval, Secteur de la faune et des parcs.

vonHoldt, B. M., Pollinger, J. P., Earl, D. A., *et al.* (2011). A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Research* **21**, 1294-1305.

vonHoldt, B. M., Cahill, J. A., Fan, Z., *et al.* (2016). Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Science Advances* **2**.

vonHoldt, B. M., Brzeski, K. E., Wilcove, D. S., Rutledge, L. Y. (in press). Redefining the Role of Admixture and Genomics in Species Conservation. *Conservation Letters* **0**.

Wang, J. L. (2016). Individual identification from genetic marker data: developments and accuracy comparisons of methods. *Molecular Ecology Resources* **16**, 163-175.

Appendice 1. Valeurs d'assignations populationnelles des Grands Canidés remis par le **Premier Laboratoire** pour des valeurs de K=2, K=3 et K=5. #1 Coyotes (*Canis latrans*); #2 Loups gris (*Canis lupus*) de l'Ontario; #3 Loups gris (*Canis lupus*) du Québec; #4 Loups de l'Est (*Canis lycaon*) de l'Ontario; #5 Loups de l'Est (*Canis lycaon*) du Québec; et finalement #6 Grands Canidés inconnus incluant ceux du Mont-Tremblant.

No Admixture																	
	K1		K2		K3		K1		K2		K3		K4		K5		
	mean	SD															
#1	0,04	0,15	0,89	0,26	0,07	0,19	0,60	0,34	0,01	0,06	0,03	0,10	0,31	0,29	0,04	0,11	
#2	0,49	0,38	0,08	0,13	0,44	0,31	0,02	0,03	0,21	0,34	0,55	0,43	0,04	0,07	0,18	0,25	
#3	0,71	0,39	0,01	0,03	0,28	0,38	0,00	0,01	0,54	0,45	0,13	0,29	0,03	0,09	0,30	0,39	
#4	0,30	0,33	0,29	0,40	0,42	0,32	0,06	0,15	0,12	0,28	0,32	0,40	0,34	0,36	0,17	0,26	
#5	0,28	0,37	0,20	0,35	0,52	0,40	0,10	0,22	0,16	0,32	0,33	0,40	0,15	0,25	0,26	0,37	
#6	0,14	0,30	0,53	0,43	0,33	0,38	0,29	0,36	0,10	0,25	0,16	0,31	0,34	0,36	0,12	0,25	
MT	0,11	0,27	0,48	0,38	0,41	0,36	0,18	0,26	0,05	0,17	0,16	0,30	0,43	0,38	0,18	0,32	

Admixture alpha=1																	
	K1		K2		K3		K1		K2		K3		K4		K5		
	mean	SD															
#1	0,09	0,13	0,80	0,22	0,11	0,15	0,53	0,29	0,05	0,07	0,06	0,08	0,28	0,24	0,08	0,09	
#2	0,51	0,29	0,14	0,15	0,36	0,23	0,08	0,07	0,27	0,31	0,48	0,33	0,08	0,09	0,09	0,10	
#3	0,65	0,34	0,05	0,07	0,29	0,32	0,03	0,04	0,51	0,36	0,16	0,24	0,06	0,08	0,24	0,28	
#4	0,35	0,27	0,25	0,27	0,40	0,24	0,07	0,10	0,13	0,20	0,29	0,32	0,29	0,27	0,22	0,25	
#5	0,30	0,31	0,19	0,26	0,51	0,34	0,10	0,16	0,18	0,24	0,32	0,33	0,16	0,20	0,25	0,31	
#6	0,17	0,25	0,50	0,35	0,33	0,30	0,27	0,29	0,11	0,20	0,18	0,25	0,31	0,29	0,12	0,18	
MT	0,15	0,21	0,46	0,29	0,39	0,28	0,17	0,19	0,09	0,16	0,17	0,24	0,41	0,31	0,16	0,23	

Admixture alpha=0.33																	
	K1		K2		K3		K1		K2		K3		K4		K5		
	mean	SD															
#1	0,09	0,13	0,80	0,22	0,11	0,15	0,53	0,29	0,05	0,07	0,06	0,08	0,28	0,24	0,08	0,09	
#2	0,50	0,29	0,14	0,15	0,36	0,23	0,08	0,07	0,27	0,31	0,48	0,33	0,08	0,09	0,09	0,10	
#3	0,65	0,34	0,05	0,07	0,29	0,32	0,03	0,04	0,51	0,36	0,16	0,24	0,06	0,08	0,24	0,28	
#4	0,34	0,27	0,25	0,27	0,41	0,24	0,07	0,10	0,13	0,20	0,29	0,32	0,29	0,27	0,22	0,25	
#5	0,30	0,31	0,19	0,26	0,51	0,34	0,10	0,16	0,18	0,24	0,32	0,33	0,16	0,20	0,25	0,31	
#6	0,17	0,25	0,50	0,35	0,33	0,30	0,27	0,29	0,11	0,20	0,18	0,25	0,31	0,29	0,12	0,18	
MT	0,15	0,21	0,46	0,29	0,39	0,28	0,17	0,19	0,09	0,15	0,17	0,24	0,41	0,31	0,16	0,23	

Appendice 2. Valeurs d'assignations populationnelles des Grands Canidés remis par le **Second Laboratoire** pour des K=4 et K=5. EC : Eastern Coyotes; EW : Eastern Wolfes; GLW : Great-Lake Wolfes; NWT : North-West Territories; et DOG : Domestic Dog.

	K1		K2		K3		K4	
	mean	SD	mean	SD	mean	SD	mean	SD
EC	0,95	0,06	0,02	0,02	0,01	0,01	0,02	0,05
EW	0,01	0,02	0,96	0,05	0,01	0,01	0,01	0,04
GLW	0,01	0,01	0,02	0,02	0,97	0,03	0,01	0,02
DOG	0,01	0,01	0,01	0,01	0,01	0,02	0,97	0,03
Inconnus	0,27	0,33	0,47	0,33	0,23	0,24	0,03	0,04

	K1		K2		K3		K4		K5	
	mean	SD								
EC	0,95	0,06	0,02	0,02	0,01	0,01	0,01	0,01	0,02	0,05
EW	0,01	0,01	0,96	0,04	0,01	0,01	0,01	0,02	0,01	0,02
GLW	0,01	0,01	0,02	0,02	0,11	0,22	0,85	0,23	0,01	0,02
NWT	0,00	0,00	0,01	0,00	0,97	0,04	0,02	0,04	0,01	0,00
DOG	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,97	0,03
Inconnus	0,27	0,33	0,41	0,31	0,07	0,14	0,23	0,25	0,02	0,03