

Outil d'identification génétique des sous-espèces d'abeilles mellifères européennes

Melanie Parejo, Jean-Daniel Charrière, Andone Estonba
Agroscope, Centre de recherche apicole, 3003 Berne
Laboratoire génétique, Université du Pays Basque (UPV/EHU),
Leioa, Espagne

Dans le cadre du projet de recherche européen SmartBees, auquel le CRA a contribué, un outil d'identification génétique des sous-espèces a été développé sur la base d'un échantillonnage étendu de la diversité des abeilles mellifères en Europe. Cet outil utilise un modèle d'apprentissage automatique pour classer les échantillons d'abeilles en 14 sous-espèces européennes différentes sur la base d'environ 4000 SNP.

Protéger la diversité des abeilles mellifères

L'abeille mellifère européenne (*Apis mellifera*) présente des variations considérables dans sa zone de distribution biogéographique et compte au moins 30 sous-espèces appartenant à différentes lignées évolutives. L'Europe héberge une grande partie de cette diversité avec de nombreuses sous-espèces endémiques représentant quatre de ces lignées évolutives, à savoir la lignée d'Europe centrale et orientale (C), la lignée d'Europe occidentale et septentrionale (M), la lignée africaine (A) et la lignée du Proche-Orient et d'Asie centrale (O). Or, l'activité humaine a progressivement réduit non seulement la diversité génétique, mais aussi la zone de distribution naturelle de certaines abeilles européennes. L'une des craintes liées au commerce et à l'importation de reines à grande échelle ainsi qu'à la transhumance sur de longues distances, est que les populations d'abeilles mellifères indigènes adaptées à leur environnement risquent d'être réduites ou perdues en raison de leur remplacement par des abeilles non indigènes ou simplement par hybridation. Il a été démontré que les abeilles mellifères adaptées localement ont une meilleure capacité de survie. Promouvoir l'élevage d'abeilles locales est donc une étape importante vers la conservation à long terme des populations d'abeilles.

De nombreux projets de zones de conservation et d'élevage des abeilles mellifères indigènes ont été lancés dans toute l'Europe. Le succès de ces programmes de conservation et de sélection dépend des mesures mises en place pour contrôler les fécondations et de la surveillance continue de l'ascendance génétique de la population. Un outil d'analyse génétique rapide, précis et facile à utiliser est indispensable à cet effet.

Le projet SmartBees a été lancé dans le but de développer de nouveaux outils pour décrire et conserver la diversité des abeilles mellifères en Europe. Sur la base d'un échantillonnage étendu de la diversité génétique des abeilles mellifères, nous avons conçu un outil moléculaire composé de marqueurs de polymorphisme nucléotidique (SNP « single nucleotide polymorphisms » en anglais) hautement informatifs, qui permet de déterminer la sous-espèce des abeilles mellifères européennes analysées. Les bases de la détermination avec la technique SNP ont été décrites dans un article dans la revue SAR de novembre-décembre 2018.

Procédure

Au total, 22 populations ont été échantillonnées, représentant les quatre lignées évolutives européennes ainsi que 14 sous-espèces, récoltées dans leur zone de répartition indigène en Europe et dans les régions adjacentes (fig. 1), y compris 100 abeilles *A. m. mellifera* de Suisse. Chaque population sélectionnée comprenait une centaine d'abeilles ouvrières provenant de colonies non apparentées, soit au total plus de 2000 échantillons finaux, ce qui constitue l'échantillonnage le plus complet des abeilles européennes à ce jour.

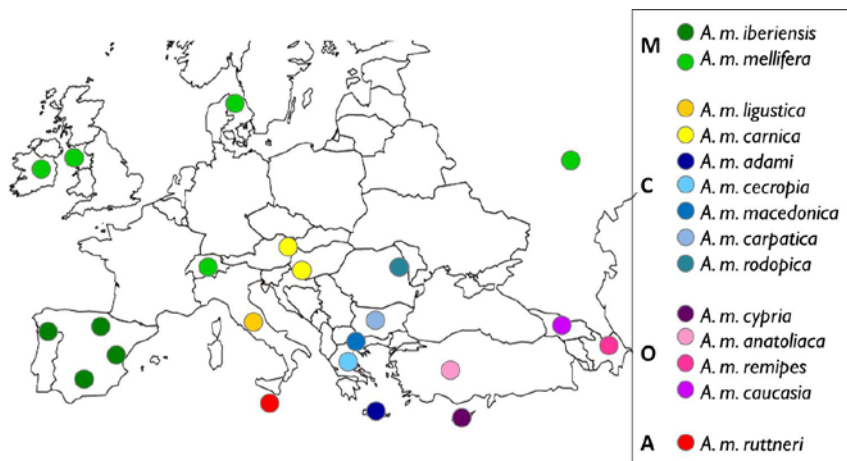


Figure 1 : Origine des 22 populations échantillonnées initialement. Dans chaque population, environ 100 abeilles ouvrières ont été récoltées. Toutes proviennent de colonies non apparentées et, si possible, de ruchers différents. Au total, plus de 2000 abeilles ont été échantillonnées.

Sur la base des données de séquençage, nous avons sélectionné les marqueurs génétiques les plus informatifs pour déterminer l'ascendance, c'est-à-dire les marqueurs qui distinguent le mieux ces 14 sous-espèces d'abeilles européennes. Au total, il existe plus de 4000 polymorphismes nucléotidiques simples (SNP). Nous avons ensuite génotypé individuellement les 100 abeilles de chaque population, soit un total de plus de 2000 analyses. Ces données de génotypage ont été visualisées à l'aide d'un diagramme appelé « t-SNE manifold plot », qui regroupe les échantillons de manière aussi compacte que possible dans une carte bidimensionnelle dans laquelle chaque sous-espèce est représentée par un symbole différent (fig. 2). Dans ce diagramme, les échantillons se regroupent dans des clusters isolés (nuages de points) qui correspondent à leur lignée évolutive ou à leur sous-espèce. La seule sous-espèce de la lignée A dans notre étude, *A. m. ruttneri*, est placée au milieu, entre les autres clusters. Dans la lignée O, les abeilles *A. m. cyprina* sont complètement séparées de *A. m. anatoliaca*, *A. m. caucasia* et *A. m. remipes*, qui sont moins différenciées. Les deux sous-espèces de la lignée M sont bien différenciées, les populations d'*A. m. mellifera* étant regroupées en trois sous-ensembles représentant des régions d'échantillonnage éloignées (région de Burzyan, Russie, sous-groupe supérieur d'*A. m. mellifera* dans la figure 2) ou isolées (île de Læsø, Danemark, sous-groupe inférieur d'*A. m. mellifera*). Les échantillons d'*A. m. mellifera* suisses sont situés dans le groupe

central avec les échantillons d'*A. m. mellifera* d'Irlande et de l'île de Man. Les échantillons de la lignée C sont regroupés en trois sous-groupes : (i) *A. m. ligustica*, (ii) *A. m. carnica* incluant une partie des échantillons d'*A. m. carpatica* et (iii) un sous-groupe hétérogène d'*A. m. macedonica*, *A. m. cecropia*, *A. m. adami*, *A. m. rodopica* et le reste des abeilles *A. m. carpatica*.

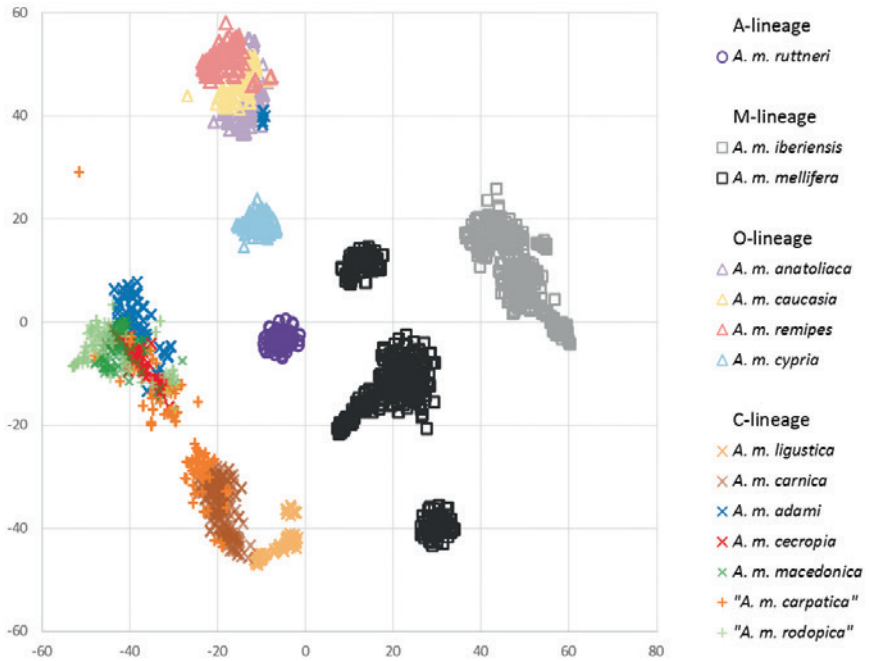


Figure 2: Visualisation (t-SNE manifold plot) des données de génotypage des échantillons de la figure 1 génotypés individuellement avec plus de 4000 SNP. Il est important de noter qu'avec ce type de graphique on entend regrouper les échantillons de manière aussi compacte que possible afin d'obtenir un regroupement optimal. Pour cette raison, les distances relatives entre les clusters et les échantillons individuels ne représentent pas la différenciation évolutive ou la parenté génétique. Par exemple, *A. m. cypria* et le sous-groupe supérieur de *A. m. mellifera* sont génétiquement très éloignés, même s'ils apparaissent comparativement proches dans ce graphique t-SNE.

Echantillons de test supplémentaires et modèle d'apprentissage automatique

En outre, 1900 autres échantillons d'abeilles ouvrières provenant de différents ruchers d'élevage SmartBees à travers l'Europe ont été génotypés. Ces 1900 échantillons ainsi que les 2000 déjà génotypés ont été utilisés pour construire un modèle statistique qui classe les abeilles européennes en utilisant des algorithmes d'apprentissage automatique. Ce modèle calcule la probabilité qu'un échantillon donné appartienne à l'une des 14 sous-espèces. L'avantage d'utiliser un modèle d'apprentissage automatique est qu'il ne repose sur aucune hypothèse préalable, ce qui permet de révéler des différences subtiles. Ceci était particulièrement pertinent pour notre étude en raison du grand nombre de sous-espèces étroitement liées que nous voulions discriminer. Plus de détails sur les méthodes de classification spécifiques peuvent être trouvés dans notre publication scientifique (Momeni *et al.* 2021).

Précision de la classification dans une sous-espèce

Le modèle a correctement classé la majorité des échantillons avec une précision moyenne de 96,2%. Ces résultats sont mieux visualisés par une matrice de confusion, qui indique les pourcentages d'échantillons dont le modèle a prédit correctement ou incorrectement l'attribution à une sous-espèce (fig. 3). Les chiffres de la diagonale centrale indiquent le pourcentage d'échantillons correctement classés, tandis que les triangles supérieurs et inférieurs indiquent les échantillons mal classés. Une classification erronée est considérée comme telle lorsque la classification dans une sous-espèce faite par le modèle est différente de celle faite a priori pour l'échantillon. Ces erreurs ne sont pas obligatoirement une « erreur » du modèle, mais peuvent aussi provenir d'une mauvaise attribution a priori des échantillons de test.

Classification avec seuil de 90%		Sous-espèce prédite										# non attribué					
		A		M		O		C									
		Ruttneri	Iberiensis	Mellifera	Cypria	Caucasia	Anatoliaca	Remipes	Carnica	Ligustica	Cecropia	Macedonica	Rodopica	Carpatuca	Adami		
Sous-espèce étiquetée	A	Ruttneri	100	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	M	Iberiensis	0	100	0	0	0	0	0	0	0	0	0	0	0	0	6
		Mellifera	0	0	100	0	0	0	0	0	0	0	0	0	0	0	48
	O	Cypria	0	0	0	100	0	0	0	0	0	0	0	0	0	0	3
		Caucasia	0	0	0	0	100	0	0	0	0	0	0	0	0	0	11
		Anatoliaca	0	0	0	0	0	100	0	0	0	0	0	0	0	0	6
		Remipes	0	0	0	0	0	0	100	0	0	0	0	0	0	0	4
	C	Carnica	0	0	0	0	0	0	0	99	0	0	0	0	0	0	70
		Ligustica	0	0	0	0	0	0	0	0	100	0	0	0	0	0	45
		Cecropia	0	0	4	0	0	0	0	0	0	92	4	0	0	0	55
		Macedonica	0	0	0	0	0	0	0	0	0	0	100	0	0	0	125
	C	Rodopica	0	0	0	0	0	0	0	0	0	0	0	100	0	0	9
		Carpatuca	0	0	0	0	0	0	0	0	0	0	0	0	100	0	15
		Adami	0	0	0	0	0	0	0	0	0	0	0	0	0	100	8
																407	

Figure 3 : Matrice de confusion pour les échantillons analysés, reflétant les pourcentages (arrondis) d'individus réellement assignés (diagonale) et les pourcentages d'individus assignés à une sous-espèce différente en utilisant un seuil de probabilité de la prédiction de 90%. Certains échantillons sont considérés comme « non attribués » et sont exclus de la matrice de confusion, comme c'est le cas pour Ruttneri, où 100% des échantillons ont été correctement classés, mais où il y avait deux échantillons non attribués. C'est également le cas pour Carnica, avec 99% des échantillons correctement attribués, mais 70 non attribués. Ce n'est pas le cas avec Cecropia, où 92% des échantillons ont été correctement attribués, tandis que 4% ont été mal classés comme Macedonica et 4% comme Mellifera.

Pour permettre l'application du modèle dans les programmes de conservation et d'élevage, nous avons fixé un seuil minimum de 90% de compatibilité. Si la probabilité de la prédiction d'un échantillon est inférieure à ce seuil, il est considéré comme « non attribué », tandis que s'il dépasse le seuil, il est attribué à la sous-espèce correspondante.

Les défis de la diversité et les prochaines étapes

La diversité considérable de l'abeille européenne a constitué un défi pour le développement d'un outil d'identification des sous-espèces. La grande divergence génétique entre les lignées évolutives (M, C, A, O) permet une distinction facile à l'aide de quelques

SNP. Cependant, la différenciation des sous-espèces est plus difficile, car la divergence entre les sous-espèces de la même lignée évolutive s'est produite plus récemment et, par conséquent, elles sont génétiquement très similaires. En outre, il existe certaines régions d'Europe où la variation des sous-espèces d'*A. mellifera* n'a pas encore été décrite de manière exhaustive, tandis que dans d'autres, l'introduction artificielle par l'homme d'autres sous-espèces brouille

les zones de répartitions naturelles des sous-espèces. Les programmes nationaux d'élevage ont également pu perturber le flux génétique naturel, altérant ainsi la base génétique des sous-espèces d'origine. Pour cette raison, certaines sous-espèces étaient plus faciles à classer que d'autres et certains échantillons n'ont pas pu être classifiés avec notre outil d'identification. Cependant, il s'agit d'un outil dynamique qui peut être actualisé afin d'étendre la base de données de référence et/ou d'inclure davantage de sous-espèces. À cet égard, les recherches en cours indiquent que cette méthode est applicable à *A. m. siciliana* de Sicile.

Utilité de cette analyse

Grâce à cet outil, il est donc possible de tester de nouveaux échantillons d'abeilles et de déterminer leur origine génétique, ce qui peut être utile entre autres dans les domaines suivants : pour les apicultrices et apiculteurs qui souhaitent connaître la sous-espèce de leurs abeilles ou savoir si elles sont croisées/hybridées ; pour les gestionnaires de la protection de la nature en Europe, pour lesquels l'identification des sous-espèces est essentielle pour surveiller le taux d'hybridation des colonies dans les zones protégées ; pour les vétérinaires afin de contrôler le commerce des reines ; pour les éleveuses et éleveurs afin de certifier l'appartenance de leurs reines à une sous-espèce et pour les apicultrices et apiculteurs afin de prouver que leur miel a été produit par une sous-espèce particulière.

Remerciements

Ce travail n'a été possible que grâce aux précieux échantillons fournis par de nombreux apicultrices et apiculteurs, éleveuses et éleveurs et autres collaboratrices et collaborateurs. Nous leur adressons nos sincères remerciements. Pour les échantillons suisses spécifiquement, nous remercions Padruot Fried et mellifera.ch. Le projet SmartBees a été financé par la Commission européenne dans le cadre de son programme FP7 KBBE (2013.1.3-02, Grant nr. 613960). Melanie Parejo a reçu une subvention du Gouvernement basque (IT1233-19).

Pour accéder à la publication scientifique complète :

Momeni, J., Parejo, M., Nielsen, R. O., Langa, J., Montes, I., Papoutsis, L., ... & Estonba, A. (2021). Authoritative subspecies diagnosis tool for European honey bees based on ancestry informative SNPs. *BMC genomics*, 22(1), 1-12.

<https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-021-07379-7>