

SEQUENÇAGE DU GENOME ENTIER ET DIVERSITE GENETIQUE DE POPULATIONS SAUVAGES ET DOMESTIQUES DE PINTADE

Vignal Alain¹, Boitard Simon¹, Thebault Noémie¹, Dayo Guiguigbaza-Kossigan², Yapi-Gnaore Valentine², Youssao Issaka³, Berthouly-Salazar Cécile⁴, Pálinkás-Bodzsár Nóra⁵, Guémené Daniel⁶, Thibaud-Nissen Françoise⁷, Warren Wesley C.⁸, Tixier-Boichard Michèle⁹ and Rognon Xavier⁹

¹GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 CASTANET TOLOSAN, ²CIRDES, 01 BP 454 BOBO DIOULASSO, ³Ecole Polytechnique d'Abomey-Calavi, Université d'Abomey-Calavi, BP 2009, COTONOU, ⁴UMR DIADE, IRD, Université de Montpellier, 34090 MONTPELLIER, ⁵Research Centre for Farm Animal Gene Conservation, H-2100 GÖDÖLLÖ, ⁶SYSAAF, Centre INRA Val-de-Loire, 37380 NOUZILLY, ⁷National Center for Biotechnology Information, NIH, BETHESDA, MD 20894, ⁸McDonnell Genome Institute, Washington University School of Medicine, SAINT LOUIS, MS 63108, ⁹GABI, INRA, AgroParisTech, Université Paris-Saclay, 78352 JOUY-EN-JOSAS

xavier.rognon@agroparistech.fr

RÉSUMÉ

La pintade, *Numida meleagris*, est une espèce appartenant à l'ordre des Galliformes. Son aire de répartition recouvre une grande partie de l'Afrique sub-saharienne, du Sénégal à l'Érythrée et du Tchad à l'Afrique du Sud. Elle a probablement été domestiquée dans le Nord de cette zone, il y a environ 2 000 ans. La pintade domestique a ensuite été largement répandue dans le monde méditerranéen, puis à partir du XVI^{ème} siècle, dans le monde entier. Aujourd'hui elle est toujours élevée, dans le cadre d'une aviculture villageoise en Afrique et de façon plus formalisée dans d'autres régions du monde, principalement en France. Jusqu'à récemment, le manque de marqueurs moléculaires a fortement limité les études génétiques concernant la pintade. Pour initier de telles études, nous avons produit la première séquence entière du génome de la pintade. Nous l'avons utilisé comme référence pour une étude par une approche Pool-seq de 12 populations sauvages et domestiques d'Europe et d'Afrique. La domestication a entraîné une perte de diversité génétique. Le statut sauvage ou domestique sépare mieux les populations africaines que leur localisation géographique. Plusieurs régions génomiques montrant des signatures de sélection, liées à la domestication, à l'importation en Europe ou aux programmes de sélection plus intensifs en Europe, ont été détectées. Elles incluent des gènes candidats intéressants, dont *EDNRB2* qui est connu pour induire des phénotypes de dilution ou de taches blanches chez les oiseaux.

ABSTRACT

Whole-genome sequencing and genetic diversity of wild and domestic guinea fowl populations

The guinea fowl, *Numida meleagris*, belongs to the Galliformes order. Its natural range includes a large part of the sub-Saharan Africa, from Senegal to Eritrea and from Chad to South Africa. The domestication of this species may have occurred about 2000 years BP in the north of this area. Domestic guinea fowl was first introduced in the Mediterranean world and since the 16th century, throughout the rest of the world. Today, it is still farmed in Africa, mainly as a village poultry. More formal livestock farming has also been developed in other countries, mainly in France. Until now, few genetic studies have been conducted on this species, and very few molecular tools are currently available. To initiate such studies, we produced a first-whole-genome sequence of guinea fowl. We used it as a reference for a Pool-seq study of 12 wild and domestic populations from Europe and Africa. Domestication has led to a loss of genetic diversity. Wild or domestic status separates African populations better than their geographical location. Several genomic regions showing selection signatures related to domestication, importation to Europe or more intensive breeding programs in Europe, were also found. They contain interesting candidate genes, including *EDNRB2* which is known to induce white spotting or extreme diffusion of plumage coloration in domesticated birds.

INTRODUCTION

La pintade, *Numida meleagris* (famille des Numidae), est originaire d'Afrique. Son aire de répartition recouvre une grande partie de la zone sub-saharienne, du Sénégal à l'Érythrée et du Tchad à l'Afrique du Sud. Elle a probablement été domestiquée, il y a environ 2 000 ans, dans le Nord de cette zone (Larson et Fuller, 2014). Elle apparaît en Europe dès l'antiquité, puis en disparaît. A partir du XVI^{ème} siècle, cette espèce sera diffusée en Europe, puis à travers le monde. Aujourd'hui elle est toujours élevée, dans le cadre d'une aviculture villageoise en Afrique et de façon plus formalisée dans d'autres régions du monde, en particulier en France, 1^{er} producteur mondial.

Jusqu'à récemment, le manque de marqueurs moléculaires a fortement limité les études génétiques concernant cette espèce. Nous avons produit un premier assemblage de la séquence du génome de la pintade. Cet assemblage a été utilisé comme référence pour une étude de la diversité génétique, par séquençage de mélanges d'ADN (Pool-seq) d'individus de populations sauvages et domestiques d'Europe et d'Afrique. L'objectif était d'appréhender les processus de domestication et de sélection ayant conduit à la diversité des populations actuelles.

1. MATERIELS ET METHODES

1.1. Le génome de référence

La séquence de référence a été produite à partir de l'ADN d'un mâle consanguin, issu de la lignée G44 (Galor). Le séquençage Illumina et l'assemblage des régions contiguës, ou scaffold, (logiciel ALLPATHS 2) ont été réalisés en collaboration avec l'Université Washington (Saint-Louis, USA). L'assemblage des scaffolds en chromosomes a été réalisé par alignement sur la référence *Gallus gallus*-5.0 du génome de la poule à l'aide du logiciel LASTZ, en tenant compte des réarrangements chromosomiques connus par des études de cytogénétique (Shibusawa et al., 2002).

1.2. Echantillonnage et séquençage

Afin d'évaluer la diversité génétique chez la pintade, nous avons étudié des individus provenant de 10 populations (Tableau 1), à savoir 3 populations sauvages africaines (1 d'Afrique du Sud et 2 du Burkina Faso), 4 populations domestiques traditionnelles africaines (3 du Burkina Faso et 1 du Bénin), et 5 populations domestiques européennes, dont 2 populations traditionnelles hongroises et 3 échantillons de lignées commerciales françaises. Les individus de la population BEG-s correspondent à un croisement commercial (1 lignée Beghin x 1 lignée Grimaud Frères Sélection) et ceux des échantillons GAL-s et GRI-s représentaient un mélange de lignées sélectionnées par les entreprises Galor et Grimaud Frères Sélection.

L'ADN a été extrait individuellement et quantifié afin de réaliser un mélange en quantités égales des échantillons. Le séquençage a été réalisé sur une plateforme Illumina HiSeq. Les lectures ont été alignées sur le génome de référence et les polymorphismes identifiés par Pool-hmm (<https://forge-dga.jouy.inra.fr/projects/pool-hmm>) par population puis dans une approche commune.

1.3. Analyses

Les fréquences alléliques des polymorphismes dans les populations ont été estimées avec la méthode Pool-HMM (Boitard et al., 2013), qui tient compte des spécificités du protocole Pool-seq (mélange des individus et erreurs de séquençage). La diversité génétique de chaque population a été évaluée d'une part par l'hétérozygotie moyenne, déduite de ces fréquences, d'autre part à partir des lectures brutes par une variante de l'estimateur de Watterson spécifique aux données Pool-seq implémentée dans le logiciel PoPoolation2 (Kofler et al., 2011). Les relations génétiques entre les populations ont été étudiées par Analyse en Composantes Principales (ACP) des fréquences alléliques.

Deux approches différentes ont été appliquées pour détecter des signatures de sélection. Premièrement, nous avons cherché à mettre en évidence les traces de sélection au niveau du génome de chaque population (Boitard et al., 2012). Cette approche recherche des régions génomiques présentant une diversité génétique réduite, tout en tenant compte des spécificités des données Pool-seq. Deuxièmement, nous avons cherché les régions génomiques différenciant fortement les populations (Fariello et al., 2017). Cette approche calcule la probabilité d'obtenir la diversité observée par hasard, les très faibles probabilités indiquant alors un événement de sélection.

2. RESULTATS ET DISCUSSION

2.1. Le génome de référence

Nous avons obtenu 2728 scaffolds pour une longueur totale de 1,04 Gb. Les N50 pour les contigs et les scaffolds sont de 232 kb et 7,8 Mb respectivement. L'assemblage final des scaffolds sur le génome étroitement apparenté du poulet est composé de 29 autosomes (seuls utilisés pour la suite de l'étude), du chromosome Z, d'un groupe de liaison (LGE64) et de l'ADN mitochondrial.

2.2. Diversité génétique

Les fréquences alléliques ont été estimées pour 10 205 115 SNP bi-alléliques. La diversité génétique, (H_e et Θ_{ws}) est moins importante dans les populations domestiques (tableau 1), en lien avec la réduction de la taille efficace résultant de la domestication.

Sur les 8 sous-espèces de Pintade, 2 sont représentées à l'état sauvage dans notre échantillonnage, *N. m. coronata* (Afrique du Sud) et *N. m. galeata* (Burkina Faso). La structure génétique mise en évidence par

l'ACP (Figure 1) montre sur le premier axe une différenciation entre *N. m. coronata* et les autres populations, en accord avec l'hypothèse de domestication initiale à partir de *N. m. galeata* en zone soudano-sahélienne (Blench, 2000 ; Larson et Fuller, 2014) et suivie par une diffusion ultérieure en Europe à partir de l'Afrique de l'Ouest (Belshaw, 1985). Parmi les populations du Burkina Faso, le statut domestiqué ou sauvage est plus discriminant que la localisation géographique. Les populations KOF-w et YAB-w se chevauchent, bien que le lieu d'échantillonnage de YAB-w soit plus proche de celui des populations domestiques SDA-t et SKO-t. Ceci laisse penser qu'il existe un très faible flux de gènes entre populations sauvages et domestiques, bien que l'élevage villageois de la pintade soit en totale liberté, à la différence de la situation observée pour la poule au Vietnam (Berthouly et al., 2009). Le deuxième axe oppose les populations ouest-africaines et européennes, avec une séparation des populations hongroises des lignées commerciales françaises.

2.3. Signatures de sélection

Dans l'approche intra-population, des milliers de régions génomiques montrent une réduction de diversité et sont donc potentiellement sous sélection. Parmi celles-ci, 5 sont potentiellement liées à la domestication, puisqu'elles ont été détectées dans au moins 6 (sur 7) des populations domestiques, sans montrer de signal significatif dans aucune des 3 populations sauvages (Figure 2a pour un exemple sur le chromosome 7). De même, 31 régions étaient potentiellement liées à l'importation en Europe, car détectées dans au moins 4 (sur 5) des populations européennes (Figure 2b pour un exemple sur le chromosome 1) et 64 à la sélection récente pour les caractères de production dans les lignées commerciales (détectées dans les 3 lignées).

Avec l'approche inter-populations, 8 régions ont été détectées. Deux de ces régions recouvrent des signatures de domestication détectées par l'approche intra-population, et une recouvre une signature européenne. Les données obtenues dans les 5 autres régions suggèrent que 3 d'entre elles sont également liées à la domestication ou à l'importation en Europe.

2.4. Régions/gènes candidats

Parmi les 223 gènes trouvés dans les régions détectées, nous avons concentré notre étude sur la présence d'allèles à fréquences extrêmes : $>0,75$ dans les populations sélectionnées et $<0,25$ dans les autres. Une liste de 58 gènes contenant de telles mutations, dont 25 avec annotations, a été établie. Nous présentons ici 2 de ces gènes candidats potentiellement impliqués dans la domestication.

La petite région de domestication de 21 kb en position 7,3 Mb sur le chromosome 7 (Figure 2a) contient *PAPPA2* comme seul gène. 27 SNP sur une portion de 20 kb du gène présentent des différences de fréquence extrêmes entre les populations sauvages et domestiquées, l'une d'entre elles modifiant la

séquence de la protéine (variant faux-sens). Ce pic étroit est observé dans les populations traditionnelles africaines du Burkina Faso et du Bénin, alors que dans les populations européennes, la signature de sélection englobe une région beaucoup plus large, de 219 kb. Une signature de sélection étroite suggère un événement ancien, alors qu'une région plus grande suggère une sélection récente. Une hypothèse expliquant cette observation pourrait être qu'une première étape de sélection aurait pu avoir lieu lorsque la pintade a été domestiquée en Afrique et qu'une deuxième s'est produite après l'importation de l'espèce en Europe, affectant un autre gène à proximité ou l'un des haplotypes restants dans *PAPPA2*. Un lien entre *PAPPA2* et le contrôle génétique de la taille corporelle a été rapporté chez la souris (Christians et al., 2006), l'homme (Lango Allen et al., 2010) et les bovins (Bouwman et al., 2018). Sa mise en évidence comme signature de domestication chez la pintade est à mettre en relation avec une domestication pour la production de viande.

Sur le chromosome 8, le locus LOC110403465, annoté *EDNRB2-like* contient aussi un variant faux sens présentant des fréquences alléliques extrêmes entre sauvages (0) et domestiques (1) et une analyse par BLAST confirme sa forte identité avec *EDNRB*. Ce gène est impliqué dans la pigmentation et plusieurs mutations codantes ont été associées à des taches blanches ou à l'extension du plumage blanc chez les poulets (Kinoshita et al., 2014).

L'observation sur les phénotypes individuels des pintades échantillonnées au Burkina Faso et au Bénin a confirmé l'implication possible de *EDNRB* dans la domestication : (1) toutes les pintades domestiques présentaient de grandes plaques de peau blanche sur la tête, tandis que les pintades sauvages avaient une peau noire sur tout le corps et la peau des pattes présentait fréquemment des zones blanches ou jaunes sur les pintades domestiques alors que celles des oiseaux sauvages étaient toujours noires; (2) 16 oiseaux sur 31 présentaient des taches blanches sur le ventre et plusieurs plumes des ailes étaient entièrement blanches, imitant le phénotype des taches blanches rencontré sur certains poulets marbrés porteurs d'une mutation de *EDNRB2*; (3) 5 oiseaux sur 31 présentaient un phénotype dilué allant du gris pâle au blanc, où un motif de taches fantômes pouvait être distingué, imitant la mutation 'mo' de *EDNRB2* décrite dans une race de poulets japonaise entièrement blanche par Kinoshita et al. (2014). Par ailleurs, *EDNRB* est exprimé dans les mélanocytes provenant de la crête neurale et est particulièrement mentionné par Wilkins et al. (2014) qui proposent qu'une altération minime des cellules de la crête neurale durant le développement embryonnaire puisse influencer de nombreux caractères morphologiques, physiologiques et comportementaux liés au syndrome de domestication.

CONCLUSION

Cette étude a révélé que les populations sauvages ont une diversité génétique plus importante que les domestiquées et que le statut sauvage ou domestique sépare mieux les populations africaines que leur localisation géographique. Le flux de gènes entre domestiques et sauvage est limité. Plusieurs régions génomiques montrant des signatures de sélection, liées à la domestication, à l'importation en Europe ou à la sélection commerciale, ont été détectées. Parmi les gènes candidats localisés dans ces régions, *EDNRB2* est connu pour induire des phénotypes de perte de coloration chez les oiseaux domestiques.

Remerciements

Les auteurs remercient les financeurs de cette étude (AGENAVI (Programme SeqVol), Département de Génétique Animale de l'INRA, AgroParisTech et Prix CSRS 2013-Fonds Eremitage pour la recherche scientifique en partenariat), ainsi que les partenaires (Entreprises de sélection, Direction Provinciale de l'Environnement du Houet, agents techniques sur le terrain) pour leur aide dans la collecte des échantillons et la plateforme Genotoul pour le séquençage des échantillons.

REFERENCES BIBLIOGRAPHIQUES

- Belshaw R.H.H., 1985. Guinea Fowl of the World. Nimrod Book Services, Liss, Hampshire.
- Berthouly C., Leroy G., Van T.N., Thanh H.H., Bed'hom B. et al., 2009. BMC Genet., (10), 1.
- Boitard, S., Kofler, R., Françoise, P., Robelin, D. et al., 2013. Mol. Ecol. Resour. 13, 337–340
- Blench R.M., 2000. In : The Origins and Development of African Livestock: archaeology, genetics, linguistics and ethnography (Blench R.M., MacDonald K.C. edit.) Routledge, New York, pp314-338.
- Boitard S., Schlotterer C., Nolte V., Pandey R. V., Futschik A., 2012. Mol. Biol. Evol., (29), 2177–2186.
- Bouwman A.C., Daetwyler H.D., Chamberlain A.J., Ponce C.H. et al., 2018. Nature Genet., (50), 362–367.
- Christians J.K., Hoeflich, A., Keightley P.D., 2006. Genetics, (173), 1547–1553.
- Fariello M.I., Boitard S., Mercier S., Robelin D., Faraut T. et al., 2017. Mol. Ecol., (26), 3700–3714.
- Kinoshita K., Akiyama T., Mizutani M., Shinomiya A., Ishikawa A., et al., 2014. PloS One, (9), e86361
- Kofler R., Pandey R.V., Schlotterer C., 2011. Bioinforma. Oxf. Engl. 27:3435–3436.
- Lango Allen H., Estrada K., Lettre G., Berndt S.I., Weedon M.N. et al., 2010. Nature, (467), 832–838.
- Larson G., Fuller D.Q., 2014. Annu. Rev. Ecol. Evol. Syst., (45), 115–136.
- Shibusawa M., Nishida-Umehara C., Masabanda J. et al., 2002. Cytogenet. Genome Res., 98, 225–230.
- Wilkins A.S., Wrangham R.W., Fitch W.T., 2014. Genetics, (197), 795–808.

Tableau 1 : Description de l'échantillonnage utilisé pour les analyses en Pool-seq

Les localisations géographiques et le nom des entreprises de sélection sont indiqués avec le nom de l'échantillon.
 *A cause de la faible profondeur de séquençage, ces 3 échantillons seront fusionnés en un seul échantillon 'Burkina Faso', nommé BUR-t, pour les analyses de signature de sélection.

Echantillon	Statut	Nombre d'individus	Profondeur de séquençage	H_e	Θ_w
AFS-w: Afrique du Sud	sauvage	3	14,47	0,1736	0,0035
KOF-w: Koflandé, Burkina Faso	sauvage	8	17,64	0,1319	0,0034
YAB-w: Yabé, Burkina Faso	sauvage	8	20,08	0,1499	0,0034
SDA-t: Sara-Dan, Burkina Faso*	traditionnel	5	9,21	0,0885	0,0021
SKO-t: Sarakongo, Burkina Faso*	traditionnel	5	13,01	0,1097	0,0024
DOR-t: Dori, Burkina Faso*	traditionnel	5	7,59	0,0941	0,0022
BEN-t: Bénin	traditionnel	15	15,69	0,1150	0,0026
GOD-t: Godollo, Hongrie	traditionnel	30	16,76	0,0944	0,0023
HAR-t: Hortobagy, Hongrie	traditionnel	30	18,04	0,1037	0,0022
BEG-s: Béghin, France	sélectionné	12	15,06	0,0999	0,0022
GAL-s: Galor, France	sélectionné	29	37,01	0,1125	0,0020
GRI-s: Grimaud Frères Sélection, France	sélectionné	20	38,71	0,1077	0,0019

Figure 1. (A) localisation des populations d’Afrique de l’Ouest. (B) Analyse en Composantes Principales (ACP) réalisée à partir des fréquences des polymorphismes identifiés. Voir le tableau 1 pour le nom complet des populations. Code couleur : bleu = populations sauvages ; vert = populations domestiques traditionnelles africaines ; orange = populations domestiques traditionnelles européennes ; rouge = lignées commerciales sélectionnées.

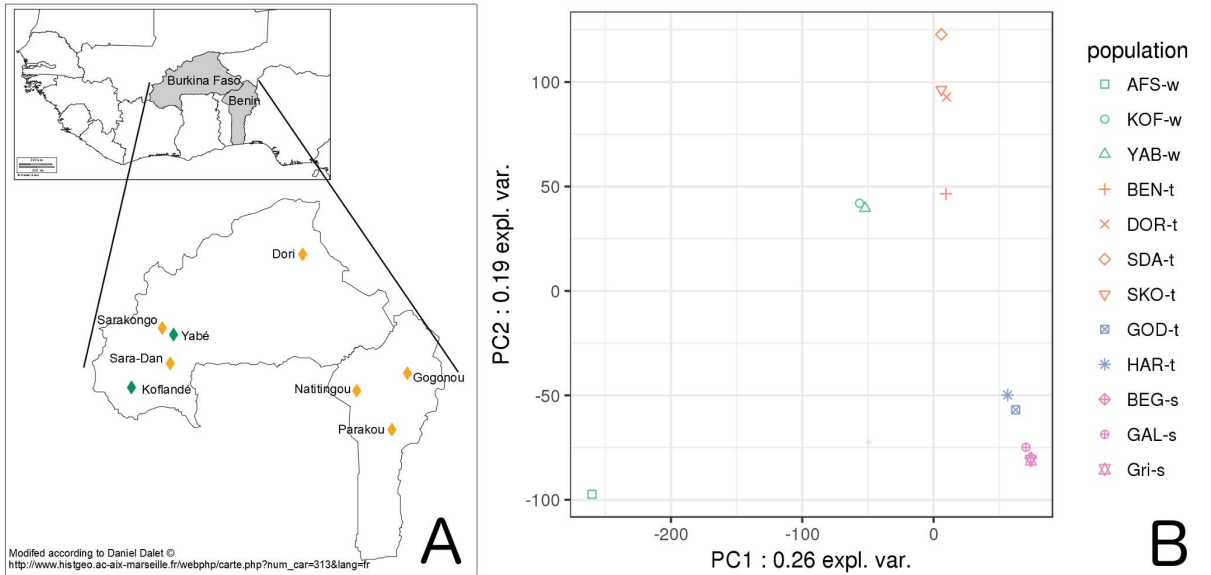


Figure 2. Profils de diversité génétique pour deux régions sous sélection, et dans les 10 populations. (2a) région localisée sur le chromosome 7, en lien avec la domestication ; (2b) région localisée sur le chromosome 1, en relation avec l’importation des pintades vers l’Europe. Pour chacune des populations, la fréquence de l’allèle majeur (axe y) est représentée pour tous les variants bi-alléliques de la région, selon leur position génomique (axe x). Les points rouges indiquent la région détectée dans chaque population et les lignes rouges indiquent l’union des régions détectées lors de la combinaison de toutes les populations sélectionnées.

