

Des signatures génomiques de domestication homologues chez la chèvre et le mouton

François Pompanon

Université Grenoble Alpes, Université Savoie Mont-Blanc, CNRS, LECA, 38000 Grenoble, France.

francois.pompanon@univ-grenoble-alpes.fr

> Les processus de domestication ont joué un rôle essentiel dans l'histoire de l'humanité en permettant la sédentarisation des sociétés humaines. Après le chien, domestiqué il y a plus de 15 000 ans, les premières espèces domestiquées au néolithique dans le croissant fertile ont permis l'émergence de l'agriculture : le blé et l'orge pour les cultures, la chèvre et le mouton pour l'élevage. On peut considérer la domestication comme la phase finale d'un processus d'intensification des relations entre l'homme et des populations animales ou végétales [1] qui se sont différenciées de celles issues de la même espèce mais demeurées sauvages. Ce processus a modelé profondément la morphologie, la physiologie et le comportement des animaux captifs par rapport à leurs homologues sauvages, par le biais de sélections que l'homme a opérées pour répondre à ses besoins, qu'ils soient liés à l'alimentation (comme le lait ou la viande), à l'habillement (comme la laine ou le cuir), ou au travail (par exemple la force motrice ou la fabrication d'outils).

Le syndrome de domestication

Dès 1868, Charles Darwin [2] identifiait des modifications qui étaient retrouvées chez pratiquement tous les mammifères domestiques : la réduction de la taille, le pelage pie, les poils bouclés, les oreilles tombantes, etc. Un tel ensemble de convergences¹ définit le syndrome de domestication qui suggère l'existence, chez tous les mammifères, d'un même

groupe de gènes sélectionnés parallèlement lors de leurs différentes domestications. Les trajectoires de domestication ont suivi depuis des millénaires différentes étapes d'amélioration zootechnique par la sélection intentionnelle de caractères liés à la productivité, comme la masse musculaire ou la qualité du lait. On pense que, *a contrario*, les caractères constituant ce syndrome de domestication auraient été sélectionnés involontairement lors des étapes précoces de la domestication, alors que la docilité ou les taux de développement étaient ciblés. La compréhension fine de ces phénomènes nécessite l'étude des mécanismes génétiques qui en sont responsables et de nouvelles voies de recherche très prometteuses ont été ouvertes dans ce domaine ces dernières années, grâce aux développements des techniques de séquençage à haut-débit de l'ADN et de méthodes d'inférence des processus évolutifs (→ Voir la Chronique génomique de B. Jordan *m/s* n° 12, décembre 2012, page 1139). Il est maintenant possible de caractériser les génomes de dizaines, voire de centaines, d'individus d'une même espèce afin d'identifier les régions génomiques différenciant les animaux sauvages des animaux domestiques. On peut ainsi citer, de façon non exhaustive, la découverte de gènes impliqués dans le processus de domestication en lien avec la capacité à digérer l'amidon, chez le chien [4], ou avec la taille et le comportement agressif, chez le cochon [5], ou la morphogenèse et le développement du système nerveux, chez le lapin [6].

Chèvres et moutons : un cas d'école pour étudier les convergences

À ce jour, aucune étude n'avait encore exploré la possibilité que des mêmes régions génomiques aient été impliquées dans des processus de domestication indépendants. Pour tester cette hypothèse, notre consortium (constitutif du projet NEXTGEN²) a tiré parti des trajets de domestication parallèles retrouvés chez deux espèces phylogénétiquement très proches : la chèvre (*Capra hircus*) et le mouton (*Ovis aries*) [7]. Les ancêtres sauvages de ces petits ruminants de la famille des Caprinae ont divergé il y a environ 6 millions d'années. Ils ont été domestiqués il y a environ 10 500 ans dans la même région du Moyen-Orient (l'Anatolie et les monts Zagros d'Iran). Depuis cette période, chèvres et moutons ont suivi les migrations humaines et ont colonisé toute la planète. Les espèces sauvages, l'aegagre (*Capra aegagrus*) et le mouflon asiatique (*Ovis orientalis*), sont, elles, restées dans leur région native. Afin de comparer les bases génomiques de la domestication de ces deux espèces, nous avons réalisé, pour chacune d'elles, plus de 70 analyses de génomes complets individuels. L'échantillonnage est représentatif de populations sauvages iraniennes et de races domestiques issues de différentes régions du monde (incluant la zone de domestication initiale) et de différents modes d'élevage (traditionnel ou industriel). Nous avons ainsi identifié plus de 33 millions

¹ Convergence : évolution indépendante vers un même état dans des lignées évolutives différentes.

² <https://nextgen.epfl.ch>

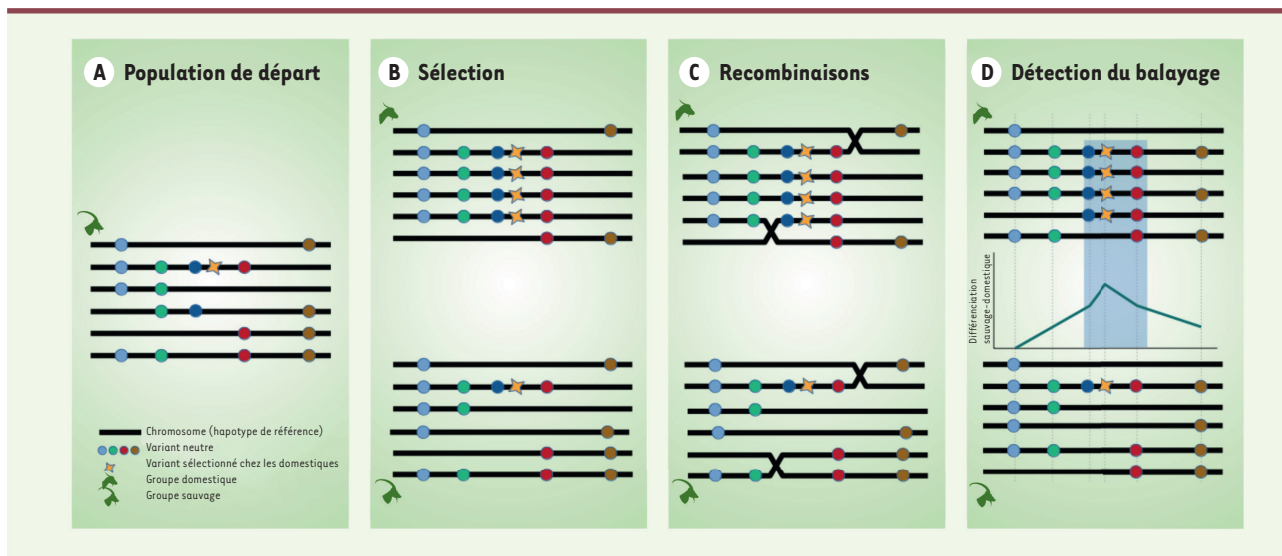


Figure 1. Balayage sélectif : mécanisme et détection. **A.** Représentation simplifiée d'un échantillon d'haplotypes représentant la variabilité génétique d'une population sauvage. **B.** La population est scindée en 2 groupes. Dans le groupe domestique, le variant jaune lié à un caractère d'intérêt est sélectionné, entraînant avec lui la portion de chromosome qui l'entoure. L'haplotype portant cette mutation augmente en fréquence au fil des générations. Dans le groupe sauvage, les fréquences des haplotypes varient aléatoirement (dérive génétique). **C.** Parallèlement aux variations des fréquences, les recombinaisons chromosomiques se produisant lors des méioses modifient les associations entre variants au sein des haplotypes, d'autant plus que les variants sont physiquement éloignés. **D.** La région génomique touchée par le balayage sélectif est détectée car elle correspond à une zone où la différenciation entre les groupes sauvage et domestique augmente localement par rapport au reste du génome. Recombinaison : formation d'une nouvelle association d'allèles au sein d'un haplotype par échange d'ADN entre les bras de chromosomes homologues lors de la méiose.

de variants génomiques (des SNP³ et de courtes insertions/délétions) chez le mouton et plus de 23 millions chez la chèvre. L'étude de cette variabilité a confirmé, pour les deux espèces, une nette différenciation génétique globale entre individus sauvages et domestiques, traduisant l'absence d'hybridation entre les deux populations dans leur histoire récente, ce qui aurait pu compromettre la validité des analyses ultérieures.

Les gènes impliqués dans le processus de domestication

L'objectif de notre étude était de rechercher les régions génomiques permettant de différencier l'ensemble des groupes domestiques des populations sauvages.

³ Balayage sélectif : processus par lequel la sélection d'un allèle favorable augmente également la fréquence des allèles qui sont physiquement liés réduisant la variation génétique autour du locus sélectionné. Ce processus se produit lorsque la sélection est suffisamment rapide pour ne pas permettre aux recombinaisons de rompre le lien entre le locus sous sélection et les locus proches.

Ces zones de forte différenciation dans les génomes sont en effet les signatures de balayages sélectifs⁴ (Figure 1). Nous avons ainsi identifié 90 régions génomiques comme cibles probables de la domestication (46 chez le mouton et 44 chez la chèvre), pour lesquelles les haplotypes⁵ présentaient un fort degré de similarité pour tous les groupes domestiques (Figure 2). Plus de 80 % de ces régions montrent une diversité génétique réduite dans les haplotypes domestiques par rapport aux sauvages, ce qui est compatible avec une sélection stabilisante⁶ chez les domestiques. Pour les autres régions

⁴ Balayage sélectif : processus par lequel la sélection d'un allèle favorable augmente également la fréquence des allèles qui sont physiquement liés, réduisant la variation génétique autour du locus sélectionné. Ce processus se produit lorsque la sélection est suffisamment rapide pour ne pas permettre aux recombinaisons de rompre le lien entre le locus sous sélection et les locus proches.

⁵ Haplotype : combinaison d'allèles sur une portion de chromosome (génotype haploïde).

⁶ Sélection stabilisante : sélection favorisant l'apparition de phénotypes intermédiaires et entraînant une baisse de la diversité génétique au locus sélectionné.

génomiques, la diversité s'est révélée plus importante chez les domestiques, suggérant une sélection diversifiante⁷ dans ce groupe, ou le relâchement d'une pression de sélection qui n'aurait agi que sur les sauvages. On attend en effet que la captivité conduise à la disparition des facteurs sélectifs en lien avec la vie en milieu sauvage (la prédation par exemple), qui contraignaient certains caractères phénotypiques (par exemple une coloration homogène du pelage chez tous les individus en lien avec la nécessité de camouflage). Pour 59 des régions génomiques identifiées, nous disposons d'annotations fonctionnelles qui permettent de déterminer les gènes sous sélection et leur fonction. Il apparaît ainsi que la plupart des gènes affectés sont connus, chez les animaux d'élevage, pour affecter l'expression de phénotypes liés 1) à la réponse immu-

⁷ Sélection diversifiante : sélection favorisant l'existence de plusieurs phénotypes et responsable d'une forte diversité génétique au locus sélectionné.

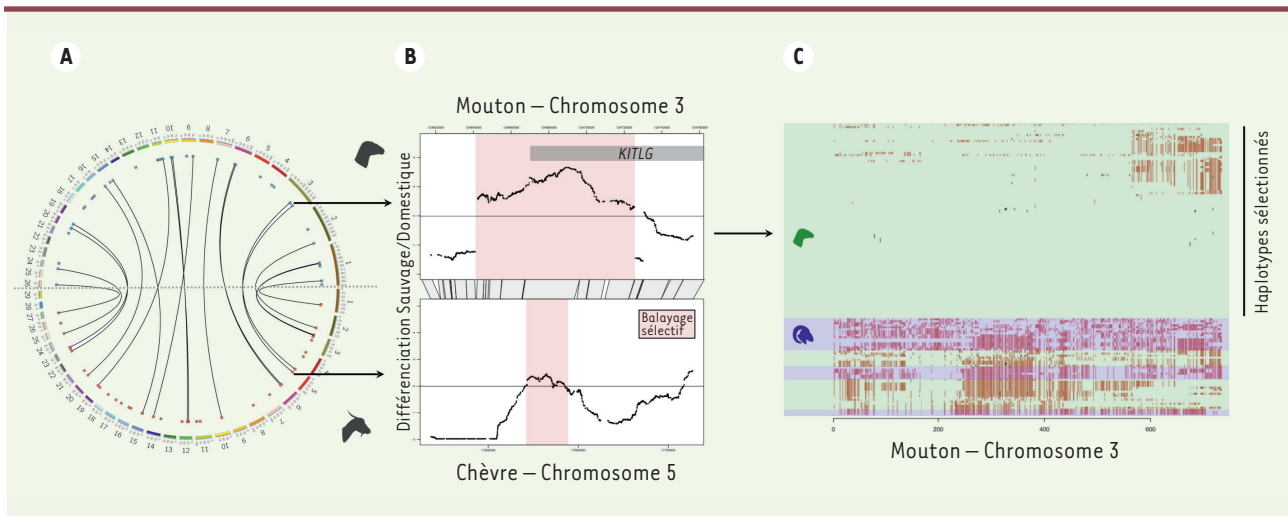


Figure 2. Caractérisation des régions génomiques influencées par la domestication chez la chèvre et le mouton. **A.** Localisation des signatures de sélection liées à la domestication. Chaque tiret de couleur numéroté représente un chromosome de mouton (moitié supérieure) ou de chèvre (moitié inférieure) ; les points rouge et bleu localisent les régions sélectionnées qui sont reliées lorsqu'elles sont homologues entre les deux espèces. **B.** Exemple de deux régions homologues présentant un balayage sélectif. La différenciation entre groupes domestique et sauvage le long de la portion de chromosome est estimée en fonction du taux de détection de faux positifs (*false discovery rate*) exprimé en $-\log_{10} \text{FDR-val-}q$ (la valeur $-q$ est l'équivalent de la valeur p des tests statistiques dans une approche de faux positifs), considéré significatif au seuil de $\text{val-}q = 10^{-2}$, soit un pourcentage de faux positifs de 1 %. Le balayage sélectif inclue une portion du gène *KITLG* (*KIT proto-oncogene receptor tyrosine kinase ligand*) dont la position est représentée par une barre grisée. **C.** Haplotypes de moutons dans la portion sous sélection. Chaque ligne représente un individu (bleu : sauvage, vert : domestique), l'alternance des couleurs représente la variation de la séquence d'ADN par rapport au génome de référence du mouton et les points noirs représentent les données manquantes. Les haplotypes sont ordonnés selon la proximité de leurs séquences.

nitaire (11 gènes), 2) à la productivité (le lait, 11 gènes ; la viande, 11 gènes ; le pelage, 4 gènes ; la fertilité, 2 gènes), 3) au développement neural et au système nerveux (5 gènes). La presque totalité des variants génomiques retrouvés dans les régions sélectionnées sont dans des séquences non codantes, que ce soit dans des régions intergénomiques (36 %), introniques (50 %), ou en amont ou en aval de gènes (14 %). Seuls 3 variants, sur plus de 1 000 identifiés, sont dans des exons. L'ampleur de ce résultat ne peut s'expliquer entièrement par les lacunes, encore existantes, des annotations fonctionnelles des génomes de référence des espèces étudiées. Elle laisse supposer que des séquences régulatrices de l'expression des gènes seraient une cible privilégiée de la sélection [7].

Des signatures génomiques de domestication trans-spécifiques

Le résultat majeur de cette étude est que près de 45 % des régions liées à la

domestication sont homologues entre la chèvre et le mouton ! Les principaux gènes pour lesquels un rôle a été identifié dans les 2 espèces ont des effets pléiotropes, dont les fonctions sont connues chez plusieurs espèces domestiques. Ainsi, *KITLG* (*KIT proto-oncogene receptor tyrosine kinase ligand*) est associé au développement neural, à la pigmentation du pelage et à la fertilité [8], *HMGI-C* (*high mobility group AT-hook 2*) est lié au nanisme [6], *MTMR7* (*myotubularin-related protein 7*) à la qualité de la graisse, *NBEA* (*neurobeachin*) à la neurotransmission et à la qualité de la laine [9]. Ce résultat constitue un premier pas vers la compréhension des mécanismes responsables du processus de domestication. Il montre que des événements de domestication indépendants ont impliqué des régions génomiques homologues. On peut penser que la nature pléiotrope des gènes sélectionnés a facilité, lors des premières étapes du processus de

domestication, la sélection d'un large spectre de caractères en lien avec le comportement, le développement ou la productivité, alors même que certains n'étaient pas ciblés volontairement.

Conclusion et perspectives

Il est probable que les gènes, dont les allèles sélectionnés sont retrouvés dans une importante proportion d'animaux au niveau mondial (Figure 2), aient été sélectionnés avant la dissémination des populations domestiques à la surface du globe, c'est-à-dire lors des phases précoces de la domestication en marge du croissant fertile. Cependant, on ne peut exclure des événements plus tardifs, éventuellement d'origine géographique différente, qui auraient été fixés secondairement dans les populations actuelles, consécutivement à des remplacements *via* des vagues de colonisation successives. De tels phénomènes de remplacement sont documentés, notamment chez le cochon [10]. L'ana-



lyse de génomes actuels et de génomes anciens reconstitués à partir d'ossements retrouvés dans des sites archéologiques permettra de répondre à ces questions, et plus largement de reconstituer la chronologie des événements de sélection et les trajectoires évolutives des génomes lors de la domestication. Les travaux à venir devront également permettre de préciser, sur un plus large ensemble d'espèces, les cibles génétiques du syndrome de domestication. ♦

Homologous genomic signatures of domestication in goat and sheep

REMERCIEMENTS

Ce travail a été financé par l'Union européenne (projet NEXTGEN, FP7 n°244356), l'ANR (FACCE ERA-

NET Plus project CLIMGEN, ANR-14-JFAC-0002-01), le LabEx OSUG@2020 (Investissements d'avenir - ANR10LABX56), le Wellcome Trust (WT108749/Z/15/Z) et l'European molecular biology laboratory (EMBL).

LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

1. Vigne JD. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *C R Biol* 2011 ; 334 : 171-81.
2. Darwin C. *The variation of animals and plants under domestication*. London : John Murray, 1868 : vol 1 et 2.
3. Jordan B. Néanderthal et *Homo sapiens* : to meet or not to meet ? *Med Sci (Paris)* 2012 ; 29 : 1129-1132.
4. Axelsson E, Ratnakumar A, Arendt M-L et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 2013 ; 495 : 360-4.
5. Frantz LAF, Schraiber JG, Madsen O, et al. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat Genet* 2015 ; 47 : 1141-8.
6. Carneiro M, Rubin C-J, Di Palma F, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 2014 ; 345 : 1074-9.
7. Alberto FJ, Boyer F, Orozco-terWengel P, et al. Convergent genomic signatures of domestication in sheep and goats. *Nat Commun* 2018 ; 9 : 813.
8. An XP, Hou JX, Lei YN, et al. Two mutations in the 5'-flanking region of the KITLG gene are associated with litter size of dairy goats. *Anim Genet* 2015 ; 46 : 308-11.
9. Wang Z, Zhang H, Yang H, et al. Genome-wide association study for wool production traits in a Chinese Merino sheep population. *PLoS One* 2014 ; 9 : e107101.
10. Ottoni C, Flink LG, Evin A, et al. Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. *Mol Biol Evol* 2013 ; 30 : 824-32.

NOUVELLE

La protéine L du virus Ébola porte une nouvelle activité enzymatique impliquée dans la méthylation interne des ARN

Baptiste Martin^{1,*}, Coralie Valle^{1,*}, Bruno Coutard¹, Bruno Canard¹, Françoise Debart², Étienne Decroly¹

¹AFMB, CNRS, Université Aix-Marseille, UMR 7257, Case 925, 163 Avenue de Luminy, 13288 Marseille cedex 09, France

²IBMM, Université de Montpellier, CNRS, ENSCM, Montpellier, France

*Contribution égale

etienne.decroly@afmb.univ-mrs.fr

► L'infection par le virus Ébola conduit à des fièvres hémorragiques souvent fatales. Depuis sa découverte en 1976 en République Démocratique du Congo, plus d'une vingtaine d'épidémies d'Ébola ont été recensées en Afrique. La plus importante a eu lieu en 2015-2016, entraînant le décès de plus de 11 000 personnes [1]. En raison du risque de réémergence de ce virus, et en l'absence de vaccins et d'agents thérapeutiques approuvés, la recherche de médicaments ciblant le virus Ébola est une priorité pour l'Organisation mondiale de la santé. Le virus Ébola est un virus enveloppé possédant un ARN génomique simple brin de polarité négative et appartient à l'ordre des *Mononegavirales*. Cet ordre est divisé en 5 familles de

virus contenant des pathogènes humains importants, comme les virus de la rage, de la rougeole, ou encore le virus respiratoire syncytial (VRS). Le virus Ébola fait partie de la famille des *Filovirus* (→) [2] et appartient au genre *Ebolavirus* qui contient 5 espèces : Zaïre, Soudan, Taï Forest, Bundibugyo et Reston. Le génome des filovirus code sept protéines virales produites à partir d'ARN messagers (ARNm) transcrits de manière séquentielle. Parmi ces protéines, la protéine « Large » (L) est un des acteurs majeurs de la propagation de l'infection puisqu'elle contient les activités enzymatiques nécessaires à la réplication/

(→) Voir la Synthèse de B. Martin et E. Decroly, *m/s* n° 8-9, août-septembre 2018 page 671

transcription et à la maturation des ARN viraux [3]. N'ayant pas d'homologues cellulaires, la protéine L constitue une cible antivirale attractive. Cette protéine contient six régions conservées (CR) entrecoupées de régions variables. La structure, récemment résolue par cryo-microscopie, de la protéine L du virus VSV (virus de la stomatite vésiculaire) indique que cette protéine est organisée en 4 domaines structuraux distincts (Figure 1) [4]. L'extrémité N-terminale de la protéine (régions CRI à CRIII) forme le domaine ARN polymérase ARN-dépendante (RdRp) intimement intriqué avec le domaine Cap (régions CRIV et CRV) qui porte l'activité polyribonucléotidyltransférase (PRNTase)