Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com

Trajectories of genetics, 150 years after Mendel/Trajectoires de la génétique, 150 ans après Mendel

Genetic and epigenetic variation of human populations: An adaptive tale

Variabilité génétique et épigénétique des populations humaines : une histoire adaptative

Lluis Quintana-Murci

Unit of Human Evolutionary Genetics, CNRS URA3012, Institut Pasteur, 25–28, rue du Docteur-Roux, 75015 Paris, France

A B S T R A C T

The evolutionary history of modern humans means much more than their demographic past. It includes the way in which humans have had to genetically adapt to the different environments they have encountered—nutritional, climatic or pathogenic—as well as the different epigenetic responses elicited by such environmental cues. Detecting how natural selection has affected human genome variability has proven to be a powerful tool to delineate genes and biological functions having played a key role in human adaptation, a variation which can also be involved in phenotypes of medical relevance. This article reviews several examples that illustrate well how different environmental pressures, particularly those imposed by pathogens and infectious diseases, have shaped the patterns of genetic and epigenetic variability currently observed in human populations.
© 2016 Académie des sciences. Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

R É S U M É

L'histoire évolutive de l'Homme signifie bien plus que son histoire démographique. Elle inclut également son adaptation génétique aux divers environnements qu'il a rencontrés — nutritionnels, climatiques ou pathogéniques — ainsi que les différentes réponses épigénétiques mises en place pour y faire face. Détecter la façon dont la sélection naturelle a influencé la variabilité du génome humain représente un outil puissant pour identifier des fonctions biologiques ayant joué un rôle majeur dans l'adaptation et la survie de notre espèce et qui peuvent également être associées à des phénotypes variables d'intérêt médical. Cet article présente différents exemples qui illustrent la façon dont les pressions environne-mentales, tout en particulier celles exercées par les agents pathogènes et les maladies infectieuses, ont influencé la diversité génétique et épigénétique des populations humaines.
© 2016 Académie des sciences. Publié par Elsevier Masson SAS. Cet article est publié en Open Access sous licence CC BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The wide range of phenotypic variation observed in human populations may reflect distinctive processes of

genetic adaptation to variable environmental conditions. Over the past decade, the advent of genome-wide single-nucleotide polymorphisms (SNPs) and whole-genome sequence datasets has enabled one to test different hypotheses concerning how natural selection, in its different forms and intensities, has influenced the variability of the human genome. Genome-wide scans for selection have identified numerous candidate genes under selection, increasing knowledge of the adaptive history of humans and providing new tools for delineating genomic regions associated with phenotype variation, both benign and disease-related [1–4].

In addition to genome-wide approaches, studies of candidate genes have also provided evidence for the action of selection, particularly when functional evidence is available, with an increasing number of selected genes being documented in relation to phenotypes associated with adaptation to nutritional resources, different climates or pathogen presence [1,5–7]. For example, iconic cases of genetic adaptation to diet have been well described for milk consumption, starch-rich diets or bitter-taste perception. Likewise, genetic adaptation to changing environments is provided by the exposure of ancestral populations to colder climates and lower levels of sunlight after early migrations out of Africa. These changes led to variation in the quantity, type and distribution of melanin in the skin, resulting in the various levels of skin pigmentation observed in present-day human populations. Another interesting case of selection is adaptation to high altitude, for which different mutations in different genes have been reported as evolving adaptively to avoid hypoxia.

2. Forms of natural selection

Natural selection can manifest in different forms (Fig. 1A), each of them leaving distinctive molecular signatures in the targeted genomic region (reviewed in [6]). *Purifying selection*, or *negative selection*, refers to the process by which deleterious mutations are culled from the population, and is the most pervasive form of selection. At the population level, the reduced number of non-synonymous SNPs observed, as compared with the non-synonymous mutation rate, reflects the elimination of many non-synonymous mutations through purifying selection. Selection also occurs when a novel mutation is favorable, as is referred to as *positive selection*, which is thought to be one of the ways in which adaptive evolution occurs. Most approaches to detect positive selection rely on the fact that a beneficial allele will increase to a high frequency within the population at a rate that is much faster than that of a neutrally-evolving allele. Finally, *balancing selection* refers to a selective regime in which two, or multiple alleles, at a given locus are maintained in the population, leading to an overall increase in genetic diversity. Balancing selection can maintain polymorphism through heterozygote advantage, in which individuals who are heterozygous at a particular locus have a greater fitness than homozygous individuals (e.g., HbS [sickle-cell] variant), or frequency-dependent selection, where the fitness of a phenotype is dependent on its frequency relative to other phenotypes in a given population. In

humans, it appears that positive selection is more pervasive than balancing selection, although the latter regime has been particularly documented in genes involved in immune functions.

3. Approaches for detecting the effects of selection

Each type of selection leaves a distinctive molecular signature (e.g., nucleotide diversity, allele frequency spectrum, haplotype length, etc.) in the genome concerned (Fig. 1B). Such molecular signatures can be detected with an increasing number of statistical tests that can be broadly subdivided into those that search for selection at the inter-species level (e.g., human vs. chimpanzees) and those that focus on particular aspects of within-species data (for a review, see [6]). The latter are used to detect selection within and between human populations, and can be further subdivided into distinct groups, each one focusing on different aspects of the genetic data. These include: (i) frequency-based methods (e.g., Tajima's D and derivatives, Fay and Wu's H tests), which determine whether the frequency spectrum of mutations conforms to the expectations of the standard neutral model; (ii) population differentiation-based methods (e.g., F_{ST} and LSBL), which test for altered levels of differentiation between populations. For example, when positive selection occurs in only a subset of populations, the frequency of the selected variant may differ across populations to a greater extent than that predicted under neutrality (increased F_{ST}) (Fig. 1C); (iii) haplotype-based methods (e.g., iHS, LDD, XP-EHH), which examine the patterns of haplotype homozygosity associated with particular alleles. For example, an allele targeted by recent positive selection would be expected to have an unusually long haplotype for its population frequency, because the advantageous allele increases in frequency too rapidly for recombination to have a major effect on haplotype length; and (iv) composite methods, which combine different, independent tests into a single composite score, increasing power and minimizing the detection of false positive signals [6].

Some of these tests are sensitive to the confounding effects that other factors, in particular demography, have on the patterns of genetic diversity. However, it is possible to overcome this caveat, as demographic events affect the whole genome, whereas selection acts locally and is restricted to particular genomic regions. Demographic models that consider realistic scenarios for the demographic history of human populations (e.g., population expansion, bottlenecks, etc.) can be incorporated into neutral expectations. Likewise, empirical procedures can be used to compare the value of a given statistic for the gene of interest (e.g., Tajmas's D, F_{ST}, etc.) with background expectations for that statistic generated from genome-wide data, which should reflect neutrality. Thus, simulation-based or empirical procedures can be used to distinguish between the effects of demographic factors and those of natural selection events targeting specific genomic regions, providing evidence of the true effects of selection in the human genome.
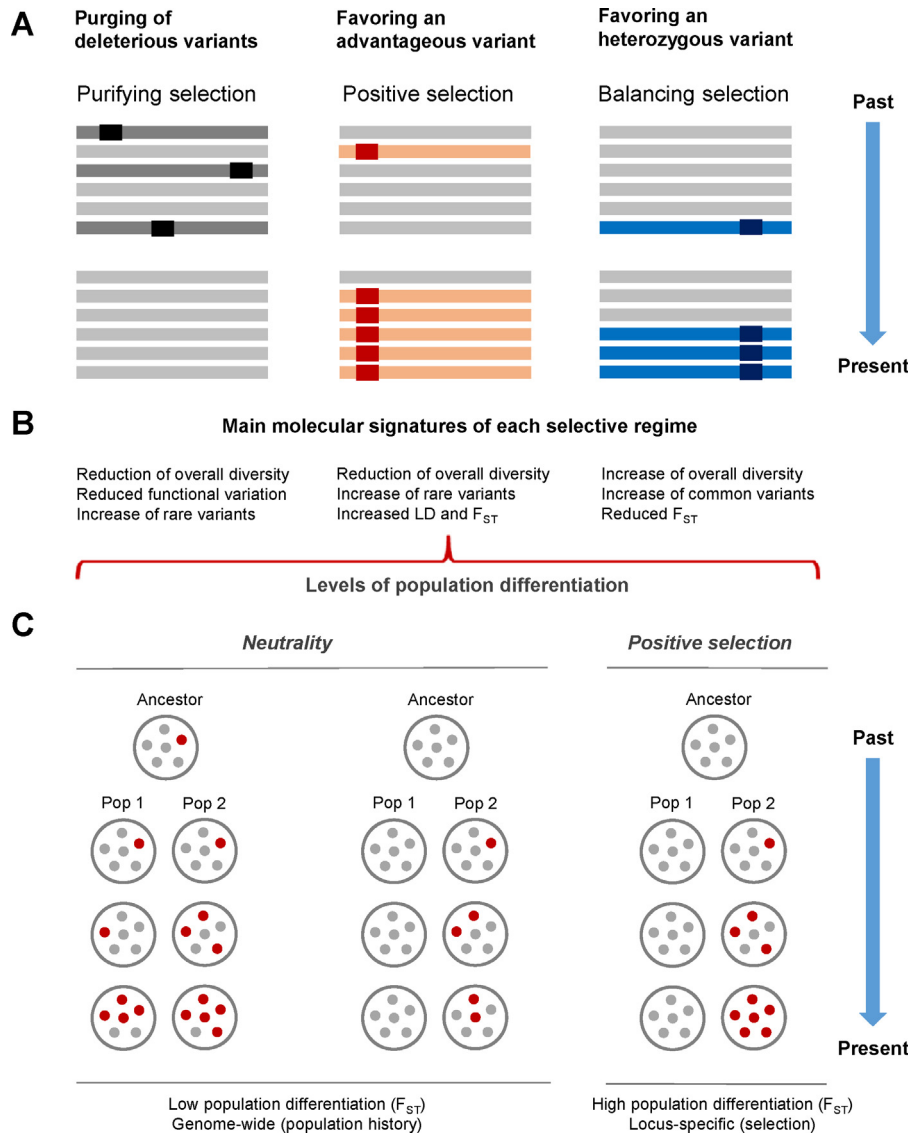
Fig. 1. **Modes of natural selection and molecular signatures of each selective regime.** A. Purifying selection leads to the removal of deleterious alleles (in black) from the population. Positive selection favors the increase of a given allele (in red) in the population. Balancing selection, for example, can favor the presence of heterozygotes (in blue) in the population. B. The main molecular signatures of each selection type are described (LD: linkage disequilibrium). C. An iconic example of a signature of positive selection, i.e., increased levels of population differentiation, is provided. Under neutrality (left), the frequency of a given mutation (in red) will fluctuate across generations by simple genetic drift, reflecting the demographic history of the population concerned. Under a scenario of population-specific positive selection (right), the advantageous mutation (in red) will increase in frequency in population 2, leading to high levels of population differentiation (high F_{ST}) for the mutation concerned.

4. Pressures imposed by pathogens and infectious diseases

Probably the most important selective pressure that has confronted humans is that imposed by infectious diseases, as pathogens have been, and still are in regions in which antibiotic treatment, vaccine administration and hygiene improvements are limited, a major cause of human mortality. Numerous studies have shown that genes involved in immunity and host defense are privileged targets of selection, increasing our understanding of how pathogens have exerted pressure on human genome variability [1–3,5,8,9]. In humans, scans for positive selection, bolstered by the advent of genome-wide datasets, have detected more than 5000 loci presenting signatures of positive selection (see [1,10] for reviews). Of these, more than 300 genes with immune-related functions have been identified, with more than half of them being detected as targets of positive selection by at least two independent studies [1]. This group of "selected genes" may display functional variation that is differentially distributed between populations and is therefore likely to be involved in the present-day differences in susceptibility to infectious, chronic

inflammatory, and autoimmune diseases, observed in human populations [4].

The most obvious selection pressure on immunity genes is the presence of pathogens, i.e., pathogen-driven selection. Proof of the importance of pathogen-driven selection comes from studies correlating genetic variability in human populations and pathogen diversity in the corresponding geographic regions, with significant correlations being detected for the Human Leukocyte Antigen (HLA) class-I genes, blood group antigens, and interleukin-related genes. Other studies have identified genetic variation in host genes that correlates with specific groups of microbes, such as viruses, protozoa, and parasitic worms. Furthermore, when testing for genetic correlations with a large variety of environmental variables, including climate, subsistence strategies, diets and pathogen load, it has been found that pathogens are still the primary drivers of local adaptation [11]. That genes under pathogen-driven selection are enriched in functions such as innate immunity and inflammatory response supports the major role played by pathogens in human evolution, particularly that of the immune response.

5. From population genetics to human immunology

The additional insight brought by studies of natural selection is that they enable the delineation of the biological relevance of immunity genes *in natura* (i.e., their degree of essentiality, redundancy or adaptability), and the prediction of their involvement in infectious or immunity-related diseases [3,12,13]. Genes evolving under purifying selection are likely to be involved in essential mechanisms of host defense, variation in which should lead to severe disorders [13]. This is supported by genome-wide studies, as Mendelian disease genes are enriched in signals of purifying selection [14]. Focusing on innate immunity, it has been recently shown that innate immunity genes have evolved under stronger evolutionary constraints than the remainder of the genome [15]. For example, microbial sensors such as endosomal Toll-like receptors (TLRs) and many Nod-like receptors (NLRs), adaptors such as MYD88 and TRIF, and effectors such as some type-I IFNs and IFN-γ have been targeted by purifying selection, attesting to the unique, essential nature of the mechanisms—immunological or otherwise—involved (reviewed in [3]).

Clinical genetic studies further support this notion, as rare mutations underlying severe diseases have been found in highly constrained genes and pathways. For example, mutations in the TLR3-TRIF, TIR-MYD88, and IFN-γ pathways have been associated with life-threatening infections during childhood, including HSV-1 encephalitis, pyogenic bacterial infections and MSMD, respectively (see [16] and references therein). Conversely, genes evolving under weak negative selection are likely to be involved in more redundant processes [1,12,16]. For example, among innate immunity receptors that sense nucleic acids, the weaker constraints characterizing the RIG-I-like receptor (RLR) family, with respect to endosomal TLRs, point to some redundancy of RLR-mediated antiviral immunity. Extreme cases of immunological redundancy

are provided by molecules such as MBL or TLR5, for which loss-of-function alleles can increase to very high population frequencies [3].

The action of positive or balancing selection, in turn, attests to more dynamic mechanisms, variations of which have been beneficial to the host over different evolutionary timescales. Selection can increase the frequency of some mutations in specific populations, as they can exert a protective, almost Mendelian, effect against infections [7]. Notable examples are provided by the HbS heterozygotes in Africa, independent G6PD deficiency variants worldwide, the DARC null allele in Africa, and the various FUT2 deficiency alleles in different populations. Positive selection can also increase the frequency of alleles associated with more complex traits or diseases, such as the TLR1 I602S hypo-responsiveness mutation in Europe [17], suggesting an advantage associated with weak TLR1-mediated responses, or variants in type III IFN genes in Eurasians [18], some of which have been associated with the clearance of HCV infection. A recent study focusing on > 1500 innate immunity genes has shown that their patterns of diversity result from different demographic and selective events, including Neanderthal introgression and hard sweeps at some loci in specific populations occurring mostly during the Neolithic transition [15].

6. Trade-offs of past selection: maladaptation

In some cases, past selection may result in maladaptation and immune dysfunction, such as inflammation and autoimmunity. The present increased incidence of chronic immunity-related disorders appears to be concomitant with the "pathogenic sterilization" of modern societies during the 20th century [19]. The hygiene hypothesis postulates that a decrease in the diversity of microbes we are exposed to has led to an imbalance in the immune response, promoting chronic inflammation [20]. Population genetics has provided support for this hypothesis, as several immunity-related genes, variants of which confer a higher risk of inflammatory bowel disease, celiac disease, type-I diabetes, multiple sclerosis, or psoriasis, have been targeted by positive selection. The higher frequency of alleles conferring greater susceptibility to some of these diseases in populations exposed to high microbial/viral loads suggests that these variants play an otherwise beneficial protective role in host defense [20]. Furthermore, risk alleles for celiac disease, in genes such as *IL12A*, *IL18RAP*, and *SH2B3*, have been targeted by positive selection and individuals carrying these alleles benefit from protection against some infections [1]. More generally, strong population differentiation has been observed for some risk alleles associated with several autoimmune conditions [21], supporting further the connection between past adaptation and current disease risk.

7. Population epigenetics: the case of DNA methylation variation

Besides genetic adaptation, humans, as well as other organisms, have alternative ways to respond to environmental pressures. In this context, epigenetic variation,

including histone modifications, RNA-based mechanisms and DNA methylation, plays a crucial role at the interface between the environment and the genome [22]. DNA methylation is perhaps the best understood component of the epigenetic machinery [23], and can be affected by inherited DNA sequence variation and environmental factors, such as nutrition, toxic pollutants and social environment. DNA methylation differences exist between major ethnic groups, highlighting the potential contribution of epigenetic modifications to phenotypic variation, including physical appearance, drug metabolism, sensory perception, and disease susceptibility [24]. These studies have also shown that DNA methylation differences between populations result from a combination of differences in allele frequencies of genetic variants associated with DNA methylation variation (methylation quantitative trait loci, meQTL) and gene–environment (G × E) interactions.

Recent work has evaluated the impact that temporal changes in habitat and lifestyles, together with genetic diversity, have on epigenetic variation [25]. By comparing the genome-wide DNA methylation profiles of rainforest hunter-gatherers and sedentary farmers from Central Africa, it appears that methylation variation associated with recent changes in habitat (urban/rural vs. forest) mostly concerns immune functions, whereas that associated with historical lifestyle (farming vs. hunting and gathering) affects primarily developmental processes. Furthermore, DNA methylation changes that correlate with historical lifestyle show strong associations with genetic variants that, moreover, are enriched in signals of natural selection. All these studies increase our understanding of the relative impacts that population genetic variation and differences in lifestyles and ecologies have on the human epigenome, and illustrates the utility of DNA methylation as a marker to track variation in regulatory activity following environmental change.

8. Concluding remarks

Population genetic studies have collectively helped to delineate functionally important loci responsible for the genetic adaptation, or epigenetic responses, of human populations to environmental pressures and lifestyle transitions. Likewise, the investigation of how natural selection, in its different forms and intensities, has targeted particular genes and biological functions has proven a useful tool to inform the relationship between genetic diversity, adaptive phenotypes and disease, providing an indispensable complement to clinical and epidemiological genetic studies. Such multidisciplinary, integrative efforts are required to clarify the relationship between natural selection and disease and to improve our understanding of the evolutionary mechanisms accounting for the present-day disparities in disease susceptibility, resistance or progression observed, both at the individual and population levels.

Disclosure of interest

The author declares that he has no competing interest.

References

[1] L.B. Barreiro, L. Quintana-Murci, From evolutionary genetics to human immunology: How selection shapes host defence genes, Nat. Rev. Genet. 11 (1) (2010) 17–30.
[2] E.K. Karlsson, D.P. Kwiatkowski, P.C. Sabeti, Natural selection and infectious disease in human populations, Nat. Rev. Genet. 15 (6) (2014) 379–393.
[3] L. Quintana-Murci, A.G. Clark, Population genetic tools for dissecting innate immunity in humans, Nat. Rev. Immunol. 13 (4) (2013) 280–293.
[4] J.F. Brinkworth, L.B. Barreiro, The contribution of natural selection to present-day susceptibility to chronic inflammatory and autoimmune disease, Curr. Opin. Immunol. 31 (2014) 66–78.
[5] S.R. Grossman, K.G. Andersen, I. Shlyakhter, S. Tabrizi, S. Winnicki, A. Yen, D.J. Park, D. Griesemer, E.K. Karlsson, S.H. Wong, M. Cabili, R.A. Adegbola, R.N. Bamezai, A.V. Hill, F.O. Vannberg, J.L. Rinn, E.S. Lander, S.F. Schaffner, P.C. Sabeti, Identifying recent adaptations in large-scale genomic data, Cell 152 (4) (2013) 703–713.
[6] J.J. Vitti, S.R. Grossman, P.C. Sabeti, Detecting natural selection in genomic data, Annu. Rev. Genet. 47 (2013) 97–120.
[7] L. Quintana-Murci, L.B. Barreiro, The role played by natural selection on mendelian traits in humans, Ann. N. Y. Acad. Sci. 1214 (2010) 1–17.
[8] L.B. Barreiro, G. Laval, H. Quach, E. Patin, L. Quintana-Murci, Natural selection has driven population differentiation in modern humans, Nat. Genet. 40 (3) (2008) 340–345.
[9] M. Fumagalli, M. Sironi, Human genome variability, natural selection and infectious diseases, Curr. Opin. Immunol. 30C (2014) 9–16.
[10] J.M. Akey, Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res. 19 (5) (2009) 711–722.
[11] M. Fumagalli, M. Sironi, U. Pozzoli, A. Ferrer-Admetlla, L. Pattini, R. Nielsen, Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution, PLoS Genet. 7 (11) (2011) e1002355.
[12] L. Quintana-Murci, A. Alcais, L. Abel, J.L. Casanova, Immunology in natura: Clinical, epidemiological and evolutionary genetics of infectious diseases, Nat. Immunol. 8 (11) (2007) 1165–1171.
[13] A. Alcais, L. Quintana-Murci, D.S. Thaler, E. Schurr, L. Abel, J.L. Casanova, Life-threatening infectious diseases of childhood: Single-gene inborn errors of immunity? Ann. N. Y. Acad. Sci. 1214 (2010) 18–33.
[14] R. Blekhman, O. Man, L. Herrmann, A.R. Boyko, A. Indap, C. Kosiol, C.D. Bustamante, K.M. Teshima, M. Przeworski, Natural selection on genes that underlie human disease susceptibility, Curr. Biol. 18 (12) (2008) 883–889.
[15] M. Deschamps, G. Laval, M. Fagny, Y. Itan, L. Abel, J.L. Casanova, E. Patin, L. Quintana-Murci, Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes, Am. J. Hum. Genet. 98 (1) (2016) 5–21.
[16] J.L. Casanova, L. Abel, L. Quintana-Murci, Immunology taught by human genetics, Cold Spring Harb. Symp. Quant. Biol. 78 (2013) 157–172.
[17] L.B. Barreiro, M. Ben-Ali, H. Quach, G. Laval, E. Patin, J.K. Pickrell, C. Bouchier, M. Tichit, O. Neyrolles, B. Gicquel, J.R. Kidd, K.K. Kidd, A. Alcais, J. Ragimbeau, S. Pellegrini, L. Abel, J.L. Casanova, L. Quintana-Murci, Evolutionary dynamics of human toll-like receptors and their different contributions to host defense, PLoS Genet. 5 (7) (2009) e1000562.
[18] J. Manry, G. Laval, E. Patin, S. Fornarino, Y. Itan, M. Fumagalli, M. Sironi, M. Tichit, C. Bouchier, J.L. Casanova, L.B. Barreiro, L. Quintana-Murci, Evolutionary genetic dissection of human interferons, J. Exp. Med. 208 (13) (2011) 2747–2759.
[19] J.F. Bach, The effect of infections on susceptibility to autoimmune and allergic diseases, N. Engl. J. Med. 347 (12) (2002) 911–920.
[20] M. Sironi, M. Clerici, The hygiene hypothesis: An evolutionary perspective, Microbes Infect. 12 (6) (2010) 421–427.
[21] E. Corona, R. Chen, M. Sikora, A.A. Morgan, C.J. Patel, A. Ramesh, C.D. Bustamante, A.J. Butte, Analysis of the genetic basis of disease in the

context of worldwide human relationships and migration, PLoS Genet. 9 (5) (2013) e1003447.

[22] R. Jaenisch, A. Bird, Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals, Nat. Genet. 33 (Suppl.) (2003) 245–254.

[23] J.A. Law, S.E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals, Nat. Rev. Genet. 11 (3) (2010) 204–220.

[24] H. Heyn, M. Esteller, DNA methylation profiling in the clinic: Applications and challenges, Nat. Rev. Genet. 13 (10) (2012) 679–692.

[25] M. Fagny, E. Patin, J.L. MacIsaac, M. Rotival, T. Flutre, M.J. Jones, K.J. Siddle, H. Quach, C. Harmant, L.M. McEwen, A. Froment, E. Heyer, A. Gessain, E. Betsem, P. Mouguiama-Daouda, J.M. Hombert, G.H. Perry, L.B. Barreiro, M.S. Kobor, L. Quintana-Murci, The epigenomic landscape of african rainforest hunter-gatherers and farmers, Nat. Commun. 6 (2015) 10047.