Genetic drift in mammals

JEAN-FRANÇOIS BACH

Académie des Sciences, 23 Quai de Conti, 75006 Paris, France

How to cite: BACH JF. 2019. Genetic drift in mammals. An Acad Bras Cienc 91: e20190339. DOI 10.1590/0001-3765201920190339.

Abstract: Genetic drift is the fortuitous occurrence of genetic events that when they become fixed modify the genome of populations. They can take the form of mutations of single nucleotides (SNPs), the insertion or deletion of short sequences (Indels) or the repetitions of short sequences (CNV i.e. copy number variants) or long insertions or deletion (structural modifications). Their frequency is 10^{-9} to 10^{-8} depending on the species, or 50 to 100 per birth in humans. The incidence of these de novo mutations is higher when the father is old at conception. It thus appears that genetic drift, which constitutes the initial element of evolution, has a very strong dynamics. Its intervention in the appearance or disappearance of some major phenotypes is complicated by the uncertainties about the genetic mechanisms in heritability which, paradoxically, are only partially understood.

Key words: genetic drift, genome of populations, *de novo* mutations, inbred mouse strains.

INTRODUCTION

The time line of evolution is long. It is counted in million or hundreds of million years depending on the context envisaged. Until recently, knowledge about genetic drift occurring in a short time frame was limited by the difficulty of access to genome sequence. Modern genome sequencing techniques have completely changed the situation. It is in the light of the data obtained by these techniques that we will try to take a new look at the dynamics of evolution in mammals.

Genetic drift is an unavoidable phenomenon leading to changes in the genome of populations, whether these changes are fixed or disappear. Mutations that occur unexpectedly are neutral, not subject to a selection force. They can take different genetic aspects: mutations of single nucleotides (SNPs), the insertion or deletion of short sequences (Indels) or the repetitions of short sequences (CNV i.e. copy number variants) or long insertions or deletion (structural modifications).

INBRED MOUSE STRAINS

Before discussing the recent observations that have been made in humans, we will evoke remarkable facts concerning laboratory mouse strains. Pure strains of mice have been produced in various laboratories since the early 20th century. These pure mouse strains can be easily obtained by systematically breeding siblings of the same litter over at least 20 generations. At the 20th generation, it is admitted that mice are inbred and genetically identical, although not totally identical because

Correspondence to: Jean-François Bach
E-mail: jean-francois.bach@academie-sciences.fr
ORCid: https://orcid.org/0000-0002-1282-6067

there persists among individuals up to 1.6% polymorphisms (notably SNPs and Indels). Mice are for the most part homozygous at all locis and all identical. The C57BL/6 strain was introduced to the Jackson Laboratory in 1948. Shortly thereafter, in 1951, some of these mice were raised in parallel at the National Institute of Health (NIH) with no thought that the two mouse colonies would genetically evolve separately. The problem became apparent in 2013 when the genome sequences of mice raised in the Jackson laboratory (C57BL/6J) and those raised at the NIH (C57BL/6N) were compared. More than 300 differences in the genome were observed including SNPs, indels and some structural modifications (Simon et al. 2013). At the same time, phenotype differences were noted, particularly concerning photoreceptors and certain behaviors. In particular, it was observed that the two mouse colonies behaved differently to cocaine: whereas in C57BL/6J mice cocaine significantly accelerated the running speed of mice placed on a treadmill, this acceleration was significantly diminished in C57BL/6N mice (Kumar et al. 2013). From this phenotypic difference, it was possible by a quantitative trait loci (QTL) analysis to locate the gene responsible for this difference. It was a gene coding for a cytoplasmic protein involved in cocaine receptor signaling (Kumar et al. 2013). Unraveling this difference between these two strains of C57BL/6 mice was all the more remarkable as, as mentioned above, for years no attention was paid by investigators to the use of the one or the other sub-strain. It is interesting to note that the International Knockout Mouse Consortium (IKMC) had used the C57BL/6N mouse but that many users mixed them with C57BL/6J mice in their experiments, which makes the retrospective analysis of certain results obtained in these genetically invalidated mice highly questionable.

DE NOVO MUTATIONS IN MICE AND MAN

The occurrence of de novo mutations is, in fact, well known in both mice and humans. An example in mice is the mutation that leads to a decrease in resistance to infections and more particularly to the effect of lipopolysaccharide. This mutation observed in C3H mice was found to correspond to a gene of major importance that was part of the Toll like receptors (TLR) genes. More specifically, it was the TLR4 encoding gene (Poltorak et al. 1998). The same is true in humans, where accidental mutations have been at the origin of most monogenic diseases. In many cases, it was possible to trace the founder. Thus, for severe juvenile glaucoma, it was possible in a family in northern France and among Greek immigrant populations in different countries of the world to identify the existence of a founder, in the 17th century and in the Middle Age respectively (Brézin et al. 1998, Hewitt et al. 2007). It is indeed possible to find in subjects suffering from this disease the precise mutation of the founder, possibly associated with haplotype sequences due to linkage disequilibrium. Other observations of the same type have been reported knowing, however, that in many cases, including cystic fibrosis, numerous distinct mutations of the gene involved were reported which suggests the existence of several founders.

The question arises as to how often these *de novo* mutations occur. This question has been the subject of numerous studies mainly in drosophila and also humans in more recent years (Keightley et al. 2014, Campbell and Eichler 2013, Smeds and Qvarnström 2016). The approach is simple: it consists in taking trios i.e. the father, the mother and one of their direct descendants to compare the complete genome sequence of the three individuals. In drosophila, very often, rather than taking a single trio, studies address the cumulative effect of genetic drift over several generations. The results obtained are remarkable. If one evaluates

the frequency of de novo mutations per nucleotide, one finds values in the order of 10^{-9}, slightly less in humans (about 5×10^{-8}). When one looks at the few articles concerning other species, one finds similar figures with nevertheless a higher frequency of *de novo* mutations in *Caenorhabditis elegans* and in *Saccharomyces cerevisiae* (Smeds and Qvarnström 2016).

Remarkably, studies conducted in humans have shown that the frequency of *de novo* mutations, which is 50 to 100 per individual, is all the greater the older the father is at the time of conception (Goldmann et al. 2016). There is in fact an almost linear relationship between the frequency of *de novo* mutations and the age of the father. There is also some effect of the mother's age but much less. Analysis of *de novo* mutations observed in humans showed that they most often include SNPs or Indels but also mobile genes, CNV and, in some cases, major structural modifications, even aneuploidy (Campbell and Eichler 2013).

An even more singular observation was made. Studying the entire genome of two monozygotic twin pairs, while one would expect to find an absolute identity of the sequences, a small number of mutations present in one twin but absent in the other were observed (Dal et al. 2014). This phenomenom is defined as a post-zygotic mutation. Its existence could have been suspected for many years because geneticists are well aware that certain genetic diseases, some of which are well known such as cystic fibrosis, Duchenne myopathy or hemophilia A, can be found in a single monozygotic twin.

It is therefore clear that the genome is constantly and highly evolving. Depending on the mixing of the populations, mutations that appear in this way can become fixed or, on the contrary, disappear and may or may not give rise to a visible phenotype.

How to place the role of these *de novo* mutations in the genetic control of physiological traits or even diseases? The example of complex polygenic diseases is particularly interesting because it has given rise to numerous studies and in-depth analyses. If we take the example of autoimmune insulin-dependent diabetes, heredity is strong. When a subject is affected by the disease, his children have a 5% risk of being affected by the disease. When a subject has autoimmune diabetes in a family, siblings have a 7% risk of developing the disease and even nearly 15% if the probant and the sibling are identical at some risk haplotypes in the HLA complex. Faced with such a heredity one should expect to easily find the predisposition genes with modern genetic techniques. If we exclude HLA, whose role is very specific to autoimmune diabetes, the very numerous studies on a very large number of subjects that have been conducted over the past 15 years using GWAS (Genome Wide Associated Studies) have allowed the identification of a very large number of chromosomal regions associated with diabetes. Similarly, in other complex diseases the situation became complicated when it was realized that the number of potential susceptibility regions were considerable for a given disease, more than 100 or even more than 200 (Fortune et al. 2015). In addition, the relative risk associated with each of these predisposing chromosomal regions is extremely low, most often <1.3 or 1.5. Finally, with GWAS regions the single genes involved have not been identified. It is presently considered that the contribution to complex diseases heritability of GWAS findings does not exceed 15 or 20%, thus opening the discussion on missing heritability (Manolio et al. 2009, Nolte et al. 2017). Similar observations have been made for physiological traits, particularly height (Eichler et al. 2010). It therefore appears necessary to define other approaches to identify the genetic factors responsible for heredity, more generally for heritability. There are many leads but they have not really come to fruition yet. Among the possibilities: are: the existence of rare variants not recognized by GWAS, variants with low penetrance,

variants resulting from recent *de novo* mutations segregating in families, structural modifications, epistasis (i.e. interaction between several genes), epigenetics. One should also highlight the possible disease heterogeneity from one family to another, which to a large extent invalidates studies carried out on very large cohorts, independently of family groupings.

Last but not least environment may also modify the expression of certain genes. Some environmental factors can have a positive stimulating effect on phenotypes; this may be the case of some viruses in autoimmunity. In other cases, the environment may be protective; this is the basis of hygiene hypothesis in autoimmune and allergic diseases (Bach 2018). A role for the intestinal microbiota, the composition of which is known to correlate with certain chronic diseases cannot be excluded (Bach 2018).

To conclude on the dynamics of evolution, it is quite obvious that the means and markers studied to follow this evolution are far from simple and thus explain the uncertainties that are presented here. The problem remains at the population level to know how many years it takes to a *de novo* mutation that appear after genetic drift to become fixed. These are the challenges of the coming years, where genomic analysis of species and subspecies will hopefully answer these questions.

REFERENCES

BACH J-F. 2018. The hygiene hypothesis in autoimmunity: the role of pathogens and commensals. Nat Rev Immunol 18: 105-120.

BRÉZIN AP ET AL. 1998. Founder effect in GLC1A-linked familial open-angle glaucoma in Northern France. Am J Med Genet 76(5): 438-445.

CAMPBELL CD AND EICHLER EE. 2013. Properties and rates of germline mutations in humans. Trends Genet 29: 575-584.

DAL GM, ERGÜNER B, SAĞIROĞLU MS, YÜKSEL B, ONAT OE, ALKAN C AND ÖZÇELIK T. 2014. Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. J Med Genet 51: 455-459.

EICHLER EE, FLINT J, GIBSON G, KONG A, LEAL SM, MOORE JH AND NADEAU JH. 2010. Missing heritability and strategies for finding the underlying causes of complex diseases. Nat Rev Genet 11: 446-450.

FORTUNE MD ET AL. 2015. Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. Nat Genet 47: 839-846.

GOLDMANN JM ET AL. 2016. Parent-of-origin-specific signatures of de novo mutations. Nat Genet 48: 935-939.

HEWITT AW ET AL. 2007. Investigation of founder effects for the Thr377Met Myocilin mutation in glaucoma families from differing ethnic backgrounds. Mol Vis 13: 487-492.

KEIGHTLEY PD, NESS RW, HALLIGAN DL AND HADDRILL PR. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. Genetics 196: 313-320.

KUMAR V ET AL. 2013. C57BL/6N mutation in cytoplasmic FMRP interacting protein 2 regulates cocaine response. Science 342: 1508-1512.

MANOLIO TA ET AL. 2009. Finding the missing heritability of complex diseases. Nature 461: 747-753.

NOLTE IM ET AL. 2017. Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. Eur J Hum Genet 25: 877-885.

POLTORAK A ET AL. 1998. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science 282: 2085-2088.

SIMON MM ET AL. 2013. A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. Genome Biol 14: R82.

SMEDS L AND QVARNSTRÖM A. 2016. Direct estimate of the rate of germline mutation in a bird. Genome Res 26: 1211-1218.