Genetic polymorphism and selection

Auteur :
VEUILLE Michel, Directeur d'Etudes, Ecole pratique des hautes études, Paris

23-03-2019

Polymorphism has caused controversy about its role in evolution. But if it essentially follows a neutral evolution, it serves as a reference, in contrast, for the study of natural selection. It is also used by ecologists in conservation biology to reconstruct the past history of species.

1. Mutations, random drift and neutral evolution

Polymorphism consists of mutations that escape DNA repair systems over cell divisions. Their rate of appearance is therefore a biological variable. In humans and chimpanzees, it is $\mu \approx 10^{-8}$ mutations by nucleotide\textsuperscript{basic element of a nucleic acid such as DNA or RNA. It is composed of a nucleic base (or nitrogenous base), a ose with five carbon atoms, called pentose, whose association forms a nucleoside, and finally one to three phosphate groups.} and by generation. The considerable amount of sperm produced by male mammals means that there is much more cell division in male germ line\textsuperscript{all cells from stem cells to gametes} than in female germ line: 380 against 23 at age 30 (i.e. 16 times more), and even more so when men age (840 against 23 at age 50, i.e. 36 times more). This means that in these species the mutations are mainly produced in male lines and depend on the age of the father. Each birth produces about 100 new mutations per genome\textsuperscript{genetic material of a living organism. It contains genetic information encoding proteins. In most organisms, the genome corresponds to DNA. However, in some viruses called retroviruses (e.g. HIV), the genetic material is RNA.} but because only a small part of the genome is coding\textsuperscript{describes the part of the DNA or RNA of a gene translated into protein. Represents only a part of the gene from which it originates, as well as the mRNA in which it is registered.} 99% of them have no effect on survival or fertility. They are called neutral. A new allel\textsuperscript{two homologous genes are called alleles when they have different shapes, distinguishable at a given level of observation. An allele can therefore correspond to a single sequence, or to a set of sequences that are different but not distinguishable at the phenotype level. (e.g. blue/brown/green eye colour but at the nucleotide level there are many more different alleles, several per colour).} can be neutral, harmful or advantageous. Neutral mutations are the most studied, as they allow predictive models to be written to explore population history. Their distribution also serves as a null hypothesis\textsuperscript{refers to the basic point of view, to the default position regarding a given phenomenon. In general, hypotheses opposing the null hypothesis are called alternative hypotheses.}
Nous pouvons penser que dans un **gène comprenant uniquement des allèles neutres**, le décalage des fréquences alleliques pourrait compenser pour un flou dû à une variation sur le côté et que l’haplotypique diversité serait stable à long terme. Mais cette impression est fausse. Graduellement, **la diversité est en train de disparaître**. Ce phénomène est très similaire à la perte de diversité des noms de famille, un phénomène qui est lent mais significatif en population humaine comme les villages à distance. Quand une famille ne possède pas un fils, cela n’influe pas sur son nom de famille. Mais le nom de famille devient plus rare avec le temps. Ceci est dû non pas à un phénomène biologique propre au chromosome Y, qui accompagne les naissances masculines. Chance est d’autant plus importante pour expliquer. Ce phénomène reflète la réalité du fait que la constitution d’une génération futale d’un parent est soumise au principe d’un échantillonnage avec remise [in-text] et le nombre de gènes est donné par une loi de Poisson de paramètre 1 à savoir que ces gènes ne sont pas mis échantillons dans un tas contenant un nombre d’objets n, donc le premier gène, puis un deuxième gène, puis un troisième gène, etc., jusqu’au p-ième gène. Cela signifie choisir p objets parmi n avec remise (vous pouvez choisir le même objet plusieurs fois). Le résultat de ce processus est que le nombre d’ancêtres équivalents est égal à n × n × n × ... × n = np. [end-tooltip]

Comme les chromosomes Y, certains gènes de la génération parentale ne sont pas dermés, et ne sont pas trouvés dans la population des dérivés. Si ce n’était pas le cas, l’ancêtre de la population serait différent sans consanguinité. En conséquence, il existe deux conséquences à ce: premier, la polymorphisme d’une espèce est toujours "récent" sur l’échelle de la durée de l’espèce, car il dépend sur les mutations qui ont restauré le polymorphisme malgré l’érosion de la diversité qui accompagne le décalage des fréquences alleliques. Second, le niveau de polymorphisme est un compromis entre deux mécanismes opposants, créant le **neutral mutation-drift balance**.

La disparition de polymorphisme sur le temps peut être mesurée de la manière opposée à la direction: quand nous retournons dans l’histoire, il existe toujours un ancêtre commun entre deux gènes de la même locus [end-tool-tip]. Position du gène sur le chromosome. En population génétique, tous les gènes homologues (classe de homologie). Deux chromosomes ou deux gènes sont homologues s'ils se correspondent au niveau de l'haplotypique. Ce qui est appelé par John Kingman le processus de coalescence. L’ancêtre n’est pas le même pour différents locus, car la sexualité multiplie le nombre d’ancêtres, donc aussi les gènes les plus communs des ancêtres. Si la probabilité d’avoir un ancêtre commun à la génération précédente  \( q = 1/N_e \), reste constante au fil du temps, la distribution d’ancêtres suit une loi exponentielle  \( t = q e^{-qt} \). L’âge attendu de ces ancêtres est égal à  \( N_e \). Deux gènes seront génétiquement similaires si aucune mutation n’a eu lieu depuis. Mais il est important que la mutation a eu lieu dans l’un des traits alléliques émergés depuis les ancêtres. Cela peut être déduit que le nombre de bases nucléotidiques se partager entre ces deux gènes est  \( \hat{\theta} = N_e x 2\mu \), où \( \mu \) est la **taux de mutation neutre**. Cette valeur, définie comme  \( \hat{\theta} = 2N_e\mu \), est un paramètre fondamental de la population génétique.
The neutral evolution of natural populations is very important in conservation biology, as it allows the history of species to be reconstructed. Geneticists have long known that random genetic drift allows them to infer models of population differentiation and species structure in space (Figure 1). During the second half of the 20th century, the most commonly used indicator to study the structuring of a population into sub-populations was the $F_{ST}$ of the formula:

$$F_{ST} = 1 - \frac{H_S}{H_T}$$ (5)

where $H_S$ is the average of the diversities of the sub-populations and $H_T$ is the diversity of the total population [2].

In the 21st century, the age of numerical genome analysis, the theory of coalescence [3], independently developed by Kingman, Hudson and Tajima in 1982-83, makes it possible, in addition to studying structuring, to determine whether populations have remained stable or have undergone demographic changes (Figure 2).

2. Neutral model and biodiversity management
Figures 1 and 2 illustrate how genetic variation profiles are affected by population history: spatial structuring, colonization, migration, population change are all events that impart a specific signature in the molecular polymorphism of species, and allow ecologists to work in ecology. The job of an ecologist is to study the relationships between organisms and the surrounding world. Should not be confused with the ecologist, who campaigns to protect ecology, to trace its history. During the Quaternary era - the current geological period - the world’s climates changed cyclically, resulting in periodic changes in the coastline, a north-south shift of biological associations and glaciers, and periods of wet or dry climate at all latitudes. The resulting movements, decreases, increases, invasions of populations, indices of species’ responses to changes in their environment, are systematically recorded by population biologists before any natural population management initiative is undertaken. Most of the applications of population genetics today are in conservation biology.

3. Harmful mutations

Because genes code for proteins, most mutations in coding regions modify the protein sequence (about 3/4 of the mutations, a proportion that varies according to the composition of the sequence). In the human lineage, about 40% of these changes are deleterious, i.e. they are missing when the evolution of the genome of this species is assessed since its separation from the chimpanzee lineage. If a mutation were neutral, it would have a 1/N_e chance of replacing the other genes present at this locus one day (in a population of effective size N_e, the other genes taken together are in a 1-1/N_e proportion, and each also has a 1/N_e chance of replacing all the others). But a mutation can be harmful and affect the health or fertility of the individuals who carry it. Its frequency may fluctuate for a few generations by random drift before disappearing by selection (forty generations on average in Drosophila). All members of a species are carriers of deleterious mutations. You and I are. They are almost always in the heterozygous state. This characterizes an organism that has two different alleles of the same gene at the same locus for each of its homologous chromosomes. Because if a mutation has a frequency, for example, of 1/1000, it will have a thousand times fewer representatives in the homozygote state, characterized by an organism that has two identical alleles of this gene at the same locus for each of its homologous chromosomes than in the heterozygote state. It is the slight disadvantage of heterozygotes that eliminates the mutation rather than the often much greater disadvantage of the homozygous. Since the effects of deleterious mutations on several locus are cumulative, the mutation burden becomes a quantitative variable like any other whose additive effects may be undetectable, but nevertheless effective over the long term to purge the genome permanently. This explains why proteins remain functional and harmful mutations remain of low frequency. They are probably one of the factors that explain the maintenance of genetic recombination. This makes it possible both to group harmful mutations together to eliminate them and to limit the consequences of their elimination on adjacent regions of the chromosomes.

4. Advantageous mutations

What are the 60% of mutations affecting proteins without deleterious effect? Like mutations affecting other regions of the genome, they can be "neutral", i.e. without any effect on health or fertility in a particular environment and in a particular communication system of a species. Their frequency fluctuates randomly in natural populations. But if conditions change, they can be advantageous. They are then part of the natural selection and sexual selection imagined by Darwin, but also of the selection in the first sense of the word, i.e. the selection made by man on his domestic species. There are two types of polymorphism selected: transient polymorphism and balanced polymorphism.

Figure 3. Selective scanning. Recombination makes it possible to decouple the evolution of adjacent regions of the genome. If there was no selection, neutral polymorphism would reach comparable values along the chromosome through a very slow equilibrium process. When a mutation is advantageous in a region and sets at the frequency of 1, the process is very fast, and it sweeps away the neutral polymorphism of that region, but not that of adjacent regions. The contrast of the level of neutral polymorphism in the scanned regions and in the neutral regions makes it possible to affirm that it is indeed the selection that has acted in the former, and excludes the circular reasoning that would admit that “what is adapted is what we see”. This example shows two contiguous areas of selective scanning on the X chromosome of Drosophila simulans. They make it possible to identify two complexes of genes that act simultaneously to modify for their benefit the Mendelian proportions in the offspring of fruit flies (see ref [6]) (so-called "selfish" genes). In these two zones (SR1 and SR2), selection has eliminated neutral polymorphism.

Transient polymorphism is the case of an advantageous mutation that gradually "fixes" itself by eliminating alternative alleles,
which can lead to a frequency of 1. This is the case, for example, of insecticide resistance genes in mosquitoes, antibiotic resistance in bacteria and antimalarial drug resistance in the malaria parasite: these mutations would probably not have had an advantage under natural conditions, but in the environmental context imposed by medicine, these alleles increase in frequency. This is also the case for the three alleles that regulate the expression of lactase, an enzyme that allows humans to digest milk sugar (lactose) not only in the newborn state, as in other mammals, but also in adults. These mutations have become beneficial in livestock populations, while our hunter-gatherer ancestors only had the opportunity to digest fruit sugar (sucrose) as adults. In all these cases of transient polymorphism, the locus to which the selection relates is "betrayed" by a signature in the genome: the rapid expansion of its frequency makes the adjacent neutral variation on the chromosome disappear. This is a case of selective scanning, which makes it possible to affirm that the fixation of an allele is not due to random drift, but to selection (Figure 3, [4]).

![Figure 4. Balanced polymorphism](image)

**Balanced polymorphism** refers to situations where two alleles coexist because each is favoured under certain conditions, but where neither can prevail over the other in all circumstances of time or space. An example is given by cases where the selective advantage of a genotype increases due to its inverse frequency. This is called frequency-dependent selection. Such situations of balanced polymorphism are frequent in cases of sexual selection (Figure 4, [5]).

**5. Is polymorphism useful?**

In the 1930s to 1960s, natural population geneticists discovered an increasing number of polymorphisms in nature. They wanted to assess its extent and discover its potential utility in terms of evolution. Debates opposed researchers who considered that genetic diversity conferred an advantage in itself and that selection maintained it at high levels, to researchers who considered that selection led to a phenotype {ind-text} All the observable characteristics of a fairly homogeneous wild individual, the remaining variations being rather harmful. None of them were right. The Frenchman Gustave Malécot had already demonstrated in the 1950s that neutral polymorphism was a consequence of Mendel's laws {ind-text} Lois concerning the principles of biological heredity, set out by the Czech monk and botanist Gregor Mendel (1822-1884). {end-tooltip} in a finite size population [6]. It was finally the discovery in 1966 of extremely high levels of molecular polymorphism, which could not be explained by natural selection alone [7], that allowed the Japanese Kimura and Ohta to put forward the neutralist theory [8]. It was realized that the alternative to Darwin's theory of natural selection was not the fixity of species (as thought by Darwin's opponents, for example) but a continuous genetic change predicted by the neutral model, similar to the random walk of a diffusion phenomenon in physics. This vision was definitively accepted in the 1980s. However, the very low value of the effective population size measured in all species, compared to the reproductive population size, indicates that forces are eroding genetic diversity much more than neutralist models predict. This erosion is due in part, still poorly estimated, to natural selection, which eliminates harmful mutations and fixes advantageous variations, and thus increases the effects of drift on the neutral variation. Although extremely important for the future of the species, the selected polymorphisms certainly represent only a small fraction of the cases of polymorphism.

Neutral molecular polymorphism provides the basic theory, the reference model, from which the selection and history of populations are studied. The paradox is that, from now on, the molecular signatures of natural selection are sought in the genome using neutralist theory.

The existence of selective forces that maintain the recombination system, Mendel's laws [9], and the genetic mixing of sexuality is an argument for considering that polymorphism, which they maintain in this way, has a short-term advantage in natural populations.
References and notes

[1] Until about 2000, the effective size was expressed in individuals and not in chromosomes, so the effective size of the chromosomes was $2N_e$ for autosomes, $1.5N_e$ for $X$ chromosomes, and $0.5N_e$ for $Y$ chromosomes and mitochondria, provided that the number of males and females at breeding is the same. These formulas can be found in manuals.

[2] This formula, here very general, takes several forms and denominations according to the genetic model used: Wright's $F_{ST}$ (for two alleles), Nei's $G_{ST}$ (its generalization, of which formula 5 above is a variant), $\Phi_{ST,QST}$, etc. It can be replaced by statistics with similar properties: $D_{XY}$, AMOVA. This redundancy shows above all the success of "$F$-statistics" in ecology. Because of the dependence of the estimate on the sample size, the use of unbiased estimators must also take into account the particularities of the observation design. Cf. Weir B.S. & Cockerham C.C. (1984) Estimating $F$-statistics for the analysis of population structure. Evolution 38:1358-1370


L'Encyclopédie de l'environnement est publiée par l'Université Grenoble Alpes - www.univ-grenoble-alpes.fr


Les articles de l'Encyclopédie de l'environnement sont mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.