Perspective

Darwinian selection within an individual or somatic selection: facts and models

There are known cases where a cell subpopulation bearing a selective advantage tends to or eventually replaces another cell subpopulation within an organism. This process is supposed to underlie the purging of the most frequent mitochondrial DNA pathogenic variant from leukocytes and other tissues. It is also likely to explain the normal increase of sex chromosome aneuploidy in human leukocytes with age or under pathological conditions and the expansion of a cell subpopulation after the spontaneous in vivo reversion to normal of an inherited pathogenic variant. Furthermore, it is the rationale behind gene therapy for certain diseases. This (total or partial) population replacement process can be described by a logistic function. Here, I also explore the commonalities between somatic selection and the classical case of directional selection in population genetics.

When one cell subpopulation replaces another one within an individual

Alterations of the mitochondrial genome (mtDNA) can lead to mitochondrial diseases. Each eukaryotic cell has hundreds to thousands of mitochondria, each of which contains several copies of mtDNA circles (Satoh and Kuroiwa, 1991). Thus, an individual with a variant in mtDNA can have a mixture of mitochondrial genomes, a phenomenon called heteroplasmy. Moreover, the stochastic segregation of mitochondria in the oocytes of a mother carrying a mtDNA variant may lead to germ cells more or less enriched in variant mitochondria. This translates into a variability of heteroplamsy levels in her offspring (de Laat et al., 2013).

The mtDNA variant m.3243A>G is one of the most frequent causes of mitochondrial disease and the most prevalent pathogenic alteration of the mitochondrial genome. It is responsible for a variable phenotype ranging from no symptoms to myopathy, encephalopathy, lactic acidosis and stroke-like episodes, also known as MELAS (Goto et al., 1990; van den Ouweland et al., 1992; Laloi-Michelin et al., 2009). The degree of heteroplasmy correlates with the severity of the symptoms (Flierl et al., 1997; Laloi-Michelin et al., 2009). Thus, patients with a higher proportion of mutated mitochondrial genome are expected to have more severe manifestations. Interestingly, an increasing number of studies have shown that the proportion of mutated mitochondria in leukocytes declines with age ('t Hart et al., 1996; Rahman et al., 2001; Pyle et al., 2007; Langdahl et al., 2018). In contrast, skeletal muscle, which is a terminally differentiated tissue, displays much higher proportions of mutated mtDNA compared to leukocytes from the same individuals (McDonnell et al., 2004; Frederiksen et al., 2006). This shows that the elimination of mutated mtDNA (or the corresponding increase of the proportion of wild-type mitochondria) in leukocytes is related to the high proliferation rate of their progenitors.

To understand this process, one can broadly divide the stem cell population into rather-wild-type cells (i.e. those with heteroplasmy levels below a certain threshold), which would have a selective advantage and those rather-mutated, with above-threshold values of heteroplasmy (Veitia, 2018a). This can be explained by biases in mitochondrial segregation during somatic cell division, which produces stem cells with different mutation loads. The selective advantage of the rather-wild-type cells (i.e. enriched in normal mitochondria) can derive from either a higher mitotic capacity and/or a lower specific death rate. With time, cells enriched in wild-type mitochondria can replace those enriched in mutated mitochondria (Veitia, 2018a). This is in line with the views of (Rajasimha et al., 2008). Their simulations showed that under certain assumptions there is a slow decay of the proportion of mutated mitochondria during childhood followed by an exponential-like decrease at older ages. This behavior mirrors the S-shaped rise of the proportion (P) of cells enriched in wild-type genomes. Figure 1 shows the evolution of the proportion of cells enriched in either mutated or wild-type mitochondria with time and how the latter can invade the population even when their initial proportion (P_0) is low. This S-shaped curve is generated by the logistic function (Verhulst, 1838), which will be further discussed below.

The decrease of heteroplasmy in individuals bearing m.3243A>G, that we can take as a proxy for the proportion of mutated cells, has been proposed to follow an exponential decay (Rajasimha et al., 2008; Grady et al., 2018). However, further analyses have shown that often the exponential models yield physically meaningless values, such as a proportion of mutated cells >100% (Veitia, 2018a, 2019). This cannot happen when the logistic function is used to model the process. For low values of P_0 (<<0.5) and t sufficiently small, there is an exponential increase of the proportion of rather-wild-type cells that eventually outnumber and replace the rather-mutated cells, thus reaching a maximum $P = 1$ or 100% (Figure 2).

Further cases of partial or total subpopulation replacement

Another phenomenon leading to an increase in the proportion of a cell subpopulation

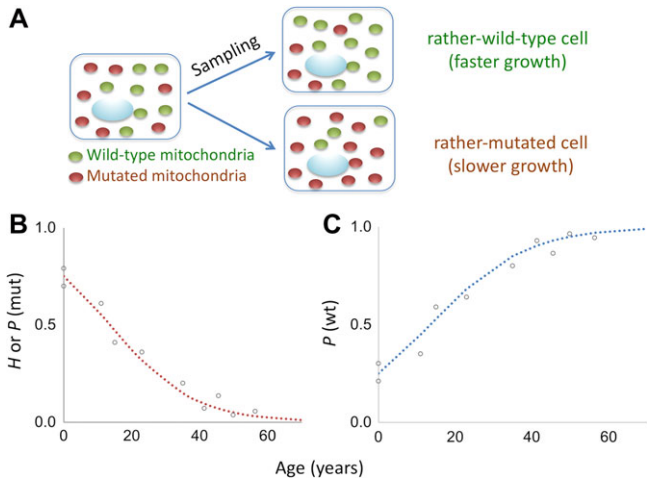

Figure 1 (**A**) Stochastic segregation of mitochondria carrying wild-type or mutated genomes due to sampling effects during cell division. This leads to rather-wild-type or rather-mutated cells with different fitnesseses. (**B**) Graph representing the decrease of heteroplasmy (H) of the mtDNA variant m.3243A>G with age obtained by pooling data from several individuals (redrawn from Veitia, 2018a). Note that H is taken as a proxy for the proportion of rather-mutated cells. (**C**) Time course of rather-wild-type cells invading a population of leukocytes for an initial proportion (P_0) of 0.25. The S-like shape of the curve would be more pronounced for lower values of P_0 (see Figure 2).

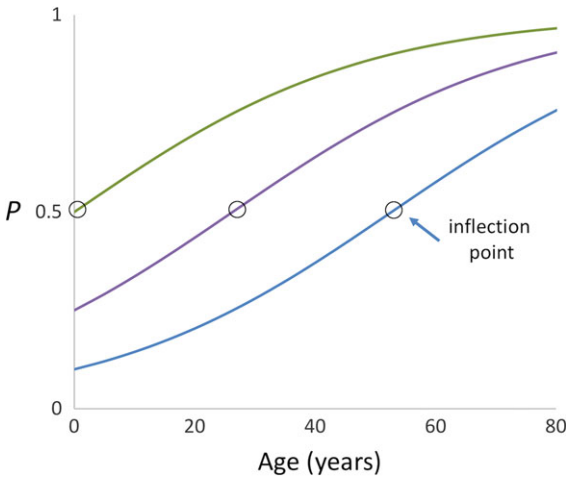

Figure 2 How rather-wild-type cells (having purged m.3243A > G) invade the population for different starting values of P_0 (for $k' = 0.0417$ years^{-1}, the most updated value of the parameter in Equation (1) according to Veitia (2019). Note that the smaller the proportion of healthier cells is, the more S-shaped the curve is, because the inflection point occurs at $P = 0.5$.

concerns the increasing aneuploidy levels of human leukocytes with age and in particular for the sex chromosomes (Guttenbach et al., 1995). In the case of males, the Y chromosome is increasingly lost in the leukocytes of normal men (Figure 3). According to a recent model, cells having lost the Y chromosome (i.e. X0 cells) in an XY background have a selective advantage, but they cannot invade the population because they are produced at a very slow rate (Veitia, 2018b). The selective advantage might be due to the haploinsufficiency of gene(s) encoding negative effector(s) of cell proliferation present on both sex chromosomes or to the plain absence of Y-specific gene(s). In pathological conditions, the proportion of X0 cells can become alarmingly high (Herens et al., 1999; Forsberg et al., 2014; Zink et al., 2017). For instance, a study of leukemic or preleukemic male patients showed that more than 3% of them had a 45,X0 clone in their bone marrow. This phenomenon called clonal hematopoiesis was striking in about 1% of the patients who had more than 90% of X0 cells in their bone marrow (Herens et al., 1999). The proportion of X0 cells also rises in elderly women (Guttenbach et al., 1995).

Linked to the above, biased X chromosome inactivation has been observed in cases of severe combined immunodeficiency (SCID), which is a disease occurring almost exclusively in males, females being healthy mutation carriers. Indeed, early studies have shown that the mothers, heterozygous carriers of *IL2RG* pathogenic variants responsible for SCID, had a highly skewed T-cell X inactivation pattern with the mutated X chromosome being inactive (Puck et al., 1997).

Another, less common example of somatic selection is the expansion of a cell subpopulation after the spontaneous *in vivo* reversion to normal of an inherited damaging variant, in which the revertant cells have an advantage. This has been reported for adenosine deaminase deficiency, which also leads to a recessive form of SCID. The reversion was shown by the absence of the deleterious variant inherited from the mother in most of the B cell lines studied and in a substantial proportion of the alleles genotyped in blood DNA (Hirschhorn et al., 1996). Interestingly, this condition leads to increased concentrations of potentially mutagenic metabolites. This and the fact that it concerns rapidly dividing cells provides a fertile ground for the appearance and spread of spontaneous revertants leading to non-negligible levels of somatic mosaicism. Reversion has also been reported in other conditions, such as Wiskott-Aldrich syndrome (Davis et al., 2010), Fanconi anemia (Gregory et al., 2001) and T-cell immunodeficiencies (Rieux-

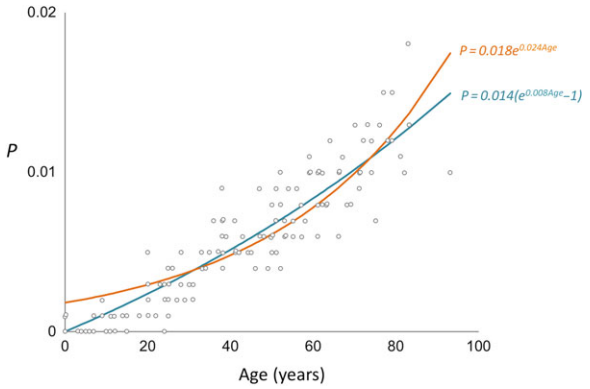

Figure 3 Plot of the proportion P of cells having lost their Y chromosome (X0) vs. age in normal men (data from Guttenbach et al., 1995). The blue line corresponds to the predictions made by a specific exponential model discussed in Veitia (2018b) (Pearson's correlation coefficient, $R = 0.90$, $P < 0.001$). The orange line corresponds to predictions based on a population-genetics formalism (P_0e^{st}, $R = 0.88$, $P < 0.001$). Although both models are suitable, the former model fits data marginally better than the one based on population genetics (this technical discussion is beyond the scope of this paper).

Laucat et al., 2006; Uzel et al., 2008; Kuijpers et al., 2013). Again, most of such spontaneous reversions occur in rapidly proliferating tissues.

Mathematical models describing the selection/replacement process

Let us now briefly explore mathematical models of somatic selection by considering two co-existing cell subpopulations: one having a selective advantage (say, *a*) over the other one (*b*). This can be envisioned as the expansion of a cell lineage bearing an advantageous variant or as the expansion of wild-type cells in a mutated background (as for the rather-wild-type cells mentioned in the case of m.3243A>G). In the context of this mtDNA variant, a model not considering interactions between the two subpopulations has been published (Veitia, 2018a). Accordingly, the competitors simply limit room or resources for growth, yet at some point one of them (*a* cells) will outnumber and practically erase the other (*b* cells). As shown in Supplementary material, the proportion P of *a*-type cells can be described by the logistic equation:

$$P = \frac{P_0}{P_0 + (1 - P_0)e^{-k't}}, \qquad (1)$$

The parameter k' corresponds to the difference $k_a - k_b$, where k_a and k_b are the specific proliferation rates of the relevant cell types and contain information on both cell division and death. Although this equation describes a sigmoid, the populations *a* and *b* are supposed to increase exponentially. It can be argued that there cannot be unrestricted exponential growth within an organism. However, one can consider a parallel between stem cell proliferation and differentiation and the principle of a turbidostat, where the number of growing cells is kept constant by continuously removing a proportion of them. Indeed, as stem cells divide a fraction also differentiates, which limits the exponential growth of the subpopulations *a* and *b*. However, this subtlety is not taken into account in the model and this is the price to pay to obtain a simple solution (i.e. Equation (1)).

The logistic function also describes population growth when the extent of the available resources is limited (Verhulst, 1838). It is the solution of the differential equation:

$$\frac{dN}{dt} = kN(1 - N/K), \qquad (2)$$

where N is the number of individuals in the population, k is a specific growth rate and K is the highest number of individuals that can be sustained by the environment.

We can also derive P versus t by replacing N by P, k by s and making $K = 1$ (or 100%, if P is expressed as a percentage) in Equation (2). P and $1 - P$ are the frequencies of (cells carrying) alleles *a* and *b*, respectively, and s is the selection coefficient, which corresponds to the fitness difference between cells/organisms carrying alleles *a* and *b*. The now classical equation $dP/dt = sP(1 - P)$ (Crow and Kimura, 1970) shows the mathematical convergence with the reasoning leading to Equation (1) explained in Supplementary material and describes directional selection favoring allele *a* in asexual haploid individuals (which can be extended to diploids). It applies to both somatic and gametic selection. The advantage of the population-genetics formalism is that it is less dependent on the assumption made above that *a* and *b* cells grow exponentially. Indeed, it assumes that the processes regulating population size affect both genotypes similarly.

Let us apply the population genetics formalism to the normal loss of the Y chromosome with age (Figure 3). As mentioned above, the initial segment of the logistic curve is basically exponential. In the present context, as P_0 is close to 0 and the extent of the increase of P is small (<1.5%), the logistic function reduces to $P = P_0e^{st}$. If we explore the same dataset as in Veitia (2018b), we obtain $P_0 = 0.0018$ and $s = 0.024$. Now, we can also compare this value of s with the difference between the specific rates of proliferation and death of X0 versus XY cells calculated according to a formalism based on *ad hoc* differential equations similar to those described in the Supplementary material (see details in Veitia, 2018b). We obtain a selective advantage of 0.008. Both models show that losing the Y chromosome does provide a selective advantage. However, the extent of the advantage is different according to the two models because of their different assumptions and formalisms. Figure 3 shows that the one based on ad hoc differential equations is (marginally) better. However, loosely speaking, both types of models are convenient and the choice will depend on the background of the biologist doing the modeling and on the specific question being addressed, as further discussed in Supplementary material for the m.3243A>G variant.

Conclusions and perspectives

This paper emphasizes that somatic selection can be studied using both population-genetics and other system-dynamics models. Such models have practical applications as previously discussed for the m.3243A>G variant (Veitia, 2018a), because having reliable estimates of heteroplasmy at birth from leukocytes is clinically important and can make this determination minimally invasive for patients and avoid routine muscle biopsy. The growth advantage discussed here is also the central tenet underlying gene therapy, particularly for SCID. Indeed, it has been shown that less than 1% of corrected hematopoietic progenitor cells are able to replenish the T cell pool (Cavazzana-Calvo and Fischer, 2007). An estimation of the selection parameter k' or s might be useful in assessing the efficiency of gene therapy protocols. Although not discussed, selection of resistant cells escaping chemotherapy or the contamination of a cell culture with highly proliferative cells are further examples of the logistic replacement process.

In sum, this paper shows that the replacement of a cell type by another bearing a selective advantage can be conveniently modeled by a logistic function according to different formalisms and assumptions. Furthermore, given the continuous nature of the population replacement process, we can say that the replacement curve should be sigmoidal (for low P_0 values) whether the underlying function is the classical logistic equation or not.

[Supplementary material is available at Journal of Molecular Cell Biology online. Work in R.A.V.'s lab is supported by University Paris Diderot and by the Centre National de la Recherche Scientifique.]

Reiner A. Veitia[1,2,*]

[1]Institut Jacques Monod, Université Paris Diderot, 75013 Paris, France
[2]Université Paris-Diderot, 75013 Paris, France
*Correspondence to: Reiner A. Veitia, E-mail: reiner.veitia@ijm.fr

References

Cavazzana-Calvo, M., and Fischer, A. (2007). Gene therapy for severe combined immunodeficiency: are we there yet? J. Clin. Invest. *117*, 1456–1465.

Crow, J.F., and Kimura, M. (1970). An Introduction to Population Genetic Theory. New York: Harper and Row.

Davis, B.R., Yan, Q., Bui, J.H., et al. (2010). Somatic mosaicism in the Wiskott-Aldrich syndrome: molecular and functional characterization of genotypic revertants. Clin. Immunol. *135*, 72–83.

de Laat, P., Koene, S., Heuvel, L.P., et al. (2013). Inheritance of the m.3243A>G mutation. JIMD Rep. *8*, 47–50.

Flierl, A., Reichmann, H., and Seibel, P. (1997). Pathophysiology of the MELAS 3243 transition mutation. J. Biol. Chem. *272*, 27189–27196.

Forsberg, L.A., Rasi, C., Malmqvist, N., et al. (2014). Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. Nat. Genet. *46*, 624–628.

Frederiksen, A.L., Andersen, P.H., Kyvik, K.O., et al. (2006). Tissue specific distribution of the 3243A→G mtDNA mutation. J. Med. Genet. *43*, 671–677.

Goto, Y., Nonaka, I., and Horai, S. (1990). A mutation in the $tRNA^{Leu(UUR)}$ gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. Nature *348*, 651–653.

Grady, J.P., Pickett, S.J., Ng, Y.S., et al. (2018). mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease. EMBO Mol. Med. *10*, pii: e8262.

Gregory, J.J., Jr, Wagner, J.E., Verlander, P.C., et al. (2001). Somatic mosaicism in Fanconi anemia: evidence of genotypic reversion in lymphohematopoietic stem cells. Proc. Natl Acad. Sci. USA *98*, 2532–2537.

Guttenbach, M., Koschorz, B., Bernthaler, U., et al. (1995). Sex chromosome loss and aging: in situ hybridization studies on human interphase nuclei. Am. J. Hum. Genet. *57*, 1143–1150.

't Hart, L.M., Jansen, J.J., Lemkes, H.H., et al. (1996). Heteroplasmy levels of a mitochondrial gene mutation associated with diabetes mellitus decrease in leucocyte DNA upon aging. Hum. Mutat. *7*, 193–197.

Herens, C., Brasseur, E., Jamar, M., et al. (1999). Loss of the Y chromosome in bone marrow cells: results on 1907 consecutive cases of leukaemia and preleukaemia. Clin. Lab. Haematol. *21*, 17–20.

Hirschhorn, R., Yang, D.R., Puck, J.M., et al. (1996). Spontaneous in vivo reversion to normal of an inherited mutation in a patient with adenosine deaminase deficiency. Nat. Genet. *13*, 290–295.

Kuijpers, T.W., van Leeuwen, E.M., Barendregt, B. H., et al. (2013). A reversion of an IL2RG mutation in combined immunodeficiency providing competitive advantage to the majority of $CD8^+$ T cells. Haematologica *98*, 1030–1038.

Laloi-Michelin, M., Meas, T., Ambonville, C., et al. (2009). The clinical variability of maternally inherited diabetes and deafness is associated with the degree of heteroplasmy in blood leukocytes. J. Clin. Endocrinol. Metab. *94*, 3025–3030.

Langdahl, J.H., Larsen, M., Frost, M., et al. (2018). Leucocytes mutation load declines with age in carriers of the m.3243A>G mutation: A 10-year Prospective Cohort. Clin. Genet. *93*, 925–928.

McDonnell, M.T., Schaefer, A.M., Blakely, E.L., et al. (2004). Noninvasive diagnosis of the 3243A>G mitochondrial DNA mutation using urinary epithelial cells. Eur. J. Hum. Genet. *12*, 778–781.

Puck, J.M., Pepper, A.E., Henthom, P.S., et al. (1997). Mutation analysis of IL2RG in human X-linked severe combined immunodeficiency. Blood *89*, 1968–1977.

Pyle, A., Taylor, R.W., Durham, S.E., et al. (2007). Depletion of mitochondrial DNA in leucocytes harbouring the 3243A→G mtDNA mutation. J. Med. Genet. *44*, 69–74.

Rahman, S., Poulton, J., Marchington, D., et al. (2001). Decrease of 3243 A→G mtDNA mutation from blood in MELAS syndrome: a longitudinal study. Am. J. Hum. Genet. *68*, 238–240.

Rajasimha, H.K., Chinnery, P.F., Samuels, D.C., et al. (2008). Selection against pathogenic mtDNA mutations in a stem cell population leads to the loss of the 3243A→G mutation in blood. Am. J. Hum. Genet. *82*, 333–343.

Rieux-Laucat, F., Hivroz, C., Lim, A., et al. (2006). Inherited and somatic CD3ζ mutations in a patient with T-cell deficiency. N. Engl. J. Med. *354*, 1913–1921.

Satoh, M., and Kuroiwa, T. (1991). Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Exp. Cell Res. *196*, 137–140.

Uzel, G., Tng, E., Rosenzweig, S.D., et al. (2008). Reversion mutations in patients with leukocyte adhesion deficiency type-1 (LAD-1). Blood *111*, 209–218.

van den Ouweland, J.M., Lemkes, H.H., Ruitenbeek, W., et al. (1992). Mutation in mitochondrial $tRNA^{Leu(UUR)}$ gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness. Nat. Genet. *1*, 368–371.

Veitia, R.A. (2018a). How the most common mitochondrial DNA mutation (m.3243A>G) vanishes from leukocytes: a mathematical model. Hum. Mol. Genet. *27*, 1565–1571.

Veitia, R.A. (2018b). On the loss of human sex chromosomes in lymphocytes with age: a quantitative treatment. Eur. J. Hum. Genet. *26*, 1875–1878.

Veitia, R.A. (2019). Further quantitative insights into the decrease of heteroplasmy of m.3243A>G with age in leukocytes. Clin. Genet. *95*, 542–543.

Verhulst, P.-F. (1838). Notice sur la loi que la population poursuit dans son accroissement. Corresp. Mathématique Phys. *10*, 113–121.

Zink, F., Stacey, S.N., Norddahl, G.L., et al. (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood *130*, 742–752.