



**HAL**  
open science

# What does the Distribution of Fitness Effects of new mutations (DFE) reflect? Insights from plants

Jun Chen, Thomas Bataillon, Sylvain Glémin, Martin Lascoux

► **To cite this version:**

Jun Chen, Thomas Bataillon, Sylvain Glémin, Martin Lascoux. What does the Distribution of Fitness Effects of new mutations (DFE) reflect? Insights from plants. *New Phytologist*, 2022, 233 (4), pp.1613-1619. 10.1111/nph.17826 . hal-03414519

**HAL Id: hal-03414519**

**<https://hal.science/hal-03414519v1>**

Submitted on 4 Jun 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License



## Tansley insight

# What does the distribution of fitness effects of new mutations reflect? Insights from plants

Author for correspondence:  
Martin Lascoux  
Email: martin.lascoux@ebc.uu.se

Received: 1 June 2021  
Accepted: 28 September 2021

Jun Chen<sup>1</sup> , Thomas Bataillon<sup>2</sup> , Sylvain Glémin<sup>3,4</sup>  and Martin Lascoux<sup>4</sup> 

<sup>1</sup>College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China; <sup>2</sup>Bioinformatics Research Centre, Aarhus University, C.F. Möllers Allé 8, Aarhus C DK-8000, Denmark; <sup>3</sup>Centre National de la Recherche Scientifique (CNRS), ECOBIO (Ecosystèmes, Biodiversité, Evolution) – Unité Mixte de Recherche (UMR) 6553, Université de Rennes, Rennes F-35000, France; <sup>4</sup>Program in Plant Ecology and Evolution, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala 75236, Sweden

## Contents

Summary	1613	IV. Conclusions	1617
I. Introduction	1613	Acknowledgements	1618
II. DFE and related quantities	1614	References	1618
III. The DFE and the biology of plants	1615		

*New Phytologist* (2022) **233**: 1613–1619  
doi: 10.1111/nph.17826

**Key words:** demography, distribution of fitness effects, life history traits, mating systems, nearly neutral theory.

## Summary

The distribution of fitness effects (DFE) of new mutations plays a central role in molecular evolution. It is therefore crucial to be able to estimate it accurately from genomic data and to understand the factors that shape it. After a rapid overview of available methods to characterize the fitness effects of mutations, we review what is known on the factors affecting them in plants. Available data indicate that life history traits (e.g. mating system and longevity) have a major effect on the DFE. By contrast, the impact of demography within species appears to be more limited. These results remain to be confirmed, and methods to estimate the joint evolution of demography, life history traits, and the DFE need to be developed.

## I. Introduction

Without mutations there is no evolution. The path evolution will take depends, to a large extent, on how many mutations enter the population each generation and how they affect fitness. This general picture was, by and large, firmly established by the work of Fisher, Wright, and Haldane, and its main tenets have not changed much since then (Charlesworth *et al.*, 2017). However, until recently it was difficult to assess the relative parts played by evolutionary forces at each of these steps since data were not available. This has drastically changed today with the availability of genomic data in many species. Though mutation rates and effects remain intrinsically very difficult to quantify, it is nonetheless possible today to obtain both indirect (i.e. phylogeny-based) and direct (i.e. pedigree-based) genome-wide estimates of

mutation rates – see Moorjani *et al.* (2016) and Chintalapati & Moorjani (2020) and references therein – as well as to characterize the distribution of fitness of mutations. We will focus on the latter in this article.

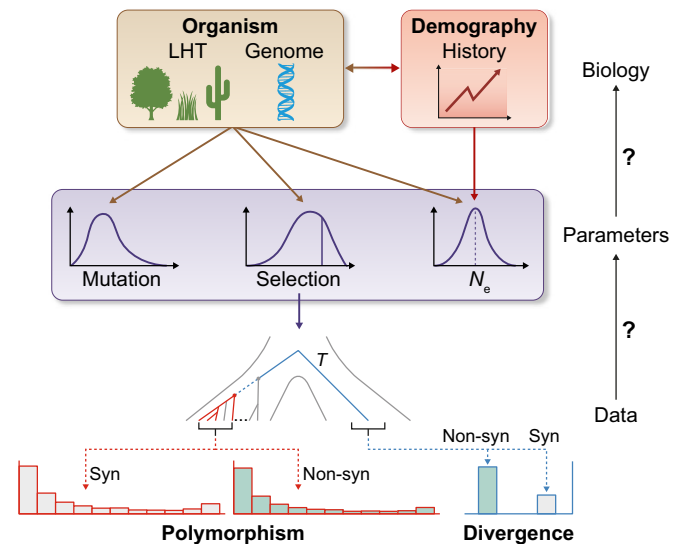
Mutations can be classified according to their fitness effects: deleterious, neutral, and beneficial. All agree that most mutations are most often either deleterious or neutral and that the contribution of highly deleterious and beneficial mutations to standing variation is limited (Eyre-Walker & Keightley, 2007). The highly deleterious mutations are quickly eliminated from the population and the beneficial mutations are likely rare. However, agreement ends here, and there are contrasted views on the relative size and role in evolution of the different fitness classes (Jensen *et al.*, 2018; Kern & Hahn, 2018). In part, these disagreements reflect different views on evolution and on the relative importance of the forces driving it.

But, in no lesser part, these disagreements simply stem from the inherent difficulty to characterize the fitness effects of mutations.

Different methods have been devised to characterize the fitness effect of mutations from sequence data. A first group of methods assumes that mutations occurring at nucleotide sites that are highly conserved across species are likely to be deleterious (Ng & Henikoff, 2003; Davydov *et al.*, 2010). Accordingly, these methods assign a score to a site that measures how much the site departs from the variation that would be observed in a multispecies alignment if the sites were evolving neutrally. The resulting score is effectively measuring how deleterious the mutations are at a given site. The pros of this approach are that it makes single-sites predictions, is readily available for an increasing number of species, can incorporate covariates from functional genomic studies, and does not depend on elusive population genetics parameters (e.g. effective population size). However, it can be misleading for predictions on the fitness effect of genetic variation currently segregating as it does not directly estimate fitness effects and focuses exclusively on deleterious mutations. This approach was recently used to characterize the impact on fitness of amino acid changing mutations in a host of plant species; for example, poplars (*Populus ilicifolia*; Z. Chen *et al.* 2020) and sorghum (*Sorghum bicolor*; Valluru *et al.*, 2019).

In contrast to methods using scores based on conservation/divergence across species, methods in the second group leverage polymorphism within species and estimate an underlying distribution of fitness effects (DFE) of new mutations. This is achieved by modelling the site frequency spectrum (SFS) at putatively neutral and selected sites, for instance, synonymous and nonsynonymous sites. The DFE of new mutations is the distribution of the relative frequencies of selection coefficients  $s$ , extending from strongly and weakly deleterious, through neutral mutations, to slightly and strongly advantageous (Moutinho *et al.*, 2020). The most recent estimation methods consider both deleterious and beneficial mutations, although we stress that estimating the fraction of beneficial mutations is intrinsically challenging simply because beneficial mutations are rare and have a modest impact on the SFS. In contrast to methods relying exclusively on conservation scores, methods based on the DFE estimate the effect of mutations currently segregating and do not assume conservation of effects across species. Estimates of the DFE from SFS data have been reported across many animal and plant species (Galtier, 2016; Chen *et al.*, 2017; Castellano *et al.*, 2019).

Being able to estimate the DFE is a prerequisite, and this remains a difficult statistical problem. Understanding what factors determine the DFE would greatly help focusing our efforts on the most important sources of variation (Fig. 1). It should be noted here that though being able to estimate a DFE reliably is crucial to assessing competing models of molecular evolution, obtaining a reliable estimate of the DFE is also important to characterize the evolution of species: mutations rates and the DFE are not fixed entities but instead are under the same evolutionary forces as any other complex phenotypic trait. As they evolve, so will other key biological features of organisms; for instance, life history traits, mating systems, ploidy level or genome structure, and population history (demography, environment). In turn, changes of these features will influence the evolution of the DFE.



**Fig. 1** Inverting the map: exploiting genome-wide patterns of polymorphism and divergence to learn about mutation and their fitness effects. The left side of the figure sketches a causal path from proximal evolutionary causes at the top to the patterns in the data used to estimate the distribution of fitness effects (DFE) of new mutations at the bottom. The right side pictures the inference path, from data to the biological factors that have shaped the DFE.  $T$  is the divergence time between the focal species and the outgroup species. LHT, life-history traits;  $N_e$ , effective population size; Non-syn, nonsynonymous; Syn, synonymous.

## II. DFE and related quantities

The estimation of the fitness effect of new mutations is based on a fundamental intuition: the fitness effect of a new mutation influences the frequency at which it segregates in a population. However, demography also affects allelic frequencies, making it crucial to tease apart the effect of demography and selection. In brief, this is done by comparing the SFS of putatively neutral mutations (typically synonymous ones) only affected by demography to mutations under selection (typically non-synonymous ones) affected by both demography and selection, as further described in Box 1. Importantly, as usual in population genetics, these methods provide estimation of the population-scale selection coefficient  $N_e s$ , the product of effective population size  $N_e$  and selection coefficient  $s$ , not directly  $s$  alone. It means that in small populations, a higher proportion of mutations is mostly affected by drift and behaves like neutral ones, which is the central idea of the nearly neutral theory of molecular evolution (Ohta, 1973).

Many studies choose to focus on quantities related to the DFE, which may be simpler to calculate. Two important statistics are the ratio of nucleotide diversity at nonsynonymous and synonymous sites  $\pi_N/\pi_S$  and the proportion of adaptive nonsynonymous substitutions  $\alpha$  (Smith & Eyre-Walker, 2002; Galtier, 2016). The ratio  $\pi_N/\pi_S$  roughly measures the proportion of effectively neutral mutations; that is, mutations for which  $|N_e s| \lesssim 1$  (Ohta, 1973). It gives a global measure of the efficacy of selection relative to drift. The proportion of adaptive nonsynonymous substitutions  $\alpha$  has attracted a lot of attention. Initially,  $\alpha$  was inferred using an

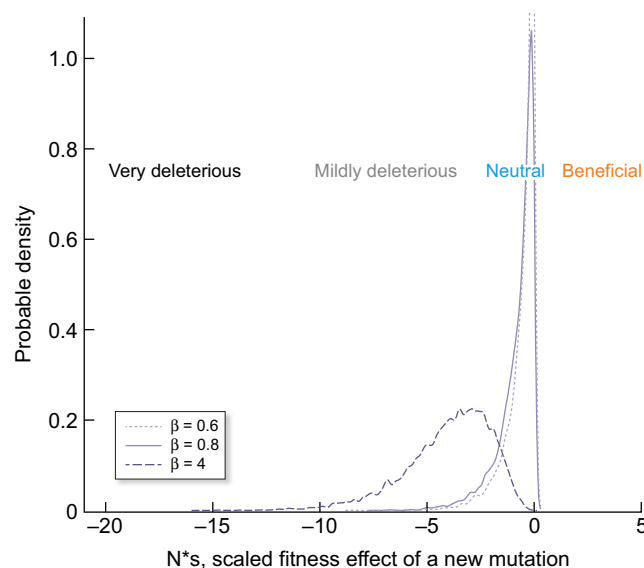
extension of the McDonald–Kreitman test, a test of neutrality that compares the number of polymorphisms with the number of substitutions in two types of sites, synonymous and nonsynonymous.  $\alpha$  can then be defined as  $1 - (D_s P_n) / (D_n P_s)$  ( $D$ , divergence;  $P$ , polymorphism;  $s$ , synonymous;  $n$ , nonsynonymous). A related quantity is the rate of adaptive nonsynonymous substitutions relative to the mutation rate,  $\omega_a$  (Galtier, 2016). However, these estimates were shown to be biased – in particular by the presence of slightly deleterious mutations – and have since then been replaced by estimates derived from the DFE and accounting for demography (Keightley & Eyre-Walker, 2007; Galtier, 2016; Tataru *et al.*, 2017; Tataru & Bataillon, 2019; and see Box 1).

### III. The DFE and the biology of plants

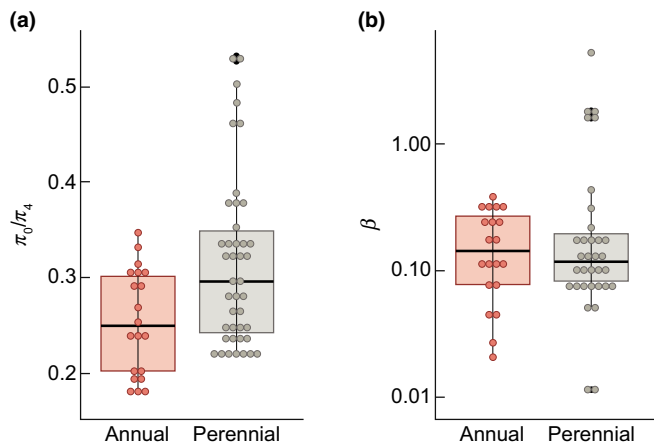
Plants harbor a large diversity of life-history traits, genome features, and demographic histories, and their impact on the DFE has just started to be investigated. First, because the DFE depends on  $N_e$ , we certainly expect demography to matter. However, it is important to distinguish contingent demographic processes – for instance, changes related to glacial cycles – from those that are related to the intrinsic properties of the species. Historical demographic events will seriously affect the estimation of the DFE and their interpretation, but they can, in principle at least, be accommodated (Brandvain & Wright, 2016) and will not have permanent effects.

#### Box 1 Estimation of the distribution of fitness effects.

Most estimation methods model information from the site frequency spectrum (SFS) for two classes of nucleotide changes: putatively neutral ones, generally synonymous or four-fold degenerate ones, and putatively selected ones, classically nonsynonymous or four-fold degenerate changes (Bataillon & Bailey, 2014; Moutinho *et al.*, 2020). The SFS describes polymorphism in a sample of chromosomes by counting the number of segregating sites at which a mutation is represented by one copy, two copies, and so on. If we can infer the ancestral/derived allele at a polymorphic site, we will obtain an unfolded SFS with  $n - 1$  classes for a sample of size  $n$ . If we do not know the ancestral site, the SFS will be folded with  $n/2$  or  $(n + 1)/2$  classes depending on whether  $n$  is even or odd. Unfolded SFSs are more informative, but identifying the ancestral state is still fraught with difficulties and misidentification can bias distribution of fitness effects (DFE) estimates (Keightley & Jackson, 2018). The nonsynonymous SFS will be affected by selection. For instance, in the presence of slightly deleterious mutations, one expects an excess of rare variants compared with the neutral SFS in a population at mutation drift equilibrium. Unfortunately, selection is not the only evolutionary force affecting the SFS. Demography, too, will influence it, and, for example, a similar excess of rare variants will result from population expansion or a not too recent bottleneck. One main challenge, therefore, is to distinguish between the effects of selection and demography. This is achieved by assuming that synonymous sites are evolving under neutrality and can be used to infer demography and correct for it when estimating the effect of selection from the nonsynonymous SFS. This assumption is central to all methods inferring the DFE from the SFS. Different maximum likelihood methods exist to estimate the DFE. Usually, they assume that the effects of deleterious mutations are gamma distributed, with a mean  $N_e s$  ( $N_e$ , effective population size;  $s$ , selection coefficient) and a shape parameter  $\beta$  (see Box Fig. 1). Otherwise, they primarily differ in the way they correct for the effects of demography and whether they require divergence data and consider beneficial mutations (Eyre-Walker *et al.*, 2006; Keightley & Eyre-Walker, 2007; Galtier, 2016; Tataru *et al.*, 2017; Tataru & Bataillon, 2019). Moutinho *et al.* (2020) reviewed the different methods. Generally, the DFE is estimated at the species or population level, but recent implementations also allow testing for invariance or change of the DFE across species or gene categories (Tataru & Bataillon, 2019) or correlation of DFEs across populations of the same species (Huang *et al.*, 2021).



**Box Fig. 1** Three hypothetical, gamma distributed negative DFE. The overall shape of the DFE is governed by the parameter  $\beta$ .



**Fig. 2** Box plots of (a) the  $\pi_0/\pi_4$  ratio and (b) the shape parameter  $\beta$  of the negative part of the distribution of fitness effects (DFE) for annual and perennial plant species. Data are from J. Chen *et al.* (2017, 2020, 2021).  $\pi_0/\pi_4$  is the ratio of nucleotide diversity at zero-fold and four-fold degenerate sites and is used as a proxy for  $\pi_N/\pi_S$ . When estimating the shape of the negative part of the DFE,  $\beta$ , one assumes that the negative DFE follows a gamma distribution with shape parameter  $\beta$ . The horizontal line gives the median, circles are individual values, and the box gives the interquartile range.

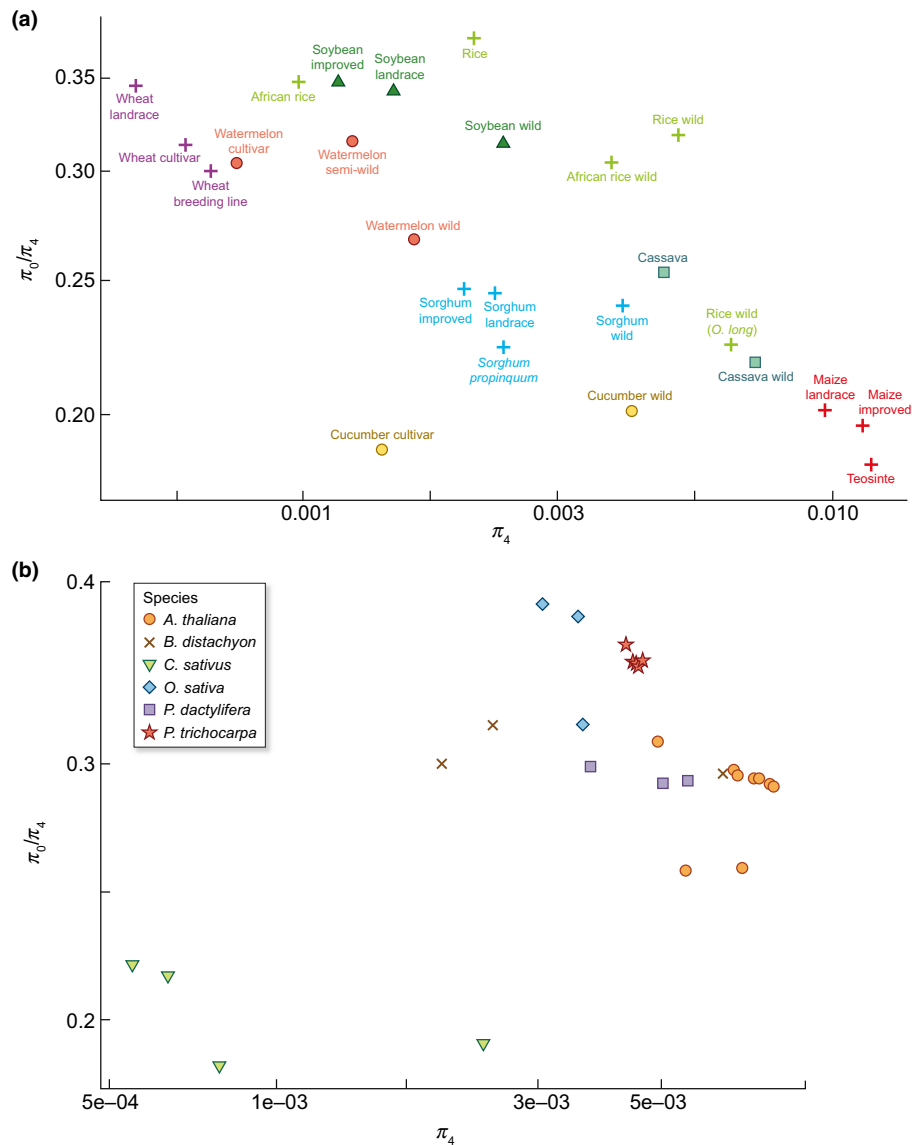
Life-history traits can also affect the effective size of a species, and thereby the proportion of nearly neutral mutations. For example,  $N_e$  is predicted to be much lower in selfing than in outcrossing species (Charlesworth, 2003). A second idea is that life-history traits can affect the strength of selection experienced by a species and the speed of adaptation to changing environment. For example, long-lived species likely experience stronger and more variable selective pressure and may adapt more slowly (Petit & Hampe, 2006).

So far, most studies have not made explicit comparisons of the DFE but instead have focused on specific aspects of the DFE in relation to a given hypothesis. In a comprehensive study aimed at testing the prediction of the nearly neutral theory of molecular evolution, using 34 animal and 28 plant species, Chen *et al.* (2017) investigated the effect of different life-history traits on the  $\pi_N/\pi_S$  ratio. As already discussed herein, we expect the  $\pi_N/\pi_S$  ratio to reflect the proportion of effectively neutral mutations, and so to be inversely related to the effective population size of a species. The  $\pi_N/\pi_S$  ratio should, therefore, be higher in self-fertilizing species than in outcrossing ones. Indeed, this is what is observed, with the other factor having a major effect being longevity. In fact, although the proportion of effectively neutral mutations was much smaller in outcrossing annuals than in self-fertilizing ones, the proportion of effectively neutral mutations was similar to that in self-fertilizing annuals, indicating a strong effect of longevity. An effect of the reproductive system was also observed by Muyle *et al.* (2020), who tested the effect of dioecy vs gynodioecy vs hermaphroditism on the DFE and on derived statistics in the genus *Silene*. As for selfing vs outcrossing, the change in sexual system is associated with a change in the DFE: dioecious species harbored more diversity, more efficient purifying selection (low  $\pi_N/\pi_S$  ratio), and evolved with a higher rate of adaptive nonsynonymous substitutions  $\omega_a$ . The impact of life-history traits on other properties of the DFE, such as

its shape parameter for the negative part (Box 1), was less clear, but a further study (J. Chen *et al.*, 2020) found a clear difference between selfers and outcrossers (see also Douglas *et al.*, 2015). Based on some of the species used in J. Chen *et al.* (2017, 2020) and a few additional ones, we recently estimated the full DFE in annuals and perennial species. A comparison of two DFE parameters between these two groups is given in Fig. 2(a,b). As already observed, there is a clear difference in the  $\pi_N/\pi_S$  ratio; but the DFE shape is similar, with perennials tending to have a wider range of shapes (Fig. 2b).

In contrast to the profound effect of life-history traits on the DFE and derived summary statistics such as  $\pi_N/\pi_S$ ,  $\alpha$ , or  $\omega_a$ , the effect of demography within species or domestication seems more limited. Particularly striking is the limited effect of domestication on  $\pi_N/\pi_S$  (J. Chen *et al.*, 2017, 2020). As shown in Fig. 3(a), domestication led to a very severe decrease in synonymous nucleotide diversity  $\pi_S$ , but this was not accompanied by a similar change in  $\pi_N/\pi_S$ , which, on the contrary, appears extremely stable. Similarly, variation in levels of neutral nucleotide diversity across populations in a range of different species was not associated with commensurate changes in  $\pi_N/\pi_S$  (Fig. 3b). Lack of marked differences between populations within the same species does not preclude differences between species that have experienced very different demographic histories. For example, Jaramillo-Correa *et al.* (2020) compared two parapatric Mediterranean pine species, stone pine (*Pinus pinea*) and maritime pine (*Pinus pinaster*). The former went through a severe bottleneck and has today an extremely limited genetic variation compared with maritime pine ( $\pi_S = 0.00043$  vs  $\pi_S = 0.0045$ , with corresponding  $\pi_N/\pi_S = 0.45$  vs  $\pi_N/\pi_S = 0.26$ ). The DFEs inferred in each species were similar, but stone pine had a significantly higher proportion of slightly deleterious mutations and maritime pine a higher proportion of highly deleterious mutations. However, stone pine can be divided into a western and an eastern cluster, and no differences were observed between the DFEs of these two clusters in spite of marked differences in effective population sizes ( $N_e(\text{west}) = 5620$  vs  $N_e(\text{east}) = 841$ ). It therefore seems that the differences observed between the two pine species may not be entirely due to differences in demographic history, at least not the recent one. Another striking illustration of the joint effect of demography and mating system is provided by the comparison of populations of *Arabidopsis alpina* (Laenen *et al.*, 2018). Southern European populations have a nucleotide diversity varying between 0.0046 and 0.0080, whereas Scandinavian ones have a mean nucleotide diversity more than ten-fold smaller (0.0002). Within the southern European populations,  $\pi_N/\pi_S$  varied between 0.27 and 0.29, whereas the value was 0.39 in Scandinavia. Scandinavian populations, unlike their southern European counterparts, experienced a severe bottleneck during the Last Glacial Maximum. Perhaps more importantly, they exhibit a much higher selfing rate. Considering results obtained in other species, the differences in DFE between the two groups of populations might primarily reflect the change in mating system rather than be a consequence of differences in past demography. A similar conclusion can be drawn when comparing deleterious load between sorghum (a selfer) and maize (*Zea mays*, an outcrosser) (Lozano *et al.*, 2021).

Finally, genome properties will also affect the DFE. For instance, GC-biased gene conversion in both *Arabidopsis thaliana*,



**Fig. 3** Difference in the  $\pi_0/\pi_4$  ratio (a) between domesticated species and their wild relatives and (b) between populations within species (*Arabidopsis thaliana*, *Brachypodium distachyon*, *Crocus sativus*, *Oryza sativa*, *Phoenix dactylifera*, *Populus trichocarpa*). Data are from Chen *et al.* (2017).  $\pi_0/\pi_4$  is the ratio of nucleotide diversity at zero-fold and four-fold degenerate sites and is used as a proxy for  $\pi_N/\pi_S$ .

*Arabidopsis lyrata*, and *Capsella grandiflora* contribute to a genome-wide signal of decreased efficacy of selection (Hämälä & Tiffin, 2020). The effect of polyploidy has not been studied extensively yet, but available studies indicate relaxed selection in polyploids (Douglas *et al.*, 2015) and a strong parental legacy effect on the subgenomes (Kryvokhyzha *et al.*, 2019). One could therefore expect the DFE of the subgenomes to differ among themselves and also from the parental species' DFEs.

In general, it appears that the DFE is primarily determined by the biological properties of the species, here captured by traits such as mating system or longevity (and maybe polyploidy), and, secondarily, by their demographic history – but see Willi *et al.* (2020) for a case where demography and mating system apparently have a similar impact. A possible reason is that the DFE is weakly sensitive to short-term demographic variations (Brandvain & Wright, 2016) but influenced more by long-term trends, well captured by basic

biological characteristics of a species, as also proposed for polymorphism patterns in animals (Romiguier *et al.*, 2014). Of course, this is a somewhat arbitrary separation, as a great deal of the demography of a species is also determined by its biology and it is, intrinsically, difficult to separate the effect of selection and demography, even though it would be highly desirable (Li *et al.*, 2012).

#### IV. Conclusions


The first attempts to estimate the DFE only date back to the early 2000s, and much development in estimation methods and gathering of data has already occurred. Given the rate at which high-quality genomic data sets become available, we expect numerous estimates of the DFE across plant species to be available soon. This should soon allow us to combine phylogenetic and

population genetics approaches within an integrative framework (Brevet & Lartillot, 2021). This will be an important step for a more robust understanding of the factors shaping the DFE and assessing its evolutionary stability. A truly comparative approach should allow us to ascertain some of the trends observed through comparisons of population genetic estimates of the DFE across species. Another promising path is to use knowledge about the genome structure to improve estimation of the DFE. Whereas some methods rely on machine learning and require a large amount of information that is only available in a handful of species (Huang & Siepel, 2019), others should be easily implemented in numerous less well studied species (Chen *et al.*, 2021). Ultimately, this should help to tease apart the effects of life history and genomic traits vs demography (basically  $N_e$ ) on the DFE. Finally, when possible (i.e. with species with short generation times), these efforts should be supplemented by more direct experimental approaches, such as mutation accumulation experiments (Böndel *et al.*, 2019).


## Acknowledgements


Martin Lascoux acknowledges support from the Swedish Research Council (grant no. 2019-00806) and the Swedish Research Council for Sustainable Development (Formas) (grant no. 2020-01456).

## ORCID

Thomas Bataillon  <https://orcid.org/0000-0002-4730-2538>

Jun Chen  <https://orcid.org/0000-0003-3187-7668>

Sylvain Glémin  <https://orcid.org/0000-0001-7260-4573>

Martin Lascoux  <https://orcid.org/0000-0003-1699-9042>

## References

- Bataillon T, Bailey SF. 2014. Effects of new mutations on fitness: insights from models and data. *Annals of the New York Academy of Sciences* 1320: 76–92.
- Böndel KB, Kraemer SA, Samuels T, McClean D, Lachapelle J, Ness RW, Colegrave N, Keightley PD. 2019. Inferring the distribution of fitness effects of spontaneous mutations in *Chlamydomonas reinhardtii*. *PLoS Biology* 17: e3000192.
- Brandvain Y, Wright SI. 2016. The limits of natural selection in a nonequilibrium world. *Trends in Genetics* 32: 201–210.
- Brevet M, Lartillot N. 2021. Reconstructing the history of variation in effective population size along phylogenies. *Genome Biology and Evolution* 13: evab150.
- Castellano D, Macià MC, Tataru P, Bataillon T, Munch K. 2019. Comparison of the full distribution of fitness effects of new amino acid mutations across great apes. *Genetics* 213: 953–966.
- Charlesworth D. 2003. Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358: 1051–1070.
- Charlesworth D, Barton NH, Charlesworth B. 2017. The sources of adaptive variation. *Proceedings of the Royal Society B: Biological Sciences* 284: e20162864.
- Chen J, Bataillon T, Glémin S, Lascoux M. 2021. Hunting for beneficial mutations: conditioning on SIFT scores when estimating the distribution of fitness effect of new mutations. *Genome Biology and Evolution*. doi: 10.1093/gbe/evab151.
- Chen J, Glémin S, Lascoux M. 2017. Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution* 34: 1417–1428.
- Chen J, Glémin S, Lascoux M. 2020. From drift to draft: how much do beneficial mutations actually contribute to predictions of Ohta's slightly deleterious model of molecular evolution? *Genetics* 214: 1005–1018.
- Chen Z, Ai F, Zhang J, Ma X, Yang W, Wang W, Su Y, Wang M, Yang Y, Mao K *et al.* 2020. Survival in the tropics despite isolation, inbreeding and asexual reproduction: insights from the genome of the world's southernmost poplar (*Populus ilicifolia*). *The Plant Journal* 103: 430–442.
- Chintalapati M, Moorjani P. 2020. Evolution of the mutation rate across primates. *Current Opinion in Genetics & Development* 62: 58–64.
- Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology* 6: e1001025.
- Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R, Ågren JA, Hazzouri KM, Wang W *et al.* 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proceedings of the National Academy of Sciences, USA* 112: 2806–2811.
- Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nature Reviews Genetics* 8: 610–618.
- Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173: 891–900.
- Galtier N. 2016. Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genetics* 12: e1005774.
- Hämälä T, Tiffin P. 2020. Biased gene conversion constrains adaptation in *Arabidopsis thaliana*. *Genetics* 215: 831–846.
- Huang X, Fortier AL, Coffman AJ, Struck TJ, Irby MN, James JE, León-Burguete JE, Ragsdale AP, Gutenkunst RN. 2021. Inferring genome-wide correlations of mutation fitness effects between populations. *Molecular Biology and Evolution* 38: 4588–4602.
- Huang Y-F, Siepel A. 2019. Estimation of allele-specific fitness effects across human protein-coding sequences and implications for disease. *Genome Research* 29: 1310–1321.
- Jaramillo-Correa JP, Bagnoli F, Grivet D, Fady B, Aravanopoulos FA, Vendramin GG, González-Martínez SC. 2020. Evolutionary rate and genetic load in an emblematic Mediterranean tree following an ancient and prolonged population collapse. *Molecular Ecology* 29: 4797–4811.
- Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D, Charlesworth B. 2018. The importance of the neutral theory in 1968 and 50 years on: a response to Kern and Hahn 2018. *Evolution* 73: 111–114.
- Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177: 2251–2261.
- Keightley PD, Jackson BC. 2018. Inferring the probability of the derived vs. the ancestral allelic state at a polymorphic site. *Genetics* 209: 897–906.
- Kern AD, Hahn MW. 2018. The neutral theory in light of natural selection. *Molecular Biology and Evolution* 35: 1366–1371.
- Kryvokhyzha D, Salcedo A, Eriksson MC, Duan T, Tawari N, Chen J, Guerrina M, Kreiner JM, Kent TV, Lagercrantz U *et al.* 2019. Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid *Capsella bursa-pastoris* (Brassicaceae). *PLoS Genetics* 15: e1007949.
- Laenen B, Tedder A, Nowak MD, Toräng P, Wunder J, Wötzel S, Steige KA, Kourmpetis Y, Odong T, Drouzas AD *et al.* 2018. Demography and mating system shape the genome-wide impact of purifying selection in *Arabidopsis alpina*. *Proceedings of the National Academy of Sciences, USA* 115: 816–821.
- Li J, Li H, Jakobsson M, Li S, Sjödin P, Lascoux M. 2012. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular Ecology* 21: 28–44.
- Lozano R, Gazave E, dos Santos JPR, Stetter MG, Valluru R, Bandillo N, Fernandes SB, Brown PJ, Shakoob N, Mockler TC *et al.* 2021. Comparative evolutionary genetics of deleterious load in sorghum and maize. *Nature Plants* 7: 17–24.
- Moorjani P, Gao Z, Przeworski M. 2016. Human germline mutation and the erratic evolutionary clock. *PLoS Biology* 14: e2000744.
- Moutinho AF, Bataillon T, Dutheil JY. 2020. Variation of the adaptive substitution rate between species and within genomes. *Evolutionary Ecology* 34: 315–338.
- Muyle A, Martin H, Zemp N, Mollion M, Gallina S, Tavares R, Silva A, Bataillon T, Widmer A, Glémin S *et al.* 2020. Dioecy is associated with high genetic diversity and adaptation rates in the plant genus *Silene*. *Molecular Biology and Evolution* 38: 805–818.

- Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* 31: 3812–3814.
- Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246: 96–98.
- Petit RJ, Hampe A. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology and Systematics* 37: 187–214.
- Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N *et al.* 2014. Comparative population genomics in animal uncovers the determinants of genetic diversity. *Nature* 515: 261–263.
- Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415: 1022–1024.
- Tataru P, Bataillon T. 2019. POLYDFEv2.0: testing for invariance of the distribution of fitness effects within and across species. *Bioinformatics* 35: 2868–2869.
- Tataru P, Mollion M, Glémin S, Bataillon T. 2017. Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. *Genetics* 207: 1103–1119.
- Valluru R, Gazave EE, Fernandes SB, Ferguson JN, Lozano R, Hirannaiah P, Zuo T, Brown PJ, Leakey ADB, Gore MA *et al.* 2019. Deleterious mutation burden and its association with complex traits in sorghum (*Sorghum bicolor*). *Genetics* 211: 1075–1087.
- Willi Y, Fracassetti M, Bachmann O, Buskirk JV. 2020. Demographic processes linked to genetic diversity and positive selection across a species' range. *Plant Communications* 1: e100111.



## About New Phytologist

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Foundation, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.
- Regular papers, Letters, Viewpoints, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <23 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**