Moderating the neutralist–selectionist debate: exactly which propositions are we debating, and which arguments are valid?

Menno J. de Jong[1,*] , Cock van Oosterhout[2] , A. Rus Hoelzel[3] and Axel Janke[1,4,5]

[1]*Senckenberg Biodiversity and Climate Research Institute (SBiK-F), Georg-Voigt-Strasse 14-16, Frankfurt am Main 60325, Germany*
[2]*Centre for Ecology, Evolution and Conservation, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK*
[3]*Department of Biosciences, Durham University, South Road, Durham, DH1 3LE, UK*
[4]*Institute for Ecology, Evolution and Diversity, Goethe University, Max-von-Laue-Strasse 9, Frankfurt am Main 60438, Germany*
[5]*LOEWE-Centre for Translational Biodiversity Genomics (TBG), Senckenberg Nature Research Society, Georg-Voigt-Straße 14-16, Frankfurt am Main 60325, Germany*

ABSTRACT

Half a century after its foundation, the neutral theory of molecular evolution continues to attract controversy. The debate has been hampered by the coexistence of different interpretations of the core proposition of the neutral theory, the 'neutral mutation–random drift' hypothesis. In this review, we trace the origins of these ambiguities and suggest potential solutions. We highlight the difference between the original, the revised and the nearly neutral hypothesis, and re-emphasise that none of them equates to the null hypothesis of strict neutrality. We distinguish the neutral hypothesis of protein evolution, the main focus of the ongoing debate, from the neutral hypotheses of genomic and functional DNA evolution, which for many species are generally accepted. We advocate a further distinction between a narrow and an extended neutral hypothesis (of which the latter posits that random non-conservative amino acid substitutions can cause non-ecological phenotypic divergence), and we discuss the implications for evolutionary biology beyond the domain of molecular evolution. We furthermore point out that the debate has widened from its initial focus on point mutations, and also concerns the fitness effects of large-scale mutations, which can alter the dosage of genes and regulatory sequences. We evaluate the validity of neutralist and selectionist arguments and find that the tested predictions, apart from being sensitive to violation of underlying assumptions, are often derived from the null hypothesis of strict neutrality, or equally consistent with the opposing selectionist hypothesis, except when assuming molecular panselectionism. Our review aims to facilitate a constructive neutralist–selectionist debate, and thereby to contribute to answering a key question of evolutionary biology: what proportions of amino acid and nucleotide substitutions and polymorphisms are adaptive?

Key words: neutral theory, molecular evolution, neutralist–selectionist debate, molecular clock, Lewontin's paradox, cost of selection, McDonald–Kreitman test, codon usage bias, mutation bias, mutation walk.

CONTENTS

* Author for correspondence (Tel.: +49 69 7542 1842; E-mail: menno.de-jong@senckenberg.de).

I. INTRODUCTION

In the 1960s, analysis of molecular data sets generated with the newly developed techniques of protein sequencing (Zuckerkandl & Pauling, 1965; de Chadarevian, 1999) and gel electrophoresis (Harris, 1966; Hubby & Lewontin, 1966; Lewontin & Hubby, 1966) led to the foundation of the 'neutral mutation–random drift theory' (Kimura & Ohta, 1971*c*), now better known as the 'neutral theory of molecular evolution' (Dietrich, 1994). Half a century later and exabytes of sequencing data further, the debate between its proponents and opponents continues unabated (Kern & Hahn, 2018; Jensen *et al.*, 2019).

The neutral theory of molecular evolution is a set of ideas built around a single hypothesis (Fig. 1), known in full as the 'neutral mutation–random drift hypothesis of molecular evolution and polymorphism' (Kimura, 1976). Apart from this main hypothesis, the neutral theory comprises a number of arguments thought to support the hypothesis (Figs 2 and 3). These arguments are of the hypothetico-deductive type: predictions derived from the neutral mutation–drift hypothesis are compared to empirical findings. The neutralist–selectionist debate questions the validity of this shell of supporting arguments (Figs 2 and 3). The debate has many aspects to it, but the recurring question is: are empirical observations consistent with the neutral mutation–drift hypothesis, or instead with the opposing selectionist hypothesis?

In this review, we argue that the neutralist–selectionist debate has been hampered by the co-existence of different

Fig. 1. The (nearly) neutral hypotheses. (A) Schematic overview of the nearly neutral hypotheses and competing hypotheses. Colours represent relative proportions of five categories: deleterious (red), slightly deleterious (light-red), neutral (grey), positively selected alleles (green) and alleles under balancing selection (orange). (B) Different approaches to estimate the importance of positive selection yield vastly different results. α, proportion of substitutions resulting from positive selection; β, proportion of polymorphisms resulting from balancing selection; d, uncorrected genetic distance; f_0, proportion of mutations that are selectively neutral (functional constraint); He, heterozygosity per site; N_e, effective population size; q, proportion of deleterious mutations that reach fixation; t, time to most recent common ancestor; μ, mutation rate.

interpretations of the meaning and implications of the neutral mutation–drift hypothesis. We highlight four aspects of the hypothesis that are subject to multiple interpretations (Fig. 4), namely: the extent to which the hypothesis acknowledges natural selection (Section II); the measurement unit in which the hypothesis is defined (Section III); the definition of selective neutrality (Section IV); and the functional importance of neutral molecular changes, along with its implications for phenotypic divergence (cladogenesis, Section V) and evolutionary optimisation (anagenesis, Section VI). Lastly, we discuss how the ambiguous nature of the neutral hypothesis has complicated the process of deducing and evaluating testable predictions (Sections VII–IX).

Validation or falsification of the neutral hypothesis cannot occur in a meaningful manner if the proposition under scrutiny is ambiguous and open to different interpretations (Fig. 1B). This review is meant as an objective, impartial treatise which aims to contribute to a consensus interpretation of the neutral hypothesis and thereby to facilitate a constructive neutralist–selectionist debate. Our review is not meant to advocate a dogmatic stance. On the contrary, we emphasise that in essence neutralists and selectionists agree on many aspects, except for on the frequency of positive selection events (Fig. 5). Estimating this frequency is still as relevant today as it was when the neutral theory was founded.

For brevity, throughout this review we will refer to the neutral mutation–random drift hypothesis as simply the neutral hypothesis. We will use this as an umbrella term, which includes three versions of the neutral hypothesis (i.e. the 'original', 'revised' and 'nearly' neutral hypothesis), but which, crucially, excludes the 'strict neutrality' hypothesis (Fig. 1A).

II. TO WHAT EXTENT DOES THE NEUTRAL HYPOTHESIS ACKNOWLEDGE NATURAL SELECTION?

(1) A misnomer: the neutral hypothesis does not deny selection

The controversy between neutralists and selectionists concerns the question of what proportions of mutations, polymorphisms and substitutions are adaptive (Fig. 1). In this context, and throughout this review, mutations are defined as *de novo* alleles that have been introduced by recent mutation events. Polymorphisms are defined as alleles with a certain arbitrary minimum allele frequency (e.g. MAF > 0.05) and substitutions as alleles that have reached fixation (Kimura, 1991a).

The neutral hypothesis states that most mutations are not adaptive but instead neutral or deleterious, and that most

The neutral theory of molecular evolution
and the neutralist–selectionist debate

Box (definitions):
- d: uncorrected genetic distance
- F_{ST}: $(d - He)/d$
- f_0: proportion of neutral mutations
- He: heterozygosity
- m: migration rate
- N_e: effective population size
- q: proportion of deleterious mutations that drift to fixation
- r: recombination rate
- μ: mutation rate per site per generation
- t_g, t_y: time to most recent common ancestor, in number of generations (t_g) or years (t_y)

Central rings: INTERPRETATION | OBSERVATION | PREDICTION TESTED / DEDUCED | HYPOTHESIS

Center — neutral hypothesis of protein evolution: most amino acid mutations are neutral or deleterious; most substitutions are neutral

Equations:
$$He \propto r \quad (1)$$
$$d \propto He \quad (2)$$
$$He = 4N_e\mu f_0$$
$$d = 2\mu t_g f_0$$
$$He \propto N_e \quad (6)$$
$$F_{ST} \propto \frac{t_g}{mN_e} \quad (5)$$
$$d \propto f_0 \quad (3)$$
$$d \propto t_g \quad (4)$$

Radial labels: Genetic distances: WITHIN GENES / BETWEEN GENES; Factors explaining variation: ACROSS LOCI / ACROSS GENES

1. *Begun & Aquadro (1992):* genomic regions with high recombination rates (and thus less indirect selection) have higher genetic diversity

1. Polymorphism levels *vs* recombination rates
Hill & Robertson (1966), Birky & Walsh (1988): a locus linked to another locus under selection experiences a reduction of N_e.
Begun & Aquadro (1992): positive correlation between polymorphism levels and recombination rates suggests linked positive selection.
Charlesworth et al. (1993): observed correlation could also be caused by linked negative selection ('background selection').

2. *Skibinski & Ward (1982):* protein He correlates with genetic distance (d)

2. Polymorphism *vs* divergence levels
Skibinski & Ward (1982): at He = 0, d > 0, while neutral hypothesis predicts a zero intercept.
Chakraborty & Hedrick (1983); Ward & Skibinsky (1985): non-zero intercept because relationship between d and He is non-linear.
Hudson et al. (1987): predicted correlation not observed for adh-locus and the 5' flanking sequence of two related *Drosophila* species.

3. *Miyata & Yasunaga (1981), Li et al (1981), Hendriks et al. (1987):* pseudogenes have high genetic distances

3. Functionally constraint loci evolve slower
King & Jukes (1969), Dickerson (1971), Kimura & Ohta (1974): observed variation in amino acid substitution rates across loci reflects differences in functional constraints.
Li et al. (1981): pseudogenes have highest substitution rates.
Epstein (1967), Grantham (1974), Graur (1985): functional constraint (f_0) depends on protein stability (i.e. amino acid composition), not on functional importance.
Hurst & Smith (1999): rates do not differ between essential and non-essential genes.

4. *Margoliash (1963), Zuckerkandl & Pauling (1965), Dickerson (1971):* corrected genetic distance is proportional to divergence time in years (d ∝ t_y)

4. Molecular clock (substitution rate constancy)
King & Jukes (1969): observed rough substitution rate constancy is unexpected if most substitutions are selective.
Lewontin (1974), Gillespie (1984a, 1986): long-term rate averages obscure episodes of selection.
Langley & Fitch (1974): rates significantly deviate from constancy.
Kimura & Ohta (1974): neutral theory predicts rough, not strict, rate constancy.
Ohta & Kimura (1971c): molecular clock is in years (d ∝ t_y), not generations (d ∝ t_g).
Ohta & Kimura (1971c), Ohta (1972a): molecular clock is in years because $\mu \propto 1/N_e$ and $q \propto 1/N_e$.

5. *Lewontin & Hubby (1966):* many proteins with low F_{ST}

5. Allele frequency differences
Prakash et al. (1969), Maynard-Smith (1970b), Ayala & Gilpin (1974): similarity in allele frequencies across populations (low F_{ST}) suggests balancing selection.
Kimura & Ohta (1971c): similarity is caused by gene flow
Li & Nei (1977): genetic drift is weak in large populations, and hence allele frequencies may remain similar to ancestral population for a long time.
Lewontin & Krakauer (1973), Ayala & Gilpin (1974): F_{ST} variation across loci cannot be explained by demography (drift or gene flow).
Chakraborty et al. (1978): F_{ST} variation across loci agrees with neutral expectations.

6. *Lewontin (1974):* similar He across species

6. Lewontin's Paradox
Maynard-Smith (1970a): the observed He constancy would imply N_e constancy across species, which clearly cannot be true.
Kimura & Ohta (1971c): N_e and mutation rate per generation are inversely related (μ ∝ $1/N_e$), and hence the product $N_e\mu$ is roughly constant.
Maynard-Smith & Haigh (1974): He constancy can only partly be attributed to linked selection.
Nei et al. (1975): heterozygosity levels need a long time to recover from population bottlenecks.
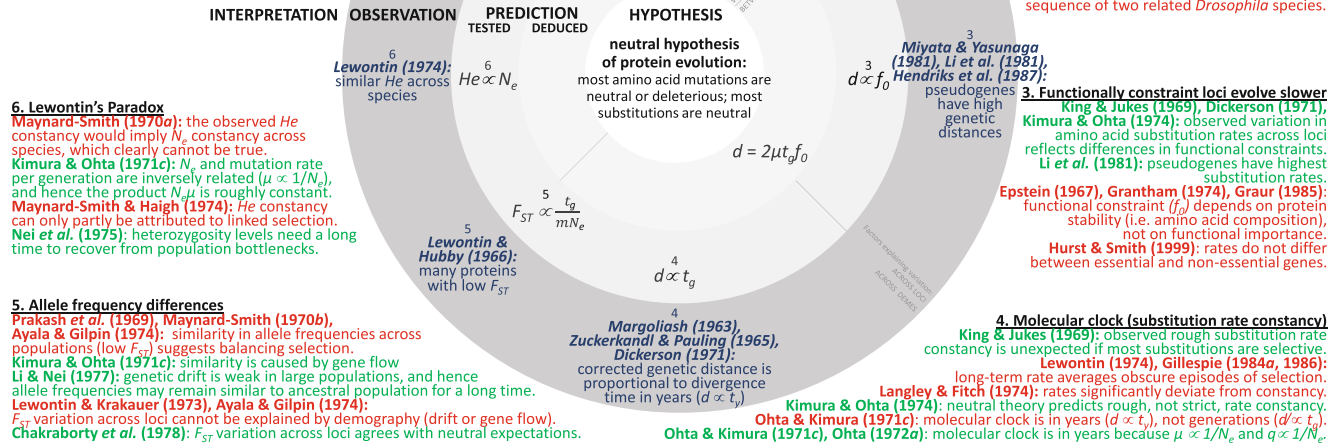
Fig. 2. Schematic overview of the neutral theory and the neutralist–selectionist debate. Green text presents arguments in favour of the neutral hypothesis, red text lists opposing views. The arguments shown relate to the neutral hypothesis of protein evolution (i.e. proportion of adaptive *amino acid* substitutions). The figure does not include arguments relating to the neutral hypothesis of DNA evolution (i.e. proportion of adaptive *nucleotide* substitutions), such as the prevalence of silent and non-coding substitutions. The validity of arguments is evaluated with respect to the revised neutral hypothesis, which acknowledges purifying selection.

polymorphisms and substitutions are neutral, resulting from segregation and fixation of (nearly) neutral mutations through random genetic drift (Fig. 1) (Crow, 1972a; Hughes, 2008; Hahn, 2008; Razeto-Barry, Díaz & Vásquez, 2012; Jensen et al., 2019). A selectionist, on the other hand, believes that most polymorphisms and substitutions are adaptive, and result from balancing selection and positive selection acting on adaptive mutations (Fig. 1). Kimura (1977, p. 275), focusing on substitutions, wrote: '*According to the neutral mutation–random drift hypothesis of molecular evolution and polymorphism, most mutant substitutions detected through comparative studies of homologous proteins (and the nucleotide sequences) are the result of random fixation of selectively neutral or nearly neutral mutations. This is in sharp contrast to the orthodox neo-Darwinian view that practically all mutant substitutions occurring within species in the course of evolution are caused by positive Darwinian selection.*'

The difference between the neutral hypothesis and the opposing selectionist view is subtle. Although selectionists and neutralists disagree on the importance of positive selection and balancing selection, they agree on the importance of purifying selection (Hughes, 2008). This consensus was reached soon after Kimura (1968a) first proposed the neutral hypothesis, following an almost immediate and '*important revision*' (Kimura & Ohta, 1971c, p. 469) which was first suggested by King & Jukes (1969) and subsequently adopted by Kimura & Ohta (1971c), and which acknowledged that many mutations are deleterious. In the words of Kimura (1976, p. 248): '*I must emphasise that there is no fundamental disagreement between the neutralists and the selectionists with respect to the prevalence of negative selection; both agree that deleterious mutations are numerous and these mutants tend to be eliminated by negative selection*'. Furthermore, the neutral hypothesis does not deny the occurrence of adaptive substitutions or polymorphisms, but instead poses they are relatively rare. Kimura & Ohta (1974, p. 2851) stated: '*Adaptive changes due to positive Darwinian selection do no doubt occur at the molecular level, but we believe that definitely advantageous mutant substitutions are a minority when compared with a relatively large number of "non-Darwinian" type mutant substitutions, that is, fixations of mutant alleles in the population through the process of random drift of gene frequency*'.

Neutralists even admit that many or even most of the alleles that are effectively neutral (i.e. randomly drifting in the population), may not be truly neutral, but instead have a small fitness effect. The nearly neutral hypothesis, which represents a second revision of the neutral hypothesis (Ohta & Kimura, 1971c; Ohta, 1972a,c, 1973, 2003; Akashi, Osada & Ohta, 2012), was introduced to formally acknowledge the implications, in particular that small populations are expected to accumulate not only neutral but also slightly deleterious substitutions (Fig. 1).

Kimura (1976) argued that the nearly neutral theory accentuated the difference between neutralists and selectionists. He
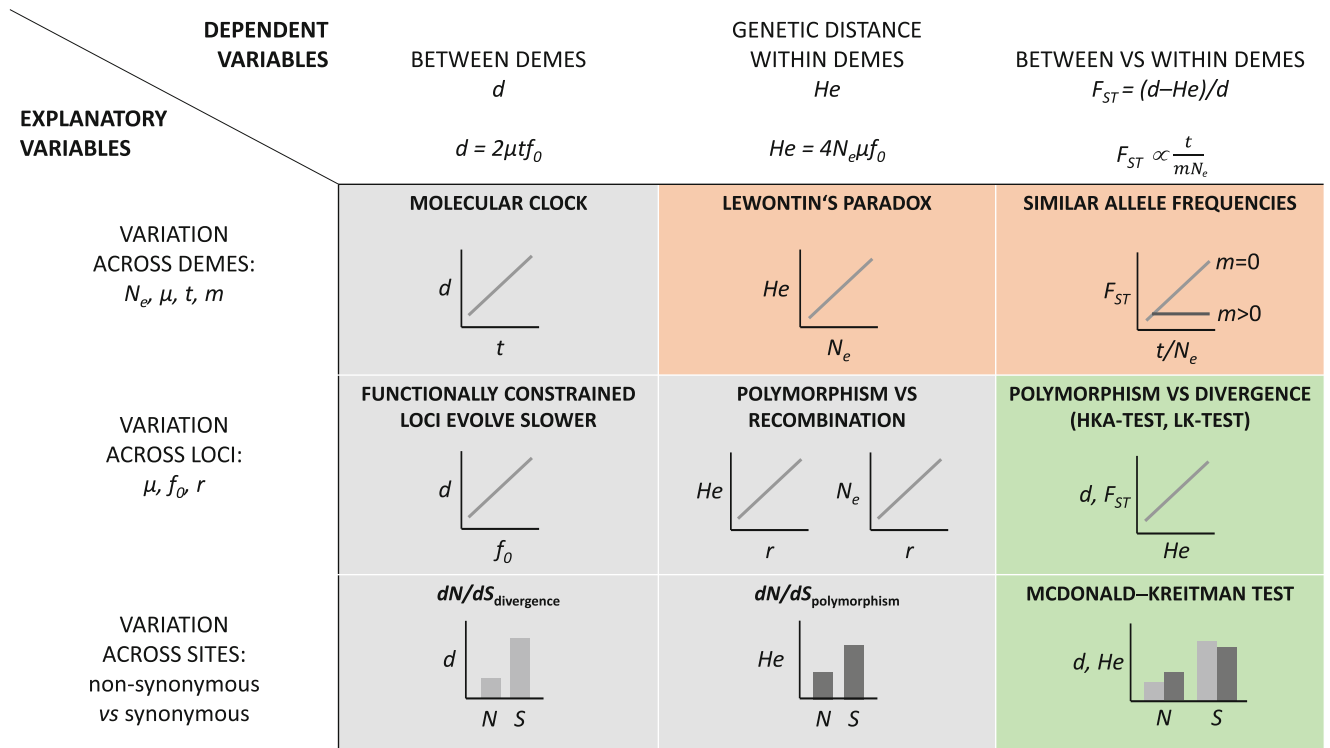
Fig. 3. Systematic classification of neutral theory predictions, and associated neutralist–selectionist arguments. The neutral theory assumes that sequence dissimilarity within and between populations – genetic distance (d) and heterozygosity (He) respectively – depend on the proportion of neutral mutations (functional constraint, f_0), migration rate (m), effective population size (N_e), recombination rate (r), split time (t), mutation rate (μ), and the functional importance of a nucleotide site [e.g. synonymous (S) *versus* non-synonymous (N)]. Colour coding indicates usefulness of the tested predictions. Red: tested prediction is not useful because it does not necessarily follow from the neutral hypothesis. Grey: tested prediction is indiscriminate, meaning it may (possibly) also be deduced from the selectionist hypothesis. (For considerations relating to the molecular clock argument, see Section IV.2). Green: correct, discriminative prediction. In general, discriminative power increases when both dependent variables (d and He) are examined in combination. Here, levels of population differentiation (F_{ST}) are defined as (d–He)/d. Not depicted: interaction between the explanatory factors (e.g. N_e and r *versus* functional importance of a site) may explain additional variation, such as variation of the ratio between non-synonymous (N) and synonymous (S) substitutions ($dN/dS_{\text{divergence}}$) and polymorphisms ($dN/dS_{\text{polymorphism}}$) across demes and loci.

wrote (p. 249): '*If we adopt it, we deviate still further from the "selectionist" camp. In fact, if the neutral theory is "non-Darwinism," Ohta's theory amounts to "anti-Darwinian" to the extent that random fixation of very slightly deleterious mutations occurs frequently in evolution*'. But if so, the second revision only marginally undid the concession of the first revision. By conceding the importance of purifying selection, neutralists had moved a considerable distance towards the selectionist standpoint (Lewontin &
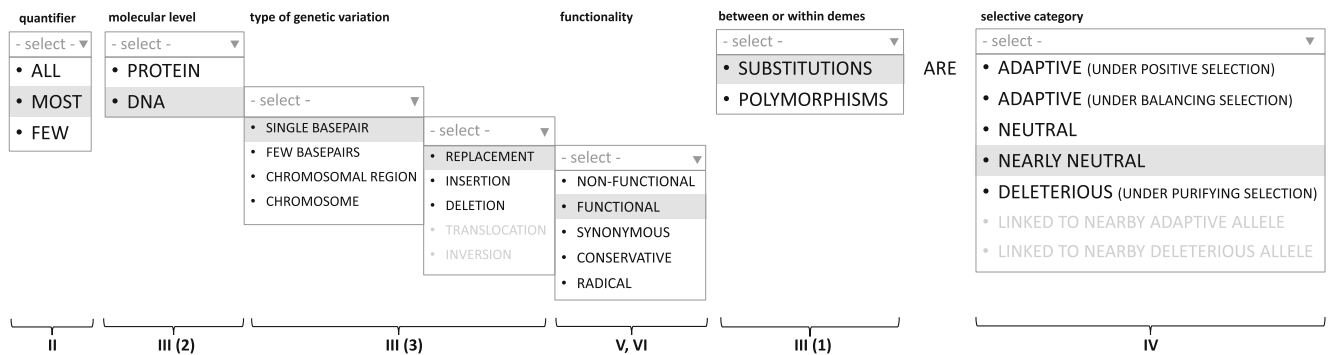
Fig. 4. The neutral hypothesis can be interpreted in various ways. Grey shading indicates one possible interpretation, namely: 'Most substitutions involving single-base replacements in functional DNA are nearly neutral'. The roman numbers denote the sections of this review in which each potential source of ambiguity is discussed.

Are you a neutralist or a selectionist?

Neutral hypothesis of genomic DNA evolution
A large proportion of the genome (at least for species with limited reproductive capacities) is non-coding and non-regulatory. For many species the functional fraction is smaller than the non-functional fraction, meaning that most nucleotide substitutions across the genome result from random fixation of neutral mutations.

Background selection reduces genetic variation in genomic regions which flank coding and regulatory regions.

Neutral hypothesis of functional DNA evolution
Due to the redundancy of the genetic code, many substitutions within protein-coding DNA are silent and therefore nearly neutral (though not completely neutral). Hence, within protein-coding DNA, many substitutions result from random fixation of nearly neutral mutations at silent sites.

A proportion of replacement substitutions are conservative: they affect the primary structure of proteins (i.e. amino acid sequence) but not higher order structures, and hence barely alter protein function.

Most mutations within protein-coding and regulatory DNA are deleterious. In the case $|s|>>1/(2N_e)$, they are effectively removed from the population by purifying selection. In the case $|s|<<1/(2N_e)$, genetic drift could occassionaly cause fixation. Some mutations within protein-coding and regulatory DNA are beneficial. These mutations are favoured by natural selection and responsible for the adaptation of organisms to their environment.

TRUE / FALSE

Neutral hypothesis of protein evolution
The vast majority of *amino acid substitutions in proteins* result from random fixation of (nearly) neutral mutations.

INCOMPATIBLE WITH NEUTRALIST MODEL

INCOMPATIBLE WITH SELECTIONIST MODEL

TRUE / FALSE

Neutral hypothesis of phenotypic evolution
Genetic drift not only causes fixation of silent and conservative mutations, but also of non-conservative mutations which alter protein function and hence alter phenotype.

TRUE / FALSE

EXTENDED NEUTRALIST
Observed differences between species are predominantly caused by the **majority** of **nearly neutral** substitutions.

NARROW NEUTRALIST
Observed differences between species are predominantly caused by the **minority** of **adaptive** substitutions.

SELECTIONIST
Observed differences between species are predominantly caused by the **majority** of **adaptive** substitutions.

Fig. 5. Flowchart depicting (dis)agreements between selectionists and neutralists. Selectionists and neutralists agree on many aspects of molecular evolution, except for on the question whether most substitutions and polymorphisms in proteins and functional DNA are caused by genetic drift or instead by positive and balancing selection. N_e denotes effective population size, and s denotes the selection coefficient.

Hubby, 1966; Lanks & Kitchin, 1970; Clarke, 1970*a,b*). From the perspective of a selectionist, acknowledging that even randomly drifting alleles are not truly neutral was just another step of neutralists accepting the ubiquity of natural selection (Kimura, 1976).

Whereas the outcome of the second revision became known as the 'nearly neutral theory' or the 'slightly deleterious theory', no new names were introduced during the first revision. The labels 'neutral theory' and 'neutralists' persisted, even though they had become misnomers.

Lewontin (1974, p. 197–198) wrote: '*The suggestion that most, if not all, of the molecular variation in natural populations is selectively neutral has unfortunately led to the widespread use of the terms "neutral mutation theory" and "neutralists" to describe the theory and its proponents [...]. But these rubrics put the emphasis in just the wrong place [...]. It is not claimed that nearly all mutations are neutral or that evolution proceeds without natural selection, chiefly by the random fixation of neutral mutations. [...] On the contrary, the claim is that many mutations are subject to natural selection, but these are almost exclusively removed from the population. [...] In addition, the theory allows for the rare favourable mutation, which will be fixed by natural selection, since after all adaptive evolution does occur. But it supposes this event to be uncommon*'.

(2) Testing the neutralist strawman of strict neutrality

None of the three versions of the neutral hypothesis (original, revised, and nearly neutral) excludes selection events entirely (Fig. 1). This is in sharp contrast to a fourth hypothesis called 'strict neutrality', which assumes complete absence of selective forces (Fig. 1). While most researchers will be well aware of this distinction, it is nevertheless frequently overlooked when interpreting empirical data.

Because the strict neutrality hypothesis assumes stochastic processes only, it can be translated into a statistical null model, and therefore offers practical convenience for testing purposes (Kreitman, 1996; Hey, 1999). However, as with all null models, interpretations need to be in the context of the simplifying assumptions made. Clearly, rejection of strict neutrality predictions does not equate to rejection of the revised neutral or nearly neutral hypothesis, and not even to rejection of the original neutral hypothesis. The neutral hypothesis is, like the selectionist hypothesis, an alternative hypothesis which opposes the null hypothesis of strict neutrality. While rejection of strict neutrality predictions is indicative of selection, it does not reveal the nature of the selection pressures (purifying or positive), and hence it cannot serve to decide between the neutral hypothesis and the selectionist hypothesis. As commented by Kimura & Ohta (1974, p. 2851): '*The existence of selective constraint, often inferred from non-randomness in amino acid or nucleotide sequences, does not contradict the neutral mutation–random drift hypothesis*'. And in the words of Lewontin (1974, p. 198): '*The neoclassical theory cannot be refuted by erecting a neutralist strawman and refuting that*'. Even so, many popular arguments against the neutral hypothesis do just that: they refute strict neutrality, not one of the versions of the neutral hypothesis (see Section VIII).

(3) The neutral hypothesis acknowledges purifying selection, and thus background selection

The neutralist–selectionist debate originally concerned the proportion of adaptive amino acid substitutions, but the debate gradually shifted to the proportion of nucleotide substitutions, first within genes, later also outside genes. The arrival of next generation sequencing data stretched the neutral theory to its limits. Even though the effects of linked selection were investigated concurrently with the foundation of the neutral theory (Hill & Robertson, 1966; Maynard-Smith & Haigh, 1974), predictions from the neutral hypothesis were originally derived for individual loci in isolation from the rest of the genome (Charlesworth & Charlesworth, 2018). In the words of Hey (1999, p. 36): '*The theory that had been constructed on the basis of allelic data was not up to the task of revealing the role of natural selection in shaping the pattern of this new kind of variation*'.

Against this historical backdrop, it is perhaps no surprise that challenging the neutral hypothesis based on falsification of strict neutrality predictions has become especially commonplace in the study of genome-wide genetic variation. Evidence for linked selection (i.e. allele frequency changes due to linkage to a nearby allele under selection) is routinely referred to as evidence against the neutral hypothesis, even if the nature (positive or purifying) of the selection events has not been established. For example, it has been stated that the neutral hypothesis is violated by the finding that '*many loci reveal nonneutral local patterns of variation*' (Hey, 1999, p. 37), by the finding that '*natural selection plays a dominant role in shaping patterns of neutral molecular variation in the genome*' (Corbett-Detig, Hartl & Sackton, 2015, p. 13) and by the finding that '*at almost every locus, there has recently been a selected allele nearby (whether advantageous or deleterious)*' (Kern & Hahn, 2018, p. 1368).

This interpretation of genome-wide variation assumes that the neutral hypothesis denies linked selection. For example, Hahn (2008, p. 257) writes: '*The linked selection claim of the neutral theory is that linked selection does not affect a vast majority of loci, and therefore that variation in nature reflects the predictions of neutral models*'. Sella *et al.* (2009, p. 2) claim: '*Neutral theory [...] states that the effects of both positive and negative selection at linked loci on the dynamics of neutral alleles can be ignored*'. Similarly, Phung, Huber & Lohmueller (2016, p. 4) claim: '*Selection affecting linked neutral diversity and divergence is at odds with the neutral and nearly neutral theories*'.

These claims are misleading. It follows logically that a hypothesis that advocates the prevalence of purifying selection also advocates the prevalence of background selection (i.e. allele frequency changes due to linkage to a nearby deleterious allele). Thus, neutralists can readily accept background selection as the null model for explaining levels of genome-wide genetic variation without having to modify their main proposition (Booker, Jackson & Keightley, 2017; Comeron, 2017).

(4) Selectionists do not advocate panselectionism

Early papers on the neutral theory challenged the idea of panselectionism. For instance, King & Jukes (1969) explicitly disagreed with Simpson (1964, p. 1537), who wrote that '*it seems highly improbable that proteins, supposedly fully determined by genes, should have non-functional parts*'. King & Jukes (1969, p. 789) argued: '*To hold that selectively neutral iso-alleles cannot occur is equivalent to maintaining that there is one and only one optimal form for every gene at any point in evolutionary time*'.

Kimura ([1979b](#)) also contrasted the neutral hypothesis to molecular panselectionism. He wrote (p. 98): '*The Darwinian theory of evolution through natural selection is firmly established among biologists. […] In this view any mutant allele, or mutated form of a gene, is either adaptive or less adaptive than the allele from which it derived*'. Jukes ([1991](#), p. 484) explicitly stated: '*The real choice is between panselectionism and near-neutrality*'.

However, opposing the neutral theory does not automatically make one a panselectionist (Ayala, [1974](#)). The view that most amino acid substitutions are adaptive is compatible with the view that certain amino acids can be replaced without altering protein function. Therefore, claiming that selectionists believe that '*each amino acid has a unique survival value in the phenotype of the organism*' (Jukes, [1991](#), p. 480) is almost as inaccurate as mistaking the neutral hypothesis for strict neutrality. Just as the neutralist hypothesis cannot be refuted by refuting strict neutrality (Section [VIII](#)), the selectionist hypothesis cannot be refuted by refuting panselectionism. Unfortunately, many observations thought to support the neutral hypothesis are in reality incompatible with panselectionism, but not with the selectionist hypothesis (Section [IX](#)).

(5) Neutrality during stabilising and directional selection

Given that neutralists acknowledge that many mutations are deleterious, it is, in fact, self-evident that neutralists also accept that some mutations are adaptive. By incorporating purifying selection into their theory (King & Jukes, [1969](#); Kimura & Ohta, [1971c](#)), neutralists implicitly acknowledged the occurrence of positive selection events. As pointed out by Goodman ([1981](#), p. 114), whenever purifying selection operates, '*somewhere in the past Darwinian selection must have spread the amino acid substitutions which are now preserved*'. In the words of Gillespie ([1994](#), p. 943): '*Natural populations evolve to the point where most mutations are deleterious through the substitution of advantageous mutations. At such time when all mutationally accessible advantageous alleles are exhausted, all newly arising mutations will be deleterious*'. Wagner ([2008](#), p. 970) similarly remarked: '*The effect of a mutation exists only in the context of the mutations preceding it*'.

Once a gene or protein converged to its optimum configuration, functionally important sites are expected to experience an equal number of slightly deleterious substitutions and slightly beneficial back-substitutions or compensatory substitutions (Fig. [6](#), Table [1](#)) (Charlesworth & Eyre-Walker, [2007](#)), a process which has been referred to as '*selection without adaptation*' (Hartl & Taubes, [1996](#)). In the words of Gillespie ([1994](#), p. 943): '*While the […] majority of borderline mutations are deleterious, of those that fix, precisely half are deleterious and half are advantageous*'. Meanwhile, functionally less-important sites will continue to accumulate neutral substitutions, turning adaptive substitutions into a minority (Table [1](#)).

It has been argued that, like deleterious mutations, also neutral mutations ultimately result from past Darwinian selection. This idea, known as the drift-barrier hypothesis (Sung *et al.*, [2012](#)) or 'selection–mutation–drift'

hypothesis (Bulmer, [1991](#)), holds that a phenotypic trait under Darwinian selection will converge to its optimum, but may never reach it, because the final selective increments are so small that the causative mutations are effectively neutral (Fig. [6](#)). This would be especially true if the relationship between fitness score and trait value is best described by a plateauing function in which trait increments are associated with diminishing fitness returns (Hartl, Dykhuizen & Dean, [1985](#); Akashi *et al.*, [2012](#)). The inevitable outcome of the adaptation process would be near-neutrality of subsequent mutations. Hartl *et al.* ([1985](#), p. 669) wrote: '*We infer from this assumption that natural selection will continue until an enzyme activity is achieved beyond which an additional increase in enzyme activity results in a negligible increase in fitness. At this point the fate of new mutations that have small effects on enzyme activity will be determined principally by the effects of random genetic drift. Thus, we deduce that evolution by means of the random genetic drift of neutral or nearly neutral mutations is not only consistent with the Darwinian theory of natural selection but also that the existence of nearly neutral alleles follows as a consequence of long continued natural selection of an enzyme gene. In this somewhat ironic sense, the greater the selectionist one might be, the greater the neutralist one should become*'.

These considerations do not, however, imply that the neutralist hypothesis is valid only during the phase of stabilising selection, once a phenotypic trait, and there with the underlying genetic architecture, has converged to its optimum (Fig. [6](#)). In reality, and depending on the number of critical sites in a protein or regulatory sequence, it is theoretically very much possible that even during the adaptation process itself (i.e. directional selection phase), most amino acid substitutions are neutral (Table [1](#)).

III. IN WHICH MEASUREMENT UNIT IS THE NEUTRAL HYPOTHESIS DEFINED?

(1) Proportion of adaptive substitutions or adaptive polymorphisms?

The neutral hypothesis makes statements about two evolutionary processes: the maintenance of genetic variation within demes, and the genetic divergence of demes. Prior to the neutral theory, it was believed that these two processes were mainly regulated by different mechanisms, namely balancing selection and positive selection (Dietrich, [1994](#); Ohta, [2003](#)). When viewed in the light of the neutral theory, these two processes reflect to a large extent two sides of the same coin: genetic drift of (nearly) neutral alleles. In the words of Kimura & Ohta ([1971c](#), p. 469): '*Protein polymorphism and molecular evolution are not two separate phenomena, but merely two aspects of a single phenomenon caused by random frequency drift of neutral mutants in finite populations*'.

The potential of the neutral theory to unite two seemingly unrelated processes – genetic divergence and genetic variability – into a single theoretical framework is one of its great appeals (Higgins, [2004](#)). In the words of
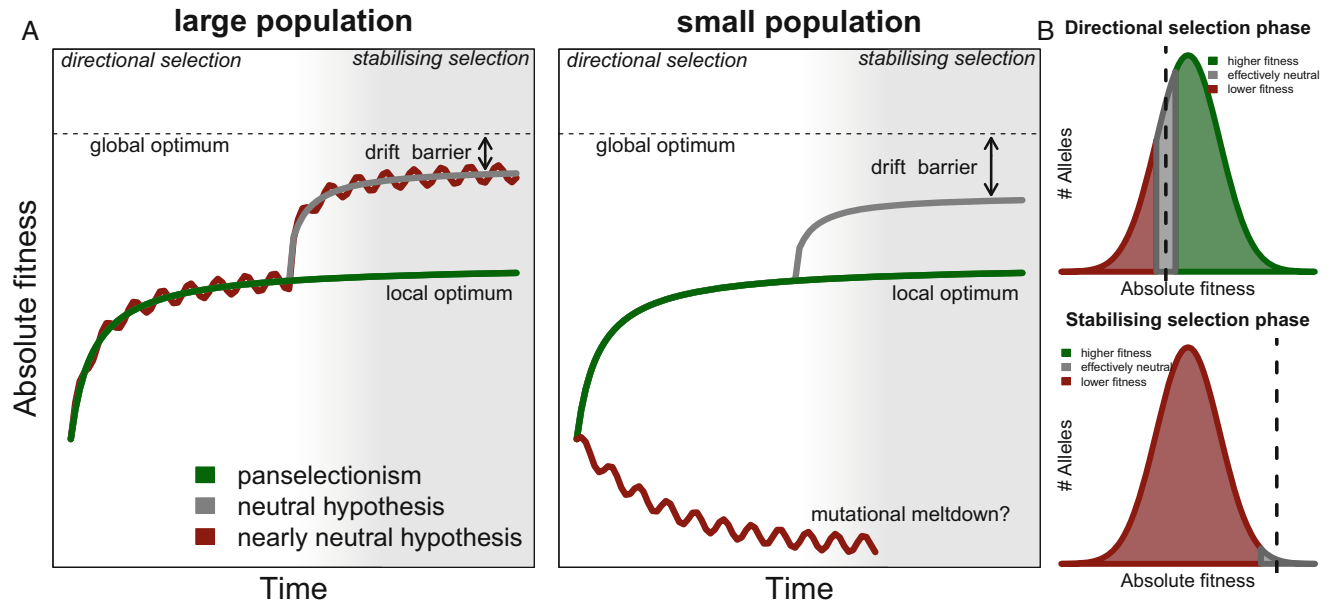
Fig. 6. Evolutionary optimisation. (A) Directional selection on a certain trait (e.g. protein stability or binding specificity) is expected to enhance protein performance over time through the serial replacement of higher-fitness alleles. However, the search through the multi-dimensional sequence space is constrained by the availability of neighbouring alleles which are single-step mutations away. If all mutations have a fitness effect (panselectionism), the protein configuration may become trapped in a local optimum and remain suboptimal. Neutral mutations provide neutral pathways through the sequence space, providing an escape route out of local optima. Because small fitness increments are effectively neutral, a protein can only converge, yet never fully reach, the global optimum. The magnitude of this drift barrier depends on the population size. In theory, the fixation of slightly deleterious alleles (nearly neutral hypothesis) can in small populations initiate a mutational meltdown. Larger populations can avoid this through the fixation of back and/or compensatory mutations, resulting in an equal number of deleterious and advantageous substitutions during the stabilising selection phase. (B) The selective values of all alleles in sequence space (here arbitrarily depicted as being normally distributed) are relative to the allele currently present in the population (dashed line). Once the protein has converged to its optimum state (stabilising selection phase), all potential mutations are either deleterious or effectively neutral.

Table 1. Estimation of alpha. Simplified, hypothetical example of protein evolution, illustrating different methods to estimate the proportion of adaptive substitutions (α, in bold): a per-generation estimate, a cumulative estimate, and a relative estimate (here relative to start sequence). The protein consists of five amino acids, in which the letters S, N, D, and A refer to start, neutral, deleterious and adaptive base respectively. The colours grey, red and green refer to substitution events where the new base is neutral, deleterious or adaptive respectively, assuming a stable environment. Owing to the occurrence of multiple substitutions per site, the relative estimate will overestimate the cumulative proportion of adaptive substitutions. Once the protein has reached its optimal composition/configuration (asterisk), the evolution of the protein reaches a new phase of stabilising selection in which the number of deleterious substitutions at functionally important sites is expected to equal the number of adaptive back-substitutions. Meanwhile, functionally less important sites (site 1) will continue to accumulate neutral substitutions, thereby further decreasing the cumulative proportion of adaptive substitutions. Here, a substitution with a deleterious allele is counted as a neutral event.

GEN.	Site					Per-generation			Cumulative			Relative to generation 0		
	1	2	3	4	5	Neutral	Adaptive	α	Neutral	Adaptive	α	Neutral	Adaptive	α
0	S	S	S	S	S									
1	N	S	S	S	A	1	1	**0.50**	1	1	**0.50**	1	1	**0.50**
2	D	N	N	A	A	3	1	**0.25**	4	2	**0.33**	3	2	**0.40**
3	N	D	A	A	A	2	1	**0.33**	6	3	**0.33**	2	3	**0.60**
4*	S	A	A	A	A	1	1	**0.50**	7	4	**0.36**	1	4	**0.80**
5	N	A	A	A	A	1	0	**0**	8	4	**0.33**	1	4	**0.80**
6	S	A	D	A	A	2	0	**0**	10	4	**0.29**	2	3	**0.60**
7	N	A	A	A	A	1	1	**0.50**	11	5	**0.31**	1	4	**0.80**
8	S	A	A	A	A	1	0	**0**	12	5	**0.29**	1	4	**0.80**

Maynard Smith (1970b, p. 231): '*A hypothesis which kills two birds with one stone [...] is attractive*'. However, it does mean that neutralists and selectionists are fighting a battle on two fronts, involving slightly different claims (Fig. 1).

Selectionists stress the importance of positive selection when it comes to genetic divergence, but the importance of balancing selection when it comes to polymorphisms. Because positively selected mutations reach fixation rapidly, they will constitute a minor proportion of observed polymorphisms, even if they occur frequently. Furthermore, the rapid fixation of adaptive alleles causes reduction of genetic variation, which conflicts with the observed high levels of genetic diversity – the trigger of the neutralist–selectionist debate on polymorphism data (Hubby & Lewontin, 1966; Kimura & Ohta, 1971c; Dietrich, 1994). Mutations under balancing selection, on the other hand, can cause a strong increase in polymorphism levels (Kimura, 1960b; Dietrich, 1994; Roff, 1998).

Thus, the yardstick for the neutralist–selectionist debate on genetic divergence is the proportion of substitutions resulting from positive selection (known as alpha, α) (Booker *et al.*, 2017), whereas the yardstick for the debate on polymorphism data is the proportion of polymorphisms resulting from balancing selection (here denoted as beta, β) – not the proportion of polymorphisms that are under purifying or positive selection.

(2) Does the neutral hypothesis refer to protein evolution or DNA evolution?

The neutral hypothesis was originally proposed to explain patterns in protein data, namely the rough constancy of amino acid substitutions inferred from protein sequence data sets, the great diversity of primary structure of homologous proteins across species, and the unexpectedly high levels of polymorphisms in protein gel electrophoresis data sets (Kimura, 1968a; King & Jukes, 1969; Maynard-Smith, 1970b). One could therefore argue that originally the hypothesis that most substitutions are neutral pertained to amino acid sequences (Gillespie, 1995). This agrees with early formulations of the neutral hypothesis in the literature, such as that '*most of the changes that have occurred in evolution at the level of protein sequences might be meaningless noise rather than adaptive change*' (King, 1983, p. 25), that '*a majority of amino acid substitutions that occurred in these proteins are the result of random fixation of selectively neutral or nearly neutral mutations*' (Kimura, 1969b, p. 1181), that '*a considerable proportion of amino acid substitutions in proteins are selectively neutral*' (Maynard-Smith, 1970b, p. 231), and that '*random drift of neutral mutations in finite populations can account for observed protein polymorphisms*' (Kimura & Ohta, 1971c, p. 467).

On the other hand, throughout the 1960s, and prior to the onset of the neutralist–selectionist debate, Kimura repeatedly discussed genetic diversity at the level of nucleotide sequences (Kimura & Crow, 1964; Kimura, 1968b). He continued this line of reasoning in his papers on the neutral theory, including his landmark paper on Haldane's dilemma (Kimura, 1968a), where he attempted to convert amino acid substitution rates into genome-wide nucleotide substitution rates. In a second paper released in the same year, synonymous mutations played a central role in his argument that many mutations must be neutral (Kimura, 1968b), and they would continue to do so (Kimura, 1977, 1980, 1991a). For instance, Kimura (1991b, p. 374) recalled: '*The first truly favourable evidence for the neutral theory was the finding that synonymous base substitutions, which do not cause amino acid changes, occur almost always at much higher rates than non-synonymous, that is, amino-acid altering substitutions*'.

Thus, the neutral theory of molecular evolution can be considered an umbrella term for two parallel theories: the neutral theory of DNA evolution and the neutral theory of protein evolution (Crow, 1972a). The neutral hypothesis of DNA evolution holds that most *nucleotide* substitutions are fixed by random drift, whereas the neutral hypothesis of protein evolution holds that most *amino acid* substitutions are fixed by random drift (Fig. 5). The neutral hypothesis of DNA evolution can be further subdivided into the neutral hypothesis of genomic DNA evolution and the neutral hypothesis of functional DNA evolution, previously referred to as the 'weak' and 'strong' version, respectively (Hahn, 2018). The former holds that most nucleotide substitutions across the entire genome are fixed by random drift. The latter holds that this claim remains true even when only considering coding and regulatory regions (Fig. 5).

Inevitably, this ambiguous nature of the neutral theory has clouded the neutralist–selectionist debate. This confusion surfaces mainly during discussions of silent sites and non-coding DNA. Substitutions and polymorphism in these regions are relevant to the neutral hypothesis of DNA evolution, but irrelevant when estimating the proportion of adaptive amino acid substitutions (Crow, 1972a). It may be argued that by steering the discussion to the high substitution rates in silent sites and non-coding regions (Kimura, 1991a; Ohta & Gillespie, 1996), neutralists unintentionally committed a red herring fallacy. Hey (1999, p. 36) reflected: '*The finding that nonfunctional sequences evolved fastest and harboured lots of DNA sequence variation was definitely a neutralist prediction, and books written by neutralists in the middle 1980s proclaimed victory to varying degrees. To a neutralist from this time [...], it would not have seemed fair to label these findings a distraction from the debate, but that is what they seemed to selectionists. Even if one did accept that junk DNA could have neutral mutations, nonfunctional mutations in nonfunctional DNA did not bear on the original questions about natural selection on protein variation*'.

The recent debate between Kern & Hahn (2018) and Jensen *et al.* (2019) illustrates that this ambiguity continues to cause apparent disagreements today. Kern & Hahn (2018, p. 1367) write: '*Application of the MK test to data from protein-coding genes has revealed a predominant role for adaptive natural selection*'. In reply, Jensen *et al.* (2019, p. 112) accuse Kern & Hahn (2018) of circular reasoning because the '*inferred frequencies of adaptive substitutions mostly concern only the small fraction of the genome that codes for proteins*'. This disagreement can be resolved by identifying whether arguments are meant to evaluate the neutral hypothesis of protein evolution, the neutral hypothesis of

genomic DNA evolution, or the neutral hypothesis of functional DNA evolution.

In this context it is worth noting that the founders of the neutral theory, despite often considering the entire genome, did not believe that the neutral hypothesis stands or falls by the proportion of mutations in non-functional DNA (Kimura, 1991*a*). Although King & Jukes (1969) did discuss the presence of non-coding DNA, they did so merely as a correction to Kimura's 'cost of selection' calculations (Okazaki *et al.*, 2021). In a subsequent study, Ohta & Kimura (1971*c*, p. 24) wrote: '*We are quite sure that the adaptive gene substitutions constitute a small fraction of the total substitutions. This is true even if we restrict our consideration to informational part of the DNA. Since nucleotide substitutions in the non-informational part must be largely non-adaptive, the fraction of adaptive substitutions becomes still smaller if we consider the total DNA*'. And to quote Kimura (1991*b*, p. 370): '*If the neutral theory is valid, a large fraction of evolutionary nucleotide substitutions occurring at functionally important parts of the genome are also selectively neutral [...]. Thus, neutral evolution is by no means restricted to "junk" part of the genome*'.

The ongoing debate mostly concerns the original question, the verity of the neutral hypothesis of protein evolution. In the words of Crow (1997, p. 262), who used a slightly different terminology: '*The neutral theory can be thought of in weak and strong forms. The weak form asserts that most non-coding and non-regulatory DNA evolves by random drift. The strong form asserts that amino acid changes are also dominated by random drift. I think it is fair to say that the weak form is widely accepted. [...] As for the strong form, the jury is out*'.

Even so, both neutral hypotheses of DNA evolution are not undisputed either. The majority of substitutions within genes may be synonymous, but codon usage bias suggests that at least some synonymous substitutions are adaptive. With regard to the neutral hypothesis of genomic DNA evolution, it is thought that mutational load puts an upper limit of 20–25% on the functional fraction of the human genome, challenging (along with other considerations) claims of the ENCODE project that this proportion could be as high as 80% (Dunham *et al.*, 2012; Eddy, 2012; Graur *et al.*, 2013; Rands *et al.*, 2014; Graur, 2017). But for species with much higher reproductive capacities, mutational load is less of a limiting factor (Elliott & Gregory, 2015), and in many branches of the tree of life the proportion of functional DNA (protein-coding or regulatory) may outweigh the proportion of non-functional DNA (Halligan & Keightley, 2006; Land *et al.*, 2015). For these organisms, the verity of the neutral hypothesis of genomic DNA evolution is not self-evident, and instead depends on the frequency of adaptive and non-adaptive substitutions within functional regions.

(3) Which types of mutations?

When Kimura (1968*a*) compared the empirical substitution rate to Haldane's theoretical upper limit, he expressed both estimates in nucleotide base pairs per year. More specifically, he calculated that '*one nucleotide pair has been substituted in the population roughly every 2 year*' (Kimura, 1968*a*, p. 625). Based on similar considerations, Ohta & Kimura (1971*c*, p. 18) provided what may be regarded a quantitative formulation of the neutral hypothesis, by stating that '*approximately 10% of the amino acid substitutions of average cistrons [i.e. genes] might be adaptive*'. These statements indicate that the claim that most substitutions are neutral, initially referred particularly to single-base substitutions resulting from point mutations.

King & Jukes (1969) also focussed mostly on single-base substitutions, as evidenced by their discussions of synonymous substitutions and mutation rates. However, they did acknowledge that single-base mutations are not the only source of genetic variation, writing (p. 797): '*Protein molecules are subjected to incessant probing as a result of point mutations and other DNA alterations*'.

Kimura (1968*b*, p. 259) briefly mentioned indels as potential source of genetic variation, stating: '*It is now known that mutation is caused by changes in DNA base arrangements, namely, substitutions, gains and losses. Among them, addition or loss of a single base pair causes a shift of reading frame ("frame shift") and will produce far more drastic effects than single-base alterations*'. Nevertheless, his discussion on the 'Nature of mutant alleles' dealt almost exclusively with describing different types of single-base substitutions, not with other types of mutations.

Kimura & Ohta (1974), inspired by the work of Ohno, widened the scope and argued that gene duplication must always precede the emergence of a novel gene. They wrote (p. 2850): '*The existence of two copies of the same gene enables one of the copies to accumulate mutations and to eventually emerge as a new gene, while another copy retains the old function required by the species for survival through the transitional period*'. Kimura (1991*b*, p. 382) commented: '*As a kind of mutation, gene duplication is constantly fed into the population, and many of them have so little deleterious effect that they become fixed in the population through neutral evolution, that is, by random drift under gene duplication pressure*'. He added (p. 382): '*In considering the problem of progressive evolution, we should not forget the possibility that gene duplication and subsequent accumulation of new mutations in the duplicated genes must have played a very important role*'.

This haphazard addition to the neutral theory, also known as 'the neutral theory of gene duplication' (Kondrashov, 2012), opened a new way to estimate the proportion of adaptive substitutions (α). This proportion originally referred to single-base substitutions, a concept which is still widely in use today, for instance implemented in the McDonald–Kreitman test (Eyre-Walker, 2006). But alternatively, the proportion of adaptive substitutions may also be estimated as the proportion of adaptive *versus* neutral gene duplication events (Kondrashov, 2012). While an increasing number of studies have reported evidence for positive selection on higher or lower gene dosage (Kondrashov, 2012; van't Hof *et al.*, 2016; Yan *et al.*, 2023), evidence also exists for random fluctuations of gene copy number. According to Nei (2007, p. 12236), for certain genes it appears that '*after the copy number reached a required level, the number has fluctuated by random duplication and deletion of genes. We may call this event random genomic drift, in analogy with random genetic drift of allele frequencies in population*

genetics'. Ohta made a similar argument with regard to regulatory elements (Ohta, 2002, 2003).

While α estimates are still most reliably obtained from data on single-base substitutions, the most comprehensive picture of the relative contributions of neutral and selective forces on genetic variation is obtained when considering all mutation types (replacements, insertions, deletions, inversions, translocations), ranging in size from single bases to entire chromosomes.

IV. WHEN SHOULD A SUBSTITUTION OR POLYMORPHISM BE CONSIDERED NEUTRAL?

(1) Asking the wrong question: do truly neutral mutations exist?

The neutral hypothesis states that most polymorphisms and substitutions are neutral, but what exactly is meant by 'neutral'? Completely neutral mutations do not alter protein structure, function or regulation in any way and hence do not have a fitness effect (Perutz, 1984; Bowie *et al.*, 1990). Nearly neutral mutations, on the other hand, may affect protein structure and function, but with negligible fitness effects. In the words of Kimura (1991*b*, p. 370): '*By selectively neutral I mean selectively equivalent: namely, mutant forms can do the job equally well in terms of survival and reproduction of the animals possessing them*'.

The co-existence of two definitions of selective neutrality [i.e. completely neutral ($s = 0$) *versus* nearly neutral ($s \longrightarrow 0$), with s denoting the selection coefficient (Table 2A)], is yet another potential source of confusion in the neutralist–selectionist debate (Stoltzfus, 1999). For example, according to Hahn (2008, p. 256), the neutral theory claims that '*the vast majority of polymorphism within species and fixed differences between species have no effect on fitness – that is, there is no direct selection on them*'. By contrast, Jensen *et al.* (2019, p. 111) write: '*The neutral theory of molecular evolution asserts that most de novo mutations […] are under such weak selection that they may become fixed as a result of genetic drift*'. These statements suggest that Hahn (2008)

refers to completely neutral mutations, whereas Jensen *et al.* (2019) refer to nearly neutral mutations.

The ambiguous interpretation of selective neutrality dates back to the foundation of the neutral theory. Kimura (1968*a*, p. 625) explicitly referred to '*neutral or nearly neutral mutations*', and stressed that population-genetic theory predicts that the behaviour of these two mutation types is indiscernible (Kimura, 1968*a*,*b*, 1991*a*). By contrast, King & Jukes (1969) focused their discussion on 'truly neutral' mutations. While Kimura (1968*b*, p. 260) remarked that '*probably not all synonymous mutations are neutral, even if most of them are nearly so*', King & Jukes (1969, p. 789) wrote instead: '*As far as is known, synonymous mutations are truly neutral with respect to natural selection*'. They furthermore explicitly disagreed with Simpson (1964, p. 1537), who had stated: '*The consensus is that completely neutral genes or alleles must be very rare if they exist at all*'.

Owing to the attempt of King & Jukes (1969) to identify truly neutral genetic variants, the neutralist–selectionist debate headed off in the wrong direction. Proponents and opponents got bogged down in discussions on whether silent and conservative mutations are truly neutral or not, and whether iso-alleles do exist. Critics pointed out, rightfully, that the selective neutrality of a mutation is not determined by its effect on protein structure and protein function only, but also by other factors such as translation efficiency, and that therefore it is premature to assume that silent and conservative mutations are truly neutral (Richmond, 1970; Clarke, 1970*a*).

These points were all valid, but the relevant question was – and is – not if and how many alleles have a fitness effect, but rather how many alleles have a fitness effect large enough for selection to outforce stochastic factors controlling the fate of alleles (i.e. genetic drift) (Crow, 1972*a*). Theory predicts that this depends on the effective population size (N_e), which is why population geneticists assume that '*neutrality depends not only on* s *but also on* N_e'. (Kimura, 1968*b*, p. 262). In the words of Kreitman (1996, p. 680): '*It is not so much a question of whether strictly neutral mutations exist – probably no mutation is absolutely neutral – but whether the strength of selection for or against that mutation is much smaller or greater than the strength of genetic drift*'.

Table 2. (A) Three different definitions of neutrality. N_e denotes effective population size, s denotes the selection coefficient, and z denotes an arbitrary threshold value, typically assumed to range between 0 and 2. (B) The neutral theory of molecular evolution and the Darwinian theory of evolution by natural selection are not directly comparable because they apply to different levels of biological organisation.

A		Unconditional	Conditional on N_e
	$|s| = 0$	Completely neutral	–
	$|s| \longrightarrow 0$	Nearly neutral	Effectively neutral ($|s| < z/N_e$)
B		Molecular evolution	Phenotypic evolution
	Neutral theory	Few substitutions are caused by positive selection	??
	Darwinian theory	??	Phenotypic differences are caused by positive selection

The importance of N_e cuts both ways. As stressed by Lanks & Kitchin (1970, p. 754), in large populations even the smallest selection coefficients can make a difference: '*An arbitrarily small selective advantage or disadvantage may be undetectable in a single individual but quite apparent in a sufficiently large population. […] It is certainly not valid to maintain that a mutated protein is equivalent to the wildtype on the basis of such gross evaluation as enzyme activity assays or the absence of clinical symptoms*'. But the opposite is equally true: in small populations strong fitness effects can be overruled by genetic drift. In the words of Kimura & Ohta (1974, p. 2851): '*Difference in function at the molecular level does not necessarily lead to effective natural selection at the levels of individuals within a population*'.

Confusingly, in spite of explicit statements that neutrality depends on s and N_e (Kimura, 1968b), and in spite of explicit references to 'neutral and nearly neutral mutations' Kimura (1968a, p. 625), neutral theory predictions have been derived assuming neutral mutations only (see Section VII). This inconsistency has been corrected in the nearly neutral theory, but persists to date in the neutral theory. In the words of Kimura (1991b, p. 377): '*The neutral theory assumes that the mutations can be classified into two distinct groups, namely, the completely neutral class (with the fraction* f_0*), and the definitely deleterious class (fraction* $1-f_0$*)*'.

Kreitman (1996) used this criterion to distinguish between the 'completely neutral theory' (i.e. original and revised neutral hypothesis) and the 'slightly deleterious model' (i.e. nearly neutral hypothesis) – not to be confused with the distinction between the neutral hypothesis and strict neutrality (Fig. 1). He wrote (p. 678): '*The issue is whether the neutral theory applies to those models in which there is absolutely no selection (completely neutral mutations only), or to models in which genetic drift dominates the fate of a mutation (effectively neutral mutations). The distinction is not trivial, as this slight semantic change in the theory has been shown to have severe consequences on the predicted patterns of genetic variation and the substitution rate*'.

(2) Should hitchhiking alleles be considered neutral?

Traditionally, the proportion of adaptive substitutions and polymorphisms has been calculated without considering genetic hitchhiking (Ohta & Kimura, 1971a; Crow, 1972b). However, Kern & Hahn (2018, p. 1367) ask the question: '*How much of the genome is directly or indirectly influenced by adaptive natural selection?*'. The phrase 'directly or indirectly' suggests that in their opinion hitchhiking alleles should be factored into the equation. This is also how the following statement of Wagner (2008, p. 965) could be interpreted: '*According to selectionism, beneficial mutations are abundant: most mutations that go to fixation in a population would be beneficial, or are at least linked to abundantly occurring beneficial mutations*'. Clearly, including hitchhiking sites in the estimate of α will lead to very different conclusions, especially regarding the validity of the neutral hypothesis of genomic DNA evolution.

Although the question posed by Kern & Hahn (2018) is indeed relevant for other theoretical questions and practical applications, such as demographic inference (Gillespie,

2001; Pouyet *et al.*, 2018; Harris, 2018; Johri *et al.*, 2021; Buffalo, 2021), we argue that when evaluating the neutral hypothesis, the proportion of adaptive substitutions should be evaluated considering direct selection only. This is historically more correct, since the statement that most polymorphisms and substitutions are selectively neutral has traditionally always been about direct selection only. It is also theoretically more correct, because the effects of linked selection (also known as 'genetic draft') mimic the effects of genetic drift: linked selection decreases the fixation probabilities of adaptive alleles, while increasing the fixation probabilities of deleterious alleles (Hill & Robertson, 1966; Birky & Walsh, 1988; Gillespie, 2001; Jensen *et al.*, 2019). Thus, classifying hitchhiking alleles as alleles under positive selection would produce a biased, inflated estimate of α.

It has been estimated that up to 85% of the human genome is affected by direct or linked natural selection (Pouyet *et al.*, 2018). Harris (2018, p. 2) commented: '*Superficially, this might seem like a death knell for the neutral theory, but it is nothing of the kind*'. We agree. The implications of such a value for the neutral hypothesis depend on two factors: (*i*) what number/proportion of amino acid (or nucleotide) substitutions under direct selection is responsible for indirect genome-wide effects; and (*ii*) how many of these substitutions result from positive selection (causing genetic hitchhiking) and how many from negative selection (causing background selection)? As long as these two factors are unknown, the implications for the neutral hypothesis cannot be determined.

(3) Delimiting nearly neutral mutations

Neutralists consider mutations neutral if their '*selection coefficients are so small that their behaviour may not be very different from the strictly neutral mutants*' (Ohta & Kimura, 1971c, p. 21). This begs the question: is there an objective threshold? Or, in the words of Crow (1972b, p. 307): '*How similar in fitness must two genes be to be regarded as selectively equivalent?*'

Kimura (1968b, p. 262) posited: '*A mutant gene may be called almost neutral if* $|2N_e \cdot s|$ *is much smaller than unity*'. As pointed out by Kreitman (1996), Kimura did not apply a consistent definition, using the phrases 'much smaller' and 'less than', as well as the symbols '$\ll$' and '$<$', interchangeably. Ohta (1976, p. 258) used a different threshold, writing: '*Generally, when the population size is small, very slightly deleterious alleles become effectively neutral (*$N_e s \ll 1$*) and the strict neutral theory is practically valid, whereas when the population size gets large such that* $N_e s \gg 1$*, these alleles are selected against*'. Ohta (2002, p. 16134) used another threshold still: '*The nearly neutral mutations are defined such that their fate in the population depends on both selection and drift. Thus, the absolute value of the product,* $N_e s$*, should be small, e.g. not larger than 2*'. In yet another study, Kimura & Ohta (1971b) assumed that mutations are selectively nearly neutral when their absolute selection coefficient is below $1/(4N_e)$ (Kimura & Ohta, 1971a; Ohta & Kimura, 1971c; Hartl *et al.*, 1985; Wagner, 2008).

Apart from the inconsistency, there are two problems with any of these proposed ranges meant to delimit nearly neutral mutations. First, the set limits are arbitrary and in need of justification. In previous studies, Kimura derived mathematically that the fixation probability (u) of a novel mutation in a diploid population is given by the formula $u = (1 - e^{-2s})/(1 - e^{-4N_e s})$, which in the case of small selection coefficient (s) simplifies to: $u = 2 s/(1 - e^{-4N_e s})$ (Kimura, 1957, 1962). Kimura (1962, p. 716) commented: '*For a positive* s *and very large* N$_e$, *we obtain the known result that the probability of ultimate survival of an advantageous mutant gene is approximately twice the selection coefficient. On the other hand, if we let* s → 0, *we obtain* u = 1/(2N$_e$), *the result known for a neutral gene*'. However, it remains arbitrary at what point fixation probabilities differ sufficiently from neutral mutations to no longer be considered nearly neutral. Kimura (1968a, p. 625) stated: '*In the special case of* |2Ns| ≪ 1 [...], *the probability of fixation* [...] *is roughly equal to its initial frequency*'. But without an objective criterion, the same could be said to be true for any of the other suggested thresholds.

Second, the question may be asked how biologically meaningful these suggested narrow ranges are. Theory predicts that the average time to fixation for a nearly neutral mutation is close to $4N_e$ generations (Kimura & Ohta, 1969). Given the spatiotemporal variation of selective pressures (Wallace, 1975; Reznick, 2016), fluctuations of the selection coefficient during these prolonged periods are likely to be of more importance to the fate of the allele than the negligible mean value itself (Ohta & Kimura, 1971c; Ohta, 1972b; Nei, 2005b). Fisher (1931, p. 220) commented: '*The neutral zone of selective advantage in the neighbourhood of zero is so narrow that changes in the environment, and in the genetic constitution of species, must cause this zone to be crossed and perhaps recrossed relatively rapidly in the course of evolutionary change, so that many possible gene substitutions may have a fluctuating history of advance and regression before the final balance of selective advantage is determined*'. For this reason, near-neutrality is better defined as $N_e \sigma < 1$ (instead of $|N_e s| < 1$), in which σ denotes the standard deviation of the selection coefficient (Ohta, 1972b; Ohta & Tachida, 1990).

Defining the limits of near-neutrality is not a mere semantic or technical issue. As we will discuss next, the implications of the neutral theory for phenotypic evolution depend critically on the selective values (i.e. functional consequences) of the molecular changes assumed to be governed by random genetic drift.

V. WHAT ARE THE NEUTRALITY PREDICTIONS FOR PHENOTYPIC EVOLUTION?

(1) The nature of neutral mutations

The neutral hypothesis raises a fundamental question for evolutionary biologists: if most differences between species at the molecular level result from genetic drift, then what are the implications for evolution at the phenotypic level?

Does it mean that most observed phenotypic differences within and between species are, likewise, neutral?

The neutral theory of molecular evolution applies primarily to the genotype (including protein sequences) while the Darwinian theory of evolution by natural selection applies primarily to the phenotype (Table 2B). To assess whether the neutral theory and Darwinian theory are compatible or in conflict, their main propositions need to be translated into the same unit of measurement. That is, either Darwinian theory must be expressed in terms of genotypic change, or alternatively the neutral theory must be expressed in terms of phenotypic change. For the latter approach, one first needs to know the exact nature (or functional importance) of the molecular changes which neutralists consider to be neutral.

King & Jukes (1969) argued that protein-coding DNA contains two types of substitutions which do not affect higher protein structure and function and hence can be considered neutral: synonymous (also known as silent) and conservative substitutions. They wrote (p. 797) that '*most proteins contain regions where substitutions of many amino acids can be made, without producing appreciable changes in protein function*'. In line with this reasoning, Kimura (1991b, p. 380) wrote: '*A replacement of an amino acid within a protein often preserves the activity of that protein unaltered*'. Jukes (1991, p. 477) similarly remarked: '*Enzymes have a small active center containing a few key amino acids, and the rest of the molecule supplies bulk, size and other general properties that are more or less nonspecific. This larger region can accommodate many changes without losing its structure and function*'.

But do silent and conservative substitutions represent all that is at stake in the neutralist–selectionist debate? If the neutral hypothesis is taken to mean that many substitutions within protein-coding DNA are silent and many others conservative, or that certain nucleotides within regulatory sequences can be replaced without affecting protein regulation, one might understand the opinion of one of the reviewers of the manuscript of King & Jukes (1969), according to which the '*idea was obviously true and therefore trivial*' (King, 1983, p. 25). No biologist – selectionist or neutralist alike – will disagree with the statement that synonymous sites accumulate more substitutions than non-synonymous sites, or that alteration of amino acid sequence does not necessarily imply a change of protein function.

If neutrality is meant to refer not only to completely neutral mutations but also to effectively neutral mutations, it is evident that the neutral hypothesis is not just about the prevalence of silent and conservative polymorphism and substitutions. It is also about the prevalence of non-conservative (i.e. radical) polymorphisms and substitutions – changes which *do* affect protein function and/or regulation (Stoltzfus, 1999). This, in turn, has implications for phenotypic evolution, both for cladogenesis (discussed in this section) and for anagenesis (see Section VI).

(2) Ecological phenotypic divergence

In theory, the neutral hypothesis leaves open the possibility that all or most of the phenotypic differences observed

between species are caused by a minority of adaptive substitutions (Ayala, 1974; Perutz, 1984; Komiyama *et al.*, 1995; Nei, 2005*b*; Zhang, 2018). This decoupling of genomic and phenotypic differences was emphasised by King & Jukes (1969). From their introductory remarks it is evident that although they argued that most nucleotide substitutions are neutral, they believed that most species differences at the phenotypic level are adaptive. They wrote (p. 788): '*Evolutionary change at the morphological, functional and behavioural levels results from the process of natural selection, operating though adaptive changes in DNA. It does not necessarily follow that all or most evolutionary change in DNA is due to the action of Darwinian natural selection. There appears to be considerable latitude at the molecular level for random genetic changes that have no effect upon the fitness of the organism*'. Kimura (1976, p. 249) also emphasised this decoupling between genomic and phenotypic evolution: '*When we come to the phenotypic level far removed from the molecular constitution of genes, I have little doubt that Darwin's principle of natural selection will prevail in the evolutionary changes of form and function. What I am claiming is that, deep down at the level of the internal structure of genetic material, there is a great deal of evolutionary change propelled by random genetic drift*'.

The belief that molecular evolution is mostly neutral whereas phenotypic evolution is mostly adaptive rests on the assumption that the neutral hypothesis refers solely, or at least predominantly, to the occurrence of non-coding, silent and conservative substitutions. Protein sequences and DNA sequences contain a great deal of redundant information, which means that many nucleotides and even amino acids can be replaced by random genetic drift without affecting protein function or regulation, and hence without substantially affecting phenotype (Perutz, 1984; Bowie *et al.*, 1990; Komiyama *et al.*, 1995; Bloom & Arnold, 2009).

If we accept this view, the neutral theory of molecular evolution is complementary, or 'supplementary' (Kimura, 1976), to Darwin's concept of ecological speciation. Darwinian theory describes evolutionary change at the phenotypic level, whereas the neutral theory of molecular evolution describes meaningless evolutionary change ('noise') at the molecular level (Ayala, 1974). In the words of Kimura (1979*b*, p. 126): '*Darwinian, or positive, selection cares little how [...] phenotypes are determined by genotypes. [...] Even if Darwin's principle of natural selection prevails in determining evolution at the phenotypic level, down at the level of the internal structure of the genetic material a great deal of evolutionary change is propelled by random drift*'. Kimura (1983, p. ix) also supported the complementary perspective, and wrote: '*The neutral theory is not antagonistic to the cherished view that evolution of form and function is guided by Darwinian selection, but it brings out another facet of the evolutionary process by emphasizing the much greater role of mutation pressure and random drift at the molecular level*'.

(3) Neutral phenotypic divergence

An alternative interpretation is that the neutral theory of molecular evolution opposes the concept of ecological divergence, at least to a certain extent, namely by implying the existence of phenotypic differences between species that result from random genetic drift instead of positive selection. This interpretation posits that the random fixation of non-conservative substitutions has far-reaching consequences for phenotypic divergence. Once we admit that genetic drift can cause differential fixation or loss of alleles that alter protein function and regulation, and thus cause functional changes, there is no way back from admitting that disconnected populations can diverge without adaptation, through the accumulation of non-conservative substitutions resulting from random genetic drift. Thus, we are left with having to accept that at least part of the phenotypic differences observed between species can result from genetic drift rather than from natural selection (Lande, 1976; Lynch & Hill, 1986; Stoltzfus, 1999). If we take a final additional step, by assuming that the random fixation of non-conservative mutations in disconnected populations can cause reproductive barriers, then the neutral theory can even be regarded to advocate the reality of neutral allopatric speciation, as an alternative to ecological speciation (Hedges *et al.*, 2015).

Kimura contemplated this possibility in his last papers on the subject. Kimura (1991*a*, p. 5972): '*How can we understand evolution at two levels – that is, molecular and phenotypic – in a unified way? It is generally believed that, in contrast to the neutralist view of molecular evolution, evolutionary changes at the phenotypic level are almost exclusively adaptive and caused by Darwinian positive selection. However, I think that even at the phenotypic level, there must be many changes that are so nearly neutral that random drift plays a significant role, particularly with respect to quantitative characters*'. Kimura (1991*a*, p. 5972) concluded: '*If the neutral theory is valid so that a great majority of evolutionary changes at the molecular level are controlled by random genetic drift under continued input of mutations, it is likely that selectively neutral changes have played an important role [...] in phenotypic evolution*'.

Based on the above discussion, we propose here that a distinction can be made between two types of the neutral hypothesis: a narrow and extended version (Fig. 5). The narrow neutral hypothesis holds that even though most substitutions within proteins, genes and regulatory factors result from random genetic drift, these are solely or predominantly silent or conservative substitutions that do affect protein function and hence do not affect the phenotype (and have no fitness consequences either). The extended neutral hypothesis, by contrast, which may be regarded a '*generalisation of the neutral theory to the phenotypic level*' (Lynch & Hill, 1986, p. 915), holds that genetic drift can cause fixation of nearly neutral, non-conservative mutations (which do affect protein function and regulation and hence do affect phenotype), implying that accumulation of these types of substitutions can cause between-species phenotypic differences.

(4) Testing for neutral phenotypic divergence

If the accumulation of nearly neutral substitutions indeed causes phenotypic divergence (as posited by the extended neutral hypothesis), then levels of molecular and phenotypic evolution are expected to correlate: both would depend on

the time to most recent common ancestor (Khaitovich *et al.*, 2004). In terms of variability of divergence rates across traits, the neutral hypothesis predicts lower phenotypic divergence for functionally important (i.e. conserved) traits compared to less important traits (Ho, Ohya & Zhang, 2017). If, instead, phenotypic divergence is caused solely by a few (large-effect) selected mutations (as posited by the narrow neutral hypothesis) (Nei, 2007), we expect to find no correlation between levels of molecular and phenotypic divergence (Kimura & Ohta, 1971*a*).

Zuckerland & Pauling (1965, p. 148) leaned towards the last view, stating: '*There is no reason to expect that the extent of functional change in a polypeptide chain is proportional to the number of amino acid substitutions in the chain. Many such substitutions may lead to relatively little functional change, whereas at other times the replacement of one single amino acid residue by another may lead to a radical functional change*'. Kimura (1969*b*, p. 1187) similarly commented: '*If amino acid changes are often due to chance, then these should be established as frequently in evolutionary conservative species as in those that undergo rapid changes in morphology. […] It would support the hypothesis of this paper if haemoglobins and other proteins show the same rate of amino acid substitution in […] living fossils as in rapidly evolving species*'.

Under strict neutrality conditions (i.e. when the genetic variance underlying a phenotypic quantitative trait is determined solely by neutral mutations and genetic drift), levels of phenotypic variation within and between lineages are predictable, conditional on a hypothetical distribution of fitness effects (DFE) (Lynch & Hill, 1986; Lynch, 1988; Hodgins-Davis, Rice & Townsend, 2015). The actual variation observed both within and between lineages, as for instance measured in terms of gene expression levels, is typically found to be smaller than predicted by the strictly neutral model of phenotypic evolution (Lemos *et al.*, 2005; Denver *et al.*, 2005; Hodgins-Davis *et al.*, 2015). As discussed in Section II, such violations of strict neutrality predictions may result from purifying selection, and hence do not necessarily refute the neutral hypothesis.

VI. WHAT ARE THE NEUTRALITY PREDICTIONS FOR FITNESS LEVELS?

(1) Neutral mutations may facilitate adaptation

Despite the designation 'non-Darwinian evolution' (King & Jukes, 1969), the founders of the neutral theory did not deny the importance of Darwinian selection for evolutionary innovations. In the words of Kimura & Ohta (1974, p. 2851): '*There is not a slightest doubt that the marvellous adaptations of all the living forms to their environment have been brought about by positive Darwinian selection*'. And in the words of Lewontin (1974, p. 199): '*The neoclassical theory [i.e. the neutral theory] cannot be disposed of by pointing to the elephant's trunk and the camel's hump. The theory does not deny adaptive evolution but only that the vast quantity of molecular variation within populations and, consequently, much of* the molecular evolution among species, has anything to do with that adaptive process'.

For this reason, it has been argued that the neutral hypothesis may just as well as be ignored by evolutionary biologists interested in phenotypic evolution. In the words of Crow (1972*a*, p. 2): '*A biologist may well say that if these changes are so nearly neutral as to be governed by chance […] they are not really of much interest. He is more interested in processes that affect the organism's ability to survive and reproduce, and which have brought about such exquisite adaptations to diverse environments*'.

However, even though neutral mutations do not affect fitness directly, this does not mean they cannot affect the adaptation process indirectly. Theoretical considerations suggest that random mutations may indirectly facilitate, or even enable, the adaptation process. This relevance of neutral mutations is generally accepted in the context of environmental change. According to this 'adaptive potential' hypothesis, the genetic variation accumulated by the random drift of neutral mutations may become adaptive following an environmental change, allowing populations to react promptly to new environmental demands (Haldane, 1957; Kimura, 1960*a*; Teixeira & Huber, 2021).

Moreover, the indirect importance of neutral mutations for the adaptation process is not limited to environmental change scenarios alone – it also applies to populations occurring in a stable environment, without selection coefficient fluctuations. The reasoning is that in the absence of neutral changes (i.e. if all mutations have a certain fitness effect) populations are likely to become trapped in local optima (Fig. 6). Neutral mutations, on the other hand, allow populations to wander the fitness landscape freely (Wright, 1932; Crow, 1972*b*; Kauffman & Levin, 1987; Gavrilets, 1997). As an analogy, the countless conceivable series of neutral mutations can be thought of as a dense network of raised mutational pathways crisscrossing the fitness landscape and connecting high-fitness regions (Huynen, 1996; van Nimwegen & Crutchfield, 2000; Poelwijk *et al.*, 2007; Lenormand, Roze & Rousset, 2009). In the words of Sewall Wright: '*Changes in wholly non-functional parts of the molecule would be the most frequent ones but would be unimportant, unless they occasionally give a basis for later changes which improve function in the species in question which would then become established by selection*'. (Huynen, 1996, p. 165).

This conceptual idea of (near-)neutral substitutions facilitating the evolutionary process is supported by protein network modelling and protein engineering studies. These studies suggest that evolution by means of natural selection may be severely impeded if most amino acid substitutions have a strong phenotypic effect, because this would prohibit an explorative 'random walk' through the hyperdimensional space of all possible protein sequences (Maynard-Smith, 1970*a*; Gillespie, 1984*a*; Lipman & Wilbur, 1991; Wagner, 2008; Bloom & Arnold, 2009). It also suggests that neutrality dictates the pace and even the direction of evolution, as the search through genotype space is determined by the available neutral paths (i.e. series of single-step mutations involving amino acids with similar physicochemical

properties) (van Nimwegen & Crutchfield, 2000). A random walk may accidently hit upon an entry into an adaptive pathway, the evolutionary starting point for the next adaptation event (Bloom & Arnold, 2009).

Others have gone even a step further, by arguing that complexity can also arise solely through random fixation of neutral mutations, without Darwinian selection (Stoltzfus, 1999; Muñoz-Gómez *et al.*, 2021). This hypothesis, known as 'constructive neutral evolution', posits that '*a novel attribute appears initially as an excess capacity and later becomes a contributor to fitness, due to a neutral change at some other locus that creates a dependency on it*' (Stoltzfus, 1999, p. 176).

Kimura (1979*b*) initially argued that the scientific value of his theory should not be weighted by the biological relevance of neutral mutations. He wrote (p. 126): '*People have told me, directly and indirectly, that the neutral theory is not important biologically because neutral genes are not involved in adaptation. My own view is that what is important is to find the truth, and that if the neutral theory is a valid investigative hypothesis, then to establish the theory, test it against the data and defend it is a worthwhile scientific enterprise*'. Later, however, Kimura (1991*b*) argued that the neutral variation accumulated through random genetic drift may provide the raw material needed for adaptation following an environmental change. Kimura (1991*b*, p. 383) stated: '*No one would be able to say then, that neutral changes are by definition not concerned with adaptation, and that therefore the neutral theory is biologically not very important*'.

(2) Can slightly deleterious mutations cause degeneration?

The random fixation of neutral mutations allows populations to traverse between high-fitness peaks in a cost-free way. The fixation of slightly deleterious mutations may also reduce the possibility of remaining trapped in local optima, but this is a riskier strategy. Before accidently reaching the base of a higher adaptive peak, populations first have to descend into a fitness valley. In the words of Lande (1976, p. 320): '*Genetic drift can thus be thought of as a process of random exploration of the adaptive zones in a temporarily maladaptive way, on the chance that a new phenotype may be found which will be better adapted*'.

In theory, a population risks never finding the way up again and to enter a 'slippery slope'. Continuous accumulation of slightly deleterious mutations could draw populations into a vicious circle known as the 'extinction vortex' or 'mutational meltdown'. A positive feedback loop between population size and the probability of fixation of deleterious mutations could see the population heading for extinction (Lynch, Conery & Burger, 1995). In the words of Kimura and Ohta (1974, p. 2851): '*Accumulation of very slightly deleterious mutations by random drift is essentially equivalent to the deterioration of environment, and definitely adaptive gene substitutions must occur from time to time to save the species from extinction*'.

In reality, empirical evidence for mutational meltdowns in sexually reproducing populations is sparse (Whitlock, Griswold & Peters, 2003; Teixeira & Huber, 2021). This may simply indicate that mutational meltdowns are usually

averted by compensatory mutations (Lande, 1998; Poon & Otto, 2000; Whitlock *et al.*, 2003), but some authors regard this discrepancy as potential evidence against the nearly neutral hypothesis. For instance, Crow (1997, p. 262) commented: '*An unappealing aspect of the nearly neutral theory is its implication that slightly deleterious mutations accumulate with a consequent (very slow) deterioration of the population*'. Nei (2005*b*, p. 2325) agreed: '*A problem with Ohta's theory is that if deleterious mutations accumulate in a gene, the gene gradually deteriorates and eventually loses its function. If this event occurs in many important genes, the population or species would become extinct*'.

(3) Genetic load is not an optimality measure

Fitness differences are typically quantified in terms of the genetic load, which denotes the mean difference in fitness in a population relative to the most fit genotype present at a given time (Muller, 1950; Kimura, 1960*a*; Brues, 1964, 1969; Graur, 2017; Bertorelle *et al.*, 2022).

The genetic load consists of several components (Brues, 1969; Bertorelle *et al.*, 2022), but its definition was originally formulated to measure one particular type of genetic load, the 'mutational load' (Muller, 1950). This load results from a mutation-purifying selection balance in which '*selection tends to eliminate alternative alleles but mutation restores them*' (Lewontin & Hubby, 1966, p. 606). Because neutralists and selectionists both acknowledge that many mutations are deleterious, they agree that populations carry such a mutation load. However, they disagree about other components of the genetic load, in particular drift load, segregation load and substitutional load.

Considering all components, the predictions of neutralists imply a lower genetic load than the predictions of selectionists do. This seemingly contradictory outcome is due to the definition of genetic load, which, following the classical definition, does not measure the difference between the observed genotypes and the theoretical optimal genotype, but instead the difference between the observed genotypes and the fittest genotype present in the population (Brues, 1964). This definition leads to counterintuitive estimates in certain evolutionary contexts (Van Valen, 1963; Brues, 1964, 1969).

For instance, whenever an adaptive mutation occurs, a genetic load (i.e. 'substitutional load') is created – even though the new allele causes individuals in the population to be better adapted. The substitutional load peaks immediately after the mutation event and disappears only when the beneficial allele has spread throughout the entire population (Van Valen, 1963; Brues, 1964; Kimura, 1968*a*). In the words of Brues (1969, p. 1135): '*The genetic load involved in the substitutional situation is in fact an artifact [...] if we adhere to the definition of fitness in terms of the optimum genotype. The appearance of a new advantageous gene in even the smallest numbers creates a new optimum genotype in relation to which the formerly optimum genotype is demoted, accused of contributing a large amount of load, and blamed for a loss of population fitness. Actually, the appearance and multiplication of a new and advantageous gene leads to an increase in population fitness*'.

The above considerations explain why the selectionist's model, in which most polymorphisms and substitutions result from balancing and positive selection respectively, may imply a higher genetic load than the neutralist hypothesis. Indeed, one of the original arguments underlying the foundation of the neutral theory, the 'cost of selection' argument, was that the selectionist's model would predict an intolerably high genetic load and therefore cannot be true (Kimura & Crow, 1964; Hubby & Lewontin, 1966; Dietrich, 1994).

(4) Reasons underlying suboptimality

Genetic load quantifies the fitness differences between observed genotypes, but more relevant for the neutralist–selectionist debate is the question how observed genotypes, and the phenotypic traits they underlie, compare to the theoretical optimum (i.e. maximised fitness) (Fig. 6). This theoretical optimum reflects a trade-off at which, given all abiotic and biotic environmental conditions, the total costs are minimised and, hence, the net fitness effect maximised. This trade-off is evident at various levels of biological organisation. For instance, the optimum limb morphology of an aquatic mammal reflects a trade-off between mobility in water and mobility on land; while the optimum sprint speed of a cheetah reflects a trade-off between hunting success rate and its ability to defend a catch against other predators. On the molecular level, the optimum protein configuration reflects a trade-off between protein functionality, stability, production and regulation, while the optimum codon usage represents a trade-off between factors such as translation efficiency, metabolic costs and messenger RNA (mRNA) stability (Hey, 1999; Akashi & Gojobori, 2002; Scott et al., 2012; Kessler & Dean, 2014).

The complexity of these trade-offs makes it difficult to determine whether observed trait values are close to optimality. At any point in time the realised genotype or trait value might be suboptimal for a number of reasons. One obvious explanation is that the trait is still evolving (under directional selection) and will reach its optimum if given enough time. Another possible reason, relating to the phase of stabilising selection, is practical limitations: even though, in theory, higher sprint speeds will increase a cheetah's hunting success rate, these predators may have already reached the limit of what is physically attainable for quadrupedal locomotion. A third possible explanation, also relating to the phase of stabilising selection, is the drift-barrier hypothesis (Sung et al., 2012): the cheetah's sprint speed may have converged so close to its optimum value that the effects of further adaptive mutations are too small to overcome the effects of genetic drift (Fig. 6). Once this point has been reached, genetic drift will prevent the spread of slightly beneficial alleles (which would reduce the distance between achieved and optimum value), and promote the fixation of slightly deleterious alleles (which move the trait further from its optimum value) (Fig. 6).

(5) The (sub)optimality of the mutation rate

The subject of trait suboptimality, and its underlying reasons, plays a role in the discussion on mutation rate heterogeneity across lineages. It has long been recognised that the mutation rate (μ) is in itself a phenotypic trait under the control of natural selection (Sturtevant, 1937; Hubby & Lewontin, 1966; Kimura, 1967; Crow, 1972a; Lynch & Conery, 2003; Graur, 2017). Mutation rates are known to vary across lineages, causing substitution rate variation (Gojobori, Moriyama & Kimura, 1990; Kimura, 1991a; Sung et al., 2012). There are currently three main competing hypotheses to explain mutation rate variation: the 'cost of fidelity' hypothesis, the 'physical limit' hypothesis, and the 'drift-barrier' hypothesis (Kimura, 1967; Sniegowski et al., 2000; Martincorena & Luscombe, 2013; Lynch et al., 2016).

The 'cost of fidelity' hypothesis states that mutation rates are at their optimum values. This optimum value reflects the trade-off between the costs and benefits associated with increasing or decreasing the mutation rate. The cost of lowering the mutation rate is not thought to be the loss of adaptive potential (because natural selection only cares about the here and now, not about the future), but the time and energy needed to maintain this fidelity. Kimura (1967, p. 31) states: '*An elaborate apparatus that must be developed for checking and eliminating errors in replication might be physiologically so costly relative to the gain thereby achieved that it did not pay in adaptive evolution*'. Thus, the 'cost of fidelity' hypothesis holds that genomic mutation rates will tend to evolve towards to the equilibrium point with minimum total cost of the mutation rate (which is the sum of the cost of fidelity and the cost of deleterious mutations).

The remaining two hypotheses hold that mutation rates in nature may be suboptimal, but assume different causes. The 'physical limit' hypothesis states that selection only strives to lower the mutation rates (i.e. zero is the optimal value), but that the mutation rate gets stuck at a suboptimal value above zero, namely at the lower limit that is practically attainable. Kimura (1967, p. 31) suggested that '*mutation as the replication error of the genetic material cannot entirely be eliminated because of physical or physiological limitations*'.

The 'drift-barrier' hypothesis (Lynch, 2010; Sung et al., 2012) suggests that mutation rate may be suboptimal due to genetic drift. It asserts that the lower limit is not determined by practical feasibility, but instead by the efficacy of natural selection given the effective population size (i.e. the magnitude of genetic drift). In the words of Lynch et al. (2016, p. 712): '*Selection typically operates to minimise the mutation rate, with the efficiency of such downward movement being eventually overcome by the power of genetic drift. […] The drift barrier is typically encountered before any insurmountable biophysical or biochemical limits to replication fidelity*'. This hypothesis is consistent with the empirical finding that species-specific mutation rates are negatively correlated with effective population size (Lynch, 2010; Sung et al., 2012).

It is evident that the drift-barrier hypothesis fits into the scheme of the nearly neutral hypothesis: namely, by

assuming that the suboptimal (above zero) mutation rates result from repeated random fixation of slightly deleterious mutations, in this particular case specifically in DNA polymerase and repair genes which compromise the fidelity of these enzymes. But what about the other two hypotheses, which assert that mutation rates are kept at their practical or even theoretical optimum? While these hypotheses, if true, might seem more compatible with the selectionist view than with the neutral hypothesis, there are two reasons why this conclusion could be premature.

First, and as discussed above, even proteins that have been fine-tuned to their optimal configuration (in this case DNA polymerase) are not necessarily at odds with the neutral hypothesis. It may still be the case that the majority of amino acid substitutions during the evolution of a protein did not (substantially) alter the protein function. Once the protein has reached its optimal configuration, the protein is kept in this optimal state by the removal of deleterious alleles through purifying selection – a process which is compatible with both neutralist and selectionist views.

Second, while the performance of DNA polymerase affects global mutation rates across the entire genome, these proteins make up a small subset of the entire proteome. Thus, even if most substitutions in these particular proteins during a considerable time span and in a wide range of organisms have been adaptive, the neutral hypothesis does not stand or fall with these proteins alone. The proportion of adaptive substitutions should be considered across all proteins, or at least a substantial proportion of all proteins. Similar considerations apply to the adjustment of local mutation rates through any other hypothetical mechanisms, such as modification of epigenomic features in genomic regions enriched with functionally constrained genes (Monroe *et al.*, 2022).

(6) The (sub)optimality of codon usage

The frequency of triplets underlying amino acids deviates significantly from neutral expectations, even after correcting for GC content (Sharp & Li, 1986; Li *et al.*, 2015). This finding indicates that synonymous mutations are not completely neutral but instead subject to weak selection (Richmond, 1970; Clarke, 1970*a*; Ikemura, 1981; Akashi, 1995) – contrary to the claim of King & Jukes (1969), but as correctly predicted by Kimura (1968*b*) and Kimura & Ohta (1974). As a naïve prediction, one could postulate that the ultimate outcome of this adaptation process is a genome in which each amino acid is encoded for by the preferred codon exclusively. In reality, relative synonym codon usage (RSCU) values are mostly below 1.5, indicating that the preferred codon is only slightly more frequent than expected under strict neutrality.

One possible explanation is that the observed codon usage is in fact (nearly) optimal, for instance because the preferred codon may differ across genes and across sites depending on certain unknown factors. An alternative explanation is that codon usage is suboptimal, and that '*codon usage patterns result from the balance in a finite population between selection favouring an optimal codon for each amino acid, and mutation together with drift*

allowing the persistence of nonoptimal codons' (Bulmer, 1991, p. 897). This hypothesis is consistent with the finding that codon usage bias is less pronounced in genes with low expression levels, which is indicative of a drift barrier, where the weaker purifying selection on less-used genes is more frequently overruled by genetic drift (Bulmer, 1991; Akashi *et al.*, 2012).

Eventually, what matters for the neutralist–selectionist debate are the proportions of adaptive, neutral and deleterious substitutions at degenerate sites. While codon usage bias is indicative of pervasive selection, it does not necessarily imply that most substitutions at degenerate sites are or have been adaptive (see Table 1).

(7) The (sub)optimality of the genetic code

King & Jukes (1969) pointed out that the variation in the relative frequencies of each amino acid type is largely explained by the redundancy of the genetic code (King & Jukes, 1969; Gilis *et al.*, 2001). For instance, the amino acid serine, which is encoded for by six DNA codons (TCT, TCC, TCA, TCG, AGT, AGC), is about three times more abundant than the amino acid tyrosine, which is encoded for by two DNA codons (TAT, TAC) (King & Jukes, 1969). To this, selectionists replied that the genetic code is a product of natural selection itself, moulded to provide the most codons for the amino acids in most demand (Richmond, 1970; Clarke, 1970*a*; Kimura & Ohta, 1974; Xia & Li, 1998; Gilis *et al.*, 2001). However, even if this hypothesis is correct, the question of how the universal genetic code evolved during the origin of life and whether it reached an optimal configuration is very different from the question of which evolutionary processes have shaped the primary structure of proteins in vertebrates many millions of years later. Evaluating amino acid type frequencies, in the context of the available genetic code, and how these frequencies affect substitution rates (Graur, 1985), provides insight into the latter question without challenging the claim that the genetic code itself is adaptive.

VII. IS THE NEUTRAL HYPOTHESIS FALSIFIABLE?

(1) Strict neutrality predictions

The neutral theory has been praised for its usefulness as null model, and neutralists and selectionists agree that as such, regardless of its validity, it has greatly benefited the field of molecular biology (Ayala, 1974; Higgins, 2004). Kimura (1979*b*, p. 126) wrote: '*Because our theory is quantitative it is testable and therefore much more susceptible to refutation when it is wrong than are selectionist theories, which can invoke special kinds of selection to fit special circumstances and usually fail to make quantitative predictions*'. Ayala (1974, p. 694) commented: '*The neutrality theory is a hypothesis with rich empirical content, according to Popper's criterion of falsifiability. It makes precise predictions about the nature and*

pattern of protein polymorphisms in populations and of protein differences between populations'. Maynard-Smith (1978, p. 37) wrote: 'The neutral hypothesis is a good "Popperian" one; if it is false, it should be possible to show it'. Kreitman (1996, p. 678) famously stated: 'The neutral theory is dead. Long live the neutral theory!' He argued (p. 678): 'It has provided empiricists with a strong set of testable predictions and hence, a useful null hypothesis against which to test for the presence of selection'. Crow (1997, p. 262) wrote: 'It has had great heuristic value. Being concrete, mathematical and simple, it leads to testable predictions, and this has been one of its greatest merits'.

However, while there can be little doubt that the neutral hypothesis is more parsimonious than the selectionist hypothesis, and while the above comments might be mostly true for strict neutrality, they are somewhat flattering for the neutral hypotheses, in particular the nearly neutral hypothesis.

In the complete absence of selective pressures, coalescent theory predicts that the genetic distance (d) between two haplotypes is given by: $d = 2\mu T$, in which μ and T denote the mutation rate and time to coalescence respectively. For two haplotypes drawn from two distantly related species, the expected coalescence time roughly equals the divergence time (t), implying: $d = 2\mu t$ (Figs 2 and 3) (Gillespie, 2001). For two haplotypes drawn from the same population, the expected coalescence time equals $2N_e$, and thus $d = 2\mu(2N_e) = 4N_e\mu$. In the case of panmixia, the mean genetic distance between any two haplotypes within a population (also known as nucleotide diversity, π) equals heterozygosity, and hence: $He = 4N_e\mu$ (Figs 2 and 3). Kimura, who used a different approach, concluded: $He = (4N_e\mu)/(1 + 4N_e\mu)$, which for realistic values of N_e and μ returns very similar estimates (Kimura & Crow, 1964; Kimura, 1969a, 1991b; Tajima, 1996).

One difficulty is that the exact values of the parameters t, μ and N_e are typically unknown (Maynard-Smith, 1978). Therefore, neutrality predictions can often only be evaluated globally – $d \propto t$ instead of $d = 2\mu t$, and $He \propto N_e$ instead of $He = 4N_e\mu$ – and in the latter case only if assuming that certain species characteristics, such as range size, body size, generation time or IUCN status, may serve as N_e proxies (Bolívar et al., 2019) (Fig. 2).

A further complication is that each function contains multiple explanatory variables, meaning that variation of dependent variables may be attributed to different factors, plus potential interaction effects, leaving room for interpretation. For instance, variation in genetic distances between lineages may not only be attributed to the time to most recent common ancestor, but also to mutation pressure (Kimura, 1991b). Similarly, variation in genetic diversity across lineages may reflect differences in effective population size, but also differences in global mutation rates (Hodgkinson & Eyre-Walker, 2011), or an interaction effect between these two factors (Kimura & Ohta, 1971c).

Another complication is that discrepancies between strict neutrality predictions and actual observations might not refute the hypothesis, but instead be attributable to confounding factors. In the words of Kimura (1991b, p. 379): 'The real biological world […] is very complicated, containing some disturbing or complicating factors that make actual observations depart from the neutrality predictions'. The most prominent confounder is linked selection. Strict neutrality predictions are derived for individual loci in isolation from the rest of the genome, and hence ignore the effects of linked selection (Charlesworth & Charlesworth, 2018). While linked selection does not affect substitution rates (Birky & Walsh, 1988), it is able to cause genetic diversity to deviate from neutral theory predictions (Gillespie, 2001; Buffalo, 2021).

Another example of a confounder is directional mutation pressure: variation in GC content across species and across loci may result from differences in conversion rates from AT bases to GC bases and vice versa (Sueoka, 1962, 1988; Freese, 1962; Jukes, 1991). A process with similar effect is GC-biased gene conversion (BGC), where during recombination events GC bases are favoured over AT bases, regardless of their selective values (Galtier et al., 2001; Kern & Begun, 2005; Pouyet et al., 2018; Harris, 2018). As a consequence, codon frequencies might deviate from strict neutrality predictions even in the absence of selective pressure, and hence codon usage bias can only be interpreted as such after correcting for GC content (Sharp & Li, 1986; Li et al., 2015).

The greatest difficulty for testing the strict neutrality model, however, is violation of underlying assumptions. Strict neutrality predictions are based on the Wright–Fisher model, which assumes constant population sizes and non-overlapping generations (Akashi et al., 2012). A mismatch between predicted and observed values could be interpreted as evidence against the strict neutrality model, but may alternatively simply reflect violation of Wright–Fisher model assumptions, particularly demographic non-equilibrium (Nei, Maruyama & Chakraborty, 1975; Pollak, 1982; Nei, 2005b; Balloux & Lehmann, 2012; Akashi et al., 2012; Müller, Kaj & Mugal, 2022).

For instance, the predicted correlation between heterozygosity and effective population size ($He \propto N_e$) holds only in the case of long-term constant population sizes, when the gain and loss of alleles has reached an equilibrium, a prerequisite which in reality will rarely be met (Nei et al., 1975; Nei, 2005a). In the words of Kimura (1979b, p. 120): 'One can show mathematically that the genetic variability due to neutral alleles can be greatly reduced by a population bottleneck from time to time, after which it takes millions of generations for the variability to build up again to the theoretical level characteristic of a very large population maintained constantly over a long period'. While, in theory, N_e may be considered the harmonic mean of historic population sizes, in practice this interpretation makes it almost impossible to evaluate the relationship between genetic diversity and N_e proxies ($He \propto N_e$), as the latter are usually based on contemporary indicators.

This footnote to the neutral theory concerns not only predictions on genetic diversity, but also predictions on genetic divergence from standing variation. In the absence of selective pressures, allele frequencies between disconnected sister populations are expected eventually to become uncorrelated. However, it may take genetic drift many generations to remove all traces of shared ancestry and to establish the

new migration–drift equilibrium (Ayala, 1974; Li & Nei, 1977).

(2) Neutral theory predictions

The neutral hypothesis adds another factor to the equation: functional constraint f_0, which is defined as the proportion of neutral mutations. In the absence of positive and balancing selection, or at least if assuming '*they are so rare that they may be neglected from our consideration*' (Kimura, 1991*b*, p. 371), the expected genetic distances and heterozygosity levels are now given by $d = 2\mu f_0$ and $He = 4N_e\mu f_0$ (Figs 1–3), with μf_0 denoting the neutral mutation rate (μ_0) (Kimura, 1977, 1991*a*).

The addition of an extra parameter opens up room for interpretation. Owing to potential differences in functional constraint across lineages, the theory now allows for substitution rate variation. Similarly, variation of polymorphism levels across loci may now not only be attributed to mutation rate variation or differences in effective population size (as a result of linked selection) (Hill & Robertson, 1966; Birky & Walsh, 1988), but also to functional constraint (Fig. 3). Consequently, the positive relationship $d \propto He$, which according to the strict neutrality hypothesis arises from local mutation rate variation, could according to the neutral hypothesis also reflect variation in f_0 across loci (Hudson, Kreitman & Aguadé, 1987).

Another question is: how to estimate f_0? Among the factors traditionally thought to determine functional constraint are functional density (i.e. the proportion of critical sites) (Dickerson, 1971; Zuckerkandl, 1976) and the physicochemical differences between amino acid sequences and their 'single-mutation derivatives' (Epstein, 1967; Grantham, 1974; Kimura, 1983; Graur, 1985; Tang *et al.*, 2004; Chen *et al.*, 2019). In practice, the functional constraint of a protein can be measured based on the second criterion (e.g. by calculating physicochemical distances to all conceivable single-mutation derivatives), but it is rarely feasible to determine the proportion of critical sites. As commented by Graur (1985, p. 53), '*the "importance" of a protein or site is frequently inferred from its rate of evolution, and the argument thus becomes a circular one*'.

Kimura & Ohta (1974) suggested that functional constraint may also depend on 'functional importance', which could be interpreted to mean 'indispensability': the probability that an organism can survive and reproduce without the given protein (Wilson, Carlson & White, 1977). However, gene knockout experiments have provided little empirical support for this hypothesis, and indicate that essential genes have similar substitution rates to non-essential genes (Hurst & Smith, 1999; Zhang & Yang, 2015). It has since emerged that protein substitution rates correlate with expression levels, but the underlying mechanisms are yet to be understood (Zhang & Yang, 2015). Functional constraint is still a vague catch-all term, which compromises the falsifiability of neutral theory predictions.

(3) Nearly neutral theory predictions

The '*remarkable simplicity*' (Kimura, 1991*b*, p. 371) of neutral theory equations depends critically on the unrealistic assumption that deleterious mutations are either so detrimental that they are immediately removed from the population, or alternatively so nearly neutral that they behave like completely neutral mutations (Gillespie, 1995; Akashi *et al.*, 2012). In the words of Kimura (1991*a*, p. 5969): '$(1-f_0)$ *represents the fraction of definitely deleterious mutants that are eliminated from the population without contributing either to evolution or polymorphism, even if the selective disadvantages involved may be very small in the ordinary sense*'. This simplification causes a disparity between the proposition (which is claimed to consider completely neutral as well as nearly neutral mutations) and the tested predictions ($d = 2\mu f_0$ and $He = 4N_e\mu f_0$), which hold true if, and only if, ignoring nearly neutral mutations.

In reality, as pointed out by Ohta (1972*a*), nearly neutral, slightly deleterious mutations will segregate in the population and occasionally reach fixation, and thereby increase polymorphism levels and substitution rates. Taking both components (neutral and deleterious mutations) into account, the expected genetic distance is given by $d = 2\mu t(f_0 + q(1 - f_0))$ (Ohta, 1972*a*, 1973), with q denoting the proportion of deleterious mutations that reach fixation, a proportion which in itself depends on N_e. Similarly, polymorphism rates are a composite of polymorphisms in mutation–drift equilibrium ($He = 4N_e\mu f_0$) and polymorphisms occurring in mutation-purifying selection balance ($He \propto 1/N_e$). This second part of the equation is more difficult to predict, but likely it is incorrect to assume that the proportion of sites in mutation–drift equilibrium largely outweighs the proportion occurring in mutation–selection equilibrium (Ohta, 2003). Although deleterious variants can be relatively rapidly purged by selection, the high frequency with which they arise through mutation means they could constitute a substantial, perhaps even largest, proportion of total polymorphisms (Lewontin, 1974; Akashi & Schaeffer, 1997; Bertorelle *et al.*, 2022). Thus, substitution rates and polymorphism levels are much less predictable when slightly deleterious alleles are included in the equations (Kimura, 1979*a*).

Gradual refinement of the neutral hypothesis has generated a multi-parameter hypothesis (the nearly neutral hypothesis) which is difficult to falsify on empirical grounds, as almost any observation can be brought into accordance by adjusting one or multiple parameters (particularly f_0 and N_e, the latter determining q) (Akashi *et al.*, 2012). In the words of Kreitman (1996, p. 683): '*The battle [between neutralists and selectionists] may have shifted slightly from complete neutrality to near neutrality. But this slight shift is a quantum leap in terms of the difficulty in distinguishing between the nearly neutral model and the stronger selection models. [...] One can almost always propose a particular history of changes in population size that will account for almost any pattern of molecular variation or change. Thus, unlike the strictly neutral theory, the slightly deleterious model cannot be easily falsified*'. Crow (1997, p. 262) agreed: '*It has been remarkably difficult to distinguish experimentally or analytically between the nearly neutral theory and evolution*

by incorporation of favourable mutations in a fluctuating environment'. Kimura held '*that the nearly neutral theory might be more realistic than the […] neutral theory, but that the latter was certainly more useful than the former*'. (Ohta, 2003, p. 376).

This drawback of the nearly neutral hypothesis illustrates the delicate trade-off between simplicity and realism of a model (Higgins, 2004). The assumption of reduced efficiency of purifying selection in small populations (Ohta, 1972*c*), which makes the nearly neutral hypothesis less '*attractive as a null model*' (Kreitman, 1996, p. 678), is supported by empirical data (Hughes, 2008). For instance, comparisons across species with a range of population sizes have revealed a negative relationship between N_e and amino acid substitution rates (Akashi *et al.*, 2012; Galtier, 2016; Moutinho, Bataillon & Dutheil, 2020), between N_e and the ratio of radical to conservative substitutions (*Kr/Kc*) (Weber *et al.*, 2014; Weber & Whelan, 2019), between N_e and the ratio of non-synonymous to synonymous substitutions (Popadin *et al.*, 2007; Hughes, 2008; Kim & Yi, 2008; Weber *et al.*, 2014; Galtier, 2016; Figuet *et al.*, 2016; Bolívar *et al.*, 2019), between N_e and the difference between expected and observed heterozygosity (Corbett-Detig *et al.*, 2015), and between generation time (proxy for N_e) and extent of codon usage bias (Subramanian, 2008).

(4) Predictions based on a continuous DFE

For simplicity, the neutral theory assigns mutations artificially into discrete categories, namely: (slightly) deleterious, neutral, and beneficial. In reality, of course, the fitness effects of mutations are better described by a continuous distribution, ranging from lethal to highly beneficial (Crow, 1972*a*; Ohta, 1977; Kimura, 1979*a*; Ohta & Gillespie, 1996; Keightley & Eyre-Walker, 2010; Razeto-Barry *et al.*, 2012), meaning that '*the borderline between deleterious and neutral mutations is vague*' (Ohta & Kimura, 1971*b*, p. 393).

Ohta (1977) and Kimura (1979*a*) used mathematical modelling to derive neutralist predictions which are not based on the erroneous assumption of discrete categories, but instead on a more realistic, continuous distribution of fitness effects (DFE). This increases the precision of neutral theory predictions, but only if the hypothesised underlying DFE is accurate. Ohta (1977) assumed the DFE is best described by an exponential distribution, while Kimura (1979*a*) argued in favour of a gamma distribution.

On the downside, not much remains of the elegant simplicity of the neutral theory once its equations have been rewritten to derive predictions from a continuous distribution. The complexity of this type of model is illustrated by a conceptual error which was uncovered by Gillespie (1995). Both Ohta (1977) and Kimura (1979*a*) originally defined the fitness effects of novel mutations in their models relative to the dominant allele, and as a consequence the DFE shifted every time the dominant allele was substituted. To overcome the shortcomings of these 'shift models', Ohta & Tachida (1990, p. 220) introduced a 'fixed model' (also known as house-of-cards model) (Gillespie, 1995; Razeto-Barry

et al., 2012), in which '*the distribution of fitness coefficients is fixed regardless of the allelic state occupying the population*'. However, Gillespie (1995) discovered that this model, which was supposed to represent Ohta's nearly neutral hypothesis, in fact equated to a selectionist hypothesis: it predicted half of all substitutions to be advantageous.

(5) A theory under construction

The difficulty to falsify the neutral hypothesis is illustrated by the history of the neutralist–selectionist debate, which is characterised by repeated attempts of neutralists to reconcile neutrality predictions with empirical observations through adjustment of the four parameters f_0, μ, N_e, and q (and their interdependencies).

For instance, adjusting the parameter f_0 (functional constraint, or proportion of neutral mutations) from large to small (first revision), served to bring the neutral hypothesis into agreement with lines of evidence for pervasive purifying selection. For example, the observation that chemically dissimilar amino acids replace each other less frequently than chemically similar amino acids (Epstein, 1967), a fingerprint of purifying selection, was initially cited against the neutral hypothesis (Clarke, 1970*b*; Grantham, 1974), but after the revision as support for the neutral hypothesis (Kimura & Ohta, 1971*b*). Lewontin (1974, p. 228) complained: '*Thus the neoclassicists have the best of both worlds. Both randomness and non-randomness are interpreted as evidence in their favour. They do not tell us what observations might not confirm the theory*'.

As a second example, King & Jukes (1969) argued that the correlation between codon degeneracy and amino acid type frequencies indicated stochasticity, but they were unable to explain the observed deviations (Gilis *et al.*, 2001). These deviations from strict neutrality expectations were seized upon by selectionists as evidence against the neutral theory (Richmond, 1970; Clarke, 1970*a*), until modelling studies suggested that they could also arise from purifying selection (Kimura & Ohta, 1971*b*; Ohta & Kimura, 1971*b*). As a final example, King & Jukes (1969) claimed that the number of amino acid substitutions at variant sites within proteins follows a Poisson distribution. This claim, rejected by Clarke (1970*a*), Richmond (1970) and Fitch & Markowitz (1970), was tacitly retracted following the first revision of the neutral hypothesis (Kimura & Ohta, 1971*c*), which from then on correctly predicted fewer substitutions in functionally important protein regions (Kimura & Ohta, 1974; Cooper *et al.*, 2005; Hughes, 2008).

A purported inverse relationship between the parameter μ (mutation rate per generation) and N_e (effective population size) served to bring the prediction for protein polymorphism (i.e. $He = (4N_e\mu)/(1 + 4N_e\mu)$) in accordance with the observation that heterozygosity is relatively constant across species. Kimura & Ohta (1971*c*) speculated that differences in effective population size are cancelled out by differences in mutation rate (measured per generation, μ_g). They wrote (p. 469): '*The species with short generation time [and hence lower μ] tends to have small body size and attain a large population number, while the species*

which takes many years for one generation [and hence has higher μ*] tends to have a small population number*'. Therefore, the product $4N_e\mu$ '*should be less variable among different organisms than its components*'. (Kimura & Ohta, 1971*c*, p. 469).

The novel assumption that N_e is inversely related to generation time (G) was subsequently used by Kimura & Ohta (1971*c*) and Ohta (1972*a*) to explain another discrepancy between observations and neutral theory predictions: the absence of a generation time effect. In the formula $d = 2\mu t$, t should be interpreted as number of generations (t_g), not number of years (t_y), because this is what counts for biological evolution (Fig. 2) (Ohta, 2003). Likewise, μ should be interpreted to mean mutation rate per generation (μ_g), not per year (μ_y). For any given time interval (t_y), species with short generation times (G) proceed through more generations (higher t_g) relative to species with longer generation times. Because the number of germ line cell divisions does not scale to generation time (Drost & Lee, 1995; Weller & Wu, 2015), the mutation rate per generation is not proportional to mutation rate per year ($\mu_g \neq \mu_y G$), meaning $t_g = t_y/G$ is not simply cancelled out. Instead, species with short generation times are expected to accumulate more substitutions.

Yet, the empirical data from which the molecular clock hypothesis had been inferred, indicated that substitution rates were constant over time across lineages, independent of generation time (Dickerson, 1971; Crow, 1972*a*). This was inconsistent with neutral theory predictions (Ohta, 2003), unless maintaining the questionable assumption $\mu_g = \mu_y G$ (Kimura, 1968*a*; Gillespie, 1995). In the words of Gillespie (1995, p. 65): '*Thus, the neutral theory did not, as is commonly claimed, "predict" the molecular clock. Rather, a biologically unrealistic assumption had to be introduced to bring neutrality and the protein molecular clock into agreement*'.

Ohta & Kimura (1971*c*) and Ohta (1972*a*) added the parameter q, the proportion of deleterious mutations that reach fixation, to reconcile the neutral theory with the data (Laird, McConaughy & McCarthy, 1969; Ohta, 1972*a*; Kohne, Chiscon & Hoyer, 1972; Sarich & Wilson, 1973; Wilson *et al.*, 1977; Pollak, 1982; Gillespie, 1995). They argued that the absence of a 'generation time' effect stems from the negative correlation between N_e and generation time (G), combined with the negative correlation between N_e and q. In the words of Ohta & Kimura (1971*c*, p. 23): '*If generation time and effective population size are negatively correlated, the resulting substitution rate will approach constancy per year*'. In formula: if $k_y = k_g/G$ and $k_g = \mu_g q$, we obtain: $k_y = \mu_g q(1/G)$; when further assuming $q \propto 1/N_e$, we obtain: $k_y \propto \mu_g(1/N_e)$ $(1/G)$; and thus: $k_y \propto \mu_g/(N_e G)$; and finally, when assuming $G \propto 1/N_e$, we obtain $k_y \propto \mu_g$. This way neutral theory predictions could be reconciled with empirical data, but at the cost of adding to the theory two additional propositions ($\mu \propto 1/N_e$ and $q \propto 1/N_e$).

Corrections are a natural part of theory development. Early versions of the neutral hypothesis were falsified, and replaced with updated versions (i.e. the revised neutral hypothesis and nearly neutral hypothesis) which take into account the original shortcomings. Still, the practice of

continually adjusting and adding parameters, or introducing novel assumptions about their interdependency, in order to accommodate empirical findings, begs the question of which observations allow one to reject the neutral hypothesis rather than warranting a minor correction only.

Kern & Hahn (2018, p. 1369) wrote that '*in order to remove the lingering misapprehensions of the neutral theory, we must of course replace it with an explanatory theory of greater value*'. They added (p. 1369): '*A more sufficient model of genetic variation would at minimum have to account for the direct and indirect effects of selective sweeps and the direct and indirect effects of purifying selection, while simultaneously accounting for variation in population size and population structure*'.

But abandoning the neutral theory altogether just because of not agreeing with the main proposition could be a typical case of throwing the baby out with the bathwater. The neutral theory is more than the neutral hypothesis alone; it has developed into a theoretical framework which allows derivation, from an interchangeable main proposition, of a range of predictions on genetic diversity and genetic divergence (Fig. 2). While it is clear that the various assumptions underlying neutral theory predictions (e.g. absence of linked selection, constant N_e) are unrealistic, this does not render these predictions useless. As, for instance, illustrated by the events leading to the discovery of Neptune, discrepancies between predictions and observations are informative in itself, and often simply indicate a set of ideas is incomplete rather than incorrect. The neutral theory provides a theoretical infrastructure to build upon, and further improvements may ensure that predictions come to take into account the confounding effects of linked selection and non-equilibrium population demography.

VIII. INVALID ARGUMENTS AGAINST THE NEUTRAL HYPOTHESIS

(1) Molecular clock: rate heterogeneity

Arguments against the neutral hypothesis often refute strict neutrality predictions, not the neutral hypothesis itself. A prime example concerns the observed variation in substitution rates across lineages, also known as overdispersion of the molecular clock (Langley & Fitch, 1974; Cutler, 2000; Kern & Hahn, 2018). This argument against the neutral hypothesis dates back to Langley & Fitch (1974), who were among the first – together with Ohta & Kimura (1971*a*) – to present statistical evidence that rate heterogeneity across lineages exceeds stochastic variation (as expected from a Poisson model). Langley & Fitch (1974, p. 174) believed that the '*rejection of the hypothesis of constancy of [...] rates of substitution deals a [...] severe blow to the hypothesis that the amino acid sequence differences are the product of random drift of neutral mutations*'.

However, the findings of Langley & Fitch (1974) merely showed that the variation of evolutionary rates among branches is significantly higher than expected by pure chance. Langley & Fitch (1974) did not test whether rate

heterogeneity differs significantly from neutral theory predictions (Wilson *et al.*, 1977). Unlike the strict neutrality model, the neutral hypothesis predicts that the substitution rate depends on the proportion of neutral and deleterious mutations (f_0), which can differ between environments and thus between lineages. The nearly neutral hypothesis adds that the substitution rate depends as well as on the efficacy of purifying selection, which varies across lineages with $\mathcal{N}_e$. Therefore, the neutral hypotheses allow for a certain degree of rate heterogeneity across lineages greater than that predicted by a simple stochastic process (Wilson *et al.*, 1977; Kimura, 1977). This prediction has been confirmed by modelling and simulation studies, which have indicated that rate heterogeneity can indeed occur in the absence of positive selection and hence does not contradict the neutral hypothesis (Cutler, 2000; Bastolla, Vendruscolo & Roman, 2000). In other words, rejection of a strict molecular clock does not necessarily imply rejection of the neutral theory (Nei, Suzuki & Nozawa, 2010).

Wilson *et al.* (1977, p. 615) commented: '*The use of the Poisson distribution as the probabilistic model for the neutral-mutation hypothesis may have been an oversimplification. It seems reasonable to propose, for example, that the rate of occurrence of neutral mutations [the factor "μf_0"] might itself be subject to fluctuations. [...] Thus, although the variation of the evolutionary clock is greater than that of a Poisson process, this does not invalidate the neutral hypothesis*'. Kimura (1991a, p. 5971) agreed: '*A universally valid and exact molecular evolutionary clock would exist only if, for a given molecule, the mutation rate for neutral alleles [μf_0] were exactly equal among all organisms at all times (which is rather unlikely in nature). [...] In other words, the variance of the evolutionary rates among different lineages for a given molecule may tend to become larger than expected from the simple Poisson distribution*'.

(2) Lewontin's paradox

Arguing against the neutral hypothesis based on Lewontin's paradox of variation is yet another example of erecting and refuting the neutralist strawman of strict neutrality. Strict neutrality predicts a mutation–drift equilibrium in which heterozygosity depends on $\mathcal{N}_e$, as given by $He = 4\mathcal{N}_e\mu$ (Kimura & Crow, 1964; Kimura, 1969a, 1991a). In reality, levels of genetic variation appear to differ much less among species with varying $\mathcal{N}_e$ than predicted by this equation (Maynard-Smith, 1970b; Lewontin, 1974; Nei, 2005b; Buffalo, 2021). This violation of strict neutrality predictions is regularly equated with violation of neutral hypothesis predictions, cultivating claims that Lewontin's paradox is incompatible with the neutral hypothesis (Hahn, 2008; Corbett-Detig *et al.*, 2015; Kern & Hahn, 2018). However, as discussed in Section VII, the neutral hypothesis predicts some alleles to be in mutation–drift equilibrium and others in mutation–selection equilibrium, causing deviations of heterozygosity levels (Kimura, 1979b; Gillespie, 2000; Charlesworth & Charlesworth, 2018). Furthermore, even if all mutations are neutral, deviations from strict neutrality predictions may still be expected due to the confounding

effects of linked selection (Gillespie, 2001; Buffalo, 2021), as well as due to violation of the assumption of population equilibrium conditions (Nei *et al.*, 1975; Nei, 2005b) (see Section VII).

(3) Polymorphism levels *versus* recombination rates

A third example of arguing against the neutral hypothesis based on rejection of strict neutrality predictions concerns the observed correlation between recombination rates and polymorphism levels. Linked selection reduces the local effective population size and therefore allows to test the prediction $He \propto \mathcal{N}_e$ by comparing polymorphism levels across loci. More specifically, proteins or genes occurring in genomic regions with low recombination rates should harbour less genetic variation than proteins or genes in regions with high recombination rates (Begun & Aquadro, 1992; Begun *et al.*, 2007; Hughes, 2008; Wagner, 2008; Lohmueller *et al.*, 2011; Cutter & Payseur, 2013; Charlesworth & Jensen, 2021). Originally, it was argued that such a positive correlation between recombination rates and polymorphism levels is a hallmark of linked positive selection (Kaplan, Hudson & Langley, 1989; Begun & Aquadro, 1992; Hahn, 2008; Kern & Hahn, 2018). However, it was soon recognised that this correlation can also result from linked negative selection (Charlesworth, Morgan & Charlesworth, 1993; Hudson & Kaplan, 1995; Lohmueller *et al.*, 2011; Frankham, 2012; Charlesworth & Jensen, 2021).

Thus, the observed correlation between recombination rates and polymorphism levels is of little use in assessing the validity of the neutral hypothesis. In the words of Kreitman (1996, p. 682): '*An alternative model of hitchhiking has now been proposed in which neutral mutations are eliminated by virtue of their linkage to frequently occurring deleterious mutation, and the data are now thought to be largely consistent with this theory. Therefore, what was initially thought to be incontrovertible evidence for positive selection is now explained by deleterious selection*'. Kreitman (1996, p. 682) concluded: '*The unexpected strong correlation of nucleotide polymorphism levels and recombination means that selection and hitchhiking must be operating to shape patterns of variation [...], but we may not be able to resolve whether that selection is mostly positive, mostly negative, or a mixture of the two*'.

(4) Similar allele frequencies across populations

An early argument against the neutral theory was the claim that the observed similarity of allele frequencies in geographically separated populations suggested balancing selection (Stone *et al.*, 1968; Prakash, Lewontin & Hubby, 1969; Maynard-Smith, 1970b; Clarke, 1970b; Ayala, 1974; Ayala & Gilpin, 1974; Chakraborty, Fuerst & Nei, 1978). It was thought that the neutral hypothesis predicts high allele frequency differences between isolated populations, due to the random segregation of alleles. Li & Nei (1977) showed that this prediction only holds true for a system of populations in migration–drift equilibrium, which may take many generations to set in. Recently disconnected populations

can retain the same alleles at similar frequencies for prolonged periods of time, provided that population sizes are large. They concluded (p. 912): '*Recent electrophoretic surveys of proteins have shown that two related species often share many common alleles. Some investigators interpreted this as an indication of the selective maintenance of the alleles. […] The present study, however, shows that two related species may share common alleles for a long time after their separation even if there is no selection*'.

Because the level of population differentiation (F_{ST} value) depends on many, often unknown, demographic factors (i.e. split time, effective population size, migration rate), allele frequency differences do in itself not allow to readily test the neutral hypothesis. But there is a potential work-around: evaluating the variation of F_{ST} values across loci. Genetic drift, as a function of demographic changes, affects all loci, unlike natural selection (Cavalli-Sforza, 1966; Lewontin & Krakauer, 1973). Therefore, strict neutrality predicts that all loci should have roughly similar F_{ST} values, with the neutral hypothesis allowing for few loci to stand out from this neutral distribution (de Jong, Lovatt & Hoelzel, 2021).

(5) Polymorphism levels *versus* substitution rates

The neutral hypothesis makes a testable prediction regarding the covariance of divergence and polymorphisms across loci (e.g. proteins or genes), namely '*that molecules or parts of one molecule which are more important in function, and which therefore evolve more slowly, will show a lower level of heterozygosity*' (Hudson *et al.*, 1987, p. 153).

A locus that is free to evolve without positive or negative selection will have high levels of variation (e.g. high nucleotide diversity) within populations, as well as high levels of divergence (e.g. high sequence dissimilarity) between populations. A locus that is functionally constrained experiences purifying selection and is expected to have low levels of variation within populations, as well as low levels of divergence between populations (i.e. low *He*, low *d*) (Chen, Wang & Cohen, 2007). By contrast, a locus under positive selection will contain low variation within the selected population due to rapid fixation of the adaptive allele, but will differ from homologous loci in sister populations (i.e. low *He*, high *d*), assuming the locus is selected for in only one population. Furthermore, a locus under balancing selection will exhibit high levels of genetic variation within populations, but relatively low levels of genetic variation between populations (i.e. low *He*, low *d*). In summary, genetic drift and purifying selection cause a positive relationship between genetic distance and polymorphism levels (albeit for different reasons), whereas positive and balancing selection cause a negative relationship between – or 'uncoupling of' – genetic distance and polymorphism levels. These different predictions permit the neutral hypothesis and the selectionist hypothesis to be put to the test (Berry, Ajioka & Kreitman, 1991).

Begun *et al.* (2007) tested this prediction on genomic data from two closely related *Drosophila* species, and found a negative relationship between divergence and polymorphism levels. It has been claimed that this finding is inconsistent with the neutral theory (Begun *et al.*, 2007; Hahn, 2008; Wagner, 2008). However, the reasoning underlying the predicted polymorphism–divergence correlation applies only to direct selection, not to linked selection. Background selection and selective sweeps reduce intraspecies genetic variation, but do not alter substitution rates of linked neutral alleles (Birky & Walsh, 1988; Berry *et al.*, 1991; Phung *et al.*, 2016). Because linked selection does not affect substitution rates, genome-wide polymorphism–divergence correlations (Begun *et al.*, 2007) cannot distinguish between the neutral hypothesis and the selectionist hypothesis (Jensen *et al.*, 2019). Thus, the predicted positive correlation between polymorphism levels and divergence levels applies to protein or gene data only (Skibinski & Ward, 1982; Chakraborty & Hedrick, 1983; Ward & Skibinski, 1985; Hudson *et al.*, 1987). Challenging the neutral hypothesis based on the absence of such a relationship in genome-wide data is another example of testing a prediction that does not logically follow from the neutral hypothesis (Begun & Aquadro, 1991; Berry *et al.*, 1991; Begun *et al.*, 2007).

IX. INVALID ARGUMENTS AGAINST THE SELECTIONIST HYPOTHESIS

(1) Cost of selection

Useful predictions are discriminative: they logically follow from only one of the competing hypotheses. In the words of Gillespie (1984*b*, p. 733): '*It is not enough to argue that the neutral theory is compatible with the observations; it must also be shown that selection is incompatible*'. In this last section we will question four observations from molecular data for which neutralists claim that they are consistent with the neutral hypothesis and inconsistent with the selectionist hypothesis.

One of the earliest arguments against the selectionist hypothesis was the 'cost of selection' argument. Kimura (1968*a*) calculated from protein sequence alignment data that substitution rates in nature greatly exceed Haldane's theoretical upper limit of sustainable adaptive substitutions (one in every 300 generations), as inferred from substitutional load considerations. Kimura (1968*a*) argued that this discrepancy implies that most substitutions cannot be adaptive, and instead must result from the fixation of neutral mutations through random genetic drift.

Similarly, considerations on segregation load appear to suggest an upper limit on the number of balanced polymorphisms (Kimura & Crow, 1964; Hubby & Lewontin, 1966; Dietrich, 1994). In theory, the fitness differences between individuals in a population with many polymorphisms under balancing selection would vary beyond reproductive capacities. Kimura (1968*a*) argued that the high levels of polymorphisms observed by Hubby & Lewontin (1966) therefore implied that most polymorphisms had to be neutral.

However, the 'cost of selection' argument has been challenged on theoretical grounds by numerous studies (Sved, 1968; Maynard-Smith, 1968; Brues, 1969;

Felsenstein, 1971; Ewens, 1972), including by King & Jukes (1969). These criticisms are supported by simulation studies, which suggest that the upper rate of adaptive substitutions is far less strict than Haldane's limit (Nunney, 2003; Weissman & Barton, 2012; Hickey & Golding, 2019). One obvious mistake made by Kimura (1968a), and pointed out by King & Jukes (1969), was that he calculated the substitution rate over the entire genome, rather than for functional DNA only, causing an overestimate of the cost of selection (Crow, 1972b; Dietrich, 1994; Nei, 2005b). Kimura (1968a) also did not discuss the differences between adaptation from standing variation after environmental change compared to adaptation from a novel mutation in a stable environment (Van Valen, 1963; Brues, 1964, 1969).

With respect to balancing selection, several mechanisms have been proposed that potentially explain, by relaxing or adjusting assumptions underlying cost of selection calculations, how high numbers of adaptive polymorphisms can be maintained without causing an unbearable load (Milkman, 1967; Sved, Reed & Bodmer, 1967; King, 1967; Wallace, 1975; Charlesworth, 2013; Reznick, 2016). One such explanation is that the fitness of genotypes should not be compared to the theoretical optimal genotype (heterozygous for all loci), because this genotype is very unlikely to occur anyway (Milkman, 1967; Sved *et al.*, 1967; King, 1967; Dietrich, 1994).

Hence, a key argument against the selectionist hypothesis, put forward in the landmark founding paper of the neutral theory (Kimura, 1968a) and the line of evidence that first led Kimura '*to consider seriously the possibility*' (Kimura, 1988, p. 5), appears to be conceptually or technically flawed. The 'cost of selection' argument backfired on neutralists, and selectionists have asked the question: if the original line of evidence has faltered, then why still cling to the theory it once inspired? (Dietrich, 1994; Kern & Hahn, 2018).

However, refutation of the 'cost of selection' argument does not eliminate evidence in support of the neutral hypothesis, but instead undoes rejection of the selectionist hypothesis. This means that rebuttal of the 'cost of selection' argument does not invalidate the neutral hypothesis, which was first inferred from the discovery of the molecular clock and iso-alleles (Ingram, 1963; Sved, 1968; Ohta & Gillespie, 1996; Ohta, 2003; Nei, 2005b). As suggested by Maynard-Smith (1968, p. 1116), Kimura (1968a) could have been right for the wrong reasons: '*Kimura's conclusion that a large proportion of amino acid substitutions are selectively neutral and have occurred by drift, although it may be true, does not follow necessarily from the cost of selection argument*'.

Furthermore, Kimura (1968a) might have underestimated the possible rate of adaptive substitutions, but he correctly estimated the possible rate of neutral substitutions. It had been unclear whether random genetic drift could lead to a sufficient number of substitutions to account for the observed sequence divergence between species (Margoliash & Smith, 1965; Dietrich, 1994). Kimura (1968a) removed this doubt by showing that in the absence of selective pressures, the substitution rate equals the mutation rate. This conclusion remains, even after refutation of the 'cost of selection' argument.

(2) Molecular clock: rough rate constancy

Neutralists have pointed out that if most substitutions were indeed adaptive, one would expect to find a much higher level of rate heterogeneity across lineages than actually observed (King & Jukes, 1969; Kimura & Ohta, 1971c), such that phylogenetic reconstruction from molecular data would be impossible altogether (Kimura & Ohta, 1971a; Kimura, 1991b).

As commented by Ohta & Gillespie (1996, p. 139), as of yet selectionists '*cannot explain why there is not even more variation in the rate of substitution. For example, why have we not found some lineages on which hemoglobin evolution has stopped?*' Van Valen (1974, p. 89) commented that rate constancy '*is so general a result, even with the exceptions, that it should be predictable as such from any theory that attempts to explain evolution. The neutralist theory of Kimura […] does this; the existing selectionist theory does not*'. Kimura (1983, p. 85) therefore stated that '*emphasizing local fluctuations as evidence against the neutral theory, while neglecting to inquire why the overall rate is intrinsically so regular or constant is picayunish. It is a classic case of "not seeing the forest for the trees"*'.

While rough rate constancy across lineages appears indeed to be more parsimoniously explained by the neutralist hypothesis, selectionists have proposed an alternative explanation which according to modelling studies might be plausible. This dates back to Lewontin (1974), who argued that long-term averages may obscure short-term deviations caused by episodes of selection. Phrased differently, Lewontin (1974) claimed that temporal substitution rate variation within lineages can in the long run produce apparent substitution rate constancy across lineages. He pointed out that the average divergence time between species pairs in protein data sets available at the time was approximately ∼100 million years (Margoliash, 1963; Zuckerkandl & Pauling, 1965; Sarich & Wilson, 1967, 1973; Dickerson, 1971; Ohta & Kimura, 1971a). Lewontin (1974, pp. 228–229) commented: '*Since that time every phyletic line has undergone numerous episodes of rapid and slow evolution, so that the substitution rates shown in the total are averages over vast periods of time and many cycles of speciation, extinction and phyletic evolution. The claimed "constancy" is simply a confusion between an average and a constant*'.

This claim was dismissed by Kimura (1979b, p. 104), who wrote: '*The death rates characteristic of man and of an insect do not become equal by merely being averaged over a long period of time or over a large number of individuals; there is no reason to expect two averages to converge on each unless the intrinsic factors shaping them are the same*'. However, Gillespie arrived at a different conclusion after performing extensive mathematical modelling intended to investigate the consequences of an episodic clock (Gillespie, 1984a, 1986, 1994). Unlike Kimura's models, Gillespie's models allowed the substitution rate (k) to alternate between two states, reflecting periods of evolutionary stasis ($k = 0$) interspersed with bursts of episodic selection ($k > 0$). He envisioned the short bursts to reflect rapid

adaptation to '*major environmental changes, such as recent ice ages*', thought to '*occur on time scales of tens of thousands or hundreds of thousands of years*' (Gillespie, 1986, p. 139). These models suggested, as hypothesised by Lewontin (1974), that '*large variabilities in rate of evolution are not necessarily manifest in large variabilities in the numbers of substitutions in different lineages*'. (Gillespie, 1986, p. 139). In fact, the shorter the duration of episodic selection events, the lower the predicted rate variability across lineages, with variability across lineages even dropping below those observed in nature (Gillespie, 1994; Ohta & Gillespie, 1996; Cutler, 2000).

(3) Functionally constrained loci evolve more slowly

Neutralists hold that substitution rate variation across loci results from differences in functional constraint (i.e. f_0, the proportion of neutral mutations), with the least constrained sites approaching the upper limit set by the global mutation rate (King & Jukes, 1969; Arnheim & Taylor, 1969; Dickerson, 1971; Kimura & Ohta, 1974; Kimura, 1977). In the words of Kimura (1979*b*, p. 104): '*Intrinsic evolutionary rates are essentially determined by the structure and function of molecules and not by environmental conditions*'.

The alternative explanation is variation in local mutation rates (Martincorena & Luscombe, 2013), but this explanation is less consistent with the finding that pseudogenes, including fossil genes (genes freed from purifying selection but remaining in the same location of the genome), also have much higher substitution rates than operative genes (Miyata & Yasunaga, 1981; Li, Gojobori & Nei, 1981; Hendriks *et al.*, 1987; Kimura, 1991*a*; Zhang & Yang, 2015).

Neutralists have argued that this negative relationship between substitution rates and functional constraint would not be observed if most substitutions are adaptive (Kimura, 1977, 1991*a*; Li *et al.*, 1981). For example, Kimura (1977, p. 276) wrote: '*Various observations suggest that as the functional constraint diminishes the rate of evolution converges to that of synonymous substitutions. If this is valid, such a convergence (or plateauing) of molecular evolutionary rates will turn out to be strong supporting evidence for the neutral theory. On the other hand, if we adhere to the selectionist position that practically all the mutant substitutions in evolution are caused by positive natural selection, there can be no upper limit to the evolutionary rate at the molecular level*'.

However, while this might be true for molecular panselectionism, it misportrays the selectionist hypothesis. Selectionists do not deny that genetic drift can cause many substitutions in non-functional regions. Thus, it is incorrect to state that selectionists believe that in pseudogenes '*virtually no nucleotide substitution occurs because they are functionless and there is no way for positive selection to operate*' (Nei, 2005*b*, p. 2323). Furthermore, if the vast majority of mutations at functional sites are deleterious, the small proportion of non-deleterious mutations that do reach fixation could still be predominantly adaptive even though the substitution rates at functional sites are below the rate at unconstrained sites. Thus, it is also not necessarily true that according '*to the selectionist view, the most important regions of proteins would be expected to evolve fastest*'. (Hughes, 2008, p. 164).

Both the neutral and the selectionist hypotheses acknowledge that many coding mutations are deleterious (Fig. 5), and therefore evidence for pervasive purifying selection at functional sites cannot be used to discriminate between the two hypotheses. In the words of Kreitman (1996, p. 680): '*The observed polymorphisms and fixed differences in proteins must, by definition, be ones that have escaped strong constraining selection. The question is, have they escaped by virtue of their selective neutrality (or near-neutrality) or by virtue of their selective advantage*'.

(4) $dN/dS < 1$

The dN/dS ratio, also known as omega (ω), quantifies the difference in substitution rates or polymorphism levels between two categories of sites which differ in functional importance – for instance, between non-synonymous and synonymous sites in protein-coding regions, or between protein-binding and non-protein-binding sites in regulatory regions (Eyre-Walker, 2006; Charlesworth & Eyre-Walker, 2008). Estimation of gene-specific dN/dS ratios indicates that substitution rates at synonymous sites are considerably higher than observed for non-synonymous sites (Sonneborn, 1965; Kimura, 1968*b*, 1977; King & Jukes, 1969; Kimura & Ohta, 1974; Jukes, 1978; Kreitman, 1983; Nei & Gojobori, 1986; Nei *et al.*, 2010), with typical dN/dS values among species pairs ranging between 0.1 and 0.3 (Kosiol *et al.*, 2008; Wolf *et al.*, 2009; Nei *et al.*, 2010).

Kimura (1991*b*, p. 375) argued that this observation is inconsistent with the selectionist model: '*If positive selection is the driving force of evolution, one should expect that synonymous and other silent changes should show lower evolutionary rates than amino-acid altering changes*'. However, the claim that selectionists would predict $dN > dS$ is incorrect. In reality, selectionists acknowledge that purifying selection drives dN/dS values of genes below 1.

Gene-wide dN/dS values are largely unaffected by the few non-synonymous changes that do reach fixation, regardless of whether they are neutral or adaptive (Nielsen & Yang, 1998; Eyre-Walker, 2006). Consequently, the detection of positively selected genes using a threshold of $dN/dS = 1$ (Hughes & Nei, 1988; Kimura, 1991*a*), is yet another example of testing strict neutrality, albeit this time leading to unjustified rejection of the selectionist hypothesis.

As a simplified example, consider a gene consisting of 20 non-synonymous sites and 10 synonymous sites. If this gene accumulates one non-synonymous and five synonymous substitutions, the dN/dS ratio would be $(1/20)/(5/10) = 0.1$, which is well below the threshold of $dN/dS = 1$. If we assume that the single non-synonymous substitution results from positive selection, the proportion of adaptive amino acid substitutions (α) is 1. The observation that dN/dS values are generally below one, demonstrates '*the prevalence of purifying selection on protein-coding regions and the comparative rarity of positive selection*' (Hughes, 2008, p. 164), but not the rarity of adaptive substitutions relative to neutral substitutions.

These considerations have inspired the development of methods which aim to estimate dN/dS ratios per codon rather than per gene (Yang & Nielsen, 1998, 2008; Yang, Wong & Nielsen, 2005; Nei *et al.*, 2010; Murrell *et al.*, 2012; Spielman & Wilke, 2015). Another potential work-around is to estimate for each gene a corrected dN/dS threshold, that is the dN/dS ratio expected after correcting for purifying selection. The McDonald–Kreitman test (McDonald & Kreitman, 1991) uses the dN/dS ratio observed within demes (Kreitman, 1983; Hughes & Nei, 1988; Hughes *et al.*, 2003; Hughes, 2008) as a proxy for the corrected dN/dS threshold, to which between-population dN/dS values (Hughes & Nei, 1988; Mugal, Wolf & Kaj, 2014) are compared, resulting in a neutrality index (Stoletzki & Eyre-Walker, 2011). In theory, the excess of the between-demes dN/dS ratio relative to this proxy of the corrected dN/dS threshold represents an estimate of α, the proportion of adaptive substitutions. In formula: $\alpha = (\omega_{between} - \omega_{within})/\omega_{between}$ (McDonald & Kreitman, 1991; Smith & Eyre-Walker, 2002; Bustamante *et al.*, 2005; Nei *et al.*, 2010).

However, this work-around is not flawless. Because deleterious alleles are unlikely to fix but do segregate within a population before being removed, the inferred dN/dS threshold is possibly an overestimate of the true corrected dN/dS threshold (albeit less severely as in the case of the threshold of $dN/dS = 1$). The outcome would be consistent underestimation of the proportion of adaptive substitutions (Fay, Wyckoff & Wu, 2002; Eyre-Walker, 2006; Charlesworth & Eyre-Walker, 2008; Andolfatto, 2008; Parsch, Zhang & Baines, 2009; Nei *et al.*, 2010; Galtier, 2016; Booker *et al.*, 2017; Murga-Moreno *et al.*, 2019). Violation of the assumption of constant population sizes could occasionally cause further bias, in either direction. Overestimation of α could occur in the case of recent population expansion (if slightly deleterious mutations previously were fixed but recently became more effectively removed), whereas underestimation would occur in the case of recent population reduction (if slightly deleterious mutations previously were removed but recently became less effectively removed) (McDonald & Kreitman, 1991; Fay *et al.*, 2002; Eyre-Walker, 2002; Parsch *et al.*, 2009; Nei *et al.*, 2010).

X. CONCLUSIONS

(1) The history of the neutralist–selectionist debate is characterised by continuous adjustment of four key parameters (i.e. effective population size, mutation rate, selective constraint, and efficiency of selection). These corrections, meant to reconcile neutrality predictions with empirical observations, have compromised the falsifiability of the neutral hypothesis.

(2) Two adjustments led to official revisions of the neutral hypothesis, as originally proposed by Kimura (1968*a*). The first and lesser-known revision acknowledged that many mutations are deleterious and are removed by purifying selection (King & Jukes, 1969; Kimura & Ohta, 1971*c*). The second revision acknowledged that in small populations deleterious mutations may contribute to polymorphism levels and substitution rates (Ohta & Kimura, 1971*c*; Ohta, 1972*a*).

(3) The neutral hypothesis of protein evolution is a statement about amino acid proportions, whereas the neutral hypothesis of genomic DNA evolution and the neutral hypothesis of functional DNA evolution are statements on nucleotide proportions across the entire genome and in functional regions, respectively.

(4) The neutralist–selectionist debate is a battle on two fronts – the proportion of substitutions resulting from positive selection (α) and the proportion of polymorphisms resulting from balancing selection (beta) – not the proportions of polymorphisms under directional (positive or purifying) selection.

(5) The neutralist–selectionist debate does not concern the question of whether truly neutral mutations exist, but the question of how often the selective value of a mutation permits genetic drift to override selection. If we acknowledge that genetic drift can govern the fate of non-conservative mutations, the neutral hypothesis has implications beyond the domain of molecular evolution, and predicts that certain phenotypic differences between species may be neutral.

(6) The phrase 'non-Darwinian evolution' is a misnomer. Neutral mutations indirectly facilitate the adaptation process, by providing an escape route out of local fitness optima. Evidence that certain traits have been fine-tuned by natural selection (such as codon usage or local mutation rates), should not be mistaken for evidence against the neutral theory.

(7) Many arguments against the neutral hypothesis – including arguments relating to the molecular clock and Lewontin's paradox – refute the null hypothesis of strict neutrality rather than any version of the neutral hypothesis. *Vice versa*, observations claimed to support the neutral hypothesis – such as the high substitution rates in non-functional regions and synonymous sites – are often equally consistent with the selectionist hypothesis.

(8) The neutral hypothesis of protein evolution and the hypothesis of functional DNA evolution are most clearly interpreted through meta-analyses of protein-coding genes and regulatory regions, rather than from indirect inferences obtained through evaluation of genome-wide patterns.

XII. AUTHOR CONTRIBUTIONS

This review has been written by MJ, under the supervision of CO, RH and AJ.

XIII. REFERENCES

Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in *Drosophila* DNA. *Genetics* **139**, 1067–1076.

Akashi, H. & Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 3695–3700.

Akashi, H., Osada, N. & Ohta, T. (2012). Weak selection and protein evolution. *Genetics* **192**, 15–31.

Akashi, H. & Schaeffer, S. W. (1997). Natural selection and the frequency distributions of 'silent' DNA polymorphism in *Drosophila*. *Genetics* **146**, 295–307.

Andolfatto, P. (2008). Controlling type-I error of the McDonald–Kreitman test in genomewide scans for selection on noncoding DNA. *Genetics* **180**, 1767–1771.

Arnheim, N. & Taylor, C. E. (1969). Non-Darwinian evolution: consequences for neutral allelic variation. *Nature* **223**, 900–903.

Ayala, F. J. (1974). Biological evolution: natural selection or random walk? *American Scientist* **62**, 692–701.

Ayala, F. J. & Gilpin, M. E. (1974). Gene frequency comparisons between taxa: support for the natural selection of protein polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 4847–4849.

Balloux, F. & Lehmann, L. (2012). Substitution rates at neutral genes depend on population size under fluctuating demography and overlapping generations. *Evolution International Journal of Organic Evolution* **66**, 605–611.

Bastolla, U., Vendruscolo, M. & Roman, H. E. (2000). Structurally constrained protein evolution: results from a lattice simulation. *European Physical Journal B* **15**, 385–397.

Begun, D. J. & Aquadro, C. F. (1991). Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the yellow-achaete region. *Genetics* **129**, 1147–1158.

Begun, D. J. & Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519–520.

Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y.-P., Hahn, M. W., Nista, P. M., Jones, C. D., Kern, A. D., Dewey, C. N., Pachter, L., Myers, E. & Langley, C. H. (2007). Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology* **5**, e310.

Berry, A. J., Ajioka, J. W. & Kreitman, M. (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**, 1111–1117.

Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H. E. & van Oosterhout, C. (2022). Genetic load: genomic estimates and applications in non-model animals. *Nature Reviews Genetics* **23**, 1–12.

Birky, C. W. & Walsh, J. B. (1988). Effects of linkage on rates of molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 6414–6418.

Bloom, J. D. & Arnold, F. H. (2009). In the light of directed evolution: pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9995–10000.

Bolívar, P., Guéguen, L., Duret, L., Ellegren, H. & Mugal, C. F. (2019). GC-biased gene conversion conceals the prediction of the nearly neutral theory in avian genomes. *Genome Biology* **20**, 5.

Booker, T. R., Jackson, B. C. & Keightley, P. D. (2017). Detecting positive selection in the genome. *BMC Biology* **15**, 98.

Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990). Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**, 1306–1310.

Brues, A. M. (1964). The cost of evolution vs. the cost of not evolving. *Evolution* **18**, 379–383.

Brues, A. M. (1969). Genetic load and its varieties. *Science* **164**, 1130–1136.

Buffalo, V. (2021). Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's paradox. *eLife* **10**, e67509.

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**, 897–907.

Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M. T., Glanowski, S., Tanenbaum, D. M., White, T. J., Sninsky, J. J., Hernandez, R. D., Civello, D., Adams, M. D., Cargill, M. & Clark, A. G. (2005). Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153–1157.

Cavalli-Sforza, L. L. (1966). Population structure and human evolution. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **164**, 362–379.

Chakraborty, R., Fuerst, P. A. & Nei, M. (1978). Statistical studies on protein polymorphism in natural populations II. Gene differentiation between populations. *Genetics* **88**, 367–390.

Chakraborty, R. & Hedrick, P. W. (1983). Heterozygosity and genetic distance of proteins. *Nature* **304**, 755.

Charlesworth, B. (2013). Why we are not dead one hundred times over. *Evolution International Journal of Organic Evolution* **67**, 3354–3361.

Charlesworth, B. & Charlesworth, D. (2018). Neutral variation in the context of selection. *Molecular Biology and Evolution* **35**, 1359–1361.

Charlesworth, B. & Jensen, J. D. (2021). Effects of selection at linked sites on patterns of genetic variability. *Annual Review of Ecology, Evolution, and Systematics* **52**, 177–197.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Charlesworth, J. & Eyre-Walker, A. (2007). The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proceedings of the National Academy of Sciences* **104**, 16992–16997.

Charlesworth, J. & Eyre-Walker, A. (2008). The McDonald-Kreitman test and slightly deleterious mutations. *Molecular Biology and Evolution* **25**, 1007–1015.

Chen, C. T. L., Wang, J. C. & Cohen, B. A. (2007). The strength of selection on ultraconserved elements in the human genome. *American Journal of Human Genetics* **80**, 692–704.

Chen, Q., Lan, A., Shen, X. & Wu, C.-I. (2019). Molecular evolution in small steps under prevailing negative selection: a nearly universal rule of codon substitution. *Genome Biology and Evolution* **11**, 2702–2712.

Clarke, B. (1970a). Darwinian evolution of proteins. *Science* **168**, 1009–1011.

Clarke, B. (1970b). Selective constraints on amino-acid substitutions during the evolution of proteins. *Nature* **228**, 159–160.

Comeron, J. M. (2017). Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **372**, e20160471.

Cooper, G. M., Stone, E. A., Asimenos, G., NISC Comparative Sequencing Program, Green, E. D., Batzoglou, S. & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research* **15**, 901–913.

Corbett-Detig, R. B., Hartl, D. L. & Sackton, T. B. (2015). Natural selection constrains neutral diversity across a wide range of species. *PLoS Biology* **13**, e1002112.

Crow, J. (1997). Motoo Kimura, 13 November 1924–13 November 1994. *Biographical Memoirs of Fellows of the Royal Society* **43**, 254–265.

Crow, J. F. (1972a). Darwinian and non-Darwinian evolution. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Darwinian, Neo-Darwinian, and non-Darwinian Evolution* **6**, 1–23.

Crow, J. F. (1972b). The dilemma of nearly neutral mutations. How important are they for evolution and human welfare? *The Journal of Heredity* **63**, 306–316.

Cutler, D. J. (2000). Understanding the overdispersed molecular clock. *Genetics* **154**, 1403–1417.

Cutter, A. D. & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews. Genetics* **14**, 262–274.

de Chadarevian, S. (1999). Protein sequencing and the making of molecular genetics. *Trends in Biochemical Sciences* **24**, 203–206.

de Jong, M. J., Lovatt, F. & Hoelzel, A. R. (2021). Detecting genetic signals of selection in heavily bottlenecked reindeer populations by comparing parallel founder events. *Molecular Ecology* **30**, 1642–1658.

Denver, D. R., Morris, K., Streelman, J. T., Kim, S. K., Lynch, M. & Thomas, W. K. (2005). The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nature Genetics* **37**, 544–548.

Dickerson, R. E. (1971). The structures of cytochrome c and the rates of molecular evolution. *Journal of Molecular Evolution* **1**, 26–45.

Dietrich, M. R. (1994). The origins of the neutral theory of molecular evolution. *Journal of the History of Biology* **27**, 21–59.

Drost, J. B. & Lee, W. R. (1995). Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environmental and Molecular Mutagenesis* **25**, 48–64.

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B.-K., Pauli, F., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.

Eddy, S. R. (2012). The C-value paradox, junk DNA and ENCODE. *Current Biology* **22**, 898–899.

Elliott, T. A. & Gregory, T. R. (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, e20140331.

Epstein, C. J. (1967). Non-randomness of amino-acid changes in the evolution of homologous proteins. *Nature* **215**, 355–359.

Ewens, W. J. (1972). The substitutional load in a finite population. *The American Naturalist* **106**, 273–282.

Eyre-Walker, A. (2002). Changing effective population size and the McDonald-Kreitman test. *Genetics* **162**, 2017–2024.

Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. *Trends in Ecology & Evolution* **21**, 569–575.

FAY, J. C., WYCKOFF, G. J. & WU, C.-I. (2002). Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**, 1024–1026.

FELSENSTEIN, J. (1971). On the biological significance of the cost of gene substitution. *The American Naturalist* **105**, 1–11.

FIGUET, E., NABHOLZ, B., BONNEAU, M., MAS CARRIO, E., NADACHOWSKA-BRZYSKA, K., ELLEGREN, H. & GALTIER, N. (2016). Life history traits, protein evolution, and the nearly neutral theory in amniotes. *Molecular Biology and Evolution* **33**, 1517–1527.

FISHER, R. A. (1931). XVII.—the distribution of gene ratios for rare mutations. *Proceedings of the Royal Society of Edinburgh* **50**, 204–220.

FITCH, W. M. & MARKOWITZ, E. (1970). An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochemical Genetics* **4**, 579–593.

FRANKHAM, R. (2012). How closely does genetic diversity in finite populations conform to predictions of neutral theory? Large deficits in regions of low recombination. *Heredity* **108**, 167–178.

FREESE, E. (1962). On the evolution of the base composition of DNA. *Journal of Theoretical Biology* **3**, 82–101.

GALTIER, N. (2016). Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genetics* **12**, e1005774.

GALTIER, N., PIGANEAU, G., MOUCHIROUD, D. & DURET, L. (2001). GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* **159**, 907–911.

GAVRILETS, S. (1997). Evolution and speciation on holey adaptive landscapes. *Trends in Ecology & Evolution* **12**, 307–312.

GILIS, D., MASSAR, S., CERF, N. J. & ROOMAN, M. (2001). Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biology* **2**, e0049.

GILLESPIE, J. H. (1984*a*). Molecular evolution over the mutational landscape. *Evolution International Journal of Organic Evolution* **38**, 1116–1129.

GILLESPIE, J. H. (1984*b*). The status of the neutral theory: the neutral theory of molecular evolution. *Science* **224**, 732–733.

GILLESPIE, J. H. (1986). Natural selection and the molecular clock. *Molecular Biology and Evolution* **3**, 138–155.

GILLESPIE, J. H. (1994). Substitution processes in molecular evolution. III. Deleterious alleles. *Genetics* **138**, 943–952.

GILLESPIE, J. H. (1995). On Ohta's hypothesis: most amino acid substitutions are deleterious. *Journal of Molecular Evolution* **40**, 64–69.

GILLESPIE, J. H. (2000). Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* **155**, 909–919.

GILLESPIE, J. H. (2001). Is the population size of a species relevant to its evolution? *Evolution* **55**, 2161–2169.

GOJOBORI, T., MORIYAMA, E. N. & KIMURA, M. (1990). Molecular clock of viral evolution, and the neutral theory. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 10015–10018.

GOODMAN, M. (1981). Globin evolution was apparently very rapid in early vertebrates: a reasonable case against the rate-constancy hypothesis. *Journal of Molecular Evolution* **17**, 114–120.

GRANTHAM, R. (1974). Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864.

GRAUR, D. (1985). Amino acid composition and the evolutionary rates of protein-coding genes. *Journal of Molecular Evolution* **22**, 53–62.

GRAUR, D. (2017). An upper limit on the functional fraction of the human genome. *Genome Biology and Evolution* **9**, 1880–1885.

GRAUR, D., ZHENG, Y., PRICE, N., AZEVEDO, R. B. R., ZUFALL, R. A. & ELHAIK, E. (2013). On the immortality of television sets: 'function' in the human genome according to the evolution-free gospel of ENCODE. *Genome Biology and Evolution* **5**, 578–590.

HAHN, M. (2018). *Molecular Population Genetics*. Oxford University Press, Oxford, New York.

HAHN, M. W. (2008). Toward a selection theory of molecular evolution. *Evolution* **62**, 255–265.

HALDANE, J. B. S. (1957). The cost of natural selection. *Journal of Genetics* **55**, 511.

HALLIGAN, D. L. & KEIGHTLEY, P. D. (2006). Ubiquitous selective constraints in the Drosophila genome revealed by a genome-wide interspecies comparison. *Genome Research* **16**, 875–884.

HARRIS, H. (1966). Enzyme polymorphisms in man. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **164**, 298–310.

HARRIS, K. (2018). The randomness that shapes our DNA. *eLife* **7**, e41491.

HARTL, D. L., DYKHUIZEN, D. E. & DEAN, A. M. (1985). Limits of adaptation: the evolution of selective neutrality. *Genetics* **111**, 655–674.

HARTL, D. L. & TAUBES, C. H. (1996). Compensatory nearly neutral mutations: selection without adaptation. *Journal of Theoretical Biology* **182**, 303–309.

HEDGES, S. B., MARIN, J., SULESKI, M., PAYMER, M. & KUMAR, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution* **32**, 835–845.

HENDRIKS, W., LEUNISSEN, J., NEVO, E., BLOEMENDAL, H. & DE JONG, W. W. (1987). The lens protein alpha A-crystallin of the blind mole rat, *Spalax ehrenbergi*: evolutionary change and functional constraints. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 5320–5324.

HEY, J. (1999). The neutralist, the fly and the selectionist. *Trends in Ecology & Evolution* **14**, 35–38.

HICKEY, D. & GOLDING, B. (2019). Sex solves Haldane's Dilemma. *Genome* **62**, 761–768.

HIGGINS, E. T. (2004). Making a theory useful: lessons handed down. *Personality and Social Psychology Review* **8**, 138–145.

HILL, W. G. & ROBERTSON, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research* **8**, 269–294.

HO, W.-C., OHYA, Y. & ZHANG, J. (2017). Testing the neutral hypothesis of phenotypic evolution. *Proceedings of the National Academy of Sciences* **114**, 12219–12224.

HODGINS-DAVIS, A., RICE, D. P. & TOWNSEND, J. P. (2015). Gene expression evolves under a house-of-cards model of stabilizing selection. *Molecular Biology and Evolution* **32**, 2130–2140.

HODGKINSON, A. & EYRE-WALKER, A. (2011). Variation in the mutation rate across mammalian genomes. *Nature Reviews Genetics* **12**, 756–766.

HUBBY, J. L. & LEWONTIN, R. C. (1966). A molecular approach to the study of genic heterozygosity in natural populations. I. the number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**, 577–594.

HUDSON, R. R. & KAPLAN, N. L. (1995). Deleterious background selection with recombination. *Genetics* **141**, 1605–1617.

HUDSON, R. R., KREITMAN, M. & AGUADÉ, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

HUGHES, A. L. (2008). Near-neutrality: the leading edge of the neutral theory of molecular evolution. *Annals of the New York Academy of Sciences* **1133**, 162–179.

HUGHES, A. L. & NEI, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170.

HUGHES, A. L., PACKER, B., WELCH, R., BERGEN, A. W., CHANOCK, S. J. & YEAGER, M. (2003). Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 15754–15757.

HURST, L. D. & SMITH, N. G. (1999). Do essential genes evolve slowly? *Current Biology* **9**, 747–750.

HUYNEN, M. A. (1996). Exploring phenotype space through neutral evolution. *Journal of Molecular Evolution* **43**, 165–169.

IKEMURA, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *Journal of Molecular Biology* **146**, 1–21.

INGRAM, V. M. (1963). *The Hemoglobins in Genetics and Evolution*. Columbia University Press, New York and London.

JENSEN, J. D., PAYSEUR, B. A., STEPHAN, W., AQUADRO, C. F., LYNCH, M., CHARLESWORTH, D. & CHARLESWORTH, B. (2019). The importance of the Neutral Theory in 1968 and 50 years on: a response to Kern and Hahn 2018. *Evolution* **73**, 111–114.

JOHRI, P., RIALL, K., BECHER, H., EXCOFFIER, L., CHARLESWORTH, B. & JENSEN, J. D. (2021). The impact of purifying and background selection on the inference of population history: problems and prospects. *Molecular Biology and Evolution* **38**, 2986–3003.

JUKES, T. H. (1978). Neutral changes during divergent evolution of hemoglobins. *Journal of Molecular Evolution* **11**, 267–269.

JUKES, T. H. (1991). Early development of the neutral theory. *Perspectives in Biology and Medicine* **34**, 473–485.

KAPLAN, N. L., HUDSON, R. R. & LANGLEY, C. H. (1989). The 'hitchhiking effect' revisited. *Genetics* **123**, 887–899.

KAUFFMAN, S. & LEVIN, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology* **128**, 11–45.

KEIGHTLEY, P. D. & EYRE-WALKER, A. (2010). What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 1187–1193.

KERN, A. D. & BEGUN, D. J. (2005). Patterns of polymorphism and divergence from noncoding sequences of *Drosophila melanogaster* and *D. simulans*: evidence for nonequilibrium processes. *Molecular Biology and Evolution* **22**, 51–62.

KERN, A. D. & HAHN, M. W. (2018). The neutral theory in light of natural selection. *Molecular Biology and Evolution* **35**, 1366–1371.

KESSLER, M. D. & DEAN, M. D. (2014). Effective population size does not predict codon usage bias in mammals. *Ecology and Evolution* **4**, 3887–3900.

KHAITOVICH, P., WEISS, G., LACHMANN, M., HELLMANN, I., ENARD, W., MUETZEL, B., WIRKNER, U., ANSORGE, W. & PÄÄBO, S. (2004). A neutral model of transcriptome evolution. *PLoS Biology* **2**, e132.

KIM, S.-H. & YI, S. V. (2008). Mammalian nonsynonymous sites are not overdispersed: comparative genomic analysis of index of dispersion of mammalian proteins. *Molecular Biology and Evolution* **25**, 634–642.

KIMURA, M. (1957). Some problems of stochastic processes in genetics. *The Annals of Mathematical Statistics* **28**, 882–901.

KIMURA, M. (1960a). Optimum mutation rate and degree of dominance as determined by the principle of minimum genetic load. *Journal of Genetics* **57**, 21–34.

KIMURA, M. (1960b). Relative applicability of the classical and the balance hypotheses to man. Especially with respect to quantitative characters. *Journal of Radiation Research* **1**, 155–164.

KIMURA, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719.

KIMURA, M. (1967). On the evolutionary adjustment of spontaneous mutation rates. *Genetics Research* **9**, 23–34.

KIMURA, M. (1968a). Evolutionary rate at the molecular level. *Nature* **217**, 624–626.

KIMURA, M. (1968b). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research* **11**, 247–269.

KIMURA, M. (1969a). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903.

KIMURA, M. (1969b). The rate of molecular evolution considered from the standpoint of population genetics. *Proceedings of the National Academy of Sciences of the United States of America* **63**, 1181–1188.

KIMURA, M. (1976). The neutral theory as a supplement to Darwinism. *Trends in Biochemical Sciences* **1**, 248–249.

KIMURA, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275–276.

KIMURA, M. (1979a). Model of effectively neutral mutations in which selective constraint is incorporated. *Proceedings of the National Academy of Sciences* **76**, 3440–3444.

KIMURA, M. (1979b). The neutral theory of molecular evolution. *Scientific American* **241**, 98–126.

KIMURA, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111–120.

KIMURA, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, New York.

KIMURA, M. (1988). Thirty years of population genetics with Dr. Crow. *Idengaku Zasshi* **63**, 1–10.

KIMURA, M. (1991a). Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 5969–5973.

KIMURA, M. (1991b). The neutral theory of molecular evolution: a review of recent evidence. *Idengaku Zasshi* **66**, 367–386.

KIMURA, M. & CROW, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

KIMURA, M. & OHTA, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771.

KIMURA, M. & OHTA, T. (1971a). On the rate of molecular evolution. *Journal of Molecular Evolution* **1**, 1–17.

KIMURA, M. & OHTA, T. (1971b). Population genetics, molecular biometry and evolution. In *Darwinian, Neo-Darwinian, and Non-Darwinian Evolution* (ed. J. F. CROW), pp. 43–68. University of California Press, Berkeley.

KIMURA, M. & OHTA, T. (1971c). Protein polymorphism as a phase of molecular evolution. *Nature* **229**, 467–469.

KIMURA, M. & OHTA, T. (1974). On some principles governing molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 2848–2852.

KING, J. (1983). This week's citation classic. *Current Contents* **34**, 25.

KING, J. L. (1967). Continuously distributed factors affecting fitness. *Genetics* **55**, 483–492.

KING, J. L. & JUKES, T. H. (1969). Non-Darwinian evolution. *Science* **164**, 788–798.

KOHNE, D. E., CHISCON, J. A. & HOYER, B. H. (1972). Evolution of primate DNA sequences. *Journal of Human Evolution* **1**, 627–644.

KOMIYAMA, N. H., MIYAZAKI, G., TAME, J. & NAGAI, K. (1995). Transplanting a unique allosteric effect from crocodile into human haemoglobin. *Nature* **373**, 244–246.

KONDRASHOV, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences* **279**, 5048–5057.

KOSIOL, C., VINAR, T., DA FONSECA, R. R., HUBISZ, M. J., BUSTAMANTE, C. D., NIELSEN, R. & SIEPEL, A. (2008). Patterns of positive selection in six mammalian genomes. *PLoS Genetics* **4**, e1000144.

KREITMAN, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412–417.

KREITMAN, M. (1996). The neutral theory is dead. Long live the neutral theory. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* **18**, 678–683.

LAIRD, C. D., McCONAUGHY, B. L. & McCARTHY, B. J. (1969). Rate of fixation of nucleotide substitutions in evolution. *Nature* **224**, 149–154.

LAND, M., HAUSER, L., JUN, S.-R., NOOKAEW, I., LEUZE, M. R., AHN, T.-H., KARPINETS, T., LUND, O., KORA, G., WASSENAAR, T., POUDEL, S. & USSERY, D. W. (2015). Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics* **15**, 141–161.

LANDE, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314–334.

LANDE, R. (1998). Risk of population extinction from fixation of deleterious and reverse mutations. *Genetica* **102–103**, 21–27.

LANGLEY, C. H. & FITCH, W. M. (1974). An examination of the constancy of the rate of molecular evolution. *Journal of Molecular Evolution* **3**, 161–177.

LANKS, K. W. & KITCHIN, F. D. (1970). Conservative mutations in homologous proteins. *Nature* **226**, 753–754.

LEMOS, B., MEIKLEJOHN, C. D., CÁCERES, M. & HARTL, D. L. (2005). Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* **59**, 126–137.

LENORMAND, T., ROZE, D. & ROUSSET, F. (2009). Stochasticity in evolution. *Trends in Ecology & Evolution* **24**, 157–165.

LEWONTIN, R. C. (1974). *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

LEWONTIN, R. C. & HUBBY, J. L. (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595–609.

LEWONTIN, R. C. & KRAKAUER, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195.

LI, J., ZHOU, J., WU, Y., YANG, S. & TIAN, D. (2015). GC-content of synonymous codons profoundly influences amino acid usage. *G3: Genes|Genomes|Genetics* **5**, 2027–2036.

LI, W.-H., GOJOBORI, T. & NEI, M. (1981). Pseudogenes as a paradigm of neutral evolution. *Nature* **292**, 237–239.

LI, W. H. & NEI, M. (1977). Persistence of common alleles in two related populations or species. *Genetics* **86**, 901–914.

LIPMAN, D. J. & WILBUR, W. J. (1991). Modelling neutral and selective evolution of protein folding. *Proceedings: Biological Sciences* **245**, 7–11.

LOHMUELLER, K. E., ALBRECHTSEN, A., LI, Y., KIM, S. Y., KORNELIUSSEN, T., VINCKENBOSCH, N., TIAN, G., HUERTA-SANCHEZ, E., FEDER, A. F., GRARUP, N., JØRGENSEN, T., JIANG, T., WITTE, D. R., SANDBÆK, A., HELLMANN, I., ET AL. (2011). Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genetics* **7**, e1002326.

LYNCH, M. (1988). The rate of polygenic mutation. *Genetics Research* **51**, 137–148.

LYNCH, M. (2010). Evolution of the mutation rate. *Trends in Genetics* **26**, 345–352.

LYNCH, M., ACKERMAN, M. S., GOUT, J.-F., LONG, H., SUNG, W., THOMAS, W. K. & FOSTER, P. L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics* **17**, 704–714.

LYNCH, M., CONERY, J. & BURGER, R. (1995). Mutational meltdowns in sexual populations. *Evolution* **49**, 1067–1080.

LYNCH, M. & CONERY, J. S. (2003). The origins of genome complexity. *Science* **302**, 1401–1404.

LYNCH, M. & HILL, W. G. (1986). Phenotypic evolution by neutral mutation. *Evolution* **40**, 915–935.

MARGOLIASH, E. (1963). Primary structure and evolution of cytochrome c. *Proceedings of the National Academy of Sciences of the United States of America* **50**, 672–679.

MARGOLIASH, E. & SMITH, E. L. (1965). Structural and functional aspects of cytochrome c in relation to evolution. In *Evolving Genes and Proteins* (eds V. BRYSON and H. J. VOGEL), pp. 221–242. Academic Press, New York.

MARTINCORENA, I. & LUSCOMBE, N. M. (2013). Non-random mutation: the evolution of targeted hypermutation and hypomutation. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* **35**, 123–130.

MAYNARD-SMITH, J. (1968). 'Haldane's dilemma' and the rate of evolution. *Nature* **219**, 1114–1116.

MAYNARD-SMITH, J. (1970a). Natural selection and the concept of a protein space. *Nature* **225**, 563–564.

MAYNARD-SMITH, J. (1970b). Population size, polymorphism, and the rate of non-Darwinian evolution. *The American Naturalist* **104**, 231–237.

MAYNARD-SMITH, J. (1978). Optimization theory in evolution. *Annual Review of Ecology and Systematics* **9**, 31–56.

MAYNARD-SMITH, J. & HAIGH, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23–35.

McDONALD, J. & KREITMAN, M. (1991). Adaptive protein evolution at the Adh locus in *drosophila*. *Nature* **351**, 652–654.

MILKMAN, R. D. (1967). Heterosis as a major cause of heterozygosity in nature. *Genetics* **55**, 493–495.

MIYATA, T. & YASUNAGA, T. (1981). Rapidly evolving mouse alpha-globin-related pseudo gene and its evolutionary history. *Proceedings of the National Academy of Sciences* **78**, 450–453.

MONROE, J. G., SRIKANT, T., CARBONELL-BEJERANO, P., BECKER, C., LENSINK, M., EXPOSITO-ALONSO, M., KLEIN, M., HILDEBRANDT, J., NEUMANN, M., KLIEBENSTEIN, D., WENG, M.-L., IMBERT, E., ÅGREN, J., RUTTER, M. T., FENSTER, C. B., ET AL. (2022). Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature* **602**, 101–105.

Moutinho, A. F., Bataillon, T. & Dutheil, J. Y. (2020). Variation of the adaptive substitution rate between species and within genomes. *Evolutionary Ecology* **34**, 315–338.

Mugal, C. F., Wolf, J. B. W. & Kaj, I. (2014). Why time matters: codon evolution and the temporal dynamics of dN/dS. *Molecular Biology and Evolution* **31**, 212–231.

Muller, H. J. (1950). Our load of mutations. *American Journal of Human Genetics* **2**, 111–176.

Müller, R., Kaj, I. & Mugal, C. F. (2022). A nearly neutral model of molecular signatures of natural selection after change in population size. *Genome Biology and Evolution* **14**, 1–17.

Muñoz-Gómez, S. A., Bilolikar, G., Wideman, J. G. & Geiler-Samerotte, K. (2021). Constructive neutral evolution 20 years later. *Journal of Molecular Evolution* **89**, 172–182.

Murga-Moreno, J., Coronado-Zamora, M., Hervas, S., Casillas, S. & Barbadilla, A. (2019). iMKT: the integrative McDonald and Kreitman test. *Nucleic Acids Research* **47**, 283–288.

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K. & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics* **8**, e1002764.

Nei, M. (2005a). Bottlenecks, genetic polymorphism and speciation. *Genetics* **170**, 1–4.

Nei, M. (2005b). Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution* **22**, 2318–2342.

Nei, M. (2007). The new mutation theory of phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12235–12242.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.

Nei, M., Maruyama, T. & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1–10.

Nei, M., Suzuki, Y. & Nozawa, M. (2010). The neutral theory of molecular evolution in the genomic era. *Annual Review of Genomics and Human Genetics* **11**, 265–289.

Nielsen, R. & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936.

Nunney, L. (2003). The cost of natural selection revisited. *Annales Zoologici Fennici* **40**, 185–194.

Ohta, T. (1972a). Evolutionary rate of cistrons and DNA divergence. *Journal of Molecular Evolution* **1**, 150–157.

Ohta, T. (1972b). Fixation probability of a mutant influenced by random fluctuation of selection intensity. *Genetical Research* **19**, 33–38.

Ohta, T. (1972c). Population size and rate of evolution. *Journal of Molecular Evolution* **1**, 305–314.

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98.

Ohta, T. (1976). Role of very slightly deleterious mutations in molecular evolution and polymorphism. *Theoretical Population Biology* **10**, 254–275.

Ohta, T. (1977). Extension of the neutral mutation drift hypothesis. In *Molecular Evolution and Polymorphism* (ed M. Kimura). National Institute of Genetics, Mishima.

Ohta, T. (2002). Near-neutrality in evolution of genes and gene regulation. *Proceedings of the National Academy of Sciences* **99**, 16134–16137.

Ohta, T. (2003). Origin of the neutral and nearly neutral theories of evolution. *Journal of Biosciences* **28**, 371–377.

Ohta, T. & Gillespie, J. H. (1996). Development of neutral and nearly neutral theories. *Theoretical Population Biology* **49**, 128–142.

Ohta, T. & Kimura, M. (1971a). Amino acid composition of proteins as a product of molecular evolution. *Science* **174**, 150–153.

Ohta, T. & Kimura, M. (1971b). Genetic load due to small mutations with very small effects. *The Japanese Journal of Genetics* **46**, 393–401.

Ohta, T. & Kimura, M. (1971c). On the constancy of the evolutionary rate of cistrons. *Journal of Molecular Evolution* **1**, 18–25.

Ohta, T. & Tachida, H. (1990). Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* **126**, 219–229.

Okazaki, A., Yamazaki, S., Inoue, I. & Ott, J. (2021). Population genetics: past, present, and future. *Human Genetics* **140**, 231–240.

Parsch, J., Zhang, Z. & Baines, J. F. (2009). The influence of demography and weak selection on the McDonald–Kreitman test: an empirical study in *Drosophila*. *Molecular Biology and Evolution* **26**, 691–698.

Perutz, M. F. (1984). Species adaptation in a protein molecule. *Advances in Protein Chemistry* **36**, 213–244.

Phung, T. N., Huber, C. D. & Lohmueller, K. E. (2016). Determining the effect of natural selection on linked neutral divergence across species. *PLoS Genetics* **12**, e1006199.

Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. & Tans, S. J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386.

Pollak, E. (1982). The rate of mutant substitution in populations with overlapping generations. *Genetical Research* **40**, 89–94.

Poon, A. & Otto, S. P. (2000). Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution* **54**, 1467–1479.

Popadin, K., Polishchuk, L. V., Mamirova, L., Knorre, D. & Gunbin, K. (2007). Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proceedings of the National Academy of Sciences* **104**, 13390–13395.

Pouyet, F., Aeschbacher, S., Thiéry, A. & Excoffier, L. (2018). Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *eLife* **7**, e36317.

Prakash, S., Lewontin, R. C. & Hubby, J. L. (1969). A molecular approach to the study of genic heterozygosity in natural populations IV. Patterns of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura*. *Genetics* **61**, 841–858.

Rands, C. M., Meader, S., Ponting, C. P. & Lunter, G. (2014). 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genetics* **10**, e1004525.

Razeto-Barry, P., Díaz, J. & Vásquez, R. A. (2012). The nearly neutral and selection theories of molecular evolution under the Fisher geometrical framework: substitution rate, population size, and complexity. *Genetics* **191**, 523–534.

Reznick, D. (2016). Hard and soft selection revisited: how evolution by natural selection works in the real world. *The Journal of Heredity* **107**, 3–14.

Richmond, R. C. (1970). Non-Darwinian evolution: a critique. *Nature* **225**, 1025–1028.

Roff, D. A. (1998). Evolution of threshold traits: the balance between directional selection, drift and mutation. *Heredity* **80**, 25–32.

Sarich, V. M. & Wilson, A. C. (1967). Immunological time scale for hominid evolution. *Science* **158**, 1200–1203.

Sarich, V. M. & Wilson, A. C. (1973). Generation time and genomic evolution in primates. *Science* **179**, 1144–1147.

Scott, A., Petrykowska, H. M., Hefferon, T., Gotea, V. & Elnitski, L. (2012). Functional analysis of synonymous substitutions predicted to affect splicing of the CFTR gene. *Journal of Cystic Fibrosis* **11**, 511–517.

Sella, G., Petrov, D. A., Przeworski, M. & Andolfatto, P. (2009). Pervasive natural selection in the *Drosophila* genome? *PLoS Genetics* **5**, e1000495.

Sharp, P. M. & Li, W. H. (1986). Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Research* **14**, 7737–7749.

Simpson, G. G. (1964). Organisms and molecules in evolution. *Science* **146**, 1535–1538.

Skibinski, D. O. F. & Ward, R. D. (1982). Correlations between heterozygosity and evolutionary rate of proteins. *Nature* **298**, 490–492.

Smith, N. G. C. & Eyre-Walker, A. (2002). Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024.

Sniegowski, P. D., Gerrish, P. J., Johnson, T. & Shaver, A. (2000). The evolution of mutation rates: separating causes from consequences. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* **22**, 1057–1066.

Sonneborn, T. M. (1965). Degeneracy of the genetic code: extent, nature, and genetic implications. In *Evolving Genes and Proteins* (eds V. Bryson and H. J. Vogel), pp. 377–397. Academic Press, New York.

Spielman, S. J. & Wilke, C. O. (2015). The relationship between dN/dS and scaled selection coefficients. *Molecular Biology and Evolution* **32**, 1097–1108.

Stoletzki, N. & Eyre-Walker, A. (2011). Estimation of the neutrality index. *Molecular Biology and Evolution* **28**, 63–70.

Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *Journal of Molecular Evolution* **49**, 169–181.

Stone, W. S., Wheeler, M. R., Johnson, F. M. & Kojima, K. I. (1968). Genetic variation in natural Island populations of members of the *Drosophila nasuta* and *Drosophila ananassae* subgroups. *Proceedings of the National Academy of Sciences of the United States of America* **59**, 102–109.

Sturtevant, A. H. (1937). Essays on evolution. I. on the effects of selection on mutation rate. *The Quarterly Review of Biology* **12**, 464–467.

Subramanian, S. (2008). Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes. *Genetics* **178**, 2429–2432.

Sueoka, N. (1962). On the genetic basis of variation and heterogeneity of DNA base composition. *Proceedings of the National Academy of Sciences of the United States of America* **48**, 582–592.

Sueoka, N. (1988). Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 2653–2657.

Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G. & Lynch, M. (2012). Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences* **109**, 18488–18492.

Sved, J. A. (1968). Possible rates of gene substitution in evolution. *The American Naturalist* **102**, 283–293.

Sved, J. A., Reed, T. E. & Bodmer, W. F. (1967). The number of balanced polymorphisms that can be maintained in a natural population. *Genetics* **55**, 469–481.

Tajima, F. (1996). The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics* **143**, 1457–1465.

Tang, H., Wyckoff, G. J., Lu, J. & Wu, C.-I. (2004). A universal evolutionary index for amino acid changes. *Molecular Biology and Evolution* **21**, 1548–1556.

TEIXEIRA, J. C. & HUBER, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences* **118**, e2015096118.

van NIMWEGEN, E. & CRUTCHFIELD, J. P. (2000). Metastable evolutionary dynamics: crossing fitness barriers or escaping via neutral paths? *Bulletin of Mathematical Biology* **62**, 799–848.

VAN VALEN, L. (1963). Haldane's dilemma, evolutionary rates, and heterosis. *The American Naturalist* **97**, 185–190.

VAN VALEN, L. (1974). Molecular evolution as predicted by natural selection. *Journal of Molecular Evolution* **3**, 89–101.

van't HOF, A. E., CAMPAGNE, P., RIGDEN, D. J., YUNG, C. J., LINGLEY, J., QUAIL, M. A., HALL, N., DARBY, A. C. & SACCHERI, I. J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**, 102–105.

WAGNER, A. (2008). Neutralism and selectionism: a network-based reconciliation. *Nature Reviews Genetics* **9**, 965–974.

WALLACE, B. (1975). Hard and soft selection revisited. *Evolution* **29**, 465–473.

WARD, R. D. & SKIBINSKI, D. O. (1985). Observed relationships between protein heterozygosity and protein genetic distance and comparisons with neutral expectations. *Genetical Research* **45**, 315–340.

WEBER, C. C., NABHOLZ, B., ROMIGUIER, J. & ELLEGREN, H. (2014). Kr/Kc but not dN/dS correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biology* **15**, 542.

WEBER, C. C. & WHELAN, S. (2019). Physicochemical amino acid properties better describe substitution rates in large populations. *Molecular Biology and Evolution* **36**, 679–690.

WEISSMAN, D. B. & BARTON, N. H. (2012). Limits to the rate of adaptive substitution in sexual populations. *PLoS Genetics* **8**, e1002740.

WELLER, C. & WU, M. (2015). A generation-time effect on the rate of molecular evolution in bacteria. *Evolution* **69**, 643–652.

WHITLOCK, M. C., GRISWOLD, C. K. & PETERS, A. D. (2003). Compensating for the meltdown: the critical effective size of a population with deleterious and compensatory mutations. *Annales Zoologici Fennici* **40**, 169–183.

WILSON, A. C., CARLSON, S. S. & WHITE, T. J. (1977). Biochemical evolution. *Annual Review of Biochemistry* **46**, 573–639.

WOLF, J. B. W., KÜNSTNER, A., NAM, K., JAKOBSSON, M. & ELLEGREN, H. (2009). Nonlinear dynamics of nonsynonymous (dN) and synonymous (dS) substitution rates affects inference of selection. *Genome Biology and Evolution* **1**, 308–319.

WRIGHT, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the XI International Congress of Genetics* **8**, 209–222.

XIA, X. & LI, W. H. (1998). What amino acid properties affect protein evolution? *Journal of Molecular Evolution* **47**, 557–564.

YAN, D., LUO, X., TANG, J., XU, S., HUANG, K., WANG, X., FENG, T., QUE, T., JIA, M., GUO, X., REHMAN, S., LI, Z., YANG, Y., LI, K., CUI, K., ET AL. (2023). High-quality genomes of pangolins: insights into the molecular basis of scale formation and adaption to myrmecophagous diet. *Molecular Biology and Evolution* **40**, msac262.

YANG, Z. & NIELSEN, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution* **46**, 409–418.

YANG, Z. & NIELSEN, R. (2008). Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Molecular Biology and Evolution* **25**, 568–579.

YANG, Z., WONG, W. S. W. & NIELSEN, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* **22**, 1107–1118.

ZHANG, J. (2018). Neutral theory and phenotypic evolution. *Molecular Biology and Evolution* **35**, 1327–1331.

ZHANG, J. & YANG, J.-R. (2015). Determinants of the rate of protein sequence evolution. *Nature Reviews. Genetics* **16**, 409–420.

ZUCKERKANDL, E. (1976). Evolutionary processes and evolutionary noise at the molecular level. I. Functional density in proteins. *Journal of Molecular Evolution* **7**, 167–183.

ZUCKERKANDL, E. & PAULING, L. (1965). Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins* (eds V. BRYSIN and H. J. VOGEL), pp. 97–166. Academic Press, New York.