Trends in Ecology & Evolution

CellPress REVIEWS

Review

The Importance of Genetic Redundancy in Evolution

Áki J. Láruson,[1,3,*,@] Sam Yeaman,[2,@] and Katie E. Lotterhos[1,3,@]

Genetic redundancy has been defined in many different ways at different levels of biological organization. Here, we briefly review the general concept of redundancy and focus on the evolutionary importance of redundancy in terms of the number of genotypes that give rise to the same phenotype. We discuss the challenges in determining redundancy empirically, with published experimental examples, and demonstrate the use of the *C*-score metric to quantify redundancy in evolution studies. We contrast the implicit assumptions of redundancy in quantitative versus population genetic models, show how this contributes to signatures of allele frequency shifts, and highlight how the rapid accumulation of genome-wide association data provides an avenue for further understanding the prevalence and role of redundancy in evolution.

Highlights

The use of the term 'genetic redundancy' has changed over the past 100 years.

The amount of redundancy in the mapping of genotype to phenotype is a critical parameter for evolutionary outcomes under a range of models.

Distinguishing the conceptual terms 'genotypic redundancy' and 'segregating redundancy' promotes clearer language to further the study of redundancy at all levels of evolutionary biology.

Empirical determination of redundancy is challenging, but there are approaches which allow for the quantitative inference of redundancy.

The *C*-score measure of evolutionary repeatability is correlated with two different definitions of redundancy and can be applied to empirical datasets.

History of Redundancy in Evolutionary Studies

Redundancy describes a situation in which there is an excess of causal components in a system, above the minimum needed for its proper function. Understanding how redundancy is built into biological systems has critical implications for how we study the way life evolves and persists. Redundancy implies flexibility and the potential for multiple solutions to a given problem, which can greatly impact the outcome of evolutionary processes. However, there are many different levels at which redundancy can be considered with respect to genetics and evolution and it is important to clarify how these various levels interact. The concept of redundancy in the context of genetics has been discussed for nearly a century, but the use of the term 'genetic redundancy' has multiple different implicit meanings across different studies. In clarifying specific applications of the term, and discussing the varying biological levels where it is relevant, we hope to promote clearer language to further the discussion of the role of redundancy in evolutionary biology.

Early use of the term 'genetic redundancy' came from studies on gene duplication, where it was used to describe the redundancy within a genome arising from more than one copy of the same gene. The phenomenon of gene duplication was first recognized following Alfred Sturtevant's discovery of unequal crossing over while studying the *Bar* mutation in *Drosophila* [1–4]. R.A. Fisher eventually suggested that multicopy loci could allow for the accumulation of beneficial mutations, as opposed to just 'sheltering' of deleterious mutations [5]. The discussion on the nature and impact of gene duplication in evolution continued when S.G. Stephens published an article titled 'Possible Significance of Duplication in Evolution' [6], where it was suggested that duplication provides yet more raw material upon which evolution can act. The exact term 'genetic redundancy' was being used variously to describe instances of gene duplication as early as the 1960s [7,8]. Seminal work on the contribution of redundant genetic elements to evolution, regardless of their origin, was published by Nowak *et al*. in the late 90s [9].

These early uses of 'genetic redundancy' hinge on a specific understanding that we define as functional redundancy: a situation where 'two or more genes perform the same **biochemical function**' (see Glossary) [10], or 'can partially or fully substitute for the function of the other'

[1]Northeastern University Marine Science Center, 430 Nahant Rd, Nahant, MA 01908, USA
[2]University of Calgary, Department of Biological Sciences, 507 Campus Drive NW, Calgary, AB, T2N 1N4, Canada
[3]https://sites.google.com/site/katielotterhos/

*Correspondence: a.laruson@northeastern.edu (Á.J. Láruson).
@Twitter: @AkiJarl (Á.J. Láruson), @sam_yeaman (S. Yeaman), and @DrK_Lo (K.E. Lotterhos).

[11], within an individual. More generally, we consider two causal genetic elements functionally redundant if a loss-of-function variant could occur in one element with no discernible changes to the phenotype of interest, providing the other element still functioned (Table 1). In more recent theoretical studies, however, the term 'genetic redundancy' has been used to describe the situation where more than one combination of genetic variants produces the same phenotype (or overall fitness) over an evolutionary timescale [12–17]. We define this phenomenon as genotypic redundancy: the number of possible **causal genotypes** that would yield the same phenotype or trait value over time (Table 1) [13]. Note that there may be no genotypic-trait redundancy (i.e., only one genotype produced a given trait value), but high genotypic-fitness redundancy (multiple genotypes produce different phenotypes with the same fitness value), or vice versa. In the simple conceptual example illustrated in Figure 1, over evolutionary time scales there are seven possible genotypes that can be achieved through mutation. Some of these genotypes exhibit genotype-trait redundancy, while other genotypes exhibit genotype-fitness redundancy [18]. Genotype-trait and -fitness redundancy do not correspond when a genotype has deleterious pleiotropic or epistatic effects on other traits. In order to distinguish between the potential polygenic redundancy that could arise through mutations (i.e., genotypic redundancy) and the actual number of redundant causal genotypes that are segregating within a population at a given time point, we will use the term segregating redundancy to refer to the latter.

Whereas functional redundancy refers to the interchangeability of genetic elements in their biochemical functions within an individual, genotypic redundancy refers to the number of genotypes that could possibly yield a given phenotype or fitness in a population. Both functional and genotypic redundancy are closely related to **evolvability** [19] and the perturbability of a phenotype to new mutations [20]. Functional redundancy is in particular related to genetic/mutational **robustness** [21–24], as it describes robustness to a loss of function mutation. When this robustness evolves directly as a response to natural selection, the terms genetic canalization and assimilation [25–27], as well as genetic buffering/degeneracy [28–30], could be applied. While robustness and functional redundancy are concerned with the perturbability of a phenotype, genotypic redundancy refers to the potential variation that could evolve. It's important to highlight that genotypic redundancy is based on the number of possible genotypes that yield a phenotype

Table 1. Definitions of Different Types of Genetic Redundancy.

Term	Level and scale	Definitions
Functional redundancy	Within individual at a particular time	A situation where two or more genes perform the same biochemical function [10], or can partially or fully substitute for the function of the other [11]. Describes the case when a loss-of-function variant has no discernible effect on the phenotype.
Genotypic redundancy	Within population over evolutionary time	When more than one genotype can produce the optimal phenotype [12], or when more than one genotype gives the same phenotype or fitness [13,18]. Includes all possible genotypes that could arise by mutation, but that are not yet present in the population.
Segregating redundancy	Within population at a particular time	Number of redundant genotypes segregating in the population, which is a subset of the possible genotypic redundancy that has arisen from evolutionary processes.
Redundant genotypes	Within population at a particular time	Different genotypes that produce the same trait or the same fitness.
Genetic redundancy	May refer to any of the above	

Glossary

Biochemical function: the biological process to which the product of a gene contributes.
Causal genotypes: genotypes that are a part of the functional genome and contribute to a trait or fitness under selection.
Complex trait: a measurable trait that exhibits a continuous scale of variation.
Constraint: context dependent, but here taken to mean the limiting of possible combinations of genotypes that produce a particular phenotype or fitness value.
Convergence: when two or more evolutionary lineages acquire the same or similar traits, independently.
Diminishing returns epistasis: a phenomenon where the trait/fitness effects of individual mutations become nonadditive when co-occurring in an individual.
Evolvability: the capacity of a system to produce adaptive genetic diversity.
Gene family: a set of genes of similar function, likely originating from duplication events.
Gene regulatory network: a suite of molecular factors influencing the expression of genes and the amount of gene products produced.
Genetic architecture: the pattern of genetic effects that build and control a given phenotypic character and its variational properties; can refer to the number of sites with mutations that contribute to the phenotype, their effect sizes, linkage relationships, as well as patterns of pleiotropy, dominance, and epistasis.
Genome wide association study (GWAS): an observation study of multiple genomes for the purpose of associating variants to an observable trait.
Genomic repeatability: the extent to which the same genetic changes are observed in independent replicates evolving under similar conditions.
Genotype–phenotype map: the documented association between a genotype and a phenotype from empirical studies. Similar association maps between genotypes and fitness measures can be made.
Michaelis-Menten dynamics: the reduction in the rate of a reaction as the concentration of a substrate increases.
Quantitative trait loci (QTL): genetic markers correlated with an observed quantitative trait in a population.

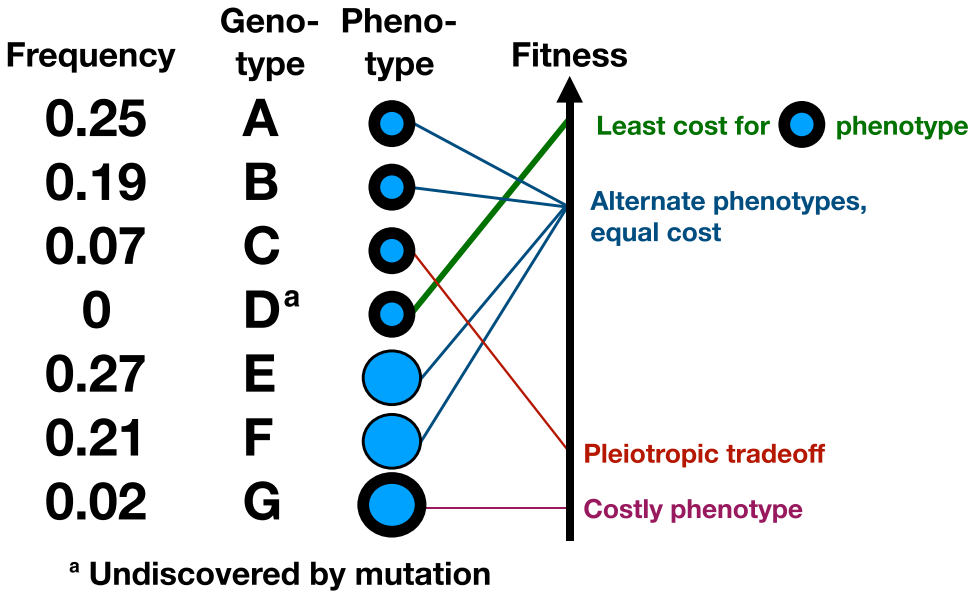

Figure 1. Genotypic and Segregating Redundancy. In this example, seven possible genotypes, occurring over an evolutionary time scale, contribute to one of three possible phenotypes ('small' blue cells with 'thick' cell walls, 'large' blue cells with 'thin' cell walls, or 'large' blue cells with 'thick' cell walls). The sets of genotypes (A, B, C, D), and (E, F) are genotypically redundant, specifically with regard to the phenotype (they each have genotype-trait redundancy). However, not all genotypes that yield a given phenotype have the same fitness. Genotype C in this example has a pleiotropic tradeoff with another trait that is not shown, resulting in lower fitness, even though it has the same phenotype as A, B, and D (this could also arise from an epistatic effect). Genotype D is the least costly for the small-cell/thick-wall phenotype (and therefore has the highest fitness), but has yet to be discovered by mutation and so is not yet segregating in the population (note the frequency of zero). So while genotype D contributes to the genotypic redundancy, it is not contributing to the segregating redundancy. Genotype set (A, B, E, F) is genotypically redundant with respect to fitness, (genotype-fitness redundancy), since the small cells with thick cell walls have the same fitness as large cells with thin cell walls. Genotype G, which is not redundant at either the trait or fitness level, is costly because investment in both large cells and thick cell walls reduces the investment in reproduction.

Robustness: a property in biological systems that allows for the maintenance of a phenotype in spite of changes to said system.
SNP: a nucleotide variant at a distinct site in a genome.
Trait optimum: a value or expression of a phenotypic trait that confers the greatest fitness.
Transgressive segregation: the observation of extreme phenotypes in hybrid offspring that are not observed in the parental lines.

or fitness value and not the number of different mutational pathways that could possibly yield each genotype, although the two may be closely related [12,13]. The concept of multiple different mutational pathways is explicitly considered in the study of empirical fitness landscapes [31] and distributed robustness [32] and also has important consequences for evolutionary dynamics, but we do not consider this further here. Instead, we focus this review on the importance of genotypic redundancy and describe its effect on evolutionary dynamics and the difficulties involved in attempting to quantify it in an empirical context.

It is important to note that functional redundancy and genotypic redundancy may both have distinct effects on the evolutionary process and the evolvability of a population, but are not mutually exclusive. If both terms are used to describe a particular trait, then an increase in functional redundancy will usually cause an increase in genotypic redundancy. However, traits with high genotypic redundancy do not always have high functional redundancy. **Complex traits** can be affected by many biochemical pathways and therefore might have a large number of genes that could mutate and contribute to evolution towards a given phenotype, without any functional redundancy within these pathways.

Functional Redundancy

Functional redundancy, as defined earlier, operates via at least two key biochemical processes: the regulation of dosage effect and pathway redundancy. Gene dosage refers to the number of

copies of a single gene within a genome, while the dosage effect is determined by the amount of expressed gene product. A duplication of a gene within an individual, in the absence of strict regulatory control, doubles that gene's product within the individual. A duplication of genes leading to a phenotype/fitness shift, however, would not be an example of functional redundancy. Functional redundancy occurs when increased dosage is mediated, for example, by **Michaelis-Menten dynamics** of enzyme kinetics or by regulatory control [33] (Figure 2). This phenomenon produces patterns of **diminishing returns epistasis** and can in fact play a significant role in establishing adaptive **genetic architectures** [17,34]. In a similar vein, pathway redundancy allows for multiple genes to contribute to a shared function through an overlap in gene interaction. Two distinct genetic elements may be functionally redundant if they contribute to the maintenance of a **gene regulatory network** of interconnected expression factors, if the loss of one element is compensated for by the other element in the expression network. Note that this scenario is directly related to distributed robustness [32]. Both dosage effect regulation and pathway redundancy allow for functional redundant genes to persist. As exemplified by the myogenic

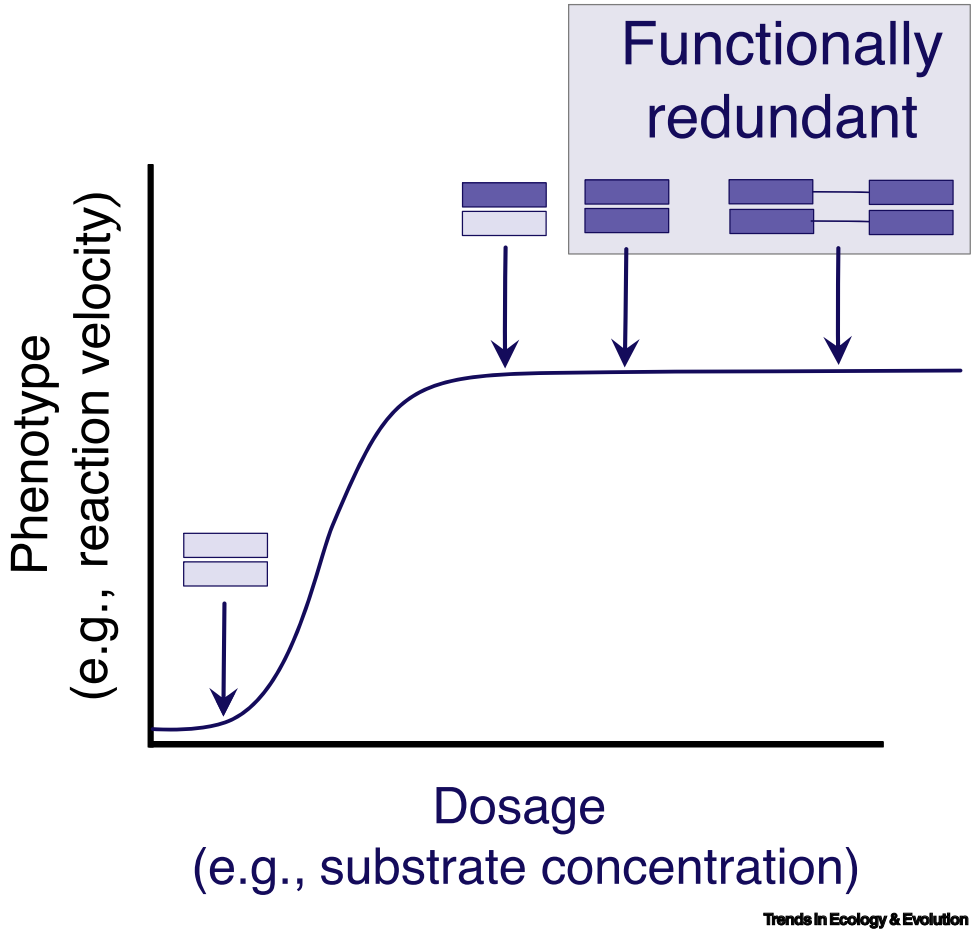

Figure 2. A Theoretical Example of a Phenotypic Value as a Function of Gene Product Dosage under Michaelis-Menten Dynamics of Enzyme Kinetics. Color indicates the allele and from left to right are shown the homozygote recessive (two gray bars), heterozygote (purple and gray bars), homozygote dominant (two purple bars), and gene duplication (four purple bars). The gene is completely dominant for the phenotype because the heterozygote has the same phenotype as the homozygote. Functionally redundant genes are those in which one copy could be deleted and give the same phenotypic value.

regulatory factor (MRF) **gene family**, dosage and pathway functional redundancy can operate in conjunction [35–38].

While much attention has been given to the eventual evolutionary fates of functionally redundant genetic elements, the origins of functional redundancy are still the subject of much study [39–41]. Origin of redundancy through duplication has received special attention, given that the genome-wide rates of gene duplication in multicellular eukaryotes have been observed in the range of 10^{-7} duplications per gene per generation [42] and locus-specific rates have been observed to be as high as 10^{-5} duplications per gene per generation [43,44]. For more comprehensive discussions of what we term 'functional redundancy' and related concepts, please see [9,33,41,45–47].

Genotypic and Segregating Redundancy

Genotypic redundancy refers to the total number of possible causal genotypes that could occur through mutation that yield a given phenotype (scaled by likelihood of occurrence, see later), while segregating redundancy refers to the realized number of different genotypes with the same phenotype that are present at a particular time point in the population or species. Genotypic redundancy therefore only evolves with changes in an organism's body plan and genome that change the underlying contribution of genotype to phenotype to fitness (which occurs slowly). Segregating redundancy is the subset of genotypic redundancy, determined by allelic variation at loci in the **genotype–phenotype map**, and is therefore constantly (and potentially rapidly) changing as a result of the interaction between all population genetic processes. From a technical standpoint, if genotypic redundancy accounts for all possible causal mutations, including insertions or deletions, the possible genotype space approaches infinity, so a more precise way to formally quantify genotypic redundancy would be to weight each genotype by its likelihood of occurring by mutation, although this is difficult.

All else being equal, an increase in genotypic redundancy can affect evolution in three main ways: (i) increase the per-trait mutation rate and therefore cause an increase in segregating redundancy, standing variation, and response to selection; (ii) increase the number of ways an organism can be well adapted, reducing **genomic repeatability** of the causal basis of evolutionary change among independent replicates (see 'Quantifying Genotypic and Segregating Redundancy' later); (iii) lead to different evolutionary dynamics, as a result of the first two factors. Early work that explicitly studied genotypic redundancy examined how it could contribute to the maintenance of genetic variation [12,48]. A recent model of adaptation to a new environment showed that high genotypic redundancy can lead to a change from a 'sweeps' to a 'shifts' regime, where instead of single alleles spreading to high frequency, a similar change in phenotype can be achieved by small shifts in allele frequency at multiple loci [17]. In models of local adaptation, high redundancy can have a range of effects on evolutionary dynamics, which we describe more fully later. As an example of these effects, consider the contrast between two studies of adaptation to heterogeneous environments: a population genetic model with no genotypic redundancy and pure divergent selection finds gradual accumulation of established differences and eventual speciation [49], while a quantitative genetic model with high redundancy and stabilizing selection finds more subtle shifts in the underlying genetic architecture and long-term local adaptation [50].

Population- and quantitative-genetic models make very different implicit assumptions about genotypic redundancy. Population genetic models typically assign selection coefficients to individual alleles and only model interactions among loci using explicit models of gene–gene interactions. To the extent that they are used to scale up to whole-genome predictions across multiple loci, such models implicitly assume that there is no genotypic redundancy, as the highest fitness

phenotype can only be achieved by fixing alleles at all selected loci. By contrast, quantitative genetic models rely upon the infinitesimal assumption, whereby evolution in phenotypic means, variances, and higher moments can be approximated by assuming that all alleles have vanishingly small and interchangeable effects on the phenotype. Such models therefore implicitly assume complete genotypic redundancy (and universal pleiotropy), although they also give useful approximations in cases with a small number of loci of moderate to large effect [51–53]. To put this another way, quantitative genetic models assume additive effects on the phenotype, with epistasis arising when selection favors intermediate phenotypes, whereas population genetic models do not model the phenotype and therefore only model epistasis directly from fitness effects.

Redundancy and Evolutionary Dynamics of Local Adaptation

To illustrate the effects of genotypic and segregating redundancy on evolution, we use a two-patch polygenic model of local adaptation, where the phenotype in each patch is evolving to a different local optimum (Box 1). In this simulation, all single step mutations have equal effect sizes and probabilities of occurring. Increasing the number of loci that contribute to the trait (from 20 to 50) increases the overall genotypic redundancy at the fitness optimum (from 1 to ~10^{12} possible genotypes that could give the optimum phenotype). All else being equal, this

Box 1. Redundancy in a Quantitative Genetic Context

Genotypic redundancy can be computed with some simplifying assumptions for a haploid life cycle. We assume n diallelic loci. Each locus has an ancestral allele that has no effect on the trait ($\alpha = 0$) and an alternate allele that has an effect on the trait. For the alternate allele, half of the loci have effect α on the trait and the other half of the loci have effect $-\alpha$.

This gives $n + 1$ possible phenotypes that we label as $k \in 0, 1, 2...n$. The phenotypic value φ for each of the $n + 1$ possible phenotypes is given by:

$$\phi(k) = \alpha\left(-\frac{n}{2} + k\right) \quad [i]$$

and the number of genotypes that produces each phenotypic value (e.g., the genotypic redundancy, r) is simply given by the number of ways that each k number of loci could be combined out of the total n (i.e., the binomial coefficient):

$$r(k) = \binom{n}{k} \quad [ii]$$

Given all else equal, different genotypic redundancy surfaces can affect evolutionary outcomes in terms of the distribution of phenotypes that can evolve, their proximity to the phenotypic optimum, and the number of redundant genotypes (see Table 1 in main text) segregating in the population (i.e., segregating redundancy). Ultimately, these outcomes will depend on the distributions of genotypic redundancy and fitness.

To illustrate this in a quantitative genetic context, we used individual-based haploid simulations of local adaptation. We assumed a two-patch model with N individuals, migration rate m equal between patches, and mutation rate μ to the alternate allele at each locus (Figure I). In a haploid life cycle the n loci are assumed to be linked on a single chromosome and inherited as a unit. We use a Gaussian (quadratic) function to describe stabilizing selection on the phenotype (z) of individual (i) in population (p) with phenotypic optimum Θ_p, and selection variance ω_p^2:

$$Wz_{i,k} = 1 - \frac{(z_{i,p} - \Theta_p)^2}{\omega_p^2} \quad [iii]$$

The two-patch model is initiated with the same environment in both patches for a burnin period, followed by an environmental shift that creates a different optimum in each patch. During the burnin period, we simulated a phenotype under weak stabilizing selection to an optimum of $\Theta_p = 0$ in both patches for $4N$ generations to allow for an accumulation of standing variation, followed by a gradual shift towards two different optima over 100 generations. Then we simulated $4N$ generations of spatially heterogeneous selection with a lower optimum environment in patch one ($\Theta_1 = -1$) and higher optimum environment in patch two ($\Theta_2 = 1$). The amount of local adaptation in the metapopulation was measured by the difference in the mean fitness in sympatry (e.g., mean fitness of individuals in their home patch) and allopatry (e.g., mean fitness of individuals in the foreign patch) [68].

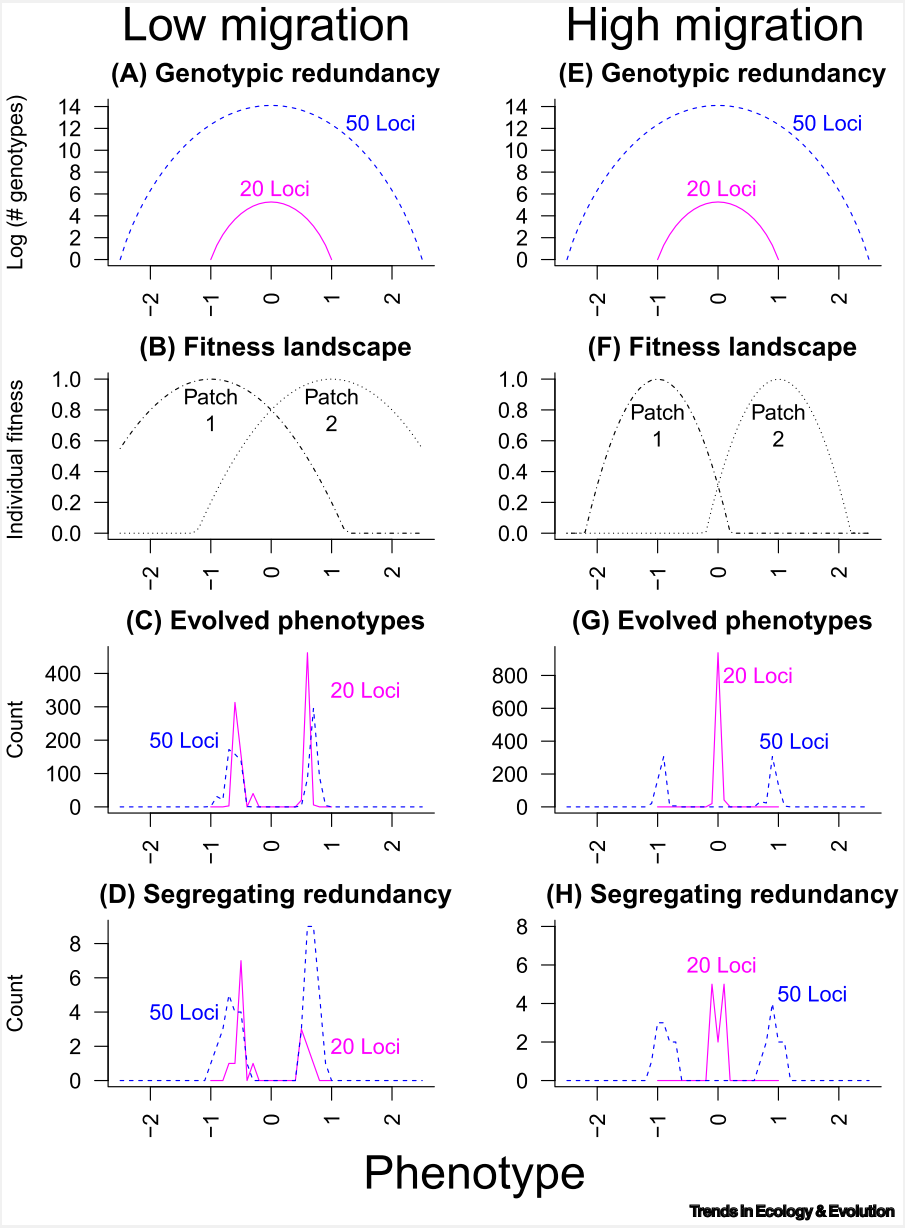

Figure I. Genotypic Redundancy in a Two-Patch Genetic Model with 20 or 50 Loci Underlying the Trait with Low and High Migration. Note that under the 20 locus model, there is only one possible genotype that gives either optimal phenotype. (A,E) Distribution of genotypic redundancy. (B,F) Individual fitness landscape. (C,G) Distribution of evolved phenotypes. (D,H) Distribution of the number of genotypes segregating in the population that give each phenotypic value at the end of the simulation. Low migration (left): simulation results with selection variance (ω_p^2) = 5, migration rate (m) = 0.2, and allele effect size (α) = 0.1. The amount of local adaptation achieved in (C) by the 20 locus model (individuals had on average 0.25 more offspring in their home patch as compared with the foreign patch) was slightly less than that achieved by the 50 locus model (0.34 more offspring in home patch compared with foreign patch). High migration (right): simulation results for ω_p^2 = 1.45, m = 0.5, and α = 0.1. Note that compared with low migration, here there is complete mixing via migration but also stronger stabilizing selection and it might be considered a case of microgeographic adaptation [69] or disruptive selection. In this case the 20 locus model did not achieve local adaptation (0.0116 more offspring in home patch compared with foreign patch), but the 50 locus model did (2.55 more offspring in home patch). In the 20 locus case there was low segregating redundancy at the end of the simulation, with the majority of individuals having a phenotypic value of 0 given by two genotypes.

additional genotypic redundancy can allow populations to achieve more local adaptation in high migration scenarios (see Figure IG,IH in Box 1). These results also illustrate the importance of understanding the distribution of genotypic redundancy across all trait values relative to the **trait optimum**, although this is rarely reported in theoretical studies.

These evolutionary dynamics are determined by the interaction of migration-selection balance, drift, and stochasticity in the size and linkage relationships of the realized mutations. With local adaptation in the face of incoming maladapted alleles, the evolutionary dynamics depend not just on the incoming migration rate, but on the level of genotypic redundancy. In Figure 3 (Key Figure), we summarize theoretical predictions for how genotypic redundancy determines the underlying quantitative genetic architecture of local adaptation in the two-patch model of Yeaman [13]. In this model, the degree of local adaptation that can be attained depends partly on genotypic redundancy for the optimum trait (i.e., the number of possible genotypes that give the optimal trait), the mutation rate (i.e., the probability of hitting a favorable combination of alleles), and the effect size of the alleles that underlie the trait. Alleles with larger additive effects on the phenotype (relative to the migration rate of incoming maladaptive alleles) are more resistant to the homogenizing effect of incoming migration (being

Key Figure

Stability of Genetic Architecture of Local Adaptation through Time for a Quantitative Genetic Model of Migration-Selection Balance, Where Multiple Loci Contribute Additively to a Phenotype [13]

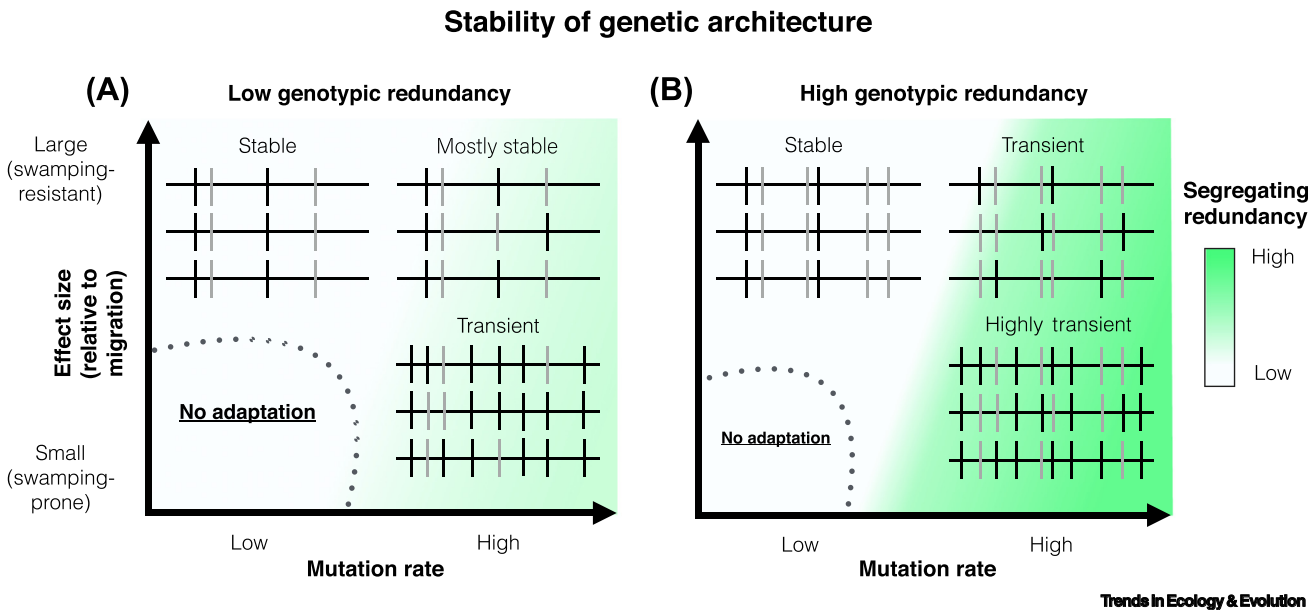

Figure 3. The x-axis represents the effect of mutation rate, under either low (A) or high (B) genotypic redundancy. The effect size (y-axis) represents the additive effect of the allele on the phenotype. This effect size parameter is relative to the migration rate, where 'swamping' indicates whether the allele would be able to evolve in the face of incoming migration. In each quadrant, a population of three individuals is shown as linear chromosomes, where the hashes represent the locations of causal alleles on the chromosome. Dark hashes are extant alleles segregating in the population, gray hashes are potentially contributing alleles not yet discovered by mutation. 'Stable' motifs have consistent architecture and relatively invariable allele frequencies through time, while 'transient' motifs have different loci contributing to phenotypic divergence through time and may also have high segregating redundancy at any particular snapshot in the evolutionary process. Note that segregating redundancy is expected to reach higher levels under a high genotypic redundancy scenario, and swamping-prone alleles can persist in the population under lower mutation rates than when genotypic redundancy is low.

'swamped'), while small individual allele effect sizes may be swamped by migration in single-locus models.

Theory predicts that when genotypic redundancy is low, and the allele effect size is large enough to be swamping-resistant, the genetic architecture will be stable through time (Figure 3A, top) [13]. In this low redundancy case, if the allele effect sizes are small relative to migration and the mutation rate is also low, locally adapted alleles are unlikely to accumulate (Figure 3A, lower left). When both genotypic redundancy and mutation rates are high, theory predicts that more transient genetic architectures will arise in the population across a range of allelic effect sizes (because new genotypes contributing to the optimum arise more frequently; Figure 3B, right side) [18]. In this case, swamping-prone alleles can persist in the population under lower mutation rates than when genotypic redundancy is low, as shifting combinations of genotypes allow for maintenance of a locally adapted phenotype (Figure 3B, lower middle) [18]. Finally, as an increase in genotypic redundancy implies an increase in the number of causal loci, it should also increase the segregating redundancy if mutation rate per locus is held constant.

Quantifying Genotypic and Segregating Redundancy

Given the practical impossibility of identifying all potential causal variants, genotypic redundancy can't be quantified empirically, but can be inferred from observations about the repeatability of adaptation among individuals or lineages [18,54–56], as evolution will be more repeatable when redundancy is low. Note that in addition to genotypic redundancy, the number of different mutational pathways that yield a given genotype can also be important for repeatability [57]. Segregating redundancy is proportional to the number of different ways a given phenotype can be created by combinations of alleles present within a population, while genotypic redundancy also includes the loci that do not currently contribute to variation, but that could contribute given a suitable mutation event. Thus, segregating redundancy can be assessed by comparing individuals within a population, while genotypic redundancy can be assessed by comparing independently evolved populations or species. When lineages share a relatively recent evolutionary history, similarities in the complement of alleles driving variation can persist due to shared standing variation, so it is important to account for this contribution. Later, we discuss four approaches that can be used to quantify which alleles are responsible for observed trait variation and draw inferences about the amount of redundancy.

While labor-intensive, one of the most direct ways of finding evidence for genotypic redundancy is to conduct numerous **quantitative trait loci (QTL)** experiments on individuals with divergent phenotypes, to identify the number of loci and the direction of their effects on phenotypes. Comparisons among crosses can then be used to provide a direct estimation of how commonly different individuals arrive at a similar phenotype via different QTLs [18,54]. For example, a study of QTLs for body shape in stickleback found evidence of alleles acting in two distinct directions of effect within a population adapting to a novel environment, suggesting that combinations of these alleles would provide different ways of achieving the same phenotype [58]. Similar approaches could be extended to **genome wide association study (GWAS)**-based methods to quantify the direction and magnitude of **SNP** effects and, indeed, most GWAS studies identify a large number of alleles underlying trait variation, which is suggestive of high redundancy [59,60]. These approaches also lend themselves to the assessment of **transgressive segregation** [61]. When a cross is made between two populations that have evolved independently in allopatry for some time, the F2 generation may show variation that exceeds the values of the parental types. One way that this transgressive segregation can arise is through high redundancy, whereby the two parental types harbor different combinations of QTLs, in which case the F1s are intermediate

but the F2 individuals can inherit an extreme dosage of alleles that either all increase or decrease a given phenotype [61]. Evidence from a survey of studies suggests that transgressive segregation is common [61] and if other explanations such as heterosis can be discounted, then the extent of transgressive segregation can be interpreted as an indicator of the extent of segregating redundancy.

Another approach is to make direct comparisons among lineages that have adapted to a similar challenge from a common ancestor. Regardless of the method of assessment, it is important to differentiate between variation that arises through mutation-selection balance within a population (which occurs with stabilizing selection) and variation that arises as a product of multiple routes to adaptation (where high overlap is commonly referred to as **convergence**). In either case, comparisons of the amount of overlap among independent evolutionary lineages will provide evidence about the amount of genotypic redundancy underlying the evolutionary trajectory. However, these different scenarios have very different implications for how evolution will proceed [18]. Yeaman *et al*. [18] introduced the *C*-score as a measure to quantify repeatability in the genomic basis of trait variation observed over independent bouts of adaptation. High values of the *C*-score occur with high repeatability, which implies that some **constraints** exist that limit the number of viable genotypes that can give rise to an adaptive phenotype. In Box 2, we demonstrate how the *C*-score is related to the amount of genotypic and segregating redundancy using simulated data and discuss some of its limitations. (All code used for the simulations and calculations in Boxes 1 and 2 is publicly available online: https://github.com/akijarl/redundant_redundancy. The R code is also included as a supplemental file.)

Finally, experimental evolution can provide insights into segregating and genotypic redundancy. If populations are initiated with substantial standing variation, then the amount of observed repeatability in the genetic basis of adaptation should be inversely proportional to the level of segregating

Box 2. Quantitative Measurement of Redundancy

Directly quantifying genotypic and segregating redundancy in real organisms is challenging, but it can be inferred based on the genomic basis of repeated adaptation, because under high redundancy the same adaptive genotype is unlikely to be repeated. The *C*-score index developed by Yeaman *et al*. [18] represents the deviation between the observed and expected amount of overlap in the loci contributing to adaptation in independently evolving lineages, under a null model where all genes could potentially contribute to adaptation (i.e., complete redundancy). We used the chi-squared version of the *C*-score ($C_{\chi 2}$), which can be interpreted as a 'constraint' score, where higher values represent higher constraint and lower redundancy (e.g., fewer available routes that contribute to adaptation). In a situation where a pair of lineages repeatedly adapt via changes in many of the same genes in response to a given selection pressure (high *C*-score), the amount of genotypic redundancy can be inferred to be low, as few adaptive solutions must have been available to evolution. Conversely, if a pair of lineages repeatedly evolve via changes in different genes (low *C*-score), the amount of genotypic redundancy can be inferred to be high and there are fewer constraints on which genes can contribute to adaptation. We compared repeated runs of the simulations described in Box 1 (width of selection function of $\omega_p^2 = 5$, allele effect size of $\alpha = 0.1$, migration rate of $m = 0.2$), across four scenarios where 2%, 3%, 4%, and 5% of an otherwise neutral 1000 locus genome contributes to the phenotype under selection. As the number of loci in the genome potentially affecting the trait under selection increased, *C*-score values declined, while the $\log_{10}$ genotypic redundancy of the median evolved phenotype increased (Figure IA), and the same pattern was observed with the total number of segregating redundant variants (Figure IB).

There is a negative relationship (adjusted R^2 between 15% and 20%, across repeated simulations at this genome size) between the *C*-score and both measures of redundancy as the percentage of the genome that could potentially contribute to the trait increases (Figure I, all *P* values $< 10^{-15}$). As the number of loci that could potentially contribute to the optimal fitness trait increases, both $\log_{10}$ of genotypic redundancy values and the number of segregating sites increase (Figure II). Because the *C*-score quantifies the result of adaptive evolution, there is noise in its mapping to redundancy, due to drift and the stochasticity of mutation occurrence. Factors such as mutation rate, genetic drift, and the interplay between migration and selection will cause segregating redundancy to vary when genotypic redundancy is held constant and further work is required to explore the relationship between *C*-score and these kinds of redundancy. In an experimental context, initializing each population from the same standing variation will allow quantification of *C*-score as it relates to segregating redundancy, while initializing each population with no standing variation will allow quantification of *C*-score as it relates to genotypic redundancy.

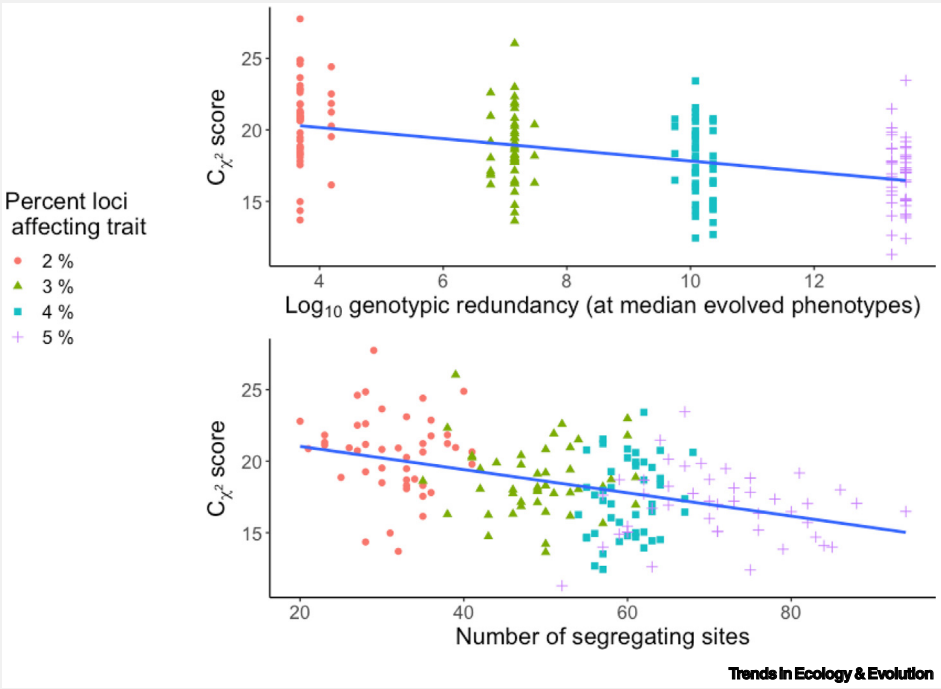

Figure I. Comparisons of the Relationship between an Index of Evolutionary Repeatability (*C*-Score) and (A) Genotypic Redundancy and (B) Segregating Redundancy. Points are color coded by the number of loci out of a 1000 locus genome that contributed to the phenotype under selection.

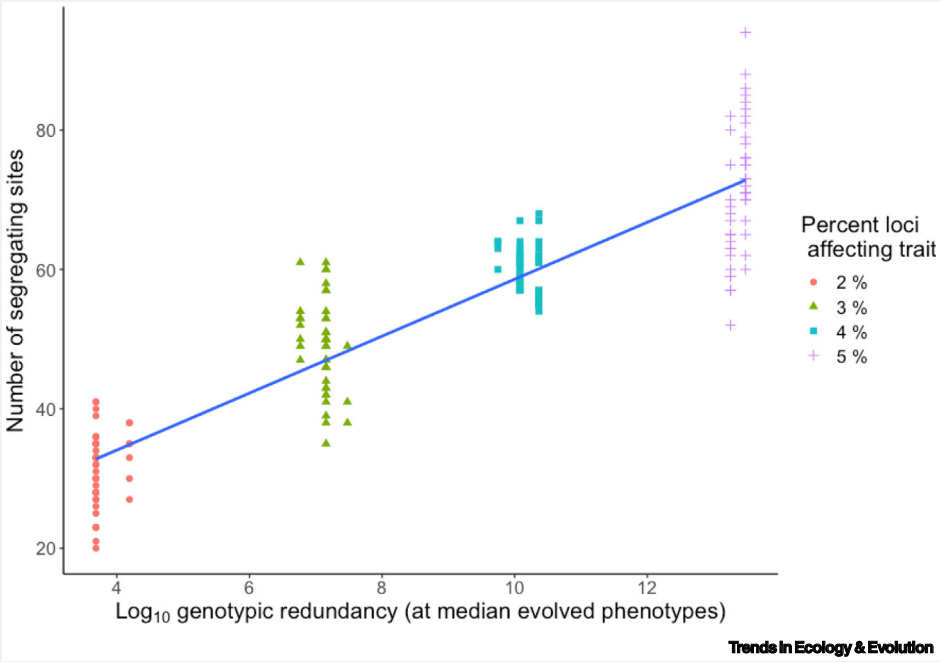

Figure II. A Strong Positive Relationship between Log_{10} of Genotypic Redundancy Values and the Number of Segregating Sites, as the Number of Loci Contributing to the Optimal Fitness Trait Increase (Adjusted R^2 Approximately 80%, all *P* Value $<10^{-15}$, across Repeated Simulations). Points are color coded by the number of loci out of a 1000 locus genome that contributed to the phenotype under selection.

redundancy at the start of the experiment. Note that segregating redundancy may increase during the experiment from new mutation and recombination, especially when genotypic redundancy is inherently high or substantial linkage disequilibrium exists among causal variants. If populations are started from an isogenic line, then all adaptation occurs from new mutations and genotypic redundancy can be inferred from replicate sampling designs to infer the number of loci that could potentially contribute to adaptation (as per Yeaman *et al*. [18]; for an example of a study initialized with abundant standing variation, see Barghi *et al*. [16], described later).

Empirical Examples of Genetic Redundancy

Bar Mutation: No Functional or Genotypic Redundancy for Phenotype with Genotypic Redundancy for Fitness

The *Bar* mutation in *Drosophila* was a historical springboard for discussions surrounding genetic redundancy. Because a duplication of the Bar gene makes the fly eye-shape more bar-like, and a triplication narrows the eye shape even more [3], this is not an example of functional redundancy or genotypic redundancy for a phenotype because the copy number determines the phenotype. Barred mutants do, however, have similar viability to the wild type in the laboratory [62], indicating there may be some genotypic redundancy for fitness, although this may be environmentally dependent.

Arabidopsis: Low Functional Redundancy for Fitness-Related Traits

In a study of over 1700 *Arabidopsis thaliana*, including 116 homozygous insertion mutation lines and ten natural accession ('ecotype') lines, Rutter *et al*. suggested that a far greater number of knockout mutations actually alter fitness-associated phenotypes than previously reported [63]. With the assumption that functional redundancy would render most knockdown mutations as effectively neutral, the prevalence of a fitness response (measured as number of fruits set) in their set of knockdown mutants lead Rutter *et al*. to conclude that the functional redundancy in *A. thaliana* was far less than previously reported.

Mc1r Coat Color: High Genotypic Redundancy within a Gene

Some traits exhibit very little genotypic redundancy at the genome-scale but considerable redundancy within a given gene. Despite a large number of different genes that can contribute to variation in melanism in vertebrates [64], light/dark color adaptation commonly involves large effect changes at *Mc1r* in a wide range of animals, from mice to lizards to mammoths [65,66]. Within *Mc1r*, however, there are many different causal mutations that have been identified that all contribute similar changes in coloration [64,66,67]. Given that fitness-driven adaptation is highly repeatable (i.e., via *Mc1r*), despite many other possible genes that could mutate to change color, this trait exhibits high genotype–phenotype redundancy and low genotype-fitness redundancy at the genome scale, but high genotype-fitness redundancy within the *Mc1r* gene.

Polygenic Adaptation to Warming: High Genotypic Redundancy across a Genome

Experimental evolution with *Drosophila simulans* has recently shown evidence of extensive segregating redundancy for traits associated with temperature-mediated laboratory adaptation [16]. In this experiment, replicate populations sampled after 60 generations evolved similar phenotypes via allele frequency changes at many different combinations of loci. Simulations comparing expected variation in allele frequency change under a model of directional selection on all loci, compared with a model with high genotypic redundancy underlying a trait under selection, found the redundant model to much better fit the observed allele frequency changes. From these results, we can infer that segregating redundancy was high prior to the beginning of the experiment (as new mutations are unlikely to have contributed much to the response) and therefore genotypic redundancy must also be high.

Concluding Remarks and Future Directions

While studying genotypic redundancy is critically important for understanding how evolutionary processes give rise to adaptive variation, many questions remain unanswered (see Outstanding Questions). For instance, will a species with more segregating redundancy evolve more rapidly in response to a changing environment? Given that the response to selection appears to be proportional to the standing genetic variation, the answer to this question would likely depend upon how redundancy affects the maintenance of variation. Very little formal theory has been conducted on this question and this would be a clear avenue for further research. Meanwhile, there is rapid accumulation of a wealth of GWAS data, yet there has been relatively little specific interpretation of these data in light of what it means for redundancy. Reassessing published GWAS data to examine the prevalence of redundancy could yield many insights into its evolutionary function. Finally, in order to truly assess the role of redundancy in evolution, it is imperative that future empirical and theoretical studies aim to quantify redundancy whenever possible.

Outstanding Questions

How and why does the level of genotypic and segregating redundancy vary among species?

Do some types of traits tend to have more genotypic redundancy than others?

What conditions favor the evolution of increased genotypic redundancy?

How well do empirical indices of repeatability map to theoretical concepts of redundancy and how can such methods be improved?

Does phenotypic plasticity or epigenetics contribute to genotypic redundancy in the genotype–phenotype or genotype-fitness map?

Under what conditions do we see a decoupling of the levels of genotypic and segregating redundancy?

What is the relationship between genotypic redundancy and evolvability?

How is the evolution of genetic architecture related to genotypic redundancy?

Acknowledgments

We would like to thank the editor Andrea Stephens, reviewer Joachim Hermisson, and another anonymous reviewer for insights and suggestions that greatly improved the quality of this manuscript. Dr Molly Albecker, Dr Thais Bitter, Alan Downey-Wall, Sara Schaal, and Victoria Sindorf also provided many helpful comments and suggestions. S.Y. is supported by Alberta Innovates and NSERC Discovery funding. Á.J.L. and K.E.L. are supported by NSF (grant no. 1655701).

Supplemental Information

Supplemental information associated with this article can be found online https://doi.org/10.1016/j.tree.2020.04.009.

References

1. Muller, H.J. (1936) Bar duplication. *Science* 83, 528–530
2. Bridges, C.B. (1936) The Bar "gene" a duplication. *Science* 83, 210–211
3. Wolfner, M.F., Miller, D.E., eds (2016) Alfred Sturtevant walks into a Bar: gene dosage, gene position, and unequal crossing over in *Drosophila*. *Genetics* 204, 833–835
4. Sturtevant, A.H. (1925) The effects of unequal crossing over at the Bar locus in *Drosophila*. *Genetics* 10, 117–147
5. Fisher, R.A. (1935) The sheltering of lethals. *Am. Nat.* 69, 446–455
6. Stephens, S.G. (1951) Possible significance of duplication in evolution. *Adv. Genet.* 4, 247–265
7. Gabriel, M.L. (1960) Primitive genetic mechanisms and the origin of chromosomes. *Am. Nat.* 94, 257–269
8. Mazia, D. (1960) The analysis of cell reproduction. *Ann. N. Y. Acad. Sci.* 90, 455–469
9. Nowak, M.A. *et al.* (1997) Evolution of genetic redundancy. *Nature* 388, 167–171
10. Ascencio, D. and DeLuna, A. (2013) Genetic redundancy. In *Encyclopedia of Systems Biology*, pp. 824–827, Springer
11. Thomas, J.H. (1993) Thinking about genetic redundancy. *Trends Genet.* 9, 395–399
12. Goldstein, D.B. and Holsinger, K.E. (1992) Maintenance of polygenic variation in spatially structured populations: roles for local mating and genetic redundancy. *Evolution* 46, 412–429
13. Yeaman, S. (2015) Local adaptation by alleles of small effect. *Am. Nat.* 186, S74–S89
14. Kokko, H. *et al.* (2017) Can evolution supply what ecology demands? *Trends Ecol. Evol.* 32, 187–197
15. McDonald, T.K. and Yeaman, S. (2018) Effect of migration and environmental heterogeneity on the maintenance of quantitative genetic variation: a simulation study. *J. Evol. Biol.* 31, 1386–1399
16. Barghi, N. *et al.* (2019) Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biol.* 17, e3000128
17. Höllinger, I. *et al.* (2019) Polygenic adaptation: from sweeps to subtle frequency shifts. *PLoS Genet.* 15, e1008035
18. Yeaman, S. *et al.* (2018) Quantifying how constraints limit the diversity of viable routes to adaptation. *PLoS Genet.* 14, e1007717
19. Payne, J.L. and Wagner, A. (2019) The causes of evolvability and their evolution. *Nat. Rev. Genet.* 20, 24–38
20. Lack, J.B. *et al.* (2016) Decanalization of wing development accompanied the evolution of large wings in high-altitude *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 113, 1014–1019
21. de Visser, J.A.G.M. *et al.* (2003) Perspective: evolution and detection of genetic robustness. *Evolution* 57, 1959–1972
22. Whitacre, J. and Bender, A. (2010) Degeneracy: a design principle for achieving robustness and evolvability. *J. Theor. Biol.* 263, 143–153
23. Plata, G. and Vitkup, D. (2014) Genetic robustness and functional evolution of gene duplicates. *Nucleic Acids Res.* 42, 2405–2414
24. Félix, M.-A. and Barkoulas, M. (2015) Pervasive robustness in biological systems. *Nat. Rev. Genet.* 16, 483–496
25. Waddington, C.H. (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565
26. Eshel, I. and Matessi, C. (1998) Canalization, genetic assimilation and preadaptation. A quantitative genetic model. *Genetics* 149, 2119–2133
27. Ehrenreich, I.M. and Pfennig, D.W. (2016) Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. *Ann. Bot.* 117, 769–779
28. Rutherford, S.L. (2000) From genotype to phenotype: buffering mechanisms and the storage of genetic information. *Bioessays* 22, 1095–1105
29. Hartman 4th, J.L. *et al.* (2001) Principles for the buffering of genetic variation. *Science* 291, 1001–1004
30. Costanzo, M. *et al.* (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353, aaf1420
31. de Visser, J.A.G.M. and Krug, J. (2014) Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* 15, 480–490
32. Wagner, A. (2005) Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays* 27, 176–188
33. Qian, W. *et al.* (2010) Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet.* 26, 425–430
34. Hansen, T.F. (2006) The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. Syst.* 37, 123–157

35. Atchley, W.R. *et al.* (1994) Molecular evolution of the MyoD family of transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 91, 11522–11526
36. Megeney, L.A. and Rudnicki, M.A. (1995) Determination versus differentiation and the MyoD family of transcription factors. *Biochem. Cell Biol.* 73, 723–732
37. Kablar, B. *et al.* (1999) Myogenic determination occurs independently in somites and limb buds. *Dev. Biol.* 206, 219–231
38. Hernández-Hernández, J.M. *et al.* (2017) The myogenic regulatory factors, determinants of muscle development, cell identity and regeneration. *Semin. Cell Dev. Biol.* 72, 10–18
39. Bergthorsson, U. *et al.* (2007) Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci. U. S. A.* 104, 17004–17009
40. Lynch, M. and Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459–473
41. Teufel, A.I. *et al.* (2019) The many nuanced evolutionary consequences of duplicated genes. *Mol. Biol. Evol.* 36, 304–314
42. Lipinski, K.J. *et al.* (2011) High spontaneous rate of gene duplication in *Caenorhabditis elegans*. *Curr. Biol.* 21, 306–310
43. Konrad, A. *et al.* (2018) Mutational and transcriptional landscape of spontaneous gene duplications and deletions in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* 115, 7386–7391
44. Keith, N. *et al.* (2016) High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. *Genome Res.* 26, 60–69
45. Wagner, A. and Wright, J. (2007) Alternative routes and mutational robustness in complex regulatory networks. *Biosystems* 88, 163–172
46. Kafri, R. *et al.* (2009) Genetic redundancy: new tricks for old genes. *Cell* 136, 389–392
47. Pfennig, D.W. and Ehrenreich, I.M. (2014) Towards a gene regulatory network perspective on phenotypic plasticity, genetic accommodation and genetic assimilation. *Mol. Ecol.* 23, 4438–4440
48. Phillips, P.C. (1996) Maintenance of polygenic variation via a migration-selection balance under uniform selection. *Evolution* 50, 1334
49. Flaxman, S.M. *et al.* (2013) Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution* 67, 2577–2591
50. Yeaman, S. and Whitlock, M.C. (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution* 65, 1897–1911
51. Turelli, M. and Barton, N.H. (2004) Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G x E interactions. *Genetics* 166, 1053–1079
52. Barton, N.H. and Turelli, M. (2004) Effects of genetic drift on variance components under a general model of epistasis. *Evolution* 58, 2111–2132
53. Turelli, M. and Barton, N.H. (2006) Will population bottlenecks and multilocus epistasis increase additive genetic variance? *Evolution* 60, 1763–1776
54. Conte, G.L. *et al.* (2012) The probability of genetic parallelism and convergence in natural populations. *Proc. Biol. Sci.* 279, 5039–5047
55. Martin, A. and Orgogozo, V. (2013) The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* 67, 1235–1250
56. Lachapelle, J. *et al.* (2015) Repeatability of adaptation in experimental populations of different sizes. *Proc. Biol. Sci.* 282, 20143033
57. Kondrashov, D.A. and Kondrashov, F.A. (2015) Topological features of rugged fitness landscapes in sequence space. *Trends Genet.* 31, 24–33
58. Albert, A.Y.K. *et al.* (2008) The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution* 62, 76–85
59. Boyle, E.A. *et al.* (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169, 1177–1186
60. Robinson, M.R. *et al.* (2014) Explaining additional genetic variation in complex traits. *Trends Genet.* 30, 124–132
61. Rieseberg, L.H. *et al.* (1999) Transgressive segregation, adaptation and speciation. *Heredity* 83, 363–372
62. Tice, S.C. (1914) A new sex-linked character in *Drosophila*. *Biol. Bull.* 26, 221–230
63. Rutter, M.T. *et al.* (2017) Fitness effects of mutation: testing genetic redundancy in *Arabidopsis thaliana*. *J. Evol. Biol.* 30, 1124–1135
64. Hoekstra, H.E. (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 97, 222–234
65. Rosenblum, E.B. *et al.* (2004) Adaptive reptile color variation and the evolution of the Mc1r gene. *Evolution* 58, 1794–1808
66. Manceau, M. *et al.* (2010) Convergence in pigmentation at multiple levels: mutations, genes and function. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 365, 2439–2450
67. Mundy, N.I. (2005) A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proc. Biol. Sci.* 272, 1633–1640
68. Blanquart, F. *et al.* (2013) A practical guide to measuring local adaptation. *Ecol. Lett.* 16, 1195–1205
69. Richardson, J.L. *et al.* (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends Ecol. Evol.* 29, 165–176