æ

6. EVOLUTION OF CELLULAR COMPLEXITY

29 November 2022

Having gained an appreciation for how various population-genetic forces interact to define the accessibility of alternative evolutionary pathways, we now turn to more specific issues relevant to the diversification of cellular features. That natural selection provides a powerful mechanism for advancing adaptive mutations is well-established, so there is no need to belabor that issue further. Likely less familiar and/or less fathomable is the idea that the nonadaptive forces of mutation and drift can often dictate the paths down which phenotypic evolution is most likely to travel, sometimes with minimal involvement from selection. In certain settings, the net result can be a gradual, passive increase in organismal complexity, with little (if any) increase in fitness throughout the process.

The goal here is to instill an appreciation for the shallowness of the assumption that natural selection is a process in relentless pursuit of biological complexity. The initial focus is on general issues regarding the evolution of complex features, with details specific to particular cellular structures and functions unfolding in subsequent chapters. To maximize the accessibility of the key points, a distinctly nonmathematical sojourn will be taken, which is not to say that the mathematical details outlined in the previous chapter are irrelevant.

Before proceeding, a brief recap of the population-genetic principles relevant to phenotypic divergence is in order. First, the classes of mutations available to selection depend on the effective population size (N_e), the inverse of which defines the power of random genetic drift. Selection will be ineffective if the randomizing potential of genetic drift is sufficiently strong. Small populations can only advance beneficial mutations with relatively large effects and cannot prevent the accumulation of deleterious mutations with small effects. Large populations are more capable of evolutionary fine tuning.

Second, owing to the granularity and directional biases of mutations, phenotypic optima may only occasionally, if at all, be attainable for cellular traits. Large-N_e species are expected to evolve higher levels of efficiency and functionality of molecular attributes. However, small N_e enables populations to move into domains that can dramatically shift the course of evolution by natural selection, with mutation playing a powerful role in directing the paths open for exploration. As these fundamental evolutionary principles are unavoidable consequences of the nature of life's genetic material, they must be kept in mind in any attempt to explain cellular diversification.

Illusions of Grandeur

A common view is that biological complexity represents the crown jewel of the awesome power of natural selection (e.g., Lane 2020), with metazoans (humans in particular) representing the pinnacle of what can be achieved. This is a peculiar assumption, as there is no evidence that increases in complexity are intrinsically advantageous. Nor is there any evidence that biology's metabolic, morphological, and behavioral features have reached a maximum level of refinement or ever will. To think that a mammal is superior to a bacterium is as meaningful as proclaiming that an Olympic athlete is superior to an award-winning cellist. In the evolutionary arena, ecological context is paramount, and the currency of natural selection (relative fitness) is only exchangeable for members of the same gene pool. Bacteria can outperform vertebrates in a myriad of ways with respect to metabolism and environmental sensing. Vertebrates can harvest different food types and have complex visual and auditory systems. However, whereas a brain can be useful in certain settings, is there any objective basis for concluding that the streamlined signal-transduction systems of prokaryotes are fundamentally inferior to the baroque and error-prone nervous systems of animals?

Although there are mathematical indices for quantifying complexity in physical systems, things are not so straight-forward in living systems, and the term is used loosely here to simply reflect differences in the numbers of unique parts and interactions within organisms. Even these measures are not always easily enumerated, rendering comparisons among closely related organisms difficult. However, aspects of cellular complexity that most pique the interest of biologists are features such as large protein complexes, the emergence of the eukaryotic cell plan from a prokaryotic ancestor, and the transition from unicellularity to multicellularity. In these cases, there is no disagreement on where things lie on the complexity gradient.

In contrast to eukaryotes, most prokaryotes have not evolved internal cell structure or complex multicellularity. Is this a sign of evolutionary inferiority, i.e., of an innate inability to generate increased morphological complexity despite the benefits that could be reaped? Given their enormous population sizes, their ability to recombine, and their presence on the planet for ~ 4 billion years, the supply of variation is hardly limiting for microbes, and as noted in Chapters 2, 3, and 24, aspects of intracellular complexity and even multicellularity have in fact emerged in some prokaryotes. Thus, the unavoidable conclusion is that morphological complexity is actively selected against in the prokaryotic world. And if that is the case, what is the evidence that increased complexity is universally advantageous in eukaryotes?

The evolution of root systems and support tissues enabled land plants to occupy ecological niches unavailable to microbes, and the evolution of predatory capacity in animals opened up new ways of living. Surely, such transitions were promoted by natural selection. However, with such transitions, other modes of living were left behind, new survivorship challenges were encountered, and rapid rates of reproduction were relinquished. Moreover, the question remains as to whether all of the underlying genetic and cellular changes in such organisms were necessary antecedents to such adaptation, as opposed to being inadvertent by-products of such changes. For example, relative to their unicellular ancestors, in just a few tens of millions of years of evolution, the genomes of metazoans and land plants independently became bloated with nonfunctional, energetically costly, and mutationally hazardous DNAs such as mobile-genetic elements and large introns (Lynch 2007). Were all such em-

bellishments essential tickets to the evolution of organismal complexity, somehow maintained in anticipation of future benefits? No credible mechanisms exist for such evolutionary prescience. More likely, many aspects of increased genome complexity simply reflect the reduced efficiency of natural selection against genomic insertions in larger organisms with reduced effective population sizes.

There are at least three reasons why cellular / organismal complexity can be suppressed in certain lineages, while passively increasing in others. First, more complex features inevitably impose greater bioenergetic costs for construction and maintenance. For small cells with relatively low total energy budgets and large effective population sizes, even minor additions to the cellular repertoire can be efficiently opposed by selection unless there are immediate benefits. In contrast, for larger cells with higher total energy budgets, a given genomic addition comprises a smaller fraction of the total energy budget. Combined with a higher power of random genetic drift (Chapter 4), resulting from populations with smaller effective population sizes, moderate-sized cellular additions will then be less visible to the eyes of natural selection, and subject to fixation in an effectively neutral fashion. These issues will be addressed more formally in Chapter 17, the main point here being that cell size alone can dictate the degree to which initially unnecessary (and sometimes weakly harmful) embellishments can become established in a population.

Second, virtually all gene-structural embellishments increase the vulnerability of genes to inactivating mutations (Lynch 2007). Typically, the increased mutational susceptibility is relatively small (on the order of the product of the mutation rate per nucleotide site, u, and the number of key nucleotide sites for proper gene function imposed by the embellishment, n). As a consequence, weakly mutationally hazardous genomic alterations will only be effectively selected against in populations with very large effective sizes. As an example, n is on the order of 25 for proper intron splicing, and u is in the range of 10^{-10} to 10^{-8}. If nu is smaller than the power of drift ($1/N_e$ for a haploid), the mutational excess associated with such a gene addition cannot be countered by purifying selection.

Finally, all other things being equal, the drift-barrier hypothesis implies that organisms with lower N_e will also evolve to have less refined structural and functional features. The negative correlation of the mutation rate with decreased N_e (Chapter 4) provides a case in point, and other examples will be encountered in subsequent chapters. In some cases, the reduced functionality of a system can open up opportunities for the establishment of additional layers of complexity, which can in turn lead to further relaxation of selection on previously established mechanisms, leading to the false impression that robust systems represent adaptive improvement (Chapter 20). This matter is taken up in further detail in the following sections.

Taken together, these arguments highlight the fact that N_e limitations, driven by fundamental constraints associated with ecology and the genetic machinery, play a central role in encouraging particular lineages to ascend up the hierarchy of complexity by nonadaptive mechanisms. That is, certain population-genetic environments are conducive to the passive operation of a complexity ratchet, with small incremental changes accruing on short time scales cumulatively leading a lineage to a new location in phenotypic space. Nevertheless, one might still expect that in moving up the ladder of biological organization – from nucleotide sequences to translated protein products to higher-order structural and biochemical features of cells, there

will be a diminishing probability of effectively neutral evolution. However, as will be seen below, the very nature of genome architecture facilitates the emergence of neutral evolutionary pathways at higher levels. Just as the third positions of codons for amino-acids with four-fold redundancy in the genetic code renders some nucleotide substitutions effectively neutral, many aspects of cellular architecture are structured in ways that provide multiple degrees of freedom for making molecular shifts with minor fitness consequences. Thus, the evolution of increased complexity need not imply increased superiority in any sense of the word, and evolution driven by nonadaptive mechanisms (mutation, recombination, and random genetic drift) need not imply a descent towards overtly maladaptive change.

Constructive Neutral Evolution

A verbal model presented by Stoltzfus (1999) and colleagues (Gray et al. 2010; Lukeš et al. 2011; Brunet and Doolittle 2018) suggests ways in which seemingly gratuitous cellular complexity might grow in the absence of direct selection for such features. The process they call constructive neutral evolution (CNE) has some antecedents in earlier verbal models of Woese (1971) and Zuckerkandl (1997).

Consider an ancestral cellular function carried out by the product of a single gene (A) (Figure 6.1). Suppose a fortuitous physical interaction then develops with another protein B, with such binding having negligible effects on both A and B's functionality. By hiding part of A's surface from the cellular environment, B may suppress the effects of future mutations arising at the A-B interface that would be destabilizing to A if exposed (Chapter 13). Over time, this permissive interfacial environment could then lead to enough mutational buildup that A would no longer be functional without B. In principle, this evolved functional dependence of A on B could be followed by a similar scenario involving a third protein, C, and so on.

Under this scenario, the intricate inter-dependencies of the components of molecular complexes need not always have been advanced by positive selection for functional improvement. Rather, they may simply be the result of a series of effectively neutral coevolutionary steps accompanied by relaxed selection against previously forbidden mutations.

Although this verbal model provides a plausible argument for the passive origin of complexity, three key assumptions underlie the CNE hypothesis. Foremost is the idea that biological systems often harbor excess capacity. In particular, the process requires that the evolutionary diversion of B molecules to A has negligible effects on any preexisting benefits of B, at least to the extent that could be opposed by natural selection. As excess capacity implies a superfluous energetic drain on the cell, why would such conditions exist? As discussed in the following section, although redundancy is unlikely to be promoted on its own merits, recurrent gene duplication may lead to a sort of quasi-equilibrium level of redundancy at the population level. The specific genes involved in duplicate form at any particular time will vary, but some such genes will nearly always be present. In addition, transient conditions may exist in which a change in environment may render the prior function of B obsolete such that its diversion has no fitness consequences.

The second issue is that the evolution of A's dependency on B requires that the fortuitous A-B interaction survives for a long enough period for A to acquire the conditionally harmful mutations essential to the development of dependency on B. This returns us to the kinds of scenarios outlined in Chapter 5 whereby a small number of mutations are required for a transition to an alternative semi-stable state.

But this brings us to the third assumption of the CNE model. Unlike the situation in which populations can shift in both directions, the transition to complexity under CNE is viewed as being a one-way street – the assumption here is that once the complex is established, the accumulation of conditionally lethal mutations become extreme enough to essentially eliminate the possibility of an evolutionary reversion to the simpler condition.

Unfortunately, the population-genetic requirements for the operation of CNE have not been formally worked out except in the case of evolution by gene duplication (covered in the following section). However, based on the theory outlined in the previous chapters, one can at least envision scenarios under which the process is most likely to proceed. All of these involve a relaxation in the efficiency of selection, in particular an initial A:B state that is no worse than very weakly deleterious, combined with a sufficiently small effective population size to render the initial transition effectively neutral.

In potential support of the CNE model, numerous examples exist in which molecular complexes with universally conserved functions have larger numbers of subunits in eukaryotes than in prokaryotes. Consider, for example, oxidative phosphorylation. Carried out in the mitochondria of eukaryotes, and on the plasma membranes of prokaryotes, this energy-generating mechanism involves multiple complexes with conserved functions throughout the Tree of Life. Well over 100 subunits encoded by different genes are distributed among the multiple electron-transport chain (ETC) complexes in eukaryotes, more than double the number found in bacteria (Hirst 2011; Huynen et al. 2013), and although most of these additions occurred prior to LECA, there have been numerous subsequent lineage-specific accruals. Nearly all of the accessory proteins are encoded in the nuclear genome. Although the favored explanation for their existence is their essential roles in maintaining structural stability of the complexes, the larger eukaryotic complexes are no more stable than those in bacteria. It has been argued that the subunit additions evolved as structural compensations for defects in the mitochondrially encoded components (resulting from deleterious-mutation accumulation in organelle genomes; Chapter 23) (Angerer et al. 2011; Hirst 2011; van der Sluis et al. 2015). However, a CNE scenario in which structural dependency arose as a consequence rather than a cause of subunit recruitment has not been ruled out.

A second example of the apparently gratuitous evolution of complexity involves the ribozyme RNase P, a complex of proteins surrounding a single catalytic RNA molecule that processes precursor transfer RNAs to their mature form. Although the RNA subunit is similar in all organisms, bacterial RNase P consists of just a single protein, whereas the archaeal and eukaryotic complexes contain five to ten proteins. This is a substantial investment in complexity for an enzyme whose sole role is to cleave a single phosphodiester bond. Again, the primary function of the additional proteins appears to be in stabilizing the overall complex, although there is no evidence that the eukaryotic RNase P is exceptionally stable (Lan et

al. 2018). Thus, such dependencies likely arose secondarily as initially fortuitous interactions became entrenched by the accumulation of otherwise harmful mutations for interfacial residues. Indeed, whereas the RNA core of the bacterial complex is internally stabilized by tertiary RNA-RNA interactions, these structural RNA features are reduced in archaeal and eukaryotic RNAs (Gopalan et al. 2018), as expected under the CNE model.

It has been argued that the evolution of higher-order RNase P complexes is a by-product of their having evolved additional cellular functions (Gopalan et al. 2018), but the possibility that any such functions could also be carried out by less elaborate structures has not been ruled out. In fact, a few bacteria and eukaryotes have lost the RNA component of RNase P and carry out the usual function solely with an enzymatic protein complex, showing that a simpler structure can indeed suffice. Complementation studies have shown that these RNA-free proteins will function with no apparent harmful effects when they are expressed in species that normally utilize RNA-containing RNase P (Weber et al. 2014; Lechner et al. 2015; Nickel et al. 2017).

Although a number of open questions remain, the simplest explanation for these observations on ETC complexes and RNase P is that excess complexity arose within eukaryotes by effectively neutral processes, the result being the conservation of ancestral functions but with increased bioenergetic cost to the organism. Other examples of apparent overdesign of eukaryotic features include the circadian clock, which typically is based on products of no more than three genes in prokaryotes (Chapter 18) but involves a complex web of many more genes in eukaryotes (Sancar 2008), and the spliceosome, a complex of five RNAs and dozens of proteins involved in intron splicing, which evolved from a single-component self-splicing intron in prokaryotes (Lynch 2007). Elaborating on earlier ideas of Stoltzfus (1999) and Lukeš et al. (2011), we now consider in more depth another potential example of CNE involving an even larger ribonucleoprotein complex, the ribosome.

Ribosomes. In all cells in all organisms, the ribosome has a single, conserved role – the translation of messenger RNAs. The catalytic core of the ribosome consists of three to four ribosomal RNAs (rRNAs), which collectively operate as a complex ribozyme. However, no ribosome can operate unless the rRNAs are coassembled with dozens of structural proteins. The question of why a molecular machine of this sort would require such a large endowment of protein components is further motivated by the substantial variation in the set of ribosomal proteins utilized in different phylogenetic lineages.

Thirty-four ribosomal proteins are universally deployed in all eukaryotes and prokaryotes and often referred to as the common core. However, there are also at least 34 ribosomal proteins shared by eukaryotes and archaea but absent from bacteria, whereas bacteria share no ribosomal protein just with eukaryotes or just with archaea (Lecompte et al. 2002; Hartman et al. 2006). This phylogenetic distribution is entirely consistent with the hypothesis that bacteria form an outgroup to archaea/eukaryotes (Chapter 3).

Not only do the protein constituents of ribosomes vary among the major domains of life, but the numbers of distinct proteins deployed vary as well. Each domain harbors unique ribosomal proteins not found in either of the other groups.

In bacteria, ~ 21 and 33 proteins are deployed in the small and large ribosomal subunits (denoted SSUs and LSUs, and respectively responsible for decoding mRNA information, and forming peptide bonds). In eukaryotes, these numbers expand to 33 and 46, respectively (Melnikov et al. 2012), with most of the additional proteins joining the external surfaces of the ribosome, like rings on an onion (Hsiao et al. 2009).

The two major rRNAs, occupying the small and large subunits, also vary in size among organisms, with an average $\sim 50\%$ expansion of both in eukaryotes relative to prokaryotes, and with weak coordination in size changes between the two subunits (Figure 6.2). Most rRNA enlargements occur by the addition of expansion segments that leave the common core structure undisturbed (Petrov et al. 2014).

In eukaryotes, separate ribosomes are deployed in the cytosol and in mitochondria, with the rRNAs generally being encoded in their respective genomes, but the proteins of both almost always being nuclear-encoded. The rRNAs deployed within mitochondria are often reduced in size relative to those in bacteria. For example, the mammalian mitochondrial LSU rRNA contains less than a third of the number of nucleotides as its counterpart in the cytosolic ribosome (1559 vs. 5347) and only half that in typical bacteria, although those in yeast and other protists can be comparable in size to those in bacteria.

Substantial modifications in the protein contribution to ribosomal structure have also evolved in mitochondria. Despite having to typically translate just a dozen or so mitochondrial genes, the protein repertoire of mitochondrial ribosomes is typically quite large. Overall, mitochondrial ribosomes contain 10 to 20 proteins not found in their alphaproteobacterial ancestors, with these again largely being distributed over the ribosome surface (Desmond et al. 2011). For example, the human mitochondrial LSU contains 48 proteins, all of which are encoded in the nuclear genome and 21 of which are mitochondrial-specific (Brown et al. 2014). Eleven of these 21 are not found in the yeast mitochondrial ribosome LSU, which nevertheless contains 39 proteins (Amunts et al. 2014).

The overall picture that one gets from the above is that ribosome expansion likely followed the emergence of eukaryotes, with further gains and losses then occurring on individual lineages, and with all such changes leaving the internal catalytic core intact. However, not just the structure of the ribosome, but also the pathways involved in ribosome biogenesis became more elaborate in eukaryotes (Strunk and Karbstein 2009). In bacteria, ribosome assembly involves no more than a handful of additional proteins, whereas on the order of 200 accessory proteins are essential for the development of mature eukaryotic ribosomes. The operation of many of these ribosome-biogenesis proteins requires hydrolysis of nucleotide triphosphates (ATP or GTP) and hence is energetically demanding. Thus, given the expanded number of nucleotides and amino acids in eukaryotic ribosomal RNAs and proteins, it is clear that the overall energetic cost of the translational machinery in eukaryotes is substantially greater than that in prokaryotes.

Given their association with organismal complexity, it has been argued that ribosome expansions and elaborations reflect a long-term pattern of adaptive divergence of ribosome architecture (Petrov et al. 2014, 2015). However, such a view is confronted with two fundamental problems: 1) the apparent inability of prokaryotes to achieve such changes despite having existed for longer periods of time and in

much larger populations; and 2) the absence of evidence that either the expansion segments of rRNAs or the additional ribosomal proteins confer any intrinsic benefits or novel functions.

The maximum rate of translation per ribosome (codons per second) in eukaryotes $\simeq 17$ in *Neurospora crassa* (Alberghina et al. 1975), 10 in *Saccharomyces cerevisiae* (Boehlke and Friesen 1975; Waldron and Lacroute 1975; Bonven and Gullov 1979), and 6 in mouse embryonic stem cells (Ingolia et al. 2011). Estimates in bacteria are 20 in *E. coli* (Forchhammer and Lindahl 1971; Dennis and Bremer 1974; Young and Bremer 1976), 16 in *Staphylococcus aureus* (Martin and Iandolo 1975), and 3 in *Streptomyces coelicolor* (Cox 2004). Although some of these estimates are likely more reliable than others, there is no indication of an elevated processing rate in larger eukaryotic ribosomes. Nor is there any indication that translation accuracy is improved in eukaryotes (Chapter 20).

These kinds of observations have not inhibited some authors from claiming that ribosomes are optimally designed. Focusing on *E. coli,* Reuveni et al. (2017) have argued that ribosomes consist of large numbers of similarly sized but unusually small proteins and have a heavy endowment of rRNA because such features maximize cellular efficiency. However, this conclusion seems to be another example of the perils of the adaptive-paradigm syndrome – the inevitable arrival at some kind of optimization argument if one searches hard enough. In fact, the proposed hypothesis is readily rejected upon a closer look at the data (Wei and Zhang 2018). Likewise, although Kostinski and Reuveni (2020) argue that the 2:1 mass ratio for rRNA: protein in bacterial ribosomes maximizes growth rate, their analysis is conditional on other ribosomal features, and fails to address why proteins are required at all.

Evolution by Gene Duplication

We now turn to a major route to the evolutionary origin of novelty and complexity with an ample body of empirical support. Although much of the theory reviewed in the previous chapter focused on small incremental changes to individual genes, such as single-nucleotide substitutions, larger-scale changes are also common. Duplications of entire genes or fragments thereof are of special interest because they generally contain fully functional domains tested under a prior history of selection. In this sense, novel gene functions do not have to be built from scratch, but more often than not can arise as elaborations of pre-existing functions. The potential contribution of gene duplication to evolutionary innovation is substantial, as individual genes duplicate at rates that are comparable to the rates at which base-substitution mutations arise at individual nucleotide sites (Lynch and Conery 2000; Konrad et al. 2018).

The fates of duplicate genes depend on the mechanisms by which they arise and the population-genetic environments within which they reside. Owing to the random breakpoints of duplicated DNA spans, duplication events will not necessarily encompass the full regulatory and/or coding regions of parental genes, and hence may have divergent features at birth (Katju and Lynch 2006). At the other extreme, exceptional cases involve whole-genome duplication events in which all genes are

simultaneously duplicated in entirety. Such events are known to have occurred in the ancestry of numerous eukaryotic lineages, including yeast (Wolfe and Shields 1997), ciliates (Aury et al. 2004; McGrath et al. 2014), vertebrates (Jaillon et al. 2004; Chain and Evans 2006; Putnam et al. 2008), arthropods (Kenny et al. 2016; Li et al. 2018), and land plants (Soltis and Soltis 2016).

Like all mutations, gene duplicates are initially present in just a single copy in a single individual. This will also be true for genes arising by other mechanisms, such as fortuitous *de novo* origin from preexisting noncoding sequence (Wissler et al. 2013; Bornberg-Bauer et al. 2015; McLysaght and Hurst 2016; Neme and Tautz 2016; Vakirlis et al. 2020) or via horizontal transfer from exogenous sources (Keeling and Palmer 2008; Vos et al. 2015). Thus, all of the population-genetic issues fundamental to the establishment of point mutations (Chapter 5), and more, apply to gene duplication. To be successful in the long term, a new gene must first drift towards fixation, and having arisen to high frequency, must then be preserved by sufficiently strong selective forces to prevent rapid loss by degenerative mutation.

The vast majority of duplicates arising by single-gene duplications are lost from populations on time scales of less than a few million generations (Lynch and Conery 2000), most never even proceeding to fixation. Basic population-genetic principles (Chapter 5) indicate why. Letting N be the population size and assuming diploidy, in the absence of immediate positive (or negative) selection, a fraction $[1 - (1/2N)]$ of newly arisen gene duplicates will be lost by random genetic drift in an average of just $\sim 2\ln(2N_e)$ generations (Kimura and Ohta 1969), a flash on the evolutionary time scale, as $2\ln(2N_e) \simeq 43$ with N_e at the upper limit of 10^9. Moreover, the small remaining fraction, $1/(2N)$, that manages to drift to fixation is also expected to fall victim to silencing mutations relatively quickly unless a preservational mechanism is acquired. Letting μ denote the rate of appearance of gene-silencing mutations, the average time to gene inactivation is on the order of the mean waiting time for the appearance of a null mutation at one of the two loci, $\simeq 1/(2\mu)$ generations, which will generally be on the order of 10^6 or so generations (Watterson 1983; Lynch et al. 2001).

Although it is often argued that an increase in gene number is a sign of evolutionary success and superiority (e.g., Lane and Martin 2010), there is little support for this point of view. Indeed, the number of genes per genome is nearly decoupled from organismal complexity (Chapter 24). For example, the genomes of the most behaviorally sophisticated animals contain fewer genes than found in many protists and only a few-fold more than in most bacteria. Only a few hundred genes are conserved across the entire Tree of Life (Tatusov et al. 2003; Koonin et al. 2004), and there can even be substantial differences in the numbers of genes among individuals within a species. This being said, the evidence is overwhelming that the repatterning of gene functions and gene locations by duplication events plays a central role in organismal diversification, although the connections often have little to do with adaptive processes.

The goals here are to summarize the ways in which gene duplication opens up novel pathways for evolutionary elaboration, provide insight into how the likelihoods of such processes are influenced by the population-genetic environment, and address some of the concerns with the more general model of constructive neutral evolution. More thorough reviews have appeared on the rates of origin, fates, and consequences

of duplicate genes (e.g., Lynch 2007; Conant and Wolfe 2008; Innan and Kondrashov 2010; Katju 2012). The small minority of duplicates that are retained for long periods of time are thought to owe their preservation to one of four mechanisms, one of which will first be dispensed with.

The masking effect. All populations harbor low-frequency, suboptimal alleles resulting from the recurrent introduction of deleterious mutations, and this has led to the common view that duplicate genes have an intrinsic selective advantage associated with their ability to mask the effects of deleterious mutations at the ancestral locus. However, the frequency at which a backup is useful is proportional to the incidence of deleterious genotypes at the opposite locus, which is on the order of the mutation rate to degenerative alleles. Thus, the selective advantage of a back-up gene is approximately equal to the rate of its own silencing by deleterious mutations. This leads to a miniscule selective advantage of the masking effect, generally smaller than the power of random genetic drift (Fisher 1935; Clark 1994; Lynch et al. 2001; Proulx and Phillips 2005).

The most serious and obvious challenge to the masking hypothesis for duplicate-gene retention is the general paucity of duplicate genes in haploid microbes despite their exceptionally high effective population sizes (which would maximize the efficiency of selection for weakly favorable redundancy). As will be discussed in Chapter 17, the energetic cost of a gene in bacterial species (relative to the total cellular energy budget) is generally sufficiently large for selection to efficiently remove redundant gene duplicates on this basis alone.

Neofunctionalization. Historically, the origin of a new gene function was thought to be the only preservational mechanism for the long-term survival of gene duplicates, with the much more common fate being the mutational silencing of one copy by degenerative mutations (Haldane 1933; Muller 1940; Ohno 1970). The idea here is that gene duplication can free one copy for evolutionary exploration and eventual acquisition of a new adaptive function. If the modifications underlying this new function are acquired at the expense of essential ancestral gene functions, the joint maintenance of both members of the pair will be enforced. A key issue here, of course, is that duplicate-gene preservation by neofunctionalization requires a setting in which there is indeed utility for a new gene function.

Neofunctionalization is expected to be more common in large populations for at least three reasons (Lynch et al. 2001; Walsh 2003). First, the larger the population size, the greater the population-level rate of origin of a rare neofunctionalizing mutation, and hence the higher the probability of fixation of such a mutation prior to one locus being silenced by a degenerative mutation. Second, in a sufficiently large population, even a duplicate gene initially destined to be lost by random genetic drift has a nontrivial chance of being rescued and propelled forward by a neofunctionalizing mutation. Third, in very large populations, the process need not depend on new neofunctionalizing mutations at all, as the requisite alleles may be maintained at low frequency by selection-mutation balance in the base population (but incapable of spreading to fixation prior to duplication because individuals lacking the essential ancestral allele are inviable). As discussed further in Chapter 13, cases

of balancing selection (e.g., heterozygote superiority) may sometimes maintain two alleles with single genetic loci at moderate frequencies (Spofford 1969).

Subfunctionalization. With the emergence of genome-sequence data in a wide variety of lineages, it became clear that the levels of retention of duplicate genes following whole-genome duplication events are far too high to be consistent with a model in which most are preserved by the evolution of novel functions. Given the mutation rate to degenerative mutations, they are also far too high to be fortuitous avoidances of gene silencing. Thus, something other than neofunctionalization must often be responsible for duplicate-gene retention. The fact that the vast majority of newly arising mutations are deleterious, combined with the emerging understanding of gene-structural complexity, suggested a resolution to this dilemma – a mechanism by which duplicate-gene preservation can be completely driven by degenerative mutations. Under the DDC (duplication-degeneration-complementation) model, both members of a gene pair acquire complementary negative changes that necessitate joint preservation (Force et al. 1999; Lynch and Force 2000; Lynch et al. 2001).

In the case of multifunctional genes, subfunctionalization can involve the partitioning of independently mutable, essential gene functions, leading to specialized copies with nonoverlapping features (qualitative subfunctionalization; Figure 6.3). Subfunctionalization can also be instigated by partial reduction in the efficiencies of the same functions in both members of a pair down to the total level required in the single-copy state (quantitative subfunctionalization) (Lynch and Force 2000; Duarte et al. 2006; Gout and Lynch 2015; Thompson et al. 2016). In both cases, subfunctionalization eliminates the need for beneficial mutations in the gene preservational process, although this need not rule out the emergence of secondary, adaptive modifications, as noted in the following section.

Contrary to the situation with neofunctionalization, the probability of subfunctionalization is expected to diminish with increasing effective population sizes, for at least three reasons (Lynch et al. 2001; Walsh 2003). First, there are prices to be paid for a pair of subfunctionalized genes. With respect to the coding region, the system will be roughly twice as mutationally vulnerable as a single-copy gene, thereby imposing a selective disadvantage equivalent to the null mutation rate per gene; and as just noted, there will also be an energetic cost of duplicate-gene maintenance and operation. As both of these costs are relatively small, they will only be opposed by selection in large populations. Second, a subfunctionalized allele en route to fixation is vulnerable to acquiring secondary silencing mutations, and the likelihood of such an effect is magnified in a large-N_e setting owing to the longer time to drift to fixation. Finally, qualitative subfunctionalization requires the presence of independently mutable regulatory mechanisms or protein domains, and as discussed below, the evolution of such modularity is reduced in large-N_e settings.

Adaptive-conflict resolution. The action of subfunctionalization and/or neofunctionalization may lead to gene copies that are not only largely distinct from each other but also have improved functionalities relative to what is possible with a single-copy ancestral gene (Piatigorsky and Wistow 1991; Hughes 1994; Stoltzfus 1999). Consider, for example, a single-copy locus subject to a "jack-of-all-trades

is a master-of-none" syndrome, i.e., with an adaptive conflict between its subfunctions. In such a situation, following duplication, complementary loss-of-subfunction mutations may alter the selective landscape experienced by the two pair members, enabling each copy to become more refined to a specific subset of tasks, potentially even opening up previously unavailable pathways to neofunctionalization. By this means, two of the most common forms of genomic upheaval, gene duplication and degenerative mutation, can provide a unique mechanism for the creation of novel evolutionary opportunities through the elimination of pleiotropic constraints. Again, however, whether such an adaptive-conflict resolution leads to a net selective advantage will depend on the degree to which the improvement(s) in gene functions exceed the cost of maintaining two genes.

A variant on the adaptive-conflict model is the IAD (innovation-amplification-divergence) model of Bergthorsson et al. (2007), which postulates that a common path to the origin of a new function starts with the duplication of a gene with a promiscuous secondary function, which in times of extreme need might suffice to provide enough functional rescue to buy time for further evolutionary refinement. Additional duplications would increase the number of mutational targets for such improvements, with deletions of the excess copies after establishment of the neofunctionalized gene eliminating the cost of gene amplification. Näsvall et al. (2012) and Newton et al. (2017) demonstrated the operation of this mechanism in the bacterium *Salmonella,* focusing on a gene involved in histidine biosynthesis with weak promiscuous involvement in tryptophan biosynthesis. When placed on a genetic background lacking the primary tryptophan synthesis pathway, evolutionary rescue was accomplished as duplicates of the histidine gene arose, and in some cases became specialized to alternative pathways leading to tryptophan. Other examples of this sort will be discussed in Chapter 19, where it will be shown that adaptive-conflict resolution and gene duplication plays a major role in the evolutionary remodeling of metabolic pathways.

The Case for Subfunctionalization

Prior to the development of the DDC model, circumstantial evidence for duplicate-gene preservation via subfunctionalization was suggested by studies of polyploid fishes, which repeatedly revealed tissue-specific expression of duplicated enzyme loci (Ferris and Whitt 1977, 1979). Such observations have now been supplemented by a wide array of investigations in other ray-finned fish lineages, zebrafish in particular, all of which arose following a whole-genome duplication event (e.g., Pasquier et al. 2017). Without an outgroup, it is difficult to determine whether duplicate-gene specialization is an outcome of neofunctionalization vs. subfunctionalization. However, the evolutionary interpretations of divergent-expression-patterns of duplicate genes in fishes have been greatly facilitated by observations of orthologous single-copy genes in tetrapods (usually mouse or chicken). These lineages, which branched off prior to the ray-finned fish-specific polyploidization event, generally reveal the presence of both gene subfunctions in their single-copy gene. Similar observations have been made in the tetraploid frog *Xenopus laevis* in comparison to its diploid

relatives (Morin et al. 2006; Sémon and Wolfe 2008), as well as in numerous land plants (Rutter et al. 2012; Freeling et al. 2015). Indeed, there are now hundreds of examples of qualitative subfunctionalization of duplicated genes via the partitioning of tissue-specific expression in multicellular organisms.

Although this particular mechanism of duplicate-gene preservation (i.e., tissue-specific expression divergence) is unavailable to unicellular species, there many other potential paths to subfunction partitioning in such organisms. For example, gene products may become specialized for use in different subcellular locations. Genes can also be regulated in modular ways with respect to timing of expression during the cell cycle or in response to different environmental conditions. In addition, proteins that assemble as homomeric multimers can acquire complementary interfacial changes after duplication, enforcing assembly as heteromers between the duplicate-gene products (Diss et al. 2017; Chapter 13).

Thus, although unicellular species often have large effective population sizes, which might be expected to reduce the incidence of subfunctionalization, the process is by no means restricted to multicellular species. Indeed, as outlined in subsequent chapters, key episodes of the process may have been facilitated during small-N_e phases in early eukaryotic history. A striking example of subfunctionalization deep in the eukaryotic phylogeny involves the dynamin family of proteins, which are used to pinch membranes. Phylogenetic analysis suggests the presence in LECA of a bifunctional dynamin with dual roles in vesicle scission from cell membranes and in mitochondrial division (Purkanti and Thattai 2015; Leger et al. 2015). Although this dual-function gene is retained in numerous eukaryotic lineages, following duplications in three independent lineages, the two copies became specialized to the two alternative ancestral functions.

Given the enormous amount of cell biological work done on yeast, and the whole-genome duplication that preceded the emergence of *Saccharomyces cerevisiae* (Wolfe and Shields 1997), much has been learned about the mechanisms preserving duplicate genes in this species. In particular, empirical studies in which *S. cerevisiae* duplicates have been replaced with the single-copy gene from a closely-related outgroup species have provided some of the most compelling evidence for subfunctionalization (van Hoof 2005). For example, Orc1 and Sir3 are sister genes in *S. cerevisiae*, with the former playing a role in chromosomal replication and the latter being part of a nucleosome-binding complex involved in chromosome-silencing functions. In *Kluyveromyces lactis,* a related taxon that branched off prior to whole-genome duplication, both functions are carried out by a single-copy gene (Hickman and Rusche 2010).

An example of subfunctionalization's role in adaptive-conflict resolution has also been revealed by molecular dissection in *S. cerevisiae,* where two sister genes are involved in galactose utilization, one (Gal3) playing a regulatory role in pathway induction and the other (Gal1) serving as a galactokinase (Hittinger and Carroll 2007). Again by reference to *K. lactis,* it was determined that the ancestral single-copy gene served both functions. Gene duplication then allowed the refinement of binding-site configurations that had previously been constrained in the ancestral gene, thereby enabling the emergence of a much more tightly regulated system (Figure 6.4).

A striking example of subfunctionalization based on structural alterations in

yeast is provided by the hexameric membrane ring for the vacuolar ATP synthase pump (Figure 6.5). In most eukaryotes, the ring consists of five copies of one protein (Vma16) and one of another (Vma3), both of which arose from an ancient gene duplication. In fungi, a third duplicate (Vma11) that arose by duplication of Vma3 replaces one subunit of Vma16, specifically residing between Vma16 and Vma3. Experimental modifications of the subunit interfaces revealed that one side of Vma3 lost the ability to bind to one side of Vma16, whereas the other side of Vma11 lost the ability to bind to Vma3 (Finnigan et al. 2012). There is no evidence that this increase in the complexity of vacuolar ATP synthase has endowed yeast with increased fitness.

As noted above, it is unlikely that duplicate genes are selectively preserved on the basis of having backup features. Nonetheless, observations from *S. cerevisiae* show that such properties can arise fortuitously as an indirect consequence of overlapping gene functions retained after partial subfunctionalization. For example, two ancient yeast paralogs, Sir2 and Hst1, operate as histone deacetylases with rather different functions in the cell (Hickman and Rusche 2007). However, when one gene is absent, the other can partially compensate by engaging in the noncognate function. Comparison with a pre-duplication outgroup species makes clear that this is a case of quantitative subfunctionalization, illustrating the risks of assuming that because duplicate genes have redundant functions, they must have been preserved on the basis of their backup capacities.

Finally, a potentially common mode of duplicate-gene preservation in eukaryotes involves the partitioning of gene functions via the modification of transit signals for localization of mRNAs and/or proteins to particular subcellular regions (Kumar et al. 2002; Silva-Filho 2003; Krogan et al. 2006). For example, immediately after transcription, eukaryotic mRNAs are typically decorated with one or more RNA-binding proteins, many of which attach to specific motor proteins for delivery to a specific subcellular location prior to translation (Besse and Ephrussi 2008; Holt and Bullock 2009; Buxbaum et al. 2015). There are numerous cases in which modifications of transit signals in post-duplication genes have lead to sub- or neolocalization. As a case in point, following duplication of one of the subunits of cytochrome c oxidase (a terminal complex in the electron transport chain) in an ancestral vertebrate, one member came to specialize on localization to the mitochondrion, whereas the other is delivered to the golgi (Schmidt et al. 2003). Likewise, NADP-dependent isocitrate dehydrogenase has been duplicated independently in both yeast and mammals, and in both cases the descendant copies partitioned their localizations to either the nucleus or the cytoplasm (Nekrutenko et al. 1998; Szewczyk et al. 2001).

Marques et al. (2008) suggest that about a third of duplicated genes surviving the whole-genome duplication in the ancestry of *S. cerevisiae* exhibit spatial subcellular partitioning, and similar estimates have been given for other taxa. For example, up to 25% of gene duplicates in the plant *Arabidopsis* (another descendant of a whole-genome duplication event) have experienced relocalization or sublocalization of their gene products (Byun and Singh 2013; Liu et al. 2014). There is, however, some uncertainty as to whether such partitioning is typically a cause or consequence of duplicate-gene preservation, as singleton genes in *S. cerevisiae* also appear to frequently acquire novel relocalization patterns (Qian and Zhang 2009). Although bacteria also exhibit spatial organization of translation (Montero Llopis et

al. 2010; Nevo-Dinur et al. 2011), this is dictated primarily by the cellular locations of genes on the chromosome. In this case, there appears to be less opportunity for partitioning subcellular localization following gene duplication, and indeed duplicate genes are rare in prokaryotes.

The Emergence of Modular Gene Subfunctions

Taken together, these results (along with many others to appear in subsequent chapters) make clear that duplicate-gene subfunctionalization has played a major role in the evolution of structural and enzymatic features of eukaryotic cells. However, few examples have been revealed in prokaryotes. One simple reason for the rarity of subfunctionalization in prokaryotes is the population-size constraint associated with the mutational and energetic costs of duplicate genes (Adler et al. 2014), but another is the general absence of independently mutable regulatory elements and localization zipcodes necessary for subfunction partitioning. This raises the broader question as to how modular gene architectural features essential to subfunctionalization actually evolve in the first place.

Resolution of this matter resides in the fact that the same types of duplication and degeneration processes that lead to the subfunctionalization of duplicate genes promote the emergence of the subfunctions themselves (Force et al. 2005). To simplify discussion, we will assume that subfunctions are defined by transcription-factor binding sites (TFBSs) or integrated regions of such sites (simply referred to here as promoters) that are separable from other such sites, both mutationally and functionally (as further elaborated on in Chapter 21). However, the same principles apply to subfunctions defined by functional motifs in coding regions, binding interfaces in multimers, or any other gene features that can be mutationally separated.

The goal here is to understand how a gene that is initially ubiquitously controlled in the same manner under all conditions comes to be regulated by more specialized mechanisms while retaining the same overall expression pattern. The process envisioned, subfunction fission, involves the progressive reconfiguration of a general-purpose enhancer via consecutive processes of partial duplication and loss of regulatory information, with each step proceeding in a nearly neutral fashion (Figure 6.6). The first phase involves the accretion of new regulatory elements, followed by the degeneration of one or more ancestral sites to yield two semi-independent promoters. The second phase involves tandem duplication of the regulatory region, followed by the formation of two entirely independent regulatory subfunctions by complementary degenerative mutations. Other than the fact that smaller DNA elements are involved, the events during the second phase are conceptually identical to those noted above for the subfunctionalization of entire genes.

Under this model, there is not necessarily a permanent allelic state, as the alternative classes of shared and semi-independently regulated alleles are free to mutate back and forth (hence, the two-way arrows in the top left of Figure 6.6). Thus, it is necessary to consider the circumstances under which semi-independently regulated alleles are likely to rise to high frequency, as this is a requirement for completing the transition to an allele with two entirely independent subfunctions.

There are three reasons why gene structure is more likely to gravitate to the subfunctionalized modular state in small populations. First, the stochastic gain of specific regulatory elements can occur either by *de novo* mutation to an appropriate motif within existing sequence or by the insertion of a pre-existing element via duplication from alternative genomic sites. The rate of *de novo* origin of an appropriate TFBS motif by mutation will depend on the mutation rate per nucleotide site and the mutational target size (amount of intergenic spacer DNA), both of which scale approximately inversely with N_e (Chapter 4; Lynch 2007). Second, the large, more gene-laden genomes of species with small N_e (e.g., eukaryotes vs. prokaryotes) have more potential sources of TFBSs for duplicative transpositions. Third, although alleles with more complex regulatory regions have a higher mutational vulnerability and impose an excess energetic cost at the DNA level, both of these are small effects that will only be efficiently opposed by selection in populations with large N_e.

The salient point here is that the same population-genetic environments that favor the subdivision of gene functions following gene duplication are expected to favor the emergence of gene-structural architectures necessary to fuel subfunctionalization. Such reinforcement provides further support for the contention that reductions in N_e, which naturally occurred as eukaryotes arose (and was further exacerbated in the metazoan and land-plant lineages), promoted a setting for the passive evolution of complexity with essentially no involvement of positive selection. Consistent with such a march towards complexity is the observation that whereas duplicate genes gradually lose their shared expression patterns over evolutionary time, the total numbers of regulatory motifs and interacting protein partners remain roughly constant for each member of the pair, suggesting an approximate balance between gains and losses of such elements. Such patterns have been observed in yeast (Papp et al. 2003; He and Zhang 2005), mammals (Huminiecki and Wolfe 2004), and *Arabidopsis* (Arsovski et al. 2015).

Taken together, these observations raise significant questions about the frequently assumed necessity and sufficiency of natural selection as a determining force in the emergence of complex patterns of gene regulation and protein deployment. In sufficiently small populations, modular forms of gene structure are expected to emerge in the absence of any direct selection for such architectural features. In sufficiently large populations, such changes are opposed by selection (unless immediately accompanied by phenotypic advantages that substantially offset the mutational and energetic disadvantages).

The Passive Origin of Species via Gene Duplication

In addition to playing a central role in the evolutionary divergence of cellular traits within lineages, gene duplication also has a powerful indirect role in the second major engine of evolution – the process of speciation, i.e., the emergence of new phylogenetic lineages (Lynch and Force 2000). Genetic theories of speciation have traditionally focused on two competing hypotheses (reviewed in Orr 1996; Rieseberg 2001; Coyne and Orr 2004). The Bateman-Dobzhansky-Muller model postulates the accumulation of lineage-specific gene-sequence changes that are mutually incompat-

ible when brought together in a hybrid genome. The chromosomal model invokes the accumulation of genomic rearrangements that result in gene loss in hybrid backgrounds.

Both models are based on rather stringent assumptions. For example, the Bateman-Dobzhansky-Muller model invokes the evolution of mutually incompatible coadaptive complexes of epistatically interacting factors, few of which have been convincingly identified as instigating the speciation process (as opposed to being downstream consequences). Chromosomal models generally focus on major rearrangements, for which within-population fixation can be greatly inhibited by the reduction in fitness in chromosomal heterozygotes owing to problems during meiosis. Notably, the gene-duplication model for speciation is consistent with both the Bateman-Dobzhansky-Muller and the chromosomal models, while requiring fewer assumptions than either of them.

The passive reassignment of gene (sub)functions to novel locations following gene duplication is central to the gene-duplication model of speciation. To see this, consider a diploid ancestral species with an unlinked pair of duplicate autosomal genes, which then experience divergent non- or subfunctionalization in two descendent species. This results in different chromosomal locations of the active gene (Figure 6.7). Because the F_1 hybrids of such species will be "presence-absence" heterozygotes at the two independently segregating loci, 1/4 of the F_1 gametes will contain null (absentee) alleles at both loci. In a predominantly haploid species, this single divergently resolved duplication would result in an expected 25% reduction in functional progeny. In a predominantly diploid species, $(1/4)^2 = 1/16$ of the F_2 offspring from the interspecific cross would lack functional alleles at both loci, and another 1/4 would carry only a single functional allele. Thus, if the gene is haploinsufficient, 5/16 of the F_2 zygotes of such a cross would be inviable (and/or sterile). With n divergently resolved duplicates, the expected fitness of hybrid progeny is $W = (1 - \delta)^n$, with δ denoting the reduction in hybrid fitness per map change. For example, with $\delta = 5/16$, as in a zygotically-acting haploinsufficient viability gene, $W = 0.024$ in the F_2 generation when $n = 10$, and 5×10^{-17} when $n = 100$.

Observed rates of gene duplication indicate that this type of process is sufficiently powerful to yield nearly complete genomic incompatibility within a few million generations of cessation of gene flow (Lynch and Force 2000; Shpak 2005). This is also the approximate time scale over which postzygotic isolation generally occurs in animals (Parker et al. 1985; Coyne and Orr 1997; Sasa et al. 1998; Presgraves 2002; Price and Bouvier 2002). Unfortunately, knowledge on the timescale of speciation in unicellular organisms is scant. However, genomic comparisons of the yeasts *S. cerevisiae* and *Candida albicans* imply an overall rate of microchromosomal rearrangement of ~ 2.3 / lineage / MY (Seoighe et al. 2003‘), likely driven in large part by divergent resolution of duplicate genes, as further discussed below.

The gene-duplication model for speciation is effectively a chromosomal model, but because the rearrangements are microchromosomal, they are unlikely to cause significant pairing problems during meiosis. Such changes can then accumulate passively without any alteration in within-species fitness, only being revealed after crossing to a lineage with a deviant gene location. The gene-duplication model also masquerades as a Bateman-Dobzhansky-Muller model, in that reassignments of genes to new locations operate like epistatic interactions because the loss-of-

function phenotype is determined by the total number of active alleles at the two duplicate loci in hybrid progeny.

A key feature of the gene-duplication model is that speciation can occur without any molecular evolution at the gene-sequence level. All that is required is the reciprocal silencing of ancestral-gene (sub)functions in sister taxa following ancestral gene duplications. Nonetheless, this process can also proceed via paths of neofunctionalization provided the latter occurs in the ancestral gene copy in one lineage (Lynch and Force 2005), and this can lead to misinterpretations regarding the underlying genetic mechanism of postzygotic isolation. Often it is assumed that speciation is a by-product of local adaptation generating physiologically incompatible alleles. However, incompatibilities resulting from the neofunctionalization of a duplicate gene need not be a direct function of adaptive changes at the neofunctionalized locus, but simply an indirect consequence of relocation of the ancestral-gene function.

Of course, the divergent resolution of duplicate genes is by no means the only possible route to the origin of post-zygotic species isolating barriers. However, given the frequency of gene duplication, it is difficult to escape the conclusion that it is a common and pervasive mechanism for speciation. As an example, a duplicate pair of a genes involved in histidine biosynthesis was present in the ancestor of the plant *Arabidopsis thaliana,* with different copies becoming silent in different *A. thaliana* sublineages. When plants containing the reciprocally silenced genes are crossed, the hybrids (presence/absence heterozygotes at both loci) segregate out different haplotypes in the next round of gametes, with progeny lacking both copies being inviable (Bikard et al. 2009; Blevins et al. 2017). A similar scenario, involving a different gene duplication, has been found in the genus *Mimulus* (Zuellig and Sweigart 2018).

The fruit fly *Drosophila* has been one of the major workhorses for research on the genetics of speciation, and here there are also well-documented examples of the involvement of duplicate genes in reproductive isolation. In two cases, a strong phase of positive selection operating on single duplicate copies has been implicated (Ting et al. 2004; Greenberg et al. 2006), suggesting the possibility of neofunctionalization. But in some *D. melanogaster* × *D. simulans* hybrids, sterility appears to be a simple consequence of the movement of an essential gene to a new chromosomal location via an intermediate phase of gene duplication (and without a change in function) (Masly et al. 2006).

Finally, it bears emphasizing that under the gene-duplication model, certain groups of organisms are expected to be more prone to speciation than others. For example, for lineages experiencing a doubling in genome size (polyploidization), the process noted above will be essentially unavoidable, owing to the very large number of gene targets. Moreover, following the first map changes induced by reciprocal silencing in sister polyploid taxa, the thousands of duplicate pairs still remaining will be free to become divergently resolved in subsequently isolated lineages, potentially yielding a large number of nested speciation events, i.e., a species radiation.

A particularly striking example of reproductive isolation by this form of divergent resolution is provided by the *Paramecium aurelia* complex, consisting of at least 14 cryptic species of ciliates. All of these emerged after two ancestral whole-genome duplication events led to hundreds of map changes as ancestral single-copy genes came to be represented by one, two, or three copies located on different chromo-

somes (McGrath et al. 2014). Remarkably, although the members of the *P. aurelia* complex have evolved unique pairs of mating types, despite $> 10^8$ years of isolation, there has been no discernible morphological divergence among lineages.

Another observation that appears to be compatible with the gene-duplication model for the origin of isolating barriers involves the yeast *S. cerevisiae* and its close relatives, which exhibit hundreds of differences in gene-order changes resulting from divergently resolved pairs of gene duplicates following a whole-genome duplication (Scannell et al. 2006). Although the haploid offspring of crosses between such species are almost always sterile, engineering of the chromosomes to partially restore large-scale colinearity increases fertility to levels of $\sim 25\%$ (Delneri et al. 2003). Restoration to complete colinearity might have even a greater effect. Notably, Selmecki et al. (2015) demonstrated that whole-genome duplication in yeast can facilitate adaptation by providing more opportunities for modifying gene balance by large deletions and/or chromosome loss, all of which will lead to the chromosomal repatterning essential to the gene-duplication model of speciation.

The key point here is that as in the case of phenotypic change within lineages, ample mechanisms exist for the passive origin of new species via nonadaptive processes. One potential example of such a key event, touched upon in Chapter 3, involves the base of the eukaryotic lineage – the colonization of LECA by the mitochondrion. Considering the very large number of organelle-to-nucleus gene transfers that apparently occurred soon after the establishment of the mitochondrial progenitor (Martin et al. 1998), divergent resolution of duplicated organelle genes may have provoked the passive development of isolating barriers among basal eukaryotic lineages (Chapters 3 and 24).

Summary

- To minimize energetic costs and mutational vulnerability, natural selection is expected to favor simplicity over complexity. Yet, many aspects of cell biology are demonstrably over-designed, particularly in eukaryotes, and most notably in multicellular species.

- Constructive neutral evolution provides a vision for how organismal complexity can emerge by nonadaptive mechanisms. The key idea is that the fortuitous development of initially neutral interactions between different gene products can alter the selective environment in ways that enable the fixation of previously forbidden mutations, thereby leading to permanent mutual dependence. Although the formalities of the theory remain to be worked out, the model provides a plausible explanation for the origin of a wide variety of cellular features, including the large number of protein subunits associated with complexes such as the electron-transport chain and the ribosome.

- Gene duplication is one of the primary mechanisms for the origin of organismal

complexity. Although neofunctionalization of one member of a pair provides a facile route to the origin of novel gene features, duplicate genes are more commonly preserved by other nonadaptive mechanisms. Most notably, subfunctionalization occurs when complementary degenerative mutations result in the partitioning of ancestral gene functions. The probability of this outcome is elevated in populations with small effective sizes.

- The same processes that lead to subfunctionalization of duplicate genes promote the evolution of modular forms of gene structure upon which the process of subfunctionalization depends. Thus, by facilitating the recurrent emergence and partitioning of gene subfunctions, reduced effective population sizes can lead to the passive increase in organismal complexity without any direct selection for such changes.

- Gene duplication also provides a powerful mechanism for the passive origin of reproductively isolated species, particularly in lineages that have experienced whole-genome duplications, as has happened repeatedly throughout the eukaryotic phylogeny.

Literature Cited

Adler, M., M. Anjum, O. G. Berg, D. I. Andersson, and L. Sandegren. 2014. High fitness costs and instability of gene duplications reduce rates of evolution of new genes by duplication-divergence mechanisms. Mol. Biol. Evol. 31: 1526-1535.

Alberghina, F. A., E. Sturani, and J. R. Gohlke. 1975. Levels and rates of synthesis of ribosomal ribonucleic acid, transfer ribonucleic acid, and protein in *Neurospora crassa* in different steady states of growth. J. Biol. Chem. 250: 4381-4388.

Amunts, A., A. Brown, X. C. Bai, J. L. Llácer, T. Hussain, P. Emsley, F. Long, G. Murshudov, S. H. Scheres, and V. Ramakrishnan. 2014. Structure of the yeast mitochondrial large ribosomal subunit. Science 343: 1485-1489.

Angerer, H., K. Zwicker, Z. Wumaier, L. Sokolova, H. Heide, M. Steger, S. Kaiser, E. Nübel, B. Brutschy, M. Radermacher, et al. 2011. A scaffold of accessory subunits links the peripheral arm and the distal proton-pumping module of mitochondrial complex I. Biochem. J. 437: 279-288.

Arsovski, A. A., J. Pradinuk, X. Q. Guo, S. Wang, and K. L. Adams. 2015. Evolution of *cis*-regulatory elements and regulatory networks in duplicated genes of *Arabidopsis*. Plant Physiol. 169: 2982-2991.

Aury, J. M., O. Jaillon, L. Duret, B. Noel, C. Jubin, B. M. Porcel, B. Ségurens, V. Daubin, V. Anthouard, N. Aiach, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature 444: 171-178.

Bergthorsson, U., D. I. Andersson, and J. R. Roth. 2007. Ohno's dilemma: evolution of new genes under continuous selection. Proc. Natl. Acad. Sci. USA 104: 17004-17009.

Besse, F., and A. Ephrussi. 2008. Translational control of localized mRNAs: restricting protein synthesis in space and time. Nat. Rev. Mol. Cell Biol. 9: 971-980.

Bikard, D., D. Patel, C. Le Metté, V. Giorgi, C. Camilleri, M. J. Bennett, and O. Loudet. 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. Science 323: 623-626.

Blevins, T., J. Wang, D. Pflieger, F. Pontvianne, and C. S. Pikaard. 2017. Hybrid incompatibility caused by an epiallele. Proc. Natl. Acad. Sci. USA 114: 3702-3707.

Boehlke, K. W., and J. D. Friesen. 1975. Cellular content of ribonucleic acid and protein in *Saccharomyces cerevisiae* as a function of exponential growth rate: calculation of the apparent peptide chain elongation rate. J. Bacteriol. 121: 429-433.

Bonven, B., and K. Gulløv. 1979. Peptide chain elongation rate and ribosomal activity in *Saccharomyces cerevisiae* as a function of the growth rate. Mol. Gen. Genet. 170: 225-230.

Bornberg-Bauer, E., J. Schmitz, and M. Heberlein. 2015. Emergence of *de novo* proteins from 'dark genomic matter' by 'grow slow and moult'. Biochem. Soc. Trans. 43: 867-873.

Brown, A., A. Amunts, X. C. Bai, Y. Sugimoto, P. C. Edwards, G. Murshudov, S. H. Scheres, and V. Ramakrishnan. 2014. Structure of the large ribosomal subunit from human mitochondria. Science 346: 718-722.

Brunet, T. D. P., and W. F. Doolittle. 2018. The generality of constructive neutral evolution. Biol. Philos. 33: 2.

Buxbaum, A. R., G. Haimovich, and R. H. Singer. 2015. In the right place at the right time: visualizing and understanding mRNA localization. Nat. Rev. Mol. Cell Biol. 16: 95-109.

Byun, S. A., and S. Singh. 2013. Protein subcellular relocalization increases the retention of eukaryotic duplicate genes. Genome. Biol. Evol. 5: 2402-2409.

Chain, F. J., and B. J. Evans. 2006. Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog *Xenopus laevis*. PLoS Genet. 2: e56.

Clark, A. G. 1994. Invasion and maintenance of a gene duplication. Proc. Natl. Acad. Sci. USA 91: 2950-2954.

Conant, G. C., and K. H. Wolfe. 2008. Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. 9: 938-950.

Cox, R. A. 2004. Quantitative relationships for specific growth rates and macromolecular compositions of *Mycobacterium tuberculosis, Streptomyces coelicolor* A3(2) and *Escherichia coli* B/r: an integrative theoretical approach. Microbiology 150: 1413-1426.

Coyne, J. A., and H. A. Orr. 1997. "Patterns of speciation in *Drosophila*" revisited. Evolution 51: 295-303.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Assocs., Inc. Sunderland, MA.

Delneri, D., I. Colson, S. Grammenoudi, I. N. Roberts, E. J. Louis, and S. G. Oliver. 2003. Engineering evolution to study speciation in yeasts. Nature 422: 68-72.

Dennis, P. P., and H. Bremer. 1974. Macromolecular composition during steady-state growth of *Escherichia coli* B-r. J. Bacteriol. 119: 270-281.

Desmond, E., C. Brochier-Armanet, P. Forterre, and S. Gribaldo. 2011. On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. Res. Microbiol. 162: 53-70.

Diss, G., I. Gagnon-Arsenault, A. M. Dion-Coté, H. Vignaud, D. I. Ascencio, C. M. Berger, and C. R. Landry. 2017. Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. Science 355: 630-634.

Duarte, J. M., L. Cui, P. K. Wall, Q. Zhang, X. Zhang, J. Leebens-Mack, H. Ma, N. Altman, and C. W. dePamphilis. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. Mol. Biol. Evol. 23: 469-478.

Ferris, S. D., and G. S. Whitt. 1977. Duplicate gene expression in diploid and tetraploid loaches (Cypriniformes, Cobitidae). Biochem. Genet. 15: 1097-1112.

Ferris, S. D., and G. S. Whitt. 1979. Evolution of the differential regulation of duplicate genes after polyploidization. J. Mol. Evol. 12: 267-317.

Finnigan, G. C., V. Hanson-Smith, T. H. Stevens, and J. W. Thornton. 2012. Evolution of increased complexity in a molecular machine. Nature 481: 360-364.

Fisher, R. A. 1935. The sheltering of lethals. Amer. Natur. 69: 446-455.

Force, A., W. A. Cresko, F. B. Pickett, S. Proulx, C. Amemiya, and M. Lynch. 2005. The origin of subfunctions and modular gene regulation. Genetics 170: 433-446.

Force, A., M. Lynch, B. Pickett, A. Amores, Y.-L. Yan, and J. Postlethwait. 1999. Preservation of

duplicate genes by complementary, degenerative mutations. Genetics 151: 1531-1545.

Forchhammer, J., and L. Lindahl. 1971. Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of *Escherichia coli* 15. J. Mol. Biol. 55: 563-568.

Freeling, M., M. J. Scanlon, and J. E. Fowler. 2015. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. Curr. Opin. Genet. Dev. 35: 110-118.

Gopalan, V., N. Jarrous, and A. S. Krasilnikov. 2018. Chance and necessity in the evolution of RNase P. RNA 24: 1-5.

Gout, J. F., and M. Lynch. 2015. Maintenance and loss of duplicated genes by dosage subfunctionalization. Mol. Bio. Evol. 32: 2141-2148.

Gray, M. W., J. Lukes, J. M. Archibald, P. J. Keeling, and W. F. Doolittle. 2010. Cell biology. Irremediable complexity? Science 330: 920-921.

Greenberg, A. J., J. R. Moran, S. Fang, and C.-I. Wu. 2006. Adaptive loss of an old duplicated gene during incipient speciation. Mol. Biol. Evol. 23: 401-410.

Haldane, J. B. S. 1933. The part played by recurrent mutation in evolution. Amer. Natur. 67: 5-9.

Hartman, H., P. Favaretto, and T. F. Smith. 2006. The archaeal origins of the eukaryotic translational system. Archaea 2: 1-9.

He, X., and J. Zhang. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169: 1157-1164.

Hickman, M. A., and L. N. Rusche. 2007. Substitution as a mechanism for genetic robustness: the duplicated deacetylases Hst1p and Sir2p in *Saccharomyces cerevisiae.* PLoS Genet. 3: e126.

Hickman, M. A., and L. N. Rusche. 2010. Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication. Proc. Natl. Acad. Sci. USA 107: 19384-19389.

Hirst, J. 2011. Why does mitochondrial complex I have so many subunits? Biochem. J. 437: e1-e3.

Hittinger, C. T., and S. B. Carroll. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. Nature 449: 677-681.

Holt, C. E., and S. L. Bullock. 2009. Subcellular mRNA localization in animal cells and why it matters. Science 326: 1212-1216.

Hsiao, C., S. Mohan, B. K. Kalahar, and L. D. Williams. 2009. Peeling the onion: ribosomes are ancient molecular fossils. Mol. Biol. Evol. 26: 2415-2425.

Hughes, A. L. 1994. The evolution of functionally novel proteins after gene duplication. Proc. Roy. Soc. Lond. B 256: 119-124.

Huminiecki, L., and K. H. Wolfe. 2004. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. Genome Res. 14: 1870-1879.

Huynen, M. A., I. Duarte, and R. Szklarczyk. 2013. Loss, replacement and gain of proteins at the origin of the mitochondria. Biochim. Biophys. Acta 1827: 224-231.

Ingolia, N. T., L. F. Lareau, and J. S. Weissman. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147: 789-802.

Innan, H., and F. Kondrashov. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11: 97-108.

Jaillon, O., J. M. Aury, F. Brunet, J. L. Petit, N. Stange-Thomann, E. Mauceli, L. Bouneau, C. Fischer, C. Ozouf-Costaz, A. Bernot, et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431: 946-957.

Katju, V. 2012. In with the old, in with the new: the promiscuity of the duplication process engenders diverse pathways for novel gene creation. Int. J. Evol. Biol. 2012: 341932.

Katju, V., and M. Lynch. 2006. On the formation of novel genes by duplication in the *Caenorhabditis elegans* genome. Mol. Biol. Evol. 23: 1056-1067.

Keeling, P. J., and J. D. Palmer. 2008. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. 9: 605-618.

Kenny, N. J., K. W. Chan, W. Nong, Z. Qu, I. Maeso, H. Y. Yip, T. F. Chan, H. S. Kwan, P. W. H. Holland, K. H. Chu, and J. H. L. Hui. 2016. Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. Heredity 116: 190-199.

Kimura, M., and T. Ohta. 1969. The average number of generations until fixation of a mutant gene in a finite population. Genetics 61: 763-771.

Konrad, A., S. Flibotte, J. Taylor, R. H. Waterston, D. G. Moerman, U. Bergthorsson, and V. Katju. 2018. Mutational and transcriptional landscape of spontaneous gene duplications and deletions in *Caenorhabditis elegans.* Proc. Natl. Acad. Sci. USA 115: 7386-7391.

Koonin, E. V., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, D. M. Krylov, K. S. Makarova, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, et al. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 5: R7.

Kostinski, S., and S. Reuveni. 2020. Ribosome composition maximizes cellular growth rates in *E. coli.* Phys. Rev. Lett. 125: 028103.

Krogan, N. J., G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, et al. 2006. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae.* Nature 440: 637-643.

Kumar, A., S. Agarwal, J. A. Heyman, S. Matson, M. Heidtman, S. Piccirillo, L. Umansky, A. Drawid, R. Jansen, Y. Liu, et al. 2002. Subcellular localization of the yeast proteome. Genes Dev. 16: 707-719.

Lan, P., M. Tan, Y. Zhang, S. Niu, J. Chen, S. Shi, S. Qiu, X. Wang, X. Peng, G. Cai, et al. 2018. Structural insight into precursor tRNA processing by yeast ribonuclease P. Science 362: eaat6678.

Lane, N. 2020. How energy flow shapes cell evolution. Curr. Biol. 30: R471-R476.

Lane, N., and W. Martin. 2010. The energetics of genome complexity. Nature 467: 929-934.

Lechner, M., W. Rossmanith, R. K. Hartmann, C. Thölken, B. Gutmann, P. Giegé, and A. Gobert. 2015. Distribution of ribonucleoprotein and protein-only RNase P in Eukarya. Mol. Biol. Evol. 32: 3186-3193.

Lecompte, O., R. Ripp, J. C. Thierry, D. Moras, and O. Poch. 2002. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain

scale. Nucleic Acids Res. 30: 5382-5390.

Leger, M. M., M. Petrů, V. Žárský, L. Eme, Č. Vlček, T. Harding, B. F. Lang, M. Eliáš, P. Doležal, and A. J. Roger. 2015. An ancestral bacterial division system is widespread in eukaryotic mitochondria. Proc. Natl. Acad. Sci. USA 112: 10239-10246.

Li, Z., G. P. Tiley, S. R. Galuska, C. R. Reardon, T. I. Kidder, R. J. Rundell, and M. S. Barker. 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. Proc. Natl. Acad. Sci. USA 115: 4713-4718.

Liu, S. L., A. Q. Pan, and K. L. Adams. 2014. Protein subcellular relocalization of duplicated genes in *Arabidopsis.* Genome Biol. Evol. 6: 2501-2515.

Lukeš, J., J. M. Archibald, P. J. Keeling, W. F. Doolittle, and M. W. Gray. 2011. How a neutral evolutionary ratchet can build cellular complexity. IUBMB Life 63: 528-537.

Lynch, M. 2007. The Origins of Genome Architecture. Sinauer Assocs., Inc., Sunderland, MA.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science 290: 1151-1154.

Lynch, M., and A. Force. 2000. The probability of duplicate-gene preservation by subfunctionalization. Genetics 154: 459-473.

Lynch, M., M. O'Hely, B. Walsh, and A. Force. 2001. The probability of fixation of a newly arisen gene duplicate. Genetics 159: 1789-1804.

Marques, A. C., N. Vinckenbosch, D. Brawand, and H. Kaessmann. 2008. Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome Biol. 9: R54.

Martin, S. E., and J. J. Iandolo. 1975. Translational control of protein synthesis in *Staphylococcus aureus.* J. Bacteriol. 122: 1136-1143.

Martin, W., B. Stoebe, V. Goremykin, S. Hansmann, M. Hasegawa, and K. V. Kowallik. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393: 162-165.

Masly, J. P., C. D. Jones, M. A. Noor, J. Locke, and H. A. Orr. 2006. Gene transposition as a cause of hybrid sterility in *Drosophila.* Science 313: 1448-1450.

McGrath, C. L., J. F. Gout, P. Johri, T. G. Doak, and M. Lynch. 2014. Differential retention and divergent resolution of duplicate genes following whole-genome duplication. Genome Research 24: 1665-1675.

McLysaght, A., and L. D. Hurst. 2016. Open questions in the study of *de novo* genes: what, how and why. Nat. Rev. Genet. 17: 567-578.

Melnikov, S., A. Ben-Shem, N. Garreau de Loubresse, L. Jenner, G. Yusupova, and M. Yusupov. 2012. One core, two shells: bacterial and eukaryotic ribosomes. Nat. Struct. Mol. Biol. 19: 560-567.

Montero Llopis, P., A. F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet, and C. Jacobs-Wagner. 2010. Spatial organization of the flow of genetic information in bacteria. Nature 466: 77-81.

Morin, R. D., E. Chang, A. Petrescu, N. Liao, M. Griffith, W. Chow, R. Kirkpatrick, Y. S. Butterfield, A. C. Young, J. Stott, et al. 2006. Sequencing and analysis of 10,967 full-length cDNA clones from *Xenopus laevis* and *Xenopus tropicalis* reveals post-tetraploidization transcriptome

remodeling. Genome Res. 16: 796-803.

Muller, H. J. 1940. Bearing of the *Drosophila* work on systematics, pp. 185-268. In J. S. Huxley (ed.) The New Systematics. Clarendon Press, Oxford, UK.

Näsvall, J., L. Sun, J. R. Roth, and D. I. Andersson. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. Science 338: 384-387.

Nekrutenko, A., D. M. Hillis, J. C. Patton, R. D. Bradley, and R. J. Baker. 1998. Cytosolic isocitrate dehydrogenase in humans, mice, and voles and phylogenetic analysis of the enzyme family. Mol. Biol. Evol. 15: 1674-1684.

Neme, R., and D. Tautz. 2016. Fast turnover of genome transcription across evolutionary time exposes entire non-coding DNA to *de novo* gene emergence. eLife 5: e09977.

Nevo-Dinur, K., A. Nussbaum-Shochat, S. Ben-Yehuda, and O. Amster-Choder. 2011. Translation-independent localization of mRNA in *E. coli*. Science 331: 1081-1084.

Newton, M. S., X. Guo, A. Söderholm, J. Näsvall, P. Lundström, D. I. Andersson, M. Selmer, and W. M. Patrick. 2017. Structural and functional innovations in the real-time evolution of new $(\beta\alpha)(8)$ barrel enzymes. Proc. Natl. Acad. Sci. USA 114: 4727-4732.

Nickel, A. I., N. B. Wäber, M. Gößringer, M. Lechner, U. Linne, U. Toth, W. Rossmanith, and R. K. Hartmann. 2017. Minimal and RNA-free RNase P in *Aquifex aeolicus*. Proc. Natl. Acad. Sci. USA 114: 11121-11126.

Ohno, S. 1970. Evolution by Gene Duplication. Springer-Verlag, Berlin, Germany.

Orr, H. A. 1996. Dobzhansky, Bateson, and the genetics of speciation. Genetics 144: 1331-1335.

Papp, B., C. Pal, and L. D. Hurst. 2003. Evolution of *cis*-regulatory elements in duplicated genes of yeast. Trends Genet. 19: 417-422.

Parker, H. R., D. P. Philipp, and G. S. Whitt. 1985. Relative developmental success of interspecific *Lepomis* hybrids as an estimate of gene regulatory divergence between species. J. Exp. Zool. 233: 451-466.

Pasquier, J., I. Braasch, P. Batzel, C. Cabau, J. Montfort, T. Nguyen, E. Jouanno, C. Berthelot, C. Klopp, L. Journot, et al. 2017. Evolution of gene expression after whole-genome duplication: new insights from the spotted gar genome. J. Exp. Zool. B Mol. Dev. Evol. 328: 709-721.

Petrov, A. S., C. R. Bernier, C. Hsiao, A. M. Norris, N. A. Kovacs, C. C. Waterbury, V. G. Stepanov, S. C. Harvey, G. E. Fox, R. M. Wartell, et al. 2014. Evolution of the ribosome at atomic resolution. Proc. Natl. Acad. Sci. USA 111: 10251-10256.

Petrov, A. S., B. Gulen, A. M. Norris, N. A. Kovacs, C. R. Bernier, K. A. Lanier, G. E. Fox, S. C. Harvey, R. M. Wartell, N. V. Hud, et al. 2015. History of the ribosome and the origin of translation. Proc. Natl. Acad. Sci. USA 112: 15396-15401.

Piatigorsky, J., and G. Wistow. 1991. The recruitment of crystallins: new functions precede gene duplication. Science 252: 1078-1079.

Presgraves, D. C. 2002. Patterns of postzygotic isolation in Lepidoptera. Evolution 56: 1168-1183.

Price, T. D., and M. M. Bouvier. 2002. The evolution of F_1 postzygotic incompatibilities in birds. Evolution 56: 2083-2089.

Proulx, S. R., and P. C. Phillips. 2005. The opportunity for canalization and the evolution of

genetic networks. Amer. Natur. 165: 147-162.

Purkanti, R., and M. Thattai. 2015. Ancient dynamin segments capture early stages of host-mitochondrial integration. Proc. Natl. Acad. Sci. USA 112: 2800-2805.

Putnam, N. H., T. Butts, D. E. Ferrier, R. F. Furlong, U. Hellsten, T. Kawashima, M. Robinson-Rechavi, E. Shoguchi, A. Terry, J. K. Yu, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453: 1064-1071.

Qian, W., and J. Zhang. 2009. Protein subcellular relocalization in the evolution of yeast singleton and duplicate genes. Genome Biol. Evol. 1: 198-204.

Reuveni, S., M. Ehrenberg, and J. Paulsson. 2017. Ribosomes are optimized for autocatalytic production. Nature 547: 293-297.

Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16: 351-358.

Rutter, M. T., K. V. Cross, and P. A. Van Woert. 2012. Birth, death and subfunctionalization in the *Arabidopsis* genome. Trends Plant Sci. 17: 204-212.

Sancar, A. 2008. The intelligent clock and the Rube Goldberg clock. Nat. Struct. Mol. Biol. 15: 23-24.

Sasa, M. M., P. T. Chippindale, and N. A. Johnson. 1998. Patterns of postzygotic isolation in frogs. Evolution 52: 1811-1820.

Scannell, D. R., K. P. Byrne, J. L. Gordon, S. Wong, and K. H. Wolfe. 2006. Rapid speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440: 341-345.

Schmidt, T. R., J. W. Doan, M. Goodman, and L. I. Grossman. 2003. Retention of a duplicate gene through changes in subcellular targeting: an electron transport protein homologue localizes to the golgi. J. Mol. Evol. 57: 222-228.

Selmecki, A. M., Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shoresh, A. L. Sorenson, S. De, R. Kishony, F. Michor, R. Dowell, and D. Pellman. 2015. Polyploidy can drive rapid adaptation in yeast. Nature 519: 349-352.

Sémon, M., and K. H. Wolfe. 2008. Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. Proc. Natl. Acad. Sci. USA 105: 8333-8338.

Seoighe, C., C. R. Johnston, and D. C. Shields. 2003. Significantly different patterns of amino acid replacement after gene duplication as compared to after speciation. Mol. Biol. Evol. 20: 484-490.

Shpak, M. 2005. The role of deleterious mutations in allopatric speciation. Evolution 59: 1389-1399.

Silva-Filho, M. C. 2003. One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. Curr. Opin. Plant Biol. 6: 589-595.

Soltis, P. S., and D. E. Soltis. 2016. Ancient WGD events as drivers of key innovations in angiosperms. Curr. Opin. Plant Biol. 30: 159-165.

Spofford, J. B. 1969. Heterosis and the evolution of duplications. Amer. Natur. 103: 407-432.

Stoltzfus, A. 1999. On the possibility of constructive neutral evolution. J. Mol. Evol. 49: 169-181.

Strunk, B. S., and K. Karbstein. 2009. Powering through ribosome assembly. RNA 15: 2083-2104.

Szewczyk, E., A. Andrianopoulos, M. A. Davis, and M. J. Hynes. 2001. A single gene produces mitochondrial, cytoplasmic, and peroxisomal NADP-dependent isocitrate dehydrogenase in *Aspergillus nidulans.* J. Biol. Chem. 276: 37722-37729.

Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, et al. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4: 41.

Thompson, A., H. H. Zakon, and M. Kirkpatrick. 2016. Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes. Genetics 202: 765-774.

Ting, C. T., S. C. Tsaur, S. Sun, W. E. Browne, Y. C. Chen, N. H. Patel, and C.-I. Wu. 2004. Gene duplication and speciation in *Drosophila:* evidence from the *Odysseus* locus. Proc. Natl. Acad. Sci. USA 101: 12232-12235.

Vakirlis, N., O. Acar, B. Hsu, N. Castilho Coelho, S. B. Van Oss, A. Wacholder, K. Medetgul-Ernar, R. W. Bowman 2nd, C. P. Hines, J. Iannotta, et al. 2020. *De novo* emergence of adaptive membrane proteins from thymine-rich genomic sequences. Nat. Commun. 11: 781.

van der Sluis, E. O., H. Bauerschmitt, T. Becker, T. Mielke, J. Frauenfeld, O. Berninghausen, W. Neupert, J. M. Herrmann, and R. Beckmann. 2015. Parallel structural evolution of mitochondrial ribosomes and OXPHOS complexes. Genome Biol. Evol. 7: 1235-1251.

van Hoof, A. 2005. Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. Genetics 171: 1455-1461.

Vos, M., M. C. Hesselman, T. A. Te Beek, M. W. J. van Passel, and A. Eyre-Walker. 2015. Rates of lateral gene transfer in prokaryotes: high but why? Trends Microbiol. 23: 598-605.

Waldron, C., and F. Lacroute. 1975. Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. J. Bacteriol. 122: 855-865.

Walsh, B. 2003. Population-genetic models of the fates of duplicate genes. Genetica 118: 279-294.

Watterson, G. A. 1983. On the time for gene silencing at duplicate loci. Genetics 105: 745-766.

Weber, C., A. Hartig, R. K. Hartmann, and W. Rossmanith. 2014. Playing RNase P evolution: swapping the RNA catalyst for a protein reveals functional uniformity of highly divergent enzyme forms. PLoS Genet. 10: e1004506.

Wei, X., and J. Zhang. 2018. On the origin of compositional features of ribosomes. Genome Biol. Evol. 10: 2010-2016.

Wissler, L., J. Gadau, D. F. Simola, M. Helmkampf, and E. Bornberg-Bauer. 2013. Mechanisms and dynamics of orphan gene emergence in insect genomes. Genome Biol. Evol. 5: 439-455.

Woese, C. R. 1971. Evolution of macromolecular complexity. J. Theor. Biol. 33: 29-34.

Wolfe, K. H., and D. C. Shields. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708-713.

Young, R., and H. Bremer. 1976. Polypeptide-chain-elongation rate in *Escherichia coli* B/r as a function of growth rate. Biochem. J. 160: 185-194.

Zuckerkandl, E. 1997. Neutral and nonneutral mutations: the creative mix – evolution of complexity in gene interaction systems. J. Mol. Evol. 44 Suppl 1: S2-S8.

Zuellig, M. P., and A. L. Sweigart. 2018. Gene duplicates cause hybrid lethality between sympatric

species of *Mimulus*. PLoS Genet. 14: e1007130.

Figure 6.1. An idealized scenario by which increased complexity might arise by constructive neutral evolution – a transition from an independently functioning molecule A to an obligatory A:B interaction. Initially, a fortuitous interaction with B suppresses subsequent deleterious mutational effects in A (denoted by the red dot, which would otherwise be eliminated by selection), rendering A dependent on B. The complex then becomes further entrenched evolutionarily, as A acquires an additional conditionally silent mutation (red) that if exposed by elimination of B leads to loss of function. In the final stage, two additional mutations (white and black), potentially refining A:B function beyond its initial state, have become established. From Lukeš et al. (2011).

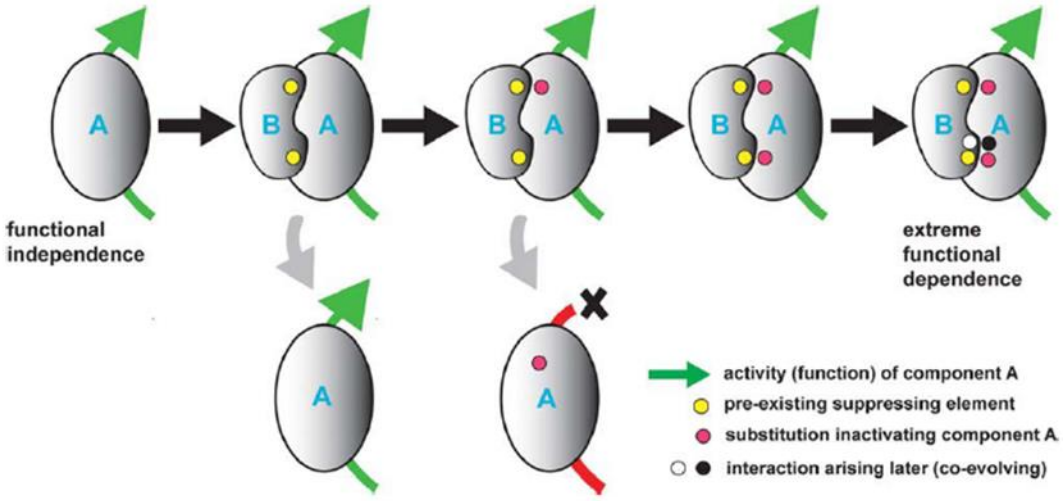
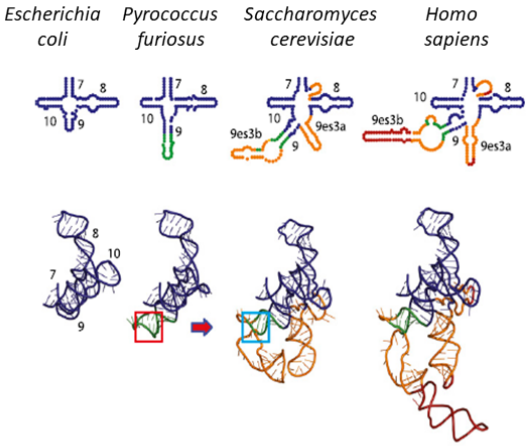


Figure 6.2. Ribosomal RNA evolution. **Left**) Adventitious growth of one particular set of helices associated with the SSU ribosomal RNA. The nearly invariant functional core is represented by the dark blue subsections. From Petrov et al. (2015). **Right**) Sizes of the small and large rRNA subunits in various taxa (in terms of sequence length). From Petrov et al. (2014).



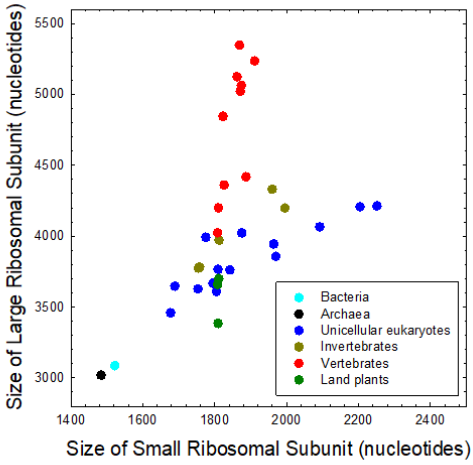


Figure 6.3. The DDC model for the alternative fates of duplicate genes. The ancestral gene is depicted as having two independently mutable subfunctions (blue and green), e.g., two regulatory elements, each driving expression in a particular tissue, subcellular location, or environmental condition. Solid boxes denote fully functional regulatory and coding regions, whereas open boxes denote loss of function, and a red box denotes the gain of a new beneficial function. Each pair of genes reflects the fixed state of the population. In this example, following the duplication event, the first degenerative mutation eliminates a subfunction from one of the copies. The second mutational event then dictates the final fate of the pair: subfunctionalization, with the second copy acquiring a complementary loss-of-subfunction mutation; neofunctionalization, with the second copy acquiring a novel, beneficial expression pattern at the expense of an ancestral subfunction; or nonfunctionalization, with the first copy losing all functional ability. From Force et al. (1999).

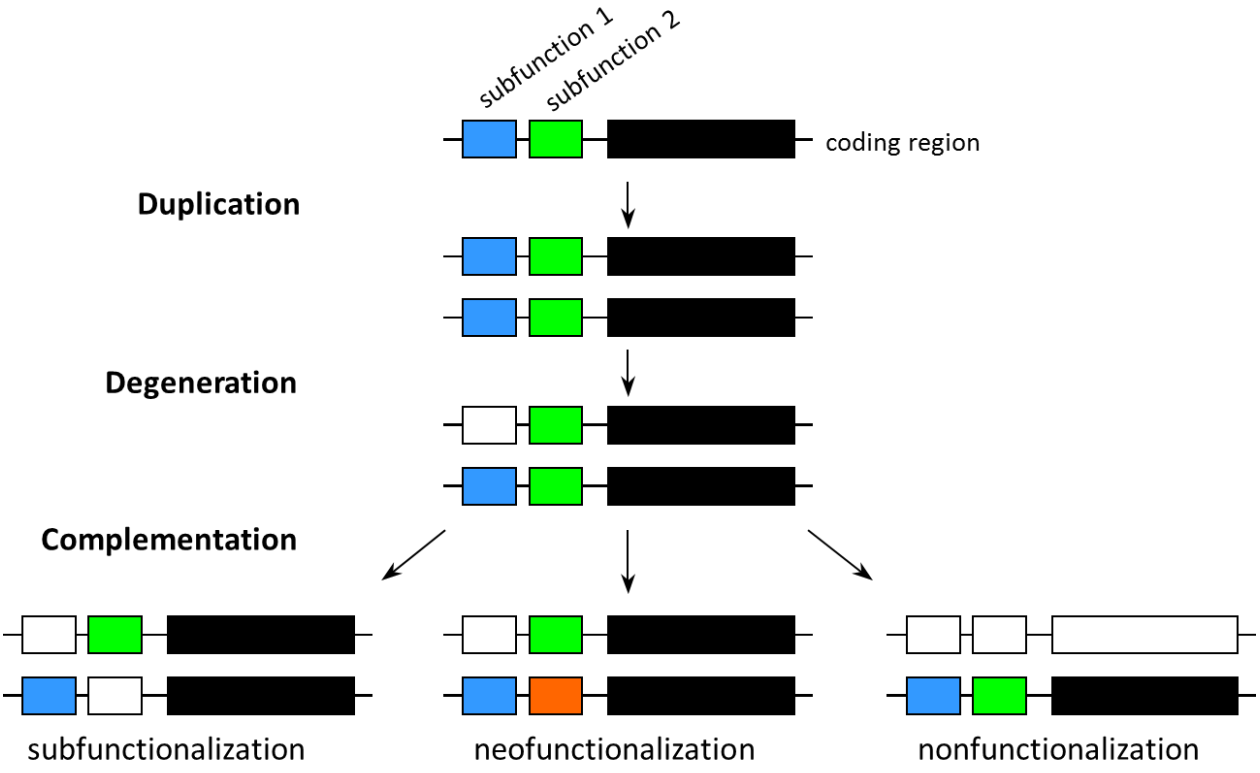


Figure 6.4. Evolution of the galactose utilization pathway following duplication and subfunctionalization of the ancestral gene (Gal1/3) in the yeast *S. cerevisiae*, giving rise to the duplicates Gal1 and Gal3. Blue represents the coding regions, and orange the binding sites of the transcription factor Gal4. The loss of binding sites in Gal3 and the rearrangement of sites in Gal1 removed an adaptive conflict involving the regulatory efficiency of the single-copy gene. From Hittinger and Carroll (2007).

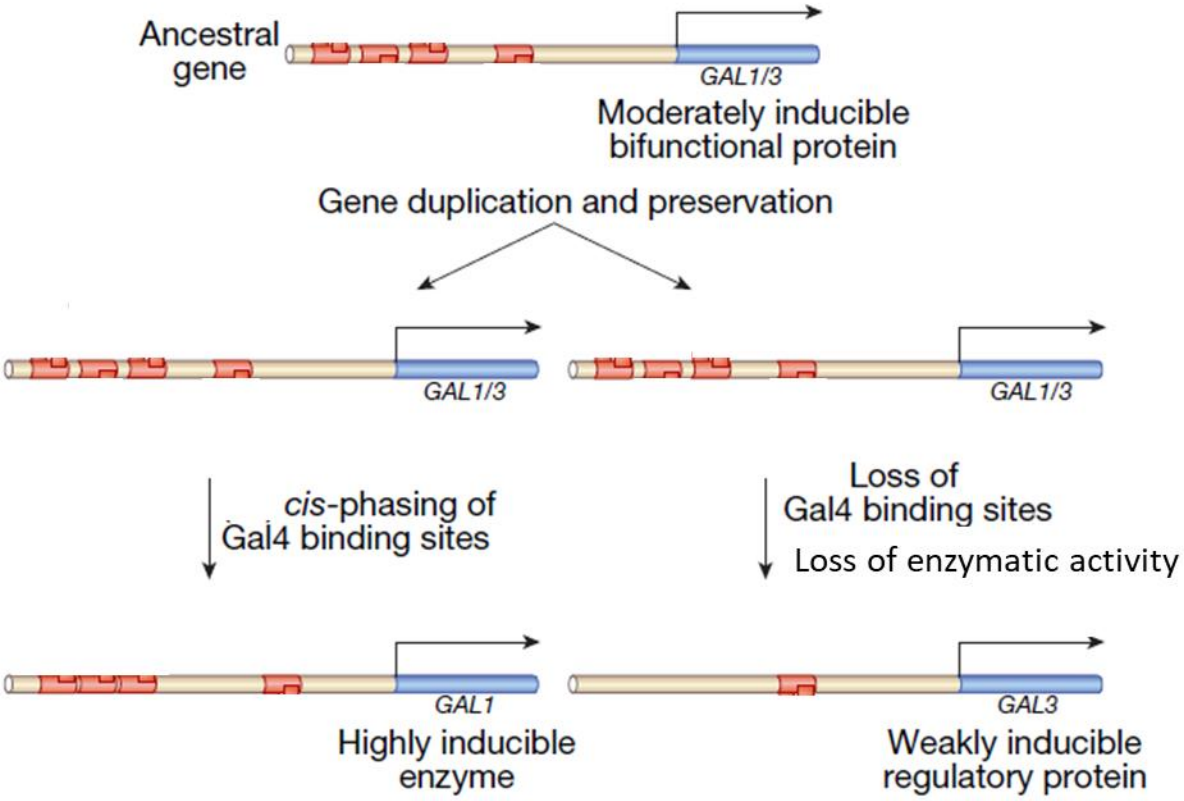


Figure 6.5. Duplication and subfunctionalization of components of the internal-membrane bound ring of vacuolar ATP synthase in the ancestral fungal genome. After their origin by gene duplication from a protein denoted by green, subunits Vma3 (blue) and Vma11 (yellow) acquired complementary changes in interface residues, preventing each from binding one side of Vma16 (pink), as demonstrated by experimental manipulation of protein sequences by Finnigan et al. (2011).

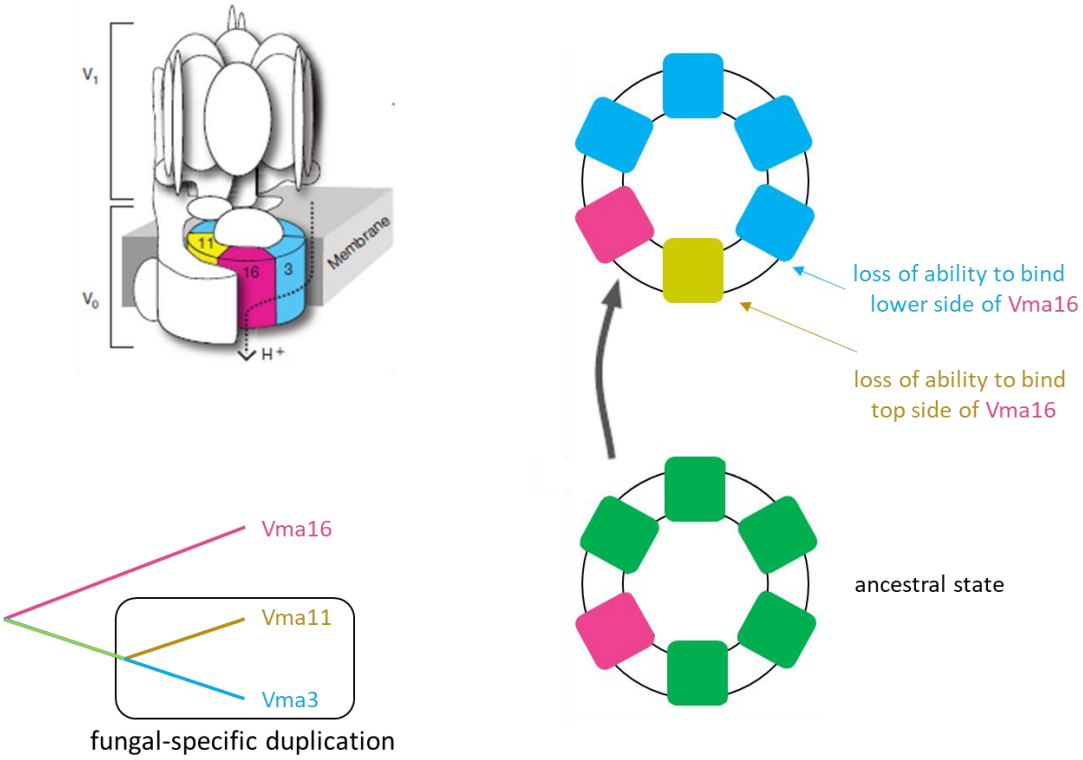


Figure 6.6. A hypothetical scenario in which a gene with two independently mutable subfunctions can arise from an ancestral state with a single generalized expression mechanism. Regulatory regions are depicted on the left, with each regulatory element color-coded according to the transcription factor that binds to it. On the right, the patterns of allele-specific utilization of transcription factors are color coded. Transcription factors denoted by black and white are ubiquitously expressed, whereas those denoted by green and red are expressed in single, non-overlapping conditions. The original gene has a promoter that requires occupancy of three transcription-factor binding sites (one white and two black) for expression. In the first phase of gene evolution, the regulatory region undergoes the sequential accretion of green and red elements, which together are redundant with respect to the white element, which is then lost in a neutral fashion by degenerative mutation. At this point, the evolved allele has a semi-independent mode of expression, as the two black elements are still essential for expression in both tissues. In the second phase, the entire enhancer region is tandemly duplicated, with each component then losing a complementary (red/green) element. The resultant gene now has two independent subfunctions denoted by the green and red open boxes, as a mutation in either region has effects that are confined to a single condition. Note that throughout all of these evolutionary steps, there has been no change in the pattern of expression of the gene; only the mechanism of expression has been altered. From Force et al. (2005).

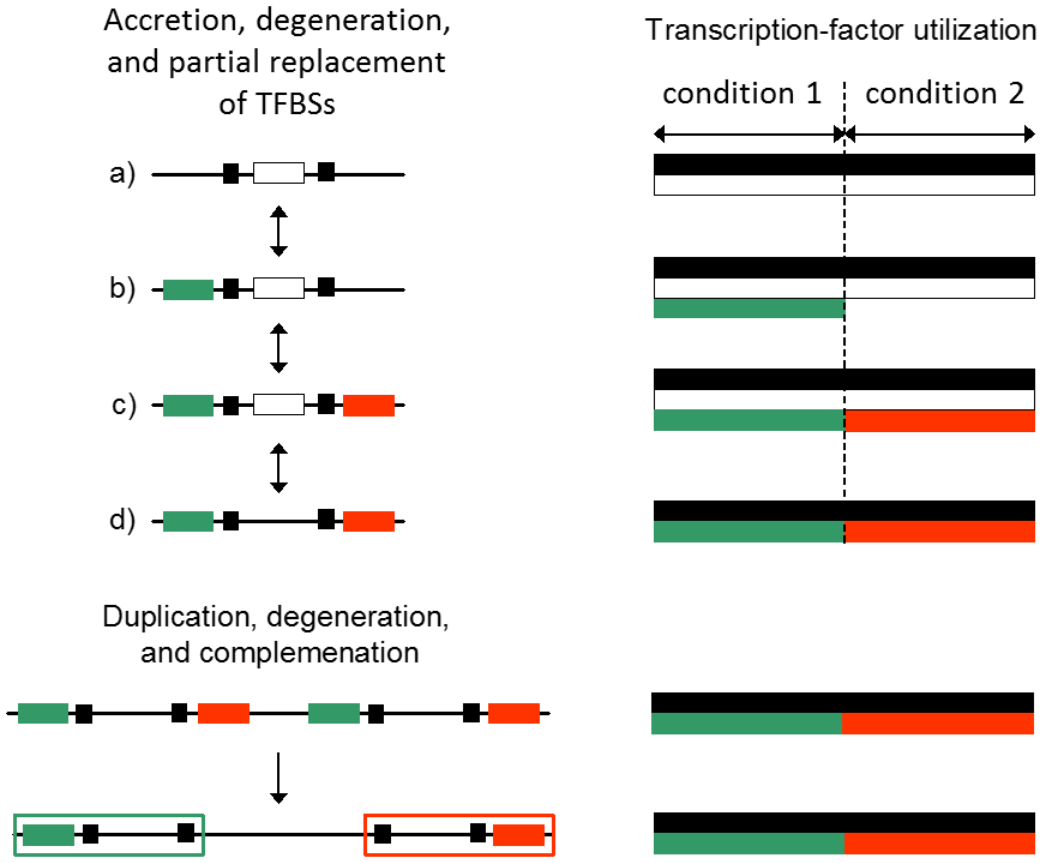


Figure 6.7. Divergent silencing of an ancestral duplicate gene in two geographically isolated lineages. One gene copy is denoted by a white box, the other by a green box, with × denoting an inactivated gene. Gene pairs represent alleles at diploid loci. Progeny genotypes other than those indicated in the bottom row might have compromised fitness, e.g., individuals with just a single active gene.