Chapter 4

Evolutionary Genetics: Progress and Challenges

Jianzhi Zhang

Genetics plays a central role in evolutionary biology, because only heritable traits can evolve. Evolution by natural selection can be succinctly described as a consequence of heritable fitness variation among individuals of a population. Charles Darwin clearly understood that a mechanism of inheritance and a source of heritable variation were necessary components of a complete theory of evolution, but both were unknown to him. Although he thought that much heritable variation arises without reference to need, he also thought that traits acquired during the lifetime of an organism could be passed on to the next generation and used this idea of Larmarckian inheritance to formulate his theory. It is of significant historical interest that the revolutionary experiments that eventually led to the revelation of the true mechanism of inheritance had already been continuing for 3 years when Darwin published *The Origin of Species*. Gregor Mendel performed his famous series of pea cross experiments between 1856 and 1863. In 1865, Mendel read his paper entitled, "Experiments On Plant Hybridization," at two meetings of the Natural History Society of Brno. When Mendel published his work in the *Proceedings of the Brno Natural History Society* (1866), he described particulate inheritance in the form of three basic laws: segregation, independent assortment, and dominance. Unfortunately, Mendel's work was unrecognized by the scientific community for the next 34 years, until it was rediscovered in 1900 by Hugo de Vries, Carl Correns, and Erich von Tschermak. The three Mendelian laws of inheritance were later integrated with the chromosomal theory of inheritance, which was developed by Thomas Morgan and colleagues in the 1910s, and formed the core of classical genetics.

It is interesting to ask why Darwin did not know about Mendel's work. First, there is no evidence that Darwin subscribed to the *Proceedings of the Brno Natural History Society*. Second, there were apparently only 11 published references to Mendel's name before 1900, but at least 3 of these references were accessible to Darwin. One of them was in the *Royal Society's*

Catalogue of Scientific Papers (1864–1873), published in 1879, 3 years before Darwin's death. As a member of the Royal Society, Darwin likely had access to this catalogue, but the catalogue gave no indication of the content of Mendel's paper. The second reference was in W. O. Focke's *Die Pflanzen-Mischlinge* (1881), of which Darwin had a copy. Focke apparently did not understand the importance of Mendel's discoveries and placed Mendel among many other plant hybridizers. The pages in which reference was briefly made to Mendel's experiments remain uncut in Darwin's copy of Focke's book. The third reference was in H. Hoffmann's *Untersuchungen zur Bestimmung des Werthes von Species und Varietät* (1869). Like Focke, Hoffmann did not recognize anything exceptional in Mendel's results. Although Darwin annotated this book and cited it in *The Effects of Cross and Self Fertilization* in 1876, he neither referred to Mendel in that work nor annotated the references to Mendel. Finally, given that Mendel's seminal work was completely ignored by the entire scientific community for so long, it is quite likely that Darwin would not have recognized its importance even if he had read it. Interested readers may find this information and more in Kritsky (1973) and Sclater (2006).

Did Mendel know about Darwin's theory and the importance of his own work to that theory? If so, why did he not contact Darwin? There is clear evidence that Mendel read *The Origin of Species* and knew about Darwin's theory. However, Mendel appeared to believe evolution by hybridization rather than evolution by natural selection (Blower 1989). Of course, he would not think that his findings provided a basis for an evolutionary theory that he regarded as incorrect.

In the first half of the twentieth century, Darwinian evolution by natural selection was synthesized with Mendelian genetics by the joint efforts of a wide array of talented evolutionary biologists, such as R. A. Fisher, J. B. S. Haldane, Sewall Wright, Theodosius Dobzhansky, Ernst Mayr, George Simpson, and G. Ledyard Stebbins (see Futuyma, Chapter 1). By the end of the 1950s, the result of this modern synthesis (often called neo-Darwinism) was widely accepted among evolutionary biologists. The basic tenets of neo-Darwinism are that:

1. evolution occurs gradually through mutation, selection, and drift, and it is explained by population genetics theories;
2. discontinuities between species are explained as originating gradually through geographical separation, divergence, and extinction, rather than saltation;
3. natural selection is the primary force driving evolutionary change;
4. genetic variation within populations is abundant and is a key contributor to evolution;
5. microevolution can be extrapolated to explain macroevolution.

In the past half century, genetics has seen extraordinary development that has led to three revolutions, two technological and one conceptual, in evolutionary genetics. Although the basic tenets of neo-Darwinism have not been significantly altered (with one exception discussed later), the understanding of evolution has been drastically widened and deepened. In this article, I review these three revolutions, summarize eight rules of evolutionary genetics that emerged from these revolutions, and discuss five major questions that I believe will be largely answered in the next few decades. I note that evolutionary genetics is an enormous field and my review describes only those subjects that I am familiar with and believe to be important. It is thus, my personal view of the progress and challenges of evolutionary genetics.

Three Revolutions in the Last 50 Years

The Molecular Revolution

The molecular basis of inheritance began to be unraveled in 1944 when Avery and colleagues first demonstrated that DNA is the material of which genes are made. In 1953, Watson and Crick revealed the structure of DNA. In 1958, Crick proposed the central dogma that describes the fundamental information flow from DNA to RNA to protein. Nirenberg and others cracked the genetic code in the early 1960s. By then, the fundamental principles of molecular biology had been established.

The great influence of molecular biology on evolutionary biology started in the second half of the 1960s on three fronts almost simultaneously. First, gel electrophoresis was introduced to detect protein polymorphisms within and between populations (Harris 1966; Hubby and Lewontin 1966). The technique was soon adopted by many population and evolutionary geneticists to survey protein polymorphisms in a great many species, and surprisingly large amounts of molecular polymorphisms were discovered. This empirical research, coupled with theoretical developments, led to the formation of a new field, known as molecular population genetics (Lewontin 1974; Nei 1975). Molecular population genetics gradually moved into the DNA era, as Sanger DNA sequencing and polymerase chain reaction became routine in the late 1980s and early 1990s. Molecular population genetics prompted the study of microevolutionary processes at the molecular genetic level.

Second, amino acid sequences of cytochrome c proteins were used to infer the evolutionary relationships of various available species and the deduced phylogeny was found to be largely consistent with traditional morphology-based phylogenies (Fitch and Margoliash 1967). This success offered the promise that evolutionary relationships among organisms could be inferred from the genetic material itself, which most faithfully records their history. In the subsequent decades, molecular phylogenetic

methodologies were developed and thousands of molecular phylogenies were constructed (Hillis et al. 1996; Nei and Kumar 2000; Felsenstein 2004). Today, systematists are mainly using DNA sequences to assemble the Tree of Life that may ultimately include all species on Earth (Cracraft and Donoghue 2004; see Hillis, Chapter 16).

Third, the molecular revolution directly led to the development of the neutral theory of molecular evolution (Kimura 1983), first proposed by Motoo Kimura in 1968 and by King and Jukes in 1969. I will discuss this theory in more detail in a later section.

The molecular revolution also provided tools that allow the identification of the molecular genetic basis of phenotypic evolution, such as the evolution of the body plan, skin color, and visual and olfactory sensitivities, although this fourth influence was unnoticed by most evolutionists until the early 1990s. A good testimony of this influence is the advent of evolutionary developmental (evo-devo) biology, which focuses on the molecular basis of developmental differences within and between species (Raff 1996; Carroll et al. 2001) and is now one of the most rapidly growing fields in evolutionary biology (see Wray, Chapter 9).

The Genomic Revolution

The first complete genome sequence of any free-living organism was determined in 1995 from the pathogenic bacterium *Haemophilus influenzae* (Fleischmann et al. 1995). As of March 2010, 1124 complete genome sequences have been published and at least 5000 genomes are in the process of being sequenced (www.genomesonline.org). With the development of the so-called next-generation sequencing technologies (Mardis 2008; Shendure and Ji 2008), many more genomes are expected to be sequenced (Genome 10K Community of Scientists 2009). In addition to genome sequencing, genomics has also provided an unprecedented amount of functional data at various levels of biological organizations (Gibson and Muse 2004; Pevsner 2009). These include genome-scale data on the concentrations of messenger RNAs, concentrations of proteins, protein half-lives, the stochastic noise of gene expression, genetic interactions, protein–protein interactions, stable protein complexes, protein subcellular localizations, phenotypes and fitness effects of gene deletions, chromatin structures, and various epigenetic modifications.

Genomic data provide substantively more genetic markers for molecular phylogenetics. While multigene phylogenies were still uncommon in systematic studies in the mid-1990s, there are virtually no single-gene phylogenies in publications today unless the purpose is to examine the evolution of the gene itself. Use of the entire genome or all genes in a genome for molecular phylogenetics is no longer rare, and the phylogenomic era has undoubtedly arrived (Delsuc et al. 2005). Genomic data also offer new types of genetic markers that may outperform the single-nucleotide substitutions that have been the primary source of phylogenetic signals in molecular

systematics (Rokas and Holland 2000). These new markers include insertions/deletions, transpositions, gene duplications, and other rare genomic events. These events can be sufficiently unique that one is unlikely to encounter homoplasy.

Genomic data allow the identification of the relative importance of alternative evolutionary mechanisms that are known from individual case studies. For example, these data have been used to address whether most stably retained duplicate genes have undergone neofunctionalization, subfunctionalization, or both (He and Zhang 2005). Another example is the relative frequency of intron gains and losses in evolution (Roy and Gilbert 2006). Such questions cannot be tackled without genome-wide data.

Genome data also revealed new mechanisms of evolution and drastically changed the view of the prevalence of certain mechanisms, such as the high incidence of duplication and deletion of genomic segments, commonly resulting in intraspecific variation in copy number (Zhang 2007). Contrary to what was once believed, genomic data also showed that transposons often contribute to the origin of new genes (Nekrutenko and Li 2001).

An Intellectual Revolution: The Neutral Theory of Molecular Evolution

One of the main tenets of the evolutionary synthesis was that natural selection is the primary force of evolution. This assertion was seriously challenged by the neutral theory of molecular evolution, which, in my view, is the only conceptual revolution in the last half-century in evolutionary genetics (and possibly all evolutionary biology). The neutral theory (Kimura 1983) claims that (1) most nucleotide differences between species result from random fixations of neutral mutations, and (2) most intraspecific polymorphisms are also neutral. Based on the evolutionary analysis of three proteins (hemoglobin, cytochrome c, and triosephosphate isomerase), Kimura reported in his classic 1968 *Nature* paper that the rate of nucleotide substitution per generation per mammalian genome is approximately two, which is 600 times what Haldane (1957) considered to be the upper-limit for the rate of adaptive evolution imposed by the cost of natural selection (Kimura 1968). To Kimura, the only solution to this contradiction is that most nucleotide substitutions are not adaptive but neutral.

Fifteen months after Kimura's paper appeared, King and Jukes (1969) published a *Science* article entitled "Non-Darwinian evolution." In this provocative paper, the authors used the available knowledge of molecular biology to argue for the prevalence of neutral genetic changes in evolution. For example, they inferred that nucleotide substitutions are faster at third codon positions than at first and second codon positions. Because a much larger fraction of nucleotide changes at third codon positions than at first and second positions are silent, their result strongly supported the neutral theory. The neutral theory was heatedly debated for most of the following two decades. If it is true, as is widely believed, that most nucleotides in

the genomes of complex eukaryotes do not code for useful product and are apparently non-functional, then many or most nucleotide substitutions and polymorphisms in genomes must be unambiguously neutral. Most of the current uncertainty about whether or not the neutral theory is correct concerns the functional part of the genome, including protein-coding, RNA-coding, and regulatory sequences.

It turned out that Kimura made a mistake in his calculation; he assumed that 100% of the mammalian genome codes for proteins, while the best contemporary estimate is ~1.5%. If he had known this, he would have calculated a per genome amino acid substitution rate only nine times Haldane's upper limit, but still consistent with the view that most amino acids substitutions must be neutral. Given the uncertainties associated with the assumed upper limit and the extrapolation of the substitution rate per genome from only three proteins, this difference probably would not have been considered to be a big surprise. This estimate can be further refined with new genome data. A comparison of the mouse and rat orthologous proteins shows that the median sequence identity is 95% (Gibbs et al. 2004). Assuming the two species diverged 18 million years ago (Gibbs et al. 2004), the mean generation time is 1 year, the mean protein size is 450 amino acids, and there are 20,000 protein-coding genes in the mammalian genome, I calculated that the amino acid substitution rate per generation, per genome is 0.0125—about four times Haldane's upper limit. Thus, the majority of amino acid substitutions are still predicted to be neutral, and the observed rate of amino acid substitution would not be incompatible with Haldane's estimate, if a quarter of amino acid substitutions are adaptive.

While the exact fraction of adaptive nucleotide changes in the functional part of the genome is still being debated, there is no doubt that many genetic differences among organisms in the functional part of the genome are neutral or nearly neutral. Today, the neutral theory usually serves as a null hypothesis and adaptation is proposed only when the expectations of neutrality are rejected by significant evidence. This practice strongly contrasts with the situation 50 years ago, when adaptation was the default explanation of almost all biological observations. Although Kimura (1983) expressly limited his neutral theory to molecular sequences and not to phenotypic traits, some authors (Gould and Lewontin 1979) have cited neutrality as a possible alternative to adaptive interpretation of some phenotypic change as well.

Emerging Rules in Evolutionary Genetics

Except for the hypothesis that natural selection is the primary force of evolutionary change, all basic tenets of neo-Darwinism are still intact. Within a framework expanded to include the neutral theory, substantial progress has been made in evolutionary genetics in the last half century. In particular, the molecular and genomic revolutions have provided such an unprecedented amount of data that there is now a much more solid and detailed understanding of evolutionary processes than before (see Kolaczkowski and Kern, Chapter 6). Here, I describe eight emerging rules revealed from the last 50 years of evolutionary genetic studies. Although substantial progress has been made in evolutionary developmental genetics and speciation genetics, I will not discuss these topics, because they are covered in other chapters of the book (see Wray, Chapter 9; Harrison, Chapter 13).

Rule 1: Life Is Fundamentally Conserved

One of the most important observations from molecular genetic and genomic studies is that life is fundamentally conserved at many levels. First, all cellular organisms use DNA as their genetic material. Second, the genetic code is largely conserved across all species, although rare variants do exist (Osawa 1995). Third, the most basic molecular cellular processes, such as DNA replication, transcription, and translation, are largely the same across all species. In addition, a large part of central metabolism is conserved in most organisms, the same signaling pathways exist in many divergent species, and many genes are shared across the three domains of life (see Lane, Commentary 4). Given that the last common ancestor of all extant species lived more than 3 billion years ago (see Lazcano, Chapter 14), the observed level of conservation is astonishing. This extreme conservation suggests that there may be only one or a very small number of ways to construct life and/or that historical contingency plays such a dominant role in evolution that descendants cannot deviate too far from the common ancestor, even with billions of years of modifications.

The high conservation leads to the prediction that purifying selection is the dominant form of natural selection. This prediction is strongly supported by comparative genomic data, which revealed significantly lower rates of nucleotide substitutions that alter the encoded amino acids (nonsynonymous changes) compared to the rates of nucleotide substitutions that do not alter the encoded amino acids (synonymous changes) for the vast majority of genes in a genome (Waterston et al. 2002).

The high conservation across all life forms, especially at molecular and cellular levels, has several significant implications. First, despite the huge diversity of life at the organismal and phenotypic levels, there are universal rules of evolution at the genetic and genomic levels; furthermore, these rules may be discovered by studying a relatively small number of species. Second, the overall conservation in evolution makes phenotypic variations among species particularly interesting, and evolutionary geneticists have been studying both the proximate causes (i.e., molecular genetic and developmental basis) and ultimate causes (i.e., selection or drift) of such variations in the last few decades. Third, many aspects of human biology may be studied using model organisms, such as the mouse *Mus musculus*, fruit fly *Drosophila melanogaster*, nematode *Caenorhabditis elegans*, flowering plant

Arabidopsis thaliana, budding yeast *Saccharomyces cerevisiae*, and bacterium *Escherichia coli*. Thus, funding agencies such as the U.S. National Institutes of Health, whose mission is to improve people's health and save lives, are willing to invest in the study of model organisms.

Rule 2: Chance Plays an Important Role in Evolution

Stochasticity plays roles at many levels of biological organization. At the level of individual organisms, survival, reproduction, and death often have large random (nonselective) components. Molecular cellular processes, such as transcription initiation, protein degradation, protein–protein interaction, and metabolism, depend on biochemical reactions that occur upon random encounters of sometimes small numbers of molecules. For example, there is a high level of intrinsic noise in gene expression revealed by variation in protein expression among isogenic cells under the same condition. This noise arises from stochastic events in processes such as transcription initiation, mRNA degradation, translation initiation, and protein degradation (Raser and O'Shea 2005; Raj and van Oudenaarden 2008). Mutation and recombination, the ultimate sources of genetic variation, are also random events. Finally, random segregation and independent assortment, two of the three Mendelian laws of inheritance, involve stochasticity.

The most fundamental impact of chance on evolution is through random genetic drift, the random sampling of alleles during the reproduction of a finite population. Genetic drift leads to random loss and fixation of alleles that are independent of the relative fitness of those alleles. The neutral theory of molecular evolution (Kimura 1983) asserts that genetic drift accounts for the majority of nucleotide substitutions in evolution. Ample evidence from molecular population genetics and molecular evolution studies supports this view. Because the neutral theory does not deny the occurrence of rare positive Darwinian selection for advantageous alleles in evolution, the occasional identification of positive selection at the molecular genetic level does not reject the theory. In the last decade, a number of authors have tried to quantify the fraction of amino acid substitutions that are adaptive, using large amounts of population genetic data. The results, however, are ambiguous and hard to interpret, in my view. Many, if not all, studies of *Drosophila* indicated that a large fraction (from 30% to 95%) of amino acid substitutions are adaptive (Fay et al. 2002; Smith and Eyre-Walker 2002; Eyre-Walker 2006; Sawyer et al. 2007; Shapiro et al. 2007; Sella et al. 2009). The fraction, however, is found to be very small in humans (Zhang and Li 2005; Eyre-Walker 2006) and effectively zero in the budding yeast (Doniger et al. 2008; Liti et al. 2009). Because genetic drift has a greater impact on evolution in small populations than in large populations (Ohta 1992), one would predict a higher contribution of positive selection in species with larger populations. However, the population size of *Drosophila* is somewhere between that of humans and yeast. So, the results thus far do not make sense. It is unclear whether this inconsistency is due to any

peculiarity of the three species, such that they do not respectively represent average species of comparable sizes.

Relative to the strength of selection, genetic drift is expected to play a more important role in small populations associated with large and complex organisms (Ohta 1992), which has led to a nearly neutral explanation of the origin of genome architecture (Lynch 2007). This explanation can be viewed as an extension of the neutral theory (Kimura 1983) and the nearly neutral theory (Ohta 1992) from molecular evolution to genomic evolution. Lynch (2007) argues that complex genomic features, such as the existence of mobile elements, gene families, split genes, and alternative splicing, may have passively originated through non-adaptive processes in small populations. When they first appeared, the features might have been very slightly deleterious, but persisted in organisms with small populations because the purifying selection against them was sufficiently weak. Through long-term evolution, these features became established and may have been modified to perform useful functions, such as the role of alternative splicing in generating multiple proteins of different functions from one gene. This novel hypothesis is certain to stimulate debate and empirical study.

Rule 3: Genomes of Complex Organisms Contain a Lot of Junk DNA

It must have been a big surprise to those who believed that organisms are perfectly or nearly perfectly adapted to their environments to learn that the genomes of many complex organisms, such as vertebrates and flowering plants, contain a large fraction of so-called junk DNA that apparently has no function. For example, only approximately 1.5% of the human genome codes proteins and only about 5% appears to be constrained to various degrees (Waterston et al. 2002). This junk DNA is largely composed of repetitive sequences that originated from transposable elements. It is likely that junk DNA has a very small fitness cost to the host and therefore, is not effectively removed by natural selection, especially in more complex organisms that often have small populations (Lynch 2007).

Is there treasure in the junk DNA? The U. S. National Human Genome Research Institute funded a large project, named The Encyclopedia of DNA Elements (ENCODE), to build a comprehensive parts list of the functional elements of the human genome. The initial analysis of 1% of the human genome provided some surprises (The ENCODE Project Consortium 2007). For example, it was found that the majority of the bases in the human genome are transcribed, although it remains unclear whether the transcription reflects biological function or simply leaky expression. It was also found that many functional elements defined by DNA-protein binding are still seemingly unconstrained across mammalian evolution, suggesting that those regions may not have physiological functions or their loss imposes no fitness reduction. Alternatively, the finding could indicate the existence of a large number of species-specific functional elements in a genome. It is interesting to note that some purportedly functionless elements of the

genome may be recruited into protein sequences. For example, it was discovered that about 4% of human proteins contain sequences that originated from transposable elements, which initially resided in introns but were later recruited to become new exons (Nekrutenko and Li 2001). There is also evidence of rare *de novo* gene origination from non-coding sequences (Levine et al. 2006; Chen et al. 2007).

Rule 4: Gene Number Does Not Predict Organismal Complexity

Analogous to the C-value paradox (Gregory 2001), which describes the lack of relationship between organismal complexity and genome size, there is no simple relation between organismal complexity and gene number. There are surprisingly few genes in the human genome (~20,000), compared to *E. coli* (4400), yeast (6000), fruit fly (13,600), nematode (19,000), sea urchin (23,500), *Arabidopsis* (27,500), and rice (41,000). While the C-value paradox has been attributed to the variation in the amount of junk DNA in different genomes, the cause of the lack of correlation between organismal complexity and gene number is yet to be determined. This being said, I emphasize the difficulty in measuring organismal complexity. If we use the number of recognizably different types of cells in an organism as a proxy for organismal complexity, vertebrates are more complex than triploblastic invertebrates, which in turn are more complex than vascular plants (Futuyma 1998). But, gene number is higher in vascular plants than in vertebrates.

Several potential mechanisms could compensate for the low gene number in highly complex organisms, such as vertebrates. First, alternative RNA splicing, prevalent in multicellular organisms, substantially increases the number of different proteins in an organism. It is estimated that over 80% of human genes are alternatively spliced (Matlin et al. 2005). However, a comparison among human, mouse, rat, cow, *D. melanogaster*, *C. elegans*, and *A. thaliana* found comparable frequencies of alternatively spliced genes across species (Brett et al. 2002). It was subsequently suggested that these results were an artifact of the different coverage of expressed sequence tags for the various organisms; when this confounding factor was removed, vertebrates show higher frequencies of alternative splicing than invertebrates (Kim et al. 2007). Nevertheless, because not all spliced forms are functional or functionally distinct, the above finding does not unambiguously demonstrate that vertebrates have substantially more functionally distinct proteins than invertebrates. The key issue in the study of alternative splicing is to estimate the proportion of alternatively spliced forms that are functionally distinct and physiologically useful.

Second, because the potential number of interactions between genes or proteins is much greater than the actual number of genes or proteins, one could hypothesize that organismal complexity depends more on the number of molecular interactions than the number of genes or proteins. Genome-wide protein–protein interactions have been surveyed in a number of model organisms, including yeast, fruit flies, and humans (Beyer et

al. 2007). Because of the incompleteness of the data and different biases in different datasets, it is hard to tell at this stage whether the number of protein interactions in a species correlates with the organismal complexity. Genetic interactions, mainly in the form of synthetic lethality or illness, have been examined by simultaneous knock-out or knock-down of pairs of genes in a few model organisms, such as budding yeast, fission yeast, and nematodes (Tong et al. 2004; Boone et al. 2007; Roguev et al. 2008; Tischler et al. 2008). The current data are still far from complete for across-species comparisons.

Finally, gene expression regulation and post-translational modification can potentially increase organismal complexity (see Wray, Chapter 9). However, there is still no good empirical data to test this hypothesis rigorously. For example, no evidence supports the proposal that gene regulation is more complex in vertebrates than in invertebrates and plants. Thus, the molecular genetic basis of organismal complexity remains largely unexplained.

Rule 5: Horizontal Gene Transfers Are Prevalent (at Least in Prokaryotes)

The Origin of Species contained only one figure, which depicted a hypothetical phylogenetic tree of 15 extant taxa. In this tree, genetic information was transmitted vertically from parents to offspring and there was no horizontal transmission of genetic information among different evolutionary lineages. We now know that horizontal gene transfers (HGTs) occur frequently among prokaryotes (Koonin et al. 2001; Gogarten et al. 2009; see Lane, Commentary 4). There is also ample evidence that they occur among eukaryotes and between prokaryotes and eukaryotes (Keeling and Palmer 2008; Gogarten et al. 2009), although the rate of occurrence appears much lower in eukaryotes. HGTs occur through three main mechanisms: transformation, conjugation, and transduction. Transformation refers to the phenomenon that cells from certain species can take up free DNA from their environments. Conjugation is the process by which a living cell transfers genetic material to another cell through the formation of a tube-like structure (i.e., pilus) between cells. Transduction refers to DNA movement from one cell to another by a virus.

HGT was first reported 50 years ago when antibiotic-resistant genes were found to be transferred across bacterial species (Ochiai et al. 1959; Akiba et al. 1960). However, it was not until the late 1990s, when multiple prokaryotic genomes were sequenced and compared, that the prevalence of HGTs in evolution became appreciated. For example, it was reported in one study that 24% of the protein-coding genes in *Thermotoga maritima*, a thermophilic eubacterium, are most similar to archaeal genes (Nelson et al. 1999). Because criteria for identifying probable horizontal gene transfer rely on unusual feature(s) of subsets of genes that distinguish them from the bulk of genes in the genome (Koonin et al. 2001), indications of HGTs

remain probabilistic and thus can sometimes be controversial. The current debate centers on the quantitative assessment of the pervasiveness and rate of HGTs (Doolittle 1999; Gogarten et al. 2002; Daubin et al. 2003). Some researchers believe that HGTs are so pervasive and frequent that the Tree of Life (at least in the prokaryotic part) becomes a network of life from which it is neither meaningful nor feasible to reconstruct species phylogenies (Doolittle 1999; Gogarten et al. 2002). More fundamentally, if genes were freely transferred across species, the species concept would collapse. Other researchers believe that a sizeable fraction of genes in the genome are incapable of HGTs, and these genes would allow the reconstruction of a species phylogeny (Daubin et al. 2003; Ciccarelli et al. 2006). For instance, it was proposed in the so-called complexity hypothesis that informational genes, which function in transcription, translation, and related processes, are horizontally transferred with a much lower rate than housekeeping operational genes, because the translational and transcriptional apparatuses are large and complex systems. In this case, a foreign gene is unlikely to be compatible in a system made of native parts (Jain et al. 1999). This hypothesis has received empirical support. For example, in an analysis of 191 species with complete genome sequences, 31 genes that are relatively immune to HGTs were found and all of them are involved in translation (Ciccarelli et al. 2006). A recent study analyzed attempted experimental movement of 246,045 genes from 79 prokaryotic genomes into *E. coli* and identified genes that consistently fail to transfer (Sorek et al. 2007). Interestingly, ribosomal proteins dominate the list of untransferable genes, and toxicity to the host is the primary cause of transfer inhibition (Sorek et al. 2007). Although different genes have different rates of HGT, the question remains whether there is a sufficiently large set of HGT-resistant genes such that a species phylogeny of prokaryotes is both meaningful and reconstructable.

Rule 6: Gene Duplication Is the Primary Source of New Genes

Although many genes involved in the most fundamental molecular cellular processes, such as protein synthesis and DNA replication, are shared among all species (Mushegian and Koonin 1996), there are probably no two species that have exactly the same set of genes. Variation in gene content is a major source of biodiversity. How new genes with novel functions originate has been a fascinating subject to many researchers, and several molecular mechanisms have been proposed. First, exon shuffling combines existing exons between different genes and generates hybrid genes with multiple exons (Gilbert 1978, 1987; Patthy 1995, 1999). The resulting protein thus exhibits additional functions conferred by the newly acquired exons, and the interactions between the amino acids encoded by different exons may also lead to entirely new protein functions. The prevalence of multi-domain proteins in high eukaryotes suggests the important contribution of exon shuffling (Patthy 1999). Second, introns (and other noncoding sequences) may, under certain circumstances, be converted to protein-coding

sequences (Nekrutenko and Li 2001). Similarly, alternative reading frames or antisense strands of functional genes may sometimes be used as the genetic material for a new gene (Yomo et al. 1992; Golding et al. 1994). However, such events are rare, because of the low probability of the occurrence of long open reading frames from random DNA sequences. The third mechanism is gene sharing. Best known in lens crystallin genes, gene sharing allows one gene to adopt an entirely different function without losing its primary function (Piatigorsky 2007). For instance, in birds and crocodiles, lactate dehydrogenase appears as an enzyme as well as a structural protein in the lens. In theory, a new gene may also arise through a *de novo* process. Although uncommon, several such examples have been reported (Chen et al. 1997; Levine et al. 2006; Chen et al. 2007). Horizontal gene transfer brings new genes from other species to a species, but this process does not generate novel genes. Except for exon shuffling, the other mechanisms seem to have minimal contributions to the origin of new genes with novel functions. However, even exon shuffling cannot account for the high rate of gene origination in evolution. In fact, most new genes were generated through gene duplication.

In 1936, Bridges reported one of the earliest observations of gene duplication in the doubling of a chromosomal band in a *D. melanogaster* mutant that exhibited extreme reduction in eye size (Bridges 1936). Evolutionary biologists quickly realized the potential of gene duplication as a mechanism of evolution of new genes (Stephens 1951). Ohno's seminal book *Evolution by Gene Duplication* (Ohno 1970) further popularized this idea among biologists. However, it was not until the late 1990s, when numerous genomes were sequenced and analyzed, that the widespread prevalence of gene duplication became clear. Virtually every genome sequenced thus far contains a high fraction of duplicate genes, and because ancient duplicate genes are difficult to recognize through sequence comparison, the true percentage in a genome is likely much higher (Zhang 2003). Gene duplication may occur through unequal crossover, retroposition, chromosomal nondisjunction, or polyploidization. These mechanisms are responsible for generating segmental duplication, retroduplication, chromosomal duplication, and genome duplication, respectively (Zhang 2003).

Retroposition was initially thought to create only pseudogenes, because in retroposition the message RNA of a gene is reverse-transcribed into complementary DNA, which is then inserted into the genome randomly. As such, the promoter of the gene is not duplicated along with the coding region of the gene; consequently, the retroduplicate is not expressed. However, recent studies showed that a small fraction of retroduplicates are by chance inserted into introns of existing genes or to genomic regions containing promoters. In such cases, a retroduplicate may become part of a functional gene or a new gene (Long et al. 2003).

A duplicated gene may experience several potential fates even when it is functional and is fixed in a population. The most common fate is

pseudogenization, which occurs when the duplicate copy is functionally redundant and therefore is not subject to any selective constraint. Several mechanisms may allow the long-term retention of a duplicate gene. First, increased dosage of certain genes (e.g., ribosomal RNA genes and histone genes) can be beneficial and lead to the retention of duplicate genes even without any change in function or expression (Zhang 2003). Second, the ancestral gene may have multiple functions that are subdivided in the daughter genes, and each of them fixes mutations (perhaps, by genetic drift) that disable some of the ancestral functions (Force et al. 1999; Lynch and Force 2000b). The joint levels of expression and patterns of activity of the two daughter genes are equivalent to those of the single ancestral gene. Consequently, both daughter genes may be stably retained. Third, a duplicate gene may neofunctionalize by acquiring a new function or a new expression pattern, such that the fitness of the organism is enhanced (Ohno 1970). Fourth, it is also possible that a duplicate gene pair experiences subfunctionalization quickly after the duplication, which permits its long-term retention and allows gradual acquisitions of new functions in evolution (He and Zhang 2005). There is no consensus on the relative contributions of these mechanisms that underlie stable retention and evolution of duplicate genes, in spite of the fact that the subject has been studied extensively. The relative roles of natural selection and genetic drift in the fixation and retention of duplicate genes is also contentious (Zhang et al. 1998). While drift is likely important in subfunctionalization, positive selection must be involved in the dosage benefit mechanism and is probable in neofunctionalization. While evolutionary geneticists are interested in the mechanism of duplicate gene retention, molecular biologists tend to focus on the functional similarities and differences among duplicate genes. These two issues are intimately related, as is clear from the above explanation of the different functional alterations invoked in the evolutionary mechanisms.

The most important contribution of gene duplication to evolution is the provision of new genetic material, upon which mutation, drift, and selection act to create either specialized or new gene function. Some of the most exquisite biological responses, such as the adaptive immune system and the olfactory and taste chemosensory systems in vertebrates, rely extensively on duplicate genes that perform similar but distinct functions (Nei et al. 1997; Shi and Zhang 2009). Recent genomic analysis in mutation-accumulation lines of yeasts showed that the spontaneous mutation rate of gene duplication is high (Lynch et al. 2008). Comparative genomics also reveals a high rate of fixed duplications (Lynch 2007). Thus, gene duplication must have contributed greatly to the genetic and phenotypic differences between different evolutionary lineages. Gene and genome duplication may have also directly contributed to speciation through the divergent resolution process, in which the random loss of a redundant gene copy in two populations could result in the missing of both copies in the gametes of their hybrids and thus reproductive isolation by hybrid sterility (Werth and Windham 1991; Lynch and Force 2000a).

Rule 7: Changes in Protein Function and Gene Expression Are Both Important in Phenotypic Evolution

For historical reasons, molecular evolutionary studies have focused more on evolutionary changes in protein sequence and function than changes in gene expression and its regulation. In 1975, King and Wilson reported that humans and chimpanzees have virtually identical protein sequences despite their large phenotypic differences. This observation prompted the authors to propose that changes in gene expression play a more important role in phenotypic evolution than changes in protein function (King and Wilson 1975). This hypothesis has been enormously influential to evolutionary biologists; many were convinced that gene expression changes are more important and have looked for both theoretical and empirical evidence for this hypothesis. We now know that between human and chimpanzee, there are on average about two amino acid differences per protein and more than 70% of proteins are non-identical (Chimpanzee Sequencing and Analysis Consortium 2005; Glazko et al. 2005). So, protein sequence differences between human and chimpanzee are numerous, which can potentially account for many of the phenotypic differences between the two species. Nonetheless, the role of gene expression changes in phenotypic evolution has been documented in many case studies (Wray 2007; Carroll 2008; Stern and Orgogozo 2009; see Wray, Chapter 9).

To answer the question of whether gene expression change is generally more important than protein function change, two research groups recently compiled cases of phenotypic evolution with known genetic mechanisms (Hoekstra and Coyne 2007; Stern and Orgogozo 2008). Although such meta-analyses are valuable in summarizing case studies and providing information about the overall empirical evidence at the present time, caution is needed because case studies are subject to ascertainment biases associated with preferences for certain methods, phenotypes, genes, and types of mutations. The empirical evidence shows that both gene expression change and protein function change are important genetic mechanisms of evolution. It is probably more productive to study whether these two types of genetic mechanisms are disproportionately used for different types of phenotypic evolution than to argue which mechanism is more important. For example, based on case studies and theoretical considerations that a distinction exists in the genetic basis of morphological and physiological evolution, it has been proposed that morphological evolution occurs mainly through gene expression changes, while physiological evolution occurs mainly through protein function changes (Carroll 2005). Because morphology and physiology are intimately connected, it may be difficult to clearly separate them. Nevertheless, one can imagine cases in which physiological changes do not require accompanying morphological changes and

vice versa. For example, response to low oxygen levels by modification of the hemoglobin sequence does not involve morphological changes, and wing pigmentation differences in some insects probably do not involve physiological changes. Thus, if morphological and physiological traits are distinguishable, Carroll's hypothesis may be tested best by using genomic data rather than case studies. That is, one could study the genes in which mutations only affect morphological traits and genes in which mutations only affect physiological traits. A recent comparison between the two types of genes in their molecular function and evolutionary pattern lends support to Carroll's hypothesis (Liao et al. 2010).

An alternative approach to studying the genetic mechanisms of phenotypic evolution is experimental evolution (Garland and Rose 2009). Genomic technologies, including high-throughput next-generation sequencing, allow cheap, quick, and accurate determinations of the genome sequences and transcriptomes, including sequencing the starting and end strains from laboratory evolutionary experiments and identifying the mutations that are responsible for the phenotypic (e.g., fitness) changes. Interestingly, two studies, one in *Escherichia coli* (Herring et al. 2006) and the other in yeast (Gresham et al. 2008), showed the prevalence of protein function and copy-number changes in physiological adaptation yet virtually no gene expression changes by *cis*-regulatory sequence alteration. One caveat is that both *E. coli* and yeast contain only short *cis*-regulatory sequences and may not represent complex organisms. Another caveat in microbial experimental evolution is that morphological changes are much more difficult to study than physiological changes.

Rule 8: Intraspecific Genetic Polymorphisms Are Abundant and Largely Neutral

The molecular revolution in evolutionary genetics resulted in a large body of literature on the genetic polymorphisms in hundreds of species. Compared to what neo-Darwinists thought, intraspecific polymorphisms at the DNA and protein sequence levels are astonishingly high (Lewontin 1974; Lynch 2007). For example, even in humans, who have a rather small effective population size estimated at approximately 10^4, the mean nucleotide diversity is on the order of 0.1%, meaning that two randomly chosen alleles of the same gene differ by 1 out of 1000 bases. This level of allelic difference is about 10% of the genetic difference between humans and chimpanzees, although probably no layperson would believe that the human–chimp difference is only 10 times that between two humans.

The neutral theory predicts that, at the mutation-drift equilibrium, neutral nucleotide diversity is given as $\pi = 4N\mu$, in which N is the effective population size and μ is the neutral mutation rate per generation. The best support for the general statement that most nucleotide polymorphisms across the genome are neutral is the clear trend for associations of polymorphism with population size or its surrogates (e.g., organisms with small body sizes have larger populations) (Nei and Graur 1984). This view is so universally accepted that measures of genetic diversity are routinely used as indirect estimates of historical population size (Roman and Palumbi 2003). Furthermore, because the neutral mutation rate μ is higher at synonymous sites than at nonsynonymous sites, the neutral theory also predicts that π is higher at synonymous sites than at nonsynonymous sites, which has been repeatedly shown (Cargill et al. 1999; Moriyama and Powell 1996). These two observations certainly indicate that most intraspecific polymorphisms are best explained by the joint forces of mutation and drift acting on neutral mutations. This finding is expected because: (1) neither deleterious nor advantageous mutations contribute much to the level of intraspecific variations and (2) these alleles are either kept at sufficiently low frequencies or become fixed quickly. Nevertheless, while this global statement is true, there are dozens of case studies in which individual polymorphisms, when examined in detail, appear to be adaptive responses to current conditions (Vasemagi and Primmer 2005; Voight et al. 2006; Linnen et al. 2009; Rebeiz et al. 2009). Detection of selection-maintained polymorphisms has been enhanced by the development of many statistical approaches, such as those associated with coalescence theory (see Kolaczkowski and Kern, Chapter 6; Wakeley, Chapter 5). These individual cases aside, balancing selection advocated by the balance school led by Dobzhansky as a major cause of most genetic polymorphism (Lewontin 1974), does not appear to account for most molecular polymorphisms. Certainly, the form of long-term balancing selection that would lead to even trans-specific polymorphism has been documented in only relatively few genes (Hughes and Nei 1988; Clark and Kao 1991; Cho et al. 2006).

Major Unsolved Questions in Evolutionary Genetics

Although evolutionary genetics has seen rapid progress in the last 50 years, a number of major unsolved questions hamper a complete and accurate understanding of evolutionary processes. Below I describe some of these questions that I think can be largely solved or will at least see significant progress in the next few decades.

Question 1: How Can the Genetic Basis of Macroevolutionary Changes Be Found?

At least in principle, it is no longer challenging to identify the nucleotide substitutions that are responsible for the phenotypic differences between individuals of the same species or closely related species, if only one locus or a small number of loci are involved. In addition to the candidate gene approach, which relies on prior knowledge of the potential roles of candidate genes in controlling a trait, positional cloning and association studies are now routinely used in model organisms and humans. Positional cloning starts from mapping the loci responsible for a trait, which is achieved

through linkage analysis in existing pedigrees or designed genetic crosses. The genomic revolution has allowed the identification in model organisms of a wealth of polymorphic genetic markers that can be inexpensively assayed in linkage analysis. Association studies look for genetic markers that are statistically correlated with a phenotypic difference. After the identification of potentially causal mutations, several molecular techniques, such as gene replacement, allow a definitive experimental verification. In the last decade, a number of genetic alterations responsible for microevolutionary phenotypic changes have been identified through the candidate gene approach, positional cloning, and association studies (Johanson et al. 2000; Sucena and Stern 2000; Takahashi et al. 2001; Shapiro et al. 2004; Yoshiura et al. 2006; Zhang 2006; Linnen et al. 2009; Rebeiz et al. 2009; Tung et al. 2009; Wittkopp et al. 2009; Chan et al. 2010).

However, these three approaches are generally difficult to apply to the study of macroevolutionary phenotypic changes, which are often most interesting and amazing, and they include some major subjects of evolutionary developmental biology. Positional cloning is simply not usable because individuals from divergent lineages cannot be crossed. Association studies fail because, statistically, all fixed genetic differences between two species are equally correlated with any fixed phenotypic difference between the species. Only the candidate gene approach may be applied, but its success depends on prior knowledge of gene function and of the extent of conservation of the molecular function and physiological role of a gene during evolution. Another obstacle is that rigorous experimental tests are difficult, even when candidate genes are available, because of large differences in genetic background between divergent species.

A better understanding of developmental pathways and gene–gene interactions in model organisms will help improve knowledge of gene function; much of the success of evolutionary developmental biology is attributable to such critical knowledge. Further, genome-wide systematic comparisons of gene function among divergent model organisms can provide some basic ideas on the conservation of gene function in evolution (Liao and Zhang 2008). These studies will likely offer candidate genes for experimental tests. Development of efficient molecular genetic techniques that allow simultaneous alterations of multiple genes in non-model organisms will also be important.

Question 2: What Is the Molecular Genetic Architecture of Multifactorial Traits?

Most traits are controlled by the developmental expression of many genes (Falconer and Mackay 1996). The molecular genetic basis of phenotypic variation in multifactorial or quantitative traits is usually difficult to discern, although considerable progress in developing the linkage-analysis–based approach to studying quantitative trait variation has been made in model organisms, such as *Drosophila* and yeast (Mackay 2001; Brem et al.

2002; Steinmetz et al. 2002; De Luca et al. 2003; Deutschbauer and Davis 2005; Mackay and Lyman 2005). In addition, genome-wide association studies of many complex human traits (mostly common diseases) have also shown the power of this method in identifying small-effect alleles, but the fact that the identified loci together explain only a small fraction of the heritability of the traits concerned (Manolio et al. 2009) is disappointing. This finding is at least in part explainable by the fact that large-effect mutations that cause diseases are kept at very low frequencies by purifying selection and thus require a sizable sample to be detected statistically. It is not clear whether association studies will be more useful in identifying genes underlying adaptive traits.

In model organisms, such as yeast, nematode, fly, and mouse, the tools are available to experimentally delete many genes individually. It is possible to comprehensively phenotype these gene-deletion strains to gain systematic knowledge about which traits are affected by which genes. For example, a recent yeast study measured 501 morphological traits in each of 4718 nonessential gene deletion haploid strains as well as the wild-type haploid strain by fluorescent imaging (Ohya et al. 2005). Because a gene deletion usually has a larger phenotypic effect than the average natural mutation, the comprehensive phenotyping of deletion strains is likely to reveal gene-trait relationships that are hard to detect in natural populations. Using the gene-trait map obtained from the systematic phenotyping as a guide, one can identify and examine candidate genes in natural populations, which can substantially increase the power of association studies. Thus, even when the purpose is to define the genetic architecture of a multifactorial trait in nature, it is useful to first have a comprehensive gene-trait map from gene deletion strains as a guide.

It is possible that gene-trait relationships in natural populations will be undetectable in gene-deletion strains, such as when a natural mutation results in the gain of a function that is different from a null mutation or deletion. The same mutation may also have different phenotypic effects in different genetic backgrounds as a result of epistasis. Furthermore, the deletion of a gene may not affect any phenotype in an artificial lab environment, but it may impact traits in the natural habitat of an organism. Nevertheless, combining the systematic examination of deletion strains and of natural variants will likely provide a more comprehensive picture of the genes responsible for a multifactorial trait.

Identifying the underlying genes of a multifactorial trait and measuring the size of their phenotypic effect are usually not the final goals of evolutionary genetics. One would also like to know the mechanisms (e.g., signaling pathways) through which certain genes control the trait being studied and explain why the phenotypic effect sizes of these genes are different. It is my contention that in addition to molecular biology, which tells us the properties of the gene products for different alleles, systems biology, which studies interactions among different parts of a system and properties

of the system brought about by these interactions, will shed light on these important yet difficult questions.

Question 3: What Is the Genomic Pattern of Epistasis and How Does this Pattern Affect Evolution?

When Bateson coined the term "epistasis" 101 years ago, he meant that the effect of a gene on a trait may be enhanced or masked by one or more other genes (Bateson and Mendel 1909; Phillips 2008). Fisher and other people extended the concept to mean non-independent (non-additive or non-multiplicative) effects of genes (Fisher 1918; Phillips 2008). The direction, magnitude, and prevalence of epistasis is important for understanding many phenomena in gene function and interaction (Hartman et al. 2001; Boone et al. 2007; Phillips 2008), speciation (Coyne 1992), the evolution of sex and recombination (Kondrashov 1988; Barton and Charlesworth 1998), evolution of ploidy (Kondrashov and Crow 1991), mutation load (Crow and Kimura 1979), genetic buffering (Jasnos and Korona 2007), human disease (Cordell 2002; Moore and Williams 2005), and drug–drug interaction (Yeh et al. 2006). Yet, epistasis is arguably the most important but least well understood phenomenon in genetics. Using Fisher's definition of statistical epistasis, recent functional studies have started generating genome-wide epistasis maps in model organisms (Tong et al. 2004; Boone et al. 2007; Roguev et al. 2008; Tischler et al. 2008). These data will provide evidence for general patterns of epistasis, which will in turn allow the testing of many important evolutionary hypotheses that depend on various assumptions about epistasis.

It should be emphasized that statistical epistasis, or non-multiplicative gene effects, may be different from true functional interactions between genes and should be distinguished (He et al. 2010). A recent study by my group revealed a surprisingly high abundance of positive statistical epistasis between deleterious mutations—in other words, two mutations together are not as bad as expected from their individual deleterious effects (He et al. 2010). This finding suggests the need for reevaluation of evolutionary theories that depend critically on overall negative epistasis, such as the mutational determinimistic hypothesis of the evolution of sexual reproduction (Kondrashov 1988) and the hypothesis of reduction in mutational load by truncation selection against deleterious mutations (Crow and Kimura 1979).

Question 4: What Is the Genomic Pattern of Pleiotropy and How Does Pleiotropy Affect Evolution?

Pleiotropy refers to the common observation that one gene (or mutation) affects multiple traits. Despite its broad implications in genetics (Wright 1968; Tyler et al. 2009), development (Hodgkin 1998; Carroll 2008), senescence (Williams 1957), disease (Albin 1993; Brunner and van Driel 2004), adaptation (Fisher 1930; Wright 1968; Waxman and Peck 1998; Orr 2000), the maintenance of sex (Hill and Otto 2007), and stabilization of cooperation (Foster et al. 2004), the genome-wide patterns of pleiotropy are unknown. Due to its central importance in many areas, the implications of pleiotropy have been extensively modeled (Fisher 1930; Turelli 1985; Wagner 1988; Waxman and Peck 1998), but because these theoretical models have virtually no empirical basis, it is unclear whether they are realistic.

Empirical data on pleiotropy are urgently needed to test some of the most fundamental hypotheses in evolutionary genetics. For example, based on Fisher's (1930) geometric model that assumes that a mutation affects all traits of an organism, Orr (2000) derived the formula for the rate of fitness increase during an adaptive walk to the optimum in an organism with n traits. He found that the adaptation rate decreases with n. In other words, complex organisms have lower adaptation rates than simple organisms—a cost of complexity. But Orr's results depend on the assumptions that (1) a mutation affects all traits of an organism and (2) the total phenotypic effect of a mutation is the same in organisms of different levels of complexity. Using a Quantative Trait Locus (QTL) study of mouse skeletal characters, Wagner et al. (2008) recently reported that neither of these assumptions is valid. They found that half of the QTL affects less than 10% of the traits examined and that the mean per-trait effect is larger for those genes influencing more traits (see Wagner, Chapter 9). These results mean that the cost of complexity, while not absent, is significantly lower than Orr's model would suppose. However, the Wagner et al. data set is relatively small (70 traits and 102 QTL), and each of their QTL may include multiple genes. A reanalysis of their data (Hermisson and McGregor 2008) did not find clear evidence that more pleiotropic genes have larger per-trait effects. This example illustrates the importance of collecting pleiotropy data for genes (not QTLs). As previously mentioned, such data can be generated by comprehensive phenotyping of gene deletion strains of model organisms. I expect that many questions related to pleiotropy will have clearer answers in the near future.

Question 5: What Are the Relative Roles of Positive Selection and Genetic Drift in Evolution?

While this question has been investigated and debated since the late 1960s, there is still no agreement among evolutionary biologists. After seeing the genomic data, some researchers are now convinced that most nucleotide substitutions are neutral (Nei 2005), while others believe that most are adaptive and even think that adaptation should now be used as the null hypothesis in explaining evolutionary observations (Hahn 2008). The answer to this question certainly depends on the definition of neutrality. According to Kimura (1983), an allele with an absolute fitness effect $|s| < 0.5/N$, in which N is the effective population size, is effectively neutral. However, Nei (2005) commented that this definition of neutrality is too stringent and suggested that neutrality should be defined by $|s| < 0.5/\sqrt{N}$. If we assume that N is, on average, approximately

10^5 in mammals, the neutrality definition is $|s| < 0.001$ (Nei 2005). The use of Nei's definition would certainly result in an increase of the fraction of substitutions that are regarded as neutral.

In the last two decades, many biologists have tried to identify the action of positive selection on individual genes, especially by inferring selective sweeps and by comparing nonsynonymous and synonymous substitution rates from sequence data (Kreitman and Akashi 1995; Nielsen 2005; Zhang 2010). Experimentalists also tried to identify nucleotide changes that significantly affect gene function or expression and thus potentially affect organismal fitness. While the occurrence of positive selection can be corroborated when the two approaches are consistent (e.g., Sawyer et al. 2005; Zhang 2006), it is important to note that they do not need to be consistent, because each method has its errors and shortcomings and because the two approaches may be measuring quite different things. For example, experimental studies are usually focused on some but not all aspects of the many functions of a gene and can easily miss important functional changes that improve fitness. It has also been reported that some sites that show function-altering substitutions are not found to be under positive selection by statistical analysis. Such inconsistency might be attributable to insufficient statistical power, but it is also possible that the functional changes detected in experiments have no impact on fitness. Experimental tests can sometimes demonstrate selection, but fitness differences that outweigh random drift are often too small to be detected statistically in the lab; moreover, it is frequently difficult to know the environment in which an evolutionary change transpired or to measure all possible components of total fitness.

Because of its central importance in evolutionary biology, the neutralist–selectionist debate is unlikely to be outgrown soon. Fortunately, the debate is becoming more quantitative than qualitative (i.e., about the percentage of fixations that are adaptive and the fitness effects of the fixed changes), and I believe that new findings will continue to emerge from genomic studies.

Outlook

Evolutionary genetics is arguably one of the fastest moving fields in biology. This phenomenon is in part because a complete understanding of any issue in biology requires an explanation of its evolutionary origin that ultimately must reach the level of molecular genetics. In turn, evolutionary genetics has benefited from having close ties with other fields of biology, especially molecular biology and genomics. Not only did evolutionary geneticists quickly adopt new technologies developed in other fields (e.g., protein electrophoresis, DNA sequencing) to address long-standing evolutionary problems (e.g., population genetics, phylogenetics), but they also have quickly identified intriguing evolutionary questions arising from new discoveries in other fields (e.g., origin of introns). Thus, in addition to the core set of long-standing questions described in the previous sections,

evolutionary geneticists constantly discover new questions. For example, codon usage bias, first reported 30 years ago (Grantham et al. 1980) and now well documented in many species, is still not fully understood. Codon usage bias is generally thought to be a result of a selection-mutation–drift balance (Bulmer 1991). Here, the word "mutation" refers to mutational bias that causes unequal equilibrium nucleotide frequencies, while the word "selection" refers to translational optimization. However, exactly what is optimized in translation that results in codon usage bias is not so clear. There is unambiguous evidence that codon usage bias is at least partly caused by selection for translational accuracy (Akashi 1994; Drummond and Wilke 2008), but whether it is also caused by selection for translational speed/efficiency remains elusive (Hershberg and Petrov 2008). Another example is the problem of the rate determinants of protein-sequence evolution. Quite surprisingly, about 10 years ago, it was found that the rate of protein-sequence evolution is determined mainly by the gene expression level rather than by the relative importance of the gene, measured by the fitness effect of gene deletion (Hurst and Smith 1999; Pal et al. 2001; Wang and Zhang 2009). Drummond and Wilke (2008) proposed a translational robustness hypothesis to explain why highly expressed proteins evolve more slowly than weakly expressed proteins. They suggest that translational errors often result in protein misfolding, which could be toxic to the cell. Thus, highly expressed proteins are expected to evolve DNA sequences that are more robust to mistranslation-induced misfolding and are thus more conserved than are lowly expressed proteins. There is still no direct evidence for the key assumptions of this provocative hypothesis, but it exemplifies how new evolutionary hypotheses arise from new molecular and genomic information.

Although many of the basic tenets of the neo-Darwinism have remained intact for the last 50 years, evolutionary genetics has made dramatic progress, thanks to the two technological revolutions and one conceptual breakthrough in the last half century. While I expect most of the basic tenets of neo-Darwinism to remain largely unchanged in the next 50 years or more, we can have full confidence that understanding in evolutionary genetics will increase greatly as a result of large amounts of genomic data and the experimental power of molecular biology, functional genomics, and systems biology. If the last 50 years of evolutionary genetics was characterized by deepening the understanding of evolution to the molecular level, the next 50 years will surely see the broadening of our understanding to the genomic scale and the systems level.

Acknowledgments

Literature Cited

Akashi, H. 1994. Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy. *Genetics* 136: 927–935.

Akiba, T., K. Koyama, Y. Ishiki, and 2 others. 1960. On the mechanism of the development of multiple-drug-resistant clones of *Shigella*. *Jpn. J. Microbiol.* 4: 219–227.

Albin, R. L. 1993. Antagonistic pleiotropy, mutation accumulation, and human genetic disease. *Genetica* 91: 279–286.

Barton, N. H. and B. Charlesworth. 1998. Why sex and recombination? *Science* 281: 1986–1990.

Bateson, W. and G. Mendel. 1909. *Mendel's Principles of Heredity: A Defence, with a Translation of Mendel's Original Papers on Hybridisation.* Cambridge University Press, Cambridge, UK.

Beyer, A., S. Bandyopadhyay, and T. Ideker. 2007. Integrating physical and genetic maps: From genomes to interaction networks. *Nat. Rev. Genet.* 8: 699–710.

Blower, P. J. 1989. *The Mendelian Revolution.* The Athlone Press, London.

Boone, C., H. Bussey, and B. J. Andrews. 2007. Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* 8: 437–449.

Brem, R. B., G. Yvert, R. Clinton, and 1 other. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* 296: 752–755.

Brett, D., H. Pospisil, J. Valcarcel, and 2 others. 2002. Alternative splicing and genome complexity. *Nat. Genet.* 30: 29–30.

Bridges, C. B. 1936. The Bar "gene" a duplication. *Science* 83: 210–211.

Brunner, H. G. and M. A. van Driel. 2004. From syndrome families to functional genomics. *Nat. Rev. Genet.* 5: 545–551.

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

Cargill, M., D. Altshuler, J. Ireland, and 14 others. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22: 231–238.

Carroll, S. B. 2005. Evolution at two levels: On genes and form. *PLoS Biol.* 3: e245.

Carroll, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* 134: 25–36.

Carroll, S. B., J. K. Grenier, and S. D. Weatherbee. 2001. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design.* Blackwell Scientific, Malden, MA.

Chan, Y. F., M. E. Marks, F. C. Jones, and 13 others. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327: 302–305.

Chen, L., A. L. DeVries, and C. H. Cheng. 1997. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc. Natl. Acad. Sci. USA* 94: 3811–3816.

Chen, S. T., H. C. Cheng, D. A. Barbash, and 1 other. 2007. Evolution of hydra, a recently evolved testis-expressed gene with nine alternative first exons in *Drosophila melanogaster*. *PLoS Genet.* 3: e107.

Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437: 69–87.

Cho, S., Z. Y. Huang, D. R. Green, and 2 others. 2006. Evolution of the complementary sex-determination gene of honey bees: Balancing selection and transspecies polymorphisms. *Genome Res.* 16: 1366–1375.

Ciccarelli, F. D., T. Doerks, C. von Mering, and 2 others. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* 311: 1283–1287.

Clark, A. G. and T. H. Kao. 1991. Excess nonsynonymous substitution of shared polymorphic sites among self-incompatibility alleles of Solanaceae. *Proc. Natl. Acad. Sci. USA* 88: 9823–9827.

Cordell, H. J. 2002. Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* 11: 2463–2468.

Coyne, J. A. 1992. Genetics and speciation. *Nature* 355: 511–515.

Cracraft, J. and M. J. Donoghue (eds.). 2004. *Assembling the Tree of Life.* Oxford University Press, New York.

Crow, J. F. and M. Kimura. 1979. Efficiency of truncation selection. *Proc. Natl. Acad. Sci. USA* 76: 396–399.

Daubin, V., N. A. Moran, and H. Ochman. 2003. Phylogenetics and the cohesion of bacterial genomes. *Science* 301: 829–832.

Delsuc, F., H. Brinkmann, and H. Philippe. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6: 361–375.

De Luca, M., N. V. Roshina, G. L. Geiger-Thornsberry, and 3 others. 2003. Dopa decarboxylase (*Ddc*) affects variation in *Drosophila* longevity. *Nat. Genet.* 34: 429–433.

Deutschbauer, A. M. and R. W. Davis. 2005. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat. Genet.* 37: 1333–1340.

Doniger, S. W., H. S. Kim, D. Swain, and 4 others. 2008. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* 4: e1000183.

Doolittle, W. F. 1999. Phylogenetic classification and the universal tree. *Science* 284: 2124–2129.

Drummond, D. A. and C. O. Wilke. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134: 341–352.

ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799–816.

Eyre-Walker, A. 2006. The genomic rate of adaptive evolution. *Trends Ecol. Evol.* 21: 569–575.

Falconer, D. S. and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics.* Pearson, London.

Fay, J. C., G. J. Wyckoff, and C.-I. Wu. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* 415: 1024–1026.

Felsenstein, J. 2004. *Inferring Phylogenies.* Sinauer Associates, Sunderland, MA.

Fisher, R. A. 1918. The correlations between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh* 52: 399–433.

Fisher, R. A. 1930. *The Genetic Theory of Natural Selection*, 2nd ed. Clarendon, Oxford.

Fitch, W. M. and E. Margoliash. 1967. Construction of phylogenetic trees. *Science* 155: 279–284.

Fleischmann, R. D., M. D. Adams, O. White, and 37 others. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269: 496–512.

Force, A., M. Lynch, F. B. Pickett, and 3 others. 1999. Preservation of duplicate genes by complementary, degenerate mutations. *Genetics* 151: 1531–1545.

Foster, K. R., G. Shaulsky, J. E. Strassmann, and 2 others. 2004. Pleiotropy as a mechanism to stabilize cooperation. *Nature* 431: 693–696.

Futuyma, D. J. 1998. *Evolutionary Biology*, 3rd ed. Sinauer Associates, Sunderland, MA.

Garland Jr., T. and M. Rose (eds). 2009. *Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments*. University of California Press, Berkeley.

Genome 10K Community of Scientists. 2009. Genome 10K: A proposal to obtain whole-genome sequence for 10,000 vertebrate species. *J. Hered.* 100: 659–674.

Gibbs, R. A., G. M. Weinstock, M. L. Metzker, and 230 others. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428: 493–521.

Gibson, G. and S. V. Muse. 2004. *A Primer of Genome Science*, 2nd ed. Sinauer Associates, Sunderland, MA.

Gilbert, W. 1978. Why genes in pieces? *Nature* 271: 501.

Gilbert, W. 1987. The exon theory of genes. *Cold Spring Harb. Symp. Quant. Biol.* 52: 901–905.

Glazko, G., V. Veeramachaneni, M. Nei, and 1 other. 2005. Eighty percent of proteins are different between humans and chimpanzees. *Gene* 346: 215–219.

Gogarten, J. P., W. F. Doolittle, and J. G. Lawrence. 2002. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19: 2226–2238.

Gogarten, M. B., J. P. Gogarten, and L. Olendzenski (eds.). 2009. *Horizontal Gene Transfer: Genomes in Flux*. Humana Press, New York.

Golding, G. B., N. Tsao, and R. E. Pearlman. 1994. Evidence for intron capture: An unusual path for the evolution of proteins. *Proc. Natl. Acad. Sci. USA* 91: 7506–7509.

Gould, S. J. and R. C. Lewontin. 1979. The spandrels of San Marco and the panglossian paradigm: A critique of the adaptationist programme. *Proc. Roy. Soc. Lond. B* 205: 581–598.

Grantham, R., C. Gautier, M. Gouy, and 2 others. 1980. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 8: r49–r62.

Gregory, T. R. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev. Camb. Philos. Soc.* 76: 65–101.

Gresham, D., M. M. Desai, C. M. Tucker, and 6 others. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* 4: e1000303.

Hahn, M. W. 2008. Toward a selection theory of molecular evolution. *Evolution* 62: 255–265.

Haldane, J. B. S. 1957. The cost of natural selection. *J. Genet.* 55: 511–524.

Harris, H. 1966. Enzyme polymorphisms in man. *Proc. Roy. Soc. Lond. B* 164: 298–310.

Hartman, J. L., B. Garvik, and L. Hartwell. 2001. Principles for the buffering of genetic variation. *Science* 291: 1001–1004.

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

He, X., W. Qian, Z. Wang, and 2 others. 2010. Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. *Nat. Genet.* 42: 272–276.

Hermisson, J. and A. P. McGregor. 2008. Pleiotropic scaling and QTL data. *Nature* 456: E3–E4.

Herring, C. D., A. Raghunathan, C. Honisch, and 8 others. 2006. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat. Genet.* 38: 1406–1412.

Hershberg, R. and D. A. Petrov. 2008. Selection on codon bias. *Ann. Rev. Genet.* 42: 287–299.

Hill, J. A. and S. P. Otto. 2007. The role of pleiotropy in the maintenance of sex in yeast. *Genetics* 175: 1419–1427.

Hillis, D. M., C. Moritz, and B. K. Mable (eds.). 1996. *Molecular Systematics.* Sinauer Associates, Sunderland, MA.

Hodgkin, J. 1998. Seven types of pleiotropy. *Int. J. Dev. Biol.* 42: 501–505.

Hoekstra, H. E. and J. A. Coyne. 2007. The locus of evolution: Evo devo and the genetics of adaptation. *Evolution* 61: 995–1016.

Hubby, J. L. and R. C. Lewontin. 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54: 577–594.

Hughes, A. L. and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167–170.

Hurst, L. D. and N. G. Smith. 1999. Do essential genes evolve slowly? *Curr. Biol.* 9: 747–750.

Jain, R., M. C. Rivera, and J. A. Lake. 1999. Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl. Acad. Sci. USA* 96: 3801–3806.

Jasnos, L. and R. Korona. 2007. Epistatic buffering of fitness loss in yeast double deletion strains. *Nat. Genet.* 39: 550–554.

Johanson, U., J. West, C. Lister, and 3 others. 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* 290: 344–347.

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

Kim, E., A. Magen, and G. Ast. 2007. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* 35: 125–131.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217: 624–626.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge, UK.

King, J. L. and T. H. Jukes. 1969. Non-Darwinian evolution. *Science* 164: 788–798.

King, M. C. and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188: 107–116.

Kondrashov, A. S. 1988. Deleterious mutations and the evolution of sexual reproduction. *Nature* 336: 435–440.

Kondrashov, A. S. and J. F. Crow. 1991. Haploidy or diploidy: Which is better? *Nature* 351: 314–315.

Koonin, E. V., K. S. Makarova, and L. Aravind. 2001. Horizontal gene transfer in prokaryotes: Quantification and classification. *Ann. Rev. Microbiol.* 55: 709–742.

Kreitman, M. and H. Akashi. 1995. Molecular evidence for natural selection. *Ann. Rev. Ecol. Syst.* 26: 403–422.

Kritsky, G. 1973. Mendel, Darwin, and evolution. *Am. Biol.Teach.* 35: 477–479.

Levine, M. T., C. D. Jones, A. D. Kern, and 2 others. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc. Natl. Acad. Sci. USA* 103: 9935–9939.

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

Liao, B., M. Weng, and J. Zhang. 2010. Contrasting genetic paths to morphological and physiological evolution. *Proc. Natl. Acad. Sci. USA* 107: 7353–7358.

Liao, B. Y. and J. Zhang. 2008. Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc. Natl. Acad. Sci. USA* 105: 6987–6992.

Linnen, C. R., E. P. Kingsley, J. D. Jensen, and 1 other. 2009. On the origin and spread of an adaptive allele in deer mice. *Science* 325: 1095–1098.

Liti, G., D. M. Carter, A. M. Moses, and 23 others. 2009. Population genomics of domestic and wild yeasts. *Nature* 458: 337–341.

Long, M., E. Betran, K. Thornton, and 1 other. 2003. The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* 4: 865–875.

Lynch, M. 2007. *The Origins of Genome Architecture.* Sinauer Associates, Sunderland, MA.

Lynch, M. and A. Force. 2000a. Gene duplication and the origin of interspecific genomic incompatibility. *Am. Nat.* 156: 590–605.

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

Lynch, M., W. Sung, K. Morris, and 8 others. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. USA* 105: 9272–9277.

Mackay, T. F. 2001. The genetic architecture of quantitative traits. *Ann. Rev. Genet.* 35: 303–339.

Mackay, T. F. and R. F. Lyman. 2005. *Drosophila* bristles and the nature of quantitative genetic variation. *Philos. Trans. Roy. Soc. Lond. B* 360: 1513–1527.

Manolio, T. A., F. S. Collins, N. J. Cox, and 24 others. 2009. Finding the missing heritability of complex diseases. *Nature* 461: 747–753.

Mardis, E. R. 2008. Next-generation DNA sequencing methods. *Ann. Rev. Genomics Hum. Genet.* 9: 387–402.

Matlin, A. J., F. Clark, and C. W. Smith. 2005. Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell. Biol.* 6: 386–398.

Moore, J. H. and S. M. Williams. 2005. Traversing the conceptual divide between biological and statistical epistasis: Systems biology and a more modern synthesis. *BioEssays* 27: 637–646.

Moriyama, E. N. and J. R. Powell. 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* 13: 261–277.

Mushegian, A. R. and E. V. Koonin. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA* 93: 10268–10273.

Nei, M. 1975. *Molecular Population Genetics and Evolution.* North-Holland, Amsterdam.

Nei, M. 2005. Selectionism and neutralism in molecular evolution. *Mol. Biol. Evol.* 22: 2318–2342.

Nei, M. and D. Graur. 1984. Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* 17: 73–118.

Nei, M. and S. Kumar. 2000. *Molecular Evolution and Phylogenetics.* Oxford University Press, New York.

Nei, M., X. Gu, and T. Sitnikova. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* 94: 7799–7806.

Nekrutenko, A. and W.-H. Li. 2001. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* 17: 619–621.

Nelson, K. E., R. A. Clayton, S. R. Gill, and 26 others. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399: 323–329.

Nielsen, R. 2005. Molecular signatures of natural selection. *Ann. Rev. Genet.* 39: 197–218.

Ochiai, K., T. Yamanaka, K. Kimura, and 1 other. 1959. Inheritance of drug resistance (and its tranfer) between *Shigella* strains and between *Shigella* and *E. coli* strains (in Japanese). *Hihon Iji Shimpor* 1861: 34.

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

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Ann. Rev. Ecol. Syst.* 23: 263–286.

Ohya, Y., J. Sese, M. Yukawa, and 22 others. 2005. High-dimensional and large-scale phenotyping of yeast mutants. *Proc. Natl. Acad. Sci. USA* 102: 19015–19020.

Orr, H. A. 2000. Adaptation and the cost of complexity. *Evolution* 54: 13–20.

Osawa, S. 1995. *Evolution of the Genetic Code.* Oxford University Press, Oxford.

Pal, C., B. Papp, and L. D. Hurst. 2001. Highly expressed genes in yeast evolve slowly. *Genetics* 158: 927–931.

Patthy, L. 1995. *Protein Evolution by Exon-Shuffling.* Molecular Biology Intelligence Unit, R. G. Landes Company/Springer, New York.

Patthy, L. 1999. Genome evolution and the evolution of exon-shuffling—a review. *Gene* 238: 103–114.

Pevsner, J. 2009. *Bioinformatics and Functional Genomics*, 2nd ed. Wiley-Blackwell, Hoboken, NJ.

Phillips, P. C. 2008. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9: 855–867.

Piatigorsky, J. 2007. *Gene Sharing and Evolution.* Harvard University Press, London.

Raff, R. A. 1996. *The Shape of Life: Genes, Development, and the Evolution of Animal Form.* University of Chicago Press, Chicago.

Raj, A. and A. van Oudenaarden. 2008. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135: 216–226.

Raser, J. M. and E. K. O'Shea. 2005. Noise in gene expression: Origins, consequences, and control. *Science* 309: 2010–2013.

Rebeiz, M., J. E. Pool, V. A. Kassner, and 2 others. 2009. Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* 326: 1663–1667.

Roguev, A., S. Bandyopadhyay, M. Zofall, and 13 others. 2008. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* 322: 405–410.

Rokas, A. and P. W. Holland. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15: 454–459.

Roman, J. and S. R. Palumbi. 2003. Whales before whaling in the North Atlantic. *Science* 301: 508–510.

Roy, S. W. and W. Gilbert. 2006. The evolution of spliceosomal introns: Patterns, puzzles and progress. *Nat. Rev. Genet.* 7: 211–221.

Sawyer, S. A., J. Parsch, Z. Zhang, and 1 other. 2007. Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 104: 6504–6510.

Sawyer, S. L., L. I. Wu, M. Emerman, and 1 other. 2005. Positive selection of primate TRIM5 identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. USA* 102: 2832–2837.

Sclater, A. 2006. The extent of Charles Darwin's knowledge of Mendel. *J. Biosci.* 31: 191–193.

Sella, G., D. A. Petrov, M. Przeworski, and 1 other. 2009. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* 5: e1000495.

Shapiro, J. A., W. Huang, C. Zhang, and 9 others. 2007. Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. USA* 104: 2271–2276.

Shapiro, M. D., M. E. Marks, C. L. Peichel, and 5 others. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428: 717–723.

Shendure, J. and H. Ji. 2008. Next-generation DNA sequencing. *Nat. Biotechnol.* 26: 1135–1145.

Shi, P. and J. Zhang. 2009. Extraordinary diversity of chemosensory receptor gene repertoires among vertebrates. In W. Meyerhof and S. Korsching (eds.), *Chemosensory Systems in Mammals, Fishes, and Insects,* pp. 1–23. Springer, Berlin.

Smith, N. G. and A. Eyre-Walker. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415: 1022–1024.

Sorek, R., Y. Zhu, C. J. Creevey, and 3 others. 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318: 1449–1452.

Steinmetz, L. M., H. Sinha, D. R. Richards, and 5 others. 2002. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 416: 326–330.

Stephens, S. G. 1951. Possible significances of duplication in evolution. *Adv. Genet.* 4: 247–265.

Stern, D. L. and V. Orgogozo. 2008. The loci of evolution: How predictable is genetic evolution? *Evolution* 62: 2155–2177.

Stern, D. L. and V. Orgogozo. 2009. Is genetic evolution predictable? *Science* 323: 746–751.

Sucena, E. and D. L. Stern. 2000. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proc. Natl. Acad. Sci. USA* 97: 4530–4534.

Takahashi, A., S. C. Tsaur, J. A. Coyne, and 1 other. 2001. The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 98: 3920–3925.

Tischler, J., B. Lehner, and A. G. Fraser. 2008. Evolutionary plasticity of genetic interaction networks. *Nat. Genet.* 40: 390–391.

Tong, A. H., G. Lesage, G. D. Bader, and 47 others. 2004. Global mapping of the yeast genetic interaction network. *Science* 303: 808–813.

Tung, J., A. Primus, A. J. Bouley, and 3 others. 2009. Evolution of a malaria resistance gene in wild primates. *Nature* 460: 388–391.

Turelli, M. 1985. Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. *Genetics* 111: 165–195.

Tyler, A. L., F. W. Asselbergs, S. M. Williams, and 1 other. 2009. Shadows of complexity: What biological networks reveal about epistasis and pleiotropy. *BioEssays* 31: 220–227.

Vasemägi, A. and C. R. Primmer. 2005. Challenges for identifying functionally important genetic variation: The promise of combining complementary research strategies. *Mol. Ecol.* 14: 3623–3642.

Voight, B. F., S. Kudaravalli, X. Wen, and 1 other. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4: e72.

Wagner, G. P. 1988. The influence of variation and developmental constraints on the rate of multivariate phenotypic evolution. *J. Evol. Biol.* 1: 44–66.

Wagner, G. P., J. P. Kenney-Hunt, M. Pavlicev, and 3 others. 2008. Pleiotropic scaling of gene effects and the 'cost of complexity'. *Nature* 452: 470–472.

Wang, Z. and J. Zhang. 2009. Why is the correlation between gene importance and gene evolutionary rate so weak? *PLoS Genet.* 5: e1000329.

Waterston, R. H., K. Lindblad-Toh, E. Birney, and 220 others. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520–562.

Waxman, D. and J. R. Peck. 1998. Pleiotropy and the preservation of perfection. *Science* 279: 1210–1213.

Werth, C. and M. Windham. 1991. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am. Nat.* 137: 515–526.

Williams, G. C. 1957. Pleiotropy, natural selection, and the evolution of senescene. *Evolution* 11: 398–411.

Wittkopp, P. J., E. E. Stewart, L. L. Arnold, and 6 others. 2009. Intraspecific polymorphism to interspecific divergence: Genetics of pigmentation in *Drosophila*. *Science* 326: 540–544.

Wray, G. A. 2007. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8: 206–216.

Wright, S. 1968. *Evolution and the Genetics of Populations.* University of Chicago Press, Chicago.

Yeh, P., A. I. Tschumi, and R. Kishony. 2006. Functional classification of drugs by properties of their pairwise interactions. *Nat. Genet.* 38: 489–494.

Yomo, T., I. Urabe, and H. Okada. 1992. No stop codons in the antisense strands of the genes for nylon oligomer degradation. *Proc. Natl. Acad. Sci. USA* 89: 3780–3784.

Yoshiura, K., A. Kinoshita, T. Ishida, and 36 others. 2006. A SNP in the ABCC11 gene is the determinant of human earwax type. *Nat. Genet.* 38: 324–330.

Zhang, J. 2003. Evolution by gene duplication—an update. *Trends Ecol. Evol.* 18: 292–298.

Zhang, J. 2006. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat. Genet.* 38: 819–823.

Zhang, J. 2007. The drifting human genome. *Proc. Natl. Acad. Sci. USA* 104: 20147–20148.

Zhang, J. 2010. Positive Darwinian selection in gene evolution. In M. Long and H. Gu (eds.), *Darwin's Heritage Today—Proceedings of Darwin 200 International Beijing Conference*, High Education Press, Beijing, in press.

Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* 95: 3708–3713.

Zhang, L. and W.-H. Li. 2005. Human SNPs reveal no evidence of frequent positive selection. *Mol. Biol. Evol.* 22: 2504–2507.