Chapter 22

Evolutionary Biology: The Next 150 Years

Hopi E. Hoekstra

Darwin was arguably the most prescient thinker that biology has ever witnessed. But, if someone had asked him in 1859 where evolutionary biology would be in 150 years, would he have guessed correctly? He might have predicted that we would have a better understanding of how traits are inherited—a prediction borne out almost 50 years later with the rediscovery of Mendel's laws in 1900. Darwin considered the lack of understanding for how traits are inherited to be the missing link in his argument for evolution by natural selection, and when pushed, he devised his own theory (i.e., pangenesis), which was one of his few major errors. Yet, the ramifications of Mendel's experiments or of subsequent discoveries, like that of the three-dimensional DNA structure by Watson and Crick (1953) a century after *The Origin of Species* was published, along with the resultant technological advances, such as the ability to sequence a complete genome in another 50 years, were unknowable in his day. Nor could Darwin have anticipated the questions that have dominated the field since, such as the relative role of drift and selection in driving molecular evolution (Kimura 1968). With the acknowledgement that technologies, discoveries, and questions will arise that, likewise, cannot be imagined, it is useful—perhaps even stimulating—to speculate about what the next 150 (or more modestly, 50 or even 20) years will hold for evolutionary biology. The organizers of the Darwin 2009 Workshop asked me to speculate on what may lie ahead. My crystal ball, if I have one, is colored by evolutionary genetics and genomics—my main research area—and so necessarily are my predictions.

To predict the advances in the field of evolutionary biology, we (i.e., evolutionary biologists) must first set a direction. One overarching goal of evolutionary biology is to understand how the diversity of life evolved, and more specifically to understand how this variation, both genetic and phenotypic, is generated and maintained in nature. This aim spans sub-

disciplines, ranging from the origin of life (see Lazcano, Chapter 14), paleontology (see P. Wagner, Chapter 17), and phylogenetics (see Hillis, Chapter 16) to theoretical population genetics (see Wakeley, Chapter 5) and evolutionary ecology (see Agrawal et al., Chapter 10). Acquiring this knowledge may also help us address some of the most pressing problems of our time, which include concerns about the future evolution (or extinction) of this biodiversity (see Gould, Chapter 21; Davis et al., Commentary 6).

The question of how diversity evolved is of course not a simple one and therefore, must be attacked from multiple angles and at several levels. At the most proximate level, we would like to know what mutations give rise to variation and how that genetic variability is maintained in populations (see Zhang, Chapter 4). Next, the question arises of *how* genetic variation actually produces variation in organisms, for example, through changes in developmental events, pathways, or processes (see Wray, Chapter 9). Deciphering how evolutionary forces act on this phenotypic variation in a given ecological context (see Agrawal et al., Chapter 10; Berenbaum and Schuler, Chapter 11)—and how different properties may constrain (see Kirkpatrick, Chapter 7) or promote (see G. Wagner, Chapter 8) phenotypic change—remains a major challenge. While the study of morphological variation offers a logical starting point, we also want to uncover the mechanisms responsible for variation and evolution of other characters (e.g., behavior) and understand how and why behavioral evolution may be similar to or different from morphological evolution (see Kokko and Jennions, Chapter 12). Moreover, we are not limited to processes occurring within a lineage; rather, our thinking must be extended to the genetic and ecological changes associated with speciation (see Harrison, Chapter 13) and macroevolutionary diversification (see Losos and Mahler, Chapter 15; Foote, Chapter 18). Finally, a temporal component must be added—changes in allele frequencies at specific loci (e.g., ancient DNA studies), genome composition (e.g., comparative genomics), phenotypes (e.g., fossils and character mapping using phylogenetics) and the environment—to understand change through time. Only when all of these pieces are taken together can we start to formulate a complete picture of evolutionary change.

Integration of these diverse approaches has long been an ideal in evolutionary biology. For example, in the introductory chapter (notably entitled "The Problem") of his 1974 book *The Genetic Basis of Evolutionary Change*, Richard Lewontin advocated a merger between scientists working on the genetic processes (e.g., population geneticists studying the impact of different evolutionary forces on changes in allele frequency) and those focusing on forces acting on phenotypes (e.g., field naturalists studying the role of differential survival and reproduction on phenotypic change across generations).

Before proceeding, however, let us review briefly some major advances that have led to this point. Not long after the rediscovery of Mendel's laws of inheritance, Thomas Hunt Morgan's mutational experiments in

Drosophila demonstrated that genes are carried on chromosomes, providing a material basis for heredity. Together, these discoveries also highlighted how phenotypic variation—whether it be the shape of peas or the eye color of flies—can be studied in the laboratory (and eventually in the field). Around the same time, the role of natural selection acting in the wild was being further documented. For example, using a simple general selection model, J. B. S. Haldane calculated the selective advantage necessary for the observed evolution of industrial melanism in peppered moths (Haldane 1924). However, it was the combination of critical contributions by Haldane, Sewall Wright, and most notably R. A. Fisher who showed how the action of many discrete genetic loci studied in the lab could result in continuous trait variation observed in nature. But, arguably, it was Dobzhansky's 1937 book *Genetics and the Origin of Species* that played a key role in bridging the gap between population geneticists and field naturalists. These works and others culminated in the modern evolutionary synthesis, that is, the union of ideas from scientists across several disciplines (from laboratory genetics to field ecology, systematics, and paleontology) about the way evolution proceeds.

From here, as is often the case, it was a technological advance that pushed the field forward. In the 1960s, Richard Lewontin moved to the University of Chicago, where he met Jack Hubby. Lewontin was an evolutionary geneticist with a question: how much genetic variation exists in natural populations? Hubby was a biochemist with a new technique: protein electrophoresis. It was the perfect union. Together, they surveyed 18 loci in *Drosophila pseudoobscura* and reported that a large fraction was polymorphic (Lewontin and Hubby 1966). This result had a great impact, as the discovery of high levels of molecular polymorphism raised the question of what evolutionary forces maintain this variation—a question that preoccupied population geneticists for decades. The ability to survey allozymes in wild populations also offered an exciting opportunity to connect genetic variation to fitness in nature. One of the earliest examples was reported by Watt (1977), who showed that variation at the phosphoglucoisomerase (PGI) locus is associated with fitness differences in natural populations of *Colias* butterflies. Yet, in the following three decades, there have been only modest steps toward cementing the link between genotype, phenotype, and the environment in any one system, that is, until very recently.

In the last few years, case studies have started to accumulate that provide a near complete picture of adaptive change, that is, the genes and mutations that underlie variation and divergence in traits that have documented fitness consequences in nature have been identified (Figure 22.1). To some, this represents the holy grail of evolutionary biology. In threespine stickleback (*Gasterosteus aculeatus*), changes in armor are associated with the invasion of freshwater lakes following the last glacial cycle, starting about 20,000 years ago, and a reduction in armor has measurable fitness consequences (Barrett et al. 2008; Marchinko 2009). At the genetic level, changes in armor are

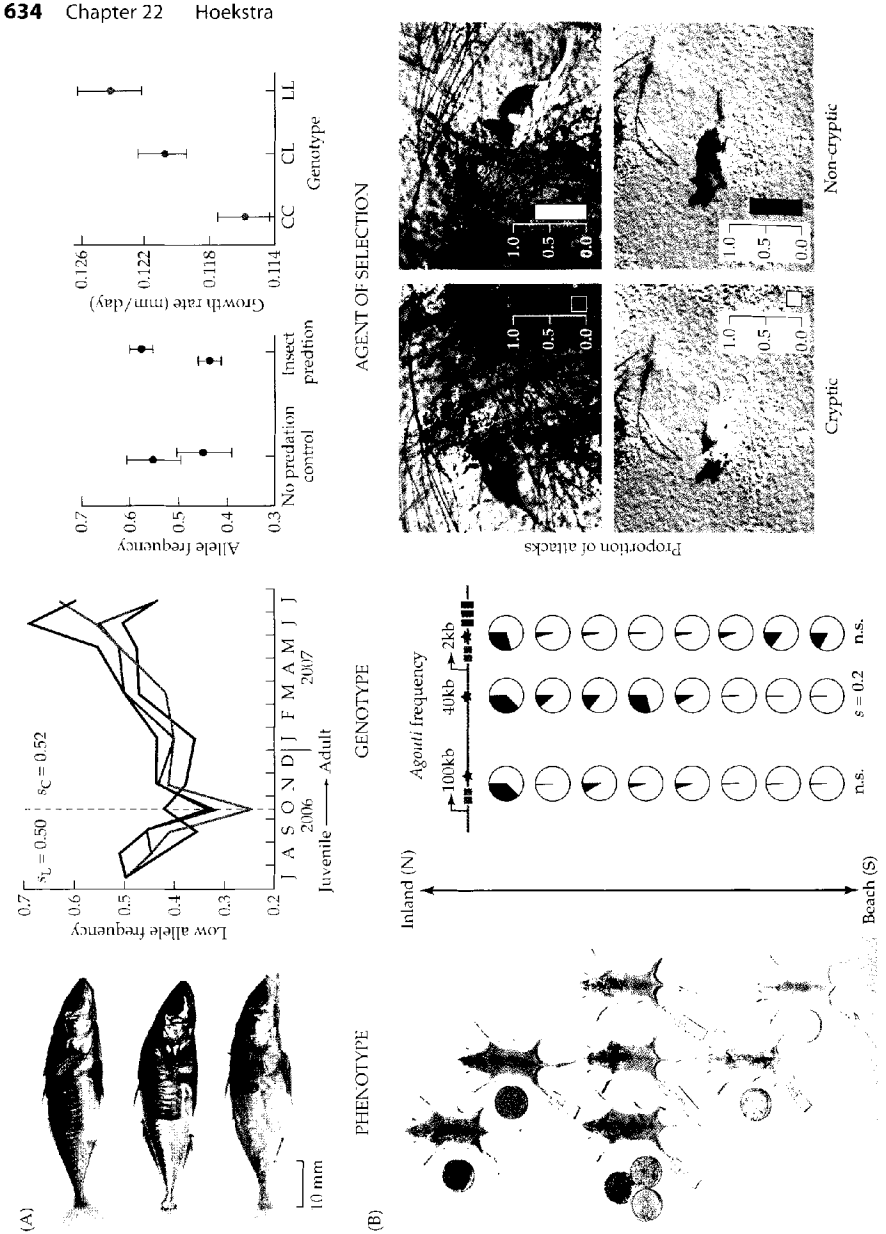

FIGURE 22.1 Two Examples for which Both the Targets (Phenotypic and Genotypic) and Agents of Natural Selection Have Been Identified (A) Selection on body armor in the threespine stickleback, *Gasterosteus aculeatus*. Left panel: complete (top), partial (middle), and low (bottom) lateral plate morphs. (From Barrett et al. 2008.) Middle panel: changes in low *Eda* allele frequency within a single generation in four replicate ponds (colored lines). Selection coefficients are given for selection against the low allele from July–October (s_L) and from selection against the complete allele from October–July (s_C). (Adapted from Barrett et al. 2008.) Right panel: relative to the complete *Eda* allele, individuals carrying the low *Eda* allele enjoy decreased predation by insects (left, adapted from Marchinko 2009) and increased growth rates in fresh water (right, adapted from Barrett et al. 2009). (B) Selection on coat color in the oldfield mouse, *Peromyscus polionotus*. Left panel: representative mice and soil sampled from collection sites along a 150-km transect from northwestern Florida (beach) to southeastern Alabama (inland). (From Mullen and Hoekstra 2008.) Middle panel: Allele frequencies at three polymorphic sites (stars) within the pigmentation gene *Agouti* (large boxes: coding exons; small boxes: untranslated exons), sampled from eight populations along the same 150-km transect. Pie charts are arranged North (top) to South (bottom), and frequency of alleles associated with light pelage are indicated in yellow and the dark pelage in red. One of the three *Agouti* single nucleotide polymorphisms (SNP; 40kb), but not the others, varies clinally; the selection coefficient is given for this SNP. (Adapted from Mullen and Hoekstra 2008.) Right panel: Increased attack rates on non-cryptic clay models relative to cryptic clay models on both light (beach) and dark (inland) soils demonstrates that visually hunting predators are an important selective agent targeting color variation within and between *P. polionotus* populations. (Modified from Vignieri and Hoekstra, in press; from Linnen and Hoekstra, in press.)

associated with changes in the *Ectodysplasin* (*Eda*; Colossimo et al. 2005) and the *PitX1* loci (Shapiro et al. 2004). Moreover, deletions of a *cis*-regulatory element in the *PitX1* locus alter pelvis-specific expression of *PitX1* during development (Chan et al. 2010). These phenotypic (and inferred allele) frequencies can be tracked through time, using both museum specimens (Bell et al. 2004; Kitano et al. 2008) and the fossil record (Hunt et al. 2008; Bell 2009). In a second example, oldfield mice (*Peromyscus polionotus*) have dorsal coat colors that closely match their local substrate, and the degree of color matching has a measurable effect on inferred predation rates (Vignieri et al., in press). Differences between a dark mainland and pale beach mouse subspecies have been attributed to three major genes (Steiner et al. 2007), and in one case, to a single amino acid mutation in the *Melanocortin-1 receptor* (*Mc1r*) gene, which affects ligand binding and receptor signaling that is associated with reduced pigmentation (Hoekstra et al. 2006). Variation in pigment allele frequencies also can be traced in space and time (Mullen and Hoekstra 2008; Mullen et al. 2009). Berenbaum and Schuler (see Chapter 11) describe a third such example, in which amino acid substitutions in a cytochrome P450 enzyme enhance the ability of swallowtail butterflies to detoxify defensive compounds in their food plants.

Such studies are now enabling us to answer long-standing questions in evolutionary biology, some first posed by the architects of the modern synthesis. These questions include, but are not limited to: Can adaptation take big leaps or does it proceed through many small steps? Are adaptive

alleles generally dominant or recessive? Is evolutionary change limited by mutation? To what extent is evolution constrained? How repeatable is evolution—are the same or different genetic solutions responsible for solving similar ecological problems? And, most recently, how do changes in or regulation of proteins during development produce phenotypic change? Answers to these general questions will come only by studying many traits in myriad species.

While these exemplars are among our most complete stories, it has taken nearly a decade to make the connection between genes, phenotype, and environment for a single trait in a single species. Moreover, at least with the stickleback and mouse examples, these studies are not unbiased, as they represent the best case scenario. In each case, there were *a priori* reasons to think that trait variation impacts fitness, and these morphological traits were easy to measure (and are influenced little, if at all, by the environment). Moreover, their genetic basis proved to be relatively simple—a few genes explain a large proportion of the phenotypic variation. How then can we be hopeful about making more progress in the future studying traits for which the functional consequences are unclear, that are challenging to measure accurately, and for which the genetic basis may be more complex? To do so, we face three major challenges: (1) describe the genomic (and other -omic) variation within and among species; (2) unravel how this genomic information is translated into phenotypic information; and (3) understand how evolutionary forces drive differentiation (and ultimately fitness) in an ecological context, which remains among the hardest tasks, even if (or perhaps because) it is not currently possible to study using high-throughput technology.

The Rise of Genomics

Our ability to comprehensively describe genetic variation by sequencing complete genomes, the basic blueprint of an organism, represents an extraordinary technological advance that is remaking the field of biology. The rate at which whole genome sequences are generated is astonishing. Complete genome sequencing started only 30 years ago when the modest 5368-base pair genome of bacteriophage fX174 was decoded (Sanger et al. 1977). It was quickly followed by several other, larger viral genomes. But, it took almost another 20 years until the first complete genome sequence of a free-living organism, *Haemophilus influenzae* (1.8 megabases), was finished because decoding a genome of this size required both technological and computational advances (Fleischmann et al. 1995).

Less than 6 years after this first complete genome sequence came the first complete *human* genome sequence—2.91 billion base pairs of euchromatic sequence (~5-fold coverage), with 2.1 million identified polymorphisms, at the reported cost of greater than $10 million (International Human Genome Sequencing Consortium [IHGSC] 2001; Venter et al. 2001) and when finished (i.e., the fold-coverage was high enough to provide reliable data),

TABLE 22.1 RAPID CHANGE IN HUMAN GENOME SEQUENCING COSTS

Year	Technology	Reference	Average reported coverage depth (fold)	Reported sequencing consumables cost	Estimated cost per 40-fold coverage
2004	IHGSC 2004	Sanger	5	$300,000,000	—
2007	Levy et al. 2007	Sanger (ABI)	7	$10,000,000	$57,000,000
2008	Wheeler et al. 2008	Roche (454)	7	$1,000,000	$5,700,000
2008	Bentley et al. 2008	Illumina	30	$250,000	$330,000
2009	Pushkarev et al. 2009	Helicos	28	$48,000	$69,000
2010	Drmanac et al. 2010	Nanoarrays	87	$8000	$3700
2010	Drmanac et al. 2010	Nanoarrays	63	$3500	$2200
2010	Drmanac et al. 2010	Nanoarrays	45	$1725	$1500

the cost was estimated to be closer to $300 million (IHGSC 2004). However, as the rate of published complete genome sequences increased, the cost decreased almost exponentially (Table 22.1). To underscore this point, on the first day of the Darwin 2009 Workshop that gave rise to this volume, a publication announced that three human genomes were sequenced at the cost of approximately $4400 (for consumables), with 45- to 87- fold coverage, identification of 3.2 to 4.5 million sequence variants per genome, and a 1-false-variant-per-100-kilobases accuracy (Drmanac et al. 2010)[1]. The rapid fall in sequencing prices is the genomic equivalent of Moore's Law, which describes the long-term trend in which the number of transistors that can be placed on computer chips doubles every 18 months, steadily driving down the cost of computing power.

Although this technology is largely driven by its applications to human disease (i.e., personalized medical genomics), the newest technologies are easily transferable to other species. At the time I write, the complete sequence is known for about 2000 viruses, 600 bacterial species, and roughly 200 eukaryotes including 60 chordates; these numbers will likely be different next week. As sequencing prices continue to fall, project proposals are becoming increasingly ambitious. For example, a recent proposal was put forth to sequence 10,000 vertebrate genomes in 5 years (Genome 10K Community of Scientists 2009). Arguably, all species[2] have the potential to be genome-enabled, just as 20 years ago it became routine to sequence the mitochondrial cytochrome b gene in any eukaryote.

[1] Published in *Science Express* on 5 November 2009.

[2] Genome sequencing of certain species undoubtedly will prove more challenging, such as those with extraordinarily large repetitive genomes or high levels of heterozygosity and those that are recent or ancient polyploids.

But with these new data flowing in at an unprecedented pace, we need tools and infrastructure in place to make them both accessible and usable. First, the need for well-maintained and long-term data repositories has never been greater (Robinson et al., in press). Second, for a genome sequence to be truly useful, it must be assembled and annotated, which is non-trivial. If, for example, the 10K Vertebrate Genome Project goes forward and reaches its goal of completion in 5 years, the processing rate will need to hold steady at five genomes per day (with no weekend or holiday breaks). Is this possible? Luckily, the field of bioinformatics—sometimes referred to as the symbiotic harmony of computer science and biology—is burgeoning. For example, Ensembl, a web-based community resource that provides genome data and analysis tools for a comprehensive set of chordate genomes, just celebrated its 10th anniversary and continues to evolve (Flicek et al. 2010). Moreover, the assembly and annotation of complete genomes likely will become more straightforward as more scaffold genomes (i.e., existing genome sequences that can be used as a reference to aid genome assembly of a new, closely related species) are available and computational algorithms become further automated. In fact, the first genome sequence based primarily on short-read, next-generation sequences, that of the giant panda, is purported to have been assembled in only 2 days, suggesting that the tools are already in place to assemble de novo a genome's worth of small DNA fragments even when a reference genome is unavailable (Li et al. 2010). Thus, the availability of genome sequences and their assembly is unlikely to be a limiting factor as we move forward.

Will the era of genomics end? Not any time soon. Instead I predict that, at least in the near future, it will continue to expand and do so exponentially. Furthermore, just as genome sequencing is becoming faster and cheaper, so too will the next level of -omics: epigenomics, trancriptomics, proteomics, metabolomics, and so forth. But, these inventories of genomic parts are limited in their utility without knowledge of the functional consequences of variation at any of these molecular levels (and we must be open to novel, perhaps unexpected, ways in which DNA sequence can confer function). For example, future studies geared at uncovering the function of noncoding or so-called junk DNA, epigenetic modification (e.g., methylation and chromatin remodeling), and spatio-temporal changes in RNA abundance on phenotype (and fitness) will undoubtedly open many windows into the evolutionary process. But, there is no doubt that evolutionary biology currently is and will continue to be transformed by our ability to sequence the genome of virtually any organism.

How to Build an Organism from Its Genomic Blueprint

Sydney Brenner is a brilliant scientist, probably best known for his contribution to the then emerging field of molecular biology in the 1960s, including his collaboration with Francis Crick to experimentally reveal the triplet nature of the genetic code. He later went on to establish *Caenorhabditis elegans* as a model system for the study of development, for which he received a Nobel Prize. Whether then it was blind enthusiasm, the naiveté of the time, or the shrewd promotion of genetics, that led him to say: "... give me the complete DNA sequence of any organism, and I can reconstruct it"[3] is unclear (R. C. Lewontin, personal communication). At the time (i.e., the 1980s), this sentiment was shared by many—the genome was the key that would unlock the secrets of biological complexity. It has become abundantly clear, however, that this assertion was indeed naïve; without the instruction manual, even with all the parts in hand, it is impossible (even at present) to reconstruct the whole organism.

Historically, molecular research has focused on identifying the parts—individual genes and proteins—and understanding their functions. While the reductionist program has been both successful and enlightening, to understand how whole organisms work, we need to consider the individual components both through developmental time (Carroll et al. 2001) and in the context of their interactions and as part of large networks (Noble 2006). By analogy, knowing all the parts of an airplane lends little to our understanding of how the plane actually functions. This daunting task falls largely under the joint purview of developmental biology and systems biology, which is the study of organisms as an integrated and interacting network of genes, proteins, and biochemical reactions (Sauer et al. 2007). Systems biology is still in its infancy, but is gaining momentum due to a host of new technologies that are high throughput, quantitative, and large-scale (Zhu and Snyder 2002).

At the moment, many of these new systems-level approaches are focused at the cellular level or conducted in relatively simple model organisms. More recently, the evolution of metabolic networks has been investigated, that is, how selection acts to optimize fitness across a landscape of networks (Pfeiffer et al. 2005), which may have implications for the early stages in the origin of life. Eventually, however, new technologies aimed at organisms that are more complex will be required, especially to automate the characterization of expression patterns (e.g., high-throughput *in situ* hybridizations) and the implementation of functional assays (e.g., RNAi, viral vectors, transformations, transgenesis; Kitano 2002). While some tools will necessarily be species- or clade-specific, the most useful advances undoubtedly will be those that are easily transferable across organisms. Thus, with new tools, we may acquire access to a fuller understanding of how genes and genomes produce phenotypes through changes in development, how historical processes have shaped that transformation, and how

[3] This is not an exact quote, but has been recounted by many, who were in attendance at Brenner's keynote address at the Cambridge University Symposium (1982) on the occasion of the 100th anniversary of Darwin's death.

constraints and opportunities may limit or promote future evolutionary change (Schwenk et al. 2009).

Thus, just as molecular biology has expanded beyond a focus on single genes or proteins, so too has developmental biology (see Wray, Chapter 9). Establishing a true understanding of how pathways and networks give rise to cellular and organismal phenotypes and how those interactions evolve will require very large experimental data sets. Thus, I predict that the largest advances will come by generating network models, predicting how they affect the phenotype, testing hypotheses derived from these models, and refining the models based on new experimentation. Importantly, understanding the functional interactions of genes, RNA, and proteins will provide a necessary first step in translating the genetic code into phenotypes and, ultimately, into fitness of organisms in nature.

The Genotype–Phenotype Connection in Nature

While systems biology may aim to unravel the details of all gene-network interactions (i.e., to identify the genes and connections necessary to produce a phenotype or whole organism), most evolutionary biologists are primarily concerned with changes in genes or development that contribute to phenotypic variation: in particular, identifying genes that are responsible for local adaptation, phenotypic novelties, and/or the promotion or restriction of diversification. To accomplish this goal, there are at least four major areas in which tremendous growth may be anticipated.

New Model Systems

Over the past several decades, tool and resource development, and hence research effort, have largely (and necessarily) focused on a handful of model organisms. However, these few species are not representative of the vast diversity of life. Thus, a broader array of organisms, replete with genome sequence data and functional tools, is needed to elucidate the genetic and developmental basis of organismal diversity (Jenner and Wills 2007). But at present, we have limitations. While genome sequencing can be done in any and all species of interest, not all species or clades can immediately become new model systems in the traditional sense; there are limited resources and a limited number of research communities to study any particular species. Just as some species are best suited to address specific biological questions,[4] others are better suited, at least initially, to become a model species—those that can be easily obtained, cultured in the lab, genetically crossed, and for which functional tests are feasible (Abzhanov et al. 2008). Yet, as technology, tools, and resources continue to develop, it is feasible to imagine that the term "model organism" eventually and finally will be eliminated from

[4] As August Krogh famously said, "for many problems there is an animal on which it can be most conveniently studied" (Krebs 1975).

biology's lexicon. Thus, as a diversity of species with appropriate genetic, genomic, developmental, and functional tools emerge, we will soon be able to ask questions at a number of levels, from variation among individuals to adaptation among populations to the evolution of novel traits between species, and understand macroevolutionary patterns by comparing across a rich array of taxa.

Population Genomics

It is inappropriate to think of "*the*" genome of a single species, because this thinking fails to capitalize on intraspecific variation. In fact, much of evolutionary biology's future promise may rest in the hands of population genomicists, those making large-scale comparisons of genome-scale DNA sequences among individuals sampled from natural populations. Why? While comparisons among phylogenetically dispersed taxa have been extremely useful in understanding genome evolution, such broad comparisons can only go so far in linking genetic to phenotypic variation. It can be argued that the comparison of genomes from collections of variable individuals is the key to understanding variation in phenotypes. Thus, I expect to see a shift in DNA sequencing efforts from single individuals in many species to many individuals in a few species, and in the future, to many, many individuals in many, many species.

The power of population genomic sampling stems from our ability to both apply population genetic tests of non-neutrality and to statistically associate genetic variants with phenotypic variation across the genome. These statistical approaches were initially designed for application to single genes, first to allozyme alleles and then to DNA sequence polymorphisms (e.g., McDonald and Kreitman 1991), but can now be applied to the whole genome (see Kolaczkowski and Kern, Chapter 6). As theoretical population geneticists and statisticians become better at identifying signatures of selection in the genome (Jensen et al. 2005; Nielsen et al. 2005), we will be better poised to identify genomic regions, genes and mutations associated with phenotypes, and the action of natural selection. The success of these statistical approaches also rests on having plenty of data. The Human HapMap project represented a major effort to generate genome-wide population data but was limited in that only single nucleotide variants and those that were most common (>~5% frequency) were considered (The International HapMap Consortium 2005). New efforts aimed at surveying and including rare variants and other forms of variation (e.g., copy-number variants) include the 1000 Genomes Project, which endeavors to create the most detailed compendium of human genetic variation (Kaiser 2008), and a recently proposed 1000 Genomes Project focused on *D. melanogaster*. Additional proposals, such as the 1001 Genomes Project for *Arabidopsis thaliana*, the workhorse of plant genetics, tout the utility of combining quantitative trait locus mapping, population genomic approaches, and genome-wide association studies (Weigel and Mott 2009).

Yet, population genomic descriptions of nucleotide variation, while a tremendously exciting advance, represent a first step. Importantly, methods for the functional validation of candidate genes and mutations identified by genome-wide approaches are needed (see previous discussion), and to assign any adaptive function at all, one must be certain that fitness is carefully measured in nature and the selective agent is established (see the following section).

Phenomics

Compared to our knowledge of genomes, our knowledge of phenotypes remains cursory. Part of the explanation for this imbalance is certainly that phenotype space is vastly more expansive than genotype space (Houle 2010); in other words, genotypes are more easily defined. However, the ability to make the link between genotype and phenotype rests, in part, on our capacity to measure phenotypes accurately and consistently. Ideally, we want to be able to measure objectively the same (homologous) trait in many individuals as well as many traits in a few individuals. While our textbook example of a phenotype is often height or weight, it can be useful to deconstruct these complex phenotypes into more precise traits. At the extreme, functional genomic outputs (e.g., mRNA or protein expression levels) may serve as readily measured quantitative phenotypes (i.e., endo-phenotype) and also can be used to assess genotype–phenotype associations. At the organismal level, high-throughput morphometrics (and accompanying databases) may be another boon, as such approaches allow for the linking of macroevolution-level collections of fossils, microevolution-level data collected from natural populations, and experimental-level altered morphologies of mutants. For example, commercial facilities now specialize in phenomics and are well equipped to do high-throughput screens for rare mutants (e.g., greenhouses that can raise thousands of seedlings) or, alternatively, multi-phenotype screens for a mutant strain (e.g., hundreds of behavioral assays run on a few mutant/transgenic animals). These high-throughput screens are revolutionizing the scale at which geneticists and neurobiologists design and implement experiments (e.g., Tecott and Nestler 2004). While evolutionary biologists have yet to fully capitalize on these large-scale approaches, it is easy to imagine that such an approach can be adapted to address questions about evolutionary diversification and constraint.

Fitness in the Wild

Measuring phenotypes in controlled laboratory environments has many advantages, but what we really want to know is how morphological, physiological, and behavioral variation translates to fitness differences in the wild. As a first step, we would like to carefully characterize changes in the environment at both the micro- and macro- spatial and temporal scales. Long term ecological research (LTER) studies have been set up to study changes in ecological processes over time, yet these studies rarely integrate

evolution. Large-scale terrestrial environmental data may soon become available through the National Ecological Observatory Network (NEON; www.neoninc.org). While still in its infancy, NEON holds the promise of providing both environmental information for terrestrial biomes and a model for characterizing additional biomes. However, describing environmental variation through time and space will not alone provide a comprehensive picture because organisms evolve in response to both abiotic and biotic factors, including those that pertain to each species individually (see Davis et al., Commentary 6). Thus, we also want to know how individuals are interacting within an environment, especially with other species.

While developing high-throughput ecological experiments is not straightforward, some initial steps are already being taken to monitor remotely the movements, survival, and reproduction of hundreds of both plant and animal individuals. New miniaturized transmitters exist to record continuously the activity and performance of organisms in nature. For example, the International Cooperation for Animal Research Using Space (ICARUS) initiative aims to establish a remote sensing platform that can track even small animals globally, enabling observations and experiments over large spatial scales (Wikelski et al. 2007). Future instrumentation to measure not only movement, but also physiological state and whole-organism performance and behavior, may capitalize on advances in microfluidics (Whitesides 2006) and imaging (Bimber 2006), which already are proving powerful in laboratory settings. Thus, organismal movement, physiology, and performance may soon be tracked remotely and recorded automatically in large databases. As with increasing amounts of -omic data, a major challenge in generating such organismal and environmental data is an effective cyberinfrastructure. One model, the iPlant Collaborative (iPlant; www.iplantcollaborative.org) has been developed by plant biologists. Done right, such programs will encourage communication among disciplines and the reuse of data models, file formats, application software, and algorithms, while fostering cross-disciplinary exchange of ideas.

However, understanding changes in the environment and even the movement and interactions of organisms within the environment is still not quite enough. In addition, we want to measure fitness, or at least components of fitness, in nature. But how? Measuring fitness in the wild for large numbers of individuals through time represents one of the most important and largest challenges to evolutionary biologists. No matter how advanced technology and tools become, it is difficult to imagine being able to replace fieldwork and the study of natural history, both at a practical and philosophical level.

The Genotype–Phenotype Synthesis

Darwin lived in a very opportune time, a time of exploration, when new specimens were continuously arriving from around the world and new

geological data were being amassed. At present, evolutionary biologists are arguably in a comparable position, as new molecular data are being generated at an unimaginable rate, and the possibility of gathering fine-scale environmental data is on the horizon. However, the real question, of course, is how will these data change the way we think about evolutionary biology, if at all? I focused previously on the technological advances that make it possible to imagine (perhaps optimistically) that we will have an ever increasing ability to link changes in the genome to changes in phenotype and/or in the environment. In the following, I will focus on four general areas (from among many), in which Darwin was clearly interested, but for which he could never have anticipated even our current depth of knowledge. These four areas also represent topics in which connecting genotype and phenotype may be especially illuminating in our quest to more fully understand the evolutionary process.

The Mechanistic Basis of Adaptation

Darwin's magnum opus was "one long argument" for the ultimate cause of adaptive change, that is, natural selection. But, even Darwin was curious about the mechanistic basis of phenotypic change—how are traits encoded and passed on from one generation to the next? Perhaps, it was a propitious augury then that Darwin's last publication arose from a collaboration with a young British naturalist by the name of Walter Drawbridge Crick (Ridley 2004), the grandfather of Francis Crick, who would contribute to the discovery of the precise mechanism of inheritance, by deciphering the three-dimensional structure of DNA.

Technological advances have had a fundamental impact on the way we study adaptation. In Lauder and Rose's 1996 book entitled *Adaptation*, only a single chapter was devoted to molecular data, whereas a decade later we are hard pressed to find new advances in the study of phenotypic adaptation without some molecular genetic contribution. Combinations of genomic and transcriptomic approaches are being applied to variation in nature, and will allow us to address fundamental questions about the adaptive process, as previously discussed. Moreover, our increasing knowledge of the genetics of adaptation in wild populations is, in turn, expanding our knowledge about natural selection (Schluter et al., in press). More specifically, genes contain information about the form, strength, timing, history, and (sometimes) the agent of natural selection (e.g., Fitzpatrick et al. 2007; Barrett et al. 2008; Linnen et al. 2009). Thus, I would suggest that some of the most enlightening future studies will combine studies of proximate (i.e., molecular and developmental mechanisms) and ultimate (i.e., selective mechanism) causes of evolutionary change, as both approaches are mutually enlightening.

While some hard-earned progress has been made identifying a handful of genes underlying adaptive variation, new genomic technologies and approaches are increasing the rate at which such genes are being identified and functionally verified. The taxonomic breath of such investigations is also expanding, as is, most importantly, the complexity of traits that are being dissected. As one example, the role of genetic constraint due to pleiotropy and the importance of epistasis in adaptation are difficult to ascertain until we have enough data, both genetic and phenotypic, to test these hypotheses in a statistically rigorous manner (see G. Wagner, Chapter 8). Thus, the lessons we have learned so far by studying a few simple traits in a handful of species may not be fully representative of adaptive change in general, and consequently, a comprehensive view of the adaptive process is yet to emerge.

Therefore, I predict (and hope) that "gene hunting" studies do not end simply with the identification of a gene, but rather continue on to determine the consequences of mutations for gene function (Dean and Thornton 2007), to understand *how* changes in gene function and regulation through development produce phenotypic change (Carroll et al. 2001), and most importantly, perhaps, to truly understand how genetic variation translates to fitness differences in the wild (Nielsen 2009; Linnen and Hoekstra, in press). With this information in hand, we should be well positioned to not only understand what happened in the past, but also start to make predictions about the genetic and phenotypic responses to known environmental change in the future.

From Instinct to Neurobiology

Darwin devoted an entire chapter of *The Origin of Species* to behavior. He concludes that because "no one will dispute that instincts are of the highest importance to each animal" (p. 243) and that behaviors can be heritable, that they can evolve by natural selection in the same way as do morphological characters. Indeed, while much research has focused on morphological adaptation, one classic view posits that change in animals' morphology is often preceded by that in behavior: "a shift into a new niche or adaptive zone is, almost without exception, initiated by a change in behavior" (Mayr 1963: 604). An understanding of the ultimate causes of behavioral variation has been the subject of study first made popular by the classic ethologists (e.g., Niko Tinbergen), who were interested in understanding both the causes and consequences of behavioral variation across many organisms in their natural environments (see Kokko and Jennions, Chapter 12).

However, despite its importance, we still know very little about the proximate evolutionary mechanisms that give rise to behavioral diversity found in nature. Thus, we remain largely ignorant of how behavior evolves and are left with many fundamental questions unanswered. For example: What are the relative contributions of genetic and environmental effects (or learning) to behavioral differences? What are the genetic changes that underlie the differences in behavior found both within and between species? Do these genetic changes act early in development to alter neural circuitry, or does the circuitry remain constant and changes in physiology (e.g.,

neurotransmitters) underlie behavioral variation? How do such changes happen at a mechanistic level? Are "behavior genes" specific to behavior? How has the extraordinary complexity of human culture evolved from simpler behavior repertoires of our primate ancestors (see Richerson and Boyd, Chapter 20)? The connections among genes, neural circuitry, and the evolution of complex and adaptive behaviors remain a major frontier in biology.

The study of behavior, unfortunately, shares many of the same obstacles as the study of the genetic basis of morphological traits but also introduces a host of new ones. Many behaviors have low heritability, and some are culturally inherited or have a learned component. Behavior is particularly prone to environmental effects; at the extreme, some behaviors require an environmental stimulus (Robinson et al. 2008). Finally, many of the most interesting behaviors (e.g., behaviors that seem to have a clear fitness consequence in the wild) occur in non-model organisms, which have been the focus of field-based behavioral ecologists (see Kokko and Jennions, Chapter 12). However, optimistically, some of these obstacles can be overcome by an increase in experimental scale, automated phenotyping in controlled environments, and the application of genomic technologies to non-model species. In addition, just as studies by developmental systems biologists are helping to elucidate the link between genes and morphology, new large-scale and high-throughput approaches being adopted by neurobiologists are setting the stage to link genes and behavior.

Neurobiologists, too, have been bitten with the genomics bug and have recently spawned a field of study entitled neurogenomics, which is the study of how the genome contributes to the evolution, development, structure, and function of the nervous system (Boguski and Jones 2004). This approach is already being adopted by those studying behavior in diverse species. For example, the songbird neurogenomics initiative (SoNG) aims to develop technical resources to study gene–brain–behavior relationships using song and songbirds as model systems. Already, transcriptome surveys and large-scale neuroanatomical mapping of gene expression are underway, and may define a new functional anatomy of the brain. Neurogenomicists may, in fact, provide the bridge between reductionist (behavioral genetics) and holist (behavioral ecology) approaches to the study of nervous system evolution, analogous to developmental genomics in morphology.

While elucidating the mechanistic details of the behavioral evolution is still in its infancy, I predict we will soon be able to draw a more complete picture of how many behaviors evolve, from the ultimate forces driving behavioral evolution down to its molecular details. This prediction assumes that with improving technologies, genes that underlie behavioral variation will become increasingly easy to identify, and that behavioral evolution (like morphology) may rely on changes to a common genetic toolkit (i.e., a set of genes shared by all animals), so that unraveling the genetic basis of behavior in one species will earn us some predictive power when studying other species.

Genomic Approaches to Studying The Origin of Species

In the opening paragraph of *The Origin of Species*, Darwin suggests his tome will "throw some light on the origin of species—that mystery of mysteries, as it has been called by one of our greatest philosophers" (Darwin 1859: 1). Despite some modest progress in understanding how new species originate (Coyne and Orr 2004; see Harrison, Chapter 13), surprises still abound, and thus, future work in speciation may demand considerable changes in our views, not just minor modifications (Orr et al. 2007).

How species originate remains one of the most fundamental questions in evolutionary biology and thus is likely to be a subject of much growth in the coming years. Perhaps the most exciting progress will be made by using genomic approaches to elucidate the processes and molecular basis underlying speciation. In particular, we would like to know: What are the forces that drive species formation, and in particular, what is the role of natural selection? How often can differentiation occur in the face of continuous or intermittent gene flow? What mechanisms of reproductive isolation are most important and which ones evolve first? What specific genes are involved in preventing gene flow between incipient species? And how do these patterns vary across organisms with different ecology, mating systems, reproductive biology, and sex chromosomes?

Genomic approaches offer some new insights into these questions in two ways. First, technical advances are increasing the rate at which specific mutations and genes responsible for reproductive isolation are being identified. Second, genomic surveys of natural populations have allowed researchers to investigate barriers to gene flow, genomic interactions, and the genetic permeability of species boundaries in the wild (Noor and Feder 2006). While much of the progress involves increases in scale and speed, as these approaches are extended to additional species, general patterns and rules might emerge. In doing so, we are no longer asking questions about whether specific types of genes and types of processes occur, but rather what their relative frequency and importance are for speciation in general (Schluter 2009).

If we are truly to understand how new species originate, two goals must be accomplished. First, the speciation process can be fully explained only if we identify the heritable underpinnings of species formation and the forces responsible for their origins. Thus, understanding genomic patterns of and specific genetic changes driving speciation likely will change the way we understand the fundamentals of diversification. Second, it is equally important to recognize, that with the growing appreciation for the role of selection in species formation, it is impossible to ignore the role of ecology (Sobel et al. 2010; see McPeek, Commentary 3). Thus, a combination of both

genomic and ecological approaches will be the key to solving Darwin's "mystery of mysteries."

Molecular Fossils: Reconstructing Evolution's Path

Evolutionary biology is in large part about reconstructing the past. But, in 1859, there was little data from the past; Darwin spends an entire chapter in *The Origin of Species* touting the importance of ancient organisms as evidence for descent with modification (Darwin 1859: Chapter 10) and another (Darwin 1859: Chapter 9) lamenting the imperfection of the fossil record. It was a great boon to Darwin's theory when, just 3 years after its publication, the canonical example of a transitional fossil, *Archaeopteryx*, was unearthed in Germany. Since then, the vastly greater and still rapidly growing knowledge of the fossil record has affected the way we think about morphological change (see P. Wagner, Chapter 17) and diversification (see Foote, Chapter 18) through time, perhaps most spectacularly in hominids (see White, Chapter 19).

With novel phylogenetic approaches used to reconstruct ancestral genes and genomes and an increasing availability of ancient DNA sequences, we now have a complementary approach to deducing past events that will stand alongside discoveries from the fossil record. First and foremost, phylogenetics (i.e., the reconstruction of the relationships among organisms) provides a means of inferring the history and pattern of past events (e.g., phenotypic character evolution). More specifically, the ancestral reconstruction of genes (Liberles 2007) and now even large stretches of genomes (Blanchette et al. 2004) provide an indirect peek into the evolutionary history of molecular function and genome evolution. Second, the combination of successes in identifying mutations that affect phenotype in extant populations and improved technologies that provide direct access to ancient genomes also allow a rare glimpse into the biology of organisms that are now long extinct. Together, these approaches have provided an increasing ability to recreate ancient phenotypes with genetic precision. For example, using phylogenetic methods, ancient opsin gene sequences that provide insight into dinosaur vision have been reconstructed (Chang et al. 2002). Using sequencing of candidate genes from ancient DNA samples, we now know that Neanderthals were lactose-intolerant and some were red headed (Lalueza-Fox et al. 2007), while some mammoths were likely blonde (Römpler et al. 2006). Now imagine reconstructing or sequencing whole genomes and being able to predict the morphology, physiology, or even behavior of many ancient creatures!

Well, we need not wait long, as these data are already rolling in. Applying phylogenetic methods to the complete genome sequences of many extant mammals, researchers have begun to develop tools to reconstruct the genome (at the nucleotide-level) of the *Ur*-mammal (Paten et al. 2008). For example, the first complete ancient human genome of a paleo-Eskimo, some 4000 years old, has already been published (Rasmussen et al. 2010).

(Note the genome sequencing was completed in just two and a half months at a cost of about $500,000.) The authors deduced that this individual was a male member of the Arctic Saqqaq, the first known culture to settle in Greenland. He likely had brown eyes, non-light skin, thick dark hair (although at risk for baldness), shovel-graded front teeth, dry (as opposed to wet) ear wax, and probably had a metabolism and body mass adapted to cold climate. This is just the beginning. Over four billion base pairs of Neanderthal genome have been sequenced (Green et al. 2010). But, even preliminary genomic results have profound implications; for example, Neanderthals might have shared some basic language capabilities with modern humans (Green et al. 2006; Noonan et al. 2008). Such studies clearly have and will continue to shed light on the phenotypic traits, genetic origin, and biological relationship to present-day populations of now extinct individuals, populations, and species.

In addition to reconstructing past organisms, we are now in a position to reconstruct Earth's history with increasing precision. Advances in biogeochemistry and isotope techniques (e.g., Robert and Chaussideon 2006; Shen et al. 2001) have revealed ancient and fine-scale changes in the environment (e.g., temperature and chemical composition)—changes, though subtle relative to large scale fluctuations over Earth's long history, that nonetheless have had dramatic effects on organisms, their physiology, lifestyles, and distributions. But this is a two-way street. Recently, geneticists have been able to provide experimental results to predict the environments of ancient life; for example, by sequencing genes from extant species, then reconstructing ancestral proteins and testing their thermostability, we can infer fluctuations in the Earth's temperature (Gaucher et al. 2008). Thus, convergent predictions from geology and biology together can be used to reliably track changes in the Earth's environment and concomitant change in the organisms living in those environments over time.

Conclusions

The future of evolutionary biology, like that of any science, is difficult to predict. At present, what is clear is that major advances are being made at a breath-taking pace. Arguably, many of these advances are being made by using interdisciplinary thinking, by taking advantage of large-scale discovery-based science, and by working at scales previously unimaginable. While I have focused here on only one slice of evolutionary biology, analogies can certainly be found in other sub-disciplines in the field.

Historically, many of the major fundamental advances in evolutionary biology have been associated with: (1) the unification of fields, such as ecology and evolution (see Agrawal et al., Chapter 10; Berenbaum and Schuler, Chapter 11; McPeek Commentary 3) or microevolution and macroevolution (see P. Wagner, Chapter 17), (2) the reconciliation of points of view, (e.g., Futuyma, Chapter 1), (3) the elaboration of new approaches such as

phylogenetics (see Hillis, Chapter 16), coalescent theory (see Wakeley, Chapter 5), genomics (see Kolaczkowski and Kern, Chapter 6), and (4) the incorporation of new fields, such as genetics (see Zhang, Chapter 4) and most recently, developmental biology (see Wray, Chapter 9). Such interdisciplinary research efforts can reap large benefits (Wake 2008). For example, paleontology and developmental genetics, two seemingly disparate fields, can illuminate each other—fossils may suggest developmental processes in basal forms, and developmental studies can aid in interpretation of the characters of extinct taxa, such as the gain of limbs of tetrapod ancestors (Shubin et al. 2009) or the loss of fins in a fish (Chan et al. 2010). Thus, integrative research that crosses traditional boundaries will undoubtedly continue to push our understanding of evolutionary biology further.

Because evolutionary biologists pride themselves on being hypothesis-driven scientists, the practice of systematically collecting data, even before knowing all of the precise ways it may be used, has often been frowned upon. However, it is important to recognize the countless evolutionary studies that already take advantage of existing comparative data, including life history characters, physiological tolerances, behavior, habitat associations, diet, geographic distribution, and morphologies (Futuyma 1998). In almost all cases, these data were collected by systematists or other biologists without any anticipation of their future use in testing new hypotheses. This descriptive natural history at the organismal level has been invaluable and now is being replayed at the genomic level. In both cases, data repositories are built and can be used for purposes not yet envisioned. Thus, researchers pursuing hypothesis-driven versus discovery-driven approaches have (or soon may) come to see each other as allies rather than antagonists (Boguski and Jones 2004). At the very least, -omic data, when viewed in the context of an appropriate evolutionary model for rigorous statistical testing, can be extremely powerful. However, large data sets will never replace imaginative hypotheses—the engines of scientific progress.

What seems increasingly novel as we move forward is the scale at which we can ask questions. We are not likely to be limited by genomic data, and both large-scale phenotypic and environmental data acquisition are not far behind. We may expect that in 50 years major progress in understanding the evolutionary process, both the proximate and ultimate mechanisms of evolutionary change, will come from a complementary (if not collaborative) effort among population genomicists, systems biologists (and/or developmental biologists, neurogenomicists), and organismal biologists. Like the happy marriages of first the Mendelians and biometricians and later of the laboratory geneticists and field naturalists, the union of genomic and organismal biology will continue to advance our understanding of evolution in years to come.

A Note

The responses by the participants of the Darwin 2009 Workshop to my presentation were surprisingly bimodal. Some said that my predictions were too cautious—that we would "know it all" in 10 years, rather than 50 or 100. Others were far less optimistic, claiming that organisms were too complicated to dissect, and thus questioned if we would *ever* be able to know it all. Perhaps not surprisingly, the type of response was highly associated with the scientists' field of study: geneticists (i.e., reductionists) tended to be more optimistic, whereas organismal biologists (i.e., holists) were generally more cautious. I look forward to looking back to see who, if either, was right.

Acknowledgments

I wish to sincerely thank the organizers of the Darwin 2009 Workshop and resulting volume, in particular Mike Bell, Walt Eanes, Doug Futuyma, and Jeff Levinton. Never before have I been given such a flattering, yet utterly daunting, task. I also wish to thank the meeting participants for engaging and stimulating discussions. Additional discussions with and comments from Andrew Berry, David Hillis, and Jonathan Losos were critical to the development of many of the ideas presented here.

Literature Cited

Abzhanov A., C. G. Extavour, A. Groover, and 4 others. 2008. Are we there yet? Tracking the development of new model systems. *Trends Genet.* 24: 353–360.

Barrett, R. D. H., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. *Science* 322: 255–257.

Barrett, R. D. H., S. M. Rogers, and D. Schluter. 2009. Environment specific pleiotropy facilitates divergence at the *Ectodysplasin* locus in threespine stickleback. *Evolution* 63: 2831–2837.

Bell, M. A. 2009. Implications of fossil threespine stickleback for Darwinian gradualism. *J. Fish Biol.* 75: 1997–1999.

Bell, M. A., W. E. Aguirre, and N. J. Buck. 2004. Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution* 58: 814–824.

Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, and 193 others. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53–59.

Bimber, O. 2006. Computational photography—the next big step. *Computer* 39: 28–29.

Blanchette, M., E. D. Green, W. Miller, and 1 other. 2004. Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Res.* 14: 2412–2423.

Boguski, M. S. and A. R. Jones. 2004. Neurogenomics: At the intersection of neurobiology and genome sciences. *Nat. Neurosci.* 7: 429–433.

Carroll, S. B., J. K. Grenier, and S. D. Wetherbee. 2001. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell, Malden.
Chan, F. Y., 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* 324: 302–305.
Chang, B. S. W., K. Jönsson, M. A. Kazmi, and 2 others. 2002. Recreating a functional ancestral archosaur visual pigment. *Mol. Biol. Evol.* 19: 1483–1489.
Colosimo, P., K. Hosemann, S. Balabhadra, and 8 others. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science* 307: 1928–1933.
Coyne, J. A. and H. A. Orr. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.
Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.
Dean, A. M. and J. W. Thornton. 2007. Mechanistic approaches to the study of evolution: The functional synthesis. *Nat. Rev. Genet.* 8: 675 688.
Dobzhanksy, T. 1937. *Genetics and the Origin of Species*. Columbia University Press, New York.
Drmanac, R., A. B. Sparks, M. J. Callow, and 62 others. 2010. Human genome sequencing using unchained base read on self-assembling DNA nanoarrays. *Science* 327: 78–81.
Fitzpatrick, M. J., E. Feder, L. Rowe, and 1 other. 2007. Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. *Nature* 447: 210–213.
Fleischmann R., M. Adams, O. White, and 7 others. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science* 269: 496–512.
Flicek P., B. L. Aken, B. Ballester, and 54 others. 2010. Ensembl's 10th year. *Nucl. Acids Res.* 38: D557–562.
Futuyma, D. J. 1998. Wherefore and whither the naturalist? *Am. Nat.* 151: 1–6.
Gaucher, E. A., S. Govindarajan, and O. K. Ganesh. 2008. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451: 704–708.
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.
Green, R. E., J. Krause, S. E. Ptak, and 8 others. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* 444: 330–336.
Green, R. E., J. Krause, A. W. Briggs, and 53 others. 2010. A draft sequence of the Neanderthal genome. *Science* 328: 710–722
Haldane, J. B. S. 1924. A mathematical theory of natural and artificial selection. Part I. *Trans. Camb. Phil. Soc.* 23: 19–41.
Hoekstra, H. E., R. J. Hirschmann, R. J. Bundey, and 2 others. 2006. A single amino acid mutation contributes to adaptive color pattern in beach mice. *Science* 313: 101–104.
Houle, D. 2010. Numbering the hairs on our heads: The shared challenge and promise of phenomics. *Proc. Natl. Acad. Sci. USA* 107: 1793–1799.
Hunt, G., M. A. Bell, and M. P. Travis. 2008. Evolution toward a new adaptive optimum: Phenotypic evolution in a fossil stickleback lineage. *Evolution* 62: 700–710.
International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* 437: 1299–1320.
International Human Genome Sequencing Consortium (IHGSC). 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.

International Human Genome Sequencing Consortium (IHGSC). 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431: 931–945.
Jenner, R. A. and M. A. Wills. 2007. The choice of model organisms in evo-devo. *Nat. Rev. Genet.* 8: 311–319.
Jensen, J., Y. Kim, V. DuMont, and 2 others. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* 170: 1401–1410.
Kaiser, J. 2008. DNA sequencing: A plan to capture human diversity in 1000 genomes. *Science* 319: 395.
Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217: 624–626.
Kitano, H. 2002. Systems biology: A brief overview. *Science* 295: 1662–1664.
Kitano, J., D. I. Bolnick, D. A. Beauchamp, and 4 others. 2008. Reverse evolution of armor plates in the threespine stickleback. *Curr. Biol.* 18: 769–774.
Krebs, H. A. 1975. The August Krogh Principle: "For many problems there is an animal on which it can be most conveniently studied." *J. Exp. Zool.* 194: 221–226.
Lalueza-Fox, C., H. Römpler, D. Caramelli, and 14 others. 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318: 1453–1455.
Lauder, G. and M. Rose. 1996. *Adaptation*. Academic Press, San Diego.
Levy, S., G. Sutton, P. C. Ng, and 30 others. 2007. The diploid genome sequence of an individual human. *PLoS Biol.* 5: e254.
Lewontin, R. C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.
Lewontin, R. C. and J. L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54: 595–609.
Li, R., W. Fan, G. Tian, and 122 others. 2010. The sequence and *de novo* assembly of the giant panda genome. *Nature* 463: 311–317.
Liberles, D. A. 2007. *Ancestral Sequence Reconstruction*. Oxford University Press, Oxford.
Linnen, C. R. and H. E. Hoekstra. In press. Measuring natural selection on genotypes and phenotypes in the wild. In *The Molecular Landscape*, Cold Spring Harbor Symposium Volume, Cold Spring Harbor Press, New York.
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.
Marchinko, K. B. 2009. Predation's role in repeated phenotypic and genetic divergence of armor in threespine stickleback. *Evolution* 63: 127–138.
Mayr, E. 1963. *Animal Species and Evolution*. Harvard University Press, Cambridge, MA.
McDonald, J. H. and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351: 652–654.
Mullen, L. M. and H. E. Hoekstra. 2008. Natural selection along an environmental gradient: A classic cline in mouse pigmentation. *Cold Spr. Harb. Symp. Evolution* 62: 1555.
Mullen, L. M., S. N. Vignieri, J. A. Gore, and 1 other. 2009. Adaptive basis of geographic variation: Genetic, phenotypic and environmental differentiation among beach mouse populations. *Cold Spr. Harb. Symp. Proc. Roy. Soc. B* 276: 3809.

Nielsen, R. 2009. Adaptationism—30 years after Gould and Lewontin. *Evolution* 63: 2487–2490.
Nielsen, R., C. Bustamante, A. G. Clark, and 10 others. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3: e170.
Noble, D. 2006. *The Music of Life: Biology Beyond the Genome.* Oxford University Press, Oxford.
Noonan, J. P., G. Coop, S. Kudaravalli, and 8 others. 2008. Sequencing and analysis of Neanderthal genomic DNA. *Science* 314: 1113–1118.
Noor, M. A. F. and J. L. Feder. 2006. Speciation genomics: Evolving approaches. *Nat. Rev. Genet.* 7: 851–861.
Orr, H. A., J. P. Masly, and N. Phadnis. 2007. Speciation in *Drosophila*: From phenotypes to molecules. *J. Hered.* 98: 103–110.
Paten, B., J. Herrero, S. Fitzgerald, and 4 others. 2008. Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.* 18: 1829–1843.
Pfeiffer, T., O. S. Soyer, and S. Bonhoeffer. 2005. The evolution of connectivity in metabolic networks. *PLoS Biol.* 3: e228.
Pushkarev, D., N. F. Neff, and S. R. Quake. 2009. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* 27: 847–850.
Rasmussen, M., Y. Li, S. Lindgreen, and 50 others. 2010. Ancient human genome sequence of an extinct Palaeo-eskimo. *Nature* 463: 757–762.
Ridley, M. 2004. Crick and Darwin's shared publication in *Nature*. *Nature* 431: 244.
Robert, F. and M. Chaussidon. 2006. A palaeotemperature curve for the Precambrian oceans based on silicon isotopes in cherts. *Nature* 443: 969–972.
Robinson, G. E., J. A. Banks, D. K. Padilla, and 18 others. In press. Empowering 21st century biology. *BioScience.*
Robinson, G. E., R. D. Fernald, and D. F. Clayton. 2008. Genes and social behavior. *Science* 322: 896–900.
Römpler, H., N. Rohland, C. Lalueza-Fox, and 6 others. 2006. Nuclear gene indicates coat-color polymorphism in mammoths. *Science* 313: 62.
Sanger, F., G. M. Air, B. G. Barrell, and 6 others. 1977. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265: 687–695.
Sauer, U., M. Heinemann, and N. Zamboni. 2007. Getting closer to the whole picture. *Science* 316: 550–551.
Schluter, D. 2009. Evidence for ecological speciation and its alternative. *Science* 323: 737–741.
Schluter, D., K. B. Marchinko, R. D. H. Barrett, and 1 other. In press. Natural selection and the genetics of adaptation in threespine stickleback. *Phil. Trans. Roy. Soc. B.*
Schwenk, K., D. K. Padilla, G. S. Bakken, and 1 other. 2009. Grand challenges in organismal biology. *Integr. Comp. Biol.* 49: 7–14.
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.
Shen, Y., R. Buick, and D. E. Canfield. 2001. Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* 410: 77–81.
Shubin, N., C. Tabin, and S. B. Carroll. 2009. Deep homology and the origins of evolutionary novelty. *Nature* 457: 818–823.
Sobel, J. M., G. F. Chen, L. R. Watt, and 1 other. 2010. The biology of speciation. *Evolution* 64: 295–315.
Steiner, C. C., J. N. Weber, and H. F. Hoekstra. 2007. Adaptive variation in beach mice caused by two interacting pigmentation genes. *PLoS Biol.* 5: e219.

Tecott, L. H. and E. J. Nestler. 2004. Neurobehavioral assessment in the information age. *Nat. Neurosci.* 7: 462–466.
Venter, J. C., M. D. Adams, E. W. Myers, and 122 others 2001. The sequence of the human genome. *Science* 291: 1304–1351.
Vignieri, S. N. and H. E. Hoekstra. In press. The selective advantage of cryptic coloration in mice. *Evolution.*
Wake, M. H. 2008. Integrative biology: Science for the 21st century. *BioScience* 58: 349–353.
Watson, J. D. and F. H. C. Crick. 1953. A structure for deoxyribose nucleic acid. *Nature* 171: 737–738.
Watt, W. B. 1977. Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: Biochemical and population aspects. *Genetics* 87: 177–194.
Weigel, D. and R. Mott. 2009. The 1001 Genome Project for *Arabidopsis thaliana*. *Genome Biol.* 10: 107–111.
Wheeler, D. A., M. Srinivasan, M. Egholm, and 24 others. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452: 872–876.
Whitesides, G. M. 2006. The origins and the future of microfluidics. *Nature* 442: 368–373.
Wikelski, M., R. W. Kays, N. J. Kasdin, and 3 others. 2007. Going wild: What a global small-animal tracking system could do for experimental biologists. *J. Exp. Biol.* 210: 181–186.
Zhu, H. and M. Snyder. 2002. "Omic" approaches for unraveling signaling networks. *Curr. Opin. Cell Biol.* 14: 173–179.