The Journal of Experimental Biology 210, 1653-1660
Published by The Company of Biologists 2007
doi:10.1242/jeb.02725

Evolvability of physiological and biochemical traits: evolutionary mechanisms including and beyond single-nucleotide mutation

Martin E. Feder

Department of Organismal Biology and Anatomy and The Committees on Evolutionary Biology, Genetics, and Molecular Medicine, The University of Chicago, 1027 E. 57th Street, Chicago, IL 60637, USA

e-mail: m-feder@uchicago.edu

Accepted 17 January 2007

Summary

A longstanding challenge for biologists has been to explain not just how organisms are adapted to diverse environments, but how these adaptations arise. Although natural selection is clearly sufficient to act on heritable variation, is this heritable variation sufficient to yield complex adaptations and how does this variation itself arise? Much prior focus has been on mutation of single nucleotides in genes. This process is common and can have dramatic phenotypes, but could be limited in its ability to culminate in complex adaptations for two kinds of reasons: (i) because natural selection is powerful, it can purge genetic variation, and (ii) evolutionary transition from the absence to the presence of a complex adaptation seemingly requires multiple mutations at the right place and time and in the right sequence, with each intermediate stage having increased overall fitness; this seems highly improbable. Because the networks that organisms comprise are hierarchical and redundant and have modular structure, however, single-nucleotide mutations can have large and tolerable impacts. Diverse mechanisms, collectively evolutionary capacitors, can shield genetic variation from the purgative of selection. These features can enable evolution to proceed *via* single-nucleotide mutation. Importantly, single-nucleotide mutation usually only modifies existing genes rather than creating new ones, and numerous other mechanisms eclipse single-nucleotide mutation in creating genetic variation. These include gene duplication (both segmental and whole-genome), lateral gene transfer, hybridization, mobile genetic elements and symbiosis. Other processes can scramble and reassemble nucleotide sequence. The mechanisms beyond single-gene mutation offer considerable promise in detailing the evolution of complex physiological and biochemical traits, and have already done so for several morphological traits.

Glossary available online at
http://jeb.biologists.org/cgi/content/full/210/9/1653/DC1

Key words: evolution, mutation, adaptation, natural selection, evolvability, gene duplication, transposable element.

Introduction

Despite the vulnerability and limitations of life processes, organisms exploit seemingly inhospitable environments and perform remarkable feats. The biochemical and physiological mechanisms that underlie these extraordinary capacities have been a major focus of experimental biologists, increasingly using the toolkits of molecular biology and genome-enabled research. Indeed, such research is increasingly revealing the inner workings of these mechanisms, as many of the other contributions to this review volume demonstrate. But how have these mechanisms themselves come into being?

Many biologists, if pressed, invoke ‘adaptation by natural selection’ as an answer to this question. For some biologists, the answer stops there. How else could the extraordinary biochemical–physiological traits that so obviously enable organisms to function in diverse environments and are so clearly ‘tuned’ to each organism’s environment have arisen except by natural selection (Fig. 1)? Consolidating this assessment are numerous examples of evolutionary convergence and parallelism, wherein functionally equivalent or identical mechanisms have arisen multiple times in response to similar selective pressures.

Other biologists will go further, and emphasize consistency with the fundamental principles of natural selection: (a) individuals in a population vary in genotype; (b) individuals in a population consequently vary in phenotype (including biochemical and physiological traits); (c) individuals in a population consequently vary in fitness; i.e. reproductive

output. Hence genotypes that encode higher fitness will come to predominate in a population. Documentation of points (a) and (b) is now abundant both within and among populations, species and higher taxa, and documentation of point (c) in natural populations is growing (Endler, 1986; Feder et al., 2000; Feder and Mitchell-Olds, 2003; Van Straalen and Roelofs, 2006). Four examples of diverse vintage are considered below.

(1) In human populations (Ingram, 1963; Williams et al., 2005), mutation of nucleotide 17 from A to G in the sequence encoding the beta-subunits of the Hemoglobin A protein converts a glutamine to a valine. As a result, hemoglobins form fibrils *via* interactions due to the valine. Consequently, the shape of the hemoglobin-containing erythrocyte changes. Individuals homozygous for the mutation develop sickle-cell disease, and have lower fitness. Individuals heterozygous for the mutation, by contrast, have enhanced resistance to malaria (*Plasmodium* infection) by some as-yet-unknown mechanism, which might include an increased immune response, filtering of infected erythrocytes by the spleen, and/or outright inhospitability to the *Plasmodium*. This is a classic example of a balanced polymorphism, wherein the advantages to the heterozygote offset the disadvantages to the homozygote mutant.

(2) The fish *Fundulus heteroclitus* inhabits brackish water habitats along the eastern coast of North America, and so experiences enormous variation in temperature. A key component of the adaptation to this thermal gradient has been evolution of the *ldh-b* gene, encoding one component of the enzyme lactate dehydrogenase (LDH). Specifically, northern populations express more of this gene's product than southern populations, which may compensate for the cooler average temperatures of the northern populations. Evolved differences between alleles are both in protein function (i.e. in coding sequence) and in gene expression (i.e. in the proximal promoter region). Indeed, as initially reported (Segal et al., 1996), the north–south variation in proximal promoter sequence is sufficient for corresponding differences in gene expression.

(3) Beach mice (*Peromyscus*) inhabit beaches with diverse sand coloration, and mismatches of fur and sand color enhance predation on mice (Hoekstra et al., 2006). Thus mouse populations on dark substrates have repeatedly evolved melanic coloration. In one case, a single-nucleotide variant (changing arginine 65 to cysteine) of the melanocortin-1-receptor gene, *Mc1r*, is strongly associated with melanism; this gene's product is key to the production of dark-pigmented fur. In other cases, no such genetic variation is evident, suggesting that melanism has arisen through at least two different genetic routes.

(4) Brewer's yeast (*Saccharomyces cerevisiae*) makes ethanol *via* the alcohol dehydrogenase enzyme ADH1, which it then feeds into the tricarboxylic acid (TCA) cycle *via* the enzyme ADH2. Thomson et al. (Thomson et al., 2005) have theorized that ancestral yeast had a single ADH that, much like vertebrate LDH and lactate, recycled NADH for use in glycolysis by producing, in this case, ethanol; ethanol could readily be lost to the environment before it poisoned the yeast. When fleshy fruits arose and *Saccharomyces* colonized them, yeast then evolved enhanced biosynthetic capacity for ethanol (and tolerance for it), which enabled yeast to triumph over microbial competitors in the fruit. If so ADH2, which is specialized for ethanol catabolism, would only have been needed *post hoc*. Using maximium likelihood techniques (Thornton, 2004), the authors inferred the nucleotide substitutions that must have transformed the common ancestor of the ADH1- and ADH2-encoding genes into their present state, synthesized the corresponding enzyme, and confirmed that it was catalytically specialized for ethanol synthesis. Importantly, this work signifies the possibility of reconstructing the sequence of single-nucleotide mutations that might transform one protein to another (see also Thornton, 2004).

These are only four of many examples of such work (Hochachka and Somero, 2002; Watt and Dean, 2000), but are noteworthy for several reasons. First, each exemplifies an unambiguous linkage among gene, trait, whole-organism function or performance, and fitness. Second, each involves alternative single nucleotide polymorphisms (SNPs), alleles or haplotypes segregating in natural populations. This variation could plausibly have arisen through the processes of single-nucleotide mutation: replication error, or damage to a base due to chemical agents or ionizing radiation followed by imperfect repair. Plausibly, multiple rounds of such random mutagenesis could have produced even the more complex haplotypes that are segregating in the latter three examples. Indeed, SNPs are common, account for a substantial fraction of phenotypic variation within populations, and are the basis for the coming era of pharmacogenomics and personalized medicine. No less a personage than Ernst Mayr (Mayr and Provine, 1998), p. 20 stated that '*...gene mutations are the only raw material of evolution*' [as cited by Ryan (Ryan, 2006)], with which many biologists concur. But is single-nucleotide mutagenesis the only process generating genetic variation that is consequential for fitness in natural populations? And is this process sufficient to account for the remarkably complex adaptations of living things?

Is simple single-nucleotide mutation sufficient to generate complex adaptations?

Although Charles Darwin understood neither heredity nor mutation in detail, he clearly anticipated that the origin of complex adaptations through the accumulation of small changes would challenge his theory (Darwin, 1859). Also well in advance of the genomic era, Richard Goldschmidt proposed that, because mutations of small effect would be insufficient to yield complex adaptations, rare mutations of large effect were needed (Goldschmidt, 1960). [As Ernst Mayr (Mayr, 1963) again pointed out, this proposition was itself implausible – and where were these mutations of large effect and how did they arise?] Even today, the 'irreducibility of complexity' (Behe, 1996) is a major talking point in the creationism and 'Intelligent

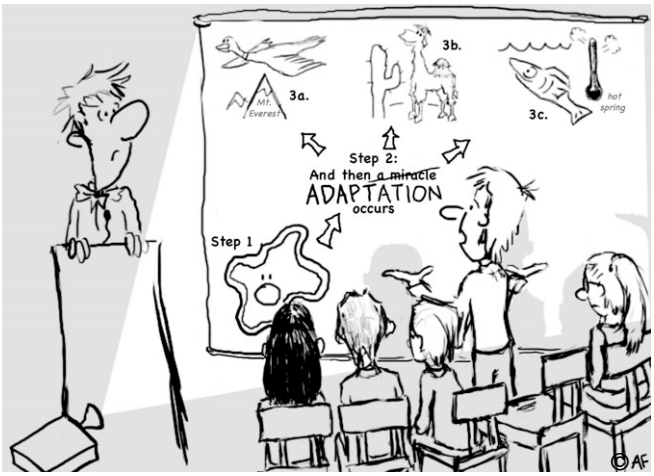
Design' political campaigns against the science of evolution – alongside the absence of fossils documenting 'missing links'. But if 'genes are living fossils' (E. Gaucher, personal communication), the molecular paleontology of modern genomics and its successor sciences have now revealed numerous mechanisms that can create massive variation in genomes upon which evolution can act. First, however, what are the evolutionary issues that these new discoveries may resolve?

Evolvability

According to Falconer (Falconer, 1981), the change in the mean value of a trait from generation to generation is the product of two variables: the intensity of natural selection and the genetic variance of the trait. Intensity of selection is a variable readily comprehensible to comparative physiologists, and needing little substantiation. Nonetheless, Falconer's equation states that no matter how intense selection may be, evolutionary change cannot ensue unless adequate genetic variation is present; all else being equal, evolvability is proportional to genetic variability. Evolvability, however, is seemingly both an independent and a dependent variable. Ongoing selection can purge a population of genetic variability if the intensity of purifying selection exceeds the ability of mutation and other variational processes to generate it – unless, that is, something shields genetic variability from selection. How, then, can genetic variability either be generated *de novo* or shielded from the purgative of selection? And how can genetic variability be generated such that at least some of it is neither so minor as to impede evolution nor so extreme as to be deleterious?

The upgrade pathway

A mechanical engineer could, in principle, transform an automobile into a powered boat by adding a hull, rudder and propeller, and by making appropriate adjustments in the transmission and steering mechanisms. Transformation of the same engine into a machine for powered flight is likewise feasible (cf. the Wright brothers, 1902), or of a propeller-driven aircraft into a jet aircraft or rocket. A computer engineer could upgrade the performance of a computer by installing a new operating system, input/output devices, and firmware. In such cases, human engineers have luxuries unavailable to evolving organisms. First, the engineers can readily power down the machine they are transforming, make changes and restart it; living things can perform the counterpart operation only with great difficulty (but see embryogenesis and metamorphosis). Second, engineers can obtain an extraordinary variety of alternative parts, or invent novel parts if necessary; if relying on simple mutation, evolving organisms can proceed only by modifying the genes already at hand (Long et al., 2003). Third, an engineer can conceive that, for example, to transform an automobile into a powered boat, requires simultaneous changes in the exoskeleton, propulsive system, steering system, etc., and implement these changes simultaneously. If relying on simple mutation, evolving organisms must either await numerous successive mutations, none of which enhances fitness without the others, or an incredibly improbable bout of massive simultaneous mutation. Weinreich et al. (Weinreich et al., 2006) addressed a relevant situation experimentally: bacteria in which five mutations confer maximal resistance against an antibiotic. These five mutations could occur in any of 5! (120) possible orders, of which only 18 successively increase antibiotic resistance. The remainder are prohibited because at least one intermediate stage decreases antibiotic resistance. Finally, an engineer can discard failed attempts and begin again, whereas organisms in which mutation decreases fitness substantially are conceivably doomed evolutionarily. How, then, does a flightless organism evolve flight, a terrestrial organism an aquatic existence, or any less-adapted organism the extraordinary features that comparative physiologists have discovered?



"…but Professor sir…how about some more detail on Step 2?"

Fig. 1. Homage to Sidney Harris (Harris, 1992). The projected image exemplifies a common attitude of biologists towards natural selection; their focus is on how adapted organisms function rather than on how adaptations have arisen. © Alison Feder, used with permission.

Network properties enable evolvability

The flowering of interactomics, network biology, and systems biology has revealed how single-nucleotide substitutions (and/or much larger mutations; see below) may in principle yield transformations of large effect, and thereby create adaptation. Although network biology is still in its infancy, several points are clear. Like any system, biological systems can be characterized as networks of interacting components, be they genes and transcription factors, messengers and signalling pathways, or species in a biological community (Barabâasi, 2003). Not all components in a biological network are equal; some are relatively highly connected (so-called 'hubs') and some less so, with the number of connections obeying a power-law distribution. This 'scale-free' pattern differs from random networks, in which the number of connections is a Poisson distribution. Not all links among components are of equal strength (Csermely, 2006). Finally, the structure of the network is consequential for its function and its resistance to perturbation (i.e. robustness).

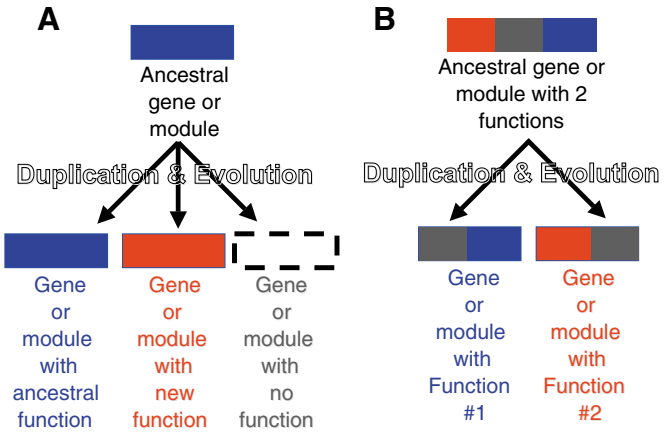


Fig. 2. Duplication–degeneration/complementation model of Force, Lynch, and colleagues (e.g. Force et al., 2005; Force et al., 1999). (A) After duplication, if one duplicate retains the original function, the other may either evolve a new function or deteriorate. (B) After duplication of a gene or module with multiple functions, the duplicates may each evolve distinct functions as long as each function of the ancestral gene/module continues.

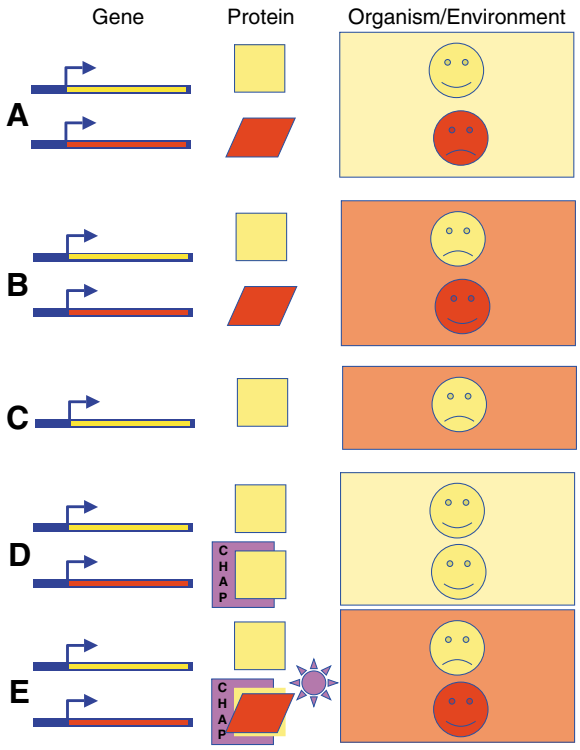


Fig. 3. Selection can purge standing variation from a population, and evolutionary capacitors can maintain it. Suppose alleles encoding 'red' or 'yellow' proteins, which fold differently to yield corresponding colors, segregate in an ancestral population. (A) In a yellow environment, conspicuousness to predators will soon eliminate the red allele from the population. (B) In a red environment, conspicuousness to predators will soon eliminate the yellow allele from a population. (C) If selection eliminates the red allele in the population in Environment A and this population *then* encounters the red environment, pre-existing variation may be insufficient for natural selection. The population may become extinct unless new variation arises. (D) An exemplary putative evolutionary capacitor, the molecular chaperone Hsp90 (CHAP), results in the folding of the red protein so that its conformation is the same as the yellow protein, yielding a yellow phenotype. Because the red phenotype is not expressed, genetic variation persists even in the yellow environment. (E) If the environment then becomes red and heat stress ensues, damaged proteins will out-compete the red proteins for chaperone, allowing the red proteins to fold to yield the red phenotype, which is beneficial in this environment. Compare the outcome here with that in C.

While a more extensive general discussion of networks is impossible here (but see, for example, Barabaasi and Csermely), several features are critical to evolvability.

Biological networks are hierarchical

Because all components in a network are connected, perturbing one component can perturb many – especially if that component is a hub. Examples are numerous: single-gene knock-outs can be lethal, as can pharmacological inhibition of a single enzyme or signaling molecule, and removal of a single 'keystone species' can devastate a natural community. More benignly, a small number of transcription factor or signaling molecules can create a cascade of impacts downstream that effect growth, development, and regulation. Such organization may be inevitable for complex systems (Csete and Doyle, 2004): the necessity for integrating numerous complex systems requires 'common currencies' such as ATP and the major second messengers.

Biological networks are modular

Networks are heterogeneous in the strength and number of connections. Some sub-organizations occur multiple times and are highly connected internally, but are somewhat interchangeable with their counterparts. These are modules, biological equivalents of the individual telephones in a telephone network, individual circuit boards in a computer or parts in a machine, or individual computers in a network. In both development and intracellular signaling, for example, specific modules repeatedly appear in diverse pathways (Carroll, 2005; Pereira-Leal et al., 2006; Pereira-Leal and Teichmann, 2005).

Biological networks are redundant and robust

Networks are more like roller coasters (in which removal of an average single strut has little impact on function) than like houses of cards (in which removal of an average card causes the entire structure to collapse) (Feder, 1996). An exhaustive account of how, at every level of biological organization, change in or removal of components does not affect the performance of the whole has recently been published (Wagner, 2005). Alternative components whose substitution has no impact are said to occupy the same 'functional space', and these functional spaces are remarkably extensive. For example, although some single-gene knockouts can be harmful or lethal (see above), the vast majority of genes can be deleted with no apparent ill effect (Feder and Walser, 2005).

Combining these features yields a recipe for evolvability. Because of redundancy and robustness, mutation can tinker with numerous individual components without jeopardizing the whole. Because of modularity, a mutated module can be substituted for an original module (Carroll, 2005; Pereira-Leal et al., 2006; Pereira-Leal and Teichmann, 2005). Because of hierarchical organization, some small changes can trigger major changes downstream. Thus, while one or more single-nucleotide mutations in the average gene may seldom suffice to generate major adaptive novelty, such mutations in key genes in the right module at the right time may have enormous impacts. Recently there have been several reviews or similar theses advanced (Kirschner and Gerhart, 2005; Carroll, 2005; Carroll et al., 2005; Davidson and Erwin, 2006).

The most relevant formalization of this thinking is the duplication–degeneration/complementation model of Force, Lynch and colleagues (Fig. 2), originally applied to duplicate genes (Force et al., 1999) but recently extended to networks in general (Force et al., 2005). It reasons that when a component is duplicated [and recall that many components and modules are already redundant], only one duplicate is necessary to maintain the pre-existing function. When the second component undergoes change, it may either degenerate (as selection need not maintain it in view of the duplicate) or take on a differing function (like the gene encoding ADH2, discussed above). The differing function can either be an entirely new one (neofunctionalization, as when change results in a gene's promoter acquiring a new response element), or the two duplicates can each take on part of the function of the original component (subfunctionalization, as exemplified by the genes encoding ADH1 and ADH2). Once refunctionalized, the former duplicate may then be preserved by selection. Again because of hierarchical organization, refunctionalization of a single component can trigger major changes downstream. Perhaps the most notorious example of this final point concerns the homeobox or *Hox* genes, which have undergone duplication/refunctionalization in development and now specify whether a body segment grows a limb and/or wing and/or haltere and/or other structure, depending on which particular *Hox* paralog is active in that segment (Carroll et al., 2005). The *Hox* paralogs have acquired the ability to respond appropriately for their position in the developing embryo *via* refunctionalization of their promoters to bind transcription factors that vary in concentration from anterior to posterior. Additional spectacular examples involve the origin of armor plating in stickleback fish (e.g. Colosimo et al., 2005) and coloration in insects (e.g. Gompel et al., 2005).

Interestingly, it has been suggested that the typical network structure of living things is itself a facilitator of evolvability, at least in theory (Oikonomou and Cluzel, 2006). These authors modeled the outputs of networks with equal numbers of components but with different topologies, random *versus* scale-free (i.e. resembling many biological networks), using Boolean threshold dynamics to compare the output with a target function. Emulating natural selection, they then randomly mutated the components of the two networks, and selected for or against mutants according to their resemblance to the target function. In this simulation, the scale-free networks evolved more rapidly and regularly toward the target function than did the random networks.

Evolutionary capacitors enable evolvability

Duplication–degeneration/refunctionalization takes time and, as noted above, natural selection is fully able to purge temporarily unsuccessful intermediates from a population before conditions favor their preservation and/or subsequent evolution. Evolutionary capacitors are mechanisms that shield variation from the purgative of selection. The term was proposed by the editors of the journal *Nature* as the title of the first report that a molecular chaperone, Hsp90, might act in this way (Rutherford and Lindquist, 1998).

Molecular chaperones are proteins (or other compounds) that interact specifically with other 'client' proteins, normally in the context of aggregation deterrence, protein folding and/or degradation; many chaperones are heat-shock or stress proteins. Hsp90 is a peculiar chaperone in that it is normally present and at relatively high concentrations in eukaryotic cells, where it interacts with numerous client proteins. It is thus not specific for a particular client. Hsp90 is also a typical chaperone in that it will bind proteins not in their native (i.e. fully and normally folded) state and assist in their proper folding. Rutherford and Lindquist hypothesized that, because of these properties, Hsp90 could bind mutant gene products that if unchaperoned would yield dramatic phenotypes – and the ensuing loss of their encoding genes due to selection. Hsp90 would thus deter the elimination of these mutations. Upon environmental stress, however, the profusion of stress-damaged proteins in the cell would titrate Hsp90 away from these mutant gene products, allowing their phenotypes to emerge. Hypothetically, if these emergent novel phenotypes were beneficial, they could then be preserved by selection (Fig. 3). Although no beneficial mutation has yet been discovered to have arisen by this mechanism in natural or experimental evolution (see Mayr's criticism of Goldschmidt, above), Lindquist and colleagues have now assembled numerous data consistent with most components of this hypothesis (Cowen and Lindquist, 2005; Queitsch et al., 2002; True et al., 2004), and their work has stimulated broad interest in evolutionary capacitors.

In hindsight, many features of living things are evolutionary capacitors or render genetic variation cryptic (Gibson and Dworkin, 2004; Masel, 2006). These include ploidy >1, recombination, segregation and sex. Like chaperones, these can protect gene products or combinations thereof from selection and suddenly reveal them. Likewise, chromatin configuration (Kornberg and Lorch, 1999; Segal et al., 2006), epigenetic modification (e.g. gene imprinting, methylation), alternative splicing and phenotypic plasticity in general, can all suppress/reveal genetic variation. Alternative splicing is especially spectacular in this regard, as a single gene can encode >38 000 splice variants (Schmucker et al., 2000). Indeed,

evidence is mounting that natural selection can manipulate splicing to express variants appropriate for a given environment (Marden, 2006). Finally, the intronic nature of eukaryotic genes provides an additional site for genetic variation. Mattick and colleagues have hypothesized that introns have enabled the evolution of encoding of RNAs that themselves do not encode proteins, and that these non-coding RNAs are in turn responsible for the complexity of eukaryotes (Mattick, 2003).

The generation of large-scale mutation

Now that techniques are available that can assess and compare most genes in multiple genomes, we are recognizing that large-scale (i.e. involving many nucleotides at once, if not entire genomes) mutations, in combination with network features and evolutionary capacitance, may play a correspondingly major role in the genesis of biological variation, especially the variations that comparative physiologists call 'adaptations'. Unlike single-nucleotide mutation, which unarguably remains an important component of the evolutionary toolkit, these readily yield plausible scenarios of large-scale evolutionary change and origin of new features (Long et al., 2003); termed 'genomic creativity' (Ryan, 2006).

Gene multiplication

Duplication of genes within the genome

The standard genetic machinery of DNA replication, recombination and segregation is not perfect. Not only does this yield single-nucleotide mutations, but also duplications of larger fractions of the genome if not entire genomes. There are several excellent introductions to the topic (Hurles, 2004; Zhang, 2003), which emphasize that gene duplications are about as frequent as single-nucleotide mutations. As a result 30–65% of the genes in eukaryotic genomes are duplicate [table 1 in Zhang (Zhang, 2003)]. Lynch and Conery (Lynch and Conery, 2000) estimated that one in a hundred genes is duplicated and fixed every million years. Thus, gene duplication is *not* a rare occurrence.

Duplications of nucleotide sequences arise primarily by two distinct mechanisms (Hurles, 2004; Zhang, 2003). In the first (tandem or segmental duplication), unequal recombination occurs between homologous sequences at two places in the genome (within the same chromatid, between sister chromatids, between different chromosomes). In the second (retroposition), a transcribed mRNA sequence is reverse-transcribed, and the resultant DNA inserted into a chromosome. Thus retroposition (a) can be into any accessible chromatin in the genome and is not limited to homologous sequence, (b) only duplicates a transcribed sequence, which contains no promoter sequence and need not contain introns, (c) duplicates genes or operons one at a time; in contrast, tandem duplications may be of sequences containing many genes and intergenic regions. Genome-wide studies reveal the extraordinary impact of these processes. Our own species is remarkable for numerous large segmental duplications (Bailey and Eichler, 2006), as is evident in whole-genome displays [see figs 2 and S4 of Bailey et al. (Bailey et al., 2002)]. The *Drosophila* genome contains hundreds of genes lacking introns but otherwise close matches of intron-containing genes on other chromosomes (Betran et al., 2002; Emerson et al., 2004; Long et al., 2003), evidently the result of retroposition.

Errors in segregation during meiosis can also result in extra copies of entire chromosomes or, the utmost in gene duplication, of the entire genome. Indeed, whole genome duplication has clearly occurred in yeast and plants, and two rounds of duplication occurred in the chordate lineage (Dehal and Boore, 2005).

Introduction of genes from outside the genome

Hybridization also has long been recognized as a potential source of genomic change. By definition, it involves the combination of two distinct genomes, typically due to the failure of reproductive isolative mechanisms. Although perhaps not an everyday occurrence, novel species have clearly arisen in nature as a result (reviewed by Ryan, 2006).

Much more frequently, species acquire new genes by lateral or horizontal gene transfer from another species, sometimes involving a biological vector and sometimes simply by uptake of DNA (Bushman, 2002). The likely magnitude of this transfer is such that it has been concluded that no real 'tree of life' exists because all 'branches' have interchanged genes and hence are genealogically indistinct (Doolittle, 1999). Again, the combination of genomic information and bioinformatic tools establishes how extensive this interchange is [see fig. 3 in Kunin et al. (Kunin et al., 2005)].

Finally, few species of eukaryotes are, strictly speaking, single species, but actually are multispecies consortia of symbiotes and their genomes (Margulis and Sagan, 2002). Whatever one's reaction to the more controversial aspects of Margulis's thesis, interactions of symbiont genomes and host genomes have clearly brought about the evolution of what physiologists classically regard as adaptations, such as the bacteriocyte of aphids (Moran and Degnan, 2006), the light organ of squids (Koropatnick et al., 2004), and the rhizosphere of plant roots (Marx, 2004).

Mobile genetic elements

Some DNA sequences encode the capacity to move within or among genomes. A first class of mobile element moves by encoding a mRNA that is reverse-transcribed and inserts elsewhere in the genome; these are aptly named retrotransposons. A second class either copies itself or excises from its original site and inserts elsewhere, but remains DNA throughout. Those unfamiliar with transposable elements tend vastly to underestimate their impact on evolutionary change (Kazazian, 2004); some (V. Walbot, personal communication; M. Evgen'ev, personal communication) attribute perhaps 40–90% of all evolutionary change to transposable elements. The simplest impact of these elements is to interrupt the sequence of their current (and often former) site in the genome;

i.e. insertional mutagenesis. Transposable elements sometimes capture portions of DNA sequence adjacent to their original insertion site and, when they excise, carry it with them. They may thus convey novel regulatory and/or coding sequence to a new insertion site, and thereby endow the new host gene with novel properties. Transposable elements now constitute large fractions of all eukaryotic genomes sequenced so far (e.g. approximately 50% in humans).

Scrambling genes within genomes

Once genomes have acquired new genes, whether by duplication or from another species, numerous processes in addition to single-nucleotide mutation may modify them, sometimes far more dramatically than can single-gene mutation. As noted, retroposition, segmental duplication and transposable elements can each transfer sequence within the genome, creating combinations of sequences that were previously non-existent. Alternative splicing can likewise create novel combinations of gene products. Additionally, mechanisms such as exon shuffling, gene fission/fusion can create new genes (Long et al., 2003). In some rare cases, novel genes can even arise from nonsense or intergenic sequence (Long et al., 2003).

Interaction among these features

The foregoing mechanisms, although often individually distinct, can interact to yield evolutionary novelty. For example (Zhang et al., 2004), the *Drosophila* gene *jingwei* arose when a first gene, *yellow emperor*, duplicated to yield a second gene, *yande*, and a third gene, *alcohol dehydrogenase*, retroposed into the third intron of *yande*. Thereafter, at least 30 non-silent single-nucleotide mutations yielded the present-day gene. Unlike its ADH ancestor, which is specialized to detoxify and assimilate ethanol, the substrate specificity of *jingwei* is for long-chain primary alcohols, which are important in hormone and pheromone metabolism.

In another example (Walser et al., 2006), the *P* element, a transposable element of *Drosophila*, needs physical access to decondensed DNA to insert itself. In many genes, the chromatin is normally condensed and in nucleosomes, except when the gene is actually being expressed (i.e. chromatin is often an evolutionary capacitor). Perhaps due to their role as an emergency response to stress, the proximal promoters of the genes encoding heat-shock proteins are constitutively decondensed and nucleosome-free. This has made the heat-shock promoters distinctively susceptible to the insertion of *P* elements, which selection may then maintain or eliminate to manipulate the expression of their host genes.

Finally, because the sequence of transposable elements is often highly conserved, their insertion throughout the genome creates opportunities for homologous recombination among distant sites, as noted above. Such recombination, in turn, may invert, duplicate or otherwise scramble genes, multi-genic nucleotide sequences, or segments of chromosomes. Observing that one class of transposable element, a form of *Alu* element, is often present at the boundaries of segmental duplications in the human genome, Bailey and Eichler (Bailey et al., 2003; Liu et al., 2003) have hypothesized that the extensive segmental duplication in the human genome is due to the proliferation of one class of *Alu* elements in the human lineage (but not in close relatives), which then allowed ectopic recombination. Thus, to the extent that our species owes its capacities to gene duplication, these elements may have underlain the evolution of humankind.

Conclusion and prospectus

The advent of molecular, genomic and bioinformatic techniques and their increasing applicability to diverse species has enormously enhanced experimental biologists' ability to understand 'how animals work' (Schmidt-Nielsen, 1972). Adaptational biology will be incomplete, however, until the understanding of how adaptations came into being is equally advanced. This understanding may well come about through sustained interaction with modern evolutionary biologists, evolutionary genomicists and evolutionary systems biologists. One clear outcome of such interaction is that single-nucleotide mutation, often the mainstay of adaptational biologists' evolutionary thinking, will become viewed as only one of several mechanisms in evolution's toolkit. The other mechanisms may be far more powerful than single-nucleotide mutation in facilitating evolvability and, although they have not done so yet, be able to explain in detail the origin of the complex traits that fascinate adaptational biologists.

The author thanks the participants of the *Journal of Experimental Biology* symposium on 'Post-Genomic Comparative Physiology' and two anonymous reviewers for insightful comments. This work was supported by NSF Grant 03-16627.

References

Bailey, J. A. and Eichler, E. E. (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* **7**, 552-564.

Bailey, J. A., Gu, Z. P., Clark, R. A., Reinert, K., Samonte, R. V., Schwartz, S., Adams, M. D., Myers, E. W., Li, P. W. and Eichler, E. E. (2002). Recent segmental duplications in the human genome. *Science* **297**, 1003-1007.

Bailey, J. A., Liu, G. and Eichler, E. E. (2003). An Alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* **73**, 823-834.

Barabâasi, A.-L. (2003). *Linked: How Everything is Connected to Everything Else and What it Means for Business, Science, and Everyday Life*. New York: Plume.

Behe, M. J. (1996). *Darwin's Black Box: The Biochemical Challenge to Evolution*. New York: Free Press.

Betran, E., Thornton, K. and Long, M. (2002). Retroposed new genes out of the X in *Drosophila*. *Genome Res.* **12**, 1854-1859.

Bushman, F. (2002). *Lateral DNA Transfer: Mechanisms and Consequences*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biol.* **3**, 1159-1166.

Carroll, S. B., Grenier, J. K. and Weatherbee, S. D. (2005). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Malden, MA: Blackwell.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D. and Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928-1933.

Cowen, L. E. and Lindquist, S. (2005). Hsp90 potentiates the rapid evotution of new traits: drug resistance in diverse fungi. *Science* **309**, 2185-2189.
Csermely, P. (2006). *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks*. Berlin, New York: Springer.
Csete, M. and Doyle, J. (2004). Bow ties, metabolism, and disease. *Trends Biotechnol.* **22**, 446-450.
Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London: J. Murray.
Davidson, E. H. and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796-800.
Dehal, P. and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, 1700-1708.
Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* **284**, 2124-2128.
Emerson, J. J., Kaessmann, H., Betran, E. and Long, M. Y. (2004). Extensive gene traffic on the mammalian X chromosome. *Science* **303**, 537-540.
Endler, J. A. (1986). *Natural Selection in the Wild*. Princeton, NJ: Princeton University Press.
Falconer, D. S. (1981). *Introduction to Quantitative Genetics*. London, New York: Longman.
Feder, M. E. (1996). Ecological and evolutionary physiology of stress proteins and the stress response: the *Drosophila melanogaster* model. In *Animals and Temperature: Phenotypic and Evolutionary Adaptation* (ed. I. A. Johnston and A. F. Bennett), pp. 79-102. Cambridge: Cambridge University Press.
Feder, M. E. and Mitchell-Olds, T. (2003). Evolutionary and ecological functional genomics. *Nat. Rev. Genet.* **4**, 651-657.
Feder, M. E. and Walser, J. C. (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J. Evol. Biol.* **18**, 901-910.
Feder, M. E., Bennett, A. F. and Huey, R. B. (2000). Evolutionary physiology. *Annu. Rev. Ecol. Syst.* **31**, 315-341.
Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-1545.
Force, A., Cresko, W. A., Pickett, F. B., Proulx, S. R., Amemiya, C. and Lynch, M. (2005). The origin of subfunctions and modular gene regulation. *Genetics* **170**, 433-446.
Gibson, G. and Dworkin, I. (2004). Uncovering cryptic genetic variation. *Nat. Rev. Genet.* **5**, 681-690.
Goldschmidt, R. B. (1960). *The Material Basis of Evolution*. Paterson, NJ: Pageant Books.
Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. and Carroll, S. B. (2005). Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**, 481-487.
Harris, S. (1992). *Chalk Up Another One: The Best of Sidney Harris*. Washington, DC: AAAS Press.
Hochachka, P. W. and Somero, G. N. (2002). *Biochemical Adaptation: Mechanism and Process in Physiological Evolution*. New York: Oxford University Press.
Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A. and Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101-104.
Hurles, M. (2004). Gene duplication: the genomic trade in spare parts. *PLoS Biol.* **2**, 900-904.
Ingram, V. M. (1963). *The Hemoglobins in Genetics and Evolution*. New York: Columbia University Press.
Kazazian, H. H. (2004). Mobile elements: Drivers of genome evolution. *Science* **303**, 1626-1632.
Kirschner, M. and Gerhart, J. (2005). *The Plausibility of Life: Great Leaps of Evolution*. New Haven, CT: Yale University Press.
Kornberg, R. D. and Lorch, Y. L. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285-294.
Koropatnick, T. A., Engle, J. T., Apicella, M. A., Stabb, E. V., Goldman, W. E. and McFall-Ngai, M. J. (2004). Microbial factor-mediated development in a host-bacterial mutalism. *Science* **306**, 1186-1188.
Kunin, V., Goldovsky, L., Darzentas, N. and Ouzounis, C. A. (2005). The net of life: reconstructing the microbial phylogenetic network. *Genome Res.* **15**, 954-959.
Liu, G., Zhao, S. Y., Bailey, J. A., Sahinalp, S. C., Alkan, C., Tuzun, E., Green, E. D. and Eichler, E. E. (2003). Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* **13**, 358-368.
Long, M., Betran, E., Thornton, K. and Wang, W. (2003). The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* **4**, 865-875.
Lynch, M. and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151-1155.
Marden, J. H. (2006). Quantitative and evolutionary biology of alternative splicing: how changing the mix of alternative transcripts affects phenotypic plasticity and reaction norms. *Heredity* doi: 10.1038/sj.hdy.6800904.
Margulis, L. and Sagan, D. (2002). *Acquiring Genomes: A Theory of the Origins of Species*. New York: Basic Books.
Marx, J. (2004). The roots of plant-microbe collaborations. *Science* **304**, 234-236.
Masel, J. (2006). Cryptic genetic variation is enriched for potential adaptations. *Genetics* **172**, 1985-1991.
Mattick, J. S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *BioEssays* **25**, 930-939.
Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, MA: Belknap Press of Harvard University Press.
Mayr, E. and Provine, W. B. (1998). *The Evolutionary Synthesis: Perspectives on the Unification of Biology*. Cambridge, MA: Harvard University Press.
Moran, N. A. and Degnan, P. H. (2006). Functional genomics of *Buchnera* and the ecology of aphid hosts. *Mol. Ecol.* **15**, 1251-1261.
Oikonomou, P. and Cluzel, P. (2006). Effects of topology on network evolution. *Nat. Phys.* **2**, 532-536.
Pereira-Leal, J. B. and Teichmann, S. A. (2005). Novel specificities emerge by stepwise duplication of functional modules. *Genome Res.* **15**, 552-559.
Pereira-Leal, J. B., Levy, E. D. and Teichmann, S. A. (2006). The origins and evolution of functional modules: lessons from protein complexes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 507-517.
Queitsch, C., Sangster, T. A. and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618-624.
Rutherford, S. L. and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336-342.
Ryan, F. P. (2006). Genomic creativity and natural selection: a modern synthesis. *Biol. J. Linn. Soc. Lond.* **88**, 655-672.
Schmidt-Nielsen, K. (1972). *How Animals Work*. Cambridge: Cambridge University Press.
Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E. and Zipursky, S. L. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671-684.
Segal, E., Fondufe-Mittendorf, Y., Chen, L. Y., Thastrom, A., Field, Y., Moore, I. K., Wang, J. P. Z. and Widom, J. (2006). A genomic code for nucleosome positioning. *Nature* **442**, 772-778.
Segal, J. A., Schulte, P. M., Powers, D. A. and Crawford, D. L. (1996). Descriptive and functional characterization of variation in the *Fundulus heteroclitus Ldh-B* proximal promoter. *J. Exp. Zool.* **275**, 355-364.
Thomson, J. M., Gaucher, E. A., Burgan, M. F., De Kee, D. W., Li, T., Aris, J. P. and Benner, S. A. (2005). Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat. Genet.* **37**, 630-635.
Thornton, J. W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366-375.
True, H. L., Berlin, I. and Lindquist, S. L. (2004). Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. *Nature* **431**, 184-187.
Van Straalen, N. M. and Roelofs, D. (2006). *An Introduction to Ecological Genomics*. Oxford, New York: Oxford University Press.
Wagner, A. (2005). *Robustness and Evolvability in Living Systems*. Princeton, NJ: Princeton University Press.
Walser, J. C., Chen, B. and Feder, M. E. (2006). Heat-shock promoters: targets for evolution by *P* transposable elements in *Drosophila*. *PLoS Genet.* **2**, 1541-1555.
Watt, W. B. and Dean, A. M. (2000). Molecular-functional studies of adaptive genetic variation in prokaryotes and eukaryotes. *Annu. Rev. Genet.* **34**, 593-622.
Weinreich, D. M., Delaney, N. F., DePristo, M. A. and Hartl, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111-114.
Williams, T. N., Mwangi, T. W., Roberts, D. J., Alexander, N. D., Weatherall, D. J., Wambua, S., Kortok, M., Snow, R. W. and Marsh, K. (2005). An immune basis for malaria protection by the sickle cell trait. *PLoS Med.* **2**, 441-445.
Zhang, J. M., Dean, A. M., Brunet, F. and Long, M. Y. (2004). Evolving protein functional diversity in new genes of Drosophila. *Proc. Natl. Acad. Sci. USA* **101**, 16246-16250.
Zhang, J. Z. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**, 292-298.

The Journal of Experimental Biology 210, 1492-1496
Published by The Company of Biologists 2007
doi:10.1242/jeb.02783

Glossary of terms

This section is designed to help readers adapt to the complex terminology associated with contemporary molecular genetics, genomics and systems biology. Fuller descriptions of these terms are available at http://www.wikipedia.org/

Ab initio prediction	methods used to predict the potential genes encoded in the genome, which are trained on datasets made of known genes, and used computationally to predict coding regions out of genome without the aid of cDNA sequence. Although their performance is improving, these algorithms perform very poorly on non-protein coding genes.
Annotation	as applied to proteins, DNA sequences or genes. The storage of data describing these entities (protein/gene identities, DNA motifs, gene ontology categorisation, etc.) within a biological database. Active projects include FlyBase and WormBase. See Gene ontology.
Assembly	the process of aligning sequenced fragments of DNA into their correct positions within the chromosome or transcript.
cDNA	complementary DNA. This is DNA synthesised from a mature mRNA template by the enzyme reverse transcriptase. cDNA is frequently used as an early part of gene cloning procedures, since it is more robust and less subject to degradation than the mRNA itself.
ChIP	chromatin immunoprecipitation assay used to determine which segments of genomic DNA are bound to chromatin proteins, mainly including transcription factors.
Chip	see Microarray.
ChIP-on-chip	use of a DNA microarray to analyse the DNA generated from chromatin immunoprecipitation experiments (see ChIP).
cis-acting	a molecule is described as *cis*-acting when it affects other genes that are physically adjacent, on the same chromosome, or are genetically linked or in close proximity (for mRNA expression, typically a promoter).
Collision-induced dissociation	a mechanism by which molecules (e.g. proteins) are fragmented to form molecular ions in the gas phase. These fragments are then analysed within a mass spectrometer to provide mass determination.
Connectivity	a term from graph theory, which indicates the number of connections between nodes or vertices in a network. Greater connectedness between nodes is generally used as a measure of robustness of a network.
CpG islands	regions that show high density of ‘C followed by G’ dinucleotides and are generally associated with promoter elements; in particular, stretches of DNA of at least 200 bp with a C–G content of 50% and an observed CpG/expected CpG in excess of 0.6. The cytosine residues can be methylated, generally to repress transcription, while demethylated CpGs are a hallmark of transcription. CpG dinucleotides are under-represented outside regulatory regions, such as promoters, because methylated C mutates into T by deamination.
Edge	as in networks. Connects two nodes (or vertices) within a system. These concepts arise from graph theory.
Enhancer	a short segment of genomic DNA that may be located remotely and that, on binding particular proteins (*trans*-acting factors), increases the rate of transcription of a specific gene or gene cluster.
Epistasis	a phenomenon when the properties of one gene are modified by one or more genes at other loci. Otherwise known as a genetic interaction, but epistasis refers to the statistical properties of the phenomenon.

eQTL	the combination of conventional QTL analysis with gene expression profiling, typically using microarrays. eQTLs describe regulatory elements controlling the expression of genes involved in specific traits.
EST	expressed sequence tag. A short DNA sequence determined for a cloned cDNA representing portions of an expressed gene. The sequence is generally several hundred base pairs from one or both ends of the cloned insert.
Exaptation	a biological adaptation where the current function is not that which was originally evolved. Thus, the defining (derived) function might replace or persist with the earlier, evolved adaptation.
Exon	any region of DNA that is transcribed to the final (spliced) mRNA molecule. Exons interleave with segments of non-coding DNA (introns) that are removed (spliced out) during processing after transcription.
Gene forests	genomic regions for which RNA transcripts, produced from either DNA strand, have been identified without gaps (non-transcribed genomic regions). Conversely, regions in which no transcripts have ever been detected are called ‘gene deserts’.
Gene interaction network	a network of functional interactions between genes. Functional interactions can be inferred from many different data types, including protein–protein interactions, genetic interactions, co-expression relationships, the co-inheritance of genes across genomes and the arrangement of genes in bacterial genomes. The interactions can be represented using network diagrams, with lines connecting the interacting elements, and can be modelled using differential equations.
Gene ontology (GO)	an ontology is a controlled vocabulary of terms that have logical relationships with each other and that are amenable to computerised manipulation. The Gene Ontology project has devised terms in three domains: biological process, molecular function and cell compartment. Each gene or DNA sequence can be associated with these annotation terms from each domain, and this enables analysis of microarray data on groups of genes based on descriptive terms so provided. See http://www.geneontology.org
Gene set enrichment analysis	a computational method that determines whether a defined set of genes, usually based on their common involvement in a biological process, shows statistically significant differences in transcript expression between two biological states.
Gene silencing	the switching-off of a gene by an epigenetic mechanism at the transcriptional or post-transcriptional levels. Includes the mechanism of RNAi.
Genetic interaction (network)	a genetic interaction between two genes occurs when the phenotypic consequences of a mutation in one gene are modified by the mutational status at a second locus. Genetic interactions can be aggravating (enhancing) or alleviating (suppressing). To date, most high-throughput studies have focussed on systematically identifying synthetic lethal or sick (aggravating) interactions, which can then be visualised as a network of functional interactions (edges) between genes (nodes).
Genome	a portmanteau of gene and chromosome, the entire hereditary information for an organism that is embedded in the DNA (or, for some viruses, in RNA). Includes protein-coding and non-coding sequences.
Heritability	phenotypic variation within a population is attributable to the genetic variation between individuals and to environmental factors. Heritability is the proportion due to genetic variation usually expressed as a percentage.
Heterologous hybridization	the use of a cDNA or oligonucleotide microarray of probes designed for one species with target cRNA/cDNAs from a different species.
Homeotic	the transformation of one body part to another due to mutation of specific developmentally related genes, notably the *Hox* genes in animals and *MADS-box* genes in plants.
Hub	as in networks. A node with high connectivity, and thus which interacts with many other nodes in the network. A hub protein interacts with many other proteins in a cell.

Hybridisation	the process of joining (annealing) two complementary single-stranded DNAs into a single double-stranded molecule. In microarray analysis, the target RNA/DNA from the subject under investigation is denatured and hybridised to probes that are immobilised on a solid phase (i.e. glass microscope slide).
Hypomorph	in genetics, a loss-of-function mutation in a gene, but which shows only a partial reduction in the activity it influences rather than a complete loss (cf. hypermorph, antimorph, neomorph, etc).
Imprinting	a phenomenon where two inherited copies of a gene are regulated in opposite ways, one being expressed and the other being repressed.
Indel	insertion and deletion of DNA, referring to two types of genetic mutation. To be distinguished from a 'point mutation', which refers to the substitution of a single base.
Interactome	a more or less comprehensive set of interactions between elements within cells. Usually applied to genes or proteins as defined by transcriptomic, proteomic or protein–protein interaction data.
Intron	see Exon.
KEGG	The Kyoto Encyclopedia of Genes and Genomes is a database of metabolic and other pathways collected from a variety of organisms. See http://www.genome.jp/kegg
Metabolomics	the systematic qualitative and quantitative analysis of small chemical metabolite profiles. The metabolome represents the collection of metabolites within a biological sample.
Metagenomics	the application of genomic techniques to characterise complex communities of microbial organisms obtained directly from environmental samples. Typically, genomic tags are sequence characterised as markers of each species to inform on the range and abundance of species in the community.
Microarray	an arrayed set of probes for detecting molecularly specific analytes or targets. Typically, the probes are composed of DNA segments that are immobilised onto the solid surface, each of which can hybridise with a specific DNA present in the target preparation. DNA microarrays are used for profiling of gene transcripts.
Model species	a species used to study particular biological phenomena, the outcome offering insights into the workings of other species. Usually, the selection is based on experimental tractability, particularly ease of genetic manipulation. For the geneticist, it is an organism with inbred lines where sibs will be >98% identical (i.e. *Drosophila*, *Caenorhabditis elegans* and mice). For genomic science, it refers to a species for which the genomic DNA has been sequenced.
miRNA	a category of novel, very short, non-coding RNAs, generated by the cleavage of larger precursors (pri-miRNA). These short RNAs are included in the RNA-induced silencing complex (RISC) and pair to the 3′ ends of target RNA, blocking its translation into proteins (in animals) or promoting RNA cleavage and degradation (in plants).
mRNA	a protein-coding mRNA containing a protein-coding region (CDS), preceded by a 5′ and followed by a 3′ untranslated region (5′ UTR and 3′ UTR). The UTRs contain regulatory elements. A full-length cDNA contains the complete sequence of the original mRNA, including both UTRs. However, it is often difficult to assign the starting–termination positions for protein synthesis unambiguously. A cDNA containing the entire CDS is often considered acceptable for bioinformatic and experimental studies requiring full-length cDNAs.
ncRNA	non-coding RNA is any RNA molecule with no obvious protein-coding potential for at least 80 or 100 amino acids, as determined by scanning full-length cDNA sequences. It includes ribosomal (rRNA) and transfer RNAs (tRNA) and is now known to include various sub-classes of RNA, including snoRNA, siRNA and piRNA. Just like the coding mRNAs, a large proportion of ncRNAs are transcribed by RNA polymerase II and are large transcripts. A description of the many forms of ncRNA can be found at http://en.wikipedia.org/wiki/Non-coding_RNA.

Node	as in networks. Objects linked by edges to create a network.
PCR	polymerase chain reaction. A molecular biology technique for replicating DNA *in vitro*. The DNA is thus amplified, sometimes from very small amounts. PCR can be adapted to perform a wide variety of genetic manipulations.
piRNA	Piwi-interacting RNA. A class of RNA molecules (29–30 nt long) that complex with Piwi proteins (a class of the Argonaute family of proteins) and are involved in transcriptional gene silencing.
PMF	peptide mass fingerprinting. An analytical technique for protein identification in which a protein is fragmented using proteases. The resulting peptides are analysed by mass spectrometry and these masses compared against a database of predicted or measured masses to generate a protein identity.
Polyadenylation	the covalent addition of multiple A bases to the 3′ tail of an mRNA molecule. This occurs during the processing of transcripts to form the mature, spliced molecule and is important for regulation of turnover, trafficking and translation.
Post-source decay	in mass spectrometry. The fragmentation of precursor molecular ions as they accelerate away from the ionisation source of the mass spectrometer. All precursor ions leaving the ion source have approximately the same kinetic energy, but fragmentation results in smaller product ions that can be distinguished from precursor ions using a ‘reflectron’ by virtue of their lower kinetic energies.
Post-translational modification	the chemical modification of a protein after synthesis through translation. Some modifications, notably phosphorylation, affect the properties of the protein, offering a means of regulating function.
Principal component analysis (PCA)	a technique for simplifying complex, multi-dimensional datasets to a reduced number of dimensions, the principal components. This procedure retains those characteristics of the data that relate to its variance.
Promoter	a regulatory DNA sequence, generally lying upstream of an expressed gene, which in concert with other often distant regulatory elements directs the transcription of a given gene.
Proteome	the entire protein complement of an organism, tissue or cell culture at a given time.
Quantitative trait	inheritance of a phenotypic property or characteristic that varies continuously between extreme states and can be attributed to interactions between multiple genes and their environment.
qPCR	quantitative real-time PCR, sometimes called real-time PCR. A more quantitative form of RT-PCR in which the quantity of amplified product is estimated after each round of amplification.
QTL	quantitative trait loci. A region of DNA that contains those genes contributing to the trait under study.
RISC	RNA-induced silencing complex. A protein complex that mediates the double-stranded RNA-induced destruction of homologous mRNA.
RNAi	RNA interference or RNA-mediated interference. The process by which double-stranded RNA triggers the destruction of homologous mRNA in eukaryotic cells by the RISC.
RT-PCR	reverse transcription–polymerase chain reaction. A technique for amplifying a defined piece of RNA that has been converted to its complementary DNA form by the enzyme reverse transcriptase. See qPCR.
siRNA	small interfering RNA, or silencing RNA. A class of short (20–25 nt), double-stranded RNA molecules. It is involved in the RNA interference pathway, which alters RNA stability and thus affects RNA concentration and thereby suppresses the normal expression of specific genes. Widely used in biomedical research to ablate specific genes.

snoRNA	small nucleolar RNA. A sub-class of RNA molecules involved in guiding chemical modification of ribosomal RNA and other RNA genes as part of the regulation of gene expression.
SNP	single nucleotide polymorphism. A single base-pair mutation at a specific locus, usually consisting of two alleles. Because SNPs are conserved over evolution, they are frequently used in QTL analysis and in association studies in place of microsatellites, and in genetic fingerprinting analyses.
SSH	suppressive subtractive hybridisation. A powerful protocol for enriching cDNA libraries for genes that differ in representation between two or more conditions. It combines normalisation and subtraction in a single procedure and allows the detection of low-abundance, differentially expressed transcripts, such as those involved in signalling and signal transduction.
Structural RNAs	a class of non-coding RNA, long known to have a structural role (for instance, the ribosomal RNAs), transcribed by RNA polymerase I or III.
Systems biology	treatment of biological entities as systems composed of defined elements interacting in defined ways to enable the observed function and behaviour of that system. The properties of the systems are embedded in a quantitative model that guides further tests of systems behaviour.
TATA-boxes	sequences in promoter regions constituted by TATAAA, or similar variants, which were considered the hallmark of Promoters. Recent data show that they are present only in the minority of promoters, where they direct transcription at a single well-defined location some 30 bp downstream of this element.
trans-acting	a factor or gene that acts on another unlinked gene, a gene on a separate chromosome or genetically unlinked usually through some diffusible protein product (for mRNA expression, typically a transcription factor).
Transcript	an RNA product produced by the action of RNA polymerase reading the sequence of bases in the genomic DNA. Originally limited to protein-coding sequences with flanking UTRs but now known to include large numbers of products that do not code for a protein product.
Transcriptome	the full set of mRNA molecules (transcripts) produced by the system under observation. Whilst the genome is fixed for a given organism, the transcriptome varies with context (i.e. tissue source, ontogeny, external conditions or experimental treatment).
Transgene	a gene or genetic material that has been transferred between species or between organisms using one of several genetic engineering techniques.
Transinduction	generation of transcripts from intergenic regions. At least some such products do not relate to a definable promoter or transcriptional start site.
Transposon	sequences of DNA able to move to new positions within the genome of a single cell. This event might cause mutation at the site of insertion. Also called ‘mobile genetic elements’ or ‘jumping genes’.
Transvection	an epigenetic phenomenon arising from the interaction between one allele and the corresponding allele on the homologous chromosome, leading to gene regulation.
TUs	transcriptional units. Used to group all of the overlapping RNA transcripts that are transcribed from the same genomic strand and share exonic sequences.
UTR	untranslated region. Regions of the mRNA that lie at either the 3′ or 5′ flanking ends of the molecule (i.e. 3′ UTR and 5′ UTR). They bracket the protein-coding region and contain signals and binding sites that are important for the regulation of both protein translation and RNA degradation.