The genetic causes of convergent evolution

David L. Stern

Abstract | The evolution of phenotypic similarities between species, known as convergence, illustrates that populations can respond predictably to ecological challenges. Convergence often results from similar genetic changes, which can emerge in two ways: the evolution of similar or identical mutations in independent lineages, which is termed parallel evolution; and the evolution in independent lineages of alleles that are shared among populations, which I call collateral genetic evolution. Evidence for parallel and collateral evolution has been found in many taxa, and an emerging hypothesis is that they result from the fact that mutations in some genetic targets minimize pleiotropic effects while simultaneously maximizing adaptation. If this proves correct, then the molecular changes underlying adaptation might be more predictable than has been appreciated previously.

Fitness
The potential evolutionary success of a genotype, defined as the reproductive success or the proportion of genes that an individual leaves in the gene pool of the next generation in a population. The individuals with the greatest fitness leave the highest number of surviving offspring.

Hybridization
Interbreeding of individuals from genetically distinct populations, regardless of the taxonomic status of the populations.

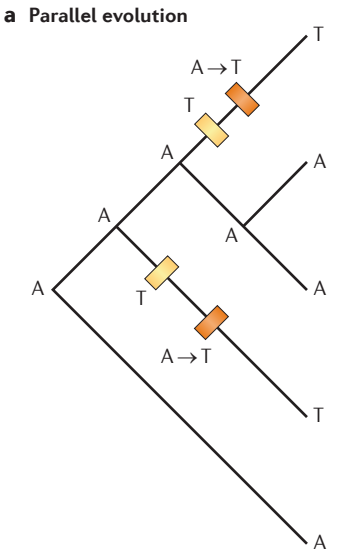
Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia 20147, USA.
e-mail: sternd@janelia.hhmi.org
doi:10.1038/nrg3483
Published online 9 October 2013

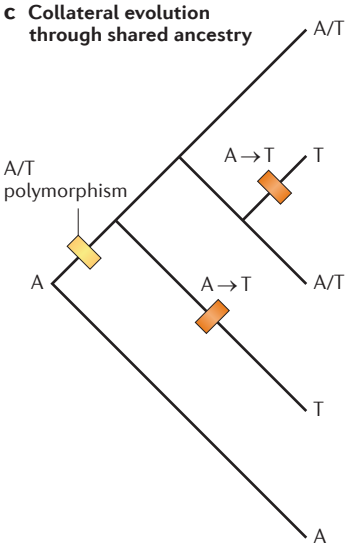
Different species often evolve similar solutions to environmental challenges. Insects, birds and bats evolved wings, and octopi, vertebrates and spiders evolved focusing eyes. Phenotypic convergence provides compelling evidence that ecological circumstances can select for similar evolutionary solutions[1,2]. Historically, convergent evolution was thought to occur primarily by divergent evolution of genetic mechanisms. For example, multiple instances of wing evolution almost certainly reflect evolution mainly through different genetic mechanisms in different taxa. Also, if convergence is considered at a sufficiently general level, such as convergence of organismal fitness to similar environmental challenges, then multiple divergent genetic mechanisms might often contribute to increasing fitness. However, at a more fine-grained level, recent studies have revealed that morphology and physiology often converge owing to the evolution of similar molecular mechanisms in independent lineages. In microorganisms, even fitness convergence often evolves through similar genetic changes. These new data reveal that genetic evolution may be more predictable than was appreciated before the application of molecular biology to evolutionary questions.

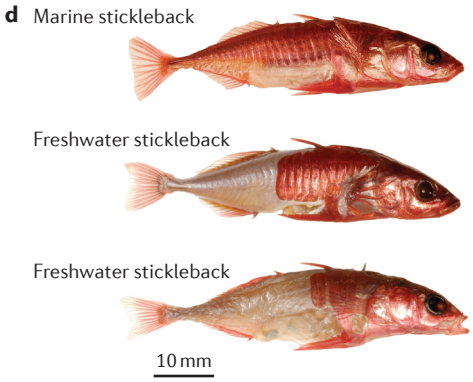
Convergent evolution at the genetic level can result from one of three processes: first, evolution by mutations that occurred independently in different populations or species; second, evolution of an allele that was polymorphic in a shared ancestral population; and third, evolution of an allele that was introduced from one population into another by hybridization, a process that is known as introgression (FIG. 1). It is worth distinguishing between these scenarios because each provides evidence for a different evolutionary path[3]. The first case, the independent origin and spread of mutations, has been called parallel genetic evolution. I suggest that the evolution of alleles which were present in an ancestral population (the second case) and the evolution of introgressed alleles (the third case) should be collectively called collateral genetic evolution (the Oxford Dictionary of English[109] defines collateral as being "descended from the same stock but by a different line"). This precedent comes from palaeontology, in which, in 1969, Shaw[4] defined collateral evolution as the simultaneous appearance of the same phenotypic forms in the stratigraphic record.

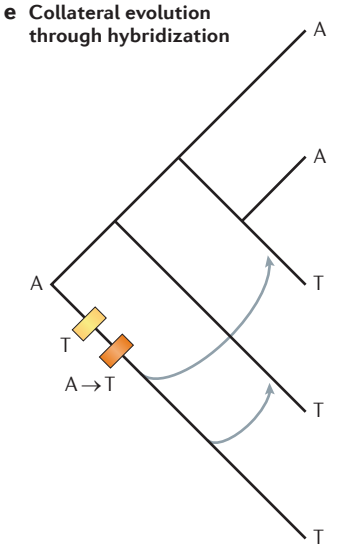
Although phenotypic convergence provides evidence for similar patterns of natural selection[1], parallel and collateral evolution can provide evidence for constraints on how variation can be generated by the genome or for natural selection that favours the fixation of some genetic variants over others, or both[3,5–8]. More than 100 cases of parallel and collateral genetic evolution have been identified in recent years[3] and have been studied using a variety of approaches (BOX 1). The abundance of parallel and collateral evolution therefore implies that genome evolution is not random, but rather that the origin or the selective consequences of genetic variants, or both, might be somewhat predictable.

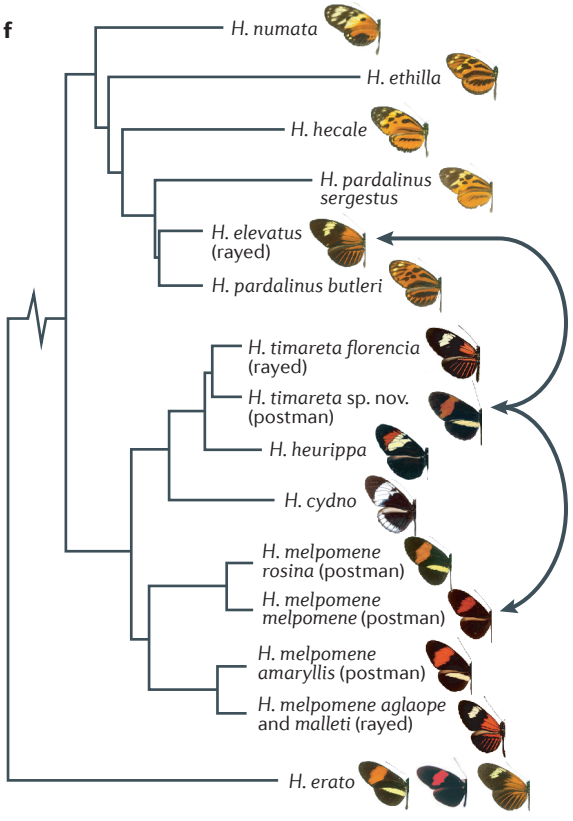
In this Review, I discuss three major topics. First, I clarify some definitions and the implications of parallel versus collateral evolution. Second, I review recent examples that illustrate major patterns in parallel and











◀ Figure 1 | **Parallel and collateral genetic evolution.** Convergent phenotypic evolution that results from similar molecular mechanisms acting in divergent taxa can occur through three historical paths, illustrated here in a phylogenetic framework. **a** | Parallel evolution refers to mutations that arise and spread in independent lineages. In this case, the ancestral state (A) independently evolved to a derived state (T) in two lineages. The yellow rectangles indicate the mutational origins of the T allele; the orange rectangles represent substitutions of the A allele by the T allele throughout the population. **b** | An example of parallel evolution is shown: the monarch butterfly caterpillar (top left), the red milkweed beetle (top right), oleander aphids (bottom left) and the large milkweed beetle (bottom right), among others, have all evolved to feed on poisonous milkweed plants through parallel mutations in the (Na^++K^+) ATPase gene[37,38]. Extant species can undergo similar evolutionary changes by collateral evolution through shared ancestry or through hybridization. **c** | In collateral evolution through shared ancestry, a mutation arises in an ancestral lineage (yellow rectangle) and later substitutes in multiple descendent lineages (orange rectangles). **d** | An example of collateral evolution through shared ancestry is shown. The loss of lateral plates in sticklebacks is illustrated by comparing a marine morph with complete body armour (top), a rare freshwater morph with an intermediate number of lateral plates (middle) and a typical freshwater morph with few lateral plates (bottom). The allele of major effect that reduces body armour in most freshwater populations is also found at low frequency in the ancestral marine populations[45]. **e** | In collateral evolution through hybridization, a mutation arises in one lineage, descendants of which then hybridize with other species and spread the new mutation to related species. The mutation does not need to become fixed in the population for it to spread by hybridization. **f** | An example of collateral evolution through hybridization is shown. The patterns of wing colours in three species of butterflies from the genus *Heliconius* reflect, at least in part, the transfer of genomic regions through hybridization[54]. The curved arrows connect species that share similar alleles of the wing colour pattern gene through hybridization. All species shown are from the genus *Heliconius*. Monarch butterfly caterpillar image courtesy of J. de Roode, Emory University, USA; red milkweed beetle image courtesy of K. Rosenthal, Walker Nature Center, Reston Association, USA; oleander aphids image © B. MacQueen, Alamy Ltd; large milkweed beetle image courtesy of J. Pippen, Duke University, USA. Images in part **d** are reproduced, with permission, from REF. 91 © (2008) American Association for the Advancement of Science. Images in part **f** are reproduced, with permission, from REF. 54 © (2012) Macmillan Publishers Ltd. All rights reserved.

collateral genetic evolution. Finally, I discuss how the structure of genes and genetic regulatory networks might constrain genetic changes underlying phenotypic evolution and might thus contribute to parallel and collateral evolution.

Parallel and collateral evolution

The terms convergence and parallelism suffer from a confusing history of alternative usages, stemming from their original definitions to describe macroevolutionary patterns[9]. Originally, convergence meant that 'unrelated' species had evolved a similar solution from different ancestral states, and parallelism meant that 'related' species evolved a similar solution from 'the same' ancestral state. It is easy to imagine the confusion sown by these definitions.

Here, I reserve the term convergence to mean that different populations or species evolved similar phenotypic solutions[1]. I use the term parallel genetic evolution only for convergent evolution at the level of the mechanisms that generate phenotypes. As parallel genetic evolution involves the independent origination of variants through new mutations, it can provide evidence that particular variants have been favoured by selection over other variants that can confer similar phenotypic changes (if such other mutations exist), or that mutational bias frequently reintroduces particular variants, or both. As discussed in further detail below, the results of mutagenesis screens in multiple species imply that mutations in many different genes can alter the phenotype in similar ways[10,11]. This suggests that, in many species, multiple mutations are accessible that could provide more than one solution to many ecological challenges. Also, convergence often occurs through divergent genetic mechanisms in nature (BOX 2). Thus, parallel evolution of mutations in the same genes in different lineages suggests that evolution favours a biased subset of mutations in these cases.

In contrast to parallel evolution, collateral evolution can provide evidence for the selection of individual variants, but it provides less compelling evidence than does parallel evolution that these variants are objectively superior to other variants with respect to fitness. During parallel evolution, all mutations have the opportunity to be selected, and their likelihood of being exposed to selection is proportional to their mutation rate. By contrast, during collateral evolution, populations do not need to wait for new mutations to arise, and pre-existing alleles can be selected even if alternative alleles would have provided superior fitness improvements[12]. In some cases, alleles that have been selected during previous bouts of selection can accumulate additional mutations that enhance their beneficial phenotypic effects and/or ameliorate deleterious effects, generating 'super-alleles' (REFS 13–15). These super-alleles may then be favoured in multiple descendant populations.

As parallel evolution and collateral evolution represent different historical processes, they leave distinct phylogenetic signatures. At the level of an individual nucleotide position, the phylogenetic signals of parallel and collateral events can be identical, but the events can be distinguished by examining the DNA sequence that surrounds a focal position. After parallel evolution, a phylogeny reconstructed from all the sites in a gene region will generally be congruent with one reconstructed from other genes in the genome, except in rare cases in which parallel evolution involves multiple identical nucleotide changes. By contrast, collateral evolution by hybridization can be detected when the phylogenetic tree reconstructed from a focal gene is inconsistent with the phylogenies inferred from other genes in the genome. A gene that has undergone collateral evolution will support a phylogeny that reflects the history of the gene transfer events between species, but other genes in the genome that were introduced during the hybridization event can be quickly lost in the face of persistent backcrossing. This leaves an anomalous phylogenetic history for only the focal gene and the closely linked genomic regions. Collateral evolution by ancestry can be detected if the focal mutations are found to have already been present in the ancestral population. This pattern can be detected when, for example, multiple related species retain the polymorphism but a subset of species has fixed the same subset of alleles. It is easier to detect collateral evolution by ancestry when the ancestral population is still extant. Species with a large central population and peripherally isolated, recently diverged descendent populations

Box 1 | **Methods for studying parallel and collateral evolution**

Five classes of methods are available for studying parallel and collateral evolution.

Experimental evolution

Experimental evolution involves the maintenance of one or more populations (starting with a single isogenic strain) in a new environment over many generations. Whole-genome resequencing allows genome-wide 'snapshots' of evolving populations to be captured. Experimental evolution combined with whole-genome resequencing provides a powerful experimental paradigm for testing components of evolutionary theory and will soon be practical even for many multicellular organisms. The major drawback of this method is that evolved populations have responded to artificially imposed and often strong selection, usually over a short period of time, compared with the time span of evolution in the wild. A second drawback is that experimental evolution is usually carried out on species that reproduce asexually, in which clonal interference might limit the prevalence of parallel evolution.

Association studies

Most of our current knowledge about parallel evolution in multicellular organisms comes from studies of single candidate genes. These association studies involve a search for correlations between DNA sequence variants and phenotypic variation within a population. The major drawback of such candidate-gene studies is that the relative contribution of the candidate gene versus other genomic regions to the phenotypic variation is not known. Thus, it may be difficult to infer whether evidence for repeated evolution of a single gene provides compelling evidence for an excess of either parallel or collateral evolution. In some cases, however, a deep understanding of protein function can provide considerable insight into the probable functional consequences of sequence variants[77]. Recently, advances in genotyping technology have allowed genome-wide association mapping[78]. Nonetheless, many of the caveats of single-locus association tests also apply to genome-wide association surveys, and additional functional tests are required to provide strong evidence to implicate individual loci and nucleotide changes.

Genetic studies

In principle, classical genetic crosses provide a powerful method to survey the entire genome for genetic regions that contribute to phenotypic differences both between strains and between species. The major advantages of genetic crosses are that the association between genotype and phenotype is tested explicitly and that the environment can be controlled. The disadvantages are that genetic approaches are reasonably time and resource consuming; at the moment, they are primarily limited by the ability to generate and rear many recombinant individuals (which can be challenging and expensive for some species) and the ability to generate robust phenotypic measures of these individuals. This approach has worked best in plants and other organisms in which recombinant inbred strains can be generated[79–81]. To demonstrate parallel or collateral evolution, one must identify multiple pairs of strains or species that can be hybridized and used in genetic experiments.

Another challenge results from the fact that most evolved phenotypic variation is caused by genetic differences at multiple loci. Quantitative trait locus (QTL) mapping can be used as a first step to identify broad genomic regions that contribute to phenotypic differences, but direct inference of parallel or collateral evolution from these QTL intervals is treacherous. Few of the QTLs identified over the past 20 years have been resolved to individual genes, and this remains a challenging method of identifying evolved loci, although in most cases it is not clear that alternative approaches are superior. QTL mapping can be carried out within or between species. In crosses of different species, several more problems are likely to arise. First, different species are normally incompletely fertile at best, which can make genetic crosses technically challenging. Second, different species often differ for large chromosomal inversions, which will thwart fine-scale mapping efforts in these regions. Third, as QTLs are resolved into their individual loci, the magnitude of the phenotypic effect conferred by a single locus may be so small that it requires the measurement of multiple individuals that are isogenic for a single recombination event. The final two issues can also confound intraspecific genetic mapping. Despite these challenges, genetic crosses provide a powerful approach for detecting loci that contribute to parallel evolution.

Transgenic assays

An interesting option for assaying parallel evolution is to carry out transgenic assays to move candidate genomic regions between species to test for evolved functions[82]. This approach has not been widely adopted, but steady improvements in transgenic technology are making this an increasingly attractive possibility.

Genome scans to detect collateral evolution

Genome-wide comparisons of allele frequencies between recently diverged populations can be used to identify physically contiguous regions of divergent allele frequencies between populations[48]. These regions may contain genes that are related to speciation or ecological adaptation. This method has become possible only recently with the development of affordable whole-genome resequencing technology. Between any pair of populations, genomic regions that exhibit strong divergence relative to the average divergence across the genome provide initial evidence that these genomic regions might have responded to natural selection in at least one of the populations. If the same region is observed to have diverged in multiple pairs of populations, then the evidence for selection is strengthened. When applied to multiple pairs of populations, this approach provides a powerful means of discovering collateral evolution. This method does not allow robust discovery of regions that experience parallel genetic evolution or divergent evolution which is specific to each population. In contrast to genetic methods, the connection between putatively selected regions and phenotypic differences is not explicitly assayed, and additional work is required to confirm these connections, including traditional genetic and transgenic assays.

Box 2 | **Why does parallelism not always occur?**

Despite the extensive and growing evidence for the importance of parallel evolution, it is thought that in many cases convergence tends to occur through the evolution of different genetic mechanisms in different lineages.

An example of parallelism is seen in the resistance to cyclodiene in at least six insect species, which has occurred through identical amino acid substitutions in the target of the insecticide, the GABA receptor[83]. By contrast, insect resistance to dichlorodiphenyltrichloroethane (DDT) has evolved through changes in multiple mechanisms, including upregulation of P450 monooxygenases, dehydrochlorination through upregulating a glutathione S-transferase, mutations in the voltage-gated sodium channel that is the target of DDT and changes in an unidentified *trans*-regulatory factor[84,85].

Similarly, experimental evolution of *Escherichia coli* coupled with whole-genome sequencing has revealed different patterns of genetic evolution depending on the selective regimen[86]. When populations were selected for growth in minimal media with glycerol, mutations in two genes predominated in five replicate populations and accounted for the majority of the evolutionary response. By contrast, when 11 populations were evolved in minimal media with lactate, 33 genes carried mutations, and most mutated genes occurred in only one population. It seems that the genetic response to selection might depend on the precise nature of selection. It is not clear why in some circumstances we observe parallel evolution, whereas in other cases we do not. This remains a substantial challenge for the future.

provide a scenario that is favourable for detecting collateral evolution by ancestry.

Examples from diverse taxa

A list of examples of parallel and collateral evolution is provided in TABLE 1, and additional examples are described in other reviews on the topic[3,5–8,16–18]. Here, I discuss several examples that demonstrate the major trends emerging from recent work.

Evidence for parallel evolution from experimental-evolution populations. Excellent evidence for genetic parallelism comes from experimental-evolution studies in which the full genome sequences of evolved strains have been determined[19]. These experiments are carried out by growing replicate populations of a single clone in one or more environments and then tracking the fate of newly arisen mutations. Mutations with no fitness effects can spread through these populations by stochastic processes at a slow rate, but mutations with positive fitness effects can spread rapidly. Strong evidence for parallel evolution in these experiments comes from the observation of repeated evolution of mutations in the same gene[20] or even affecting the same amino acid sites[21–23].

One study carried out experimental evolution of the opportunistic pathogen *Pseudomonas aeruginosa* under laboratory culture conditions that mimic cystic fibrosis lung infection in replicate populations with and without antibiotics[24]. Among 24 genotypes that evolved in the presence of the antibiotic ciprofloxacin, 44 genes carried a total of 98 mutations, 77 of which were unique. Multiple cases of parallel evolution at the gene level were observed: 20 mutations were found in the transcriptional regulator gene *nfxB*, four in DNA gyrase subunit A (*gyrA*), nine in *gyrB* and seven in the putative glycosyl transferase gene (*orfN*). Clinical isolates of *Pseudomonas aeruginosa* sometimes harbour *nfxB*, *gyrB* and *gyrA* mutations that confer fluoroquinolone resistance, providing further support for the importance of parallel changes in the evolution of antibiotic resistance in *Pseudomonas aeruginosa*. Mutations in other genes that modified the phenotypic effects of the original mutations were selected during the experiment and these modifiers ameliorated the fitness costs of resistance, such that in the absence of an antibiotic there was no correlation between the level of antibiotic resistance and growth. This result is worryingly reminiscent of observations made in clinical isolates, in which resistant strains do not experience a fitness deficit. This pattern of parallel evolution in combination with unique changes in multiple other genes has also been observed in other experimental-evolution populations[25,26].

Sometimes, gene duplication has contributed to parallel evolution during experimental evolution. One study selected six lines of *Escherichia coli* at high temperatures[27], three lines of which evolved duplications in the same genomic region. In two lines, the duplications seemed to result from the same complex homologous recombination events involving repetitive elements. Thus, mutational bias may have increased the probability that similar duplications arose in replicate lines.

A similar selection experiment was carried out on 115 populations[28], but one genome from each population was then sequenced. Few specific mutations were shared between replicate populations, but parallel evolution was observed at multiple functional levels: genes, operons and functional complexes of genes (FIG. 2a). With this substantial sample size, the authors observed both negative and positive nonrandom associations between mutations in different genes among lines (FIG. 2b). This is best explained by negative and positive epistasis, in which substitutions in one gene influenced the probability that mutations in a second gene were selectively favoured.

Another example emphasizes the importance of genetic interactions during parallel evolution. In multiple replicate populations of yeast (*Saccharomyces cerevisiae*) that were evolved under glucose limitation, two mutations — a nonsense mutation in *MTH1* and amplification of the tandemly arrayed glucose transporter genes *HXT6* and *HXT7* — occurred repeatedly in replicate populations[29]. In no single population, however, did both mutations occur together. When both mutations were combined in a single strain, the strain displayed significantly lower fitness than the ancestral strain displayed before selection. Thus, although both mutations confer a fitness advantage, the combination of both mutations is incompatible in a population that is evolving in response to glucose limitation. These kinds of interactions between mutations can limit evolutionary trajectories, as has been documented for five mutations in the *Escherichia coli* β-lactamase gene (*ampC*) that greatly increase antibiotic resistance[30].

Taxonomically widespread parallel evolution. In addition to these experimental-evolution studies, several studies of multicellular organisms have revealed cases of parallel genetic evolution[3].

Operons
Loci consisting of two or more genes that are transcribed as a unit and expressed in a coordinated manner.

Epistasis
In the context of quantitative genetics: any genetic interaction in which the combined phenotypic effect of two or more loci is less than (negative epistasis) or greater than (positive epistasis) the sum of the effects at each individual locus.

Evolutionary trajectories
In the context of this Review: the series of mutations substituted during adaptation.

Table 1 | **Selected examples of parallel and collateral genetic evolution**

Species	Kingdom	Taxonomic level	Phenotype	Types of evolution	Genes	Type of gene	Refs
ΦX174	Virus	Intraspecific (experimental evolution)	Adaptation to high temperature and a novel host	Parallel evolution	Multiple genes	NA	21,22
HIV	Virus	Intraspecific	Antiretroviral resistance	Parallel evolution	Reverse transcriptase gene	Effector	88,89
Escherichia coli	Monera	Intraspecific (experimental evolution)	Adaptation to glucose-limited medium	Parallel evolution	Multiple genes	NA	20
		Intraspecific (experimental evolution)	Adaptation to glycerol-based medium	Parallel evolution	Glycerol kinase (*glpK*) and RNA polymerase genes	Effector	86
Pseudomonas aeruginosa	Monera	Intraspecific (experimental evolution)	Adaptation to novel environments	Parallel evolution	Multiple genes	NA	24
		Intraspecific (experimental evolution)	Hyperswarming	Parallel evolution	Flagella synthesis regulator (*fleN*)	Regulatory	23
Saccharomyces cerevisiae	Fungi	Intraspecific (experimental evolution)	Adaptation to fluctuating glucose and galactose levels	Parallel evolution	*GAL80*	Regulatory	93
Diverse species of yeast	Fungi	Interspecific	Loss of galactose utilization	Parallel evolution	*GAL* genes	Regulatory	94
Ipomoea horsfalliae and *Ipomoea quamoclit*	Plantae	Intergeneric	Evolution of red flowers from blue flowers	Parallel evolution	Flavonoid 3′-hydroxylase	Effector	95
Arabidopsis thaliana and *Arabidopsis lyrata*	Plantae	Intraspecific and interspecific	Vernalization	Parallel evolution	*FRIGIDA*	Regulatory	96,97
Plants (multiple species)	Plantae	Interspecific	C_4 photosynthesis	Parallel evolution	Phosphoenolpyruvate carboxylases (PEPC) genes	Effector	43,98, 99
Alloteropsis spp. grasses	Plantae	Interspecific	C_4 photosynthesis	Collateral evolution by hybridization	PEPC and phophoenolpyruvate carboxykinase genes	Effector	52
Human (*Homo sapiens*)	Animalia	Intraspecific	Resistance to malaria	Parallel evolution	Glucose-6-phosphate dehydrogenase (*G6PD*)	Effector	100
		Intraspecific	Lactase persistence	Parallel evolution	Lactase (*LCT*)	Effector	101
Colobine leaf-eating monkeys (*Pygathrix nemaeus* and *Colobus guereza*)	Animalia	Interspecific	Enhanced digestive efficiency	Parallel evolution	RNase gene	Effector	90
Cave fish (*Astyanax mexicanus*)	Animalia	Intraspecific	Albinism	Parallel evolution	Oculocutaneous albinism II (*Oca2*)	Effector	102
		Intraspecific	Reduced pigmentation	Parallel and collateral evolution by ancestry	Melanocortin 1 receptor (*Mc1r*)	Regulatory	103
Cichlid species from Lake Tanganyika and Lake Malawi	Animalia	Interspecific	Spectral sensitivity	Parallel evolution	Rhodopsin gene	Effector	104
Pufferfish (*Takifugu rubripes* and *Tetraodon nigroviridis*) and clam (*Mya arenaria*)	Animalia	Interspecific	Tetrodotoxin resistance	Parallel evolution	Sodium channel gene	Effector	105, 106
Drosophila spp.	Animalia	Interspecific	Trichome patterning	Parallel evolution	*shavenbaby*	Regulatory	32,33, 35
Insects (multiple species)	Animalia	Interspecific	Cardenolide resistance	Parallel evolution	$(Na^+ + K^+)$ ATPase gene	Effector	37,38

Table 1 (cont.) | **Selected examples of parallel and collateral genetic evolution**

Species	Kingdom	Taxonomic level	Phenotype	Types of evolution	Genes	Type of gene	Refs
Stickleback (*Gasterosteus aculeatus*)	Animalia	Intraspecific	Pelvic spine and girdle reduction	Parallel evolution	Paired-like homeodomain transcription factor 1 (*Pitx1*)	Regulatory	50,107, 108
Sticklebacks (multiple species)	Animalia	Interspecific	Lateral plates	Parallel and collateral evolution by ancestry	*Ectodysplasin*	Regulatory	45,108
Mouse (*Mus musculus*)	Animalia	Interspecific	Warfarin resistance	Collateral evolution by hybridization	Vitamin K epoxide reductase complex, subunit 1 (*Vkorc1*)	Effector	53
Butterflies (*Heliconius* spp.)	Animalia	Interspecific	Wing colouration patterns	Collateral evolution by hybridization	*optix*	Regulatory	54

NA, not applicable.

The mouse-ear cress, *Arabidopsis thaliana*, must normally undergo a period of exposure to cold temperatures, called vernalization, to induce flowering in the spring. In many populations, however, plants have been selected to forgo vernalization and to flower immediately without experiencing cold temperatures. At least 20 times, independent mutations that incapacitate the gene *FRIGIDA*, which regulates vernalization, have spread in local subpopulations of *Arabidopsis thaliana*. Variation at *FRIGIDA* explains ~70% of the variation in flowering time in this species[31]. This example demonstrates parallel evolution between populations of a single species.

A second example illustrates parallel evolution at a higher taxonomic level, between species of a single genus. Both *Drosophila sechellia* and *Drosophila ezoana*, which diverged approximately 40 million years ago, evolved a novel pattern of larval trichomes in which naked cuticle is produced instead of the ancestral state of dense trichomes[13,32–35] (FIG. 3). In *Drosophila sechellia*, at least nine mutations in five transcriptional enhancers of *shavenbaby* (also known as *ovo*) have all contributed to the loss of trichomes[32]. In *Drosophila ezoana*, mutations in at least two *shavenbaby* enhancers that are homologous to the *Drosophila sechellia* enhancers have also evolved to contribute to the loss of trichomes, revealing parallel evolution at the level of individual transcriptional enhancers in widely divergent species[35].

Parallel evolution has been observed at an even higher taxonomic level, between insect orders. Many insect species have evolved to feed on plants that produce toxic cardenolides, which bind to and block the $(Na^+ + K^+)$ATPase pump[36]. In species belonging to four orders of insects spanning more than 300 million years of evolution, precisely the same substitution at one amino acid position has contributed to cardenolide resistance[37,38]. At a second amino acid position, various substitutions that confer cardenolide resistance have evolved at least 12 times[37,38]. In addition, in four orders of insects, the $(Na^+ + K^+)$ATPase gene has been duplicated, and the two paralogues now show differential expression between the brain and gut[38]. In these species, most of the parallel amino acid substitutions occurred in the paralogue that is expressed in the gut. This demonstrates how gene duplication can confer greater phenotypic specificity on two copies of a single gene that was previously ubiquitously expressed, and can thus presumably reduce the pleiotropic consequences of amino acid substitutions that confer resistance to a toxin. Such gene duplication of a multifunctional gene followed by specialization has been called 'escape from adaptive conflict', and several examples seem to support this model[39–41].

Other examples illustrate that gene duplications have enabled parallel evolution in other ways. Both the Antarctic notothenioid fish and the Arctic cod have evolved extremely similar antifreeze proteins. Despite these amino acid similarities, the genes in notothenioids and Arctic cod were derived from different gene duplication events. After the duplications, the genes independently evolved very similar amino acid sequences with apparently identical ice-binding functions[42].

Similarly, C_4 photosynthesis has evolved many times in plants and has required changes in a key gene that encodes phosphoenolpyruvate carboxylase (PEPC). PEPC genes occur in multigene families, only one of which is involved in C_4 photosynthesis. At least eight times in grasses, PEPC genes have independently evolved similar or identical key amino acid changes that support C_4 photosynthesis[43].

Gene duplication itself sometimes causes parallel evolution. For example, the number of copies of the salivary amylase gene (*AMY*) has increased in multiple independent human populations, apparently in response to the development of high-starch diets[44].

Collateral evolution through shared ancestry. Collateral evolution through shared ancestry might be very common, as has been documented in stickleback fish. Multiple freshwater populations of stickleback fish have evolved convergent loss of lateral ectodermal plates, which serve as body armour. Freshwater populations of sticklebacks are derived from marine sticklebacks, all of which have extensive body armour. In their new freshwater homes, natural selection repeatedly resulted in the evolution of stickleback populations with little body armour. In most populations, reduced body armour resulted from repeated fixation of the same ancestral

Trichomes
Thin, cuticular and non-sensory processes that are secreted by individual cells.

Enhancers
Regulatory DNA elements that usually bind several transcription factors; they can activate transcription from a promoter at a great distance and in an orientation-independent manner.

Paralogues
Genes in the same organism that have evolved from a gene duplication, usually with a subsequent, and sometimes subtle, divergence of function.

Pleiotropic
Pertaining to a gene having multiple developmental roles or to a mutation having multiple phenotypic effects.

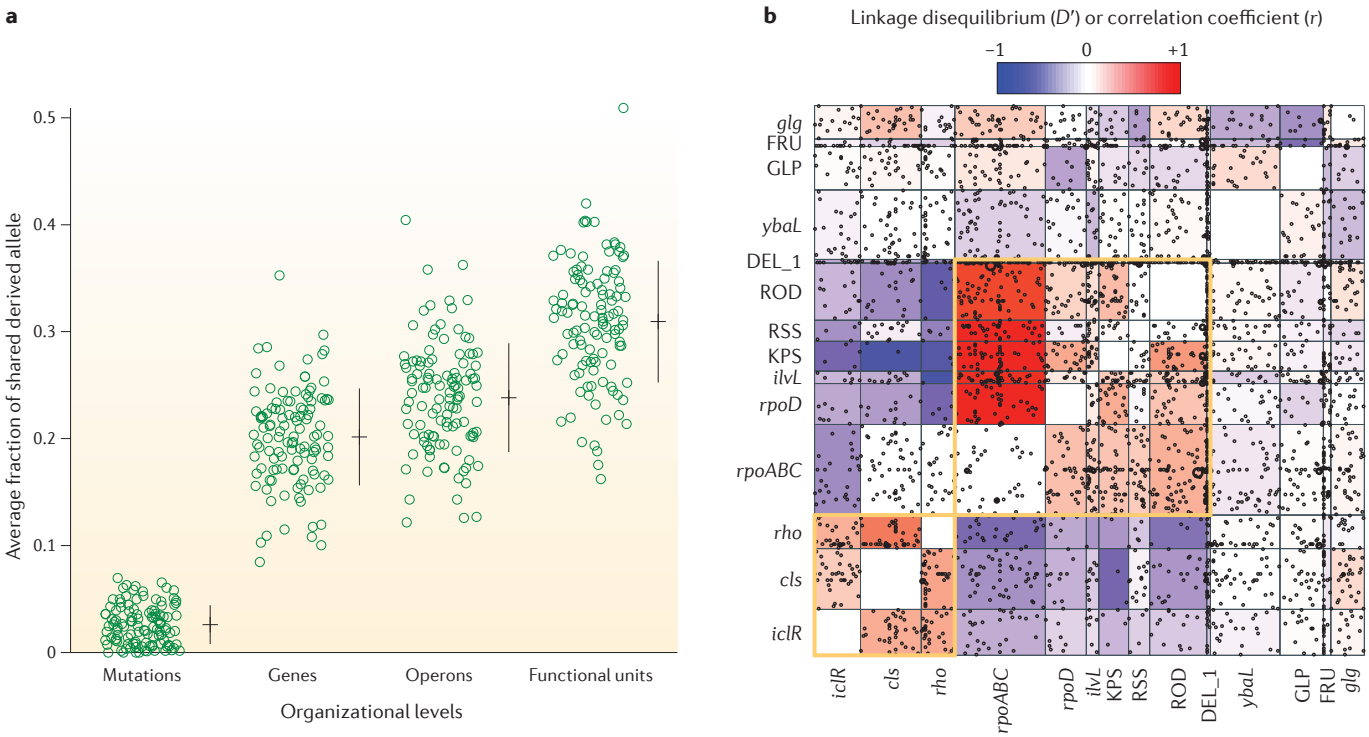


Figure 2 | **The landscape of parallel evolution.** **a** | Pairwise fraction of shared events from genome resequencing of 115 replicate *Escherichia coli* populations evolved to high temperature, shown for different organizational levels, as indicated on the *x* axis. Each circle represents the value for a replicate population, and cross bars show means ± one standard deviation. Few identical mutations occurred in replicate populations, but many parallel mutations were selected in the same genes and higher organizational levels. **b** | A plot of the co-occurrence of mutations among these replicate *Escherichia coli* lines. Each point represents a mutation in the corresponding unit (gene, operon or functional unit) on both axes. Relative excess of co-occurring mutations in some regions can be quantified as higher linkage disequilibrium (*D′*; upper diagonal) or correlation coefficient (*r*; lower diagonal). A relative deficiency of co-occurring mutations leads to *D′* or *r* values that are lower than expected. The two yellow boxes emphasize genes with an excess of positive associations within the boxes and an excess of negative associations between the boxes. This pattern can be interpreted as gene spaces that represent alternative evolutionary strategies. Part **a** is based on data from REF. 28; part **b** is modified, with permission, from REF. 28 © (2012) American Association for the Advancement of Science.

allele and, in one case, from independent evolution of a new *Ectodysplasin* allele[45].

The widespread *Ectodysplasin* allele that generates the low-plated phenotype is present at low frequency in marine populations, probably because it was introduced from freshwater populations in previous generations[45,46]. Marine sticklebacks breed in freshwater and therefore have multiple opportunities to encounter resident freshwater populations[47]. Alleles that were favoured in freshwater populations might have been introduced into marine environments during hybridization events. One can imagine many cycles of selection in freshwater habitats followed by allele leakage back into the marine environment occurring over a long period of time and in parallel in many thousands of locations around the globe, although there is currently no direct evidence for this model. This population structure and history would generate a marine population carrying multiple 'freshwater' alleles at low frequency, which would provide many opportunities for collateral evolution when new freshwater populations were established from the marine population.

Recently, genome-wide resequencing studies of sticklebacks have begun to reveal the full extent of collateral genetic evolution in this species[48,49]. One study included 20 individual sticklebacks selected from geographically diverse pairs of neighbouring marine and freshwater environments[48]. Most of the genome showed patterns of genetic differentiation that are consistent with neutral segregation of alleles between marine and freshwater populations. However, ~100–200 genomic regions encompassing <0.5% of the genome exhibited strong differentiation of allele frequencies that was associated with the habitat of the fish, and these are candidate regions for collateral evolution. Reassuringly, the region containing the *Ectodysplasin* gene was highly differentiated in this sample.

This survey provides an estimate of the proportion of the genome that has contributed to collateral evolution, but it does not provide an estimate of the proportion of the genome that has contributed to divergent local adaptation in different populations. It also does not allow an estimate of phenotypic convergence through parallel genetic evolution, which has also occurred

Resequencing
Determination of an exact DNA sequence by comparison with a known reference.

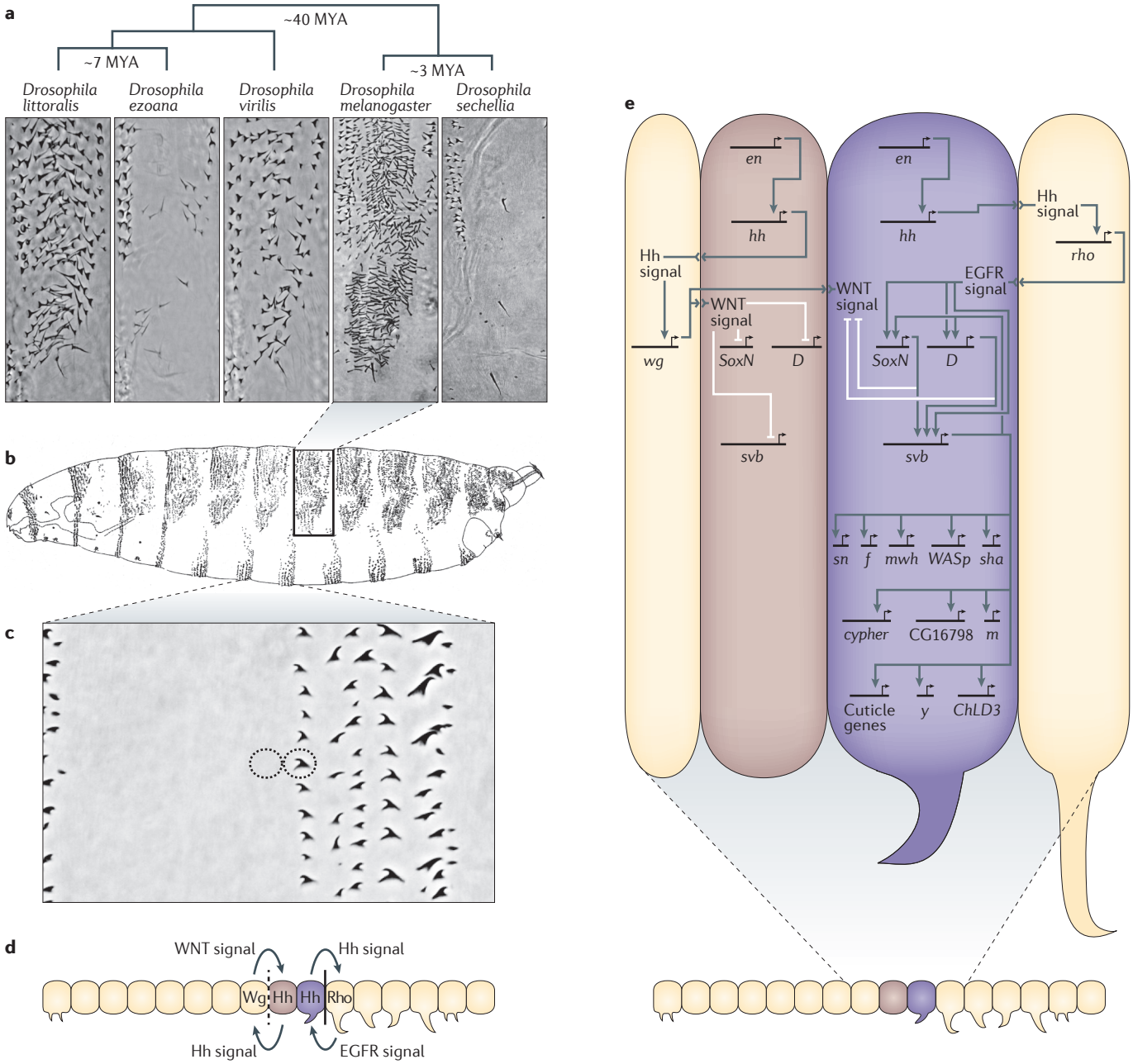


Figure 3 | **The structure of developmental networks can influence which genes underlie phenotypic evolution.** It is useful to explore developmental networks in an explicit cellular framework as 'pathworks', which highlights the roles of key input–output genes in development[5]. **a** | *Drosophila ezoana* and *Drosophila sechellia* independently evolved the loss of trichomes on the dorsal and lateral surface of first instar larvae. In both species, the evolution of dorso–lateral naked cuticle resulted from parallel evolution of orthologous *cis*-regulatory enhancers of the *shavenbaby* (*svb*) locus[13,35]. **b** | The first-instar larva of *Drosophila melanogaster* exhibits a complex pattern of trichomes. **c** | A magnified view of the ventral cuticle of a single abdominal segment, illustrating the locations of two neighbouring cells (dashed circles) that have experienced the pathwork leading to trichome differentiation in different ways. **d** | Although both cells outlined in part **c** express *hedgehog* (*hh*), they receive different signals from their respective neighbours. A Wingless (Wg) signal from the anterior ultimately causes repression of *svb* transcription, preventing trichome differentiation. By contrast, Rho, an epidermal growth factor receptor (EGFR) signal, from the posterior ultimately activates *svb* transcription, resulting in the upregulation of a cascade of genes that contribute to forming a trichome. **e** | An examination of some of the genes in the trichome formation pathwork in the two focal cells identifies *svb* as a key gatekeeper of trichome differentiation. Genes acting upstream of *svb* influence multiple other processes in these cells and their neighbouring cells; mutations in these genes, even cell type-specific changes caused by *cis*-regulatory mutations, would influence multiple processes. Mutations in genes acting downstream of *svb* cannot, on their own, cause the discrete switch between naked cuticle and trichomes[63]. By contrast, evolution of the *cis*-regulatory elements that regulate *svb* in these cells is likely to minimize deleterious pleiotropic effects of evolved changes while maximizing the phenotypic outcome[17,18,92]. This may help to explain the accumulation of many evolutionarily relevant mutations in the *svb* *cis*-regulatory region in the two species. Grey arrows indicate activation and white lines indicate repression. *D*, *Dichaete*; *en*, *engrailed*; *f*, *forked*; *m*, *miniature*; *mwh*, *multiple wing hairs*; MYA, million years ago; *rho*, *rhomboid*; *sha*, *shavenoid*; *sn*, *singed*; *SoxN*, *SoxNeuro*; *y*, *yellow*. Figure is modified, with permission, from REF. 5 © (2010) Roberts & Company Publishers.

in stickleback fish at the paired-like homeodomain transcription factor 1 (*Pitx1*) locus[50,51]. To address this issue, the authors sequenced two fish with divergent phenotypes from the extremes of an ecological gradient[48]. Among the top 0.1% of the most highly diverged regions between these two fish, ~35% included globally shared regions that are divergent between marine and freshwater stickleback populations. Thus, many, but far from all, of the locally divergent regions are also found in comparisons of divergent populations worldwide, suggesting that many genomic regions have either evolved in response to selection pressures that are unique to these populations or converged using genetic mechanisms that are not shared globally among stickleback populations.

Collateral evolution through hybridization. In the past few years, multiple examples of collateral evolution through hybridization have been detected in several taxa. Given this recent burst in the number of discoveries, it is likely that collateral evolution by hybridization is widespread in nature. Collateral evolution through hybridization has been observed in the transfer of two genes that are key elements of C_4 photosynthesis between species of the grass genus *Alloteropsis*[52]. Such transfers are estimated to have occurred at least four times. Similarly, an allele of vitamin K epoxide reductase complex, subunit 1 (*Vkorc1*) that evolved in *Mus spretus* and confers resistance to the rodenticide warfarin has spread into populations of *Mus musculus domesticus*[53].

A recent genome-wide survey of *Heliconius* spp. butterflies has provided compelling evidence for collateral evolution through species hybridization[54]. *Heliconius* is a genus of neotropical butterflies that are famous for their extensive Müllerian mimicry complexes. In Müllerian mimicry, multiple species that are distasteful to predators have evolved similar warning colouration, which allows these species to share the cost of 'educating' predators about the association between colour patterns and unpalatability. *Heliconius* spp. butterflies have evolved bold patterns of red, black, orange and white wings, and extremely similar combinations of these patterns are found in different species of the genus.

The loci controlling two of these colour patterns, a red and a yellow stripe, were previously mapped between two *Heliconius* spp.[55] Application of recently developed tests for introgression[56,57] to the genome-wide genotyping data for four mimetic species provided strong evidence that the mimetic loci had been introgressed through hybridization between two pairs of species (FIG. 1f). Extending this work even further, a more distantly related species with a similar, but not identical, mimetic pattern seems to share the same mimetic alleles through hybridization. Although several of the relevant genes causing mimicry have been identified in these butterflies, the precise molecular changes that cause mimetic phenotypes are not yet known. For example, it is possible that these mimetic alleles consist of super-alleles containing many individual nucleotide changes, as has been shown for mimetic alleles in another *Heliconius* sp.[15,58].

Effective number of participating genes

Although there are now multiple examples of parallel and collateral evolution[3], the key question is whether these types of evolution occur more often than expected by chance, or to put it another way, does convergent evolution involve a nonrandom subset of genetic changes? The probability of gene reuse underlying phenotypic convergence has been estimated to be 0.32–0.55 (REF. 16). This estimate is subject to many caveats[16], including the fact that this is a per-gene estimate. If a single gene harbours multiple substitutions (such genes have been called 'intralineage hot spots' (REF. 3)) then this may be a dramatic underestimate of the probability of gene reuse during convergence. Nonetheless, this estimate provides a starting point. We can estimate the effective number of genes participating in convergent genetic evolution as the inverse of the probability of convergence[16,59]. Then, we can compare this number (~2–3) with the number of genes that contribute to building phenotypic features during development. It is obvious that all developmental features require the activity of more than three genes, but we can generate more precise estimates for cases in which the contributions of individual genes to particular phenotypes have been revealed through mutagenesis experiments.

For example, a recent study reported that all experimental-evolution populations of *Pseudomonas aeruginosa* that evolved hyperswarming did so through substitutions in flagella synthesis regulator (*fleN*)[23], whereas a mutagenesis screen for swarming defects in this species identified 233 genes[60]. In the study of experimental evolution in the context of ciprofloxacin resistance in *Pseudomonas aeruginosa*[24], 40 of 98 mutations occurred in parallel in four genes, whereas a previous mutagenesis screen identified 114 genes that confer ciprofloxacin resistance in this species[61]. Similarly, in *Arabidopsis thaliana*, in which 70% of natural flowering time can be explained by variation at *FRIGIDA*, mutations in at least 80 genes can influence flowering time, and in a mutagenesis screen[10], only three of 50 mutations occurred in *FRIGIDA*[62]. In *Drosophila sechellia*, loss of trichomes is entirely attributable to at least nine mutations at the *shavenbaby* locus, but at least dozens of genes can influence trichome patterning in *Drosophila* spp.[11,63]. Finally, in sticklebacks, in which variation at the *Ectodysplasin* gene accounts for ~75% of the variation in armour plate number in an F_2 cross, variation in four other components of the *Ectodysplasin* signalling pathway does not contribute to phenotypic variation[64].

Thus, it seems that parallel and collateral evolution involve a restricted subset of the genes that contribute to the development of particular phenotypic features. Why might this be?

Why does parallelism occur?

For the remainder of this Review, I use the word 'locus' in a specific way that is not typically used in genetics (in which it is often taken to mean 'gene') but is closer to its original meaning; hereafter, locus refers to a contiguous region of DNA that encodes a specific function. It is possible to imagine a hierarchy of loci, for example, an operon

cis-regulatory loci
Genetic loci containing transcription factor-binding sites and other non-coding DNA elements that are sufficient to activate transcription in a defined spatial and/or temporal expression domain.

Slippage
A mutagenic process during DNA replication whereby the presence of several identical base pairs in a series causes the DNA polymerase to add or omit one base by sliding over the template.

containing multiple genes that confer a specific function, a DNA sequence that encodes all the *cis*-regulatory loci and coding information for a single protein, a single exon that encodes a protein subdomain, a transcriptional enhancer, a single transcription factor binding site, a single codon or a single nucleotide position.

From a population genetics perspective, three factors influence the probability that a locus will contribute to parallel evolution[59]: the mutation rate of the locus, the probability that mutations at the locus are net beneficial and the average magnitude of the fitness change caused by these mutational effects. The first parameter, the locus-specific mutation rate, is the product of the site-specific mutation rate and the mutational target size. I consider the second and third parameters together, because both are derived from the functional role of the locus in the cell and in development. It is important to recognize, however, that not all new advantageous alleles will automatically contribute to adaptation. Instead, the probability that a mutation will spread through a population scales with the magnitude of the net fitness improvement conferred by the mutation[65]. Additional factors, such as genome size and genome complexity, can influence the 'precision' of parallel evolution (BOX 3).

Both the probability that mutations are beneficial and the magnitude of the fitness change caused by a mutation are not invariant. In many, and perhaps all, cases, these effects depend on the genetic and environmental context. If we consider adaptation as the result of a series of substitutions[66], the adaptive value of a new mutation can depend on the particular mutations that were substituted earlier[67]. This temporal dependence of fitness effects is likely to constrain adaptive paths and enhance the probability of parallel evolution. It will also lead to the observation that any single evolving population can explore only one of several possible adaptive strategies (FIG. 2b), depending on which mutations arose first[28,30].

Mutation. Even when mutations occur randomly in a genome and in a population of finite size, it is unlikely that all possible mutations will be available at all times. In addition, mutations sometimes occur nonrandomly. For example, simple repeat regions are susceptible to slippage during replication, repetitive regions can mediate homologous recombination that generates deletions, and CpG dinucleotides are mutational hot spots in mammalian genomes[68]. In some cases, these increased mutation rates can contribute to parallel evolution[27,28,50]. Therefore, the locus-specific mutation rate can influence the probability of parallel evolution. For example, in multicellular organisms, non-coding regions, where most *cis*-regulatory loci reside, are often larger than coding regions. All else being equal, *cis*-regulatory loci may therefore provide larger mutational targets than coding loci. However, it is not yet clear whether, at the base-pair level, non-coding regions are as likely as coding regions to generate adaptive mutations. However, coding DNA requires a strict triplet code of nucleotides, whereas *cis*-regulatory DNA does not, which means that a wider variety of mutations (including SNPs, insertion–deletion events and rearrangements) can generate functional changes in *cis*-regulatory regions than in coding regions. For these reasons, it is possible that the mutation rate to functionally viable alternative alleles is higher for *cis*-regulatory regions than for coding regions, although this remains an area that requires further investigation.

In most cases, however, it seems unlikely that the mutation rate itself has limited the diversity of loci that are available for selection. For example, as discussed above, mutational screens usually reveal that many genes in the genome can be mutated to contribute to a particular phenotypic outcome. There seem to be many mutational paths available for eukaryotes, bacteria and archaea to respond to specific ecological challenges, and a biased subset of these mutations seems to be most frequently used during convergent evolution. In addition, even on theoretical grounds, it is unlikely that most populations are limited by the rate of mutation; even with a mutation rate of ~10^{-8} per base, the human population generates hundreds of mutations that are consistent with viability at every site in the genome every generation[69]. Species with even larger population sizes than humans are unlikely to experience strong mutational limitation.

Probability and magnitude of beneficial mutations. The probability that a mutation has a net beneficial effect on fitness depends on the array of phenotypic effects caused by the mutation because a mutation with a positive

Box 3 | **Precision of parallel evolution**

The precision of parallel evolution — whether parallel mutations occur at the same nucleotide position or just in the same locus — depends on multiple factors[19], including the size and functional complexity of the genome, and the functional mapping between individual molecular changes and the phenotype. The probability of parallel evolution was shown to be $2/(n+1)$, where n is the number of potentially adaptive mutations[87]. As small genomes tend to harbour fewer mutational targets that can increase fitness than do large genomes, parallel evolution through mutations of homologous sites is, correspondingly, expected to occur more commonly in organisms with smaller genomes. Similarly, even in large, complex genomes, mutations with exceedingly precise and useful consequences, such as mutations in a receptor that alters the binding of a poison, can confer considerably higher fitness on their bearers than n other mutations, and they might thus be subjected to repeated evolution in divergent taxa.

Although parallel evolution of the same genes has often been observed in experimental-evolution populations of bacteria and yeast, parallel evolution through identical mutations in a single gene is observed less often. In one example, HIV displayed remarkably precise parallel evolution in response to treatment with an antiretroviral drug. A series of the same four or five mutations occurred repeatedly, usually in the same order, in patients with AIDS who were receiving zidovudine[88,89]. In organisms with small genomes, such as HIV, there might be very few mutations that can confer an advantage in a new environment, which might help to explain the extraordinary precision of some evolutionary changes in HIV. In species with larger genomes, there may be more mutational paths to adaptation.

Nonetheless, highly specific parallel evolution is sometimes observed in eukaryotes. For example, two species of leaf-eating colobine monkeys, the Asian douc langur (*Pygathrix nemaeus*) and the African guereza (*Colobus guereza*), have independently evolved identical amino acid substitutions in duplicated ribonuclease genes[90]. These monkeys host symbiotic bacteria in their foregut. The bacteria ferment leaves and the monkeys digest the bacteria. These bacteria produce abundant RNAs, and the monkeys have evolved duplicated pancreatic ribonuclease genes to digest these RNAs. Three parallel amino acid substitutions in these duplicated genes lower the optimal pH for enzyme activity to more closely match the pH of the monkey foregut.

fitness effect on one trait might cause deleterious pleiotropic effects on other aspects of the phenotype. It is likely that the probability of mutations being beneficial decreases on average with increasing pleiotropy[70].

The pleiotropic effects of a mutation often depend on the location of the mutation within a locus. Mutations can occur in *cis*-regulatory regions or in protein-coding regions, where they can alter the encoded amino acid sequence. Mutations in coding regions have the potential to influence the function of the protein in every cell in which the protein is expressed. By contrast, most mutations in *cis*-regulatory DNA influence gene function in only a subset of the full expression domain of the gene. Primarily for this reason, *cis*-regulatory mutations will often have fewer pleiotropic consequences than mutations in protein-coding regions, and *cis*-regulatory regions might therefore contribute to morphological evolution more often than coding regions[48,71,72].

The magnitude of mutational effects that is important for evolution is the net fitness increment or decrement, not the size of the phenotypic alteration. Mutations that cause large phenotypic effects, such as many null mutations, may not be favoured by natural selection because pleiotropic effects on traits have antagonistic effects on fitness.

Some studies have provided the opportunity to assess all three factors — mutation, the probability of beneficial effect and the effect size — in a single selection regimen, albeit in a limited manner. For example, as discussed above, the experimental evolution of *Pseudomonas aeruginosa*[24] revealed multiple examples of parallel evolution. The authors of this study argued that all three population genetic factors had a role in parallel evolution in their experimental populations. For one locus, repeated observation of the same single base deletion in a homopolymeric region implicated a slippage mechanism in generating an increased mutation rate. At a second locus, loss-of-function mutations were beneficial, and because many possible mutations can generate loss-of-function alleles, the probability of beneficial mutations occurring at this locus was high. For a third locus, the observed mutations had strong fitness effects, leading to an increased probability of fixation.

Parallel evolution of regulatory and effector genes. It is useful to explore these ideas with respect to the positions of genes in regulatory networks. Genes can be broadly divided into regulatory genes and effector genes. Here, I consider these two classes of genes by viewing developmental networks in reverse, starting from the differentiated state and working backwards to earlier stages of development (FIG. 3). This approach allows the identification of paths through genetic networks that are specific to each phenotypic outcome, which I have called "pathworks" (REF. 5). Pathworks focus attention on the individual regulatory 'decisions' made by cells as they progress through development and differentiation. By contrast, genome-focused tracing of networks, which aims to illustrate all regulatory linkages in the genome, can obscure these cell type-specific regulatory architectures within vast 'spaghetti' plots. A focus on cell-based pathworks has led to the identification of 'hourglass' or 'bow tie' shapes in developmental networks, and these features can help to identify 'master regulators' of independent phenotypic outcomes and groups of effector genes downstream of these regulators[17,18,73]. These master regulators have been termed 'input–output' genes because they integrate multiple signals and regulate multiple downstream genes[74].

Theoretical considerations suggest that parallel evolution will occur more often at network locations that minimize pleiotropy and maximize the phenotypic changes[5,17,18], which is known as the hotspot hypothesis[3]. There are two locations in pathworks that can fit these criteria in different circumstances. First, input–output genes can often regulate discrete developmental alternatives largely on their own, which can both limit pleiotropy and generate a significant phenotypic change. Input–output genes often integrate regulatory information from multiple upstream regulatory genes and are often involved in regulating multiple developmental processes. Thus, mutational targets with specific effects are likely to reside in the *cis*-regulatory elements that drive specific transcriptional domains of input–output genes, rather than in the coding regions of these genes. For example, this may help to explain the parallel evolution of enhancers of the *shavenbaby* gene[34]. As individual regulatory genes can participate in multiple developmental processes, different *cis*-regulatory regions may be hot spots for different aspects of the phenotype. Similarly, regulatory genes that act upstream of input–output genes in one developmental process may themselves be input–output genes in a different process. Thus, the hotspot hypothesis for regulatory genes does not posit that limited numbers of regulatory genes in the genome are hot spots. Instead, this hypothesis is specific to each element of the phenotype. For different phenotypic features, different regulatory genes may serve as input–output genes and may thus be hot spots.

The second location in a pathwork at which genes can generate precise and substantial phenotypic effects on their own is that of downstream effector genes, of which there are many. Many effector genes have specific roles in development, behaviour or physiology such that mutations in these genes influence a very narrow subset of aspects of organismal phenotype. For example, mutations in odorant receptor genes can switch perception of discrete molecules, causing substantial changes in behaviour or physiology. In domesticated strains of both *Caenorhabditis elegans* and *Caenorhabditis briggsae*, similar deletions of neighbouring pheromone receptor genes cause resistance to pheromone-induced dauer formation[75].

Conclusions

Both parallel and collateral genetic evolution provide evidence that genetic evolution is historically predictable[3,5–8,17,18]. Recent studies provide many examples of parallel and collateral evolution, which support the hypothesis that genetic evolution displays some predictability. Quantitative studies of the probability of repeated evolution provide some support for this hypothesis[16].

Dauer
A developmentally arrested, immature, long-lived and non-feeding form of *Caenorhabditis elegans* that forms under conditions of food scarcity and high population density; it resumes development when food levels increase.

The current evidence for an abundance of parallel and collateral evolution comes from a large collection of targeted candidate gene surveys combined with a much smaller number of unbiased genome-wide surveys. The biggest challenge for the future is probably to generate more genome-wide data sets of the genetic causes of phenotypic convergence, especially in multicellular organisms (BOX 1). Only such data sets will provide the opportunity for quantitative, unbiased tests of the relative contributions of parallel, collateral and divergent evolution to convergence and the role of regulatory network structure in generating hot spots of parallel evolution. One substantial future concern is whether the genetic effects that are statistically detectable in genome-wide tests (including association tests and quantitative trait locus mapping) represent a biased, unusual set of loci[76] or are representative of the full distribution of loci that contribute to evolution[3].

If further studies continue to support the observation that parallel evolution has occurred more often than expected by chance, then we will require explicit tests of the causes of parallel evolution. I have discussed several properties of gene structure and genetic network structure that can influence the probability that different gene regions and certain genes within networks will contribute to parallel evolution. Both effector genes and genes that regulate developmental processes participate in parallel and collateral evolution (TABLE 1). At least in some circumstances, evolutionary predictability results from mutations that minimize pleiotropic effects while simultaneously maximizing the phenotypic change[5,6,17,18]. This may be a general principle of genetic evolution. Experimental-evolution studies and multiple cases of parallel evolution in the wild might provide data that are suitable to test such predictions. One important caveat that is hinted at by current data is that the probability of parallel and collateral evolution might be influenced not only by properties of genes and genetic networks but also by population genetics parameters, such as population size, population structure, and the strength and duration of natural selection[5,17,76]. Tests that explicitly measure both genetic and population genetics factors are most likely to provide insights into the multiple factors that can influence parallel and collateral evolution.

1. Endler, J. A. *Natural Selection in the Wild* (ed. May, R. M.) (Princeton Univ. Press, 1986).
2. Losos, J. B. Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840 (2011).
3. Martin, A. & Orgogozo, V. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**, 1235–1250 (2013).
A comprehensive survey of all published examples of repeated evolution. Many examples of parallel and collateral evolution are documented, along with a discussion of hotspot genes.
4. Shaw, A. B. Adam and Eve, paleontology, and the non-objective arts. *J. Paleontol.* **43**, 1085–1098 (1969).
5. Stern, D. L. *Evolution, Development, and the Predictable Genome* (Roberts & Co., 2010).
6. Gompel, N. & Prud'homme, B. The causes of repeated genetic evolution. *Dev. Biol.* **332**, 36–47 (2009).
7. Christin, P. A., Weinreich, D. M. & Besnard, G. Causes and evolutionary significance of genetic convergence. *Trends Genet.* **26**, 400–405 (2010).
8. Wood, T. E., Burke, J. M. & Rieseberg, L. H. Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**, 157–170 (2005).
9. Arendt, J. & Reznick, D. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* **23**, 26–32 (2008).
10. Levy, Y. Y. & Dean, C. The transition to flowering. *Plant Cell* **10**, 1973–1990 (1998).
11. Nusslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
12. Barrett, R. D. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
13. McGregor, A. P. *et al.* Morphological evolution through multiple *cis*-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).
14. Loehlin, D. W. & Werren, J. H. Evolution of shape by multiple regulatory changes to a growth gene. *Science* **335**, 943–947 (2012).
15. Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).
16. Conte, G. L., Arnegard, M. E., Peichel, C. L. & Schluter, D. The probability of genetic parallelism and convergence in natural populations. *Proc. Biol. Sci.* **279**, 5039–5047 (2012).
17. Stern, D. L. & Orgogozo, V. The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155–2177 (2008).
A comprehensive survey of the genetic changes contributing to evolution, revealing multiple examples of parallel evolution and suggesting that population genetic parameters influence the predictability of genetic evolution.
18. Stern, D. L. & Orgogozo, V. Is genetic evolution predictable? *Science* **323**, 746–751 (2009).
19. Dettman, J. R. *et al.* Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Mol. Ecol.* **21**, 2058–2077 (2012).
A summary of significant discoveries from whole-genome sequencing of experimental-evolution studies in microorganisms.
20. Woods, R., Schneider, D., Winkworth, C. L., Riley, M. A. & Lenski, R. E. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **103**, 9107–9112 (2006).
A beautifully designed experiment to test the importance of parallel evolution in experimental-evolution populations.
21. Bull, J. J. *et al.* Exceptional convergent evolution in a virus. *Genetics* **147**, 1497–1507 (1997).
A paper documenting many examples of parallel evolution that involve identical substitutions in evolving viral populations.
22. Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M. & Bull, J. J. Different trajectories of parallel evolution during viral adaptation. *Science* **285**, 422–424 (1999).
A study of viral populations which finds that although parallel evolution is common, the order of substitutions is less predictable.
23. van Ditmarsch, D. *et al.* Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. *Cell Rep.* **4**, 1–12 (2013).
A paper showing that *Pseudomonas aeruginosa* evolve multiple flagella when they are grown on an agar plate and that this is always because of parallel mutations in a single gene.
24. Wong, A., Rodrigue, N. & Kassen, R. Genomics of adaptation during experimental evolution of the opportunistic pathogen *Pseudomonas aeruginosa*. *PLoS Genet.* **8**, e1002928 (2012).
A study demonstrating that antibiotic resistance evolves mainly through parallel substitutions, but compensation for the costs of resistance evolves through diverse mechanisms.
25. Conrad, T. M. *et al.* Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biol.* **10**, R118 (2009).
26. Blount, Z. D., Barrick, J. E., Davidson, C. J. & Lenski, R. E. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**, 513–518 (2012).
27. Riehle, M. M., Bennett, A. F. & Long, A. D. Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **98**, 525–530 (2001).
28. Tenaillon, O. *et al.* The molecular diversity of adaptive convergence. *Science* **335**, 457–461 (2012).
This is a genomic analysis of 115 independent populations that were evolved for 2,000 generations; it provides a comprehensive documentation of parallel evolution at multiple functional levels.
29. Kvitek, D. J. & Sherlock, G. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* **7**, e1002056 (2011).
30. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
31. Shindo, C. *et al.* Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiol.* **138**, 1163–1173 (2005).
32. Sucena, E. & Stern, D. L. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by *cis*-regulatory evolution of *ovo/shaven-baby*. *Proc. Natl Acad. Sci., USA* **97**, 4530–4534 (2000).
33. Sucena, E., Delon, I., Jones, I., Payre, F. & Stern, D. L. Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* **424**, 935–938 (2003).
34. Frankel, N. *et al.* Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* **474**, 598–603 (2011).
35. Frankel, N., Wang, S. & Stern, D. L. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc. Natl Acad. Sci. USA* **109**, 20975–20979 (2012).
36. Yatime, L. *et al.* Structural insights into the high affinity binding of cardiotonic steroids to the Na^+,K^+-ATPase. *J. Struct. Biol.* **174**, 296–306 (2011).
37. Dobler, S., Dalla, S., Wagschal, V. & Agrawal, A. A. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. *Proc. Natl Acad. Sci. USA* **109**, 13040–13045 (2012).
38. Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M. & Andolfatto, P. Parallel molecular evolution in an herbivore community. *Science* **337**, 1634–1637 (2012).
39. Hittinger, C. T. & Carroll, S. B. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449**, 677–681 (2007).
40. Des Marais, D. L. & Rausher, M. D. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**, 762–765 (2008).
41. Nasvall, J., Sun, L., Roth, J. R. & Andersson, D. I. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* **338**, 384–387 (2012).

42. Chen, L., DeVries, A. L. & Cheng, C. H. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc. Natl Acad. Sci. USA* **94**, 3817–3822 (1997).
43. Christin, P. A., Salamin, N., Savolainen, V., Duvall, M. R. & Besnard, G. C_4 Photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol.* **17**, 1241–1247 (2007).
44. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nature Genet.* **39**, 1256–1260 (2007).
45. Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science* **307**, 1928–1933 (2005).
46. Hohenlohe, P. A., Bassham, S., Currey, M. & Cresko, W. A. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Phil. Trans. R. Soc. B* **367**, 395–408 (2012).
47. Hagen, D. W. Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). *J. Fish. Res. Bd. Canada* **24**, 1637–1692 (1967).
48. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
A genome-wide survey that reveals extensive evidence for collateral evolution in sticklebacks.
49. Hohenlohe, P. A. *et al.* Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* **6**, e1000862 (2010).
50. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).
51. Shapiro, M. D., Bell, M. A. & Kingsley, D. M. Parallel genetic origins of pelvic reduction in vertebrates. *Proc. Natl Acad. Sci. USA* **103**, 13753–13758 (2006).
52. Christin, P. A. *et al.* Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Curr. Biol.* **22**, 445–449 (2012).
A study showing that genes required for C_4 photosynthesis were acquired by lateral gene transfer at least four times in a genus of grass species.
53. Song, Y. *et al.* Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Curr. Biol.* **21**, 1296–1301 (2011).
54. Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
A sequencing study of *Heliconius* spp. genomes that reveals lateral exchange of chromosomal segments which control wing colour polymorphisms between species.
55. Reed, R. D. *et al. optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**, 1137–1141 (2011).
56. Green, R. E. *et al.* A draft sequence of the *Neandertal* genome. *Science* **328**, 710–722 (2010).
57. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
58. Joron, M., Jiggins, C. D., Papanicolaou, A. & McMillan, W. O. *Heliconius* wing patterns: an evo–devo model for understanding phenotypic diversity. *Heredity* **97**, 157–167 (2006).
59. Chevin, L. M., Martin, G. & Lenormand, T. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**, 3213–3231 (2010).
60. Yeung, A. T. *et al.* Swarming of *Pseudomonas aeruginosa* is controlled by a broad spectrum of transcriptional regulators, including *MetR*. *J. Bacteriol.* **191**, 5592–5602 (2009).
61. Breidenstein, E. B., Khaira, B. K., Wiegand, I., Overhage, J. & Hancock, R. E. Complex ciprofloxacin resistome revealed by screening a *Pseudomonas aeruginosa* mutant library for altered susceptibility. *Antimicrob. Agents Chemother.* **52**, 4486–4491 (2008).
62. Michaels, S. D. & Amasino, R. M. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949–956 (1999).
63. Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F. & Plaza, S. *Shavenbaby* couples patterning to epidermal cell shape control. *PLoS Biol.* **4**, e290 (2006).
64. Knecht, A. K., Hosemann, K. E. & Kingsley, D. M. Constraints on utilization of the EDA-signaling pathway in threespine stickleback evolution. *Evol. Dev.* **9**, 141–154 (2007).
65. Haldane, J. B. S. A mathematical theory of natural and artificial selection, part V: selection and mutation. *Proc. Camb. Philos. Soc.* **28**, 838–844 (1927).
66. Orr, H. A. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935–949 (1998).
67. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9**, 855–867 (2008).
68. Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
69. Kruglyak, L. & Nickerson, D. A. Variation is the spice of life. *Nature Genet.* **27**, 234–236 (2001).
70. Cooper, T. F., Ostrowski, E. A. & Travisano, M. A negative relationship between mutation pleiotropy and fitness effect in yeast. *Evol. Int. J. Org. Evol.* **61**, 1495–1499 (2007).
71. Stern, D. L. Perspective: evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
72. Carroll, S. B. Evo–devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
73. Nelson, M. D. *et al.* A bow-tie genetic architecture for morphogenesis suggested by a genome-wide RNAi screen in *Caenorhabditis elegans*. *PLoS Genet.* **7**, e1002010 (2011).
74. Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796–800 (2006).
75. McGrath, P. T. *et al.* Parallel evolution of domesticated *Caenorhabditis* species targets pheromone receptor genes. *Nature* **477**, 321–325 (2011).
76. Rockman, M. V. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**, 1–17 (2012).
77. Perutz, M. F. Species adaptation in a protein molecule. *Mol. Biol. Evol.* **1**, 1–28 (1983).
78. Kruglyak, L. The road to genome-wide association studies. *Nature Rev. Genet.* **9**, 314–318 (2008).
79. Remington, D. L., Ungerer, M. C. & Purugganan, M. D. Map-based cloning of quantitative trait loci: progress and prospects. *Genet. Res.* **78**, 213–218 (2001).
80. Seidel, H. S., Rockman, M. V. & Kruglyak, L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* **319**, 589–594 (2008).
81. King, E. G. *et al.* Genetic dissection of a model complex trait using the *Drosophila* Synthetic Population Resource. *Genome Res.* **22**, 1558–1566 (2012).
82. Yoon, H. S. & Baum, D. A. Transgenic study of parallelism in plant morphological evolution. *Proc. Natl Acad. Sci. USA* **101**, 6524–6529 (2004).
83. Ffrench-Constant, R. H. The molecular and population genetics of cyclodiene insecticide resistance. *Insect Biochem. Mol. Biol.* **24**, 335–345 (1994).
84. Hemingway, J. & Ranson, H. Insecticide resistance in insect vectors of human disease. *Annu. Rev. Entomol.* **45**, 371–391 (2000).
85. Ffrench-Constant, R. H., Pittendrigh, B., Vaughan, A. & Anthony, N. Why are there so few resistance-associated mutations in insecticide target genes? *Phil. Trans. R. Soc.* **353**, 1685–1694 (1998).
86. Herring, C. D. *et al.* Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nature Genet.* **38**, 1406–1412 (2006).
87. Orr, H. A. The probability of parallel evolution. *Evolution* **59**, 216–220 (2005).
A population genetics model revealing that, despite the apparent complexity of parallel evolution, the probability of this type of evolution at the DNA level is a simple function of the number of beneficial mutations that are possible at a gene.
88. Boucher, C. A. *et al.* Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. *J. Infect. Dis.* **165**, 105–110 (1992).
89. Kellam, P., Boucher, C. A., Tijnagel, J. M. & Larder, B. A. Zidovudine treatment results in the selection of human immunodeficiency virus type 1 variants whose genotypes confer increasing levels of drug resistance. *J. Gen. Virol.* **75**, 341–351 (1994).
90. Zhang, J. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nature Genet.* **38**, 819–823 (2006).
91. Barrett, R. D., Rogers, S. M. & Schluter, D. Natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255–257 (2008).
92. Stern, D. L. & Frankel, N. The structure and evolution of cis-regulatory regions: the *shavenbaby* story. *Phil. Trans. R. Soc. B* (in the press).
93. Segre, A. V., Murray, A. W. & Leu, J. Y. High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS Biol.* **4**, e256 (2006).
94. Hittinger, C. T., Rokas, A. & Carroll, S. B. Parallel inactivation of multiple *GAL* pathway genes and ecological diversification in yeasts. *Proc. Natl Acad. Sci. USA* **101**, 14144–14149 (2004).
95. Streisfeld, M. A. & Rausher, M. D. Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. *New Phytol.* **183**, 751–763 (2009).
96. Johanson, U. *et al.* Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**, 344–347 (2000).
97. Kuittinen, H., Niittyvuopio, A., Rinne, P. & Savolainen, O. Natural variation in *Arabidopsis lyrata* vernalization requirement conferred by a *FRIGIDA* indel polymorphism. *Mol. Biol. Evol.* **25**, 319–329 (2008).
98. Blasing, O. E., Westhoff, P. & Svensson, P. Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. *J. Biol. Chem.* **275**, 27917–27923 (2000).
99. Besnard, G. *et al.* Phylogenomics of C_4 photosynthesis in sedges (*Cyperaceae*): multiple appearances and genetic convergence. *Mol. Biol. Evol.* **26**, 1909–1919 (2009).
100. Tishkoff, S. A. *et al.* Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. *Science* **293**, 455–462 (2001).
101. Tishkoff, S. A. *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genet.* **39**, 31–40 (2007).
102. Protas, M. E. *et al.* Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genet.* **38**, 107–111 (2006).
103. Gross, J. B., Borowsky, R. & Tabin, C. J. A. Novel role for *Mc1r* in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genet.* **5**, e1000326 (2009).
104. Sugawara, T. *et al.* Parallelism of amino acid changes at the RH1 affecting spectral sensitivity among deep-water cichlids from Lakes Tanganyika and Malawi. *Proc. Natl Acad. Sci. USA* **102**, 5448–5453 (2005).
105. Bricelj, V. M. *et al.* Sodium channel mutation leading to saxitoxin resistance in clams increases risk of PSP. *Nature* **434**, 763–767 (2005).
106. Venkatesh, B. *et al.* Genetic basis of tetrodotoxin resistance in pufferfishes. *Curr. Biol.* **15**, 2069–2072 (2005).
107. Shapiro, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
108. Cresko, W. A. *et al.* Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc. Natl Acad. Sci. USA* **101**, 6050–6055 (2004).
109. Soanes, C. & Stevenson, A. (eds) *Oxford Dictionary of English* 2nd edn (revised) (Oxford Univ. Press, 2006).

Acknowledgements
The author thanks V. Orgogozo and the anonymous referees for many suggestions that improved this Review. He also thanks P. Andolfatto, O. Tenaillon and B. Gaut for providing images and raw data from their published work.

Competing interests statement
The author declares no competing financial interests.