The low cost of recombination in creating novel phenotypes

Recombination can create new phenotypes while disrupting well-adapted phenotypes much less than mutation

Andreas Wagner

Recombination is often considered a disruptive force for well-adapted phenotypes, but recent evidence suggests that this cost of recombination can be small. A key benefit of recombination is that it can help create proteins and regulatory circuits with novel and useful phenotypes more efficiently than point mutation. Its effectiveness stems from the large-scale reorganization of genotypes that it causes, which can help explore far-flung regions in genotype space. Recent work on complex phenotypes in model gene regulatory circuits and proteins shows that the disruptive effects of recombination can be very mild compared to the effects of mutation. Recombination thus can have great benefits at a modest cost, but we do not understand the reasons well. A better understanding might shed light on the evolution of recombination and help improve evolutionary strategies in biochemical engineering.

Keywords:
evolutionary innovation; genetic load; networks

DOI 10.1002/bies.201100027

Institute of Evolutionary Biology and Environmental Sciences, University of Zurich, Zurich, Switzerland

Corresponding author:
Andreas Wagner
E-mail: andreas.wagner@ieu.uzh.ch

Introduction

Proteins and gene regulatory circuits are the source of many evolutionary adaptations and innovations. Proteins form most enzymes, transport chemicals in and out of cells, are involved in locomotion and serve many other functions. Gene regulatory circuits involve genes that encode transcription factors, proteins that can bind short regulatory sequences near other genes and activate or repress their transcription. Such circuits are not only responsible for a cell's physiological response to its environments; they also pattern many body structures during the development of multicellular animals [1]. Not surprisingly, they are also involved in the evolution of many novel phenotypes, such as the evolution of limbs in vertebrates [2, 3].

Because both proteins and regulatory circuits are very important constituents of living things, it is useful to study the phenotypic effects of genetic changes in them. As I will discuss here, recombination can affect phenotypes in both system classes to a surprisingly small extent, much less than mutation does. This phenomenon contributes to recombination's role in facilitating evolutionary adaptation. Its proximate or mechanistic causes are poorly understood and call for further study.

Recombination causes greater genotypic change than mutation

Point mutations change the genotype of a biological system one small part at a time. This genotype is ultimately an RNA or DNA sequence. However, different, more compact representations of a genotype are often useful. For example, for proteins, genotypes are often represented as amino acid sequences. A point mutation in such a genotype then corresponds to a change in a single amino acid. A gene regulatory circuit can often be usefully represented through its regulatory genotype, the regulatory interactions – activating, repressing or

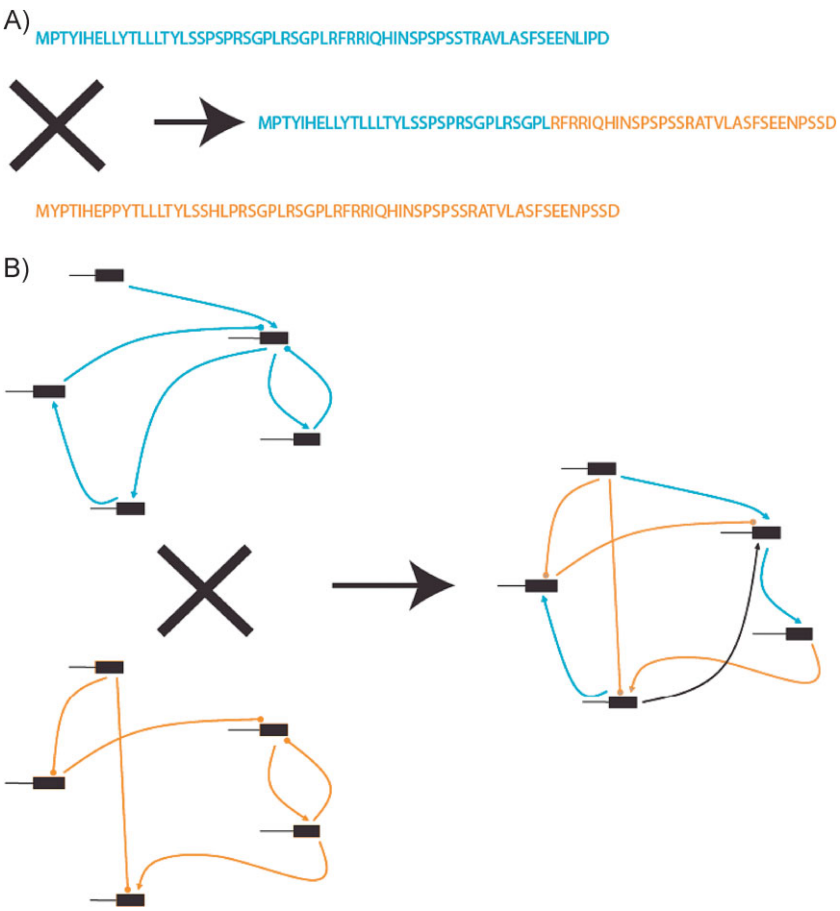

Figure 1. Recombination in proteins and regulatory circuits. **A:** The left side of the figure shows two hypothetical ''parental'' protein sequences (blue and orange). The right side shows a chimaeric protein resulting from a reciprocal recombination event between these parents. **B:** The left side of the figure shows two hypothetical regulatory circuits that differ in regulatory interactions (blue and orange curved lines) among circuit genes (black rectangles). The right side shows one of many potential chimaeric circuits created through recombination between the parents. The recombinant circuits may contain interactions that are present in one or both parent (e.g. the interaction between gene 1 and 2, when genes are numbered clockwise from the top), they may lack interactions that are present in either parent (interaction between gene 2 and 4), or they may contain interactions that do not occur in either parent (interaction between gene 4 and 2, black arrow). The possible outcomes of recombination depend on whether the transcription factors in the two parents differ in their target specificity, or whether they differ only in their regulatory regions. They also depend on whether a recombination event occurs within a gene's regulatory region.

absent – that can exist between any two of its genes [4–6]. In this case, a point mutation would correspond to a change in the strength of a single such interaction, such as might occur through a nucleotide change in a regulatory DNA region.

Recombination can rearrange genotypes more dramatically than point mutations can. In proteins, recombination can create chimaeras that differ from each parent in many amino acids (Fig. 1A). Such rearrangements can create new proteins with novel phenotypes – new tertiary structures and biochemical activities – but they can also destroy existing well-adapted phenotypes. In regulatory circuits, recombination can create chimaeric circuits that differ from each parent in multiple regulatory interactions (Fig. 1B). The phenotypes of such circuits are gene expression patterns, patterns of gene activity caused by the cross-regulatory interactions of genes within a circuit. Recombination can not only create new and potentially useful gene expression phenotypes, it can also destroy existing and well-adapted phenotypes.

Recombination may be superior to point mutation in producing novel phenotypes with desirable properties [7–11]. This observation is difficult to reconcile with the expectation that recombination usually destroys existing well-adapted phenotypes. Below I review and synthesize some recent work that challenges this expectation [12–14]. I begin by highlighting examples of the enormous success of recombination in biological engineering. Next I introduce the concept of a genotype space, which is useful to study the effects of recombination systematically, and I also discuss the organization of phenotypes in such a space. Subsequently, I examine the effects of recombination on regulatory circuits and on proteins to compare these classes of systems. Finally, I discuss how population level processes can dramatically diminish the deleterious effects of recombination.

Recombination is powerful in creating new phenotypes

DNA shuffling [15] is a widely used technique in engineering novel proteins and higher level systems in the laboratory [7, 11, 15–20]. Briefly, DNA shuffling starts from a mix of different ''parental'' variants of equally long DNA sequences, such as different alleles of a gene, and subsequently produces multiple recombined fragments of the parental DNA molecules [15]. It is an artificial, laboratory guided form of recombination.

Crameri and collaborators applied DNA shuffling to recombine genes encoding cephalosporinases. These are enzymes that confer resistance to cephalosporins, a class of antibiotics. A single DNA shuffling experiment [7] recombined four cephalosporinases that showed between 18 and 42% divergence on the DNA level. The experiment yielded a chimaeric cephalosporinase with a 270-fold increase of resistance to the cephalosporin moxalactam, as compared to the parental

sequences. By comparison, the highest improvement achievable in the same amount of time through point mutations was an eightfold increase over the parent [7]. The same approach can also shuffle DNA sequences on a much larger scale, recombining DNA containing multiple genes or entire genomes [11]. For example, recombination of entire genomes has been used to produce strains of the bacterium *Streptomyces fradiae* that produce high amounts of the antibioticum tylosin. In this approach, recombination was 20 times more effective than random mutagenesis in improving tylosin production [11]. These and other experiments show that experimental recombination of DNA sequences can rapidly generate genes, pathways and genomes with new and desirable features [7, 11, 15–20].

Although laboratory experiments like these illustrate the power of recombination and identify phenotypes with novel properties, they were not designed to examine a central problem that recombination causes in evolution: it can disrupt already existing, well-adapted phenotypes, and may thus have large deleterious effects. The genotypes with desired properties that these experiments identify might be few among an astronomical number of potentially inactive chimaeras [11].

Proteins and regulatory circuits exist in vast genotype spaces

To study the effects of point mutations and recombination systematically, it is useful to study the genotype space – the entire set of possible genotypes – that any one genotype is a part of and also the relationship between genotypes and phenotypes in this space (Fig. 2). In both proteins and regulatory circuits, we have recently gained a better understanding of this relationship. One reason is that sufficiently large amounts of empirical data exist to analyze their genotype-phenotype relationship, as in the case of proteins [21–26]. Another is that our ability to model biological systems quantitatively has recently improved, for example in regulatory circuits and metabolic networks [5, 6, 27–35]. Although experimental analysis remains the gold-standard to understand any system, such analysis is currently infeasible for the thousands or millions of genotypes one needs to analyze in order to understand a genotype-phenotype relationship comprehensively. For this purpose, computation, aided by comparative data, will remain indispensable in the foreseeable future. I will next summarize some pertinent insights from recent work on proteins and regulatory circuits [4–6, 21, 26, 36–40].

The genotypes of proteins exist in a space of amino acid sequences. Characterization of this genotype space has a long history that began with the study of lattice proteins and that has subsequently been extended to real proteins [21–26]. For polypeptides that are N amino acids long, this genotype space comprises 20^N possible amino acid sequences.

The genotype space of gene regulatory circuits comprising some N genes is represented as the space of all possible patterns of regulatory interactions that can exist between any pair of genes [4, 5]. Between N genes, of the order of N^2 regulatory interactions are possible. If one crudely classifies interactions into the three categories of activating, repressing and absent interactions, this space already comprises at least 3^{N^2} possible regulatory genotypes. Finer-grained representations of regulatory interactions would lead to an even larger genotype space. These considerations show that the genotype spaces of both proteins and regulatory circuits are extremely large and grow exponentially in size with the number of amino acids or regulatory interactions, respectively.

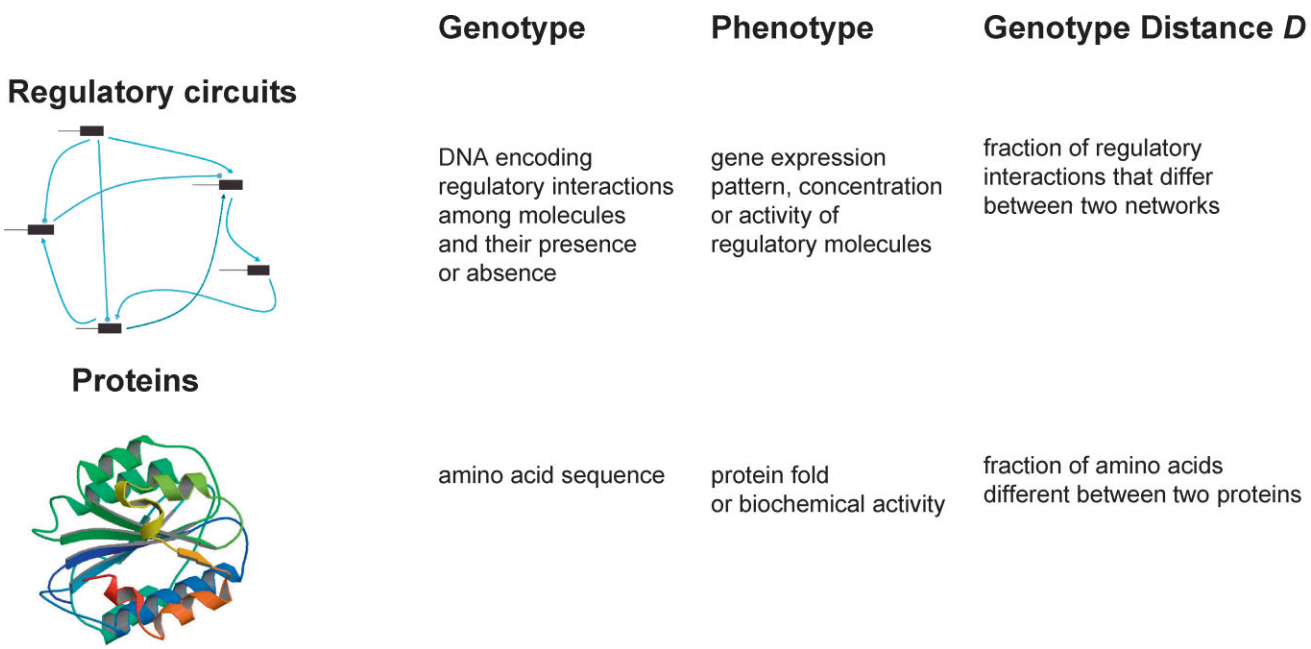

Figure 2. Genotypes and phenotypes in regulatory circuits and proteins

One of the advantages of the genotype space concept is that it makes quantification of differences between genotypes easy (Fig. 2). That is, one can define a distance D between two genotypes in the genotype space. For proteins, the simplest among several useful distance measures is the number or fraction of amino acids in which two proteins of the same length differ. For regulatory circuits, it is the fraction or number of regulatory interactions in which two circuits differ [4, 5]. Two proteins or circuits are maximally different if they do not share a single amino acid or regulatory interaction, respectively.

The genotype spaces of proteins and regulatory circuits share several other features beyond their size [4, 5, 21–24, 26]. The first pertains to a genotype's *neighbors*, genotypes that differ from it in one point mutation – a single amino acid or regulatory interaction. Individual genotypes typically have *many* neighbors in their genotype space with the same phenotype [4, 5, 41–43]. For example, random mutagenesis studies of different proteins showed that a large fraction of amino acid changes do not affect protein function [41–43]. Wherever many neighboring genotypes have the same phenotype, genotypes are to some extent *robust* to mutations. That is, their phenotype does not change in response to some mutations.

Secondly, genotypes with the same phenotype form vastly connected genotype networks that reach far through the genotype space. This means that one can step through a series of small genetic changes from one genotype to its neighbor, to the neighbor's neighbor, and so on, without ever changing a phenotype. Very different genotypes can thus have the same phenotype [4, 5, 21–24, 26, 44–46].

A third feature regards the neighborhoods of different genotypes on the same genotype network. These neighborhoods generally contain very different novel phenotypes [5, 21, 26]. More precisely, for two genotypes *G1* and *G2* on the same genotype network, consider the set ***P1*** of different phenotypes that occur in the neighborhood of *G1*, as well as the set ***P2*** of different phenotypes that occur in the neighborhood of *G2*. Then most phenotypes in ***P1*** are not in ***P2***, and vice versa. In other words, which new phenotypes can be easily accessed from a given genotype – through a minor genotypic change – depends on that genotype's location within a genotype network.

These features have important implications for how new and useful phenotypes originate in biological evolution. The existence of genotype networks means that an evolving population of proteins or regulatory circuits can gradually change its genotype through point mutations while preserving its phenotype. In doing so, it can explore different regions of genotype space. The immediate neighborhood of the population will contain very different novel phenotypes, depending on where its members are located in the genotype space. The existence of genotype networks, combined with the diversity of their neighborhoods thus permits a population to explore many novel phenotypes [89].

In this framework, recombination can be viewed as causing long jumps through genotype space (Fig. 3). By reaching into far-flung regions of this space, such jumps facilitate the exploration of different phenotypes, because different regions of genotype space contain different novel phenotypes. Thus, the phenotypic diversity of different genotype space regions explains why recombination can be more effective than point mutation for exploring new phenotypes. These long jumps, however, also have a downside. Typical genotype networks, although large in size, comprise a vanishing fraction of genotype space. For example, the number of amino acid sequences adopting the structure and function of the λ repressor, a transcription factor of the bacteriophage λ, may be of the order of 10^{56}. This astronomically large number of genotypes would nonetheless comprise only a fraction 10^{-63} of the genotype space in which this protein exists [47]. One would think that a long jump through genotype space starting from two very different recombining genotype on the same genotype network, might end on this same genotype network only extremely rarely. It might be much more likely to "jump off a cliff", landing nowhere near this genotype network. I will next discuss whether this is the case, beginning with a well-studied model of gene regulatory circuits. In doing so, I will focus on homologous recombination, as it occurs during meiosis. As opposed to various kinds of non-homologous recombination [48–52], homologous recombination leaves the length of a molecule, or the number of genes in a cellular circuit intact. Its consequences are thus much easier to analyze systematically [53].

Recombination can preserve gene expression phenotypes in model regulatory circuits much better than mutation

I will begin by examining the effects of recombination in transcriptional regulatory circuits in a well-studied computational model of such circuits [6, 13, 54–60]. Briefly, the phenotype in this model is a pattern of gene activities. This phenotype arises through cross- and autoregulatory interactions of network genes in response to regulatory input into the circuit from upstream genes [56]. Variants of the model have been successful in modeling developmental processes, especially the regulatory dynamics of early developmental genes in the fruit fly *Drosophila* [27, 61–64]. They have also helped address a broad range of conceptual questions in evolutionary biology, including why mutants often show a release of genetic variation that is cryptic in the wild-type, and how adaptive evolution of robustness can occur in regulatory circuits [54–58]. For simplicity, I will here consider circuits in haploid organisms, although extensions to higher ploidy are possible [65].

Consider two individuals ("parents"), each of which harbors a regulatory circuit genotype that produces a gene expression phenotype. Both individuals have identical phenotypes and belong to the same genotype network. Their regulatory genotypes may differ in one or more of their regulatory interactions. These two individuals produce offspring through a reciprocal exchange of their regulatory genotypes. I will discuss here a scenario where the individual genes of the circuit are not closely linked and thus can recombine freely. I choose this scenario, because here the potentially deleterious effects of recombination will be most evident. In this scenario,

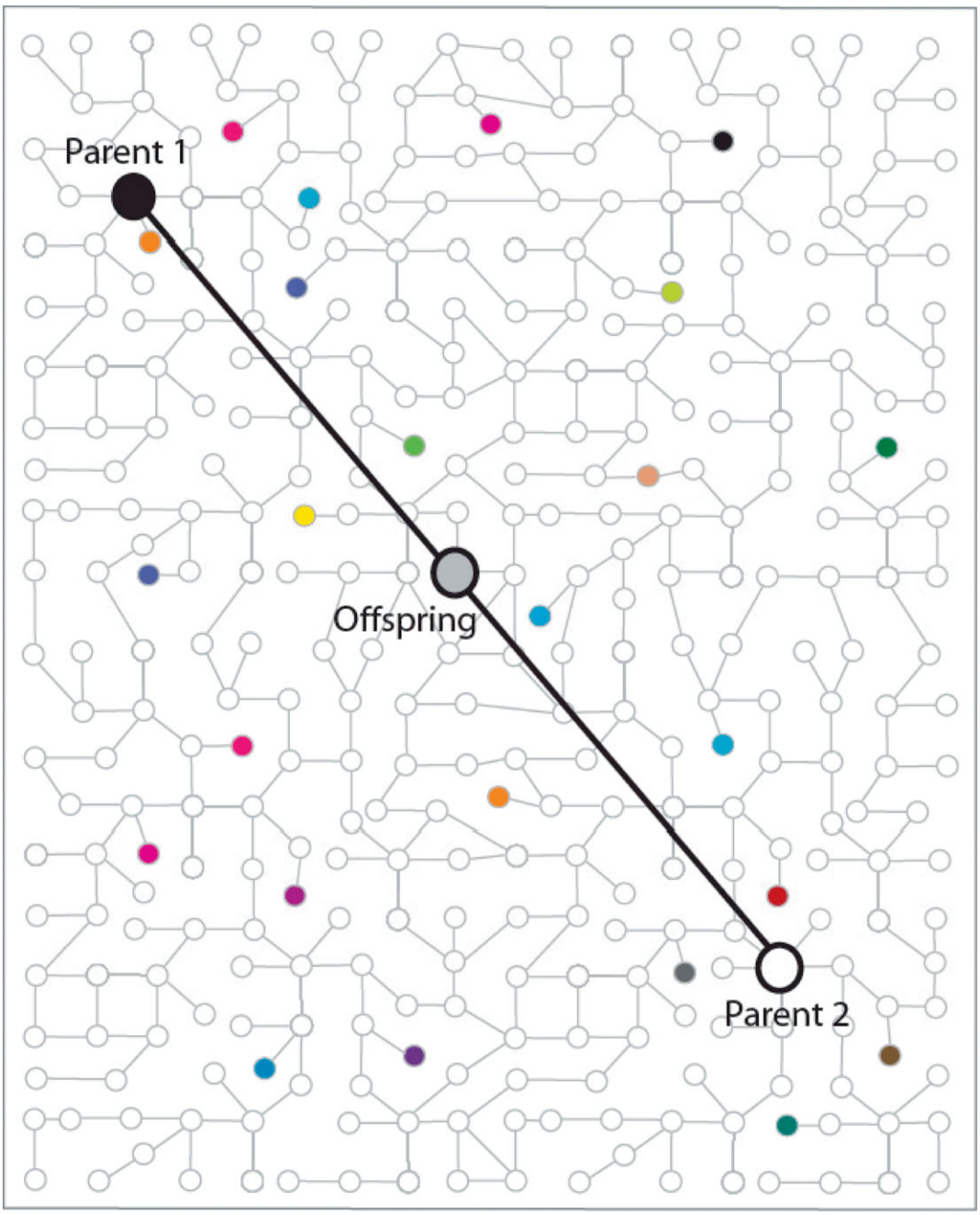

Figure 3. Recombination can cause drastic genotypic changes. The figure illustrates schematically that the offspring of a recombination event may be quite different from either parent. The large rectangle stands for genotype space. Small gray circles connected by lines indicate neighboring genotypes on one hypothetical genotype network. Colored circles indicate genotypes with a novel phenotype that are just one mutation away from genotypes on this genotype network. Different colors indicate that different neighborhoods of this genotype network and different regions of genotype space harbor different novel phenotypes. I note that genotype spaces are high-dimensional spaces whose complexity cannot be captured in two dimensions. For instance, the neighborhood of each genotype may comprise hundreds of genotypes. Also, each colored circle would be part of a different genotype network that is not shown. The large black and white circles indicated two hypothetical parental genotypes. The large gray circle stands for a recombinant offspring of the two parents. Its neighborhood in genotype space can contain novel phenotypes that do not occur near either parent, which highlights the positive role of recombination in phenotypic variability. In the image, the offspring genotype is equally distant from either parent, but in reality it may be closer to one or the other parent, depending on details of the recombination event that produced it. The offspring may lie on the same genotype network, and thus have the same phenotype as the parents, as indicated in this hypothetical example; or it may lie outside this genotype network and thus have a different phenotype.

every gene in each ''offspring'' network receives with probability one half the regulatory region of one of the parents, and with probability one half the regulatory region of the other parent. (A focus on regulatory regions is justified by the often vast size of such regions when compared to the regions encoding the transcriptional regulator itself, and by their generally more rapid evolution in comparison to the often more conserved coding region [34, 66–70]).

A quantity of interest is the probability that the offspring of recombination between two parents would no longer have the parental phenotype. This probability indicates the disruptive effects of recombination. Its value will depend on how different the parents are from one another. Recombination between genotypically similar parents will produce offspring whose genotypes are also similar to either parent. Thus, we would expect that their phenotypes are also often unchanged. Conversely, genotypically very dissimilar parents would usually produce genotypically and phenotypically diverse offspring.

One way to take parental similarity into account is to compare the offspring's genotype to one of the parents and determine the number m of regulatory interactions in which it differs from this parent. To this end, I will define the probability $R_R(m)$ as that of a recombination event whereby changing m regulatory interactions of a viable circuit does not change its phenotype. It is useful to compare this quantity to the probability $R_\mu(m)$ that m independent random changes (mutations) of individual regulatory interactions preserve the phenotype. By comparing the two quantities $R_R(m)$ and $R_\mu(m)$, we can assess how strongly recombination affects a genotype when compared to an equivalent amount of mutational change.

Figure 4A shows $R_R(m)$ and $R_\mu(m)$ for circuits sampled at random from the set of all circuits with the same expression phenotype [71]. One can see that for recombination events that change only $m = 1$ regulatory interactions, recombination is already less likely to change a circuit phenotype than point mutations. Specifically, more than 90% of recombinant offspring that differ from their most closely related parent by only one regulatory interaction preserve the parental phenotype. In contrast, only 75% of circuits where mutations changed one regulatory interaction preserve this phenotype. With increasing numbers of changes m, these differences become more drastic. For example, when a recombination event changes $m = 12$ regulatory interactions, 50% of all offspring circuits preserve the parental phenotype, whereas fewer than 8% of circuits with 12 random mutations preserve this phenotype (Fig. 4A). These observations show that exchanging regulatory interactions that are already part of a viable circuit greatly increases the likelihood to preserve the circuit's phenotype.

The following is a complementary way of examining the effects of recombination [71]. If the parent circuits differ in H regulatory interactions, then the recombinant offspring will differ from one of the parents by m regulatory interactions, whereas it will differ from the other parent by (H-m) regulatory interactions. We can then express the distance of the offspring from either parent as a fraction of H, i.e. as a recombination distance $D_R = m/H$. This recombination distance varies between 0 and 1. A value of D_R close to zero means that the offspring is close to the reference parent, whereas a value of D_R close to one means that the offspring is very distant to the reference parent, but very close to the other parent. Intermediate values of D_R mean that the offspring is approximately equally distant to either parent. Figure 4B examines the relationship between the recombination distance D_R to the probability that recombination preserves the parental phenotype. For now, I will focus on the lowest curve in this figure. The data that generated this curve were based on parental regulatory circuits that were sampled at random from a set of genotypes with the same phenotype [71]. The figure shows that offspring very similar to the parent, where D_R is close to zero or one, is very likely to preserve the parental phenotype. The likelihood that a recombination event changes the phenotype has a parabolic, U-shaped distribution, with a minimum at intermediate recombination distances D_R. This means that recombination is most likely to change a phenotype, if the recombinant circuit is maximally different from either parent.

Recombination preserves protein structure and function

The weaker effects of recombination compared to mutation in Fig. 4 might be peculiarities of transcriptional regulation circuits or models thereof. Alternatively, they may be generic properties that hold for broader classes of systems and that reflect fundamental organizational properties of genotype space. A mix of computational and experimental evidence from proteins argues for the latter possibility [14, 72]. One pertinent study focused on lattice proteins, simple computational models of protein folding [72]. Its authors studied sequences that fold into the same structure and subjected pairs of such sequences to recombination. They found that 78.9% of recombination products fold stably into a structure and that the vast majority of them (99.3%) adopt a structure identical to that of the parents. Another relevant study [14] compared the effects of recombination in both real proteins and lattice proteins. The study's authors estimated the probabilities $R_R(m)$ that a recombination event changing m amino acids preserves protein structure and compared it to the probability $R_\mu(m)$ that the same number of mutational changes preserves secondary structure. For lattice proteins with $R_\mu(1) = 0.1$, that is, where 10% of a protein's neighbors have the same structure, the authors found that the fraction of recombination events that change a single amino acid and preserve protein structure is $R_R(1) \approx 0.7$. In other words, recombination is seven times more likely than point mutation to preserve a lattice protein's structure. For mutationally more robust proteins where $R_\mu(1) = 0.5$, $R_R(1)$ exceeds 0.85. For larger numbers m of amino acid changes, mutations typically have dramatically more disruptive effects than recombination. For example, for a structure where more than 30% of recombination events that change five residues do not disrupt the structure ($R_R(5) > 0.3$), fewer than 1% of five independent mutation events leave the structure intact ($R_\mu(5) < 0.01$). This means that recombination is thirty times more likely to preserve this structure than the same amount of change caused by mutation [14].

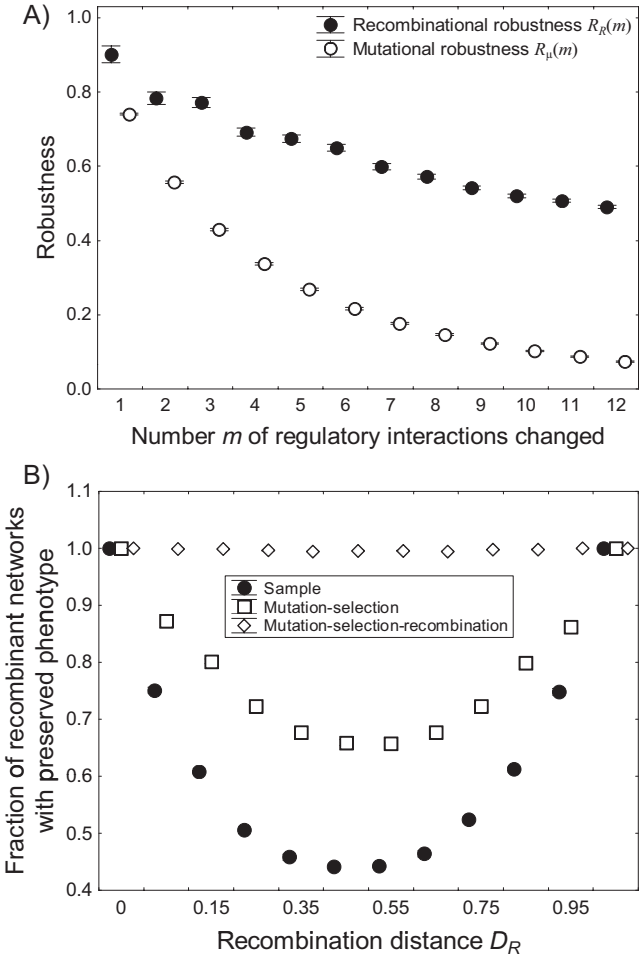

Figure 4. Recombination can exert very weak disruptive effects on phenotypes. **A:** The figure shows the probabilities $R_\mu(m)$ and $R_R(m)$ that m changes of individual regulatory interactions caused by mutation and recombination, respectively, leave a circuit's gene expression phenotype intact. The data is based on 10^6 circuits of 12 genes randomly sampled from the same genotype network [71]. A mutation changes a single regulatory interaction. **B:** The vertical axis shows the fraction of viable offspring circuits, i.e. recombinant offspring circuits with the same phenotype as the parent, as a function of the recombination distance D_R between parent and offspring (see text). The recombination distance is normalized to values ranging between zero and one. Data are shown for parental circuits sampled uniformly from the same genotype network (''sample''), for circuits from a population in mutation-selection balance, and for circuits from a population in mutation-selection-recombination balance. Note the very high fraction of viable recombinants for the population in mutation-selection-recombination balance. The middle and upper curves in (B) are based on populations of 1,000 circuits, and $\mu = 1$ mutations of regulatory interactions per circuit and generation. Lengths of error bars correspond to one standard deviation and are too small to be visible in (B). Data are for circuits of 12 genes, but circuits of different size and gene expression phenotypes show similar patterns. Figure and caption adapted from [71].

These qualitative differences between recombination and mutation have been confirmed in experimentally constructed recombinants of two well-studied proteins encoding β-lactamase. This enzyme cleaves and inactivates antibiotics that contain a four atom ring called a β-lactam. Such antibiotics include penicillins and ampicillin. β-lactamases convey bacterial resistance against these antibiotics. The experimentors used two β-lactamases called PSE-4 and TEM-1 that share 43% of their amino acids. They constructed synthetic recombinant enzymes with various amounts of amino acid change relative to either parent [14]. For comparison, they also produced enzymes with the same number of amino acid changes, but on this occasion the changes are caused by random mutation. For both classes of proteins – recombinants and mutants – they estimated what fraction of protein had preserved the ancestors' molecular activity. They did so by identifying the fraction of recombinant or mutant proteins that allowed *Escherichia coli* cells to survive treatment with the antibiotic ampicillin.

The experimentors found that a single amino acid exchange produced through recombination has a probability of $R_R(1) = 0.79$ to preserve protein function. In contrast, if random mutation causes this change, then this probability is only $R_\mu(1) = 0.54$. Thus, as in regulatory circuits and in lattice proteins, mutations are much more likely than recombination to disrupt protein structure. Increasing numbers of mutations enhance these differences dramatically, as Fig. 5 shows [14]. For example, recombinational change of some 10 amino acids has a greater than 20% chance of preserving protein function, whereas the same number of random mutations is ten times more likely to disrupt this function. More generally, the probability $R_\mu(m)$ that m mutations preserve a structure decreases exponentially with increasing m, but the same does not hold for recombination. Its effects show a parabolic distribution similar to that shown in Fig. 4 for regulatory circuits.

The effects of mutation in this system have only been measured up to some 30 mutations, but if extrapolated to the number of changes that distinguish maximally different recombinants from their parents, then recombination would be 16 orders of magnitude more likely to preserve phenotype than the same numbers of mutation [14].

Taken together, these observations suggest that the weak effects of recombination relative to mutation are not a peculiarity of one kind of system, but a generic property of different systems. A qualitative explanation may be straightforward: recombination swaps parts of a system that are able to form a given phenotype and, in this sense, have been ''pretested''. In contrast, mutation changes a system part for parts that may be incompatible with this phenotype [14, 17, 73]. In other words, recombination exchanges functional system parts for other such parts, whereas mutation may not necessarily do so.

The pattern of hydrophobic and hydrophilic amino acids on an amino acid chain serves to illustrate this principle. This pattern is necessary for the formation of a given protein structure [74]. For example, properly spaced hydrophobic amino acids may be necessary to form a protein's hydrophobic core. This means that some sequences of hydrophobic and hydrophilic amino acids are compatible with a given protein structure, whereas others are incompatible. Although the PSE-4 and TEM-1 lactamases that I just discussed have only 43% amino acid identity, if one considers only whether an amino acid is hydrophobic or polar, this identity rises to 76% [14]. Recombination will swap or exchange amino acids that

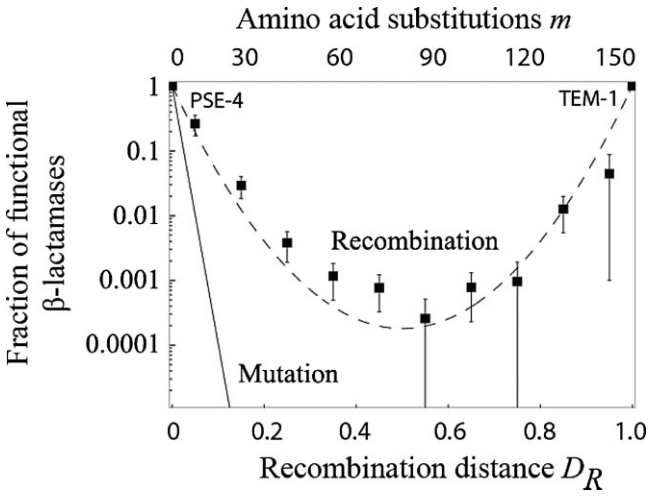

Figure 5. Recombination in β-lactamases is much more likely than mutation in order to preserve protein structure. The lower horizontal axis shows the recombination distance D_R, which is the distance of recombined β-lactamases to the PSE-4 parent, normalized to range between zero and one. The upper horizontal axis corresponds to the absolute number of amino acid changes caused by recombination between PSE-4 and TEM-1 β-lactamases. The numbers on this axis are also relative to the PSE-4 parent. Thus, the maximally possible number of 150 changes corresponds to the TEM-1 parent. The black squares show the fraction of functional recombinants binned according to their divergence from the PSE-4 parent. Error bars correspond to one standard error of the mean [14]. The line labeled ''Mutation'' is derived from mutagenesis and subsequent measurement of β-lactamase activity. Figure redrawn from [14]; original courtesy of Allan Drummond.

preserve hydrophobicity along the chain, and thus preserve compatibility with a given structure. The same considerations would hold for the volume of amino acid side chains and also for their electric charge. In regulatory circuits, a similar principle holds. Recombination swaps regulatory interactions that are compatible with a given gene activity phenotype. Examples include regulatory inputs to a gene that stabilize its expression (or repression) in an optimal expression phenotype [4, 71]. Two parental circuits may differ greatly in their genotype, but they may share such stabilizing interactions. If so, recombination involving such interactions would preserve a gene's expression state.

Robustness to recombination can greatly increase during evolution

Everything I have written thus far applies to systems that may or may not have been subject to recombination in their evolutionary history. Continued exposure to recombination and/or mutation, as it turns out, may dramatically increase the likelihood that recombination leaves a phenotype intact [13, 71]. To demonstrate this effect for regulatory circuits, one can examine populations of circuits that are subject to both repeated rounds of mutation of individual regulatory interactions and to selection preserving their gene expression pattern, until the population reaches an equilibrium between mutation and selection. For comparison, one can examine populations subjected to mutation, selection and recombination, and those that have reached a mutation-selection-recombination equilibrium. Figure 4B shows the effects of recombination in such populations [71]. It can be observed here that mutation and selection alone increase the likelihood that recombination preserves a gene expression phenotype by more than 40%, from less than 0.45 for randomly sampled circuits (black circles), to over 0.65 for populations in mutation-selection balance (open squares; both numbers are for the largest recombination distance $D_R = 0.5$). More dramatic, however, is the effect of ongoing recombination itself. The open diamonds in Fig. 4B indicate the probability that recombination leaves a gene expression phenotype unchanged, for populations in mutation-selection-recombination balance. This probability exceeds 0.995, even for recombinants with the maximal distance $D_R = 0.5$ from either parent. The same increase in robustness is evident if one examines the likelihood that a specific number of regulatory changes caused by recombination leaves a phenotype unchanged. For example, in a population existing in a mutation-selection-recombination balance, the probability that 10 independent mutations leave a phenotype intact is $R_\mu(10) = 0.49$, whereas the same probability, this time for changes in 10 regulatory interactions caused by recombination, equals $R_R(10) = 0.993$ [71]. In sum, continued exposure to recombination can dramatically increase robustness to recombination [13, 71].

In population genetic theory, the disruptive effects of recombination are conventionally expressed in terms of a ''genetic load'' [75, 76]. A population's genetic load designates a mean fitness lower than could be attained in the absence of some agent of evolutionary change, such as mutation, migration, or recombination. In the context of the regulatory circuits I just discussed, one can define the genetic load as the fraction of a generation's offspring that does not have the optimal, parental gene expression phenotype.

Think of the load as the amount of ''damage'' to the population that this agent of change causes. Recombination, one might believe, should always increase the genetic load of a population in mutation-selection-recombination balance, because it causes disruption of an optimal phenotype in at least some individuals of the population. However, this is not necessarily the case [71, 77].

Not only is the genetic load in populations subject to recombination modest, it can even be lower than in the absence of recombination, at least in large populations or in populations with a high rate of point mutations [13, 71]. The explanation of this apparently paradoxical observation is simple. Populations subject to ongoing recombination increase their robustness to recombination [13, 71]. In consequence, recombination can become a minor and mutation a major cause of deleterious phenotypic change. In addition, ongoing recombination also increases robustness to mutation, thus decreasing the genetic load (now mostly caused by mutations) compared to when recombination is absent [54, 71]. These phenomena can jointly lead to a net reduction in the genetic load [71].

These last observations are all based on circuits of transcriptional regulators. We do not yet have directly comparable evidence for proteins or other systems, but limited evidence

suggests that similar principles may exist there [17, 78, 79]. For example, a study about the effects of recombination on a regulatory and signaling circuit involved in segmenting the *Drosophila* embryo showed that continual recombination can greatly increase a circuit's robustness to mutation [79]. An unrelated study on lattice proteins showed that proteins subject to ongoing recombination can evolve a dramatically higher robustness of their structure to mutation [78]. Unfortunately, neither of these studies focused on the effects of recombination itself. In this regard, a DNA shuffling experiment of human α-interferons provides at least anecdotal pertinent evidence [80]. α-Interferons can interfere with viral infections and can inhibit cell division. In humans, they are encoded by more than 20 tandemly duplicated genes. Such tandem clusters of genes generally facilitate recombination between members of a cluster. Chang and collaborators used the human α-interferon genes in a DNA shuffling experiment. They found that most chimeric interferons were biologically active [80] and analyzed four randomly chosen recombinants in greater detail. They found that all four were at least as capable of inhibiting cell division in a human lymphoma cell line as their most active parent. These observations suggest that recombination in these proteins does not generally destroy protein function. The continual exposure of these tandemly arrayed genes to recombination may facilitate weak effects of recombination.

Conclusions and outlook

Recombination causes long jumps through genotype space. It can thus facilitate the exploration of novel phenotypes. At the same time it can destroy existing, well-adapted phenotypes. The latter feature is an obstacle to recombination's positive role in evolutionary adaptation and innovation. I showed here that the destructive role of recombination can be mild or even non-existent. This observation comes from two very different classes of systems – proteins and model regulatory circuits – suggesting that they may not be peculiarities of any one kind of system. Specifically, I firstly showed that recombination causes much weaker disruptive effects than mutation, because it exchanges system parts that are compatible with a given phenotype. Secondly, exposure of a system to recombination can dramatically increase the system's robustness to further recombination. It may even lead to a lower genetic load when compared to populations without recombination.

Despite these fundamental observations, there are still major gaps in our knowledge. The first regards experimental analyses of recombination's effects. In proteins, our knowledge comes from models of protein folding and from experiments on a few extremely well-studied proteins. We do not yet know whether other proteins also show the same phenomena. In transcriptional regulation circuits, our knowledge is even more limited, because it is extremely laborious to characterize individual circuits experimentally, let alone the offspring of multiple recombination events. Our current knowledge is thus based on quantitative models of such circuits. Experimental studies that use targeted mutagenesis of circuit architectures [81] or quantitative genetic experiments [82] to study recombination's effects on gene expression phenotypes are sorely needed. More generally, transcriptional regulation networks are only one among several classes of biological networks that include signaling networks and metabolic networks. These networks change at different rates on evolutionary time scales [32, 34, 35]. We currently know nothing about the relative impact of mutation and recombination on the phenotypes of such networks.

Beyond the general observation that recombination exchanges system parts compatible with a given phenotype, we know little about the mechanistic causes for recombination's low destructiveness in the complex phenotypes I discussed here. To investigate these causes more deeply is a worthwhile theme for future experimental research. Proteins would be a prime study object for such research, because we can manipulate them experimentally and because our ability to map genotypes to phenotypes computationally is advancing rapidly. Such work could contribute to a growing canon of explanations for the evolutionary origin and maintenance of recombination and sex [13, 83–88]. But it also has a more practical side. On the one hand, it could ask whether there are classes of protein sequences or circuit architectures that are especially resistant to recombination, while being amenable to bringing forth new phenotypes. On the other hand, it could ask whether there are modes of recombination that tend to disrupt phenotypes minimally while creating abundant new phenotypes. Answers to these questions might improve the yield of new proteins or circuits in biochemical engineering projects. In other words, such work might reward us with systems poised to innovate.

Acknowledgments

This work was supported by Swiss National Science Foundation grants 315200-116814, 315200-119697, and 315230-129708, as well as through the YeastX project of SystemsX.ch. Material in this paper is to appear in similar form in ref. 89 by the author.

References

1. **Carroll SB**, **Grenier JK**, **Weatherbee SD.** 2001. *From DNA to Diversity. Molecular Genetics and the Evolution of Animal Design*. Malden, MA: Blackwell.
2. **Tabin CJ.** 1992. Why we have (only) 5 fingers per hand: Hox genes and the evolution of paired limbs. *Development* **116**: 289–96.
3. **Burke AC**, **Nelson CE**, **Morgan BA**, **Tabin C.** 1995. Hox genes and the evolution of vertebrate axial morphology. *Development* **121**: 333–46.
4. **Ciliberti S**, **Martin OC**, **Wagner A.** 2007. Circuit topology and the evolution of robustness in complex regulatory gene networks. *PLoS Comput Biol* **3**: e15.
5. **Ciliberti S**, **Martin OC**, **Wagner A.** 2007. Innovation and robustness in complex regulatory gene networks. *Proc Natl Acad Sci USA* **104**: 13591–6.
6. **MacCarthy T**, **Seymour R**, **Pomiankowski A.** 2003. The evolutionary potential of the *Drosophila* sex determination gene network. *J Theor Biol* **225**: 461–8.
7. **Crameri A**, **Raillard S**, **Bermudez E**, **Stemmer W.** 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**: 288–91.
8. **Zhou Q**, **Zhang GJ**, **Zhang Y**, **Xu SY**, et al. 2008. On the origin of new genes in *Drosophila*. *Genome Res* **18**: 1446–55.
9. **Shapiro JA.** 2004. A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering. *Annual*

Scientific Meeting on Structural Approaches to Sequence Evolution. Dresden, Germany. pp. 91–100.
10. **Kolkman J**, **Stemmer W.** 2001. Directed evolution of proteins by exon shuffling. *Nat Biotechnol* **19**: 423–8.
11. **Zhang Y**, **Perry K**, **Vinci V**, **Powell K**, et al. 2002. Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* **415**: 644–6.
12. **Martin OC**, **Wagner A.** 2008. Multifunctionality and robustness trade-offs in model genetic circuits. *Biophys J* **94**: 2927–37.
13. **Lohaus R**, **Burch CL**, **Azevedo RBR.** 2010. Genetic architecture and the evolution of sex. *J Hered* **101**: S142–S57.
14. **Drummond DA**, **Silberg JJ**, **Meyer MM**, **Wilke CO**, et al. 2005. On the conservative nature of intragenic recombination. *Proc Natl Acad Sci USA* **102**: 5380–5.
15. **Stemmer W.** 1994. DNA shuffling by random fragmentation and reassembly – in vitro recombination for molecular evolution. *Proc Natl Acad Sci USA* **91**: 10747–51.
16. **Crameri A**, **Dawes G**, **Rodriguez E**, **Silver S**, et al. 1997. Molecular evolution of an arsenate detoxification pathway DNA shuffling. *Nat Biotechnol* **15**: 436–8.
17. **Chang C**, **Chen T**, **Cox B**, **Dawes G**, et al. 1999. Evolution of a cytokine using DNA family shuffling. *Nat Biotechnol* **17**: 793–7.
18. **Ness J**, **Welch M**, **Giver L**, **Bueno M**, et al. 1999. DNA shuffling of subgenomic sequences of subtilisin. *Nat Biotechnol* **17**: 893–6.
19. **Soong N**, **Nomura L**, **Pekrun K**, **Reed M**, et al. 2000. Molecular breeding of viruses. *Nat Genet* **25**: 436–9.
20. **Leong S**, **Chang J**, **Ong R**, **Dawes G**, et al. 2003. Optimized expression and specific activity of IL-12 by directed molecular evolution. *Proc Natl Acad Sci USA* **100**: 1163–8.
21. **Lipman D**, **Wilbur W.** 1991. Modeling neutral and selective evolution of protein folding. *Proc Biol Sci* **245**: 7–11.
22. **Thornton J**, **Orengo C**, **Todd A**, **Pearl F.** 1999. Protein folds, functions and evolution. *J Mol Biol* **293**: 333–42.
23. **Todd A**, **Orengo C**, **Thornton J.** 1999. Evolution of protein function, from a structural perspective. *Curr Opin Chem Biol* **3**: 548–56.
24. **Bastolla U**, **Porto M**, **Roman HE**, **Vendruscolo M.** 2003. Connectivity of neutral networks, overdispersion, and structural conservation in protein evolution. *J Mol Evol* **56**: 243–54.
25. **Rost B.** 2002. Enzyme function less conserved than anticipated. *J Mol Biol* **318**: 595–608.
26. **Ferrada E**, **Wagner A.** 2010. Evolutionary innovation and the organization of protein functions in sequence space. *PLoS One* **5**: e14172.
27. **Jaeger J**, **Surkova S**, **Blagov M**, **Janssens H**, et al. 2004. Dynamic control of positional information in the early *Drosophila* embryo. *Nature* **430**: 368–71.
28. **Sanchez L**, **Chaouiya C**, **Thieffry D.** 2008. Segmenting the fly embryo: logical analysis of the role of the Segment Polarity cross-regulatory module. *Int J Dev Biol* **52**: 1059–75.
29. **Albert R**, **Othmer HG.** 2003. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J Theor Biol* **223**: 1–18.
30. **von Dassow G**, **Meir E**, **Munro E**, **Odell G.** 2000. The segment polarity network is a robust development module. *Nature* **406**: 188–92.
31. **Cotterell J**, **Sharpe J.** 2010. An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. *Mol Syst Biol* **6**: 425.
32. **Shou C**, **Bhardwaj N**, **Lam HYK**, **Yan KK**, et al. 2011. Measuring the evolutionary rewiring of biological networks. *PLoS Comput Biol* **7**: e1001050.
33. **Janga SC**, **Salgado H**, **Martinez-Antonio A.** 2009. Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nucleic Acids Res* **37**: 3680–8.
34. **Lozada-Chavez I**, **Janga SC**, **Collado-Vides J.** 2006. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res* **34**: 3434–45.
35. **Babu MM**, **Teichmann SA**, **Aravind L.** 2006. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol* **358**: 614–33.
36. **Rodrigues JF**, **Wagner A.** 2009. Evolutionary plasticity and innovations in complex metabolic reaction networks. *PLoS Comput Biol* **5**: e1000613.
37. **Samal A**, **Rodrigues JFM**, **Jost J**, **Martin OC**, et al. 2010. Genotype networks in metabolic reaction spaces. *BMC Syst Biol* **4**: 30.
38. **Rodrigues JF**, **Wagner A.** 2011. Genotype networks, innovation, and robustness in sulfur metabolism. *BMC Syst Biol* **5**: 39.
39. **Schuster P**, **Fontana W**, **Stadler P**, **Hofacker I.** 1994. From sequences to shapes and back – a case-study in RNA secondary structures. *Proc Biol Sci* **255**: 279–84.
40. **Fontana W**, **Schuster P.** 1998. Continuity in evolution: On the nature of transitions. *Science* **280**: 1451–5.
41. **Huang W**, **Petrosino J**, **Hirsch M**, **Shenkin P**, et al. 1996. Amino acid sequence of beta-lactamase structure and activity. *J Mol Biol* **258**: 688–703.
42. **Rennell D**, **Bouvier S**, **Hardy L**, **Poteete A.** 1991. Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol* **222**: 67–87.
43. **Kleina L**, **Miller J.** 1990. Genetic studies of the lac repressor. 13. Extensive amino-acid replacements generated by the use of natural and synthetic nonsense suppressors. *J Mol Biol* **212**: 295–318.
44. **Rost B.** 1997. Protein structures sustain evolutionary drift. *Fold Des* **2**: S19–S24.
45. **Copley RR**, **Bork P.** 2000. Homology among ($\beta\alpha$)8 barrels: implications for the evolution of metabolic pathways. *J Mol Biol* **303**: 627–40.
46. **Goodman M**, **Pedwaydon J**, **Czelusniak J**, **Suzuki T**, et al. 1988. An evolutionary tree for invertebrate globin sequences. *J Mol Evol* **27**: 236–49.
47. **Reidhaar-Olson JF**, **Sauer RT.** 1990. Functionally acceptable substitutions in 2 alpha-helical regions of Lambda repressor. *Proteins* **7**: 306–16.
48. **Oliver KR**, **Greene WK.** 2009. Transposable elements: powerful facilitators of evolution. *BioEssays* **31**: 703–14.
49. **Li W-H.** 1997. *Molecular Evolution*. Massachusetts: Sinauer Associates.
50. **Gilbert W.** 1978. Why genes in pieces? *Nature* **271**: 501.
51. **Buljan M**, **Bateman A.** 2009. The evolution of protein domain families. *Conference on Protein Evolution Sequences, Structures and Systems*. England: Cambridge. pp. 751–5.
52. **Koonin E**, **Wolf Y**, **Karev G.** 2002. The structure of the protein universe and genome evolution. *Nature* **420**: 218–23.
53. **Stadler B**, **Stadler P**, **Shpak M**, **Wagner G.** 2002. Recombination spaces, metrics, and pretopologies. *Z Phys Chem* **216**: 217–34.
54. **Azevedo RBR**, **Lohaus R**, **Srinivasan S**, **Dang KK**, et al. 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* **440**: 87–90.
55. **Bornholdt S**, **Sneppen K.** 2000. Robustness as an evolutionary principle. *Proc Biol Sci* **267**: 2281–6.
56. **Wagner A.** 1996. Does evolutionary plasticity evolve? *Evolution* **50**: 1008–23.
57. **Siegal M**, **Bergman A.** 2002. Waddington's canalization revisited: Developmental stability and evolution. *Proc Natl Acad Sci USA* **99**: 10528–32.
58. **Bergman A**, **Siegal M.** 2003. Evolutionary capacitance as a general feature of complex gene networks. *Nature* **424**: 549–52.
59. **MacCarthy T**, **Bergman A.** 2007. Coevolution of robustness, epistasis, and recombination favors asexual reproduction. *Proc Natl Acad Sci USA* **104**: 12801–6.
60. **Huerta-Sanchez E**, **Durrett R.** 2007. Wagner's canalization model. *Theor Popul Biol* **71**: 121–30.
61. **Mjolsness E**, **Sharp DH**, **Reinitz J.** 1991. A connectionist model of development. *J Theor Biol* **152**: 429–53.
62. **Reinitz J**, **Mjolsness E**, **Sharp DH.** 1995. Model for cooperative control of positional information in *Drosophila* by bicoid and maternal hunchback. *J Exp Zool* **271**: 47–56.
63. **Reinitz J.** 1999. Gene circuits for eve stripes: reverse engineering the *Drosophila* segmentation gene network. *Biophys J* **76**: A272.
64. **Sharp DH**, **Reinitz J.** 1998. Prediction of mutant expression patterns using gene circuits. *Biosystems* **47**: 79–90.
65. **Wagner A.** 1996. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. *Biol Cybern* **74**: 557–67.
66. **Stone J**, **Wray G.** 2001. Rapid evolution of cis-regulatory sequences via local point mutations. *Mol Biol Evol* **18**: 1764–70.
67. **Wray G**, **Hahn M**, **Abouheif E**, **Balhoff J**, et al. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**: 1377–419.
68. **Tirosh I**, **Reikhav S**, **Levy AA**, **Barkai N.** 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **324**: 659–62.
69. **Wittkopp PJ**, **Haerum BK**, **Clark AG.** 2008. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet* **40**: 346–50.
70. **Wittkopp PJ**, **Haerum BK**, **Clark AG.** 2004. Evolutionary changes in cis and trans gene regulation. *Nature* **430**: 85–8.
71. **Martin OC**, **Wagner A.** 2009. Effects of recombination on complex regulatory circuits. *Genetics* **183**: 673–84.
72. **Cui Y**, **Wong W**, **Bornberg-Bauer E**, **Chan H.** 2002. Recombinatoric exploration of novel folded structures: a heteropolymer-based model of protein evolutionary landscapes. *Proc Natl Acad Sci USA* **99**: 809–14.
73. **Voigt C**, **Martinez C**, **Wang Z**, **Mayo S**, et al. 2002. Protein building blocks preserved by recombination. *Nat Struct Biol* **9**: 553–8.

74. **Branden C**, **Tooze J.** 1999. *Introduction to Protein Structure*. New York: Garland.
75. **Muller HJ.** 1950. Our load of mutations. *Genetics* **2**: 111–75.
76. **Davis B**, **Whitlock MC.** 2004. *Genetic load. Encyclopedia of Life Sciences (http://www.els.net)*. Chichester: John Wiley & Sons Ltd.
77. **Burger R.** 2000. *The Mathematical Theory of Selection, Recombination, and Mutation*. Chichester, UK: Wiley.
78. **Xia Y**, **Levitt M.** 2002. Roles of mutation and recombination in the evolution of protein thermodynamics. *Proc Natl Acad Sci USA* **99**: 10382–7.
79. **Kim KJ**, **Fernandes VM.** 2009. Effects of ploidy and recombination on evolution of robustness in a model of the segment polarity network. *PLoS Comput Biol* **5**: e1000296.
80. **Chang C**, **Chen T**, **Cox B**, **Dawes G**, et al. 1999. Evolution of a cytokine using DNA family shuffling. *Nat Biotechnol* **17**: 793–7.
81. **Isalan M**, **Lemerle C**, **Michalodimitrakis K**, **Beltrao P**, et al. 2008. Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**: 840–5.
82. **Bing N**, **Hoeschele I.** 2005. Genetical genomics analysis of a yeast segregant population for transcription network inference. *Genetics* **170**: 533–42.
83. **Barton NH**, **Charlesworth B.** 1998. Why sex and recombination? *Science* **281**: 1986–90.
84. **Otto SP**, **Lenormand T.** 2002. Resolving the paradox of sex and recombination. *Nat Rev Genet* **3**: 252–61.
85. **Otto SP**, **Gerstein AC.** 2006. Why have sex? The population genetics of sex and recombination. *Biochem Soc T* **34**: 519–22.
86. **Misevic D**, **Kouyos RD**, **Bonhoeffer S.** 2009. Predicting the evolution of sex on complex fitness landscapes. *PLoS Comput Biol* **5**: e1000510.
87. **Kouyos RD**, **Salathe M**, **Otto SP**, **Bonhoeffer S.** 2009. The role of epistasis on the evolution of recombination in host-parasite coevolution. *Theor Popul Biol* **75**: 1–13.
88. **Misevic D**, **Ofria C**, **Lenski RE.** 2006. Sexual reproduction reshapes the genetic architecture of digital organisms. *Proc Biol Sci* **273**: 457–64.
89. **Wagner A.** 2011. The origins of evolutionary innovations. *A Theory of Transformative Change in Living Systems*. Oxford, UK: Oxford University Press.