The evolution of gene duplications: classifying and distinguishing between models

Hideki Innan[*‡] *and Fyodor Kondrashov*[§]

Abstract | Gene duplications and their subsequent divergence play an important part in the evolution of novel gene functions. Several models for the emergence, maintenance and evolution of gene copies have been proposed. However, a clear consensus on how gene duplications are fixed and maintained in genomes is lacking. Here, we present a comprehensive classification of the models that are relevant to all stages of the evolution of gene duplications. Each model predicts a unique combination of evolutionary dynamics and functional properties. Setting out these predictions is an important step towards identifying the main mechanisms that are involved in the evolution of gene duplications.

Our knowledge of gene duplications at the phylogenetic, functional and genomic levels is impressive. Hardly any aspect of genome evolution or function is not somehow linked to gene duplications, which occur in all kinds of life forms[1] and have taken place since before the last universal common ancestor[2]. Some genomes contain large numbers of genes owing to whole-genome duplications[3] or lineage-specific gene expansions[4]. The duplication of genes may form a cornerstone in the evolution of biological complexity[5]. Finally, gene duplications segregate in high numbers in natural populations, and some cause disease[6] or confer an adaptive advantage[7]. These and other aspects of gene duplications have been continuously reviewed and their importance has been emphasized[2–4,8].

By contrast, our current understanding of the selective and evolutionary mechanisms involved in the emergence, maintenance and evolution of gene duplications is preliminary and fragmented. There are two main reasons for this discordance between theory and data. First, the literature contains many different models of and hypotheses about the evolution of gene duplications, which have not been described in a systematic way. Second, the development of theory on gene duplications has been driven by data, without attempts to stringently test hypotheses. Consequently, although many models for the evolution of gene duplications have been proposed, we still do not know their relative importance for describing general trends in the evolution of gene duplications or their applicability to specific gene copies.

To change this situation, we present here a classification of the different hypotheses on the emergence, maintenance and evolution of gene duplications. Following previous reviews of specific models or model types[2,9–11], we provide a comprehensive classification of all models of gene duplications. We include verbal and more formal models, and compare their predictions for the anticipated evolutionary scenarios in which gene duplications are retained, and for the fates of gene copies at the different stages of the evolution of a gene duplication. Capitalizing on the differences that we point out between the models, we also provide suggestions for future experiments aimed at distinguishing between them. We do not review evidence that supports or refutes any of the hypotheses. Rather, our aim is to bring order to the theoretical literature on gene duplications, which we hope will aid researchers in designing better studies to improve our understanding of the function and evolution of duplicated genes.

Classification and predictions

Every type of genetic change undergoes three main stages as it competes for evolutionary preservation: origin through mutation, a fixation phase when it segregates in the population and a preservation phase when the fixed change is maintained. Gene duplications follow this trajectory with one important addition; from the moment of emergence of a new gene copy, the acquisition of genetic differences between the copies can alter the chances of both copies being preserved. Although all of these life stages are functionally and evolutionarily important,

*Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan.
‡PRESTO, Japan Science and Technology Agency, Saitama 332-0012, Japan.
§Bioinformatics and Genomics Programme, Centre for Genomic Regulation, C/Dr Aiguader 88, Barcelona Biomedical Research Park Building, 08003 Barcelona, Spain.
e-mails: innan_hideki@soken.ac.jp; fyodor.kondrashov@crg.es
doi:10.1038/nrg2689
Published online 6 January 2010

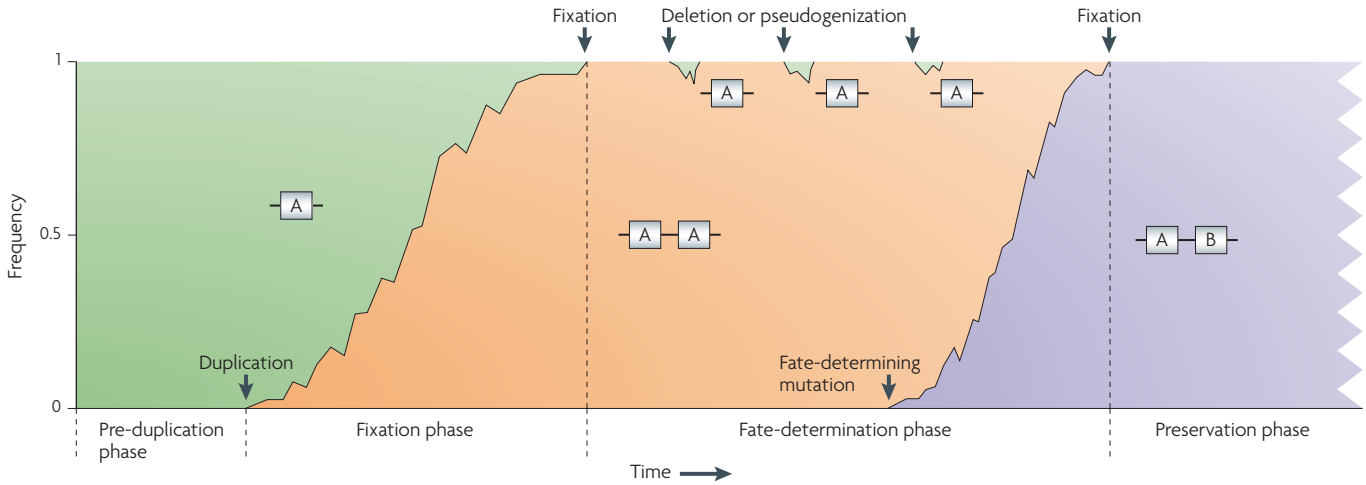


Figure 1 | **Phases leading to the stable preservation of a duplicated gene.** Typical behaviour of the frequency of a newly arisen duplicated gene is shown. Although the figure is based on the neofuntionalization model, it is applicable to all models with slight modifications. In the pre-duplication phase, the single-copy genotype (A) is fixed in the population; when a duplicate arises, the fixation phase begins. The duplicate is most likely to be lost to drift but can also achieve fixation. After the duplicated genotype (A–A) is fixed, the fate-determination phase begins and continues until the fixation of a fate-determining mutation. Note again that in some models the duplicate is likely to be pseudogenized owing to the fixation of a null mutation. Once the preservation phase is reached, the two copies are stably maintained by selection. Note that this figure shows the fixation and fate-determination phases separately; however, the two phases can overlap when a fate-determining mutation arises before the fixation of the duplicated copy or if the pre-existing allele works as a fate-determining mutation (as in models in category III). The situation in which a fate-determining mutation arises before fixation may be important when the product of the fate-determining mutation rate and the population size is large[2,26]. If the fixation and fate-determination phases overlap, multiple selective forces can operate simultaneously, and the process becomes complicated.

many models of gene duplication evolution describe the phase of acquisition of differences between gene copies as crucial in the preservation of new gene copies. Therefore, we call it the fate-determination phase (FIG. 1).

Approximately a dozen models for the evolution and maintenance of gene duplications have been proposed over the years. We describe the models as they have been articulated in the original literature (summarized in TABLE 1 and FIG. 2) and describe their development. To classify and distinguish between these models, it is convenient and useful to focus on the selective forces and evolutionary events at different stages of the life history of the duplication. However, there is substantial overlap in the descriptions and predictions of different models in the same category.

Suppose that a new duplicate gene pair (A–A) arises in a population with N random-mating diploids, in which all genomes initially have single copies of gene A. Throughout this Review, to be consistent with the models in the literature, we assume that the new duplicate has a complete set of functional motifs and is functionally indistinguishable from the original copy unless otherwise specified. The probability of the fixation of A–A and the length of the fixation phase (fixation time) theoretically depend on the relative strength of selection for the A–A and A genotypes. This is one of the most important factors that differentiate the models and we use it as the basis for our classification. When A–A confers no selective advantage or a disadvantage (that is, it is neutral), A–A will be fixed in the population at a probability of $1 / 2N$ and the fixation process takes on average $4N$ generations. This is the defining feature of the models we place in category I, including the popular neofunctionalization and subfunctionalization models (see below). By contrast, the models in categories II and III involve positive selection for the new duplicate. In these cases, the fixation probability is higher and the fixation time is shorter than in the neutral case. We place the dosage balance model in a separate category because it lacks the fixation phase and considers a pair of duplicates that are created by a whole genome duplication.

Following our systematic classification of the models, differences in the polymorphism and divergence dynamics can be used to distinguish between them. For each model we describe the most likely pattern in polymorphism levels and sequence divergence (see BOX 1 for a description of these variables), and aspects of gene function. In particular, we discuss the synonymous–non-synonymous ratios of polymorphism and divergence, ω_π and ω_K, which should represent the intensity of selection as described in BOX 2. Few models provide specific predictions about the long-term molecular evolution of the two duplicates in the final preservation phase. However, because most studies of the evolution of gene duplications focus on long-term divergence, based on these data we can provide the most likely predictions for this aspect of duplication evolution. Importantly, when making predictions we take into account the effect of gene conversion, which is common in many species[12–18] and has a strong influence on the dynamics of the evolution

Gene duplication
The emergence of a heritable copy of a gene.

Neofunctionalization
The random acquisition of a new function in the course of the accumulation of neutral mutations in duplicated genes.

Subfunctionalization
The process of the accumulation of degenerate mutations in gene copies that subdivides gene function among the duplicated genes. This term has been introduced to describe the mechanism of the duplication–degeneration–complementation model, but it is often used indiscriminately to describe any subdivision of function.

Table 1 | **Summary of the models of gene-duplication evolution**

Name	Functional evolution			Fixation phase	Fate-determination phase		Preservation phase*	
	Function of original copy	Function of new copy	Fate-determining mutation	Selection on new copy	Selection on original copy	Selection on new copy	Molecular evolution in original copy	Molecular evolution in new copy
Category I								
Neofunctionalization	Kept	Novel	Gain-of-function mutations	Neutral	Purifying selection	Neutral	α	β
DDC	Subfunctionalized	Subfunctionalized	Loss-of-function mutations	Neutral	Relaxed purifying selection	Relaxed purifying selection	β	β
Specialization or EAC	Subfunctionalized	Subfunctionalized	Gain-of-function mutations	Neutral	Relaxed purifying selection	Relaxed purifying selection	β	β
Category II								
Positive dosage	Kept	Same as original	NA	Positive selection on duplication	NA	NA	α′	α′
Shielding against deleterious mutations	Kept	Same as original	NA	Positive selection on duplication	Relaxed purifying selection	Relaxed purifying selection	NA	NA
Modified duplication	Kept	Novel	Gain-of-function mutations	Positive selection on duplication	NA	NA	α	β
Category III								
Permanent heterozygote	Subfunctionalized	Subfunctionalized	Gain-of-function mutations	Positive selection on pre-duplicational variation	NA	NA	β	β
Adaptive radiation model	Kept	Novel	Gain-of-function mutations	Positive selection on pre-duplicational variation	NA	NA	α	β
Diversifying selection	Multiple functions	Multiple functions	Gain-of-function mutations	Positive selection on pre-duplicational variation	NA	NA	o	o
Category IV								
Dosage balance	Kept	Original	NA	NA	NA	NA	α′	α′

*The predicted pattern of molecular evolution is indicated as α when the pattern is not different from that in the pre-duplication phase (as α′ when the selective pressure may be relaxed), as β when amino acid substitutions can be accelerated by positive selection and as o when amino acid substitutions are always accelerated by diversifying selection. DDC, duplication–degeneration–complementation; EAC, escape from adaptive conflict; NA, not applicable.

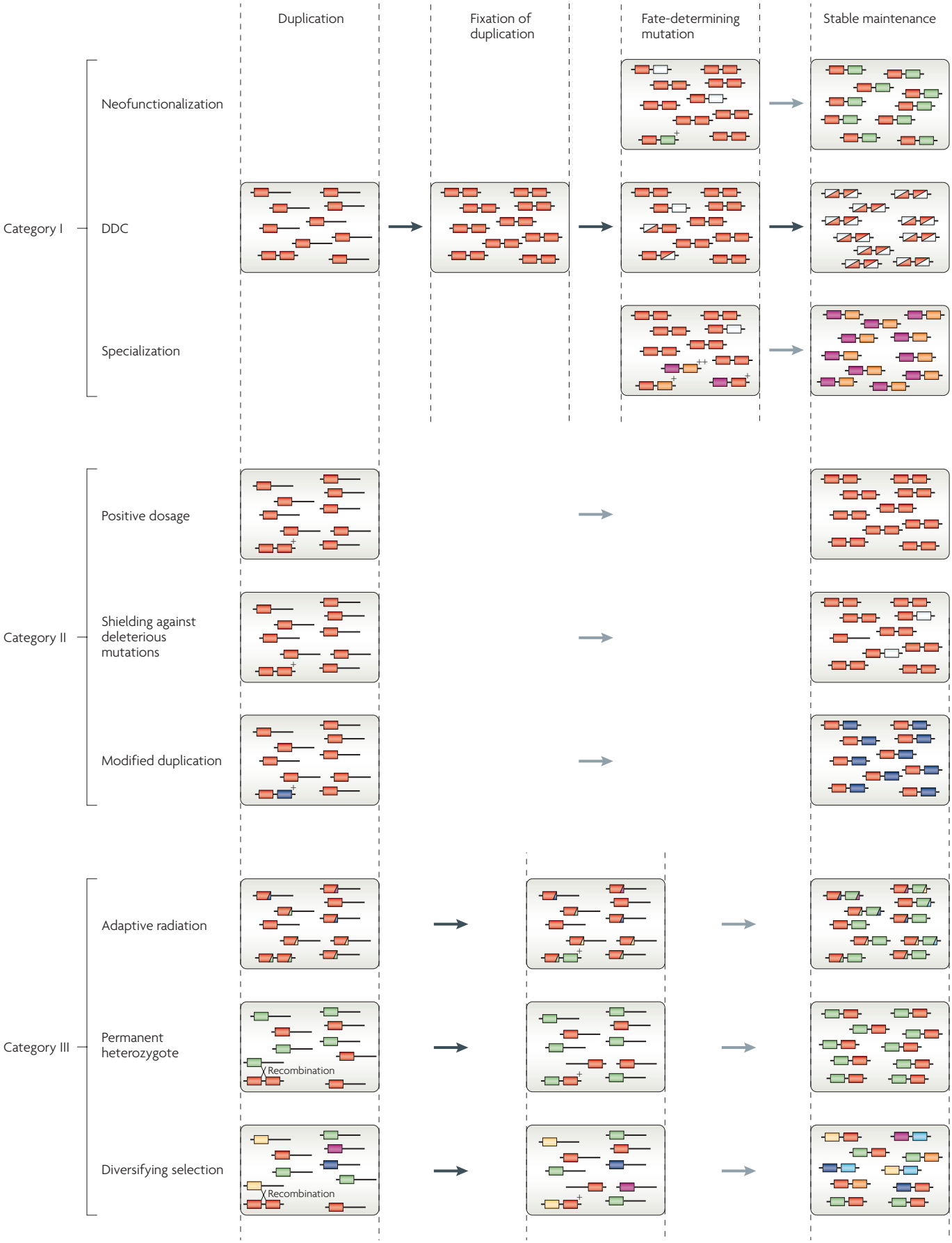
of gene duplicates in the fixation, fate-determination and early preservation phases (BOXES 3,4).

Category I

This category contains three models that assume that a duplication does not affect fitness, so that the fixation of the duplicated copy is a neutral process: the neofunctionalization, duplication–degeneration–complementation (DDC) and specialization models. A general feature of these models is that a gene duplication must go through the fate-determination phase rapidly to reach the preservation phase, otherwise one of the copies can be pseudogenized because selection is relaxed before the preservation stage. The differences among these models begin at the fate-determination phase.

Specialization
A process of improvement of different aspects of gene function in each gene copy, which is driven by positive selection.

(Category I-a) Ohno's neofunctionalization. Ohno's neofunctionalization model marked the beginning of the theoretical discussion of gene duplications[19]. He reasoned that a single gene copy is enough to fulfil the function of the gene and therefore extra copies are redundant. If such a redundant copy is fixed by drift in the population, the original copy will maintain its function, and the new copy will be relieved from negative selection[19,20]. The new copy can therefore be pseudogenized or lost through the accumulation of neutral loss-of-function mutations. However, Ohno suggested that occasionally, as the redundant, dying gene copy accumulates substitutions, it may acquire a new gene function that will be maintained by selection. It is not clear how selection can distinguish between the new and original copies of the



◀ Figure 2 | **Illustration of the models in categories I, II and III for gene-duplication evolution.** The processes that are expected to occur for each model from the birth of a gene duplication to its stable maintenance. Red boxes represent genes with the original function, functional evolution is shown by changes of colour, and white indicates a loss of gene function. The darker arrows indicate situations in which the process is neutral, and the light arrows indicate cases in which positive selection is involved. DDC, duplication–degeneration–complementation.

gene and how the new function can evolve in either copy. Nevertheless, we follow Ohno's description of the model because some of its predictions depend on the assumption that one of the copies remains under selection while the other is free from it.

This model *sensu stricto* predicts that the rate of evolution after gene duplication will be accelerated in the duplicated copy and maintained in the original one. Because the new function in the derived copy can be improved by additional advantageous mutations in the preservation phase, it is very likely that positive selection specifically targets the derived copy. Meanwhile, as the original copy keeps the ancestral function, its selective constraint may not change much throughout the process. The parameters of this model of the time to fixation and to non-functionalization (pseudogenization) have been difficult to work out because they are based on aspects of gene function, such as the fraction of loss-of-function mutations out of all possible ones, which are yet to be accurately determined. However, because most *de novo* mutations are likely to be deleterious, it is widely agreed that pseudogenization is the most probable fate of the new gene copy[19,21–23]. The rate at which a neutrally evolving gene copy may acquire a new function is the most difficult parameter to estimate and depends at least on the strength of selection in favour of the new function[24]. The length of time before a redundant gene copy is pseudogenized is likely to be short[22,23,25,26].

The predicted behaviour of polymorphism under this model is described in BOX 5. Asymmetry of divergence is expected, such that $\omega_K = \omega_{K0} < 1$ in the old copy and $\omega_K = 1$ in the new one until the fate-determining phase. Such asymmetry is also expected in the preservation phase, as ω_K for the new copy can be increased by positive selection; occasionally $\omega_K > 1$ for the new copy and $\omega_K = \omega_{K0}$ for the original copy. When the derived copy is not lost or pseudogenized, the evolution of a radically new function is expected. Gene conversion can inhibit the acquisition of a new function in this model (BOX 3).

(Category I-b) Duplication–degeneration–complementation. The DDC model was postulated by Force *et al.*[27] and can be broadly seen as an extension of Ohno's neofunctionalization model, as both models are based on the redundancy of a gene duplication. The difference between the models is that Force *et al.* suggested that the accumulation of degenerate mutations, which are neutral mutations that are damaging to gene function, can reduce the functional efficacy of both duplicated genes. After degenerate mutations have been fixed in both copies by drift, neither copy is sufficient to perform the original function, and the two copies are subfunctionalized so that both must be maintained by selection[27–29]. The division of function might be due to changes in the regulatory regions of the gene copies, which may lead to differential expression patterns of the two copies, such that the locations of expression of the original gene would be shared between the two subfunctionalized copies. Alternatively, the function of the protein encoded by the gene may be subfunctionalized.

The most likely outcome under these conditions is the same as in Ohno's neofunctionalization model: the early pseudogenization of one copy before both become essential through the fixation of null mutations[27–29]. The question of the expected time to subfunctionalization has been addressed[24,28,29]; however, estimates of this parameter are dependent on functional aspects of genes that may make them more or less prone to subfunctionalization when duplicated, which are difficult to quantify. In addition, a new function may evolve in the preservation phase[24,30].

After duplication, purifying selection is expected in both genes, but its intensity may be relaxed compared with the pre-duplication phase. This model requires at least two compensatory null mutations in different *cis*-regulatory elements or different protein domains to act as fate-determining mutations. Preservation must occur rapidly and therefore the rate of evolution of the copies should be nearly symmetrical. At the end of the fate-determination phase, the original function will be partitioned by the two genes in terms of expression or protein function in a neutral manner without the involvement of positive selection. Once the two copies have taken on different functions, the rate of amino acid evolution in one or both copies might be increased because there could be different optimum protein functions for different subfunctional roles.

It is possible to make predictions from this model; see BOX 5 for a discussion of polymorphism. A symmetrical rate of divergence is expected in the fate-determination phase, with $\omega_{K0} < \omega_K < 1$. However, in the preservation phase, divergence may target different parts of the genes and can lead to differential tissue or cellular compartmentalization of the two copies[31–35]. The combined functions of the paralogues and their contributions to fitness will be equivalent to the functions and fitness contributions of unduplicated orthologues in closely related species.

(Category I-c) Specialization or escape from adaptive conflict. Hughes[36] proposed a different verbal model of functional specialization of gene copies. This model did not challenge the complete redundancy assumption of Ohno's neofunctionalization model and assumed that gene copies are fixed by genetic drift. However, it suggested that if the original gene was performing two functions that could not be independently improved, then after duplication each gene copy can be driven by positive selection to specialize — that is, to improve one of the two functions[37]. The improvement could be owing to either differentiation of expression patterns[38] or the substantial improvement of a promiscuous function[39]. This model was recently rechristened as the escape from adaptive conflict (EAC) model[40].

Non-functionalization
The process of the accumulation of neutral mutations in a duplicated gene that renders the gene copy non-functional. Also known as pseudogenization.

Degenerate mutation
A mutation that does not affect fitness but is damaging to gene function.

Promiscuous function
A secondary, possibly neutral function that is performed by a protein with another primary function. It is an example of gene sharing.

Gene sharing
An early term describing situations in which a gene has more than one function. Modern studies describe such genes as multifunctional.

Gene dosage
The amount of product produced from a gene; broadly equivalent to gene expression.

Gene amplification
The emergence of a non-heritable extra copy of a gene in a somatic tissue. In microorganisms this term can be used interchangeably with gene duplication.

This model has not been explored formally. However, because the gene copy is fixed by drift the dynamics of the fixation phase will be the same as in the neofunctionalization and DDC models. In the fate-determination phase the divergence of the copies will be affected by positive selection for improvement of the different functions, and the adaptive mutations will act as the fate-determining events.

The predicted pattern of polymorphism for this model is explored in BOX 5. For long-term evolution after the fate-determination phase, ω_K for both copies can be increased by positive selection, which can lead to similar functional, tissue or cellular compartmentalization to the DDC model. Because the gene functions are improved compared with the original gene, the fitness contributions of the two paralogues should be greater than those of unduplicated orthologues in closely related species. Multifunctional genes should be preferentially duplicated by this mechanism. For this model to apply widely, gene sharing and antagonistic pleiotropy must be common, such that an individual function of a multifunctional gene can be improved only by harming another functional aspect of the same gene. Gene conversion will inhibit the specialization process (BOX 3).

Category II

For the models in this category, the duplication itself is advantageous. There can be at least three reasons for this type of adaptation: beneficial increases in dosage, the masking of deleterious mutations and the opportunity for the immediate emergence of a new function.

(Category II-a) Beneficial increase in dosage. Ideas relating to the effect of gene duplication on gene dosage have been developed in parallel to the neofunctionalization model (reviewed in REFS 1,19). Early suggestions of an interplay between fitness and gene copy number tended to focus on bacteria and archaea[1,41–45], and the terms gene duplication and gene amplification were often used interchangeably. However, as data from eukaryotes became more widely available, authors started to relate dosage and gene duplications in multicellular organisms[1,43,46–48]. The argument for the basis of the interplay between dosage and gene duplications is simple: if an increase of dosage of a particular gene is beneficial then a duplication of this gene may be fixed by positive selection. After fixation, the gene duplication will be maintained as long as the conditions that led to the selection for increased dosage remain. When the increase in fitness owing to the fixed gene duplication is small, the gene copies are expected to be under relaxed selection pressure[1]. The model of beneficial increase of gene dosage also intrinsically connects gene copy number differences and deleterious changes in dosage that may be caused by duplication or loss of gene copies[1,48–50].

This model may apply to three types of genes. First, genes that mediate the interaction between the organism and the environment, such as stress-response genes, sensory genes, transport genes and genes that have a metabolism-related function[1]. Second, genes that have dosage-sensitive functions owing to protein–protein interaction properties of their products or the nature of the metabolic pathways in which they act[49,50]. Third, genes that have products that are generally required in large doses, such as ribosomal or histone genes[1,51].

There have been few formal modelling attempts for this type of model[1,49,50]; however, a few general trends can be outlined. Under unvarying strong selection for increased dosage, the duplication will quickly become fixed in the population and will be preserved without undergoing a fate-determination phase. Throughout the process both copies will be under negative selection. However, if selection for the duplicated copy is weak, a null mutation might become fixed by random genetic drift, resulting in pseudogenization. This situation is similar to models in category I, in that a fate-determining mutation may be needed to reach the preservation phase. However, in contrast to category I models, weak selection for dosage increases the fixation probability and shortens the fixation phase. More importantly, the fate-determination phase can be longer, so that there is more chance of a fate-determining mutation occurring before pseudogenization takes place as a result of negative selection against null and deleterious mutations. The long-term molecular evolution in the preservation phase follows that of the models in category I.

In a variable environment, selection for increased dosage may occur but then be replaced by selection against dosage increases, leading to a cycle of gene duplication and loss[52]. Alternatively, if selection for increased dosage is strong and constant, there will be few gene duplications under near-equilibrium conditions.

It is possible to make predictions from this model; see BOX 5 for polymorphism patterns. If selection on dosage is strong, then $\omega_K < 1$ throughout the whole process. For recent duplications, the sum of the expression levels of

Box 1 | **Variables used throughout the Review**

- μ: per generation mutation rate
- *N*: population size
- θ: population variability equal to $2N\mu$
- π: the average number of pairwise nucleotide differences
- K_S: rate of synonymous evolution, measured as the number of synonymous substitutions divided by the number of synonymous sites
- K_N: rate of non-synonymous evolution, measured as the number of non-synonymous substitutions divided by the number of non-synonymous sites
- π_S: synonymous level of polymorphism, measured as the average number of pairwise synonymous differences divided by the number of synonymous sites
- π_N: non-synonymous level of polymorphism, measured as the average number of pairwise non-synonymous differences divided by the number of non-synonymous sites
- ω_K: equal to K_N/K_S, representing the strength of selection on non-synonymous substitutions
- ω_π: equal to π_N/π_S, representing the strength of selection on non-synonymous polymorphisms
- Subscript 0: the expected measurement of any of the above variables of a duplicated gene before duplication, in the pre-duplication phase. For example, π_{S0} is the expected level of synonymous polymorphisms in a gene immediately before duplication

Box 2 | **Effect of selection on polymorphism and divergence**

The most important parameters that affect the pattern of polymorphism and divergence are the mutation rate and population size. We define the neutral mutation rate as μ per site per generation. Throughout this Review, we assume a random-mating diploid population that has a constant size *N*, so that the level of polymorphism is characterized by $\theta = 4N\mu$. If the polymorphism level is measured by π, the average number of pairwise nucleotide differences, then at mutation-drift equilibrium the expectation of π at a neutral site is given by $E(\pi) = \theta$ for a single-locus gene and also in each of the duplicated genes when they reach equilibrium (but see BOX 4 when gene conversion is active). The expected rate of evolution (nucleotide substitution rate) in each copy is μ, which determines the divergence at neutral sites between the two copies.

In contrast to neutral sites, other patterns are expected at functional sites owing to the actions of positive and negative selection. A common approach for evaluating selection, which we also use in our Review, is to compare the patterns of polymorphism and divergence at synonymous and non-synonymous sites[79–81]. The rationale is that synonymous sites can be used as a neutral control and non-synonymous sites are more likely to be functional and subject to selection.

We define π_S and K_S as the observed levels of polymorphism and divergence, respectively, at synonymous sites, and π_N and K_N are the levels of polymorphism and divergence, respectively, at non-synonymous sites. ω_π and ω_K represent the expected non-synonymous–synonymous ratios of polymorphism and divergence ($\omega_\pi = E(\pi_N/\pi_S)$ and $\omega_K = E(K_N/K_S)$), respectively. Under neutrality, $\omega_\pi = \omega_K = 1$, so these measurements are commonly used as measures of selection. We add the subscript 0 to denote measurements of the single-copy gene in the pre-duplication phase (for example, π_{S0} is the level of synonymous polymorphisms in a gene immediately before duplication). Such measurements can be used as a control to quantitatively measure the change in selective pressure after gene duplication.

Negative selection constantly eliminates deleterious mutations, leading to reduction of the level of polymorphism and the rate of evolution at functional sites ($\omega_\pi < 1$ and $\omega_K < 1$). When negative selection is relaxed, both ω_π and ω_K increase up to a maximum value of 1. Positive selection can drive a beneficial mutation to fixation, temporarily reducing the level of polymorphism in the surrounding region. If beneficial mutations are constantly fixed, there is an increased rate of evolution, resulting in an increase of ω_K. Therefore, there are two possible explanations for an increase in ω_K — relaxation of negative selection and the action of positive selection — and other data are usually needed to distinguish between them. However, if $\omega_K > 1$ then the action of positive selection can be inferred.

the original and duplicated copies should exceed that of the non-duplicated orthologue of a closely related species. If strong or weak selection for extra dosage is caused by a change of environment, the functional repertoire of the duplicated genes should be relevant to the interaction of the gene with the changed environmental aspect. Two aspects of gene conversion can promote this process (the first two 'effects that make preservation more likely' in BOX 3).

(Category II-b) Shielding against deleterious mutations. Gene copies may shield one another from the accumulation of deleterious mutations[53], provided that they are at least partially redundant. An individual that has just obtained a new gene copy may have a slight selective advantage over those without one because new mutations will not negatively affect fitness, and the individual will be subject to a slightly lower mutational load[54]. However, the selective advantage of such shielding is expected to be almost insignificant (that is, of the order of the mutation rate[54–56]), but is expected to increase with the rate of mutation, the rate of duplication or the length of duplicated segments[51,56,57]. The extra gene copies are quickly destroyed by mutations and therefore duplications must occur cyclically to provide sustained sheltering against deleterious mutations[51,57]. Therefore, there is no preservation phase because the only function of the gene copy is to be destroyed by mutations.

This model predicts that repeated and rapid birth and death of extra gene copies will occur, and that each copy is present for a short time in the genome, with the new copy accumulating mutations in a neutral fashion and eventually being pseudogenized. $\omega_{K0} < \omega_K < 1$ when a copy is functional and $\omega_K = 1$ after it is pseudogenized. Gene conversion may prolong the birth and death cycle. If this model applies widely, rapid loss of synteny and many traces of pseudogenes should be observed in the genome.

(Category II-c) Gene duplication with a modified function. If the process of gene duplication itself creates a new function, the new copy can be fixed and preserved by positive selection. Partial gene duplication may fail to copy regulatory elements or other functional parts of the gene, resulting in a new function from the moment of duplication[58,59]. Similarly, a new genomic location of the gene copy may introduce new functional aspects[60], or a retrotransposed copy may recruit regulatory elements in the new location or be integrated into an existing gene, which results in the formation of a chimeric gene[61–63]. If a new and beneficial gene function emerges at the point of duplication then fixation will occur through positive selection for the new function and the gene copy will be immediately preserved. It is likely that the function of the derived copy can be improved by additional mutations, leading to a period of rapid evolution immediately after duplication.

It is predicted that a selective sweep should be observed after the duplication event (BOX 5). In some cases $\omega_K > 1$ in the new copy owing to positive selection. A radically new function will be seen for the new copy, which may have lost or gained promoters, domains or introns.

Category III

This category considers models in which duplication occurs in a gene for which genetic variation exists in the population. In some cases, these polymorphisms can immediately become fate-determining mutations that promote fixation of the duplicated copy. The duplicate and the fate-determining mutations are fixed almost at the same time, and these models do not have a fate-determining phase. A common feature of the three models in this category is that a polymorphic allele that is present in the pre-duplication phase is fixed at the same time as the new copy, which requires recombination between the two copies. If two alleles A and B are polymorphic in the ancestral single-locus gene, then a duplication event can make only A–A or B–B haplotypes, not the advantageous A–B or B–A haplotypes. An advantageous haplotype can arise through recombination; for example, between a single-copy A and duplicated B–B copies. Positive selection can then fix A–B or B–A (FIG. 2).

Box 3 | **Effect of gene conversion on the fate of duplicated genes**

Gene conversion is a non-reciprocal recombination process, which is usually described as a copy-and-paste event, and which transfers a short tract of DNA that ranges from several to (occasionally) thousands of base pairs[82,83]. Gene conversion occurs not only between orthologous regions but also between paralogous regions when they have sufficient sequence identity (as shown in the figure). We focus on paralogous gene conversion events (interlocus or ectopic gene conversion) because they have a substantial effect on the molecular evolution of duplicated genes, especially during the early stages after duplication. Gene conversion shuffles the DNA variation between paralogous duplicated regions, resulting in a substantial reduction in the divergence between duplicates. When gene conversion occurs at a reasonably high rate, the duplicates co-evolve for a long time[84] (concerted evolution).

Gene conversion can make the preservation of a duplicated gene more or less likely in the following ways.

Effects that make preservation more likely

- Frequent gene conversion keeps the sequence identity between paralogues high. This effect should promote the retention of both copies when a large amount of a single gene product is required; this can be applied to the model of the beneficial increase of dosage[51]. High sequence identity may also be preferred in the dosage-balance model.
- Gene conversion allows both copies to share a beneficial mutation, and deleterious mutations can be more efficiently removed from the population[85]. These beneficial effects hold for any model.
- Because gene conversion transfers a short DNA tract, it can create new allelic combinations that confer an advantage[86,87].

Effects that make preservation less likely

- The homogenizing effect of gene conversion should be disfavoured in all models in which a fate-determining mutation is involved[88], because selection will operate to maintain the difference between the original and duplicated copies that is created by the mutations. In such a situation, population genetic theory predicts an interesting pattern of polymorphism and divergence (BOX 4).
- Gene conversion from a pseudogenized copy to a functional copy will be deleterious. In humans, there are several genetic diseases that are caused by gene conversion from pseudogenes[89].

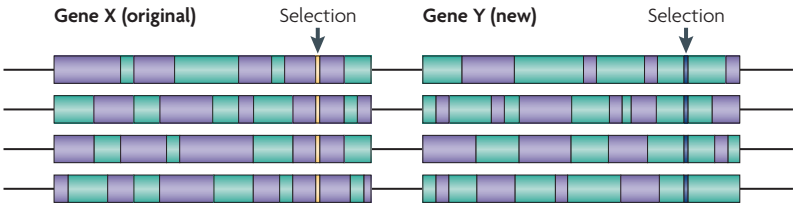
(Category III-a) Adaptive radiation model. The model proposed by Francino[52] emphasizes a period of 'pre-adaptation' in the pre-duplication phase that causes adaptive fixation of the subsequent duplication. This requires that some gene copies in the population should pre-adapt to perform new functions while still performing the original function of the gene[64] (see FIG. 2; these copies are represented by red boxes with various coloured triangles). The original article[52] explains pre-adaptation by using a gene that encodes a receptor for an environmental chemical. When a new chemical appears, the receptor may already have or will obtain the ability to bind the new chemical (although the affinity may be low) as well as the original chemical. Duplication of such a pre-adaptive allele (for example, the red gene with a green triangle in FIG. 2) allows the gene carrying this allele to evolve the full function (green box) by additional mutations. In the long term, the function of the duplicated copy could be improved by positive selection. Francino emphasizes that this process could occur repeatedly, resulting in a substantial increase in copy number. Except for the fixation phase, this model is similar to the specialization model. In addition, this model considers the benefits of the wider mutational target that is provided by several gene copies, the number of which would be higher for genes that are more prone to adaptive fixation of duplications. A formal model describing these conditions has not yet been articulated.

This model predicts asymmetrical evolution; ω_K for the new copy during the preservation phase can be increased by positive selection, and $\omega_K = \omega_{K0}$ for the original copy. The overall contribution of the original and duplicated copies to protein functionality and organismal fitness should be greater than that of the unduplicated orthologue in a closely related species. Gene conversion inhibits this process (BOX 3).

(Category III-b) Permanent heterozygote. Several authors have modelled scenarios in which heterozygote advantage would lead to the fixation and maintenance of gene duplications[65,66], as well as putting forward verbal arguments[19,67]. This model considers a population in which balancing selection maintains two alleles, A and B, of a gene (red and green boxes in FIG. 2). It is assumed that a heterozygote for a single locus is biologically equivalent to two loci that are each homozygous for a different allele. Under these conditions, the selective benefit of the two distinct loci over one heterozygous locus is the elimination of the reduced fitness of homozygotes when only a single locus is present (the segregation load). Therefore, the occurrence of a duplication in a gene for which the highest fitness is achieved in the heterozygous state is immediately beneficial[65,68] and the duplication is expected to achieve fixation faster than a neutral allele. The fixation of an extra copy of a gene under balancing selection implies that a genome with A–B (the 'permanent heterozygote') is fixed in the population. Preservation and fixation occur simultaneously because little change in the selective constraint for A and B is expected. One crucial difference between this model and the specialization model is that in the permanent heterozygote the fitness of the heterozygote is higher than either of the homozygote states in the pre-duplication phase, but in the specialization model the fitness of each homozygote is higher than that of the heterozygote.

This model predicts that there will be a high level of polymorphism between the haplotypes that harbour the two polymorphic alleles in the pre-duplication phase. Once the gene duplication is fixed, the polymorphism level of the two copies will reflect the amounts of variation in each allele in the pre-duplication phase and will gradually increase to θ in both copies at equilibrium. Gene conversion should be selected against in the preservation phase (BOX 3). The functionality of the two copies can only be replicated by a combination of two alleles of the unduplicated orthologue in a related species. In that case, the divergence between the duplicated copies would pre-date the speciation event.

Box 4 | **Effect of gene conversion on the pattern of polymorphism**



Divergence is significantly reduced by gene conversion, and its expectation is given by a function of the rates of mutation, gene conversion and recombination between the two copies of a gene. When the two copies are not tightly linked, so that there is a reasonable amount of recombination, the expected divergence is approximately given by μ/c (REFS 90,91), when the two copies reach the mutation–drift–conversion equilibrium. c is the per-site gene conversion rate per generation, that is, the rate at which a particular nucleotide site is involved in a gene conversion event.

Gene conversion increases the level of polymorphism[91]. This is easy to understand intuitively; duplicated copies can share polymorphisms that arose in one copy. The figure shows a pattern of polymorphism in a pair of duplicated genes X and Y when gene conversion is frequent. The regions that originated in genes X and Y are shown in green and purple, respectively. Each gene has a mosaic structure of the two colours owing to frequent exchanges of DNA sequences between X and Y through gene conversion. This creates several sites in which the same pair of allelic nucleotides is segregating in both X and Y (that is, shared polymorphism); this is the major reason for the increased polymorphism level in each gene. Theoretically, in a simple two-copy model the expected level of polymorphism in each copy (π) can be increased to a value that is twice as high as that of θ (REF. 91) — that is, the expectation in a single-copy gene.

One potential effect of gene conversion is that a fixed fate-determining mutation could be lost. In other words, the fixation of a fate-determining mutation does not necessarily result in a stable preservation phase when gene conversion is active. This effect of gene conversion is explained with an example of neofunctionalization. Let A and B be the original and neofunctionalizing alleles, respectively. Suppose haplotype A–B is advantageous over A–A and fixed in the population. It will then be subject to homogenization by gene conversion, which would create the deleterious haplotypes A–A and B–B. Therefore, selection has to work to eliminate the deleterious haplotypes to keep the neofunctionalizing allele. Population genetic theory[88,92] indicates that strong selection against gene conversion is required for the stable maintenance of the fate-determining mutation (and consequently of the two duplicated copies). When such strong selection protects the neofunctionalizing allele, the polymorphism pattern is likely to show a strong signature of selection, as illustrated in the figure. Gene conversion is quickly eliminated from the population if the DNA tract that is affected includes the site that specifies the neofunctionalizing allele (blue bars for gene Y in the figure). In this situation, fixation of additional mutations occurs independently in the two genes, resulting in a local peak of divergence. This effect is restricted to a short region around the target site of selection[92].

(Category III-c) Multi-allelic diversifying selection. In cases in which selection favours genetic variability, gene duplications are beneficial because they provide a larger target for mutation and selection. For a single-copy gene under multi-allelic diversifying selection, such as the major histocompatibility genes, overdominant selection operates. As a result, the maximum possible number of heterozygous individuals in the population is reached[69]. Consequently, the gene that is under selection accumulates several alleles with different functions (illustrated by boxes of various colours in FIG. 2). For a duplicate of this kind of gene, the same logic holds as that for the permanent heterozygote model, and positive selection favours the fixation of a new copy[70]. The effect of selection can be increased further compared with the permanent heterozygote model because one duplicated gene can encode up to four different alleles for a diploid. The evolutionary dynamics of this model are similar to those of the permanent heterozygote model. However, unlike the permanent heterozygote model, in the preservation phase the two copies keep accumulating new alleles. This model predicts that in the pre-duplication phase a large number of different alleles are maintained in a polymorphic state by diversifying selection or multi-allelic balancing selection. The frequency of each allele is usually low and so is the level of polymorphism in each allele. However, the overall level of polymorphism is high because of the large amount of divergence between the different alleles, resulting in very high π_{N0}, π_{S0} and ω_π, typically $\omega_\pi >> 1$. After fixation, the new copy will accumulate mutations, especially at non-synonymous sites. Gene conversion between the two copies will accelerate the recovery of polymorphism levels and creates new allele combinations (the last of the 'effects that make preservation more likely' in BOX 3), which is advantageous to the host genome. Long-term evolution involves a high rate of turnover of alleles, implying that ω_K is high, typically $\omega_K >> 1$, in both copies. New functionality is constantly gained in lineages of both duplicated and unduplicated species.

Category IV

This category is unique in that the fixation of a duplicated copy occurs as a by-product of other events, such as whole-genome (chromosome) duplication and large segmental duplications. Therefore, the model considers the fixation of the duplicate as a precondition and there is no fixation phase.

(Category IV-a) Dosage balance. For pairs or groups of genes that have optimal dosages that are dependent on each other, the genes may be duplicated or lost only synchronously. After a whole-genome duplication event, genes involved in protein complexes are preferentially kept. This model has been derived to explain the preferential retention of some duplicated genes after whole-genome duplications[71]. The original verbal model considered a functional protein that is composed of several subunits. Consider a dosage-sensitive single-copy gene, A, which has an optimum dosage dependent on the dosage of another gene, B. Gene A cannot be duplicated in isolation because of negative selection against dosage imbalance. However, if both genes have been duplicated and fixed in a single large-scale event, such as a whole-genome duplication, they will be maintained in the population because a deletion of only one of them will also cause a dosage imbalance and may be deleterious. There may be other functional scenarios in addition to the case of protein subunits in which the fitness of a duplication of a single gene is deleterious but duplication of several is not[72,73]. Therefore, if this model applies widely, only genes that are prone to dosage imbalance should be maintained after a large-scale duplication event, and such genes will not be duplicated unless their interaction partners are also duplicated.

Box 5 | **Behaviour of the level of polymorphism**

The processes that take place from the birth of a duplicated gene copy to its stable maintenance involve multiple rounds of drastic changes in the level of polymorphism in each copy of the gene. The general pattern of changes is briefly summarized here, and can be applied to most models except for those in category III. The first change occurs at the fixation of a duplicated copy. Whether fixation is achieved by genetic drift or selection, the level of polymorphism in the duplicated copy at the end of the fixation phase is expected to be reduced because the entire population shares the same recent origin of the duplicated copy. The same logic holds for the original copy if the two copies are tightly linked. When there is recombination between the two copies, the reduction in the level of polymorphism is relaxed in the original copy. The degree of the reduction of the level of polymorphism in the new copy positively correlates with the magnitude of selection for the duplicated copy. Therefore, a more drastic reduction is expected in the models in category II. After fixation, the two copies accumulate mutations and the level of polymorphism will recover to normal (that is, the expectation of π_s is θ). For gene conversion, this expectation will be higher than θ and the recovery process will be accelerated (BOX 4).

In the fate-determination phase, negative selection can be relaxed (that is, $\omega_{\pi 0} < \omega_{\pi} < 1$) in both of the two copies in all models, except for the neofunctionalization model, in which selection is completely relaxed in the duplicated copy and the selective pressure may not change in the original copy (that is, $\omega_{\pi} = \omega_{\pi 0}$ in the original copy and $\omega_{\pi} = 1$ in the duplicated copy).

The fixation of the fate-determining mutation will cause another round of reduction in the polymorphism level, which is specific to the copy of the gene in which the mutation arose (in the neofunctionalization model the fate-determining mutation occurs in the new copy). If fixation occurs through positive selection, the level of polymorphism will be substantially reduced in a large region around the mutation. Over time, the polymorphism level will recover to normal levels.

This model predicts that the system starts with a preservation phase in which negative selection is in operation; however, the intensity of negative selection may be relaxed (that is, $\omega_{\pi 0} < \omega_{\pi} < 1$ and $\omega_{K0} < \omega_{K} < 1$). Selection should also work for functional uniformity so that gene conversion is preferred (BOX 3). The functionality and fitness of unduplicated genes that function in one subunit should be the same as when all these genes are duplicated. The function of a gene duplication preserved by dosage balance should be dependent on the dosage of at least one other gene.

Conclusions and perspectives

The relative importance of the different models for the evolution and maintenance of gene duplications is hotly debated and poorly understood. As an essential part of addressing this issue, we hope that our systematic classification will guide researchers in designing rigorous studies that aim to distinguish between the mechanisms of evolution and preservation of gene duplications.

However, there are several reasons why obtaining all of the necessary data to assign specific duplications to a specific model may be an impossible task. First, some of the data are difficult to obtain or to analyse. For example, to distinguish between the DDC and specialization models, a functional characterization of two gene copies and an ancestral or orthologous single-copy gene may be needed, which is an expensive, complex and time-consuming task even for one gene duplication. Likewise, it is difficult to statistically verify deviations of the observed ω from the expected value, such as in asymmetry of evolution between two close gene copies[58]. Second, to distinguish all models it is necessary to obtain information from both long-term and short-term evolution of duplicated genes, which implies that specific gene duplications that are either too similar or too distant cannot be classified with confidence. Finally, in some cases more than one model may be applicable to specific genes. Therefore, our description may be more applicable to genome-wide studies of many duplicated genes that have the benefit of applying different tests to duplications in different stages of divergence.

A defining trend of gene-duplication evolution that emerges from our classification is that the differences between the models are almost entirely confined to the first two phases of gene-duplication evolution: the fixation and fate-determination phases. The duplicated gene is unstable in either of these phases, such that the probability that the new gene copy will reach fixation or avoid eventual loss is low. For models that assume fixation by drift, the fate-determination phase is key. Functionally and evolutionarily important changes also may occur in the short term before the preservation phase is reached, so the exploration of the short-term causes and consequences of gene duplications can be informative about their long-term evolution. We therefore propose two simple questions that are feasible to approach using the tools that are currently available and that capture a large fraction of the current debate on gene duplications: first, does the fixation of a duplicate involve selection? And second, what is the major factor in the initial preservation of duplicated genes?

To approach these questions, modern studies should shift away from the almost exclusive focus on the evolutionary divergence between gene copies. Measuring sequence divergence is a simple task, but many models predict identical dynamics of sequence divergence of gene copies (TABLE 1). Alternatively, the fate-determination phase may end before a substantial number of substitutions have accumulated and therefore measures of sequence divergence are likely to be uninformative. Sampling polymorphism data in the original and new copies can be more informative than measuring sequence divergence, and has the potential to answer the first question about the presence of selection in the fixation stage (for example, see REF. 74). Comparative genomic approaches allow us to distinguish between the original and new copies[17]. Studies of gene copy-number variation have already provided interesting data on the selection[75–77] and expression[78] of polymorphic gene duplications. However, most copy-number variation studies so far have determined the number of copies but not the sequence of each one, limiting the capacity of these studies to contribute to the debate on the evolution of gene duplications. Such studies should also measure the polymorphism pattern around the duplication and ω_{π} in polymorphisms that segregate in the original and new copies, which will be informative about whether selection is involved in the fixation of gene copies. In addition, a comparison of the expression levels of individuals with and without a segregating duplication is a test of gene dosage as a factor that contributes to the presence of the gene duplication. Such data can be complemented by data from single-copy orthologues of closely related species.

Determining the factors that affect the likelihood of preservation of duplicated genes seems more complex. In this respect, information on sequence divergence and polymorphism do not seem to be informative in isolation. Such data should be combined with genome-wide observations of the functional and evolutionary properties of duplicated genes or functional experiments. In particular, a systematic examination of functional differences between single-copy genes and their duplicated counterparts in two closely related species may provide a partial answer to the relevance of different models of the evolution of duplications on a genomic scale.

The largest hurdle to moving forward in our understanding of gene duplications is the lack of data on the population dynamics of copy-number variation and on the SNPs in and around such copies of genes. The acquisition and analysis of such data from targeted sequencing of polymorphic gene duplications will go a long way towards resolving the long-standing debates on the evolutionary mechanisms that determine the fates of duplicated genes.

1. Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I. & Koonin E. V. Selection in the evolution of gene duplications. *Genome Biol.* **3**, research0008 (2002).
A combination of a literature review and genome analysis that suggests that most recently duplicated genes are fixed by positive selection for increased dosages of genes that control responses to stress and other environmental factors.
2. Conant, G. C. & Wolfe, K. H. Turning a hobby into a job: how duplicated genes find new functions. *Nature Rev. Genet.* **9**, 938–950 (2008).
3. Van de Peer, Y. Computational approaches to unveiling ancient genome duplications. *Nature Rev. Genet.* **5**, 752–763 (2004).
4. Ponting, C. P. The functional repertoires of metazoan genomes. *Nature Rev. Genet.* **9**, 689–698 (2008).
5. Wagner, G. P., Pavlicev, M. & Cheverud, J. M. The road to modularity. *Nature Rev. Genet.* **8**, 921–931 (2007).
6. Singleton, A. B. *et al.* α-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841–841 (2003).
7. Perry, G. H. *et al.* 2007. Diet and the evolution of human amylase gene copy number variation. *Nature Genet.* **39**, 1256–1260 (2007).
8. Bailey, J. A. & Eichler, E. E. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nature Rev. Genet.* **7**, 552–564 (2006).
9. Lynch, M. *The Origins of Genome Architecture* (Sinauer Associates, Sunderland, USA, 2007).
10. Innan, H. Population genetic models of duplicated genes. *Genetica* **137**, 19–37 (2009).
11. Hahn, M. W. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* **100**, 605–617 (2009).
12. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
13. Semple, C. & Wolf, K. H. Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. *J. Mol. Evol.* **48**, 555–564 (1999).
14. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape chromosomes. *Nature* **423**, 873–876 (2003).
15. Gao, L. Z. & Innan, H. Very low gene duplication rate in the yeast genome. *Science* **306**, 1367–1370 (2004).
16. Ezawa, K., OOta, S. & Saitou, N. Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Mol. Biol. Evol.* **23**, 927–940 (2006).
17. Osada, N. & Innan, H., Duplication and gene conversion in the *Drosophila melanogaster* genome. *PLoS Genet.* **4**, e1000305 (2008).
18. Patterson, A. H. *et al.* The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**, 551–556 (2009).
19. Ohno, S. *Evolution by Gene Duplication* (Springer, New York, 1970).
A classic and influential book that outlines one of the most widely cited models of gene duplication. It includes several of the author's intuitive ideas based on his cytological studies.
20. Nei, M. Gene duplication and nucleotide substitution in evolution. *Nature* **221**, 40–42 (1969).
21. Kimura, M. & King, J. L. Fixation of a deleterious allele at one of two 'duplicate' loci by mutation pressure and random drift. *Proc. Natl Acad. Sci. USA* **76**, 2858–2861 (1979).
22. Li, W. H. Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. *Genetics* **95**, 237–258 (1980).
23. Stoltzfus, A. On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**, 169–181 (1999).
24. Walsh, J. B. How often do duplicated genes evolve new functions? *Genetics* **139**, 421–428 (1995).
A detailed theoretical analysis of the conditions necessary for the evolution of a new function under the assumptions of the neofunctionalization model.
25. Watterson, G. A. On the time for gene silencing at duplicate loci. *Genetics* **105**, 745–766 (1983).
26. Walsh, B. Population-genetic models of the fates of duplicate genes. *Genetica* **118**, 279–294 (2003).
27. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
A detailed model of preservation of gene duplications through the accumulation of mutations that are damaging to gene function. It proposes that the accumulation and maintenance of gene copies can be an entirely neutral process.
28. Lynch, M. & Force, A. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473 (2000).
29. Lynch, M., O'Hely, M., Walsh, B. & Force, A. The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804 (2001).
30. Rastogi, S. & Liberles, D. A. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol. Biol.* **5**, 28 (2005).
31. Dermitzakis, E. T. & Clark, A. G. Differential selection after duplication in mammalian developmental genes. *Mol. Biol. Evol.* **18**, 557–562 (2001).
32. Li, W. H., Yang, J. & Gu, X. Expression divergence between duplicate genes. *Trends Genet.* **21**, 602–607 (2005).
33. Ganko, E. W., Meyers, B. C. & Vision, T. J. Divergence in expression between duplicated genes in *Arabidopsis*. *Mol. Biol. Evol.* **24**, 2298–2309 (2007).
34. Byun-McKay, S. A. & Geeta, R. Protein subcellular relocalization: a new perspective on the origin of novel genes. *Trends Ecol. Evol.* **22**, 338–344 (2007).
35. Marques, A. C., Vinckenbosch, N., Brawand, D. & Kaessmann, H. Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. *Genome Biol.* **9**, R54 (2008).
36. Hughes, A. L. The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci. Lond. B* **256**, 119–124 (1994).
The first paper to propose positive selection as an important force in the long-term evolution of gene duplications.
37. Piatigorsky, J. *et al.* Gene sharing by Δ-crystallin and argininosuccinate lyase. *Proc. Natl. Acad. Sci. USA* **85**, 3479–3483 (1988).
38. Oakley, T. H., Ostman, B. & Wilson, A. C. Repression and loss of gene expression outpaces activation and gain in recently duplicated fly genes. *Proc. Natl Acad. Sci. USA* **103**, 11637–11641 (2006).
39. James, L. C. & Tawfik, D. S. Conformational diversity and protein evolution — a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **28**, 361–368 (2003).
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. Anderson, R. P. & Roth, J. R. Tandem genetic duplications in phage and bacteria. *Annu. Rev. Microbiol.* **31**, 473–505 (1977).
42. Koch, A. L. Evolution of antibiotic resistance gene function. *Microbiol. Rev.* **45**, 355–378 (1981).
43. Velkov, V. V. Gene amplification in prokaryotic and eukaryotic systems. *Genetica* **18**, 529–543 (1982).
44. Stark, G. R. & Wahl, G. M. Gene amplification. *Annu. Rev. Biochem.* **53**, 447–491 (1984).
45. Romero, D. & Palacios, R. Gene amplification and genomic plasticity in prokaryotes. *Annu. Rev. Genet.* **31**, 91–111 (1997).
46. Brown, C. J., Todd, K. M. & Rosenzweig, R. F. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* **15**, 931–942 (1998).
47. Guillemaud, T. *et al.* Quantitative variation and selection of esterase gene amplification in *Culex pipiens*. *Heredity* **83**, 87–99 (1999).
48. Kondrashov, F. A. & Kondrashov, A. S. Role of selection in fixation of gene duplications. *J. Theor. Biol.* **239**, 141–151 (2006).
49. Kondrashov, F. A. & Koonin, E. V. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplication. *Trends Genet.* **20**, 287–291 (2004).
This paper applies the metabolic dosage theory that is commonly applied to genetic dominance to study the relationship between gene dosage and gene duplications.
50. Veitia, R. A. Gene dosage balance: deletions, duplications and dominance. *Trends Genet.* **21**, 33–35 (2005).
51. Sugino, R. P. & Innan, H. Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. *Trends Genet.* **22**, 642–644 (2006).
52. Francino, M. P. An adaptive radiation model for the origin of new gene functions. *Nature Genet.* **37**, 573–577 (2005).
53. Haldane, J. B. S. The part played by recurrent mutation in evolution. *Am. Nat.* **67**, 5–19 (1933).
54. Fisher, R. A. The sheltering of lethals. *Am. Nat.* **69**, 446–455 (1935).
55. Wright, S. Physiological and evolutionary theories of dominance. *Am. Nat.* **68**, 24–53 (1934).
56. Fisher, R. A. The possible modification of the response of the wild type to recurrent mutations. *Am. Nat.* **62**, 115–126 (1928).
57. Clark, A. G. Invasion and maintenance of a gene duplication. *Proc. Natl Acad. Sci. USA.* **91**, 2950–2954 (1994).
58. Lynch, M. & Katju, V. The altered evolutionary trajectories of gene duplicates. *Trends Genet.* **20**, 544–549 (2004).
59. Katju, V. & Lynch, M. On the formation of novel genes by duplication in the *Caenorhabditis elegans* genome. *Mol. Biol. Evol.* **23**, 1056–1067 (2006).
60. Lercher, M. J., Blumenthal, T. & Hurst, L. D. Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. *Genome Res.* **13**, 238–243 (2003).
61. Long, M. & Langley, C. H. Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* **260**, 91–95 (1993).
62. Long, M., Betran, E., Thornton, K. & Wang, W. The origin of new genes: glimpses from the young and old. *Nature Rev. Genet.* **4**, 865–875 (2003).
63. Vinckenbosch, N., Dupanloup, I. & Kaessmann, H. Evolutionary fate of retroposed gene copies in the human genome. *Proc. Natl Acad. Sci. USA* **103**, 3220–3225 (2006).
64. Aharoni, A. *et al.* The 'evolvability' of promiscuous protein functions. *Nature Genet.* **37**, 73–76 (2005).

65. Spofford, J. B. Heterosis and the evolution of duplications. *Am. Nat.* **103**, 407–432 (1969).
This paper suggested that a duplication of a locus for which there is a heterozygote advantage can be immediately beneficial.
66. Proulx, S. R. & Phillips, P. C. Allelic divergence precedes and promotes gene duplication. *Evolution* **60**, 881–892 (2006).
67. Muller, H. J. Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. *Genetics* **3**, 422–499 (1918).
68. Otto, S. P. & Yong, P. The evolution of gene duplicates. *Adv. Genet.* **46**, 451–483 (2002).
69. Penn, D. J., Damjanovich, K. & Potts, W. K. MHC heterozygosity confers a selective advantage against multiple-strain infections *Proc. Natl Acad. Sci. USA* **99**, 11260–11264 (2002).
70. Doherty, P. C. & Zinkernagel, R. M. A biological role for the major histocompatibility antigens. *Lancet* **1**, 1406–1409 (1975).
71. Papp, B., Pál, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194–197 (2003).
72. Birchler, J. A. & Veitia, R. A. The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell.* **19**, 395–402 (2007).
73. Liang, H., Plazonic, K. R., Chen, J., Li, W.-H. & Fernández, A. Protein under-wrapping causes dosage sensitivity and decreases gene duplicability. *PLoS Genet.* **4**, e11 (2008).
74. Moore, R. C. & Purugganan, M. D. The early stages of duplicate gene evolution. *Proc. Natl Acad. Sci. USA* **100**, 15682–15687 (2003).
75. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
76. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
77. Emerson, J. J., Cardoso-Moreira, M., Borevitz, J. O. & Long, M. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* **320**, 1629–1631 (2008).
78. Stranger, B. E. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
79. Kreitman, M. Methods to detect selection in populations with applications to the human. *Annu. Rev. Genome Hum. Genet.* **1**, 539–559 (2000).
80. Neilsen, R. Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**, 197–218 (2005).
81. Hurst, L. D. Genetics and the understanding of selection. *Nature Rev. Genet.* **10**, 83–93 (2009).
82. Petes, T. D. & Hill, C. W. Recombination between repeated genes in microorganisms. *Annu. Rev. Genet.* **22**, 147–168 (1988).
83. Chen, J. M., Cooper, D. N., Chuzhanova, N., Ferec, C. & Patrinos, G. P. Gene conversion: mechanisms, evolution and human disease. *Nature Rev. Genet.* **8**, 762–775 (2007).
84. Teshima, K. M. & Innan, H. The effect of gene conversion on the divergence between duplicated genes. *Genetics* **166**, 1553–1560 (2004).
85. Mano, S. & Innan, H. The evolutionary rate of duplicated genes under concerted evolution. *Genetics* **180**, 493–505 (2008).
86. Ohta, T. Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. *Proc. Natl Acad. Sci. USA* **88**, 6716–6720 (1991).
87. Takuno, S., Nishio, T., Satta. Y. & Innan, H. Preservation of a pseudogene by gene conversion and diversifying selection. *Genetics* **180**, 517–531 (2008).
88. Innan, H. A two-locus gene conversion model with selection and its application to the human *RHCE* and *RHD* genes. *Proc. Natl Acad. Sci. USA* **100**, 8793–8798 (2003).
This paper considers the influence of gene conversion on the adaptive divergence of two recently duplicated genes, and establishes the conditions under which selection can overcome the homogenizing pressure of conversion.
89. Bischof, J. M. *et al.* Genome-wide identification of pseudogenes capable of disease-causing gene conversion. *Hum. Mutat.* **27**, 545–552 (2006).
90. Ohta, T. Allelic and nonallelic homology of a supergene family. *Proc. Natl Acad. Sci.* USA **79**, 3251–3254 (1982).
91. Innan, H. The coalescent and infinite-site model of a small multigene family. *Genetics* **163**, 803–810 (2003).
92. Teshima, K. M. & Innan, H. Neofunctionalization of duplicated genes under the pressure of gene conversion. *Genetics* **178**, 1385–1398 (2008).

Acknowledgements
We thank M. Lynch for insightful comments on the manuscript.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Hideki Innan's homepage:
http://www.sendou.soken.ac.jp/esb/innan/InnanLab
Fyodor Kondrashov's homepage:
http://kondrashovlab.crg.cat

ALL LINKS ARE ACTIVE IN THE ONLINE PDF