Please note the following before checking your manuscript

Dear Professor Cañestro,

Please note:

- Queries may be nested in the addresses section or in the reference list, please look out for these.
- References that could be in any way misread (for example, those that follow numbers or symbols) are usually set in brackets, on the line.
- Please check reference numbering carefully, as they may have been renumbered (for example, to avoid duplications).
- When marking up your corrections, please use the Commenting tools only (do not edit the text directly).
- If your article has Supplementary Information but you have not received the file(s) along with this proof, please let us know as soon as possible.
- We do not usually edit Supplementary Information in house, so should you require changes to your SI, please send a revised file and indicate what changes have been made.

Evolution by gene loss

Ricard Albalat and Cristian Cañestro

Abstract | The recent increase in genomic data is revealing an unexpected perspective of gene loss as a pervasive source of genetic variation that can cause adaptive phenotypic diversity. This novel perspective of gene loss is raising new fundamental questions. How relevant has gene loss been in the divergence of phyla? How do genes change from being essential to dispensable and finally to being lost? Is gene loss mostly neutral, or can it be an effective way of adaptation? These questions are addressed, and insights are discussed from genomic studies of gene loss in populations and their relevance in evolutionary biology and biomedicine.

Loss is nothing else but change, and change is Nature's delight — Marcus Aurelius, AD 121–180

Great attention has in the past been paid to the mechanisms of evolution by gene duplication (that is, neofunctionalization and subfunctionalization)[1,2]. By contrast, gene loss has often been associated with the loss of redundant gene duplicates without apparent functional consequences, and therefore this process has mostly been neglected as an evolutionary force. However, genomic data, which is accumulating as a result of recent technological and methodological advances, such as next-generation sequencing, is revealing a new perspective of gene loss as a pervasive source of genetic change that has great potential to cause adaptive phenotypic diversity.

Two main molecular mechanisms can lead to the loss of a gene from a given genome. First, the loss of a gene can be the consequence of an abrupt mutational event, such as an unequal crossing over during meiosis or the mobilization of a transposable or viral element that leads to the sudden physical removal of the gene from an organism's genome. Second, the loss of a gene can be the consequence of a slow process of accumulation of mutations during the pseudogenization that follows an initial loss-of-function mutation. This initial mutation can be caused by nonsense mutations that generate truncated proteins, insertions or deletions that cause a frameshift, missense mutations that affect crucial amino acid positions, changes involving splice sites that lead to aberrant transcripts or mutations in regulatory regions that abolish gene expression. In this Review, the term 'gene loss' is used in a broad sense, not only referring to the absence of a gene that is identified when different species are compared, but also to any allelic variant carrying a loss-of-function (that is, non-functionalization) mutation that is found within a population.

Here, we address some of the fundamental questions in evolutionary biology that have emerged from this novel perspective of evolution by gene loss. Examples from all life kingdoms are covered, from bacteria to fungi and from plants to animals, including key examples of gene loss in humans. We review how gene loss has affected the evolution of different phyla and address key questions, including how genes can become dispensable, how many of our current genes are actually dispensable, how patterns of gene loss are biased, and whether the effects of gene loss are mostly neutral or whether gene loss can actually be an effective way of adaptation. Finally, promising future perspectives on the study of gene loss are discussed. These include the development of computational pipelines to identify the complete catalogue of gene losses that have occurred during the evolution of a given species, the effect that anticipated findings have on the fields of evolutionary biology and biomedicine, and the means by which comparative population genomics approaches and the measure of 'population gene dispensability' can help to discover new genes that are relevant for human health.

Phylogenetic pervasiveness of gene loss

The field of comparative genomics, and especially our perspective on animal evolution, changed after the sequencing of the genome of various cnidarian species. These studies revealed that the ancestral eumetazoan genome was much more complex than expected and that gene loss was pervasive in many animal phyla[3–6] (FIG. 1). This new perspective superseded the traditional notion that had influenced the analyses of the first known genomes (*Caenorhabditis elegans, Drosophila melanogaster, Arabidopsis thaliana* and human), the *scala naturae*, in which attempts were made to correlate the apparent increase of biological complexity in the evolutionary ladder that leads up to humans with an increase in the number of genes[7]. The ancestral eumetazoan genome had a gene repertoire made up of at least 7,766 gene families that were putatively homologous with a gene in the sea anemone and at least one other gene

Pseudogenization
An evolutionary phenomenon whereby a gene loses its function, accumulates mutations and becomes a pseudogene.

Eumetazoan
Clade that classically includes all animals (metazoan) except sponges and Placozoa, although recent analyses of ctenophores have challenged the monophyly of this group.

Homologous
Genes that share sequence similarity because they have evolved from a common ancestral gene.

Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain.
ralbalat@ub.edu; canestro@ub.edu

doi:10.1038/nrg.2016.36
Published online DD Mmm 2016

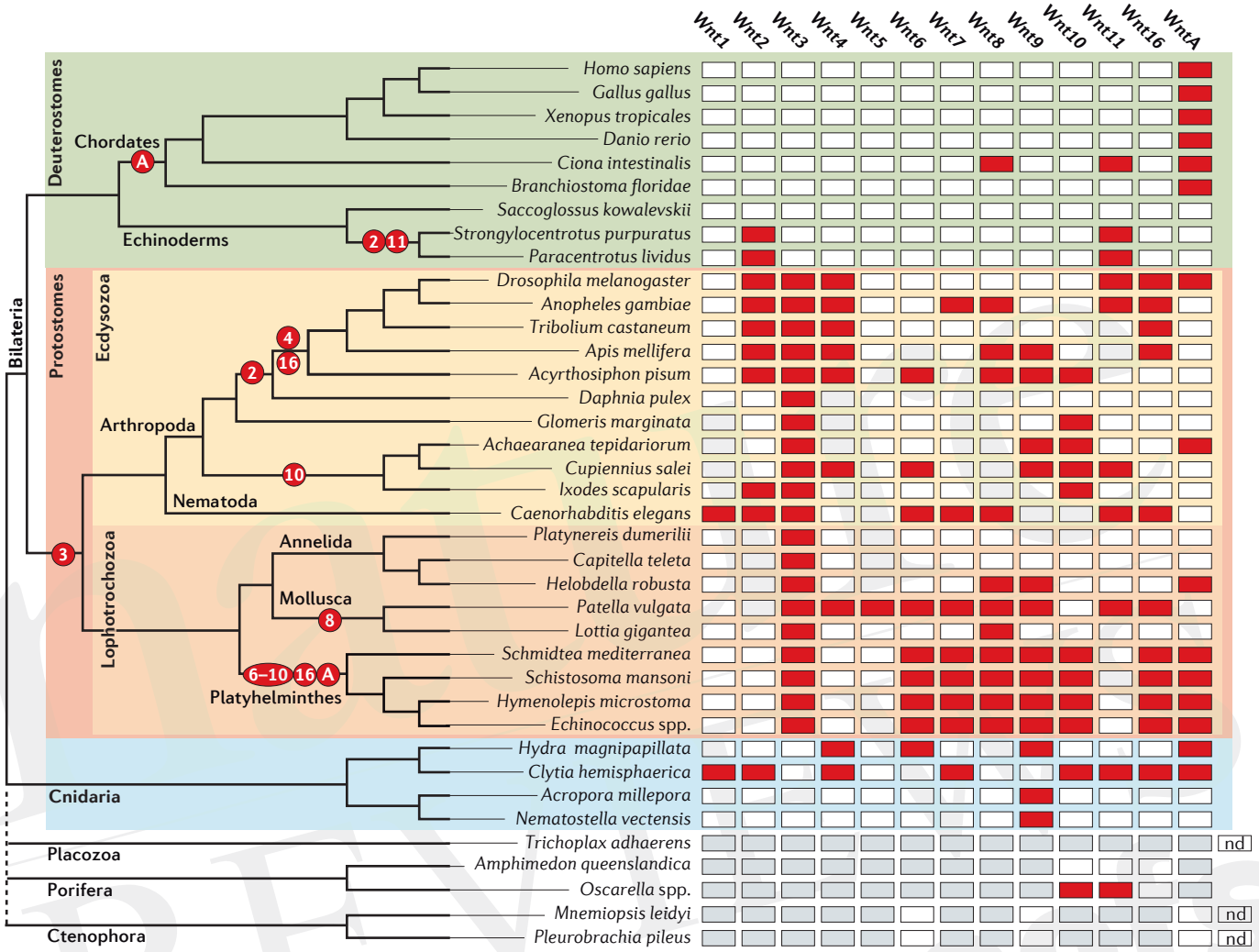


Figure 1 | **The wingless (Wnt) family: a paradigmatic example of the pervasiveness of gene loss during metazoan evolution.** In the past decade, the accumulation of fully sequenced genome data from various species has revealed great heterogeneity in the dynamics of gene loss within different animal groups. In ecdysoazoans, for instance, not all insects show the same rate of gene loss, and European honeybees (*Apis mellifera*) seem to have retained more genes than other insects (for example, species of fly and mosquito in the Diptera order)[206]. The finding, for instance, of an active DNA CpG methylation toolkit (that is, *Dnmt1*, *Dnm3a*, *Dnmt3b* and *Mdb*) in honeybees was particularly remarkable, as it has been lost in most other insects[207,208]. To date, the red flour beetle (*Tribolium casteneum*) has preserved the largest number of patchy orthologues that are also present in humans but that were lost in all other sequenced insects[209]. The genomes of crustaceans and myriapods showed less gene loss, and these groups conserved more universal bilaterian genes than insects[151,210]. In lophotrocozoans, gene loss propensity is also heterogeneous among species. Mollusc gastropods, such as *Lottia gigantea* or annelids, such as *Capitella teleta* or *Helobdella robusta*, seem to have rates of gene retention similar to those in deuterostomes[8], whereas other lophotrocozoans, such as the flatworm *Schmidtea mediterranea*, have lost approximately 40% of the ancestral gene families[8,171]. Extensive gene loss (red boxes) has affected all Wnt gene subfamilies (1 to 11; 16 and A) throughout all metazoan taxa. Some gene losses seem to be ancestral (red circles) and thereby probably relevant for the evolution of entire groups (for example, ancestral loss of *Wnt3* in the stem protostome). Other gene losses seem to occur recurrently in diverse lineages and show a patchy distribution (for example, *Wnt11* loss in some chordates, echinoderms, arthropods, nematodes, molluscs and sponges). Controversial animal phylogenies (dashed tree branches)[211,212] or uncertain gene orthologies (nd) hinder the ability to determine whether the absence of Wnt families in most basal metazoans (grey boxes) is due to gene losses or to gene gains. References for the list of Wnt genes in each species are supplied in Supplementary information S3 (box).

Bilaterian
An animal clade that includes protostomes and deuterostomes. Members of this clade are characterized by a stage during their life cycle in which they have right–left symmetry (unlike the radial symmetry present in most cnidarians and sponges).

Deuterostomes
A superphylum that includes animals in which the first opening, the blastopore, becomes the anus. This superphylum includes Ambulacraria (hemichordates and echinoderms) and Chordates (cephalochordates, urochordates and vertebrates).

in any bilaterian analysed at the time (that is, *D. melanogaster*, *C. elegans*, pufferfish, frogs and humans)[5]. Whereas deuterostomes seemed to have collectively lost only 33 gene families — a total of 0.42% — from the ancestral gene repertoire, protostomes had collectively lost 1,292 gene families (17%), which suggests a different propensity for gene loss between the two groups. The increased availability of genome sequences from a diverse range of species within protostomes, however, reveals that ecdysozoans have suffered extensive gene losses, whereas Lophotrochozoa have a gene retention rate that is similar to that of vertebrates[8] (BOX 1; FIG. 1).

Protostomes
A superphylum that includes animals in which the first opening, the blastopore, becomes the mouth. This superphylum includes two groups: Ecdysozoa (for example, arthropods and nematodes) and Lophotocozoa (for example, molluscs, annelids and platyhelminthes).

Propensity for gene loss
Proclivity of a gene to be lost during evolution of a clade, as estimated from the fraction of lineages in which a given gene has been lost and corrected by the time during which the gene was lost or preserved.

'Patchy' orthologues
Orthologues belonging to gene families that have suffered extensive gene loss during the evolution of a given clade, such that their presence is unevenly distributed and restricted to a few species in the clade.

Parahoxozoa
A hypothetical subkingdom that includes all animals apart from poriferans and ctenophores based on the absence of homeobox (*Hox*)–*ParaHox* genes from the first sequenced species of the later groups.

In deuterostomes, the sequencing of the genomes of Ambulacraria (sea urchins), Cephalochordata (amphioxus), Urochordata (ascidians and larvaceans) and several vertebrate species revealed a low propensity for gene loss in this group, with the exception of urochordate species[9–13]. Urochordates have suffered many more gene losses than other chordates, a circumstance that correlates with a morphological simplification of their body plan and the evolution of a determinative developmental mode[11–13]. Gene loss reached extreme levels in the urochordate *Oikopleura dioica*[14,15] (BOX 2).

The pervasiveness of gene loss during the evolution of the Wnt gene family exemplifies the varying trends for gene loss of different animal taxa (FIG. 1). Genome sequencing of diverse species in a given taxon has often uncovered novel 'patchy' orthologues that reveal previously hidden origins of ancestral gene families. These data allow us to differentiate ancestral gene losses from genes that have recurrently been lost in different lineages (FIG. 1). For example, comparison of the genome sequences of different sponge species shows that genome complexity predates the last common ancestor of all metazoans[16,17] and may have arisen even earlier[18,19], supporting the notion that the absence of many genes in some lineages is due to extensive gene loss. A paradigmatic example was the finding of homeobox (*Hox*) and *ParaHox* genes in two calcisponges, which led researchers to reconsider the Parahoxozoa hypothesis[20] and to validate the presence of the predicted *Hox* and *ParaHox* 'ghost' loci based on the loss of these genes in ctenophore *Mnemiopsis leidyi* and the sponge *Amphimedon queenslandica*[17,21].

Alongside animals, fully sequenced genomes from a wide range of organisms, including prokaryotes[22–26], protista[27], fungi[28,29] and plants[30], have shown that gene loss is pervasive in all life kingdoms. Recent exhaustive analyses comparing hundreds of genomes of bacteria and archaea revealed that loss of gene families has also been pervasive, dominating their evolution with frequencies, in some cases, up to three times higher than the rate of gene gain[31–33]. In plants and fungi, extensive polyploidy events have been followed by gene loss[34], and therefore these groups represent particularly useful models for understanding the dynamics and extent of gene loss after whole-genome duplication (WGD) events (for additional discussion, see REFS 30,35,36).

Taken together, the pervasiveness of gene loss in most life forms suggests that reductive evolution would not only have driven the evolution of parasitic and symbiotic species, as is classically asserted[37], but would also be a prevalent evolutionary force that affects all organisms[25].

Gene loss and dispensability

The pervasiveness of gene loss throughout evolution leads to the fundamental question of how many genes can readily be lost in a given genome. Intuitively, the answer to this question depends on how many genes are actually essential for a given organism, and therefore cannot be lost, and how many genes are to some degree dispensable, and therefore susceptible to being lost because their loss has no impact or only a slightly negative impact on fitness, at least under certain circumstances (FIG. 2).

The knockout paradox. Gene dispensability is a measure that is inversely related to the overall importance of a gene (that is, gene essentiality), and this measure has been approximated by the fitness of the corresponding gene knockout strain under laboratory conditions[38,39]. Understanding which genes are dispensable or essential by linking genotypes with phenotypes is one of the most challenging tasks in the field of genetics and biomedicine in the twenty-first-century post-genomic era. This understanding is important both theoretically, such as when defining the minimal genome for a free living organism[40], and practically, such as when identifying all essential genes that are responsible for human diseases[41].

Historically, Susumu Ohno not only pioneered the idea that gene duplication was an important evolutionary force, but in 1985 he also pondered the concept of gene dispensability and suggested that "the notion that all the still functioning genes in the genome ought to be indispensable for the well-being of the host should be abandoned" (REF. 42). The emergence of large-scale gene targeting approaches has facilitated the calculation of the number of genes that are globally dispensable in a given genome in certain conditions. Thus, systematic large-scale approaches that involve single-gene deletions in *Escherichia coli* and other bacterial species showed that only a few hundreds of genes are essential, suggesting that nearly 90% of bacterial genes are dispensable when cells are grown either in rich or minimal mediums[39,43,44]. The high degree of global gene dispensability found in bacteria is consistent with findings from systematic gene deletion screens in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. These screens revealed that approximately 80% of protein-coding

Box 1 | Gene loss shaped the evolution of vertebrates and humans

Analyses of gene loss in vertebrates are challenging because of the large amount of gene loss that followed two rounds of whole-genome duplication (WGD) during early vertebrate evolution (also known as 2R-WGD or vertebrate genome duplication (VGD) events 1 and 2) and a third round that occurred in teleost fishes (3R or TGD; reviewed in REF. 191 and REF. 101, respectively). A comprehensive phylogenetic analysis of 9,461 gene families in humans, mice, rats, chickens, frogs, zebrafish and pufferfish revealed that of the 7,350 gene families represented in at least one fish, one land vertebrate and the ascidian or fly outgroup, 5,396 families did not conserve any duplicated gene in some species. This observation implies massive post-2R gene loss events, ranging from 10,792 to 16,188 losses, depending on their chronological distribution between VGD1 and VGD2 (REF. 82). Analyses of zebrafish and pufferfish suggest that 30% of the genes that were duplicated in 2R and 20% of the genes that were duplicated in 3R were lost in each lineage. Strikingly, although all vertebrates seem to keep losing 2R-ancestral ohnologues, not all species have the same propensity to lose genes: frogs, chickens and fish, for instance, seem to have lost approximately four times as many ancestral genes as humans and rodents[82]. In mammals, from the 9,990 gene families that were inferred to be present in their most recent common ancestor, 1,421 families (14%) have zero genes in at least one extant genome[83]. Among primates, humans seem to have undergone the fewest number of gene losses, and whereas chimpanzees have undergone 729 gene losses, humans have only undergone 86 losses over the same period[83,141]. Thirty years after King and Wilson[192] recognized the apparent paradox in which "the genetic distance between humans and the chimpanzee (<2%) is probably too small to account for their substantial organismal differences", the 6% difference in their orthology complement was proposed as a fertile source of genetic change that could explain many of the differences between the two species[83,193].

Ohnologues
A term coined in honour of Susumo Ohno that refers to paralogues that originated from genome duplication (in contrast to paralogues that originated from small-scale duplications).

Polyploidy
Acquisition of additional genetic content due to whole-genome duplication.

genes are dispensable under laboratory conditions[45,46]. Following the same trend, large-scale RNA interference approaches in *C. elegans*[47,48] and *D. melanogaster*[49] suggested that 65% to 85% of genes, respectively, are dispensable in these organisms, and similar figures were obtained in mice by the Sanger Institute Mouse Genetics Project[50]. Recent attempts to test for gene essentiality in humans using gene trap and large-scale CRISPR–Cas9 screens suggest that approximately 90% of tested genes are dispensable for cell proliferation and survival, at least in human cancer cell lines[51–53]. These surprisingly high values of seemingly dispensable genes in different organisms and their tolerance to inactivation have been referred as the 'gene knockout paradox' (REF. 38). Two main factors have been provided that may account for this observed gene dispensability[54]: mutational robustness and environment-dependent conditional dispensability.

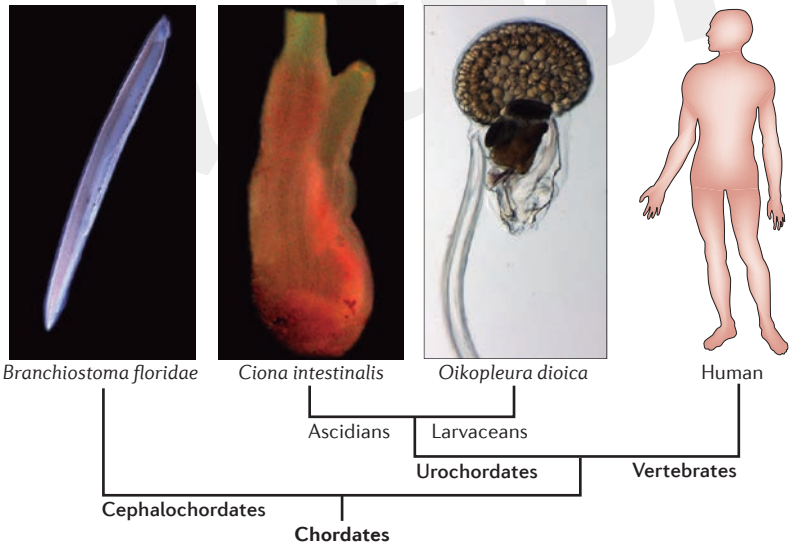
Mutational robustness. It is plausible to assume that a number of the many apparently dispensable (that is, non-essential) genes accounting for the gene knockout paradox may result from the mutational robustness of biological systems (FIG. 2; recently reviewed in REF. 55). In a simplified view, mutational robustness is increased either by the presence of redundant or backup genes — this is also termed degeneracy or genetic buffering — or by the presence of alternative pathways, which are also referred to as functional complementation or backup pathways. Redundant genes usually arise by duplication (that is, paralogues), although they may also occur by convergent evolution (that is, analogues)[56]. Alternative pathways are possible because genes and proteins are structured in regulatory and functional networks with a scale-free structure: when one part of the network fails, biological tasks can be re-routed through alternative pathways, thus conferring distributed robustness to the network[57].

Many insights into mutational robustness come from recent *in silico* and experimental studies in yeast. Large-scale *in silico* studies of yeast metabolic networks and flux re-routing have shown, for instance, that both paralogues and alternative pathways are relevant in increasing gene dispensability and ultimately facilitating gene loss (although the contribution of each factor remains controversial[58–63]). When multiple, rather than single, knockouts are considered, the number of dispensable genes drops from 80% to 25%, and the contribution of alternative pathways towards mutational robustness is more prominent than that of duplication when the knockout multiplicity required for lethality is high[54]. At the experimental level, powerful functional genomic tools have facilitated genome-wide analyses of genetic interactions and networks in yeast (as reviewed in REF. 64). Systematic gene deletion screens have experimentally shown that deletion of 9% of essential genes — the so-called 'evolvable' essential genes — can be overcome by evolution of alternative pathways, suggesting that essentiality is more of a quantitative than a qualitative property that is intrinsic to a particular gene[65]. Synthetic genetic array (SGA) analyses have found that negative genetic interactions — interactions among variants of different genes that cause severe effects on fitness, an extreme example being synthetic lethality — frequently occur between redundant genes. However, positive interactions — that is, genetic interactions that have less severe effects on fitness than would be expected — are observed between genes of alternative §pathways[66,67].

In multicellular organisms, despite several cases of gene losses having been associated with gene dispensability through the evolution of functionally overlapping paralogues[68–70] or alternative pathways[71], systematic studies are still needed into other aspects that affect gene dispensability. These other aspects include transcriptional regulation, rewiring of transcriptional regulatory circuits, robustness in translation and mechanisms accounting for cryptic variation[72].

Box 2 | ***Oikopleura dioica*: a chordate model to study gene loss effects**

Oikopleura dioica is a free-swimming planktonic larvacean urochordate that is emerging as an attractive model organism for studying gene loss in the field of evolutionary developmental biology (known as evo–devo). This is because this species occupies a key phylogenetic position within the closest sister group of vertebrates (see the figure), because the animals are easy to cultivate in the laboratory and because techniques for generating genetic knockdowns in this model are widely available[194–196]. The sequencing of *O. dioica* revealed that this animal has undergone an extreme process of genome compaction (its genome size is only 70 Mb) accompanied by extensive gene loss[15]. Notably, *O. dioica* has lost 16 of the 83 ancestral genes that are involved in DNA repair, including all the components of the non-homologous end-joining DNA repair system. This severe dismantling of the DNA repair toolkit is plausibly one of the reasons for the elevated rate of evolution and gene loss in this organism[15]. Other notable examples of gene loss in this species are those that affect the epigenetic machinery[197], the immune system[15], the microRNA repertoire[198], the apoptotic system[199] and the xenobiotic defence systems[200]. Among the key developmental genes, *O. dioica* has lost more than 30% of the homeobox gene groups[14], including all central *Hox* genes[201], and key genes involved in retinoic acid signalling[202]. This latter observation was especially surprising because retinoic acid is fundamental for anteroposterior axial patterning through *Hox* gene regulation in all chordates[203], but *O. dioica* maintains an unaltered *Hox1* expression domain (that is similar to the one found in ascidians) and a typical chordate body plan[204]. These data led to the formulation of the so-called 'inverse paradox' of evo–devo, which proposes that organisms might develop fundamentally similar morphologies (that is, phenotypic unity) despite having important differences in their genetic toolkits (that is, genetic diversity), which is especially obvious in developmental genetic toolkits that have undergone extensive gene loss[13].



The *Ciona intestinalis* photograph in the figure is from REF. 13, Nature Publishing Group.

Reductive evolution
Refers to the loss of genetic material that is usually observed during the evolution of parasitic or symbiotic species.

Fitness
The ability of a particular genotype (or phenotype) to survive and reproduce in a specific environment, which is usually expressed in relation to other possible genotypes.

Developmental genetic toolkits
Sets of genes that are required for development and that are widely shared among species.

Mutational robustness
Property of a biological system to maintain unaltered phenotypes in the face of mutations.

Synthetic genetic array
(SGA). Methodology designed to map genetic interactions on a genome-wide scale that combines arrays of mutant strains with robotic manipulations for high-throughput double-mutant construction.

Synthetic lethality
This occurs when a combination of mutations in two or more genes leads to death, but when no effects on the viability of the organism are apparent when the genes are mutated individually.

Cryptic variation
Genetic diversity within a population that does not normally generate phenotypic diversity but that does occur on environmental or genetic perturbation.

Flux balance analyses
(FBAs). Mathematical approaches for calculating the flow of metabolites through a metabolic network, which can be applied to reconstruct genome-scale metabolic networks and to predict the growth rate of an organism.

Gene Ontology
(GO). A system for classification of genes in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.

Environment-dependent conditional dispensability. In addition to mutational robustness, another explanation for the gene knockout paradox may be that genes appear to be dispensable if they are involved in processes that are only required under specific untested environmental conditions (FIG. 2). This explanation is conceivable especially if we consider that current genes in given species are the result of selective pressures of a vast range of diverse environments that act at the evolutionary scale over millions of years. Computational models of flux balance analyses (FBAs) were developed as a promising tool for testing gene dispensability in a large variety of conditions (as reviewed in REF. 38). FBA genome-scale metabolic network reconstructions in yeast predict that approximately 40–70% of metabolic genes that seem to be dispensable are a part of pathways that are inactive under the tested conditions and, therefore, could become essential if the gene deletions are simulated under other conditions[58]. The generation of more than 21,000 mutant *S. cerevisiae* strains that carry deletions of approximately 6,000 open reading frames has provided experimental support to this hypothesis through testing of more than 1,000 chemical or environmental stress conditions. The observation that 97% of gene deletions exhibited a measurable altered growth phenotype[73,74] provides experimental evidence that most of the seemingly dispensable genes of the gene knockout paradox are, in fact, required for optimal growth in at least one condition.

It remains unclear whether the mutational robustness and environmental influence underlying gene dispensability in the well-studied system of yeast are as relevant for more complex organisms (for example, mice)[75,76]. Development of further systematic studies in complex multicellular organisms will be needed to shed light on this issue.

Biased patterns of gene loss

The uneven distribution of gene loss throughout different branches of the tree of life (FIG. 1) and the differences of dispensability observed between different genes in diverse groups of organisms suggest that the evolutionary patterns of gene loss do not occur in a stochastic way. Instead, a clear bias related to gene function or genomic position is evident (FIG. 3).

Gene functional bias. Comparative analyses of gene losses according to Gene Ontology (GO) categories have revealed obvious biased patterns of gene loss. For instance, a comparison of *S. cerevisiae* and *S. pombe* reveals that not all functional categories are equally affected by gene loss but that a large fraction of lost genes belong to functional GO categories such as 'nuclear structure maintenance', 'pre-mRNA splicing', 'RNA modification', 'post-transcriptional gene silencing' and 'protein folding/processing' (REF. 77). In plants, genes that are involved in 'DNA repair and modification' or 'ancient biochemical processes' are more prone to be lost than 'transcription factors' and 'protein kinases' (REFS 78–81). In vertebrates, genes from GO categories such as 'protein modification', 'protein metabolism', 'catabolism' and 'peptidase activity' are more prone to be lost in fish than in land vertebrates, whereas genes involved in 'catalytic activity' show the opposite trend[82]. Genes from GO categories such as 'immune response', 'chemosensation', 'reproduction', 'transcription' or 'gamete interaction' are more prone to be lost in mammals than they are in other vertebrates[83,84]. These biased patterns of gene loss are likely to arise from differences in gene dispensability of each functional category. Gene dispensability is affected by differences in biological, reproductive and environmental constraints that are associated with the lifestyle of each group of organisms (for example, aquatic versus land lifestyle) (FIG. 3).

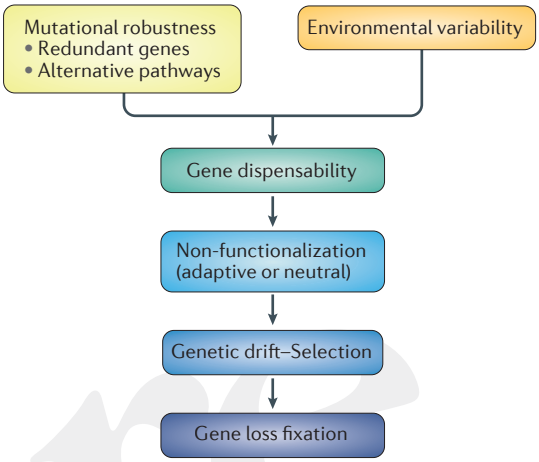


Figure 2 | **Conceptual framework for gene loss.** The loss of a gene depends on the degree of dispensability of the gene, which in turn depends on how fitness is affected by its non-functionalization. In a mutational robust system, either because of the presence of redundant genes or alternative pathways, mutations will have less impact on the fitness, therefore increasing the overall level of gene dispensability and facilitating gene loss. Gene functions are not equally essential in all environments and, therefore, environmental variability can also modify gene dispensability. Non-functionalization of a dispensable gene can either be neutral (or nearly neutral) when the gene is not needed, for instance in a new environmental condition (for example, regressive evolution), or it can can be adaptive if the loss of the function is advantageous in the new condition (for example, if it provides resistance to a disease: the less-is-more hypothesis). Finally, the balance between genetic drift, which depends on the population size, and selection will determine the probability of the fixation of gene loss.

In species that suffer relaxation of a given biological or environmental constraint, a functional bias of gene loss can often be observed that is caused by the 'co-elimination' of genes that are functionally linked in distinct pathways or complexes associated with the relaxed constraint[77,85] (FIG. 3). This co-elimination can be the result of the dismantling of a pathway within a gene network, the exception being 'hub' genes that are needed for other pathways. Examples of dismantling of pathways or complexes include: the loss of most genes of the eukaryotic translation initiation factor 3 (Eif3)–signalosome complex in *S. cerevisiae*[77]; the gastric gene repertoire in platypuses and many teleost fish (that is, *atp4a*, *atp4b*, *pga*, *pgb*, *pgc*, *pgf* and *cym*) [86,87]; the loss of teeth-specific genes encoding structural proteins that are

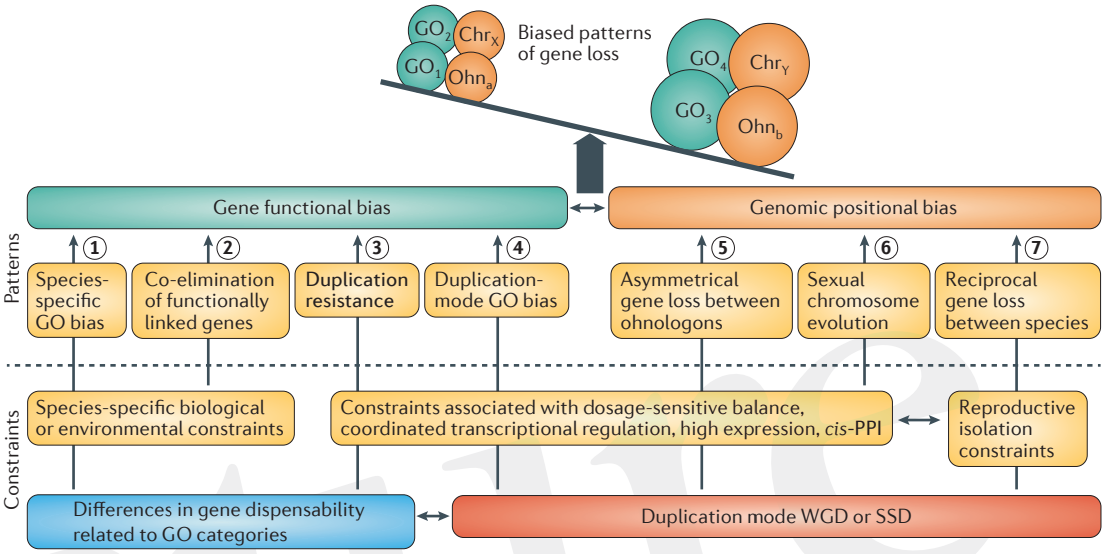


Figure 3 | **Biased patterns of gene loss.** Gene loss patterns do not seem to follow stochastic fashions, but they show clear biases related to gene function (green) or genomic position (orange). These biased patterns are mainly caused by different constrains associated with gene dispensability related to Gene Ontology (GO) categories; blue) and constraints associated with the duplication mode that precedes gene losses (red). Genes from certain GO categories are more prone to be lost in certain species than others owing to differences in biological and environmental constraints (1). Relaxation of these constraints in certain species can lead to co-elimination of genes that are functionally linked in distinct pathways or complexes (2). Duplication-resistant genes from certain GO categories that have essential cellular functions, that are highly expressed or that are sensitive to dosage balance are prone to be lost after duplication in most organisms (3). Considering that small-scale duplication (SSD), but not whole-genome duplication (WGD), alters gene stoichiometry, duplication modes bias gene loss patterns towards certain GO categories, depending on their sensitivity to dosage balance (4). After WGD, gene losses are frequently asymmetrically distributed between ohnologons (Ohn), probably owing to enrichment of genes with high levels of transcription, dose-sensitive genes, genes with coordinated transcriptional regulation and genes that code for *cis*-protein–protein interacting (PPI) products (5). An extreme case of asymmetric distribution of gene loss occurs during the evolution of sexual chromosomes, in which Y chromosomes (Chr_Y) are often depleted of most genes that were once shared with the X chromosomes (Chr_X; 6). Reciprocal distribution of gene losses between the ohnologons of species that diverged after WGD reduces the viability of hypothetical hybrids, contributing therefore to reproductive isolation (7).

crucial for enamel and dentine formation in birds (that is, *Dspp*, *Amel*, *Ambn* and *Enam*)[88]; the loss of genes underlying the urea cycle or the immunodeficiency pathway in the pea aphid[89]; and the loss of genes that are involved in DNA repair by non-homologous end-joining in the urochordate *O. dioica*[15] (BOX 2).

An interesting biased pattern associated with gene function is the loss of the so-called 'duplication-resistant genes', which, after duplication has occurred, are conserved as single-copy genes in most genomes[90] (FIG. 3). Duplication-resistant genes consistently belong to functional GO categories such as 'housekeeping roles', 'DNA repair', 'DNA recombination', 'DNA damage response' and other essential cellular functions. They are generally expressed at higher levels and in more tissues than other genes, and they also seem to be affected by the dosage-dependent balance[30].

A biased pattern in the loss (or retention) of duplicated genes also associated with GO becomes apparent when the mode of duplication is considered — that is, whole-genome duplication (WGD) versus small-scale duplication (SSD; FIG. 3; as reviewed in REF. 91). Studies from yeast, plants and vertebrates, all of which have suffered extensive WGDs, point to dosage balance constraints as a major factor affecting the loss or retention of duplicate genes[92–94]. In contrast to WSD, SSD does not maintain the stoichiometry of duplicated genes, which is consistent with the fact that genes that have functions in the following GO categories are more prone to be lost after SSDs than WGDs: 'transcriptional regulation', 'signal transduction' or 'protein–protein interacting complexes'. Notably, all these functions are more sensitive to dosage imbalance. By contrast, genes that are involved in 'DNA repair', 'RNA metabolism', 'nucleoplasm', 'apoptosis' or 'organelle functions' show the opposite trend[79,80,95–97]. The finding that gene losses between two species of yeast that diverged closely after an event of WGD are convergently biased towards certain functional categories supports the relationship between duplication mode and functional bias[97]. The overall level of gene loss therefore appears to be predictable in terms of functional bias, but unpredictable at the level of the fate of individual genes[97].

Genomic positional bias. Patterns of gene loss also seem to be biased when the genomic position of genes that have been lost is taken into consideration. This phenomenon is especially obvious for the massive losses that occur during the diploidization that follows WGD (also known as biased or asymmetric fractionation)[98,99] (FIG. 3). Comparative analyses of chromosomal regions that are duplicated by WGD (that is, ohnologons) in yeasts, plants and animals have revealed the asymmetrical distribution

Conserved synteny
Conservation of similar blocks of genes between orthologous or paralogous chromosomal regions, which can be useful in detecting gene losses after speciation or large-scale genomic duplications, respectively.

Reciprocal gene loss
Divergent resolution of gene duplicates, such that one species has lost one copy, whereas the second species has lost the other copy

Baker's rule
This rule states that self-compatible organisms are better colonizers after long-distance dispersal than self-incompatible ones.

of gene loss between ohnologons, resulting in different physical clusters of retained genes with high levels of conserved synteny[68,100–102]. The biological significance of the clusters of retained genes remains unclear, and several functional and structural explanations have been hypothesized. These include enrichment of genes with high levels of transcription, dose-sensitive genes, genes that share a coordinated transcriptional regulation by epigenetic or long-range regulatory mechanisms and genes that code for *cis*-protein–protein interacting products[99–101,103]. Interestingly, the pattern of gene loss when considering genomic positions has been shown to change over time, and although the 'choice' of which copy is discarded seems random when considering the period shortly after WGD, it becomes increasingly nonrandom as time elapses after WGD, favouring the loss of the same gene copy in independent lineages[97]. This changing pattern from initially random to non-random suggests that the process of gene loss itself imposes constraints on prospective losses.

A special case of biased pattern associated with genomic position is when gene losses occur in a reciprocal fashion between the ohnologons of two species that have diverged after WGD (FIG. 3). Patterns of reciprocal gene loss in many species of animals, plants and yeasts have been interpreted as the result of evolutionary events that have favoured reproductive isolation and speciation among different lineages[104–107], although alternative conclusions exist[108].

Finally, another pattern of loss associated with genomic position is the high frequency of gene losses during the evolution of sexual chromosomes from autosomes, which has been found in both plants[109] and mammals[110,111] (FIG. 3). The human Y chromosome, for instance, has lost nearly all of the approximately 640 genes that it once shared with the X chromosome. The exceptions to this are 36 broadly expressed genes — namely, regulators of transcription, translation and protein stability — that seem to be dosage-sensitive and, therefore, are likely under selective pressure to be retained as two copies in both sexes[111]. Empirical reconstruction of the evolution of the human male-specific Y region has revealed that each stratum transitioned from rapid exponential loss of ancestral genes to strict conservation through purifying selection[112].

Evolution by gene loss

To understand the impact of gene loss on the evolution of species, it is crucial to analyse the potential adaptability or neutrality of non-functionalization mutations that lead to gene losses (FIG. 2). Here, specific examples of gene losses are reviewed that may be either adaptive and support the 'less-is-more' hypothesis or neutral and thus occur, for instance, in the context of 'regressive evolution'. Some of the evolutionary factors are then discussed that influence the fixation of gene losses under adaptive or neutral evolution.

The less-is-more hypothesis. The less-is-more hypothesis proposes that non-functionalization represents a frequent evolutionary adaptive response, which may be of special relevance when populations are exposed to changes in the patterns of selective pressures owing to drastic shifts of environmental conditions[113,114]. Adaptive gene loss has been reported in a number of unicellular organisms. In bacteria, more than 200 examples of gene loss have been associated with adaptations to changes in environmental conditions[115]. Meta-analysis of bacterial genome-wide fitness data from transposon insertion and in-frame deletion mutations across 144 conditions shows that adaptive null mutations are extremely abundant and disproportionately affect enzymatic and regulatory pathways. Cases have even been found in which a null mutation could be adaptive in more than ten different conditions[115]. In bacteria, many instances of gene loss have also been associated with adaptive gains in pathogenicity[116]. For example, the loss of *ALL1* in *Cryptococcus neoformans*[117], *cadA* in *Shigella* spp.[118], arabinose operon genes in *Burkholderia pseudomallei* and *Burkholderia mallei*[119] or *mucA* in *Pseudomonas aeruginosa*[120] confer adaptive advantages during infection. In the pathogen *Candida glabrata*, the loss of *de novo* biosynthesis of nicotinic acid (BNA) genes has also been positively selected as a mechanism that increases pathogenicity by directing infection to the murine urinary tract. This gene family loss leads to specific expression of epithelial adhesion (EPA) genes, which encode proteins that mediate the adherence of *C. glabrata* to host cells within the renal system[121]. In yeast, different *S. cerevisiae* strains have suffered gene losses that provide them with a major fitness advantage for growth in high-sugar-level substrates, as they facilitate the co-use of various sugar sources as well as living in conditions of high acidity, high ethanol and high temperature[122,123].

Adaptive gene loss has also been reported in many multicellular organisms. In plants, adaptive gene loss has been associated with changes in pollinators: for example, loss of *AN2* leads to white flowers in *Petunia axillaris*, which has been proposed as an adaptation to pollination by nocturnal hawk moths[124]; and loss of the flavonoid 3′-hydroxylase gene leads to red flowers in *Ipomoea quamoclit*, which has been proposed as an adaptation to bird pollination[125]. In *A. thaliana*, the loss of scarecrow (*SCR*) and/or SnRK2-type protein kinase (*SRK*) genes underlies the evolution from obligate outcrossing based on self-incompatibility to a self-fertilization system — this is considered to be one of the most prevalent evolutionary transitions of flowering plants that colonize oceanic islands (known as Baker's rule)[126]. The non-functionalization of the *Desaturase 2* (*Desat2*) gene in cosmopolitan but not in tropical *D. melanogaster* strains has been associated with resistance to cold environments[127]. Also in flies, accelerated gene loss of some groups of chemoreceptors, such as gustatory or odorant receptors, has occurred under positive selection associated with the colonization of new ecological niches, loss of behaviours and the evolution of new diets in different *Drosophila* species[128–130].

Some of the most widely known cases of adaptive gene loss that support the less-is-more hypothesis have been described in humans. Loss-of-function mutations of C–C chemokine receptor type 5 (*CCR5*) and atypical chemokine receptor 1 (*ACKR1*; also known as *DUFFY*), for instance, provide resistance to AIDS[131] and vivax malaria[132], respectively. For *CCR5*, a 32-base-pair (bp)

deletion yields a non-functional allele that has been under intense positive selection[133–135]. However, AIDS is thought be a modern human disease, which suggests that the positive selection for the non-functional allele of *CCR5* has not been caused by the action of HIV itself but is caused by other viruses[136,137]. For malaria, a loss-of-function mutation abolishes the expression of Duffy antigen receptor for chemokines (*DARC*) in red blood cells and confers resistance to infection by two malarial species: *Plasmodium knowlesi* and *Plasmodium vivax* (as reviewed in REF. 138). In homozygous mutants, no receptor is present on red blood cells, thus preventing their infection. Another example of positive selection for a blood group that is attributable to its protection from malaria is the O blood group, which is a consequence of the non-functionalization of the *ABO* glycosyltransferase gene. There is substantial evidence that this mutation provides protection against *P. falciparum* malaria, and *O* alleles have higher frequencies in areas that have historic malaria. Nonetheless, it remains unclear whether the high frequencies of the *O* alleles are due to selection that is related directly to malaria exposure or whether they are caused by some other reason (reviewed in REF. 138).

Another gene loss that has been suggested to be essential for the evolutionary origins of humans is the non-functionalization of myosin, heavy chain 16 (*MYH16*). This gene probably became dispensable after the change in diet that reduced the reliance on powerful masticatory jaw muscles during the evolution of hominids. Its loss could also have been favoured by adaptive selection that led to an increase in cranial capacity and brain size that occurred during the origin of humans[139]. Other examples of gene loss that differentiate humans from other primates are the non-functionalization of CMP-*N*-acetylneuraminic acid hydroxylase (*CMAH*), which provides resistance to some pathogens[140], and the fixation of a null allele of *CASPASE12* in human populations shortly before the migration out of Africa, which confers protection from severe sepsis[141].

These examples of adaptive gene loss that affect all types of organisms show how gene loss can be a force of molecular evolution that promotes evolutionary change and generates biodiversity. However, they also lead to a new question: what fraction of adaptive mutations are loss-of-function variants? Whole-genome and whole-population sequencing experiments in yeast have shown that many of the adaptive mutations that arise under growth-restricting environments, such as limited amounts of sugar, are actually loss-of-function mutations in signalling pathways, demonstrating that gene loss can be the major adaptive strategy[142]. In addition, selection experiments of bacterial populations under different conditions have shown that adaptive loss-of-function mutations of enzymatic and regulatory functions have an important role in the adaptation of bacterial populations to new environments[31,115,143,144]. Considering that mutations that cause a loss of function are much more probable than mutations that lead to a gain of function[115], the contribution of gene loss to adaptive evolution, especially as a rapid response to environmental challenges, might be higher than previously anticipated.

Regressive evolution. The term regressive evolution refers to the loss of useless characteristics over time, and many examples of gene loss with neutral effects on fitness have been reported to occur during regressive evolution. Changing from a poor to a rich vitamin C diet is a classic example of regressive evolution accompanied by recurrent gene loss that is associated with the changes of environmental metabolic supplies[145]. The L-gulonolactone oxidase (*GLO*) gene, which encodes the enzyme that is responsible for the last step of vitamin C biosynthesis, has been independently lost in different vertebrate lineages, including some teleost fish, several passeriforme birds and some mammals, such as bats, guinea pigs and anthropoid primates. This occurred after these animals adopted a diet rich in vitamin C[146]. Gene loss associated with changes in metabolic requirements seems to be especially frequent in parasitic species, as the progressive intimate association with the hosts can lead to metabolic redundancy, resulting in the loss of many parasite genes (see Supplementary information S1 (table) for examples).

Another paradigmatic case of apparently neutral loss associated with regressive evolution is the loss of genes in species that have adapted to life in the perpetual darkness of caves, which is a phenomenon that was commented on by Darwin: "As it is difficult to imagine that eyes, though useless, could in any way be injurious to animals living in darkness, I attribute their loss solely to disuse" (REF. 147). Vision and pigmentation became dispensable features in several populations of the cavefish *Astyanax mexicanus* after the colonization of dark environments[148]. The evolution of blindness and loss of pigmentation was caused by the accumulation of non-functional mutations in ocular and cutaneous albinism 2 (*Oca2*), which does not seem to have had any other deleterious effects[148]. Supporting the relationship between the cave lifestyle and the loss of *Oca2*, independent loss-of-function mutations of *Oca2* have been observed in at least two different *Astyanax* cavefish populations[148]. Other species, such as subterranean diving beetles[149], cave isopod crustaceans[150], myriapods[151], bats[152] and rat moles[153,154], have also recurrently undergone photoreceptor or pigment regression after the colonization of dark environments, providing more examples of gene loss due to regressive evolution.

Such examples of gene loss that are associated with regressive evolution could be a priori considered to be evolutionarily neutral, as each loss can be associated with the loss of a biological feature that seems dispensable under the new environmental condition. The neutrality of certain cases of gene loss, however, is under debate. It could be argued that the loss of genes is under positive selection owing to the advantages that it provides in energy savings and spatial efficiency (for example, to avoid the replication, transcription and translation of useless genes)[155]. Comparative genomics analyses in bacteria suggest that reduction of genome size is mainly driven by genetic drift[156], and no evidence has been found to support a link between smaller cell size and environment, nor has any selective advantage of smaller genomes due to a reduced metabolic burden of replicating DNA been established[157,158]. In yeast, however, although the loss of mating genes in strains that live in environmental conditions in

Genetic drift
Stochastic changes in allele frequencies in a population due to random sampling effects through successive generations, which is therefore highly affected by the population size.

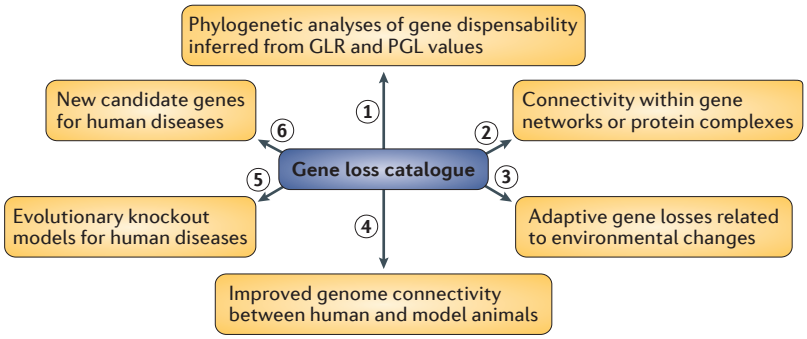


Figure 4 | **Gene loss catalogues in evolutionary biology and translational medicine.** Comparisons of the catalogues of gene losses between different species could be useful in many fields of biology. In the field of evolutionary biology, a comprehensive gene loss catalogue that covers a wide range of diverse groups of organisms can provide specific values of gene loss rate (GLR)[213] and propensity for gene loss (PGL)[214], which could help inference of the dispensability of any given gene during the evolution of any group of organisms (1). In addition, the identification of patterns of gene co-elimination make gene loss catalogues useful for predicting the functional connectivity of each gene within gene network modules or protein complexes[215] (2). Manifested recurrent and convergent patterns of gene loss in different species that evolve under similar changes of ecological conditions (for example, light exposure, temperature, salinity, food, toxic substances or pathogens) could lead to the discovery of cases of adaptive gene loss associated with those environmental changes (3). In the field of translational medicine, a gene loss database could help in improving functional connectivity between model organisms and human genomes by distinguishing orthologous from paralogous genes in the setting of reciprocal gene loss[175] (4) and could also help in the discovery of animals that are 'evolutionary knockouts' for genes related to human pathologies (5). These animals could become new disease models, as has already occurred for: the Antarctic icefish, which is a model for anaemia, osteoporosis and lipid storage disorders; the swordtail fish, which is a model for melanoma; the East African cichlid fish and Darwin's finches, which are models for craniofacial disease; some reptilian species, which are models for heterotopic ossification; and cavefish, which are models for retinal degeneration, cataracts, albinism and diabetes (as reviewed in REFS 216,217). Finally, comparison of gene loss catalogues between organisms that have suffered convergent processes of regressive evolution in structures or biological processes related to a human disease will help in discerning new candidate genes for diseases (6), as has already been demonstrated for the *BBS5* gene in Bardet–Biedl ciliopathic syndrome (as reviewed in REF. 13).

Gene loss rate
(GLR). The maximum likelihood estimate of the measure of gene loses that maximizes the probability of the phyletic pattern of presence and absence of a given gene considering estimated branch lengths of all possible ancestral phylogenetic trees for the species under study.

Antagonistic pleiotropy
This occurs when a gene controls several traits, in which at least one of these traits is beneficial to the organism's fitness and at least one is detrimental to the organism's fitness.

Hitchhiking effect
This occurs when a neutral mutation is in linkage disequilibrium with a second locus that is undergoing a selective sweep.

which sexual reproduction is abolished could be interpreted as neutral, the loss of these genes reduces the levels of expression of at least 23 genes and provides a growth-rate increase per generation of approximately 2%. This could be seen as a fitness advantage that leads to the gene loss coming under positive selection[159].

Linking gene losses to energy efficiency and spatial savings is more difficult to explain in organisms with large genomes, such as animals or plants. In these organisms, establishing whether the loss of a gene is neutral, directly selected or indirectly selected either by antagonistic pleiotropy or by a hitchhiking effect (that is, when a beneficial trait is negatively linked to the presence of a gene) is not easy and sometimes controversial. For example, an interesting debate is ongoing surrounding whether the loss of the eye in cavefish is neutral, which is supported by the independent accumulation of diverse non-functional mutations in multiple eye-related loci[160], or whether it is the indirect consequence of a hitchhiking effect promoted by positive selection on a nearby locus, which is related to the elaboration of a vibration attraction behaviour[161–163]. Although some tests from classical population genetics (as reviewed in REF. 164) could be modified to study the adaptive nature of loss-of-function alleles, it is still a challenging task to assess whether an ancestral loss in a certain lineage was adaptive or neutral. In many cases, it can merely be pointed out that a correlation exists between the loss of a gene and the evolution of new biological features (see Supplementary information S2 (table) for examples).

Gene loss fixation: adaptive or neutral? An important question for understanding the evolutionary dynamics of gene loss is whether most gene loss fixations are neutral or adaptive (FIG. 2). This question belongs to the wider and still open neutralism–selectionism debate on whether genetic variation found in populations is mostly neutral or adaptive and whether neutral variants are relevant to the emergence of evolutionary innovations[165]. Following the general principles of population genetics, the probability of fixation of neutral gene losses depends only on the population size, and thereby on genetic drift, whereas in the case of adaptive gene losses, the fixation probability also depends on the selective coefficient (FIG. 2). Even if the loss of a gene is slightly deleterious, it has a probability, albeit a small one, of becoming fixed in the population by genetic drift[166]. For free-living unicellular organisms with large population sizes, gene loss fixation seems to have been mostly adaptive and selection-driven[25,167], whereas analyses of unicellular genomes of parasitic and symbiotic species, which frequently go through bottlenecks, suggest that most gene loss fixations are neutral and driven by genetic drift[25,168]. The small effective population size of multicellular organisms, in contrast to bacteria, has led to the proposition that genetic drift is the major driving force for gene loss fixation[169]. Concordantly, comparative analysis of five vertebrate and five insect species reveals a high correlation between the rates of gene loss and the rates of molecular evolution of each species, suggesting that, overall, the fixation of most gene losses is driven by neutral evolution, despite the profound effect that the many gene losses might have on the evolution of these organisms[170].

Future directions

A future challenge in the area of gene loss research will be to use comparative genomics to map all instances of gene loss in the tree of life and to identify genes that have been lost during the evolution of any given species or taxon in relation to its last common ancestor with another given species or taxon. Comprehensive gene loss catalogues that cover a wide range of diverse groups of organisms would provide valuable information for many fields of biology, including evolutionary biology and translational medicine (FIG. 4).

To build a complete database of gene loss, however, it is necessary to develop computational strategies to identify and to map gene losses in evolutionary trees reliably. This would allow to superimpose informative layers about the evolution of biological features in each lineage to facilitate links between gene losses and key evolutionary events. Many large-scale analyses to identify gene losses use the annotated ontology of predicted proteins available in genome databases as a starting point — for example, Ensembl, notably, already includes a gene gain and loss tree. Genes can then be classified into gene families using

Box 3 | **Population gene dispensability**

To explore actual gene dispensability in natural environments resulting from evolutionary processes, the concept of population gene dispensability (PGD) is proposed here. PGD is defined as the sum of the frequencies (f_{null}) of all null alleles (n_{null}) for a given gene:

$$PGD = \sum_{i=1}^{n} f_{null_i}$$

PGD values range from 0, when no null alleles are detected, to 1 in the extreme cases of fixed non-processed pseudogenes. The number (n_{null}) and frequency (f_{null}) of null alleles depend on population mutation rate (θ), genetic drift (d) and selection (s). Assuming that θ and d equally affect all individuals within a population, differences between PGD values among genes should reflect differences in the selective pressures that affect the loss of a particular gene. In general, therefore, within a large population, it is expected that the n_{null} and f_{null} values for dispensable genes — that is, genes for which the losses are neutral, nearly neutral or adaptive — are greater than those for genes for which the losses are detrimental and in which null variants are eliminated by selection. Within the fraction of dispensable genes, in general, it is expected that n_{null} and f_{null} values for adaptive gene losses, in which null variants are enriched by selection, are greater than those values for neutral gene losses. Accordingly, the relative PGD (rPGD) value of a given gene, which is defined as the PGD of a given gene weighted by the average PGD in the population, helps to infer the adaptivity (when the value is in the high range), neutrality (in the middle range) or detrimentality (in the low range) of its non-functionalization. It is also expected that differences of rPGD for a given gene between populations exposed to different environments — for example, diet, climate, toxic substances or pathogens — will help to identify genes that are functionally associated with the different environments. The allelic frequencies of null variants of *DUFFY* (also known as *ACKR1*)and C–C chemokine receptor type 5 (*CCR5*) — which are higher in human populations exposed to malaria and HIV, respectively, than in non-exposed populations[134,136,205] — are proof of concept for these expectations.

Long-branch attraction
The phenomenon of inferring an incorrect phylogenetic tree owing to the presence of sequences that evolve rapidly and generate long branches that are mispositioned — usually attracted to the base — and thus distort the tree.

BLAST-based algorithms, phylogenetic inferences and maximum likelihood classifiers or probabilistic models based on the stochastic 'birth and death' model[82,83,171–173]. These analyses, however, are limited by the quality of genome annotations, as well as by the robustness of phylogenetic inferences, which in some cases can be affected by artefacts of long-branch attraction[174]. They are also limited by the difficulty of detecting cases of reciprocal gene losses between groups of organisms, which means that paralogues can be mistaken for orthologues, and gene losses are therefore neglected. The use of pipelines that include data on conserved synteny can be useful to overcome this limitation, as they provide solid evidence for gene losses, especially those that take place after events of genome duplication[175–177]. Pipelines that include syntenic mapping in addition to classic BLAST searches and phylogenetic inferences have shown their use in the identification of many gene losses that had previously been overlooked[178–181]. Hiller *et al.*[182] have gone one step beyond previous computational gene loss analysis methods by creating a computational strategy called 'forward genomics'. This approach is able to associate specific genomic regions in whole-genome sequences in which a gene loss has occurred with a given lost phenotype: for example, the loss of *GLO* or *ABCD4* matches loss of the ability to synthesize vitamin C or low biliary phospholipid levels, respectively[182]. Mapping gene losses but also losses of conserved non-coding DNA elements[183] seems to be a promising next step in determining how the loss of regulatory elements of genes has led to the loss of some sub-functions that have influenced the evolution of species.

Another promising line of investigation in the field of population genomics is the possibility of 'catching' ongoing processes of gene loss in natural populations and estimating actual gene dispensability in wild conditions. Analyses of 180 human genomes, after applying stringent filters to identify non-functionalized variants, have led to the inference that the genome of an average healthy person has 100 non-functionalized alleles, 20 of which are homozygous but have no apparent phenotypic consequences[184–188]. These findings show the presence of a substantial number of non-functional variants in natural populations. In this context, it is useful to consider the concept of population gene dispensability (PGD) as a new measure that could help in the estimation of the degree to which a gene is involved in the ongoing processes of gene loss in natural populations (BOX 3). The calculation of the PGD is based on the frequencies of non-functional variants of a given gene within a population. Distributions of relative PGD values could provide helpful information for inferring whether gene losses are under negative, neutral or positive natural selection. Differences in relative PGD values between populations under different environmental conditions could also make it possible to identify losses of candidate genes for which non-functionalization is adaptive and could thereby have potential interest in biomedicine. Lim *et al.*[189] have recently published a pioneering study that provides a proof of concept of the potential power of future PGD analyses for biomedical studies. After a comparative analysis of low-frequency null alleles in the exomes of 3,000 Finnish and 3,000 non-Finnish individuals, the researchers found enrichment for null alleles in the Finnish population. Interestingly, two variants enriched in the Finnish population are null alleles of *LPA*, which encodes lipoprotein A, that are associated with low levels of plasma lipoprotein A. In homozygous or compound heterozygous individuals, the presence of these null alleles provides protection against coronary artery disease and acute myocardial infarction. This work opens the door to a 'genotype-first' approach in which comparative population genomics of gene losses can be used to find target genes of therapeutic interest[190].

1. Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
2. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
3. Kortschak, R. D., Samuel, G., Saint, R. & Miller, D. J. EST analysis of the cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Curr. Biol.* **13**, 2190–2195 (2003).
4. Technau, U. *et al.* Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet.* **21**, 633–639 (2005).
5. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
The sequencing of the anemone genome changed the view of animal evolution, revealing that ancient metazoan genomes were complex and that gene losses have been pervasive throughout animal lineages.
6. Kusserow, A. *et al.* Unexpected complexity of the *Wnt* gene family in a sea anemone. *Nature* **433**, 156–160 (2005).
7. Szathmary, E., Jordan, F. & Pal, C. Molecular biology and evolution. Can genes explain biological complexity? *Science* **292**, 1315–1316 (2001).

8. Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526–531 (2013).
9. Sea Urchin Genome Sequencing Consortium *et al.* The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**, 941–952 (2006).
10. Putnam, N. H. *et al.* The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071 (2008).
11. Dehal, P. *et al.* The draft genome of *Ciona intestinalis:* insights into chordate and vertebrate origins. *Science* **298**, 2157–2167 (2002).
12. Cañestro, C., Bassham, S. & Postlethwait, J. H. Seeing chordate evolution through the *Ciona* genome sequence. *Genome Biol.* **4**, 208–211 (2003).
13. Cañestro, C., Yokoi, H. & Postlethwait, J. H. Evolutionary developmental biology and genomics. *Nat. Rev. Genet.* **8**, 932–942 (2007).
14. Edvardsen, R. B. *et al.* Remodelling of the homeobox gene complement in the tunicate *Oikopleura dioica*. *Curr. Biol.* **15**, R12–R13 (2005).
15. Denoeud, F. *et al.* Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* **330**, 1381–1385 (2010).
16. Srivastava, M. *et al.* The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010).
17. Fortunato, S. A. *et al.* Calcisponges have a *ParaHox* gene and dynamic expression of dispersed NK homeobox genes. *Nature* **514**, 620–623 (2014).
18. King, N. *et al.* The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788 (2008).
19. Suga, H. *et al.* The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* **4**, 2325 (2013).
20. Ryan, J. F. *et al.* The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *Evodevo* **1**, 9 (2010).
21. Mendivil Ramos, O., Barker, D. & Ferrier, D. E. Ghost loci imply Hox and ParaHox existence in the last common ancestor of animals. *Curr. Biol.* **22**, 1951–1956 (2012).
22. Lynch, M. Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* **60**, 327–349 (2006).
23. Moya, A., Pereto, J., Gil, R. & Latorre, A. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* **9**, 218–229 (2008).
24. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2012).
25. Wolf, Y. I. & Koonin, E. V. Genome reduction as the dominant mode of evolution. *BioEssays* **35**, 829–837 (2013).
26. Librado, P., Vieira, F. G., Sanchez-Gracia, A., Kolokotronis, S. O. & Rozas, J. Mycobacterial phylogenomics: an enhanced method for gene turnover analysis reveals uneven levels of gene gain and loss among species and gene families. *Genome Biol. Evol.* **6**, 1454–1465 (2014).
27. Loftus, B. *et al.* The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865–868 (2005).
28. Spanu, P. D. *et al.* Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**, 1543–1546 (2010).
29. Peyretaillade, E. *et al.* Extreme reduction and compaction of microsporidian genomes. *Res. Microbiol.* **162**, 598–606 (2011).
30. De Smet, R. *et al.* Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc. Natl Acad. Sci. USA* **110**, 2898–2903 (2013).
31. Koskiniemi, S., Sun, S., Berg, O. G. & Andersson, D. I. Selection-driven gene loss in bacteria. *PLoS Genet.* **8**, e1002787 (2012).
32. Puigbo, P., Lobkovsky, A. E., Kristensen, D. M., Wolf, Y. I. & Koonin, E. V. Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. *BMC Biol.* **12**, 66 (2014).
33. Nelson-Sathi, S. *et al.* Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* **517**, 77–80 (2015).
34. Wendel, J. F. Genome evolution in polyploids. *Plant Mol. Biol.* **42**, 225–249 (2000).
35. Jiao, Y. & Paterson, A. H. Polyploidy-associated genome modifications during land plant evolution. *Phil. Trans. R. Soc. B* **369**, 20130355 (2014).
36. Scannell, D. R., Butler, G. & Wolfe, K. H. Yeast genome evolution — the origin of the species. *Yeast* **24**, 929–942 (2007).
37. Andersson, S. G. & Kurland, C. G. Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268 (1998).
38. Papp, B., Notebaart, R. A. & Pal, C. Systems-biology approaches for predicting genomic evolution. *Nat. Rev. Genet.* **12**, 591–602 (2011).
This Review discusses gene dispensability in the context of systems biology and formulates the gene knockout paradox.
39. Korona, R. Gene dispensability. *Curr. Opin. Biotechnol.* **22**, 547–551 (2011).
40. Acevedo-Rocha, C. G., Fang, G., Schmidt, M., Ussery, D. W. & Danchin, A. From essential to persistent genes: a functional approach to constructing synthetic life. *Trends Genet.* **29**, 273–279 (2013).
41. Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A. & Kim, D. Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* **16**, 85–97 (2015).
42. Ohno, S. Dispensable genes. *Trends Genet.* **1**, 160–164 (1985).
43. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
44. de Berardinis, V. *et al.* A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **4**, 174 (2008).
45. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
46. Kim, D. U. *et al.* Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nat. Biotechnol.* **28**, 617–623 (2010).
47. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
48. Sonnichsen, B. *et al.* Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434**, 462–469 (2005).
49. Dietzl, G. *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
50. White, J. K. *et al.* Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* **154**, 452–464 (2013).
51. Blomen, V. A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–1096 (2015).
52. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
53. Osorio, J. Functional genomics: the genetic essence of human cells. *Nat. Rev. Genet.* **16**, 683 (2015).
54. Deutscher, D., Meilijson, I., Kupiec, M. & Ruppin, E. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat. Genet.* **38**, 993–998 (2006).
55. Felix, M. A. & Barkoulas, M. Pervasive robustness in biological systems. *Nat. Rev. Genet.* **16**, 483–496 (2015).
56. Piskur, J., Sandrini, M. P., Knecht, W. & Munch-Petersen, B. Animal deoxyribonucleoside kinases: 'forward' and 'retrograde' evolution of their substrate specificity. *FEBS Lett.* **560**, 3–6 (2004).
57. Wagner, A. Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays* **27**, 176–188 (2005).
The author discusses the contribution of gene redundancy and distributed robustness to gene dispensability.
58. Papp, B., Pal, C. & Hurst, L. D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**, 661–664 (2004).
59. Blank, L. M., Kuepfer, L. & Sauer, U. Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol.* **6**, R49 (2005).
60. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
61. Ihmels, J., Collins, S. R., Schuldiner, M., Krogan, N. J. & Weissman, J. S. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Mol. Syst. Biol.* **3**, 86 (2007).
62. Li, J., Yuan, Z. & Zhang, Z. The cellular robustness by genetic redundancy in budding yeast. *PLoS Genet.* **6**, e1001187 (2010).
63. DeLuna, A. *et al.* Exposing the fitness contribution of duplicated genes. *Nat. Genet.* **40**, 676–681 (2008).
64. Boone, C., Bussey, H. & Andrews, B. J. Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* **8**, 437–449 (2007).
65. Liu, G. *et al.* Gene essentiality is a quantitative property linked to cellular evolvability. *Cell* **163**, 1388–1399 (2015).
66. Baryshnikova, A. *et al.* Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* **7**, 1017–1024 (2010).
67. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
This work constructs a genome-scale genetic interaction map covering 75% of all genes of *S. cerevisiae* and provides experimental evidence that describes how gene redundancy and alternative pathways account for genetic robustness.
68. Cañestro, C., Catchen, J. M., Rodríguez-Marí, A., Yokoi, H. & Postlethwait, J. H. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet.* **5**, e1000496 (2009).
69. McClintock, J. M., Carlson, R., Mann, D. M. & Prince, V. E. Consequences of *Hox* gene duplication in the vertebrates: an investigation of the zebrafish Hox paralogue group 1 genes. *Development* **128**, 2471–2484 (2001).
70. Gitelman, I. Evolution of the vertebrate twist family and synfunctionalization: a mechanism for differential gene loss through merging of expression domains. *Mol. Biol. Evol.* **24**, 1912–1925 (2007).
71. Danchin, E. G., Gouret, P. & Pontarotti, P. Eleven ancestral gene families lost in mammals and vertebrates while otherwise universally conserved in animals. *BMC Evol. Biol.* **6**, 5 (2006).
72. Payne, J. L. & Wagner, A. Mechanisms of mutational robustness in transcriptional regulation. *Front. Genet.* **6**, 322 (2015).
73. Hillenmeyer, M. E. *et al.* The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365 (2008).
This work provides experimental evidence that most of the seemingly dispensable genes of the gene knockout paradox are in fact required for optimal growth in at least one condition.
74. Musso, G. *et al.* The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Res.* **18**, 1092–1099 (2008).
75. Makino, T., Hokamp, K. & McLysaght, A. The complex relationship of gene duplication and essentiality. *Trends Genet.* **25**, 152–155 (2009).
76. Liao, B. Y. & Zhang, J. Mouse duplicate genes are as essential as singletons. *Trends Genet.* **23**, 378–381 (2007).
77. Aravind, L., Watanabe, H., Lipman, D. J. & Koonin, E. V. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl Acad. Sci. USA* **97**, 11319–11324 (2000).
78. Blanc, G. & Wolfe, K. H. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **16**, 1679–1691 (2004).
79. Maere, S. *et al.* Modeling gene and genome duplications in eukaryotes. *Proc. Natl Acad. Sci. USA* **102**, 5454–5459 (2005).
80. Seoighe, C. & Gehring, C. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet.* **20**, 461–464 (2004).
81. Chen, E. C. *et al.* The dynamics of functional classes of plant genes in rediploidized ancient polyploids. *BMC Bioinformatics* **14** (Suppl. 15), S19 (2013).
82. Blomme, T. *et al.* The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **7**, R43 (2006).
83. Demuth, J. P., De Bie, T., Stajich, J. E., Cristianini, N. & Hahn, M. W. The evolution of mammalian gene families. *PLoS ONE* **1**, e85 (2006).
84. Meslin, C. *et al.* Evolution of genes involved in gamete interaction: evidence for positive selection, duplications and losses in vertebrates. *PLoS ONE* **7**, e44548 (2012).
85. Koonin, E. V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7 (2004).
86. Castro, L. F. *et al.* Recurrent gene loss correlates with the evolution of stomach phenotypes in gnathostome history. *Proc. Biol. Sci.* **281**, 20132669 (2014).

87. Ordonez, G. R. *et al.* Loss of genes implicated in gastric function during platypus evolution. *Genome Biol.* **9**, R81 (2008).
88. Sire, J. Y., Delgado, S. C. & Girondot, M. Hen's teeth with enamel cap: from dream to impossibility. *BMC Evol. Biol.* **8**, 246 (2008).
89. International Aphid Genomics Consortium. Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol.* **8**, e1000313 (2010).
90. Paterson, A. H. *et al.* Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends Genet.* **22**, 597–602 (2006).
91. Freeling, M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **60**, 433–453 (2009).
92. Qian, W. & Zhang, J. Gene dosage and gene duplicability. *Genetics* **179**, 2319–2324 (2008).
93. Conant, G. C., Birchler, J. A. & Pires, J. C. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* **19**, 91–98 (2014).
94. McLysaght, A. *et al.* Ohnologs are overrepresented in pathogenic copy number mutations. *Proc. Natl Acad. Sci. USA* **111**, 361–366 (2014).
95. Cañestro, C., Albalat, R., Irimia, M. & Garcia-Fernandez, J. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Semin. Cell Dev. Biol.* **24**, 83–94 (2013).
96. Davis, J. C. & Petrov, D. A. Do disparate mechanisms of duplication add similar genes to the genome? *Trends Genet.* **21**, 548–551 (2005).
97. Scannell, D. R. *et al.* Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc. Natl Acad. Sci. USA* **104**, 8397–8402 (2007).
98. Langham, R. J. *et al.* Genomic duplication, fractionation and the origin of regulatory novelty. *Genetics* **166**, 935–945 (2004).
99. Thomas, B. C., Pedersen, B. & Freeling, M. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* **16**, 934–946 (2006).
100. Makino, T. & McLysaght, A. Positionally biased gene loss after whole genome duplication: evidence from human, yeast, and plant. *Genome Res.* **22**, 2427–2435 (2012).
101. Braasch, I. & Postlethwait, J. in *Polyploidy and Genome Evolution* (eds Soltis, P. S. & Soltis, D. E.) 341–383 (Springer, 2012).
102. Liu, S. *et al.* The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* **5**, 3930 (2014).
103. Schnable, J. C., Springer, N. M. & Freeling, M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl Acad. Sci. USA* **108**, 4069–4074 (2011).
104. Sémon, M. & Wolfe, K. H. Reciprocal gene loss between *Tetraodon* and zebrafish after whole genome duplication in their ancestor. *Trends Genet.* **23**, 108–112 (2007).
105. Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**, 341–345 (2006).
106. Mizuta, Y., Harushima, Y. & Kurata, N. Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proc. Natl Acad. Sci. USA* **107**, 20417–20422 (2010).
107. McGrath, C. L. & Lynch, M. in *Polyploidy and Genome Evolution* (eds Soltis, S. P. & Soltis, E. D.) 1–20 (Springer, 2012).
108. Kassahn, K. S., Dang, V. T., Wilkins, S. J., Perkins, A. C. & Ragan, M. A. Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome Res.* **19**, 1404–1418 (2009).
109. Bergero, R., Qiu, S. & Charlesworth, D. Gene loss from a plant sex chromosome system. *Curr. Biol.* **25**, 1234–1240 (2015).
110. Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493 (2014).
111. Bellott, D. W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
112. Hughes, J. F. *et al.* Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82–86 (2012).
113. Olson, M. V. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**, 18–23 (1999).
The author proposes the view of gene loss as a major force of molecular evolution and formulates the less-is-more hypothesis.
114. Olson, M. V. & Varki, A. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat. Rev. Genet.* **4**, 20–28 (2003).
115. Hottes, A. K. *et al.* Bacterial adaptation through loss of function. *PLoS Genet.* **9**, e1003617 (2013).
This work carries out selection experiments on mutagenized bacteria that show how substantial adaptation can be achieved solely through gene loss.
116. Sokurenko, E. V., Hasty, D. L. & Dykhuizen, D. E. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol.* **7**, 191–195 (1999).
117. Jain, N. *et al.* Loss of allergen 1 confers a hypervirulent phenotype that resembles mucoid switch variants of *Cryptococcus neoformans*. *Infect. Immun.* **77**, 128–140 (2009).
118. Maurelli, A. T., Fernandez, R. E., Bloch, C. A. & Rode, C. K. & Fasano, A. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella spp.* and enteroinvasive *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95**, 3943–3948 (1998).
119. Moore, R. A. *et al.* Contribution of gene loss to the pathogenic evolution of *Burkholderia pseudomallei* and *Burkholderia mallei*. *Infect. Immun.* **72**, 4172–4187 (2004).
120. Yu, H., Hanes, M., Chrisp, C. E., Boucher, J. C. & Deretic, V. Microbial pathogenesis in cystic fibrosis: pulmonary clearance of mucoid *Pseudomonas aeruginosa* and inflammation in a mouse model of repeated respiratory challenge. *Infect. Immun.* **66**, 280–288 (1998).
121. Domergue, R. *et al.* Nicotinic acid limitation regulates silencing of *Candida* adhesins during UTI. *Science* **308**, 866–870 (2005).
122. Will, J. L. *et al.* Incipient balancing selection through adaptive loss of aquaporins in natural *Saccharomyces cerevisiae* populations. *PLoS Genet.* **6**, e1000893 (2010).
123. Lu, X., Wu, Q., Zhang, Y. & Xu, Y. Genomic and transcriptomic analyses of the Chinese Maotai-flavored liquor yeast MT1 revealed its unique multi-carbon co-utilization. *BMC Genomics* **16**, 1064 (2015).
124. Hoballah, M. E. *et al.* Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell* **19**, 779–790 (2007).
125. Zufall, R. A. & Rausher, M. D. Genetic changes associated with floral adaptation restrict future evolutionary potential. *Nature* **428**, 847–850 (2004).
126. Shimizu, K. K., Shimizu-Inatsugi, R., Tsuchimatsu, T. & Purugganan, M. D. Independent origins of self-compatibility in *Arabidopsis thaliana*. *Mol. Ecol.* **17**, 704–714 (2008).
127. Greenberg, A. J., Moran, J. R., Coyne, J. A. & Wu, C. I. Ecological adaptation during incipient speciation revealed by precise gene replacement. *Science* **302**, 1754–1757 (2003).
128. McBride, C. S., Arguello, J. R. & O'Meara, B. C. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* **177**, 1395–1416 (2007).
129. Clark, A. G. *et al.* Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
130. Goldman-Huertas, B. *et al.* Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc. Natl Acad. Sci. USA* **112**, 3026–3031 (2015).
131. Dean, M. *et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CKR5* structural gene. *Science* **273**, 1856–1862 (1996).
132. Tournamille, C., Colin, Y., Cartron, J. P. & Le Van Kim, C. Disruption of a GATA motif in the *Duffy* gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat. Genet.* **10**, 224–228 (1995).
133. Howes, R. E. *et al.* The global distribution of the Duffy blood group. *Nat. Commun.* **2**, 266 (2011).
134. Hodgson, J. A. *et al.* Natural selection for the Duffy-null allele in the recently admixed people of Madagascar. *Proc. Biol. Sci.* **281**, 20140930 (2014).
The authors propose that null mutations of *DUFFY* have been positively selected, supporting the hypothesis that malaria resistance drove fixation of the *DUFFY*-null allele in mainland sub-Saharan Africa.
135. Novembre, J., Galvani, A. P. & Slatkin, M. The geographic spread of the CCR5 Δ32 HIV-resistance allele. *PLoS Biol.* **3**, e339 (2005).
The authors propose that null mutations of *CCR5* gene have been positively selected and show how long-range dispersal and selection gradients have been important processes for the spread of the advantageous null allele.
136. Galvani, A. P. & Novembre, J. The evolutionary history of the *CCR5-Δ32* HIV-resistance mutation. *Microbes Infect.* **7**, 302–309 (2005).
137. Saxena, S. K. Controversial role of smallpox on historical positive selection at the CCR5 chemokine gene (CCR5-Δ32). *J. Infect. Dev. Ctries* **3**, 324–326 (2009).
138. Hedrick, P. W. Population genetics of malaria resistance in humans. *Heredity* **107**, 283–304 (2011).
139. Stedman, H. H. *et al.* Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* **428**, 415–418 (2004).
This work reveals that the loss of *MYH16* in the human lineage after the separation from chimpanzees could have facilitated an increase in the size of the brain and the human origin.
140. Chou, H. H. *et al.* Inactivation of CMP-*N*-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl Acad. Sci. USA* **99**, 11736–11741 (2002).
141. Wang, X., Grus, W. E. & Zhang, J. Gene losses during human origins. *PLoS Biol.* **4**, e52 (2006).
142. Kvitek, D. J. & Sherlock, G. Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genet.* **9**, e1003972 (2013).
This work carries out whole-genome, whole-population sequencing on replicate evolution experiments that provide experimental evidence, supporting gene loss as an important adaptive evolutionary force responding to environmental perturbations.
143. Herron, M. D. & Doebeli, M. Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. *PLoS Biol.* **11**, e1001490 (2013).
144. Cooper, V. S., Schneider, D., Blot, M. & Lenski, R. E. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J. Bacteriol.* **183**, 2834–2841 (2001).
145. Moreau, R. & Dabrowski, K. Body pool and synthesis of ascorbic acid in adult sea lamprey (*Petromyzon marinus*): an agnathan fish with gulonolactone oxidase activity. *Proc. Natl Acad. Sci. USA* **95**, 10279–10282 (1998).
This work shows a paradigmatic case of recurrent gene loss in which genes are studied that are involved in the synthesis of vitamin C. These genes have been lost in several cases during vertebrate evolution, which is associated with changes in environmental conditions (specifically, diet).
146. Drouin, G., Godin, J. R. & Page, B. The genetics of vitamin C loss in vertebrates. *Curr. Genom.* **12**, 371–378 (2011).
147. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, 1872).
148. Protas, M. E. *et al.* Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* **38**, 107–111 (2006).
In this work, the authors illuminate one of the puzzling enigmas that has existed since the time of Darwin: regressive evolution of dispensable traits in perpetual dark environments. The authors identify independent loss-of-function mutations in *Oca2*, which lead to the loss of pigmentation and vision in different cavefish populations.
149. Leys, R., Cooper, S. J., Strecker, U. & Wilkens, H. Regressive evolution of an eye pigment gene in independently evolved eyeless subterranean diving beetles. *Biol. Lett.* **1**, 496–499 (2005).
150. Protas, M. E., Trontelj, P. & Patel, N. H. Genetic basis of eye and pigment loss in the cave crustacean, *Asellus aquaticus*. *Proc. Natl Acad. Sci. USA* **108**, 5702–5707 (2011).

151. Chipman, A. D. *et al.* The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLoS Biol.* **12**, e1002005 (2014).
152. Zhao, H. *et al.* The evolution of color vision in nocturnal mammals. *Proc. Natl Acad. Sci. USA* **106**, 8980–8985 (2009).
153. Kim, E. B. *et al.* Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* **479**, 223–227 (2011).
154. Emerling, C. A. & Springer, M. S. Eyes underground: regression of visual protein networks in subterranean mammals. *Mol. Phylogenet Evol.* **78**, 260–270 (2014).
155. Cavalier-Smith, T. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Ann. Bot.* **95**, 147–175 (2005).
156. Kuo, C. H., Moran, N. A. & Ochman, H. The consequences of genetic drift for bacterial genome complexity. *Genome Res.* **19**, 1450–1454 (2009).
157. Rocha, E. P. Inference and analysis of the relative stability of bacterial chromosomes. *Mol. Biol. Evol.* **23**, 513–522 (2006).
158. Karcagi, I. *et al.* Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining. *Mol. Biol. Evol.* http:dx.doi.org/10.1093/molbev/msw009 (2016).
159. Lang, G. I., Murray, A. W. & Botstein, D. The cost of gene expression underlies a fitness trade-off in yeast. *Proc. Natl Acad. Sci. USA* **106**, 5755–5760 (2009).
160. Wilkens, H. Genes, modules and the evolution of cave fish. *Heredity* **105**, 413–422 (2010).
161. Yoshizawa, M., O'Quin, K. E. & Jeffery, W. R. Evolution of an adaptive behavior and its sensory receptors promotes eye regression in blind cavefish: response to Borowsky (2013). *BMC Biol.* **11**, 82 (2013).
162. Gunter, H. & Meyer, A. Trade-offs in cavefish sensory capacity. *BMC Biol.* **11**, 5 (2013).
163. McGaugh, S. E. *et al.* The cavefish genome reveals candidate genes for eye loss. *Nat. Commun.* **5**, 5307 (2014).
164. Nei, M., Suzuki, Y. & Nozawa, M. The neutral theory of molecular evolution in the genomic era. *Annu. Rev. Genom. Hum. Genet.* **11**, 265–289 (2010).
165. Wagner, A. Neutralism and selectionism: a network-based reconciliation. *Nat. Rev. Genet.* **9**, 965–974 (2008).
166. Moran, N. A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**, 583–586 (2002).
167. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245 (2005).
168. Mira, A., Ochman, H. & Moran, N. A. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**, 589–596 (2001).
169. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
170. Wyder, S., Kriventseva, E. V., Schroder, R., Kadowaki, T. & Zdobnov, E. M. Quantification of ortholog losses in insects and vertebrates. *Genome Biol.* **8**, R242 (2007).
171. Ptitsyn, A. & Moroz, L. L. Computational workflow for analysis of gain and loss of genes in distantly related genomes. *BMC Bioinformatics* **13** (Suppl. 15), S5 (2012).
172. Dainat, J., Paganini, J., Pontarotti, P. & Gouret, P. GLADX: an automated approach to analyze the lineage-specific loss and pseudogenization of genes. *PLoS ONE* **7**, e38792 (2012).
173. Novozhilov, A. S., Karev, G. P. & Koonin, E. V. Biological applications of the theory of birth-and-death processes. *Brief Bioinform.* **7**, 70–85 (2006).
174. Baurain, D., Brinkmann, H. & Philippe, H. Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors? *Mol. Biol. Evol.* **24**, 6–9 (2007).
175. Postlethwait, J. H. The zebrafish genome in context: ohnologs gone missing. *J. Exp. Zool. B Mol. Dev. Evol.* **308**, 563–577 (2007).
176. Catchen, J. M., Conery, J. S. & Postlethwait, J. H. Automated identification of conserved synteny after whole-genome duplication. *Genome Res.* **19**, 1497–1505 (2009).
177. Tang, H. *et al.* SynFind: compiling syntenic regions across any set of genomes on demand. *Genome Biol. Evol.* **11**, 3286–3298 (2015).
178. Zhu, J. *et al.* Comparative genomics search for losses of long-established genes on the human lineage. *PLoS Comput. Biol.* **3**, e247 (2007).
179. Zhang, Z. D., Frankish, A., Hunt, T., Harrow, J. & Gerstein, M. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol.* **11**, R26 (2010).
180. Zhao, Y. *et al.* Identification and analysis of unitary loss of long-established protein-coding genes in Poaceae shows evidences for biased gene loss and putatively functional transcription of relics. *BMC Evol. Biol.* **15**, 66 (2015).
181. Inoue, J., Sato, Y., Sinclair, R., Tsukamoto, K. & Nishida, M. Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. *Proc. Natl Acad. Sci. USA* **112**, 14918–14923 (2015).
182. Hiller, M. *et al.* A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Rep.* **2**, 817–823 (2012).
The authors introduce a computational 'forward genomics' strategy that is able to associate mutations in specific genomic regions with phenotypic losses.
183. Hiller, M., Schaar, B. T. & Bejerano, G. Hundreds of conserved non-coding genomic regions are independently lost in mammals. *Nucleic Acids Res.* **40**, 11463–11476 (2012).
184. MacArthur, D. G. & Tyler-Smith, C. Loss-of-function variants in the genomes of healthy humans. *Hum. Mol. Genet.* **19**, R125–R130 (2010).
185. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
The authors use the whole-genome sequences of 185 humans to show that there are approximately 80 heterozygous and, importantly, approximately 20 homozygous loss-of-function variants in a typical healthy individual, supporting the presence of a substantial number of non-functional variants in natural populations.
186. Burgess, D. J. Genomics: how pervasive are defective genes? *Nat. Rev. Genet.* **13**, 222 (2012).
187. Henn, B. M., Botigue, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nat. Rev. Genet.* **16**, 333–343 (2015).
188. Rivas, M. A. *et al.* Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348**, 666–669 (2015).
189. Lim, E. T. *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* **10**, e1004494 (2014).
The authors describe how loss-of-function alleles of *LPA* gene confer protection from cardiovascular disease, providing proof of concept of the potential of population gene loss analyses for biomedical studies.
190. Koch, L. Population genomics: a new window into the genetics of complex diseases. *Nat. Rev. Genet.* **15**, 644–645 (2014).
191. Cañestro, C. in *Polyploidy and Genome Evolution* (eds Soltis, P. S. & Soltis, D. E.) 309–339 (Springer, 2012).
192. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
193. Hahn, M. W., Demuth, J. P. & Han, S. G. Accelerated rate of gene gain and loss in primates. *Genetics* **177**, 1941–1949 (2007).
194. Bouquet, J. M. *et al.* Culture optimization for the emergent zooplanktonic model organism *Oikopleura dioica*. *J. Plankton Res.* **31**, 359–370 (2009).
195. Marti-Solans, J. *et al. Oikopleura dioica* culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. *Genesis* **53**, 183–193 (2015).
196. Omotezako, T., Onuma, T. A. & Nishida, H. DNA interference: DNA-induced gene silencing in the appendicularian *Oikopleura dioica*. *Proc Biol Sci* **282**, 20150435 (2015).
197. Albalat, R., Marti-Solans, J. & Cañestro, C. DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates. *Brief Funct. Genom.* **11**, 142–155 (2012).
198. Fu, X., Adamski, M. & Thompson, E. M. Altered miRNA repertoire in the simplified chordate. *Oikopleura dioica*. *Mol. Biol. Evol.* **25**, 1067–1080 (2008).
199. Weill, M., Philips, A., Chourrout, D. & Fort, P. The caspase family in urochordates: distinct evolutionary fates in ascidians and larvaceans. *Biol. Cell* **97**, 857–866 (2005).
200. Yadetie, F. *et al.* Conservation and divergence of chemical defense system in the tunicate *Oikopleura dioica* revealed by genome wide response to two xenobiotics. *BMC Genomics* **13**, 55 (2012).
201. Seo, H. C. *et al.* Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431**, 67–71 (2004).
202. Cañestro, C., Postlethwait, J. H., Gonzàlez-Duarte, R. & Albalat, R. Is retinoic acid genetic machinery a chordate innovation? *Evol. Dev.* **8**, 394–406 (2006).
203. Duester, G. Retinoic acid synthesis and signaling during early organogenesis. *Cell* **134**, 921–931 (2008).
204. Cañestro, C. & Postlethwait, J. H. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev. Biol.* **305**, 522–538 (2007).
205. Blanpain, C. *et al.* Multiple nonfunctional alleles of CCR5 are frequent in various human populations. *Blood* **96**, 1638–1645 (2000).
206. Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**, 931–949 (2006).
207. Wang, Y. *et al.* Functional CpG methylation system in a social insect. *Science* **314**, 645–647 (2006).
208. Albalat, R. Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus *Branchiostoma floridae*. *Dev. Genes Evol.* **218**, 691–701 (2008).
209. Tribolium Genome Sequencing Consortium. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**, 949–955 (2008).
210. Colbourne, J. K. *et al.* The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555–561 (2011).
211. Moroz, L. L. *et al.* The ctenophore genome and the evolutionary origins of neural systems. *Nature* **510**, 109–114 (2014).
212. Pisani, D. *et al.* Genomic data do not support comb jellies as the sister group to all other animals. *Proc. Natl Acad. Sci. USA* **112**, 15402–15407 (2015).
213. Borenstein, E., Shlomi, T., Ruppin, E. & Sharan, R. Gene loss rate: a probabilistic measure for the conservation of eukaryotic genes. *Nucleic Acids Res.* **35**, e7 (2007).
214. Krylov, D. M., Wolf, Y. I., Rogozin, I. B. & Koonin, E. V. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* **13**, 2229–2235 (2003).
215. Wolf, Y. I., Carmel, L. & Koonin, E. V. Unifying measures of gene function and evolution. *Proc. Biol. Sci.* **273**, 1507–1515 (2006).
216. Albertson, R. C., Cresko, W., Detrich, H. W. 3rd & Postlethwait, J. H. Evolutionary mutant models for human disease. *Trends Genet.* **25**, 74–81 (2009).
217. Postlethwait, J. H. "Wrecks of ancient life": genetic variants vetted by natural selection. *Genetics* **200**, 675–678 (2015).

Acknowledgements
The authors thank the interesting and helpful comments of the three anonymous thoughtful referees. The authors apologize to the researchers whose work has not been directly cited owing to space restrictions. Support is acknowledged from past grant BFU2010-14875 from Ministerio de Economía y Competitividad (Spain) and SGR2014-290 from Generalitat de Catalunya. The authors also thank the team members of the Cañestro and Albalat laboratories for fruitful discussions on *Oikopleura*'s passion for gene loss.

Competing interests statement
The authors declare no competing interests.

DATABASES
CoGe: SynFind: https://genomevolution.org/CoGe/SynFind.pl
Database of Essential Genes (DEG): http://www.essentialgene.org
Ensembl Gene gain/loss tree: http://www.ensembl.org/Help/View?id=379
Genomicus: http://www.genomicus.biologie.ens.fr/genomicus-82.01/cgi-bin/search.pl
RCSB Protein Data Bank: http://www.rcsb.org/pdb/home/home.do
Synteny Database: http://syntenydb.uoregon.edu/synteny_db

TOOLS
Gene Loss Analyzer DAGOBAH eXtension (GLADX): http://ioda.univ-provence.fr/lodaSite/gladx
TransMap Super-track settings: http://www.noncode.org/cgi-bin/hgTrackUi?hgsid=8375&c=chr8&g=transMap

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (table) | S3 (box)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Author biographies

Ricard Albalat is an associate professor in the Department of Genetics and a member of the Biodiversity Research Institute (IRBio) at the Universitat de Barcelona, Spain. He began his research career investigating the evolution of several gene families in *Drosophila* spp., before switching to analysing similar gene families in chordates. He has also investigated the impact of transposable elements and the evolution of epigenetics mechanisms in chordate genomes. His recent work has focused on the effect of gene losses on the evolution of developmental gene networks in chordates by comparative studies between vertebrates and the urochordate *Oikopleura dioica*. This emergent model animal has suffered an extreme genome compaction accompanied of massive gene loss events. Ricard Albalat's homepage: http://www.ub.edu/genetica/evolucioen/albalatre.htm

Cristian Cañestro is an associate professor in the Department of Genetics and a member of the Biodiversity Research Institute (IRBio) at the Universitat de Barcelona, Spain. His research experience is in the fields of evo–devo and genomics. During his career, he has studied several animal models — amphioxus, ascidians, larvaceans and zebrafish — to investigate the origin and evolution of our phylum, the chordates, and to develop models for human diseases. Currently, the work of his laboratory focuses on the study of *Oikopleura dioica* as a model for investigating the impact of gene loss on evo–devo, paying special attention to the heart, nervous system and maternal effect. *O. dioica* is also used as a model for exploring the power of gene loss as an adaptive evolutionary force for rapidly changing environments in the context of global warming. Cristian Cañestro's homepage: http://www.ub.edu/genetica/evo-devoen/canestrore.htm

Key points

- The recent increase in genomic data is revealing a novel perspective of gene loss as a pervasive source of genetic variation in all life kingdoms.
- Gene loss depends on gene dispensability, which in turn is affected by changes in mutational robustness and environmental conditions.
- Patterns of gene loss are not stochastic but show biases that are associated with gene functions and genomic positions.
- Although many gene losses are neutral and fixed by genetic drift, many examples support the idea that gene loss can be an adaptive evolutionary force that is especially effective when organisms are faced with abrupt environmental challenges.
- The future mapping of all instances of gene loss in the tree of life will provide valuable information for many fields of biology, including evolutionary biology and translational medicine.
- Population genomics might expose ongoing processes of gene loss in natural populations, revealing actual values of gene dispensability and identifying adaptive gene losses with potential interest in biomedicine.

Permission information

Box 2: The *Ciona intestinalis* photograph in the figure is from REF. 13, Nature Publishing Group.

Subject categories

Biological sciences / Evolution / Evolutionary genetics [URI /631/181/2474]
Biological sciences / Genetics / Population genetics / Genetic variation [URI /631/208/457/649]
Biological sciences / Genetics / Genomics / Comparative genomics [URI /631/208/212/748]
Biological sciences / Genetics / Sequencing / Next-generation sequencing [URI /631/208/514/2254
Biological sciences / Genetics / Population genetics [URI /631/208/457]

ToC blurb

000 Evolution by gene loss

Ricard Albalat and Cristian Cañestro

Gene loss is emerging as a pervasive source of genetic variation. The authors review the mechanisms by which gene loss has influenced evolution of different species and discuss insights from comparative population genomics studies of gene loss. Further, they highlight future directions for the study of gene losses and their relevance in evolutionary biology and biomedicine.