Chapter 13
Genetic Redundancies and Their Evolutionary Maintenance

Jianzhi Zhang

Abstract Genetic redundancy refers to the common phenomenon that deleting or mutating a gene from a genome has minimal or no impact on the phenotype or fitness of the organism because of functional compensation conferred by one or more other genes. Here I summarize studies of functional redundancies between duplicate genes and those among metabolic reactions that respectively represent genetic redundancies at the individual gene level and at the systems level. I discuss the prevalence of genetic redundancies in a genome, evolutionary origins of these redundancies, and mechanisms responsible for their stable maintenance. I show that genetic redundancies are highly abundant. While some of them may be evolutionarily transient, many are stable. The majority of the stable redundancies are likely to have been selectively kept, not because of their potential benefits in regard to future deleterious mutations, but because of their actual benefits at present or in the recent past. The rest are probably preserved by selection on nonredundant pleiotropic functions. The studies summarized here illustrate the utility of systems analysis for understanding evolutionary phenomena and the importance of evolutionary thinking in uncovering the functions and origins of systemic properties.

1 Introduction

There are many concepts in genetics that are inherently suitable for systems analysis, because they involve interactions among multiple components of a system. They include, for example, epistasis, pleiotropy, complex traits, and redundancy. Genetic redundancy refers to the situation where the loss of a gene can be completely or partially compensated by one or more other genes. Examples of

J. Zhang (✉)
Department of Ecology and Evolutionary Biology, University of Michigan, 1075 Natural Science Building, 830 North University Avenue, Ann Arbor, MI 48109, USA
e-mail: jianzhi@umich.edu

O.S. Soyer (ed.), *Evolutionary Systems Biology*, Advances in Experimental Medicine and Biology 751, DOI 10.1007/978-1-4614-3567-9_13,

genetic redundancies are ample in the literature and can be roughly divided into two types. The first type of redundancy occurs at the individual gene level such as that between isoenzymes, which are generated by gene duplication and differ in protein sequence but catalyze the same biochemical reactions in an organism. The second type of redundancy occurs at the systems level, due to distributed properties of networks [20, 57]. For example, glucose-6-phosphate dehydrogenase and D-ribulose-5-phosphate 3-epimerase catalyze distinct reactions and are located in alternative pentose phosphate pathways in yeast; simultaneous removal of the two enzymes is lethal, although individual removal of either enzyme is not [19].

An important consequence of genetic redundancy is robustness against genetic perturbations such as deleterious mutations. Genetic robustness is a characteristic of cellular life, observed in all domains of life and at many levels of biological organizations, from DNA replication, transcription, and translation, to metabolism, cell cycle, and embryonic development [5, 33, 59]. Thus, genetic redundancy and robustness have important relevance to development and health. In this review, I discuss the abundance of genetic redundancy as well as the mechanisms responsible for its origin and evolutionary maintenance. I focus on genome-scale empirical studies, because they provide the least biased and the most general pictures. Simple population genetic models are occasionally used to assist the description and understanding of some conceptual issues involved. I hope that these discussions also illustrate the utility of systems analysis in studying evolutionary phenomena and the tremendous value of evolutionary thinking for understanding systemic properties.

2 Functional Redundancy Between Duplicate Genes

Gene duplication is a frequent event in genome evolution across all three domains of life [62]. It has been estimated that gene duplication occurs at a rate of $\sim$1 per gene per million years [39]. However, most duplicate genes, even after they are fixed in a population, quickly pseudogenize and get lost. This is because the daughter genes generated by duplication are typically identical in function. As a result, mutations disrupting one of the duplicates can accumulate freely by genetic drift. Thus, the duplicate genes observed in a genome today are only a small fraction of those that were once present and fixed in the genome. Of the duplicates observed today, some may still possess a certain degree of functional redundancy, because it can take million of years for functional divergence between duplicates to accumulate [22]. Several early studies in mice found that knocking out one copy of a duplicate gene pair has moderate or no phenotypic effects [28, 47], prompting the hypothesis that many mouse duplicates are functionally redundant [4, 51, 52].

Andreas Wagner was the first to examine this problem at a large scale [55]. He compared the average fitness effect of deleting a singleton gene and that of deleting a duplicate gene using the then available gene deletion strains of the budding yeast *Saccharomyces cerevisiae*. He did not find a significant difference between them when 42 duplicate genes and 205 singleton genes were compared.

He also failed to observe a negative correlation between the fitness effect of deleting a duplicate gene and the sequence similarity between the deleted gene and its closest paralog, which is predicted because functional redundancy between duplicates presumably depends on their sequence similarity. Thus, he concluded that functional redundancy between duplicates must be absent or rare.

Gu and colleagues revisited this issue in yeast when the genome-wide collection of gene deletion strains became available [17]. They not only observed a significantly smaller average fitness effect of deleting a duplicate gene than deleting a singleton gene (Table 13.1) but also found the predicted negative correlation between the fitness effect of deleting a duplicate gene and its sequence similarity to the closest paralog. They further showed that deleting the more highly expressed copy of a duplicate pair has a stronger fitness effect than deleting the more weakly expressed copy, suggesting asymmetric functional compensation due to expression-level differences between duplicates. Similar findings of differences between singletons and duplicates were made in the fission yeast *Schizosaccharomyces pombe* using genome-wide gene deletion data [31], in the nematode *Caenorhabditis elegans* [3] using genome-wide RNA interference (RNAi)-based gene knockdown data, and in the flowering plant *Arabidopsis thaliana* using a moderate-size random mutagenesis study augmented with literature-curated phenotypic data of about 1,000 mutants [18] (Table 13.1).

These studies were accompanied by similar investigations in several other organisms, but the observed patterns or their interpretations are more complicated, for various reasons. For example, Hsiao and Vitkup showed that, compared to human genes with only distant paralogs, genes with a close paralog tend not to harbor disease mutations [26] (Table 13.1). The interpretation of this finding, however, is confounded by the fact that the fitness effect of disrupting a disease gene may not be greater than that of disrupting a non-disease gene, because some non-disease genes presumably cause embryonic lethality when mutated. Indeed, the same group reported that non-disease genes are more likely than disease genes to be essential (i.e., cause lethality or infertility when deleted) [13].

Analyzing the phenotypic data of $\sim$4,000 gene-knockout mice, two groups independently reported that the probability that a gene is essential (P_E) is similar among singletons and duplicates [35, 38] (Table 13.1). Further, the P_E of a duplicate gene is not correlated with the sequence similarity between the gene and its closest paralog, nor with the number of paralogs it has in the genome [38]. However, duplicates chosen for knockout studies tend to be more conserved in sequence than average duplicates in the genome, while the opposite is true for singletons [38]. Because gene essentiality and sequence conservation are positively correlated in rodents [37], the above biases led to an overestimation of P_E for duplicates and an underestimation of P_E for singletons. However, correcting these biases did not make P_E significantly lower for duplicates (53.2%) than for singletons (56.8%) [38]. These unexpected results stimulated several subsequent analyses of the mouse data. One of them noted that P_E is higher for duplicates than for singletons when genes involved in development are considered, but the opposite is true for genes with other functions [40]. The authors found that many duplicated developmental genes were

Table 13.1 Fractions of essential genes among duplicates and singletons

Species	Methods	Fractions of essential genes (no. of genes examined)		References
		Singletons	Duplicates	
Budding yeast (*Saccharomyces cerevisiae*)	Systematic gene knockout	0.290 (1275)	0.124 (1147)	[17]
Fission yeast (*Schizosaccharomyces pombe*)	Systematic gene knockout	0.307 (3816)	0.085 (1020)	[31]
Thale cress (*Arabidopsis thaliana*)	Systematic and individual insertion mutagenesis	0.445 (220)[a]	0.287 (4071)[a]	[18]
Nematode (*Caenorhabditis elegans*)	Systematic gene knockdown	0.076 (8861)	0.023 (4704)	[3]
Fruit fly (*Drosophila melanogaster*)	Individual gene knockdown	0.351 (245)[b]	0.303 (195)[c]	[2]
Mouse (*Mus musculus*)	Individual gene knockout	0.554 (785)	0.551 (3087)	[38]
Human (*Homo sapiens*)	Natural disease mutations	0.284[d]	0.082[e]	[26]

[a]Fractions of genes that cause detectable phenotypes when disrupted
[b]Fraction of essential genes among old duplicates and singletons
[c]Fraction of essential genes among newly duplicated genes
[d]Fraction of duplicates harboring disease mutations when the duplicates have only distant paralog
[e]Fraction of duplicates harboring disease mutations when the duplicates have close paralogs

produced in two rounds of genome duplication during early vertebrate evolution and suspected that their exceptionally high P_E is due to hypersensitivity to dosage balance [40]. Another study noted that mouse duplicate genes tend to have more protein interaction partners than singletons, suggesting that mouse duplicates are performing more functions than singletons do [36]. If one compares duplicates and singletons with similar numbers of protein partners, P_E is significantly lower for the former than the latter [36].

Somewhat similar to the mouse observations, Chen et al. reported comparable levels of P_E for newly duplicated genes and for old duplicates and singletons in the fruit fly *Drosophila melanogaster* (Table 13.1), based on RNAi gene knockdown data in the literature as well as from their own experiments [2].

It thus appears that data generated from systematic genome-wide surveys, ranging from fungi (*S. cerevisiae* and *S. pombe*), animals (*C. elegans*), to plants (*A. thaliana*), all suggest that on average deleting a duplicate gene has a smaller phenotypic or fitness effect than deleting a singleton gene, which is consistent with the idea that duplicate genes are functionally redundant to a certain degree. By contrast, data derived from literature curation or nonsystematic surveys either do not support or do not support unambiguously the above conclusion. While one cannot rule out the possibility that this dichotomy reflects a genuine variation among species, a more likely explanation is that the nonsystematic data are subject to numerous potential biases and are less reliable. We will probably know the answer for the mouse in the near future, as a result of the ongoing effort to knockout every single mouse gene and examine its phenotypic effects [1]. For the time being, it is reasonable to assume that the results from the systematic surveys are more trustworthy.

The inference of functional redundancy from the comparison of phenotypic effects of null mutations in singletons and duplicates is based on the assumption that the intrinsic functional importance is similar between a duplicate gene and a singleton gene such that the observed disparity in the fitness effect of deletion is entirely attributable to functional compensation that is present only between duplicates. To verify this assumption empirically, He and Zhang measured in other fungal species the duplicabilities of *S. cerevisiae* singleton genes of various functional importance [23]. They found that less important *S. cerevisiae* singleton genes have higher rates of duplication in other fungi. This phenomenon may be explained by a smaller disturbance to cellular physiology by duplication of a less important gene than that of a more important one [23]. For example, protein complex members tend to be more important than nonmembers [25], but the duplicability is lower for the former than the latter, because duplication of a protein complex member is more likely to cause dosage imbalance-induced harm than that of a nonmember [43].

Because the finding of higher duplicability of less important genes can also explain, at least qualitatively, the disparity in the fitness effect of deleting a duplicate gene and deleting a singleton gene, additional experiments are necessary to demonstrate functional compensation between duplicates. These experiments have been independently conducted in yeast by three groups [6, 8, 41]. Instead of

deleting one gene at a time, they compared the fitness effect of simultaneously deleting a duplicate pair with that of deleting individual duplicates. Functional redundancy and compensation between duplicates predict a larger fitness effect from deleting a pair of duplicates simultaneously than expected from the combined effects of individual deletions, a phenomenon that is also known as negative (or synergistic) epistasis in genetics [21]. Negative epistasis is indeed observed for many duplicate genes in yeast, some of which are phylogenetically quite old [6,8,41]. Thus, functional redundancy and compensation do exist between duplicate genes, although the amount may have been overestimated [17], due to differential duplicabilities of genes with different levels of functional importance [23].

3 Evolutionary Maintenance of Functional Redundancy Between Duplicate Genes

Because newly duplicated genes naturally share common functions, the existence of a certain degree of functional redundancy between young duplicates is expected. What is puzzling is the functional redundancy between old duplicates [6, 53, 54], some of which originated as early as a billion years ago [54]. As mentioned above, stable retention of functional redundancy between duplicate genes is unexpected, because mutations disrupting a completely redundant gene have no fitness effect and thus will accumulate in the gene, leading to its eventual degeneration.

A simple explanation of the stable preservation of functional redundancy between duplicate genes that is favored by some systems biologists is that a redundant gene copy can back up the system in case deleterious mutations occur in the other copy, very much like the function of spare tires that we carry in our cars. While cars are human-designed, biological systems are the results of evolutionary processes that are fundamentally different from engineering. Specifically, evolution and natural selection is shortsighted; it is impacted only by the current situation, not the future, although past situations may leave marks in the genome or constrain/channel subsequent evolutionary paths.

For the backup hypothesis to work, the benefit that the redundancy bestows must occur in the present. We can examine the feasibility of this scenario using population genetic theories. For simplicity, let us consider a haploid population encompassing two genotypes. The first genotype harbors two genes with the same function, while the second possesses only one of the genes. Let the null mutation rate be u per gene per generation and the fitness effect of loosing the gene function be s. For an individual with the first genotype, the probability that its two redundant genes are both nonfunctionalized by mutations is u^2 per generation, and the individual has a fitness of 1–s when this event happens; otherwise, the individual has a fitness of 1. Thus, after one generation of mutation, the expected fitness of the first genotype is $f_1 = u^2 \times (1-s) + (1-u^2) \times 1 = 1 - su^2$. For an individual with the second genotype, the probability that its nonredundant single-copy gene

is nonfunctionalized by a mutation is u per generation and the individual has a fitness of 1–s when this happens; otherwise, the individual has a fitness of 1. Thus, after one generation of mutation, the expected fitness of the second genotype is $f_2 = u \times (1-s) + (1-u) \times 1 = 1 - su$. Apparently, $f_1 > f_2$. However, in a finite population, the two genotypes are effectively neutral to each other if their fitness difference $\Delta f = f_1 - f_2 = su(1-u) \approx su$ is smaller than the inverse of the effective population size N_e. That is, the condition for a selective maintenance of redundant genes is $su > 1/N_e$, or $N_e su > 1$. Even for an essential gene (i.e., $s = 1$), it is difficult to satisfy the above condition, because in most cellular species, $N_e u$ is not greater than 1. For example, *S. cerevisiae* has an N_e of $\sim 10^7$ [58] and an u of $\sim 4 \times 10^{-8}$ per gene per generation [61], yielding an $N_e u$ of 0.4. Selective maintenance of redundant duplicates in diploids is even more difficult because most genes are haplosufficient, meaning that losing one allele has little fitness effect, compared to loosing both alleles of a gene. Under haplosufficiency, the backup hypothesis would not work unless $N_e su^2$ exceeds 1, a condition that is unlikely to exist for any diploid organism. While we considered only null mutations in the above formulation, the formula is the same when all types of deleterious mutations are considered, except that su (or su^2) is replaced with $\sum s_i u_i$ (or $\sum s_i u_i^2$), where u_i is the rate of ith type of mutation and s_i is the associated fitness effect.

Given the above theoretical consideration, it is surprising that some empirical data appear to support the backup hypothesis in explaining the persistence of redundant duplicate genes in yeast. For example, Kafri et al. reported that P_E is lower for yeast duplicates with intermediate levels of expression similarity than those with high similarity [29]. This unexpected finding was explained by transcriptional reprogramming mediated backup, in which a pair of duplicate genes are normally expressed differently; but, when one of them is deleted, the other can compensate it by changing its expression to that of the deleted one [29]. This extraordinary scenario is a derivative of the simple backup hypothesis above analyzed, because it essentially assumes that natural selection preserves the reprogramming ability of a gene because the gene can then be used as a backup for its paralog in case the paralog is damaged. The above theoretical analysis suggests that this scenario is highly unlikely. Indeed, He and Zhang found that Kafri et al.'s unexpected observation was due to a confounding factor of the number of protein interactions per gene [24]. Specifically, they found that genes with different levels of expression similarity are not directly comparable in terms of P_E, because they have different numbers of protein interactions, which also affect P_E. After the control of the number of protein interactions, the unexpected relationship between P_E and expression similarity that Kafri et al. observed disappears [24], removing the need for the backup hypothesis. A subsequent yeast experiment explicitly tested Kafri et al.'s hypothesis by measuring the expression change of a gene upon the deletion of its paralog [7]. While the majority of the 202 genes tested showed no detectable change of expression, 23 genes showed increased expression and 6 showed reduced expression. Interestingly, a further scrutiny of the upregulated ones showed that the upregulation is need-based. That is, the upregulation generally occurs when the gene product is useful in the medium tested. While this observation appears to support

the reprogramming-based backup hypothesis for a small number of duplicate genes, alternative explanations exist and are more likely to be true than the backup hypothesis. For example, when a pair of paralogous genes encode two enzymes of the same biosynthetic pathway, both genes may be regulated by a negative feedback from the final product of the pathway [7]. When one paralog is deleted, the final product of the pathway is reduced, triggering an increased expression of the other paralog. Apparently, one does not need to invoke the backup hypothesis to explain the existence of this reprogramming mechanism, because here the two paralogs do not even need to catalyze the same reaction in the pathway. Even if they do catalyze the same reaction, their apparent response to each other's loss is a byproduct of the feedback regulation of themselves. While this study investigated non-ribosomal protein genes, another study focused exclusively on ribosomal protein (RP) genes and revealed very different patterns [44]. They showed in yeast that removing the intron in an RP gene often results in a change of the mRNA level of the gene itself as well as that of its paralog. But the changes are often in the same direction rather than compensatory, suggesting that the transcriptional reprogramming here seldom allows a backup. In sum, in addition to the lack of theoretical basis, there is little empirical support for the backup hypothesis.

If the backup hypothesis does not work, how can we explain the long retention of functionally redundant duplicates in a genome? One possibility is the so-called piggyback hypothesis, which states that two duplicates may continue to share some functions because these shared functions are impossible to lose provided that their other unshared functions are indispensable [54]. In other words, the conservation of the redundant function is a byproduct of structural or other functional constraints, rather than a result from natural selection for redundancy.

Qian and colleagues recently put forward another mechanism that is perhaps more general [46]. They propose that after gene duplication the amount of expression of each daughter gene is reduced, relative to the expression of the progenitor gene. The expression reduction prohibits the loss of either daughter gene because the loss would render the total expression level after duplication lower than that before duplication, which would be deleterious. The expression reduction, when it is sufficiently large, would require both daughter genes to retain all ancestral functions, preventing the occurrence of functional divergence. In this model, although the two daughter genes are functionally equivalent, they are not redundant in a strict sense, because the deletion of either copy is expected to cause a fitness reduction that is sufficiently large to be disfavored by natural selection. Negative epistasis between functionally equivalent duplicates results from the well-established nonlinear relationship between gene expression level and fitness [32] (Fig. 13.1a). That is, the fitness effect of reducing the expression level by 50% is less than 50% [9]. This phenomenon is closely related to the observations that most genes are haplosufficient [9, 31] and that most wild-type alleles are dominant to loss-of-function alleles [9, 31]. Expression reduction after gene duplication can happen either neutrally or by positive selection. The former occurs when mutations reducing gene expression are not advantageous, whereas the latter occurs when such mutations are beneficial, which is possible if the gene dosage prior to duplication

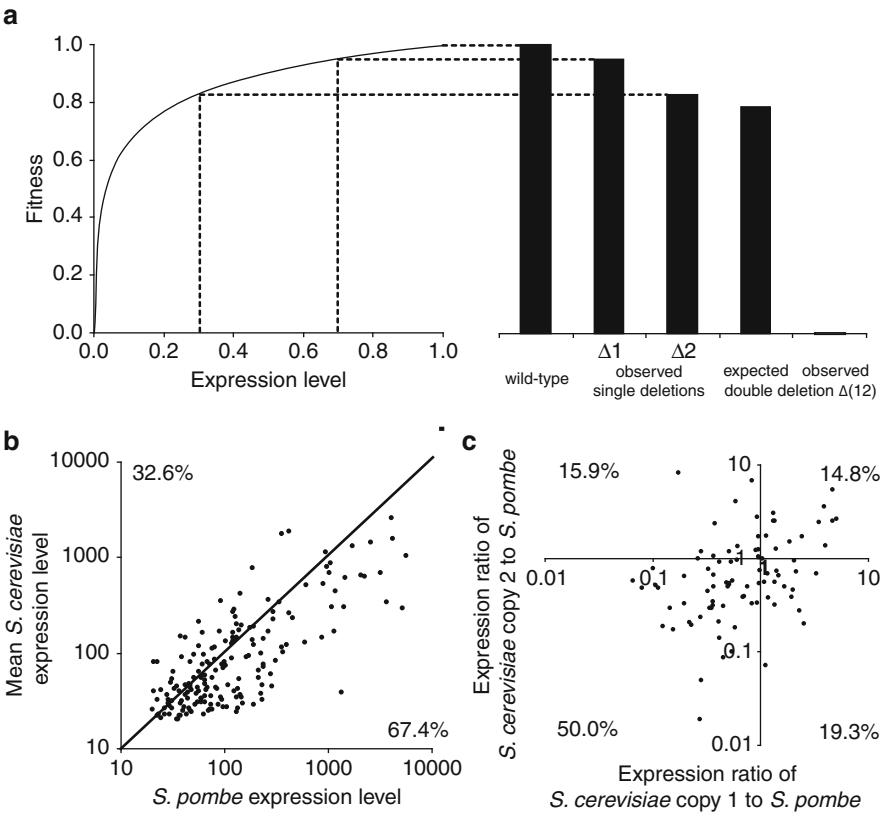

Fig. 13.1 Expression reduction after gene duplication can explain evolutionary preservation of functionally redundant duplicates. (**a**) Because fitness is a concave function of gene expression level, there is negative epistasis between duplicates with reduced expression. Synthetic lethality is observed in this hypothetical example. The fitness of the double deletion strain expected under no epistasis is calculated assuming multiplicative fitness effects of single deletions. (**b**) Each *dot* represents a two-to-one orthologous trio, for which the mean expression level of the *S. cerevisiae* duplicates and the expression level of the single-copy *S. pombe* ortholog are shown. The percentages of *dots* below and above the diagonal are presented. (**c**) Expression ratios between *S. cerevisiae* and *S. pombe* for all two-to-one orthologs. Adapted from [46] with permission

is optimal and an extra gene dose is harmful. By analyzing the RNA-sequencing-based gene expression data from *S. cerevisiae* and *S. pombe*, Qian et al. indeed found expression reduction after gene duplication [46]. They identified 227 *S. pombe* genes that were duplicated in *S. cerevisiae*. Among these two-to-one orthologs, 67.4% have lower mean expressions in *S. cerevisiae* than in *S. pombe* (Fig. 13.1b). This fraction is significantly higher than that in 891 one-to-one orthologs (52.4%). In addition to random expression changes, an excess of 31.5% of duplicate gene pairs experienced mean expression reduction after gene duplication. The median expression ratio (*S. cerevisiae*/*S. pombe*) is 0.74 for all two-to-one orthologs,

significantly lower than that (0.94) for one-to-one orthologs. Furthermore, an excess of 31.1% of duplicates experienced expression reductions in both copies (Fig. 13.1c). Interestingly, expression reduction is more pronounced for members of protein complexes than nonmembers. Because an alteration in gene dosage of complex members affect fitness more than that of a nonmember, the above result suggests that the greater expression reduction after duplication in complex members may be promoted by positive selection. These authors also found similar results when comparing human and mouse gene expression data [46]. Thus, empirical evidence supports that expression reduction after gene duplication is likely to be a general mechanism responsible for stable maintenance of functionally redundant duplicate genes.

4 Functional Redundancy in Metabolic Networks

As mentioned, genetic redundancy also exists at the systems level. But systems-level redundancies are more difficult to probe than gene-level redundancies, because genes with no common ancestry or no apparent similarity in molecular function can still be functionally redundant at the systems level. Due to this difficulty, most studies of systems-level redundancies are limited to metabolic networks because they are probably the best characterized among all biological networks and because metabolic networks can be analyzed by a series of systems biology tools such as the flux balance analysis (FBA) [45] and minimization of metabolic adjustment (MOMA) [48]. These tools allow computational predictions of metabolic phenotypes (e.g., biomass production rate and cellular fitness) from the genotype (i.e., the metabolic network) and the environment (i.e., nutrients provided). Most importantly, these computational tools, especially FBA, have been extensively verified experimentally [10–12, 15, 19, 27, 43, 49].

Studies of systems-level redundancies are fewer and more preliminary than those of gene-level redundancies. I thus focus primarily on our own work in the bacterium *Escherichia coli* and yeast *S. cerevisiae* [61]. These species were chosen because their reconstructed metabolic networks are of high quality and have been empirically verified and because they represent prokaryotes and eukaryotes, respectively. The metabolic networks of *E. coli* and *S. cerevisiae* contain 737 and 632 biochemical reactions, respectively, after the removal of dead-end reactions. We focused on biochemical reactions rather than genes encoding the enzymes that catalyze these reactions, for three reasons. First, our interest is at the systems level of a metabolic network, which is composed of reactions. Second, there is no one-to-one relationship between genes and reactions, because a reaction may be catalyzed by a multi-peptide enzyme or several isoenzymes that are encoded by multiple genes. The product of one gene may also be involved in more than one reaction. Third, annotations of enzyme genes are incomplete, making it impossible to conduct a gene-based analysis that is as comprehensive and accurate as a reaction-based analysis.

Our analysis used FBA extensively. Assuming a steady state in metabolism (i.e., no net accumulation of intermediate metabolites), we used FBA to maximize the rate of biomass production under the stoichiometric matrix of all metabolic reactions and a set of flux constraints. The FBA-optimized rate of biomass production can be regarded as the Darwinian fitness of the cell under the condition specified. If removing a reaction blocks the production of one or more biomass components, biomass production becomes zero or undefined due to imbalanced compositional stoichiometry of the biomass. In order to estimate the number (m) of metabolically redundant reactions, we need to identify the reactions whose single removal does not block the production of any biomass component under any nutritional condition. This definition defers from an earlier study in which a reaction is considered redundant if it is not used in one or a few conditions [43]. Because it is infeasible to enumerate all possible conditions, we investigated how the estimate of m changes when the number (c) of examined conditions increases. In *E. coli*, m reduced from 737 to 320 after all single-usable-carbon-source conditions were examined (Fig. 13.2a). We then created random nutritional conditions in which wide-type organisms can grow. Ten thousand conditions appeared to be sufficient for reasonably accurate estimation of m (Fig. 13.2a). Using this method, we identified 276 (37% of the network) and 295 (47% of the network) redundant reactions from *E. coli* and *S. cerevisiae*, respectively (Figs. 13.2 and 13.3).

Nonredundant metabolic reactions can be divided into two classes: always-essential and sometimes-essential. Deletion of an always-essential reaction blocks biomass production under all conditions, whereas deletion of a sometimes-essential reaction blocks biomass production under some but not all conditions. Always-essential reactions can be identified unambiguously, because a metabolic network model allows us to know all nutrients that can be used by the cell under the metabolic model. If a reaction is essential when all these usable nutrients are available, it must be essential when one or more of these nutrients are absent and hence must be an always-essential reaction. The rest of the nonredundant reactions are sometimes-essential reactions. Using this strategy, we identified 95 (13%) always-essential and 366 (50%) sometimes-essential reactions in *E. coli* (Fig. 13.3a), and 24 (4%) always-essential and 313 (49%) sometimes-essential reactions in *S. cerevisiae* (Fig. 13.3b).

While redundant reactions can be individually removed from a metabolic network without blocking biomass production, they may not be simultaneously removed. We designed an algorithm to derive a functional metabolic network with no redundancy [61]. The size of this network varies, depending on the order with which redundant reactions are removed. We randomly generated 250 zero-redundancy networks, which have on average 534 (72% of the original network) and 418 (64%) reactions for *E. coli* and *S. cerevisiae*, respectively. These results further demonstrate the high redundancy of the *E. coli* and *S. cerevisiae* metabolic networks, because as many as 28–36% of reactions can be simultaneously removed from the metabolic networks without blocking the biomass production under any condition.

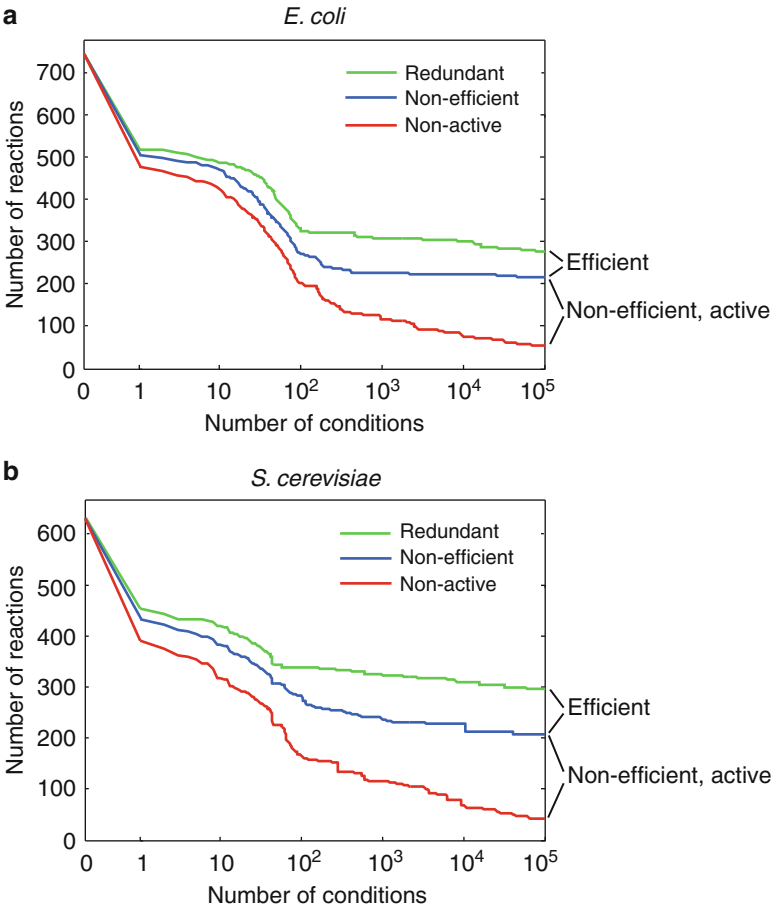

Fig. 13.2 Estimates of the numbers of various redundant reactions in (a) *E. coli* and (b) *S. cerevisiae* stabilize as the number of examined nutritional conditions increases. The first 158 conditions examined in *E. coli* and first 60 conditions examined in *S. cerevisiae* are single-usable-carbon-source conditions, whereas the remaining conditions are randomly generated following a specific sampling scheme. Note that the number of nonactive reactions might be overestimated, because the estimate continues to decline as the number of examined conditions increases. This leads to a conservative estimate of the number of active reactions. Adapted from [61] with permission

5 Mechanisms of Evolutionary Preservation of Redundant Metabolic Reactions

How can a redundant reaction be preserved during evolution? One possibility is that functionally redundant reactions have differential metabolic efficiencies under different conditions, allowing the cell to use different reactions to achieve maximal

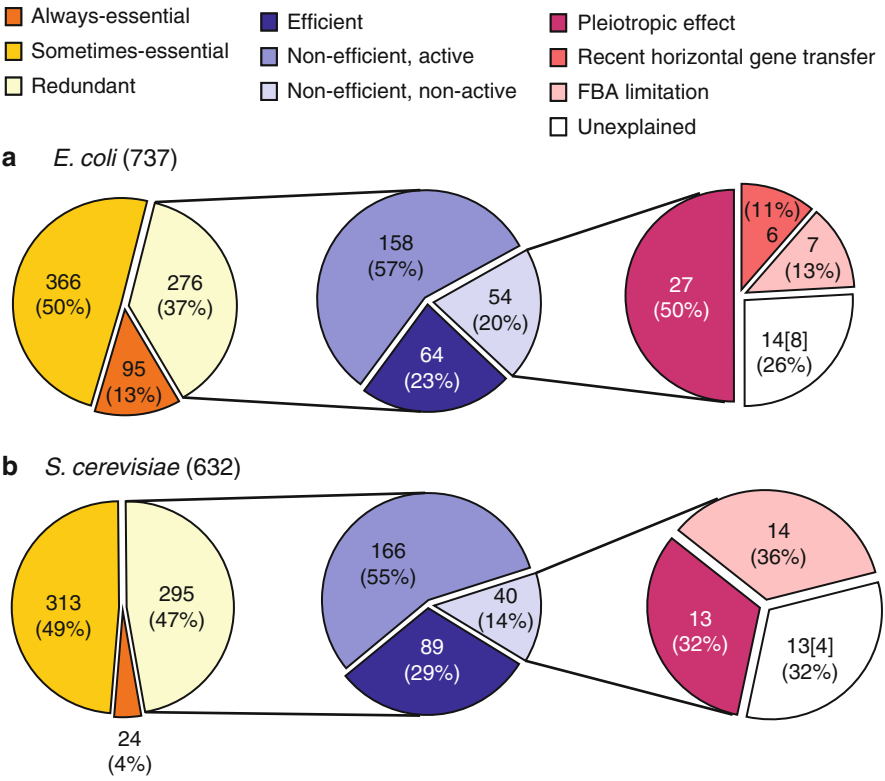

Fig. 13.3 Numbers and fractions of redundant and nonredundant reactions in (**a**) *E. coli* and (**b**) *S. cerevisiae* metabolic networks. For each species, the *middle and right circles* show various explanations for the existence of redundant reactions. Explanations in the *middle circle* are considered before those in the *right circle*; within each circle, explanations depicted with darker colors are considered before those depicted with lighter colors. For each redundant reaction, only the first applicable explanation considered is counted. The total number of reactions after the removal of dead-end reactions is given in the parentheses after the species name. An always-essential reaction is required for growth in all nutritional conditions, whereas a sometimes-essential reaction is required only in some conditions. A redundant reaction is not required in any condition. An efficient (redundant) reaction is more efficient than reactions of the same functions in at least one condition. A non-efficient, active (redundant) reaction is no more efficient than reactions of the same functions under all conditions, but has nonzero flux in at least one condition. Under certain conditions, removing such reactions causes an immediate fitness reduction, which can only be recovered via evolution by mutation, drift, and selection. A non-efficient, nonactive (redundant) reaction can be removed without causing an appreciable fitness reduction in all conditions. Pleiotropic effect refers to the situation where an otherwise dispensable reaction is preserved because its catalyzing enzyme also catalyzes another reaction that is indispensable. Recent horizontal gene transfer refers to the situation where a redundant reaction was acquired by recent horizontal gene transfer and thus may not have been stably preserved in the genome. FBA limitation refers to the situation where an indispensable reaction is misclassified as dispensable due to limitations of FBA. Because the enzyme genes associated with some reactions have yet to be identified, the number of genes known to be associated with the unexplained redundant reactions is given in brackets. Adapted from [61] with permission

growth in many environments. Under this hypothesis, deleting a redundant reaction at a given condition may reduce (but not block) biomass production when the deleted reaction is more efficient than other reactions of the same function at this condition. To understand the feasibility of this model, let us examine the condition necessary for natural selection to maintain a gene that is used infrequently. Let A collectively denote all functional alleles of the gene under study and a collectively denote all null alleles of the gene, and let p and q be the frequencies of A and a alleles, respectively. Let the null mutation rate, or the rate of mutation from A to a, be u per gene per generation. We assume that the mutation rate from a to A is zero because it is extremely unlikely for a null allele to mutate back to a functional one. Random mutations increase the frequency of a, while occasional natural selection reduces it. Let us first consider the possibility of a mutation-selection balance. At the balance, new a alleles generated by mutations are offset by those removed by selection. In haploids, let us assume that the relative fitness of A and a individuals be 1 and 1–s, respectively, and that selection occurs once every n generations. Without loosing generality, let us assume that in every cycle of n generations, selection occurs at the end of the nth generation in the form of a viability difference. Thus, when the balance is reached, in n generations, the allele frequency of a increases from q_0 to q_n by mutation, and then decreases to q_0 by natural selection. The mutational process is described by the difference equation

$$q_n = q_{n-1} + (1 - q_{n-1})u. \tag{13.1}$$

Solving (13.1), we obtained

$$q_n \approx q_0 + (1 - q_0)un. \tag{13.2}$$

In the case of haploid organisms such as *E. coli*, the selection process is described by

$$q'_n = \frac{q_n(1-s)}{(1-q_n)+q_n(1-s)} = \frac{q_n(1-s)}{1-q_n s}, \tag{13.3}$$

where q'_n is the frequency of a after selection. At the mutation-selection balance, we have

$$q'_n = q_0. \tag{13.4}$$

Using (13.2), (13.3), and (13.4), we can obtain

$$q_n = un/s. \tag{13.5}$$

For diploid organisms, the fitness of AA, Aa, and aa individuals are assumed to be 1, 1, and 1–s, respectively, because enzyme genes are largely haplosufficient [9,32]. Then, (13.3) can be rewritten as

$$q'_n = \frac{2p_n q_n + 2q_n^2(1-s)}{2[p_n^2 + 2p_n q_n + q_n^2(1-s)]} = \frac{q_n(1-q_n)+q_n^2(1-s)}{1-q_n^2 s} = \frac{q_n(1-q_n s)}{1-q_n^2 s}. \tag{13.3'}$$

Using (13.2), (13.3′) and (13.4), we obtain

$$q_n = \sqrt{un/s}. \tag{13.5′}$$

Thus, for both haploids and diploids, when $un/s < 1$, null alleles cannot be fixed because of occasional selection for the functional allele. In other words, functional alleles can be preserved in the population. The above mutation-selection equilibrium is a stable equilibrium, because if q is by chance slightly larger than its equilibrium value, the effect of selection in removing null alleles (qs for haploids and q^2s for diploids) becomes larger and the mutation rate per generation in generating null alleles ($(1–q)u$) becomes lower. Consequently, q will return to its equilibrium value. The same argument can be made if q is by chance slightly smaller than its equilibrium value. Thus, random genetic drift cannot push q much away from its equilibrium value. This is particularly so, given the large population size of *E. coli* and *S. cerevisiae*.

Although $un/s < 1$ can ensure that functional alleles at a locus will not be lost in evolution, in practice, one may consider a more stringent criterion of $q_n < 0.5$ so that a randomly sampled allele of the gene from the population is more likely to be functional than null. Thus, we consider that the gene can be retained by selection if $n < 0.5s/u$ for haploids or $n < 0.25s/u$ for diploids. The mean mutation rate u for *E. coli* metabolic enzyme genes is 7.7×10^{-8} per gene per generation [61]. If we use $s = 0.01$, n has to be smaller than 6.5×10^4. If we use $s = 0.1$, n has to be smaller than 6.5×10^5. The mean u for *S. cerevisiae* metabolic enzyme genes is 4.0×10^{-8} per gene per generation [61]. When $s = 0.01$, n has to be smaller than 6.3×10^4. When $s = 0.1$, n has to be smaller than 6.3×10^5. Thus, even a very rarely used gene that is of moderate benefit to the organism when used can be stably kept in the genome.

Given the above theoretical results and potential errors associated with FBA-predicted fitness, we regard a redundant reaction as indispensable if its removal reduces biomass production by more than 1% in one or more of the 10^5 conditions examined. Such indispensable redundant reactions are referred to as efficient reactions, as they are more efficient than other reactions of the same functions under at least one condition. Our analysis identifies 64 and 89 efficient reactions in *E. coli* and *S. cerevisiae*, respectively, accounting for 23–30% of all redundant reactions (Figs. 13.2 and 13.3). The remaining 70–77% of redundant reactions are as efficient as or less efficient than other reactions of the same functions under all conditions and are referred to as non-efficient reactions (Figs. 13.2 and 13.3).

In the above analysis, we assumed that when a redundant reaction is deleted, its compensating reaction is immediately activated to its optimal flux to produce the maximal biomass predicted by FBA. This assumption requires that the cell has a regulatory emergency plan for every possible reaction deletion, which seems unrealistic. In general, the growth performance of a perturbed metabolic network is suboptimal and the FBA-predicted maximal growth can only be achieved through evolution by mutation, drift, and selection [14, 27]. In other words, when a reaction

is deleted from a cell, the cell may be outcompeted by wild-type cells and has no chance to evolve to its FBA-predicted maximal fitness. To consider this possibility, we employed MOMA, a derivative of FBA that has also been empirically verified [48]. Under all the assumptions and constraints used by FBA, MOMA calculates the rate of biomass production after the deletion of a reaction by minimizing flux changes. Because MOMA minimizes flux changes while FBA does not, the biomass production predicted by MOMA is always lower than or equal to that predicted by FBA. A non-efficient reaction is considered to be indispensable if its removal reduces the MOMA-predicted biomass production by more than 1% in one or more of the 10^5 examined conditions. Such reactions are referred to as active reactions because they must have nonzero fluxes; otherwise their removal will not cause biomass reduction. We identified 158 and 166 active reactions in *E. coli* and *S. cerevisiae*, respectively, accounting for more than half of all redundant reactions or 75–80% of non-efficient redundant reactions (Fig. 13.3). The rest of non-efficient reactions are referred to as nonactive reactions because their removal does not affect MOMA-predicted biomass appreciably.

Although we showed how a non-efficient redundant reaction can be indispensable and kept in the network by natural selection, it is puzzling as why such reactions were incorporated into the metabolic network in the first place, as non-efficient reactions are never more efficient than other reactions of the same functions. We suggest that non-efficient reactions were incorporated by neutral processes. They became active reactions if they were equally efficient as their redundant reactions under some conditions. When multiple equally efficient redundant reactions exist (regulatory or structural), degenerate mutations may be fixed so that the total activity of the enzymes catalyzing the redundant reactions is optimized while the activity of each enzyme becomes insufficient for the maximal growth should the other redundant enzymes be removed.

Our analysis identified 54 (7% of the total network) and 40 (7%) nonactive redundant reactions in *E. coli* and *S. cerevisiae*, respectively (Fig. 13.3). Among them, 38 *E. coli* and 20 *S. cerevisiae* reactions are less efficient than other reactions of the same functions and have zero fluxes under all conditions. The rest may be as efficient as their redundant reactions and have nonzero fluxes, but their removal does not reduce MOMA-predicted biomass production by more than 1%.

How are the nonactive reactions maintained in the metabolic network? Some enzymes can catalyze multiple reactions, a phenomenon known as pleiotropy [60]. In *E. coli*, 266 reactions (36% of the total network), including 27 nonactive reactions, are catalyzed by pleiotropic enzymes. In *S. cerevisiae*, 171 reactions (27% of the total network), including 13 nonactive reactions, are catalyzed by pleiotropic enzymes. A nonactive reaction can be stably retained in the network if the enzyme that catalyzes it also catalyzes one or more indispensable reactions. Indeed, we find that every nonactive reaction catalyzed by pleiotropic enzymes can be retained by this "guilt-by-association" mechanism. In both *E. coli* and *S. cerevisiae*, there are only 27 redundant reactions whose retentions are unexplained (Fig. 13.3). Further investigations show that they are unexplained by FBA and MOMA simply because of the incompleteness of the reconstructed metabolic networks, limitations of the

metabolic models (e.g., lack of connection to regulatory and signal transduction networks), and existence of environments difficult to simulate (e.g., temperature changes). For instance, *E. coli* gene *otsB* encodes trehalose-6-phosphate phosphatase, which is required for cell viability at 4°C [30] and thus may be maintained by selection if *E. coli* sometimes experiences this low temperature in nature. We also observed six *E. coli* nonactive reactions that are catalyzed by enzymes encoded by genes that were recently horizontally transferred into *E. coli*. Horizontal gene transfers occur so frequently among prokaryotes [16] that the presence of some redundant genes may be attributable to this mechanism rather than preservation under purifying selection. Indeed, analyzing an *E. coli* horizontal-gene-transfer dataset [34], we find that the fraction of recently horizontally acquired genes is significantly greater among nonactive reactions (43%) than among other reactions (19%). After considering all these additional mechanisms, there are only 14 (8 with associated genes) *E. coli* and 13 (4 with associated genes) *S. cerevisiae* redundant reactions whose preservation in the metabolic networks remain unexplained (Fig. 13.3).

Our analysis showed that the vast majority of the functionally redundant reactions in *E. coli* and *S. cerevisiae* are selectively maintained because they reduce fitness when singly removed from the cell. In other words, the backup hypothesis is not needed for explaining the existence of systems-level redundancies either. Our formulation in Sect. 3 shows that the condition for the backup hypothesis is $N_e su > 1$, which predicts that important functions (i.e., with larger s) are more likely than unimportant ones to have backups should they exist. To test this prediction of the backup hypothesis, we measured the importance of reactions using zero-redundancy networks, because they are free from the confounding influence of redundant reactions. We calculated the average biomass reduction upon removal of a reaction from a zero-redundancy network across 10^3 conditions and repeated this calculation in 125 random zero-redundancy networks to obtain the mean. For *E. coli*, contrary to the prediction of the backup hypothesis, reactions that are redundant in the original metabolic network tend to perform less important jobs than reactions that are nonredundant (Fig. 13.4). But for *S. cerevisiae*, the observation appears to be consistent with the backup prediction (Fig. 13.4). These opposite patterns in *E. coli* and *S. cerevisiae* suggest that the backup hypothesis is either inadequate or wrong.

What processes can explain the opposite relations between the importance of a reaction and its degree of redundancy in the two species examined? We conducted a computer simulation to examine the impact of environmental changes [61]. We first created a random nutritional condition. A zero-redundancy metabolic network for this condition was then generated by removing redundant reactions from the original network. We repeated this process 10^3 times, each under a different condition. We then merged the 10^3 resultant zero-redundancy networks to form the final simulated metabolic network. We measured the relative importance of redundant and nonredundant reactions of this simulated network as was done for the real network. Interestingly, for both *E. coli* and *S. cerevisiae*, the results are similar between the simulated networks and their respective real networks (Fig. 13.4). Because we did not invoke selection for backup in the simulation, our result strongly suggests that the observation of higher redundancy for more important functions in

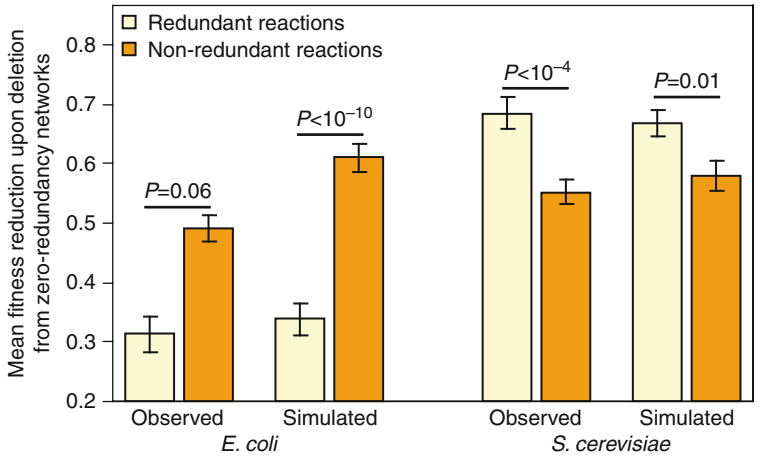

Fig. 13.4 Relationships between the importance and redundancy of metabolic reactions. *Error bars* show one standard error. *P*-values are from Mann–Whitney *U* test. The redundancy of a reaction is determined from the complete network, whereas the importance of a reaction is determined from zero-redundancy networks. Redundant reactions perform less important functions than nonredundant functions in *E. coli*, whereas the opposite is true in *S. cerevisiae*. The same patterns are recapitulated in simulated metabolic networks that are formed by merging 10^3 zero-redundancy networks that each functions in a different condition. Adapted from [61] with permission

S. cerevisiae is a byproduct of its evolutionary history. This simulation was repeated ten additional times and the above finding always holds. Taken together, natural selection arising from fluctuating environments, rather than the backup hypothesis, explains the relationship between the importance of a reaction and its degree of redundancy in both *E. coli* and *S. cerevisiae*. A recent simulation study of metabolic redundancy reaches the same conclusion [50].

6 Conclusions

Genomic scale mutagenesis studies conducted in multiple model organisms showed that many duplicate genes are functionally redundant to a certain degree. These redundancies, however, are not maintained because they could backup important cellular functions that may by chance be damaged by random mutations in the future. Rather, they are preserved either because of constraints from other nonredundant functions as explained by the piggyback hypothesis or because they improve fitness through a nonlinear fashion as explained by the model of expression reduction after gene duplication. Other models of redundancy maintenance exist [42, 56], but they appear to require more restricted conditions. For example, it has been shown mathematically that two genes sharing the same function can be stably

retained if the two genes have different efficiencies in performing the function and if the one with lower efficiency also has a lower deleterious mutation rate [42]. It will be of significant interest to investigate the relative importance of various mechanisms in preserving functional redundancies between duplicates.

The analysis of the *E. coli* and yeast metabolic networks also showed clearly that systems-level redundancies are abundant and that these redundancies need not and cannot be explained by the backup hypothesis. Rather, the vast majority of them are preserved because (1) the functionally redundant reactions have differential maximal efficiencies at different conditions, (2) their loss causes a fitness reduction that can only be recovered via evolution, or (3) they have indispensable pleiotropic functions. A small fraction of redundant reactions in *E. coli* were recently acquired from other species via horizontal gene transfers and they may be lost in the near future.

While I focused on functional redundancy between duplicate genes and among metabolic reactions in this chapter, the approaches and methodologies developed may be applied to other systems. An emerging feature, which may be obvious to evolutionary biologists but not so to systems biologists, is the important role that historical contingency plays in constraining and channeling the evolution of a system, although many events that affect subsequent evolution occur by chance. The importance of historical contingency is reflected in many properties of the genetic redundancies studied here, and I believe it will also be evident in other properties of biological systems.

It is clear that systems analyses such as FBA and MOMA provide powerful tools for dissecting complex phenomena in genetics and evolution. Similarly, evolutionary analyses such as theoretical population genetics and molecular evolutionary genetics offer invaluable insights into the functions and origins of systemic properties that may seem puzzling at first. Evolutionary systems biology, with concepts and tools of two powerful fields, holds great promises for uncovering many mysteries in biology.

Acknowledgments I thank my former and current students, especially Xionglei He, Ben-Yang Liao, Zhi Wang, and Wenfeng Qian, for their contributions to several studies summarized here. Wenfeng Qian also assisted in the preparation of Fig. 13.1 and Table 13.1. Research in my laboratory has been supported by the U.S. National Institutes of Health.

References

1. Austin CP, Battey JF, Bradley A, et al. (2004) The knockout mouse project. Nat Genet 36: 921–924
2. Chen S, Zhang YE, Long M (2010) New genes in Drosophila quickly become essential. Science 330:1682–1685
3. Conant GC, Wagner A (2004) Duplicate genes and robustness to transient gene knock-downs in Caenorhabditis elegans. Proc Biol Sci 271:89–96

4. Cooke J, Nowak MA, Boerlijst M, Maynard-Smith J (1997) Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet 13:360–364
5. de Visser JA, Hermisson J, Wagner GP, et al. (2003) Perspective: evolution and detection of genetic robustness. Evolution 57:1959–1972
6. Dean EJ, Davis JC, Davis RW, Petrov DA (2008) Pervasive and persistent redundancy among duplicated genes in yeast. PLoS Genet 4:e1000113
7. DeLuna A, Springer M, Kirschner MW, Kishony R (2010) Need-based up-regulation of protein levels in response to deletion of their duplicate genes. PLoS Biol 8:e1000347
8. DeLuna A, Vetsigian K, Shoresh N, Hegreness M, Colon-Gonzalez M, Chao S, Kishony R (2008) Exposing the fitness contribution of duplicated genes. Nat Genet 40:676–681
9. Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, Nislow C, Giaever G (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics 169:1915–1925
10. Duarte NC, Herrgard MJ, Palsson BO (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 14:1298–1309
11. Edwards JS, Ibarra RU, Palsson BO (2001) In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 19:125–130
12. Famili I, Forster J, Nielsen J, Palsson BO (2003) Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci USA 100:13134–13139
13. Feldman I, Rzhetsky A, Vitkup D (2008) Network properties of genes harboring inherited disease mutations. Proc Natl Acad Sci USA 105:4323–4328
14. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO (2005) In silico design and adaptive evolution of Escherichia coli for production of lactic acid. Biotechnol Bioeng 91:643–648
15. Fong SS, Palsson BO (2004) Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. Nat Genet 36:1056–1058
16. Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. Mol Biol Evol 19:2226–2238
17. Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH (2003) Role of duplicate genes in genetic robustness against null mutations. Nature 421:63–66
18. Hanada K, Kuromori T, Myouga F, Toyoda T, Li WH, Shinozaki K (2009) Evolutionary persistence of functional compensation by duplicate genes in Arabidopsis. Genome Biol Evol 1:409–414
19. Harrison R, Papp B, Pal C, Oliver SG, Delneri D (2007) Plasticity of genetic interactions in metabolic networks of yeast. Proc Natl Acad Sci USA 104:2307–2312
20. Hartman JL, Garvik B, Hartwell L (2001) Principles for the buffering of genetic variation. Science 291:1001–1004
21. He X, Qian W, Wang Z, Li Y, Zhang J (2010) Prevalent positive epistasis in Escherichia coli and Saccharomyces cerevisiae metabolic networks. Nat Genet 42:272–276
22. He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157–1164
23. He X, Zhang J (2006a) Higher duplicability of less important genes in yeast genomes. Mol Biol Evol 23:144–151
24. He X, Zhang J (2006b) Transcriptional reprogramming and backup between duplicate genes: is it a genome-wide phenomenon? Genetics 172:1363–1367
25. He X, Zhang J (2006c) Why do hubs tend to be essential in protein networks? PLoS Genet 2:e88
26. Hsiao TL, Vitkup D (2008) Role of duplicate genes in robustness against deleterious human mutations. PLoS Genet 4:e1000014
27. Ibarra RU, Edwards JS, Palsson BO (2002) Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420:186–189

28. Joyner AL, Herrup K, Auerbach BA, Davis CA, Rossant J (1991) Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the En-2 homeobox. Science 251:1239–1243
29. Kafri R, Bar-Even A, Pilpel Y (2005) Transcription control reprogramming in genetic backup circuits. Nat Genet 37:295–299
30. Kandror O, DeLeon A, Goldberg AL (2002) Trehalose synthesis is induced upon exposure of Escherichia coli to cold and is essential for viability at low temperatures. Proc Natl Acad Sci U S A 99:9727–9732
31. Kim DU, Hayles J, Kim D, et al. (2010) Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol 28:617–623
32. Kondrashov FA, Koonin EV (2004) A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. Trends Genet 20:287–290
33. Lenski RE, Barrick JE, Ofria C (2006) Balancing robustness and evolvability. PLoS Biol 4:e428
34. Lercher MJ, Pal C (2008) Integration of horizontally transferred genes into regulatory interaction networks takes many million years. Mol Biol Evol 25:559–567
35. Liang H, Li WH (2007) Gene essentiality, gene duplicability and protein connectivity in human and mouse. Trends Genet 23:375–378
36. Liang H, Li WH (2009) Functional compensation by duplicated genes in mouse. Trends Genet 25:441–442
37. Liao BY, Scott NM, Zhang J (2006) Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. Mol Biol Evol 23:2072–2080
38. Liao BY, Zhang J (2007) Mouse duplicate genes are as essential as singletons. Trends Genet 23:378–381
39. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155
40. Makino T, Hokamp K, McLysaght A (2009) The complex relationship of gene duplication and essentiality. Trends Genet 25:152–155
41. Musso G, Costanzo M, Huangfu M, et al. (2008) The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. Genome Res 18:1092–1099
42. Nowak MA, Boerlijst MC, Cooke J, Smith JM (1997) Evolution of genetic redundancy. Nature 388:167–171
43. Papp B, Pal C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194–197
44. Parenteau J, Durand M, Morin G, Gagnon J, Lucier JF, Wellinger RJ, Chabot B, Abou Elela S (2011) Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. Cell 147:320–331
45. Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2:886–897
46. Qian W, Liao BY, Chang AY, Zhang J (2010) Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends Genet 26:425–430
47. Saga Y, Yagi T, Ikawa Y, Sakakura T, Aizawa S (1992) Mice develop normally without tenascin. Genes Dev 6:1821–1831
48. Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 99:15112–15117
49. Snitkin ES, Dudley AM, Janse DM, Wong K, Church GM, Segre D (2008) Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. Genome Biol 9:R140
50. Soyer OS, Pfeiffer T (2010) Evolution under fluctuating environments explains observed robustness in metabolic networks. PLoS Comput Biol 6:e1000907
51. Tautz D (1992) Redundancies, development and the flow of information. Bioessays 14: 263–266
52. Thomas JH (1993) Thinking about genetic redundancy. Trends Genet 9:395–399

53. Tischler J, Lehner B, Chen N, Fraser AG (2006) Combinatorial RNA interference in Caenorhabditis elegans reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. Genome Biol 7:R69
54. Vavouri T, Semple JI, Lehner B (2008) Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. Trends Genet 24:485–488
55. Wagner A (2000a) Robustness against mutations in genetic networks of yeast. Nat Genet 24:355–361
56. Wagner A (2000b) The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. Genetics 154:1389–1401
57. Wagner A (2005a) Distributed robustness versus redundancy as causes of mutational robustness. Bioessays 27:176–188
58. Wagner A (2005b) Energy constraints on the evolution of gene expression. Mol Biol Evol 22:1365–1374
59. Wagner A (2005c) Robustness and Evolvability in Living Systems. Princeton University Press, Princeton, NJ
60. Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet 12:204–213
61. Wang Z, Zhang J (2009) Abundant indispensable redundancies in cellular metabolic networks. Genome Biol Evol 1:23–33
62. Zhang J (2003) Evolution by gene duplication: an update. Trends Eco Evol 18:292–298