Distributed Robustness Versus Redundancy as Causes of Mutational Robustness

Andreas Wagner

SANTA FE INSTITUTE

Distributed robustness versus redundancy as causes of mutational robustness.

Andreas Wagner

Department of Biology
University of New Mexico
167A Castetter Hall
Albuquerque, NM 87131-1091
Phone: (505)- 277-2021
FAX: (505)- 277-0304
Email: wagnera@unm.edu

Abstract

A biological system is robust to mutations if it continues to function after genetic changes in its parts. Such robustness is pervasive on different levels of biological organization, from macromolecules to genetic networks and whole organisms. I here ask which of two possible causes of such robustness are more important on a genome-wide scale, for systems whose parts are genes, such as metabolic and genetic networks. The first of the two causes is redundancy of a system's parts: A gene may be dispensable if the genome contains redundant, back-up copies of the gene. The second cause, distributed robustness, is more poorly understood. It emerges from the distributed nature of many biological systems, where many (and different) parts contribute to system functions. I will here discuss evidence suggesting that distributed robustness is equally or more important for mutational robustness than gene redundancy. This evidence comes from large-scale gene deletion studies, as well as from the functional divergence of redundant genes. I also ask whether one can quantify the extent to which redundancy or distributed robustness contributes to mutational robustness.

Living things are unimaginably complex, yet also highly robust to genetic change on all levels of organization. Proteins can tolerate thousands of amino acid changes, metabolic networks can continue to sustain life even after removal of important chemical reactions, gene regulation networks continue to function after alteration of key gene interactions, and radical genetic change in embryonic development can lead to an essentially unchanged adult organism [1-4]. The mutational robustness of organisms is an intriguing phenomenon, not only because complexity suggests fragility. (If you have ever built a house of cards, you will know what I mean.) Mutational robustness also affects an organism's ability for future evolution, because it may increase the rate of neutral mutations, and thus the amount of genetic variation invisible to natural selection. Here, I will explore a very basic question about mutational robustness: What are the mechanistic causes of mutational robustness, and which of these causes is most important? I will begin by illustrating the two principal causes of mutational robustness, redundancy and distributed robustness. Most of my examples regard systems above the gene level. However much of the following is equally relevant for systems on lower levels of biological organization, especially macromolecules like protein and RNA.

Redundancy and gene redundancy. Biologists use the term redundancy in more than one way. One usage invokes redundancy if a gene's activity can be changed or a system's part can be removed without affecting key system properties [5]. Another usage refers to redundancy only if two parts of a system perform the same or similar tasks. The two notions are not synonymous. A system may be unaffected by removing a part, yet no two of its parts may have identical functions. Examples include proteins, whose tertiary structure is the result of cooperative folding involving many amino acids. No two amino acids may play identical roles in the folding process, yet protein structure can be highly robust to changes in individual amino acids. I will use redundancy here strictly in the second sense – redundancy of parts. This usage is not only consistent with one of the main dictionary definitions of the word, namely "unnecessary repetition" [6]. It is also consistent with the usage of engineers concerned with designing reliable systems [7]. In engineering, redundancy is the main pillar of system reliability, partly because it is conceptually straightforward to build systems with redundant parts. But how important is redundancy for robustness in biological systems? Less important than one might think, as I will argue here.

Genes encode the most important parts of biological systems above the gene level – proteins. The notion that redundancy of genes might be important for robustness emerged with the ability to generate gene deletion, synthetic null, or gene knock-out strains of various organisms, that is, with the ability to eliminate specific genes from a genome. Many such gene deletion experiments yielded surprises. Specifically, they revealed that genes thought to be involved in key biological processes could often be removed without affecting the organism. This phenomenon exists in all organisms and all kinds of genes, be they enzymatic, structural, or regulatory [8-20]. The phenomenon is too pervasive to survey exhaustively, but I will discuss a few examples. The *CLN1* gene of budding yeast regulates the activity of the yeast cyclin dependent kinase Cdc28p [21]. Proper regulation of Cdc28p is required for the transition from the G_1-phase to the S-phase of the cell cycle. However, a null mutant in *CLN1* grows and divides normally on a minimal growth medium [22]. The *BarH2* gene from the fruit fly *Drosophila melanogaster*

encodes a transcription factor with a homeobox DNA binding domain. It is involved in sensory organ development, yet its deletion does not cause morphological defects in sensory organs [23]. The extracellular matrix protein tenascin of mice is thought to play an important role in morphogenesis during embryonic development. Yet the deletion of a gene encoding this protein leads to no detectable developmental abnormalities, to no change in the distribution of other extracellular matrix proteins, and to no detectable reduction of fertility [24]. The yeast *HMG2* gene encodes a key enzyme in the sterol biosynthesis pathway, 3-hydroxy-3-methylglutaryl-coenzyme A reductase. The end products of this pathway take part in processes ranging from electron transport to DNA replication. Yet the *HMG2* gene can be eliminated without affecting cell growth on either rich or minimal growth medium [25]. The *Drosophila* gene *knirps* encoding a key transcriptional regulator in embryonic development can be eliminated without affecting head development, during which *knirps* is expressed [20]. And finally, there are the three *TPK* genes in yeast, which encode catalytic subunits of the yeast cyclic AMP dependent protein kinase, a key molecule in cell signaling. Any two of these three genes are dispensable for cell growth [26].

All of these examples are cases of gene redundancy resulting from gene duplication. The three *TPK* genes form a family of genes that originated from two duplications of a single ancestral gene [26]. Similarly, the *CLN1* gene has two duplicates, *CLN2* and *CLN3*, with similar functions; the *BarH2* gene has a duplicate, *BarH1*; the *HMG2* gene has a duplicate *HMG1*; and *knirps* has a duplicate, *knirps-related* [20,22,23,25]. Eliminating *all* duplicates of a gene often reveals the severe phenotypic effects that mutations in the gene lack, as expected if redundancy is the cause for weak gene knock-out effects. Redundant duplicate genes with similar functions can thus help explain why knock-out mutations in important genes have no phenotypic effects.

Gene duplications are abundant by-products of recombination and DNA repair processes in most genomes. Typically, between 25 and 50 percent of eukaryotic genes have at least one duplicate within the same genome [27,28]. Immediately after a gene duplication, many pairs of duplicate genes produce the same two polypeptides with the same function. Subsequently, mutations that accumulate in either gene can lead to elimination of one duplicate, or to divergence in the duplicate's functions [26]. Although such divergence is often very limited, as the examples above show, most gene duplicates do not have completely identical functions. That is, they often encode multifunctional proteins which share some functions while they differ in others. For example, the *knirps* gene is dispensable for head development, but it is necessary for abdominal development [20]. Mig1p and Mig2p, two yeast transcriptional regulators, act redundantly in repressing the yeast gene *SUC2*, needed for metabolizing sucrose, but Mig1p is uniquely required for repressing the *GAL* genes necessary for galactose metabolism [18]. Similarly, the three *TPK* genes, any two of which are normally dispensable for cell division, have different functions under conditions where yeast forms filamentous cell aggregates called pseudohyphae. *TPK2* is an essential gene under these conditions [29]. Such observations have led to the notion that duplicate genes can have partially redundant or overlapping functions.

Distributed robustness. In distributed robustness, many parts of a system contribute to system function, but all of these parts have different roles. When one part fails or is changed through mutations, the system can compensate for this failure, but not

because a "back-up" redundant part takes over the failed part's role. Distributed robustness is a fairly poorly understood cause of mutational robustness, because it requires a detailed, quantitative understanding of the inner workings of a genetic network. I will illustrate it with two examples, one from metabolic networks, the other one from a gene regulation network.

Metabolic networks include hundreds of chemical reactions that are required to generate energy and biosynthetic building blocks for cell growth and proliferation. A variety of computational methods such as metabolic flux balance analysis can predict the maximally possible cell growth rate a metabolic network can sustain in a given nutritional environment, given stoichiometric information about all the chemical reactions that occur in the network [30-32]. The predictions of these approaches are mostly in good agreement with experimental results. Where they are not, laboratory evolution can create organismal strains that behave as predicted. A case in point is the *E. coli* strain MG1655, which grows on glycerol as sole carbon source at a rate lower than predicted as optimal by flux balance analysis. However, within 40 days or 700 generations of evolution in the laboratory, the strain increases growth substantially, and the growth rate approaches that predicted by flux balance analysis as maximal [33].

Metabolic flux balance analysis and related approaches can be used to ask how robust a metabolic network's output (e.g., cell growth) is to the elimination of individual chemical reactions in the network. Such elimination corresponds to the elimination of enzyme-coding genes from a genome, and its effects are thus a good indicator of mutational robustness in metabolic genes. Several studies have systematically examined how elimination of individual reactions would affect a metabolic network [2,34-37]. To give but one example, Edwards and Palsson assembled a reaction network comprising 436 metabolites and 736 reactions from the biochemical literature, genome sequence information, and metabolic databases [35] of the bacterium *Escherichia coli*. They assessed the effects on cell growth of eliminating each of 48 central chemical reactions in glycolysis, the pentose phosphate pathway, and the tricarboxylic acid cycle. Only seven of the 48 eliminated reactions are essential to cell growth. Of the remaining 41 nonessential reactions, 32 reduced cell growth by less than 5%, and only 9 reduced growth by more than 5%. The experimental data and computational predictions on growth/no-growth coincided in 86% of the cases. Where discrepancies occur, they may be due to the fact that wild-type *E. coli* cells have not evolved under the unusual genetic conditions of a missing enzyme-coding gene, and thus that the performance of a metabolic network is not optimal in this situation [36].

It bears emphasizing that there is no redundancy in such metabolic networks, in the sense that different enzymes generally catalyze different chemical reactions. How then, can such networks be robust to elimination of individual reactions? The answer is that they shuttle the flow of matter through parts of a metabolic network that are not affected by blocking a reaction. I will illustrate this with a concrete example, that of the enzyme glucose 6-phosphate dehydrogenase – part of the pentose-phosphate shunt – encoded by the *zwf* gene of *E. coli*. Experimental work shows that loss-of-function mutants in the E. coli gene *zwf* encoding this enzyme grow at near wild-type levels [38], and flux balance analysis predicts a similarly small growth rate reduction [34]. The elimination of this reaction has profound systemic consequences on the metabolic network. For example, before the deletion, in the wild-type state, about two-thirds of a

cell's NADPH is produced by the pentose phosphate shunt. Most of the NADH is produced by the tricarboxylic acid cycle. One of the major systemic reorganizations of metabolic flow after the mutation regards NADPH production. To compensate for the blocked oxidative branch of the pentose phosphate shunt, most of the NADPH is now produced through an increased flux in the tricarboxylic acid cycle, which generates NADH. This NADH is then transformed into NADPH via a massively increased flux through the transhydrogenase reaction. (The nonoxidative branch of the pentose phosphate shunt can still be entered through other metabolites of glycolysis, and thus still serves its other important function, to produce various biosynthetic precursors.)

In sum, the example illustrates a hallmark of distributed robustness: a systemic adjustment of network properties to compensate for a perturbed network part.

Gene regulation networks provide completely different examples of distributed robustness [3,39-42]. I will just discuss one such example, that of the network sustaining the expression of the segment polarity genes *engrailed* and *wingless* in *Drosophila melanogaster* [3]. Proper expression of the genes *wingless* and *engrailed* in 14 stripes is necessary for segmentation of the fly. The expression pattern of these genes is sustained by a network of multiple genes and their products, including *hedgehog, patched*, and *cubitus interruptus*. The properties of this network are determined by multiple biochemical parameters. These parameters include the binding constants of transcriptional regulators such as engrailed to their target gene's regulatory regions; the rate at which a ligand such as Hedgehog associates with its receptor; the stability of mRNAs and proteins expressed from network genes, and so on. Von Dassow and collaborators assembled a quantitative model of this network, a model that is grounded in empirical observations, and that correctly predicts many aspects of network function. They asked how sensitive the network is to changes in the biochemical parameters (48 in total) needed to describe its properties. The answer is that most of these parameters can be changed by more than one order of magnitude and the network still produces the *engrailed* and *wingless* gene expression patterns required for proper development. Moreover, even if one chooses random values (within a biologically sensible range) of all of the 48 biochemical parameters, there is a high probability that the network will function properly. Finally, the network is also highly robust to gene duplications, which can drastically change the expression levels of network proteins.

Again, there is no redundancy involved in this representation of the segment polarity network. Every gene has a different role in the network, yet the network is highly robust to changes of its regulatory interactions. Other studies hint at similar phenomena in different gene regulation networks [39-42]. The source of such robustness lies again in the distributed architecture of the network, and cannot be pinned down to any one gene. However, in many gene regulation networks gene expression and activity of gene products are cooperatively regulated by one or more proteins [43-48], an observation that may go a long way towards explaining this phenomenon [49].

These examples illustrate that a quantitative representation of a whole metabolic or gene regulation network – not local information on only a small group of genes – is necessary to study distributed robustness. We will thus understand this phenomenon only to the extent that high-quality quantitative models of such systems based on experimental data will become available.

Candidate examples for distributed robustness also exist on different levels of biological organization. They include the developmental pathways responsible for the embryonic development of many phenotypic characters. The genes acting in such pathways harbor much variation that may affect gene products but does not affect the phenotype. Such cryptic variation is an indicator of mutational robustness. Are redundant gene functions responsible for it? Probably not. If they were, then mutations that reduce this robustness would preferably affect genes with redundant duplicates. However, the best-explored mutations that reduce robustness are completely different in nature. In fruit flies, they affect the non-redundant gene encoding the heat-shock protein Hsp90, for which heterozygous loss-of-function mutations abolish robustness in a wide spectrum of developmental pathways [50]. The reason is that Hsp90 interacts with more than 80 different proteins important in development, and helps stabilize their conformation. On lower levels of biological organization, the paradigmatic examples of distributed robustness are biological macromolecules, such as proteins and RNA. Their folding in space involves the cooperation of hundreds of building blocks – nucleotides or amino acids – each of which forms unique (non-redundant) interactions with other building blocks. If a change in one amino acid leaves a protein's three-dimensional conformation unchanged, is this because the protein contained a redundant 'back-up' amino acid? Clearly not: The thermodynamic stability of any macromolecular structure is the sum of all contributions that such interactions make to the structure's free energy.

Large-scale gene perturbation studies. Examples, however many, of individual redundant genes or genetic networks with distributed robustness, are mere anecdotal evidence for either cause of robustness. They do not answer the question how prevalent gene redundancy or distributed robustness is on a genome-wide scale. I will now turn to this question. To answer it, systematic, genome-wide gene perturbation studies that eliminate many genes in a genome and systematically assess the phenotypic effect of these deletions are very helpful. The first such large-scale study created some 300 different strains of budding yeast, each of which lacked one of the genes on chromosome V [22]. This study was followed by larger scale studies in *Saccharomyces cerevisiae*, the worm *Caenorhabditis elegans*, and the plant *Arabidopsis thaliana* [51-53], studies that mutated the DNA or reduced the expression levels of thousands of genes. A caveat to comparing the results of such studies is that their perturbation approaches differ. For instance, although deletion of specific genes through homologous recombination is highly effective in yeast, it is not feasible on a large scale in higher organisms such as *C. elegans*. In such organisms temporary gene activation through interference with the expression of messenger RNA is the method of choice [52].

The first surprise of systematic gene knock-out studies is the sheer number of genes with no detectable effect. For instance, the systematic knock-out of most chromosome V genes in yeast shows that almost 40% of genes have no detectable effect on key fitness indicators such as cell division rate and sporulation efficiency in five different nutritional environments [22]. Later studies that eliminated more than 90 percent of all yeast genes yielded similar results [51,54]. Such striking numbers raise the question whether genes with no apparent effect are truly dispensable under all circumstances. Although no laboratory experiment could answer this question definitively, the likely answer is no. It is easy to conceive of genes with specific functions, such as metabolic enzymes, where a mutation may have no effect in one environment but may be fatal in

another environment. Even very rare environmental conditions, conditions that may occur only once every few thousand generations can cause a gene to be maintained in a population [55]. Also, if genes whose elimination has no effects in one or a few environments were truly dispensable, they should evolve much more rapidly than indispensable genes, because they would be able to tolerate fewer mutations. However, the statistical association between a gene's dispensability and its rate of evolution is weak at best [56-60].

Genome-scale gene deletion studies and redundancy. Even though their neutrality may only be temporary, the sheer number of genes – thousands – with no deletion effect still needs a mechanistic explanation. Does gene redundancy provide this explanation? The results of genome-scale gene deletion studies reveal the next surprise in this regard: Thousands of genes whose deletion has no detectable effect are single copy genes. They have no duplicates in the genome.

The results of an early pertinent study are summarized in Figure 1 [61]. In this figure, I grouped gene deletion mutations of yeast chromosome V into two categories, those with a non-detectable phenotypic defect, and those with a severe effect leading to a growth rate reduction of more than 25% relative to the wild type. The figure's vertical axis shows the proportion of genes in each defect category that have the number of duplicates indicated on the horizontal axis. Significantly, more than 40% of duplicate genes with weak or no fitness effect are single copy genes. A more recent study by Gu and collaborators that used more than 4,700 growth rate measurements of yeast gene deletion strains reached a qualitatively similar result. 39.6% of yeast genes with a weak or no fitness defect have no duplicate [62]. Additionally, even if a gene has a duplicate, the duplicate may have diverged in function and may not be responsible for a weak gene deletion effect. The situation in many-celled eukaryotes might be fundamentally different, because they contain more duplicates which occur in larger families [27,28,63]. However, a recent study using the results of RNA interference experiments that temporarily silenced more than 16,000 worm genes suggests otherwise. In this study, Gavin Conant and I found that more than 7,500 single copy genes show no detectable phenotypic effect when temporarily silenced, versus 4515 of genes with one or more duplicate [64].

Does all this mean that gene redundancy is unimportant in determining the effect of gene deletions? No. For genes with duplicates, the likelihood that a gene has a lethal effect is lower than for single copy genes [62,64]. Moreover, the proportion of genes with weak or no effects is higher for duplicate genes than for single-copy genes [62,64]. Third, the more similar two duplicates are, the less severe is the effect of deleting or silencing one of them [62,64]. Fourth, the larger the number of duplicates a yeast gene has, the faster the gene evolves, implying that it can tolerate more mutations [27]. And finally, pairs of duplicate genes have more similar phenotypic effects upon deletion than randomly chosen pairs of unrelated genes [62]. Thus, gene duplications clearly play a role in determining gene deletion effects.

Rapid functional divergence of duplicate genes. The above results suggest that both distributed robustness and gene redundancy play a role in mutational robustness. However, they leave unanswered the key quantitative question: what proportion of weak gene perturbation effects are due to gene duplication?. An estimate of this proportion based on large-scale gene deletion experiments suggests that in yeast only between 23%

and 59% of weak gene perturbation effects are due to gene duplication [62]. In a higher organism, the worm *C. elegans*, this proportion may be even lower, between 3% and 36% [64]. Both intervals show that gene redundancy is clearly not responsible for the vast majority of weak gene perturbation effects, although the numbers in the worm have to be taken with a grain of salt, because they are based on results of RNA interference assays. The true percentage of weak yeast gene deletion effects would be close to the upper bound of the above interval (59%) if most duplicates retained similar functions long after their duplication. It would be close to the lower bound (23%) if most gene duplicates diverged rapidly in function. In the following section, I will argue that most duplicate genes diverge rapidly in important indicators of gene function. This suggests that the true proportion of weak gene deletion effects due to gene duplication is nowhere near the hypothetical upper bound.

Estimating the rate at which gene duplicates diverge in function faces two technical problems. The first of them is to measure the time that elapsed since a gene duplication. The crudest indicator of this time is the number of nucleotide changes that lead to amino acid changes in either of the duplicates' gene product. However, because natural selection can strongly affect amino acid changes, their rate of accumulation can vary greatly over time. The problem can be alleviated by estimating instead the number of synonymous (silent) nucleotide substitutions that occurred in two genes since the duplication. Silent substitutions are substitutions that do *not* lead to amino acid changes in an encoded protein. They are subject to weaker selection than amino acid replacement substitutions, and thus do not pose the same problem as amino acid substitutions in estimating time since divergence [65]. In practice, synonymous (silent) divergence of two duplicate genes is measured as the fraction K_s of silent substitutions that occurred since the duplication at synonymous nucleotide positions [65].

The second, much more serious, problem is to identify whether two gene functions are identical, similar, or completely different. To identify one gene's function is already difficult, because gene function can mean many different things. For instance, gene function can be defined in biochemical terms, such as through a gene product's DNA binding and catalytic activity, or via biological processes, such as through a gene's role in liver metabolism or brain development [66]. Another reason is that even presumably very well-characterized genes such as those encoding central metabolic enzymes or aminoacyl-tRNA transferases can perform utterly surprising additional functions in processes like cellular signaling or gene expression regulation [67]. To make matters worse, assessing whether gene duplicates *generally* retain similar functions can only be answered by characterizing hundreds or thousands of gene duplicates. The only current solution to this problem is to use crude indicators of gene function, indicators that are measurable for many genes. Functional genomics provides a wealth of such indicators, including where and when a gene is expressed, which proteins regulate its transcription, and which other molecules its protein products interact with. Each such indicator captures only one dimension of gene function, but it has the advantage that it can be measured for thousands of genes at once. I will now discuss evidence for the divergence of various gene function indicators after gene duplication.

One important indicator of similarity in gene function is similarity in gene expression. Some well-studied gene duplicates, for example, have preserved their biochemical function but differ in their spatial expression pattern, and can thus not be

deleted without deleterious effects [68]. Genome-scale analyses of expression divergence after gene duplication show that even recently duplicated genes show highly different mRNA gene expression patterns, so much so that there is at best a weak correlation between the silent divergence of two duplicate genes and the similarity in their expression patterns [69,70]. Figure 2 shows an example from the nematode worm *C. elegans* [64]. The figure is based on an analysis of more than 500 different microarray gene expression measurements and sequence information from the whole worm genome. Many gene duplicates with only moderate silent divergence (and thus very low amino acid divergence) show very different mRNA expression patterns, almost as different as genes that have duplicated a long time ago.

A similar pattern holds for the transcriptional regulators that bind at regulatory regions of duplicate genes [71,72]. After a gene duplication, the regulatory regions of two duplicate genes diverge. Much of the initial divergence occurs through the disappearance of binding sites for different transcription factors, because random mutations are more likely to eliminate such binding sites than to create new ones. Binding of transcriptional regulators to a large fraction of yeast genes has been studied using the technique of chromatin immunoprecipitation [73]. Figure 3 shows how the similarity in the spectrum of transcriptional regulators that bind to the regulatory region of two duplicate genes depends on the silent (synonymous) divergence between the duplicates [71]. For yeast, a synonymous divergence of $K_s=0.1$ among duplicated genes corresponds to a duplication age of approximately 10 million years [74,75]. The solid horizontal line indicates the average fraction of shared transcriptional regulators between pairs of genes chosen at random from the yeast genome. The figure demonstrates that the regulatory regions of many closely related duplicate genes are bound by different transcriptional regulators. This is further indication of a rapid functional divergence among duplicate genes.

Thirdly, genes whose products interact with other proteins show a rapid divergence of common protein interactions [72,74,75]. An example of such rapid divergence is shown for duplicate yeast genes in Figure 4. Even among the most recent gene duplicates, with a silent divergence of $K_s<0.2$, only some 60 percent of gene products share a protein interaction partner. Among older duplicates, fewer than 20 percent have any common protein interaction partners. Qualitatively, these findings are independent of which among various genome-scale datasets on protein interactions [76-79] one uses in this analysis.

Finally, the only moderately conserved indicator of functional similarity among duplicate genes is the localization of their protein products. Protein localization data based on fluorescence microscopy of proteins tagged with green fluorescent protein has recently become available for over 4,000 yeast genes. The available data classifies a protein's localization according to one or more cellular compartments (nucleus, endoplasmatic reticulum etc.) in which it occurs. Figure 5 shows the proportion of duplicate yeast genes with a given synonymous divergence that occur in identical cellular compartments (Figure 5a) or completely different cellular compartments (Figure 5b). Clearly, the fraction of genes with identical localization localization of duplicate genes at all levels of divergence localize to the same cellular compartments.

Each genome-scale data set on indicators of gene function has its own source of error, but taken together, the above evidence suggests that duplicate genes do not, as a rule, retain indications of similar function long after a duplication. It provides further

evidence that gene redundancy may not be the dominant factor determining robustness to mutations. This does, however, not mean that genes never retain similar functions long after duplication. For instance, some of the gene families listed above, families whose members retained similar functions, result from very ancient gene duplications. Examples include the yeast CLN family of cell-cycle regulators, and the TPK gene family of protein kinase catalytic subunits, in which individual members are dispensable for growth. The most closely related genes in these families show a silent divergence of $K_s > 1$, indicating that they may be more than 100 million years old [75]. The observation that some gene duplicates retain similar functions long after others have completely diverged represents an intriguing evolutionary puzzle waiting to be solved.

Quantifying redundancy and distributed robustness? Can we quantify the extent of redundancy in a genetic system whose inner workings are perfectly well understood? On a crude level, this seems quite feasible. For example, metabolic reactions in living cells are driven by enzymes, some of which may be the products of duplicate genes. One could simply determine the fraction of enzymes in a metabolic pathway that are encoded by more than one gene, and use this fraction as an indicator of redundancy in the network. However, can one go beyond that? Specifically, for any specific indicator of a system's performance, can we say to what extent its robustness is caused by redundancy or by distributed robustness? The answer to this question must depend on how one defines robustness. For instance, a flux-balance representation of metabolic networks, and a differential equation representation of a regulatory gene network are fundamentally different, and will therefore also employ different measures of robustness. But more disturbingly, redundancy and distributed robustness can have indistinguishable effects on a system. I will use an example from a simple metabolic system to make this point.

Metabolic control analysis is concerned with the behavior of metabolic pathways – chains of chemical reactions catalyzed by enzymes – when enzyme and metabolite concentrations change. An important quantity in this regard is an enzyme's flux control coefficient [80, p 4]. The metabolic flux through an enzymatic reaction is the rate at which an enzyme converts substrate into product per unit time. Flux is a key variable that links the activity of a metabolic enzyme to the phenotype. The reason is that the flux through some chemical reactions is strongly associated with fitness components like cellular growth or biomass production. An enzyme's flux control coefficient indicates the sensitivity of flux through an enzymatic reaction to changes in the activity of some enzyme in the pathway. Consider the linear metabolic pathway of Figure 5a. In steady-state, the metabolic flux through each reaction of the pathway must be identical to some value F. The flux control coefficient C_j of each enzyme E_j in the pathway is usually described as a differential quotient [80, p 4], which measures the response of flux F to a very small change in enzyme activity E_j.

$$ C_j = \frac{\partial F}{\partial E_j} \frac{E_j}{F} \tag{1} $$

If all enzymes in the pathway are unsaturated and show Michaelis-Menten kinetics, the flux coefficient takes on a particularly simple form [81]

$$C_j = \frac{1/E_j}{1/E_1 + 1/E_2 + 1/E_3 + \cdots + 1/E_n} \quad 1 \le j \le n \tag{2}$$

For my purpose, it will be sufficient to consider an even simpler case, where every enzyme E_j ($1 \le j \le n$) has the same activity e. In this case, the flux coefficient of each enzyme simply becomes $C_j = (1/n)$ ($1 \le j \le n$). A pathway like this is robust, in the sense that a small (mutationally caused) reduction in the activity of any one enzyme activity changes the steady-state pathway flux by an amount that is a factor n smaller. Notice that redundancy, as defined above, is not involved here, because each enzyme catalyzes a different chemical reaction. The observation that each and every enzyme E_j has the same effect on flux in this scenario raises the question whether such even distribution of control is a reasonable way to make the concept of distributed robustness exact, and to distinguish it from redundancy.

To see whether this is the case, imagine that we introduce redundancy into this pathway, as shown in Figure 6b. Specifically, we duplicate the gene encoding one of the enzymes, say enzyme E_i. Two copies of the gene produce twice the amount of enzyme, and as a result, this enzyme's activity doubles to *2e*. (In engineering, this kind of redundancy would be called parallel and active redundancy, because both copies of the system part are active – expressed – and their products act in parallel. [7, Chapter 6]) How does this duplication affect robustness of the pathway? For enzyme E_i, the control coefficient (2) now calculates as $1/(2n\text{-}1)$. If n is large, then the control coefficient of this enzyme has been effectively cut in half. In other words, the pathway is now nearly twice as robust to changes in activities of this enzyme. What about the remaining enzymes, whose genes have not been duplicated? Their control coefficient (2) increases to *2/(2n-1)*, a very small increase if the number of enzymes n is sufficiently large. If we were to introduce not one but *k>1* additional copies of the gene for E_i, then the pathway would become increasingly more robust to changes in the activity of E_i. In sum, the net effect of introducing gene redundancy into this metabolic pathway is to render the previously uniform distribution of control coefficients (2) non-uniform. From this perspective, one might propose to distinguish redundancy from distributed robustness through an uneven distribution of control in the pathway: Redundancy causes uneven distribution of flux control across parts of the pathway; the lower an enzyme's control coefficients, the higher its redundancy.

The fundamental flaw in this argument is that uneven distribution of control can be achieved by means that have nothing to do with redundancy. For example, even if each enzyme is encoded by only one gene, and even if each enzyme-coding gene expresses the same numbers of enzyme molecules, each enzyme molecule could have different catalytic activity, which would cause an uneven distribution of flux control. Similarly, if we duplicated all genes in the model pathway of Figure 5a, every enzyme in the pathway would have the same flux control coefficient – flux control would be perfectly evenly distributed – but now all enzymes are encoded by two redundant genes. In sum, a pathway could have a uniform distribution of control coefficients with much gene redundancy, or a non-uniform distribution without any gene redundancy.

All this implies that our ability to determine how much redundancy a system contains depends on our perspective. If we view a metabolic pathway from the point of

view of the genes that encode its enzymes, it may be simple to determine how much redundancy it contains – just count the redundant genes. The same holds if we view the pathway from the perspective of the individual protein molecules that do the chemical work: It is in principle easy to estimate how many of the molecules catalyzing the same reaction a cell contains, and all of these can be viewed as redundant. However, as I just showed, assessing the extent of redundancy becomes impossible on the level of enzyme activities, the aggregate variables best suited to understand a pathway's robustness to small changes in enzyme activity.

This problem is not a peculiarity of metabolic pathways. For instance, quantities suitable to measure the robustness of gene regulatory networks exist that are exactly analogous to metabolic control coefficients [82,83]. One can characterize the robustness of any property of such a network – the expression level of its genes, their products' activities or phosphorylation state, or their affinity to other molecules – by such control coefficients. And for any network design devoid of redundant genes, there may be a design including redundant genes that shows exactly the same distribution of control coefficients. Even more generally, pertinent and analogous insights exist in control theory, a discipline aiming at understanding and controlling the behavior of any system, whether natural or engineered. An important concept in control theory is that of the transfer function, which specifies how a system's output behaves as a function of its input. General theorems of control theory imply that for any system with redundant parts, one can design a system without redundancy but the same transfer function, i.e., the same behavior [84, Chapter 4].

All this implies that the qualitative approach to estimate the extent of redundancy in a biological system – enumerating redundant parts – may be the only one. Qualitative, graph-based representations of metabolic and gene regulation networks lend themselves ideally to such enumeration. Beyond such enumeration, however, it may be impossible to distinguish distributed robustness and redundancy, because the two have indistinguishable signatures on the variables necessary to understand a system's function and robustness.

In summary, redundancy as a source of robustness is especially prominent for systems whose parts are genes and gene products. Its importance results from the abundance of gene duplications in most genomes. In such systems, gene redundancy is partly responsible for increased robustness to mutations and an increased incidence of neutral mutations. However, redundancy may not be the most prominent cause of robustness on this level of biological organization. This is because many genes whose elimination has no effect on the organism are genes without duplicates. Also, typically fewer than half of a genome's genes are duplicates, and many of these duplicates diverge rapidly in function after duplication.

Acknowledgments

Figure Captions

Figure 1: Many genes with no detectable effect on fitness indicators are single copy genes. All genes on yeast chromosome V are divided here into two categories, those with the most severe fitness effects when eliminated from the genome (white bars), and those with a phenotype indistinguishable from wild-type (black bars). Numbers shown above bars are percentages of genes on chromosome V with the number of related genes in the yeast genome shown on the horizontal axis. Numbers of similar genes are based on a BLAST [85] search with a cut-off score of *E*=0.01. From Figure 5 in [61].

Figure 2: Even closely related duplicate genes diverge can have very different mRNA expression patterns. mRNA gene expression similarity (vertical axis) of duplicate genes in the worm *C. elegans*,. The horizontal axis shows synonymous divergence K_s of each duplicate gene pair. K_s is the fraction of synonymous substitutions that occurred since the duplication at synonymous nucleotide positions that do not lead to amino acid replacements. mRNA gene expression similarity is measured by the Pearson product-moment correlation coefficient of logarithmically transformed gene expression ratios from 522 microarray experiments [86]. Gene duplicates with K_s<0.2 are excluded, because microarray experiments cannot distinguish differences in expression among such closely related duplicates. Most duplicate gene pairs in the silent divergence interval shown would have an amino acid divergence of less than 20%. From [64].

Figure 3. Duplicate genes diverge rapidly in the binding of transcriptional regulators to their regulatory regions. a) After a gene duplication, the regulatory regions of two duplicate genes diverge. Much of this divergence occurs through the disappearance of binding by individual transcription factors (indicated by the small rectangular boxes), because random mutations are more likely to eliminate transcription factor binding sites than to create new ones. **b)** The vertical axis shows the fraction *f* of transcriptional regulators that bind to the regulatory region of *both* genes of a pair of yeast gene duplicates whose synonymous divergence is shown on the horizontal axis. If d_1 transcriptional regulators bind to the regulatory region of gene 1, if d_2 transcriptional regulators bind to the regulatory region of gene 2, and if d_{12} transcriptional regulators bind to the regulatory regions of both genes, then this fraction *f* calculates as *f=d_{12}/(d_1+d_2-d_{12})*. A value of *f=1* indicates that the same transcriptional regulators bind the two genes' regulatory regions. A value of *f=0* indicates that completely different regulators bind the two genes' regulatory regions. The horizontal lines indicate the average (solid) plus one standard deviation (dashed) fraction *f* of shared transcriptional regulators between pairs of genes chosen at random from the yeast genome. The synonymous (silent) divergence of two duplicate genes is measured by the fraction K_s of synonymous substitutions that occurred since the duplication at synonymous nucleotide positions [74,75]. From [71], which uses transcription factor binding data based on chromatin immunoprecipitation [73].

Figure 4: Duplicate genes whose products interact with other proteins diverge rapidly in their common interactions. a) Shortly after a gene duplication, the products P and P* of two duplicate genes interact with the same proteins. Eventually, some or all of the common interactions will be lost, and new interactions may be gained by either protein. The number of common interaction partners is a crude indicator of functional similarity among proteins. **b)** The horizontal axis corresponds to duplicate gene pairs in the protein interaction network of the yeast *Saccharomyces cerevisiae*, binned according to their evolutionary distance, as measured by the fraction of synonymous (silent) substitutions at synonymous sites, K_s[65]. The vertical axis shows the proportion of gene pairs in each bin whose products have *at least* one interacting protein in common. The horizontal line indicates the fraction of protein pairs chosen at random from the yeast genome that have at least one common interaction partner. Protein interaction data is taken from high-throughput genome-scale screens of protein interactions [76,77] and from a public database of such interactions (MIPS), as described in [78]. Similar results (not shown) are obtained when these data sets are considered separately. Qualitatively similar results are also obtained when experimental data of protein complex composition [79] are used in such an analysis.

Figure 5.: Moderate conservation of protein localization after gene duplication. The horizontal axes show synonymous divergence K_s of duplicate genes in the yeast *Saccharomyces cerevisiae*. The vertical axes show **a)** the fraction of duplicates in each divergence bin whose protein products have completely identical subcellular localization, and **b)** the fraction of duplicates with completely different (non-overlapping) subcellular localization. Subcellular localization data is based on fluorescence microscopy of more than 4,000 *Saccharomyces cerevisiae* proteins tagged with green flourescent protein [87].

Figure 6.: a) An enzymatic pathway consisting of *n* enzymatic reactions catalyzed by enzymes E_1 through E_n. **b)** The same pathway, but now the gene encoding enzyme E_i has been duplicated, introducing redundancy into the pathway.
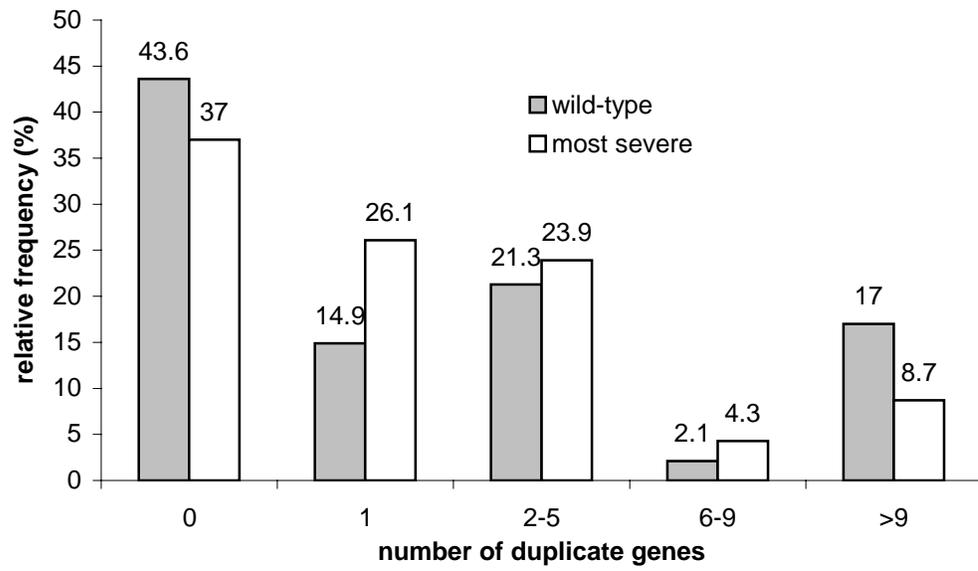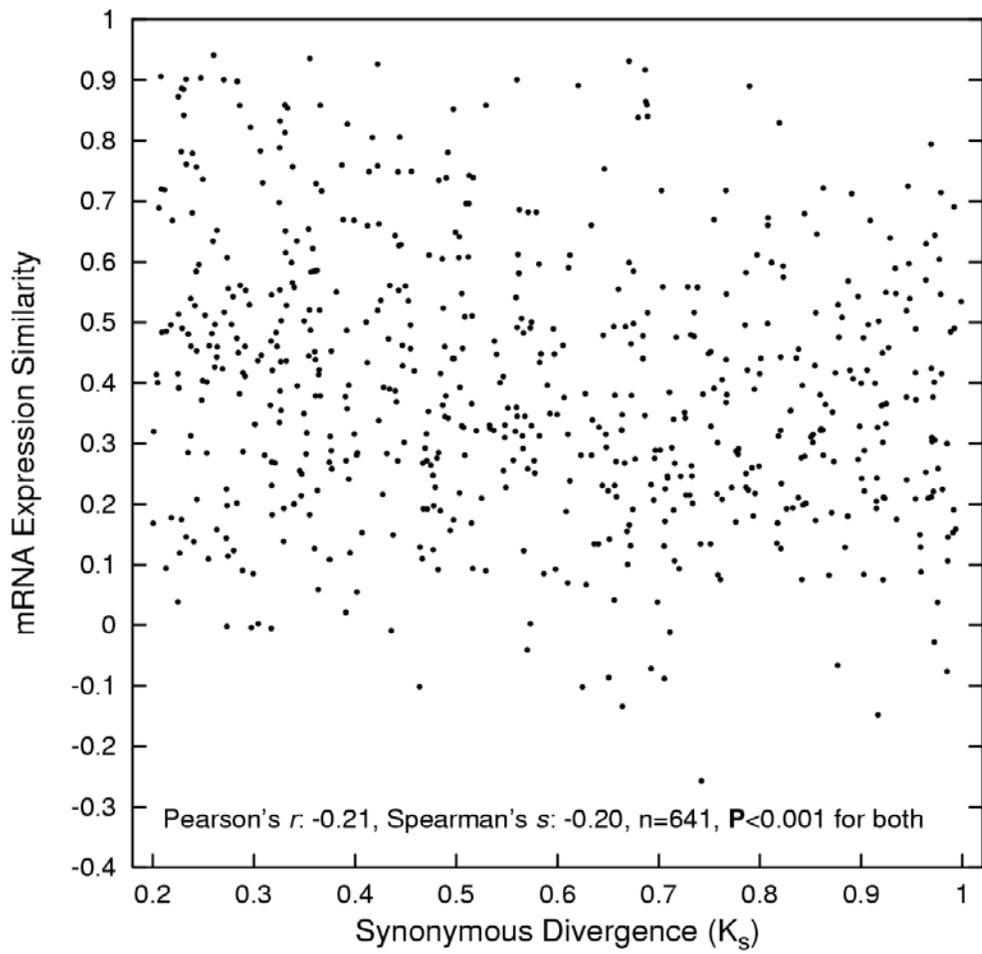
Figure 1

Pearson's *r*: -0.21, Spearman's *s*: -0.20, n=641, **P**<0.001 for both

Figure 2

a)

Duplication and divergence

b)

$\tau = -.234$ **P** $= 6.65 \times 10^{-19}$
$n = 646$

Fraction of Shared Regulators

Randomly chosen gene pairs

Synonymous Divergence (K_s)

Figure 3

a)

Duplication

Divergence

Lost interaction

New interaction

P   P   P*   P   P*

b)

Figure 4

a)

b)

Figure 5

a)

$E_1 \rightarrow E_2 \rightarrow \ldots \rightarrow E_i \rightarrow \ldots \rightarrow E_{n-1} \rightarrow E_n$

b)

$E_1 \rightarrow E_2 \rightarrow \ldots \nearrow E_i \searrow \ldots \rightarrow E_{n-1} \rightarrow E_n$
$\searrow E_i \nearrow$
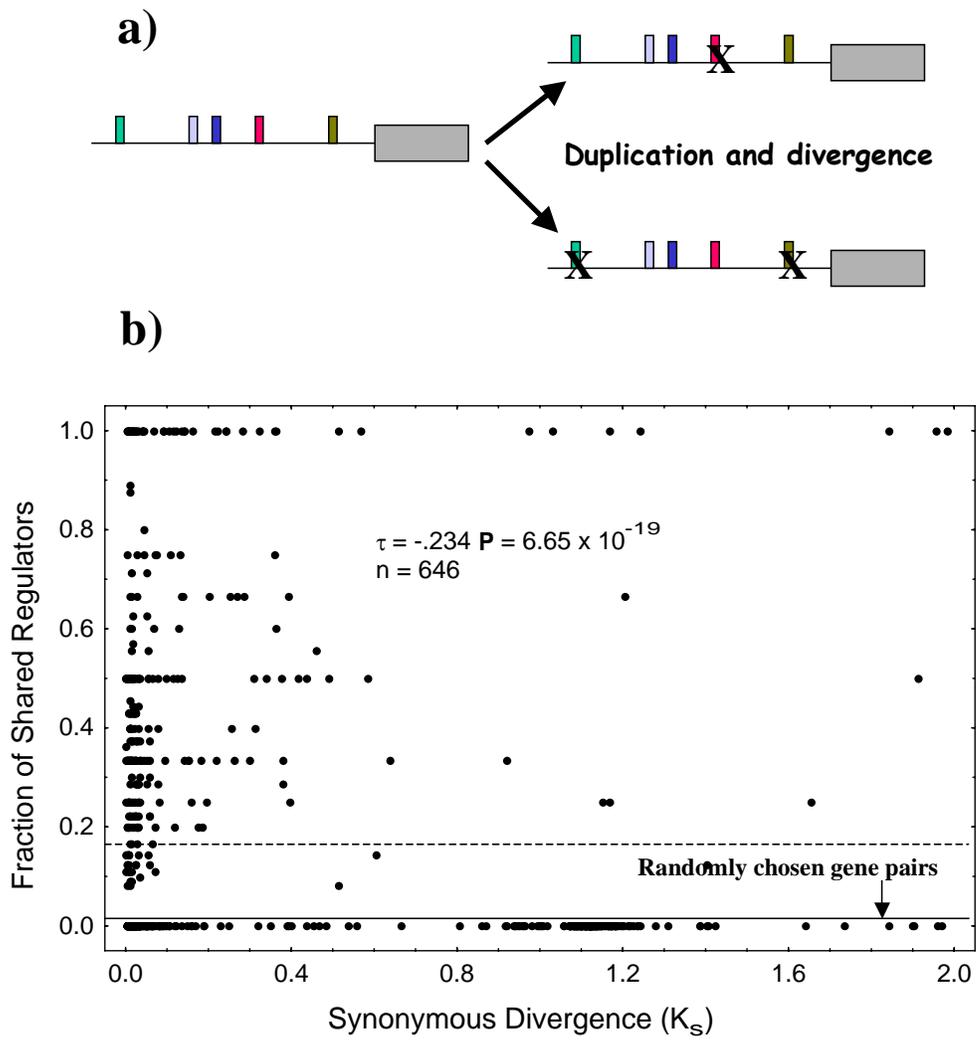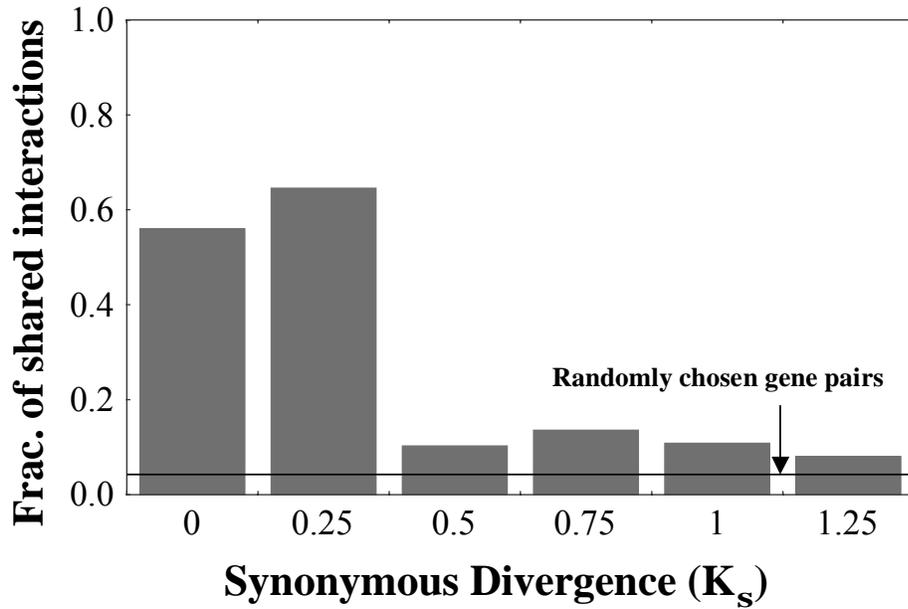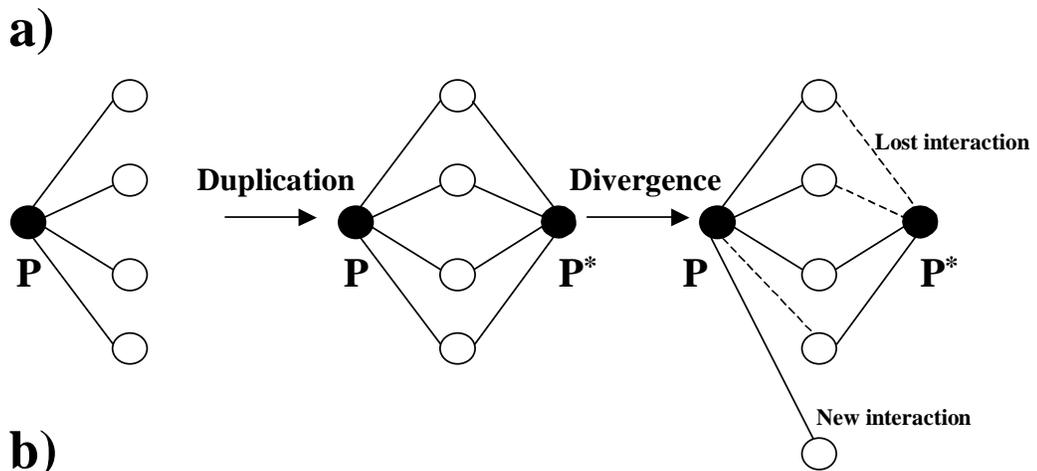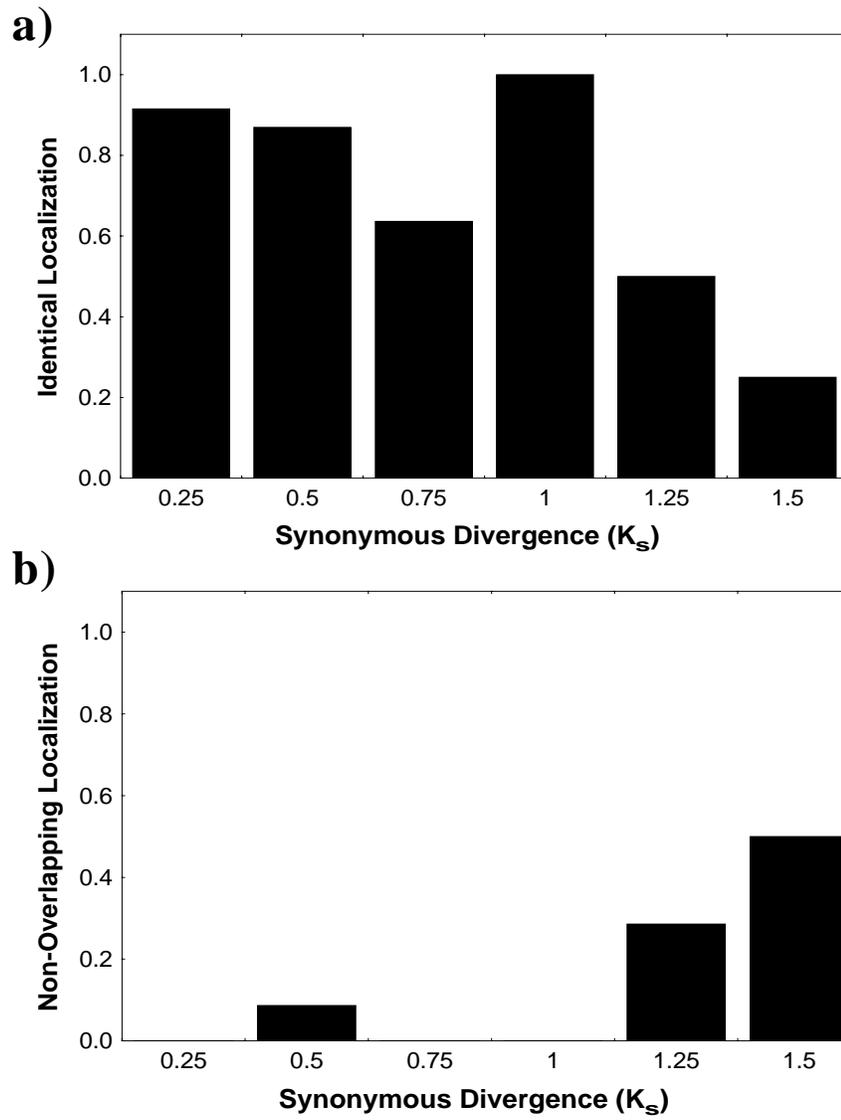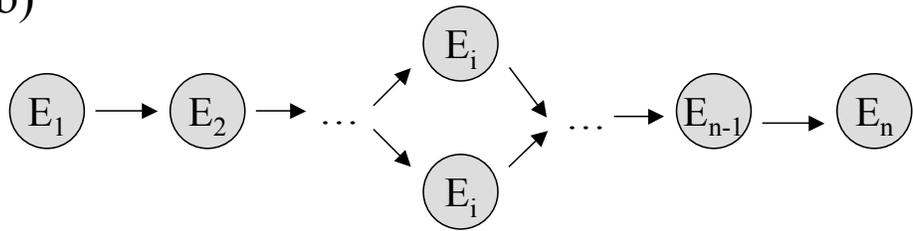
Figure 6

Literature Cited

1. Rennell D, Bouvier S, Hardy L, Poteete A. Systematic mutation of bacteriophage T4 lysozyme. Journal of Molecular Biology 1991;222(1):67-87.
2. Edwards JS, Palsson BO. Robustness analysis of the Escherichia coli metabolic network. Biotechnology Progress 2000;16(6):927-939.
3. von Dassow G, Meir E, Munro E, Odell G. The segment polarity network is a robust development module. Nature 2000;406(6792):188-192.
4. Raff RA. The shape of life. Genes, development, and the evolution of animal form. Chicago, IL: The University of Chicago Press; 1996.
5. Krakauer D, Plotkin J. Redundancy, antiredundancy, and the robustness of genomes. Proceedings of the National Academy of Sciences of the U.S.A. 2002;99(3):1405-1409.
6. The American heritage college dictionary. Boston, MA: Houghton Mifflin; 1993.
7. Fuqua NB. Reliability engineering for electronic design. New York, NY: Marcel Dekker; 1987.
8. Cadigan KM, Grossniklaus U, Gehring WJ. Functional redundancy : the respective roles of the 2 sloppy paired genes in Drosophila segmentation. Proceedings of the National Academy of Sciences of the United States of America 1994;91(#14):6324-6328.
9. Goldstein LSB. Functional redundancy in mitotic force generation. Journal of Cell Biology 1993;120(#1):1-3.
10. Hoffmann FM. Drosophila abl and genetic redundancy in signal transduction. Trends in Genetics 1991;7(#11-1):351-356.
11. Goodson HV, Spudich JA. Identification and molecular characterization of a yeast myosin-I. Cell Motility and the Cytoskeleton 1995;30(#1):73-84.
12. Joyner AL, Herrup K, Auerbach BA, Davis CA, Rossant J. Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the En-2 homeobox. Science 1991;251(#4998):1239-1243.
13. Thomas JH. Thinking about genetic redundancy. Trends in Genetics 1993;9(#11):395-399.
14. Wang YK, Jaenisch R. Myogenin can substitute for Myf5 in promoting myogenesis but less efficiently. Development 1997;124(#13):2507-2513.
15. Wang YK, Schnegelsberg PNJ, Dausman J, Jaenisch R. Functional redundancy of the muscle-specific transcription factors Myf5 and Myogenin. Nature 1996;379(#6568):823-825.
16. Lohnes D, Kastner P, Dierich A, Mark M, Lemeur M, Chambon P. Function of retinoic acid receptor gamma in the Mouse. Cell 1993;73(#4):643-658.
17. Lufkin T, Lohnes D, Mark M, Dierich A, Gorry P, Gaub MP, Lemeur M, Chambon P. High postnatal lethality and testis degeneration in retinoic acid receptor-alpha mutant mice. Proceedings of the National Academy of Sciences of the United States of America 1993;90(#15):7225-7229.
18. Lutfiyya LL, Iyer VR, Derisi J, Devit MJ, Brown PO, Johnston M. Characterization of 3 related glucose repressors and genes they regulate in *Saccharomyces cerevisiae*. Genetics 1998;150(#4):1377-1391.
19. Tautz D. Redundancies, development and the flow of information. Bioessays 1992;14(#4):263-266.
20. Gonzalez-Gaitan M, Rothe M, Wimmer EA, Taubert H, Jackle H. Redundant functions of the genes knirps and knirps-related for the establishment of anterior Drosophila head structures. Proceedings of the National Academy of Sciences of the United States of America 1994;91(#18):8567-8571.
21. Nasmyth K. Control of the yeast cell cycle by the Cdc28 protein kinase. Current Opinions in Cell Biology 1993;5:166-179.
22. Smith V, Chou KN, Lashkari D, Botstein D, Brown PO. Functional analysis of the genes of yeast chromosome V by genetic footprinting. Science 1996;274(#5299):2069-2074.
23. Higashijima S, Michiue T, Emori Y, Saigo K. Subtype determination of Drosophila embryonic external sensory organs by redundant homeo box genes BarH1 and BarH2. Genes & Development 1992;6(#6):1005-1018.

24.   Saga Y, Yagi T, Ikawa Y, Sakakura T, Aizawa S. Mice develop normally without tenascin. Genes & Development 1992;6(#10):1821-1831.

25.   Basson ME, Thorsness M, Rine J. Saccharomyces cerevisiae contains 2 functional genes encoding 3-hydroxy-3-methylglutaryl-coenzyme-A reductase. Proceedings of the National Academy of Sciences of the United States of America 1986;83(#15):5563-5567.

26.   Toda T, Cameron S, Sass P, Zoller M, Wigler M. Three different genes in *S. cerevisiae* encode the catalytic subunits of the cAMP-dependent protein kinase. Cell 1987;50:277-287.

27.   Conant GC, Wagner A. GenomeHistory: a software tool and its applications to fully sequenced genomes. Nucleic Acids Research 2002;30:1-10.

28.   Rubin GM, Yandell MD, Wortman JR, Miklos GLG, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W and others. Comparative genomics of the eukaryotes. Science 2000;287(5461):2204-2215.

29.   Robertson LS, Fink GR. The three yeast A kinases have specific signaling functions in pseudohyphal growth. Proceedings of the National Academy of Sciences of the United States of America 1998;95(23):13783-13787.

30.   Edwards JS, Ramakrishna R, Schilling C, Palsson BO. Metabolic flux balance analysis. In: Lee S, Papoutsakis E, editors. Metabolic engineering. New York: Marcel Dekker; 1999. p 13-57.

31.   Schuster S, Hilgetag C, Woods J, Fell D. Elementary modes of functioning in biochemical networks. In: Cuthbertson R, Holcombe M, Paton R, editors. Computation in cellular and molecular biological systems. Singapore: World Scientific; 1996. p 151-164.

32.   Schuster S, Dandekar T, Fell DA. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends in Biotechnology 1999;17(2):53-60.

33.   Ibarra RU, Edwards JS, Palsson BO. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 2002;420:186-189.

34.   Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. Journal of Biological Chemistry 1999;274(25):17410-17416.

35.   Edwards JS, Palsson BO. The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. Proceedings of the National Academy of Sciences of the United States of America 2000;97(10):5528-5533.

36.   Segre D, Vitkup D, Church G. Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences of the U.S.A. 2002;99(15112-15117).

37.   Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. Metabolic network structure determines key aspects of functionality and regulation. Nature 2002;420(6912):190-193.

38.   Fraenkel DG. Glycolysis. In: Neidhardt FC, editor. Escherichia coli and Salmonella: cellular and molecular biology. Volume 1. Washington, D.C.: American Society of Microbiology Press; 1996. p 189-198.

39.   Raya A, Kawakami Y, Rodriguez-Esteban C, Ibanes M, Rasskin-Gutman D, Rodriguez-Leon R, Buscher D, Feijo J, Izpisua-Belmonte J. Notch activity acts as a sensor for extracellular calcium during vertebrate left-right determination. Nature 2004;427:121-128.

40.   Alon U, Surette M, Barkai N, Leibler S. Robustness in bacterial chemotaxis. Nature 1999;397(6715):168-171.

41.   Barkai N, Leibler S. Biological rhythms - Circadian clocks limited by noise. NATURE 2000;403(6767):267-268.

42.   Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla E. Robustness and temporal pattern of floral organ determination as emergent properties of the *Arabidopsis thaliana* floral network dynamics. (submitted). 2003.

43.   Jiang G, Sladek F. The DNA binding domain of hepatocyte nuclear factor 4 mediates cooperative, specific binding to DNA and heterodimerization with the retinoid X receptor alpha. Journal of Biological Chemistry 1997;272(2):1218-1225.

44.   Pinckard J, Sheehan K, Schreiber R. Ligand-induced formation of p55 and p75 TNF receptor heterocomplexes: A mechanism underlying TNF receptor cooperativity. FASEB Journal 1996;10(6):1794.

45.   Brandt M, Vickery L. Cooperativity and dimerization of recombinant human estrogen receptor hormone-binding domain. Journal of Biological Chemistry 1997;272(8):4843-4849.

46. Stoecklin E, Wissler M, Moriggl R, Groner B. Specific DNA binding of Stat5, but not of glucocorticoid receptor, is required for their functional cooperation in the regulation of gene transcription. Molecular and Cellular Biology 1997;17(11):6708-6716.

47. Klinge C. Estrogen receptor binding to estrogen response elements slows ligand dissociation and synergistically activates reporter gene expression. Molecular and Cellular Endocrinology 1999;150(1-2):99-111.

48. Wanant S, Quon M. Insulin receptor binding kinetics: Modeling and simulation studies. Journal of Theoretical Biology 2000;205(3):355-364.

49. Wagner A. Does Evolutionary Plasticity Evolve? Evolution 1996;50(#3):1008-1023.

50. Rutherford S, Lindquist S. Hsp90 as a capacitor for morphological evolution. Nature 1998;396(6709):336-342.

51. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B and others. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 2002;418(6896):387-391.

52. Kamath R, Fraser A, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M and others. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature 2003;421(6920):231-237.

53. Alonso J, Stepanova A, Leisse T, Kim C, Chen H, Shinn P, Stevenson D, Zimmerman J, Barajas P, Cheuk R and others. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 2003;301(5633):653-657.

54. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H and others. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 1999;285(#5429):901-906.

55. Wagner A. Risk management in biological evolution. Journal of Theoretical Biology 2003;225:45-57.

56. Jordan I, Rogozin I, Wolf Y, Koonin E. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Research 2002;12(6):962-968.

57. Hirsh A, Fraser H. Rate of evolution and gene dispensability - Reply. Nature 2003;421(6922):497-498.

58. Hirsh A, Fraser H. Protein dispensability and rate of evolution. Nature 2001;411(6841):1046-1049.

59. Pal C, Papp B, Hurst L. Rate of evolution and gene dispensability. Nature 2003;421(6922):496-497.

60. Yang J, Gu Z, Li W. Rate of protein evolution versus fitness effect of gene deletion. Molecular Biology and Evolution 2003;20(5):772-774.

61. Wagner A. Mutational robustness in genetic networks of yeast. Nature Genetics 2000;24:355-361.

62. Gu Z, Steinmetz L, Gu X, Scharfe C, Davis R, Li W. Role of duplicate genes in genetic robustness against null mutations. Nature 2003;421(6918):63-66.

63. Qian J, Luscombe NM, Gerstein M. Protein family and fold occurrence in genomes: Power-law behaviour and evolutionary model. Journal of Molecular Biology 2001;313(4):673-681.

64. Conant GC, Wagner A. Duplicate genes and robustness to transient gene knockouts in Caenorhabditis elegans. Proceedings of the Royal Society of London Series B 2003;271:89-96.

65. Li W-H. Molecular Evolution. Massachusetts: Sinauer; 1997.

66. Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J and others. Gene Ontology: tool for the unification of biology. Nature Genetics 2000;25(1):25-29.

67. Jeffery C. Moonlighting proteins. Trends in Biochemical Sciences 1999;24(1):8-11.

68. Li XL, Noll M. Evolution of distinct developmental functions of 3 Drosophila genes by acquisition of different cis-regulatory regions. Nature 1994;367(#6458):83-87.

69. Gu Z, Nicolae D, Lu H, Li W. Rapid divergence in expression between duplicate genes inferred from microarray data. Trends in Genetics 2002;18(12):609-613.

70. Wagner A. Decoupled evolution of coding region and mRNA expression patterns after gene duplication: Implications for the neutralist-selectionist debate. Proceedings of the National Academy of Sciences of the U.S.A. 2000;97:6579-6584.

71. Evangelisti A, Wagner A. Molecular evolution in the transcriptional regulation network of yeast (submitted). 2004.

72.	Maslov S, Sneppen K, Eriksen K, Yan K-K. Upstream plasticity and downstream robustness in evolution of molecular networks. BMC Evolutionary Biology 2004;4:9.

73.	Lee T, Rinaldi N, Robert F, Odom D, Bar-Joseph Z, Gerber G, Hannett N, Harbison C, Thompson C, Simon I and others. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 2002;298(5594):799-804.

74.	Wagner A. How large protein interaction networks evolve. Proceedings of the Royal Society of London Series B. 2003;270:457-466.

75.	Wagner A. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Molecular Biology and Evolution. 2001;18:1283-1292.

76.	Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences of the United States of America ; 2001;98(8):4569-4574.

77.	Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P and others. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 2000;403(6770):623-627.

78.	Mewes HW, Heumann K, Kaps A, Mayer K, Pfeiffer F, Stocker S, Frishman D. MIPS: a database for genomes and protein sequences. Nucleic Acids Research 1999;27:44-48.

79.	Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM and others. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002;415(6868):141-147.

80.	Fell D. Understanding the control of metabolism. Miami: Portland Press; 1997.

81.	Kacser H, Burns JA. The molecular basis of dominance. Genetics 1981;97(3-4):639-666.

82.	Savageau MA. Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. Nature 1971;229:542-544.

83.	Savageau MA. Biochemical systems analysis: a study of function and design in molecular biology. Reading, MA: Addison-Wesley; 1976.

84.	Leigh JR. Control theory: a guided tour. London: IEE Publishing; 1992.

85.	Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. Gapped Blast and Psi-Blast : a new generation of protein database search programs. Nucleic Acids Research 1997;25(#17):3389-3402.

86.	Kim S, Lund J, Kiraly M, Duke K, Jiang M, Stuart J, Eizinger A, Wylie B, Davidson G. A gene expression map for *Caenorhabditis elegans*. Science 2001;293(5537):2087-2092.

87.	Huh W-K, Falvo J, Gerke L, Carroll A, Howson R, Weissman J, O'Shea E. Global analysis of protein localization in budding yeast. Nature 2003;425:686-691.