Gene duplications, robustness and evolutionary innovations

Wagner, A
Gene duplications, robustness and evolutionary innovations

Abstract

Mutational robustness facilitates evolutionary innovations. Gene duplications are unique kinds of mutations, in that they generally increase such robustness. The frequent association of gene duplications in regulatory networks with evolutionary innovation is thus a special case of a general mechanism linking innovation to robustness. The potential power of this mechanism to promote evolutionary innovations on large time scales is illustrated here with several examples. These include the role of gene duplications in the vertebrate radiation, flowering plant evolution and heart development, which encompass some of the most striking innovations in the evolution of life.
Gene duplications, robustness, and evolutionary innovations

Andreas Wagner

Department of Biochemistry, University of Zurich
Building Y27, Winterthurerstrasse 190
CH-8057 Zurich, Switzerland
Email: aw@bioc.unizh.ch
Phone (Secr.): + 41-44-635-6142
FAX: +41-44-635-6144

Abstract

Mutational robustness facilitates evolutionary innovations in biological systems. Gene duplications are unique kinds of mutations, in that they generally increase such robustness. The frequent association of gene duplications in regulatory networks with evolutionary innovation thus exemplifies a general mechanism linking innovation to robustness. I illustrate the potential power of this mechanism on large time scales with the role of gene duplications in the vertebrate radiation, flowering plant evolution, and heart development, which encompass some of the most striking innovations in the evolution of life.
Mutational robustness

Mutational robustness is a biological’s system ability to withstand mutations. Such robustness exists on multiple levels of biological organization. A case in point are random mutagenesis experiments of various proteins. They suggest that only a small fraction of mutations affect protein function adversely. For instance, a study of the bacteriophage T4 lysozyme generated more than 2000 random amino acid changes in the protein. Only 16% of them affected lysozyme function 1. Other examples come from regulatory gene networks, such as the molecular network specifying fruit fly segments. Such networks may tolerate much quantitative variation in interactions among network genes 2-5. Examples at the highest level of organization include macroscopic traits. Even substantial genetic variation – ultimately caused by mutations –may leave such traits unchanged. Take the vulva of the nematode worms *Caenorhabditis elegans* and *Pristionchus pacificus* 6. These organisms shared a common ancestor 200-300 million years ago. Their vulvae are very similar, yet the genetic and cellular networks producing them have diverged greatly. For example, whereas in *C. elegans* one specific cell – the anchor cell – induces vulva development, multiple gonadal cells are responsible for this induction in *P. pacificus* 7. Similarly, the same key signaling molecules, such as Wnt, may play a positive role in the network for vulval induction in *C. elegans*, but a negative role in *P. pacificus* 8,9. In sum, mutational robustness is everywhere, from proteins to organisms.

Gene duplications cause robustness

Many different kinds of mutations can affect a genome. They include point mutations, insertions, deletions, and chromosome rearrangements. Among them, duplications of genes and genomes are unique: Only they, as a rule, increase mutational robustness.

Two principal lines of evidence are germane. The first comes from efforts to eliminate (“knock out”) a gene’s expression to help determine gene function. To the chagrin of many a graduate student, such gene knock-outs often do not show a phenotypic effect, rendering them of limited use in some functional studies 10. Gene duplications are often responsible for such absent effects 11,12. Genome-scale efforts to eliminate each of thousands of genes in a genome lead to similar results 13,14, namely that only a fraction of genes have a phenotypic effect in the laboratory.

A second line of evidence comes from molecular evolution studies. Duplicate genes experience relaxed selection shortly after their duplication. They can tolerate more nucleotide changes than their single copy counterparts. The phenomenon is evident most clearly on a whole-genome scale 15,16, where recent gene duplicates in various eukaryotes tolerate 10-fold more amino acid changes than old duplicates 15. Eventually, the accumulating changes may cause duplicates to diversify their function, and sometimes quite rapidly 11,17,18.

Robustness facilitates evolutionary innovations

Narrowly defined, a biological system is *evolvable* if it can produce non-lethal, heritable phenotypic variation through mutations. More broadly defined, evolvability is the ability
to produce phenotypic diversity, novel solutions to the problems organisms face, and evolutionary innovations. Evolvability in the narrow sense is a prerequisite for evolvability in the broader sense. The requirement that phenotypic variation is non-lethal hints at an intimate link between evolvability and mutational robustness. This link is increasingly evident through laboratory evolution studies. For instance, recent experiments studied the evolvability of the enzyme cytochrome P450. In response to a given number of random mutations induced through error-prone PCR, the thermodynamically stable and mutationally robust variant of this enzyme evolved more readily the ability to hydroxylate several new substrates. Another example involves two ribozymes, the class III self-ligating ribozyme, and the Hepatitis D virus antigenomic ribozyme. These ribozymes are very different in sequence, structure, and in their biochemical activity (ligation versus cleavage). One of them is a product of laboratory design, the other is of biotic origin. Despite these differences a laboratory experiment succeeded at transforming one into the other by following a mutational path that involved some 40 point mutations. Importantly, along about half of this mutational path, the catalytic function of the evolving molecule did not change significantly relative to the starting sequence, indicating robustness. Halfway on this mutational path, however, a series of only four nucleotide changes allowed the molecule to adopt the activity of the target enzyme. While traversing this narrow region, one enzyme’s catalytic activity was transformed into that of the other.

While molecular traits are perhaps easiest to explore in evolution experiments, more complex traits have also been subject to similar analyses. One line of such work, beginning in the 1950s, focuses on morphological traits that normally show very little variation in wild populations. Through specific mutations, drug treatment, or environmental stress, such variation can be induced in a population. Subsequent artificial selection lets only organisms with the trait reproduce. Over multiple generations, populations usually respond with phenotypic change to such selection. Their response shows that the generated variation is heritable. It reflects genetic variation, caused by past mutations, but variation to which the organism is normally robust. Only in the special conditions of the experiment does that variation become expressed and available to natural selection.

In sum, multiple lines of experimental evidence show that robustness facilitates a system’s ability to produce new variation and evolutionary innovation. The principle at work is simple. It can be understood in terms of the genotype space that evolving populations explore. A population of robust individuals can explore this space rapidly, because fewer mutations affect them adversely, causing greater genotypic diversity among them. Mutations in a genotypically more diverse population, in turn, will produce more diverse phenotypes.

Gene duplications facilitate evolvability

Gene duplications cause robustness. Robustness, in turn, facilitates evolvability. Syllogistic necessity would dictate, then, that gene duplications facilitate evolvability. This last assertion is not new. Some 40 years ago, Ohno already made it for gene and genome duplications. According to him, duplications are key to evolutionary innovation.
What is new, however, is that now we can understand this role of gene duplications as a special case of a more general principle, namely that robustness facilitates evolvability. This principle manifests itself in the laboratory experiments discussed above, experiments that successfully evolved phenotypes ranging from new enzyme functions to morphological characters.

Laboratory experiments can demonstrate that gene duplications cause robustness, for example through systematic gene deletions. Laboratory evolution experiments can also demonstrate that robustness facilitates evolutionary innovations. However, laboratory experiments have one key weakness. They can only study evolutionary innovations that arise on modest laboratory time scales, such as modifications of existing enzymatic functions. To what extent does robustness promote evolutionary innovations on larger, geological time scales? Comparative studies, although they do not provide the conclusive proof of the laboratory, can address this question. They can examine spectacular evolutionary innovations that arose on time scales of hundred millions of years. I will now briefly review several such innovations and their association with gene duplications. They help us extrapolate from the laboratory to larger time scales. They give us a glimpse of how powerful robustness as an enabler of evolvability might be, when acting over hundreds of millions of years.

**Flowering plant evolution**

Flowering plants (angiosperms) are the most diverse and evolutionary successful group of land plants. Their approximately 250000 species outnumber those of all other plant taxa. Since their great radiation some 100 million years ago, flowering plants have come to dominate terrestrial ecosystems. Many of their key evolutionary innovations relate to reproductive functions. Among them are closed carpels that shield the female germ cells and prevent self-fertilization; the endosperm, a triploid tissue that nourishes a seedling; and, most visibly, flowers themselves. The prototypical angiosperm flower consists of four different floral organs – sepals, petals, stamens, and carpels – that arise sequentially from a floral meristem. A myriad variations exist on the number, arrangement, and synorganization of these four organs. Together, they account for the most visible aspects of angiosperm diversity 33.

The key to understanding angiosperm diversity lies in understanding angiosperm development, in particular the development of the flower. The identity of floral organs is specified combinatorially by a network of transcription factors that are expressed in the developing flower. The earliest and simplest incarnation of this insight is the so-called ABC model of flower development, established first in *Arabidopsis thaliana* and *Antirrhinum majus* 34. According to this model, the combined action of three classes of transcription factors called A, B, and C are necessary to specify floral organ identity. A class A transcription factor expressed by itself specifies sepals; A and B are jointly necessary to specify petals; B and C jointly specify stamens, whereas C alone specifies carpels. Accumulating evidence required some model modifications 35, but the model’s central notion, combinatorial specification of organ identity, is well corroborated.

Most of the well-studied transcription factors involved in flower organ specification are MADS box proteins. MADS box proteins are ubiquitous in eukaryotes. Flowering plants have experienced a wave of duplication in these genes 36–40 (Figure 1).
Yeast, nematode, and fruit fly genomes contain only between two and four MADS box genes \(^{40,41}\); the most recent common ancestor of gymnosperms and angiosperms may have had as few as 7 MADS box genes \(^{36,42}\). In contrast, the two completely sequenced genomes of the angiosperms *Arabidopsis thaliana* and rice each contain more than 70 MADS box genes \(^{38,39}\).

Some duplicate MADS box genes have preserved identical functions since their duplication, which underscores the notion that duplication causes robustness. Examples include the *SEPALLATA* genes, of which Arabidopsis contains several duplicates (*SEP1-4*). These genes are jointly responsible for converting leaf-like structures into petals, stamens, and carpels \(^{43-45}\). Loss-of-functions of individual members of this family, however, yield no strong phenotypic effects, indicating their redundancy and robustness to such mutations \(^{43,45}\). Another example is the *CAULIFLOWER (CAL)* gene. A loss of function in this gene has no phenotype. However, in combination with mutations in its closely related duplicate *APETALA1 (API)* gene, mutations in *CAL* give a characteristic cauliflower-like phenotype \(^{46}\).

In addition to such redundancy, many duplicate MADS box gene functions have also diversified within and among species \(^{47,48}\). This is expected if robustness caused by gene duplications provides the substrate for morphological evolution. Examples involve again the *SEP* genes. While redundant in Arabidopsis, *SEP* homologs have adopted different functions in other plants. A case in point is a tomato *SEP* homolog that is involved in fruit ripening but not in floral organ identity \(^{49}\). The *SEP* gene family has experienced further expansion in the monocotyledons. Based on divergent expression patterns in different grasses, it has been suggested that *SEP*-like genes may have influenced the morphological diversification of grass inflorescences \(^{50}\).

Another example involves the *AGAMOUS (AG)* gene family, whose name derives from *AGAMOUS*, a class C gene involved in carpel and stamen formation. This gene has experienced a duplication in the lineage leading to the eudicotyledons, creating two duplicate gene lineages. \(^{48,51}\) In Antirrhinum, ectopic expression of the AG family member *PLENA* transforms sepals into carpels, but ectopic expression of its paralog *FARINELLI (FAR)* does not \(^{52}\). The different loss of function phenotypes in the two genes show that they have adopted different functions \(^{52}\). Conversely, Arabidopsis contains two paralogs of *AG*, the *SHATTERPROOF* genes *SHP1* and *SHP2*, which have adopted new functions in fruit ripening \(^{53}\).

Taken together, examples like these suggest that the robustness originally caused by a duplication has facilitated evolutionary diversification on the molecular level. Such diversification is a prerequisite for morphological evolution.

**Vertebrate diversification**

No evolutionary account of gene duplications would be complete without Hox genes. Hox genes show a tightly linked (clustered) organization in many organisms. Their spatiotemporal expression pattern along the head-tail axis is colinear with their chromosomal order in a cluster. Hox genes are involved in the patterning of many structures along the head-tail body axis, including the hindbrain, the vertebral column and the limbs \(^{54}\). Many invertebrates have a single tightly linked cluster of Hox genes that underwent at least two duplications during vertebrate evolution. This means that many
Vertebrates have four Hox gene clusters labeled a-d. The cluster of the most recent common vertebrate ancestor likely had 14 Hox genes, of which 13 are left in vertebrates. The genes in the 4 vertebrate clusters are thus subdivided into 13 paralogous groups labeled 1 through 13.

Vertebrates are characterized by numerous innovations relative to their chordate ancestors. These include a more elaborate brain with three specialized regions (fore-, mid-, and hindbrain), cartilage, and mineralized structures – bone and teeth – that serve many roles from support to feeding. The evolution of bone in turn gave rise to the most obvious and striking vertebrate innovations. These include a differentiated vertebral column, hinged jaws, and paired appendages. The latter permit many different forms of locomotion, including walking, swimming, and flying, that made many ecological niches accessible when they first arose. Various duplicate Hox genes are critical for the proper embryonic development of these traits, suggesting important roles for Hox genes in morphological evolution.

Again, despite their duplication hundreds of million years ago, many Hox gene duplicates have retained partially redundant functions, remnants of the robustness that gene duplications cause. For example, zebrafish Hoxa2 and Hoxb2 function redundantly in embryonic patterning of the second pharyngeal arch; and the mouse Hox8 genes have redundant roles in positioning of the hindlimbs. While some aspects of Hox gene function are conserved, others have diverged. Here, a recurring theme is functional divergence through diverging gene expression rather than diverging biochemical function. A case in point are the duplicate Hox genes Hoxa3 and Hoxd3. The developmental defects found in loss-of-function mutations of either gene are very different. Hoxa3 mutants are defective in pharyngeal tissues, whereas Hoxd3 mutants show malformed cervical vertebrae. Their biochemical functions appear identical, such that quantitative expression changes may be responsible for their functional differences. Similarly, the coding regions of the duplicate mouse Hoxa1 and Hoxb1 genes are nearly identical, yet they have different functions in hindbrain development mediated by different spatiotemporal gene expression.

In sum, Hox genes have been duplicated early in the vertebrate radiation. The remnants of the resulting mutational robustness are still visible. The vertebrate radiation has produced a myriad innovations and great morphological diversity. Because Hox genes play critical roles in the development of the very traits involved in this radiation, it would be highly surprising if their diversification had played no role in the vertebrate radiation.

Heart Evolution

Gene duplications have been associated not only with spectacular evolutionary radiations, but also with evolutionary innovations in individual traits. One of them is the heart. In organisms too large for diffusion to distribute nutrients and oxygen, a pump driving fluid circulation through the body becomes necessary. The prototypical invertebrate heart and that of ancestral chordates is a simple contractile tube with bidirectional blood flow. In contrast, the amniote (reptile, bird, and mammalian) heart is a sophisticated four-chambered pump with two atria and two ventricles that separate oxygen-poor from oxygen-rich blood. During the evolution of vertebrates, the heart grew increasingly
complex. Fish hearts have a single atrium and a single ventricle, whereas amphibian hearts have two atria and one ventricle. Additional vertebrate innovations include septae to separate the heart’s chambers, valves to enforce unidirectional flow, as well as a conduction system for synchronized and powerful pumping \(^{65}\).

Heart development in vertebrates and invertebrates is controlled by a core network of transcription factor genes, including NK2, MEF2, GATA Tbx, and Hand (reviewed in \(^{66,67}\)). Like many other genes, these genes have more duplicates in vertebrates than in their chordate ancestor \(^{66}\) (Figure 2). One of these genes, MEF2 (myocyte enhancer factor 2), is involved in the expression of contractile muscle proteins. The fruit fly *Drosophila* has only one MEF2 gene. Loss of its expression eliminates expression of contractile proteins in muscle cells \(^{68,69}\). In vertebrates, there are four MEF2 duplicates showing partially redundancy \(^{70}\), a remnant of the robustness caused by their ancient duplication. Loss of function of MEF2c, one of these duplicates, eliminates a subset of contractile proteins in muscle cells \(^{68,69}\). In vertebrates, there are four MEF2 duplicates showing partially redundancy \(^{70}\), a remnant of the robustness caused by their ancient duplication. Loss of function of MEF2c, one of these duplicates, eliminates a subset of contractile proteins in the heart, and also abolishes formation of the right ventricle \(^{71}\). The population of cells from which the right ventricle is formed is specific to amniotes. The MEF2c function in it is thus probably a new acquisition. This example illustrates again the theme that single developmental regulators have broad functions, but their paralogs after duplication may adopt more specialized, restricted, and yet novel functions. Perhaps the most striking example of this principle is the *Hand* (heart and neural crest derivatives expressed transcript) gene. Zebrafish and amphibians, both of which have only one ventricle, express a single copy of this gene. The zebrafish *Hand* gene is necessary for ventricle formation \(^{72}\). Mice express two duplicates of *Hand*. Among other defects, loss-of-function mutants in *Hand1* are defective in left ventricle formation, whereas loss of function mutants in *Hand2* fail to form the right ventricle \(^{73-76}\). The functions of two duplicates have become partitioned such that each is associated with formation of a morphological partition of an organ.

**Conclusions**

In all of the above examples gene duplications lead to robustness, and this robustness may have allowed subsequent molecular and morphological diversification. Does this mean that gene duplications are **causal** to the radiations I discussed? The answer is no. They may be necessary, but they are certainly not sufficient. In any evolutionary process, both natural selection and variation are required. Variation without natural selection leads nowhere. The nematode vulva discussed earlier may be a case in point. We know that the vulva development network is robust to genetic change, because it has changed substantially in the last 200 million years. However, in this vast amount of time, the vulva itself has changed little. This is an example where robustness exists – although not necessarily caused by gene duplications –, where robustness allows genetic variation to occur and a network to change, but where the key impetus for morphological evolution – natural selection – may be missing.

Variation without selection does not lead to innovation, but the same holds for selection without (the right kind of) variation. One could view gene duplications as just one of many sources of genetic variation. However, gene duplications are unique and different from the many point mutations, deletions, and rearrangements that genomes are bombarded with. Only they increase robustness, and thus facilitate the production
evolutionary innovations. The flowering plant radiation, vertebrate evolution, and complexification of hearts indicate how powerful the principle of robustness as a facilitator of evolvability might be.

Acknowledgment

I am grateful for support through grant 315200-116814 from the Swiss National Foundation, as well as support from the Santa Fe Institute.

Figure Captions

Figure 1: highly simplified plant phylogeny, together with numbers of MADS box genes from representatives of key taxa, including yeast, nematodes, and fruit flies \(^{40,41}\); the green algae *Coleochaete scutata*\(^{77}\), and the dicotyledon *Arabidopsis thaliana* \(^{41}\). Numbers of MADS box genes should be understood as minimal numbers and could fluctuate within taxonomic groups. The images depict *C. scutata* (Permission and proper reference needed), and a flower of *A. thaliana* (Permission and proper reference needed.)

Figure 2: Number of duplicates for key members of the cardiac developmental gene network, together with the number of chambers in vertebrate hearts, and a highly simplified vertebrate phylogeny. After ref. \(^{78}\). (Permission needed)
Literature Cited


69. Ranganayakulu G, Zhao B, Dokidis A, Molkentin JD, Olson EN, Schulz RA. A series of mutations in the D-Mef2 transcription factor reveal multiple functions in


<table>
<thead>
<tr>
<th>Duplicate Cardiac Transcription factors</th>
<th>Heart Chambers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amniotes (Reptiles, birds, mammals)</td>
<td>1  2  3  4</td>
</tr>
<tr>
<td>Amphibians</td>
<td>1  2  3  4</td>
</tr>
<tr>
<td>Fish</td>
<td>1  2  3  4</td>
</tr>
<tr>
<td>Cephalochordates (Amphioxus)</td>
<td>1  2  3  4</td>
</tr>
<tr>
<td>Hand</td>
<td>1  1  1  2</td>
</tr>
<tr>
<td>MEF2</td>
<td>1  4  4  4</td>
</tr>
<tr>
<td>GATA</td>
<td>2  3  3  3</td>
</tr>
<tr>
<td>Tbx</td>
<td>?  ≥4 ≥5 ≥7</td>
</tr>
<tr>
<td>NK2</td>
<td>1  ≥3 ≥3  ≥2</td>
</tr>
</tbody>
</table>

Figure 2