Implications of duplicated cis-regulatory elements in the evolution of metazoans: the DDI model or how simplicity begets novelty

Senda Jiménez-Delgado, Juan Pascual-Anaya and Jordi Garcia-Fernández
Advance Access publication date 3 August 2009

Abstract
The discovery that most regulatory genes were conserved among animals from distant phyla challenged the ideas that gene duplication and divergence of homologous coding sequences were the basis for major morphological changes in metazoan evolution. In recent years, however, the interest for the roles, conservation and changes of non-coding sequences grew-up in parallel with genome sequencing projects. Presently, many independent studies are highlighting the importance that subtle changes in cis-regulatory regions had in the evolution of morphology through the Animal Kingdom. Here we will show and discuss some of these studies, and underscore the future of cis-Evo-Devo research. Nevertheless, we would also explore how gene duplication, which includes duplication of regulatory regions, may have been critical for spatial or temporal co-option of new regulatory networks, causing the deployment of new transcriptome scenarios, and how these induced morphological changes were critical for the evolution of new forms. Forty years after Susumu Ohno famous sentence ‘natural selection merely modifies, while redundancy creates’, we suggest the alternative: ‘natural selection modifies, while redundancy of cis-regulatory elements innovates’, and propose the Duplication–Degeneration–Innovation model to explain the increased evolvability of duplicated cis-regulatory regions. Paradoxically, making regulation simpler by subfunctionalization paved the path for future complexity or, in other words, ‘to make it simple to make it complex’.

Keywords: cis-regulatory evolution; gene duplication; Evo-Devo; metazoan evolution; DDI model

INTRODUCTION
More than a century after his death, Darwin’s rationale can be revisited with our current knowledge of genetics and development. It is as simple as to notice that, if animals and their fitness are the result of embryonic development, and embryonic development is the result of the deployment of complex genetic networks, then evolution of the wide diversity of metazoans has had to occur through changes and bricolage of these genetic systems and networks. In the particular case of bilaterian metazoans, these changes heavily concentrated in a relatively short period of time, during the Cambrian period, around 450 million years ago, in which almost all the current animal morphologies were already established. Understanding the changes that allowed such impressive explosion of animal forms is the central ambition that currently occupies many evolutionary scientists’ brains and hands.
A PARADOX THAT IS NOT A PARADOX

Since the identification and characterization of the anterior–posterior patterning Hox genes in *Drosophila* [1], hordes of other master developmental genes have been characterized, and BMP’s, Wnt’s, FGF’s, TGF’s and many other gene family abbreviations are incorporated in any literature regarding embryonic development in almost any animal model [2]. The major generalization of decades of research in developmental genetics at the end of the twentieth century was clear: major regulatory genes are conserved in almost all metazoans. This genetic toolkit patterns major body parts and structures in different animals and was probably set up in the first metazoans, and surely was present by the time of the Cambrian Explosion. The exciting result of commonality of developmental gene networks, however, raises a most puzzling conundrum: if all metazoans use the same genetic networks, how can they be so different?

A first proposal to solve the paradox is thinking it is not a paradox: maybe these generalizations are biased to particular species (fly, mouse, nematode) and to particular genes, and a deep analyses will show up other different genes and gene networks waiting to be discovered controlling the development of different groups.

However, if current generalizations match certainties, the proposal to solve the paradox is rather obvious: slight changes in proteins and protein functions, that is, *trans*-changes (as opposed to *cis*-regulatory changes for a given gene, Figure 1). In this manner, a developmental pathway could change and a new or modified trait be positively selected by evolution. Nonetheless, most developmental genes are involved in multiple pathways, given raise to several, often unrelated, traits. Thus structural changes affecting functional parts of these proteins would collectively affect several traits, and it seems unlikely that all these new traits would be co-selected simultaneously [3], hence that mutations affecting functionally the coding sequence are likely to be selected negatively. This argumentation is enhanced by two lines of evidence; first, functional domains within developmental proteins (e.g. homeodomain, b–HLH) have slower evolutionary rate than the remaining of the coding sequence [4], and second, although there are not many interspecies protein equivalence studies reported, in most of these cases, mutations of different developmental genes can be rescued by the expression of the orthologous proteins of a close or distant species. For example, human *HoxD4* in *Drosophila* substitutes the normal functions of its homologue *Deformed* [5] or the Hydra *achaete-scute* protein forms heterodimers with the endogenous *Daughterless* gene product in *Drosophila*, allowing the correct formation of sensory organs [6].

The former does not imply that there is no evolution of protein sequences at all, and that *trans-*evolution has never been at the base of morphological evolution. For instance, in insects the protein engrailed has gained a new groucho-interaction motif, not present in the rest of Arthropods, that represses *wingless* *in vivo* [7]. In addition, a now classic demonstration of the relevance of protein modifications is the gain in the insect lineage of a repression domain in the homeotic protein Ubx, which then represses *Distal-less* expression and leg formation in the abdomen, explaining why fruit flies only have six legs [8]. Moreover, amino–acid sequence or structural conservation between species is not purely static: different modifications that affect protein function are widespread, like the presence of distinct isoforms generated by alternative splicing, or post-translational modifications. For example, depending on differential splicing, the transcription factor AML1 may act as an activator or as a repressor [9].

Recent advances on the relevance of alternative splicing in morphological evolution can be found elsewhere [10]. Even though these and other examples, the vast majority of data shows that structural and functional conservation of proteins between species is extremely high, hence something else has to account for morphological evolution. Evolution of morphology has to be caused by changes at levels other than that which affects developmental proteins.

Therefore, if changes in the coding regions of DNA cannot account for most of the morphological diversification, and the genes are the same, changes in temporal and spatial expression (*cis* changes) of such developmental genes, that is, the regulation of their expression, has to correlate with evolutionary change. This idea was firstly noted by King and Wilson [11], who concluded, after analysing a large set of proteins from human and chimpanzee, that the behavioural and anatomical differences between us and them could not be explained by the small molecular differences they exhibited. Thirty years after King and Wilson seminal work, the sequencing of genomes from a great diversity of animal taxa has made possible a large number of comparative
Figure 1: The DDI model for regulatory evolution or how simplicity after duplication and degeneration turns into plasticity and innovation. The scheme shows a pleiotropic gene (A) regulated by four enhancers that drive expression in distinct temporal—spatial territories. The gene is evolutionary constrained, as modifications of the coding region or the complex regulatory region may affect the functionality of the gene in multiple territories. After gene duplication, changes in coding regions (C), as suggested by Sosumo Ohno, would allow the expression of a new protein in all the territories in which the original gene was expressed. For high pleiotropic genes, this will likely be negatively selected, as it will affect too many processes. (B) DDI: after the differential degeneration of regulatory enhancers [30] in the duplicated copies (subfunctionalization), the resulting genes will had its pleiotropy reduced, and the regulation of a particular copy will be simpler than originally. This will turn into higher plasticity or evolvability of these regions, thus facilitating the appearance of new regulatory regions which will allow the recruitment of the gene and its associated gene network in a new territory [5], paving the path to morphological innovations.
genomic and development studies, which make regulatory or cis-evolution the current hypothesis to explain morphological evolution [12]. In a recent report, Wilson and collaborators [13] approached regulatory evolution using a genomic approach. They take advantage of transgenic mice that harboured an extra human chromosome 21. Then they analysed wide expression profiles and showed that the expression of the genes in that human chromosome mimicked that of the human chromosome 21 and not that of the mouse orthologous regions. Given that the transcriptional machinery and cellular components were murine, the conclusion was that the cis-regulatory sequences were determinant for the species-specific expression profiles.

**Main features of regulatory evolution: cis-regulatory elements**

The cis-regulatory region of a particular gene is composed by a number of cis-regulatory elements (CREs) with binding sites for one or multiple transcription factors, organized in a modular manner. The spatial and temporal window of expression of every gene is controlled by its CREs [14]. The modular organization of the regulatory regions has a strong implication on how developmental genes can evolve, since it feasibly allows changes that affect different pathways independently. Therefore, CRE evolution gives a great flexibility for morphological evolution, by changes affecting the patterning and building of one part of the body, without affecting others [15].

CREs can evolve in different ways. One of them is by sequence mutations, which affect the binding site for a transcription factor by preventing, modifying the affinity or allowing de novo the binding of a transcription factor and consequently altering the spatial and/or temporal expression of the gene without implying changes in coding sequences. For instance, the sexual pigmentation of different fruit fly species is due to changes in the cis-regulatory region of the yellow locus, through gaining binding sites for much conserved transcription factors of wing development, like Engrailed, which controls the generation of a spot in male D. biaennis wings [16]. Interestingly indeed, gaining of such cis-regulatory blocks has happened independently in distinct regions of the yellow locus, in different Drosophila lineages [17]. This is one of the first reported cases of proper convergence studied at deep molecular level. Furthermore, those CREs have been independently lost in distinct Drosophila lineages. Another reported case is the binding site of Abd-B in the same yellow locus, which involves pigmentation of the posterior part of the male abdomen of D. melanogaster. The secondary lost of these new Abd-B binding sites in other Drosophila species results in the lost of such pigmentation [18]. Moreover, a striking case of regulatory evolution is found in the network involved in endoderm determination in echinoderms. Both sea stars and sea urchins use the same core gene network, but upstream regulators and downstream targets of the core are radically distinct in both clades, which indicates that cis-regulatory changes happened at least at two levels in these gene networks [19].

Besides these kind of mutations, with involved de novo appearance or lost of CREs, there are examples of CRE reshuffling. The gene sim, which regulates the patterning of a narrow ventral midline in arthropods, is a classical example. The regulatory region of this gene has been studied in different insects with a curious result. In the honeybee Apis mellifera sim is regulated only by Twist. In the case of Anopheles gambiae sim is regulated by the Notch pathway. Curiously, D. melanogaster sim is regulated by both Twist and Notch, while sim regulation in D. pseudoobscura and D. virilis resembled A. gambiae regulation. sim enhancers in these insects are in a dynamic turnover: whereas A. mellifera contains several Twist binding sites, it lacks binding sites for Notch effectors. And though in D. pseudoobscura and D. virilis sim is regulated only by Notch, there is no difference in quality or number of binding sites for Twist and Notch between D. melanogaster and the other two drosophilids [20]. This example clearly shows that restructuring in CREs is very dynamic and takes advantage of its modular nature. Furthermore, the reorganization and modular architecture of CREs make it possible that sequences of highly divergent regulatory regions promote equivalent downstream responses at gene expression level. For example, the regulatory region of the even-skipped gene from scavenger flies (Sepsidae), although show undetectable sequence conservation with its homologue in fruit flies, drives expression pattern in transgenic Drosophila identical to that of the endogenous gene of Drosophila [21].

An additional manner of CRE evolution is the recruitment of pre-existing CREs in other places within the genome, making possible the regulation of a new gene. Transposable elements (TEs) can be the cause of such movement. Bejerano et al. [22]
described a SINE transposable element in the coelacanth *Latimeria menadoensis* 0.5 Mb upstream of the *ISL1* gene. This element was found to recapitulate *in vivo* part of the expression of the gene, showing enhancer properties. The element has also been reported in mammalian genomes [22]. Furthermore, *in silico* searches of putative transcription factors binding sites suggest multiple target sites, including binding sites for ESR1, TP53, POU5F1, SOX2 and CTFC within distinct families of transposable elements [23]. A further analysis in mammalian genomes reveals that almost 25% of regulatory regions under study contain TE-derived sequences [24], highlighting the importance of mobile DNA for the evolution and dynamics of CREs.

Nonetheless, CRE mutations and reorganizations are under a certain selective pressure, as it happens with protein coding sequences: a gene miss expressed in space or time during development will probably be negatively selected, as it will, in principle, alter the body plan or body homeostasis of the organism. However, as we will suggest below, a way to gain some flexibility in mutations of cis-regulatory sequences is Ohno's duplication potential idea, but applied to cis-regulatory sequences.

**Duplication of CREs as a source of variability and evolvability**

When a gene is duplicated, not only the coding region is affected, but also are the surrounding regions (Figure 1A). Susumu Ohno’s work predated most studies of gene regulation, thus in his book *Evolution by Gene Duplication*, he focussed on the high variation potential of new proteins [25]. The duplicated genes and regulatory sequences reduce the constraints to change, since the probability for lethal mutations in all copies at once is very low. Some copies are lost because these mutations can accumulate in a duplicated copy, resulting in a pseudogene, or else, a copy might suffer mutations in its regulatory sequences that turn off its expression. Nonetheless, besides this degeneration process, duplicated genes might suffer other interesting changes.

**Neofunctionalization**

The neofunctionalization concept is controversial. It can be viewed as the process by which a protein changes its functional properties, and in this way is common in genes studied by biochemists, e.g. steroid receptors [26], but developmental biologists also include here the acquisition of new functions through the co-option of the gene in a new territory or in a distinct time of development, as this co-option might eventually generate a new embryological or physiological function. The first case is not very common on major developmental genes, because the structural characteristics, usually based in the protein domains, determine its biochemical function and those changes generally have a deep impact in highly critical developmental processes. Nonetheless, proteins might gain new domains by exon-shuffling, as it happened with the bilaterian gene *Hedgehog*, which originated by a combination of a hedge-domain and a more ancient hog/intein-domain [27]. The occurrence of the second case is much more common and is based on changes of the regulatory regions, e.g. the gain of a new CRE motif in a given promoter region alters (often, by recruitment) the spatial and/or temporal expression of the gene. Indeed, when a duplication event occurs, neofunctionalization by amino-acid substitutions is facilitated, because duplicated copies are not under selection constraints. Through accumulation of mutations, duplicated copies might acquire new functional properties (Figure 1C). However, bibliography is scarce for such sort of mutations. We here suggest that (i) duplication also facilitates neofunctionalization by means of changes in regulatory elements, and (ii) this phenomenon was widespread, particularly at the origin and early evolution of Vertebrates. For instance, the involvement of *Hox* genes in fin/limb development [28] or the recruitment of *Sonic Hedgehog* for determination of digits in mammals [29] are nice examples of the gain of a new expression territory, after gene duplication. Neofunctionalization of regulatory regions does not need duplication, but we propose that the extra-regulatory material generated after duplication might increase the chances of such neofunctionalization events.

**Subfunctionalization**

An interesting process that might occur when a gene is duplicated is the subsequent copy subfunctionalization, a process described by the Duplication–Degeneration–Complementation (DDC) model [30]. The model described the differential loss of CREs in two duplicated copies, which divided their function, so that only the expression of the
two genes, as a group, recapitulated the expression of the original ancestral gene. Several cases of DDC have been clearly identified. For example, the zebra fish Hoxb1a and Hoxb1b subfunctionalized the expression domains of the ancestral gene Hoxb1 without altering the protein functionality [31]. Another illustrative example is the expansion up to eight copies of the Hairy genes in the cephalochordate amphioxus, whereas mice bear a single hairy gene. The combined expression pattern of four out of those eight genes mimics the expression of the single mouse Hairy gene [32]. Computing analyses of the 5' region of these duplicates suggested complex events of differential degeneration between the amphioxus paralogues, [33]. Finally, the Xenopus laevis genome has retained ~40% of the duplicated genes produced in a recent whole genome duplication (WGD) event when compared to its unduplicated relative X. tropicalis, 68 out of 1300 pairs of duplicated genes suffered a reduction of gene expression in some tissues, and one-third of them had followed a process of subfunctionalization [34].

The DDI model
Here we propose that subfunctionalization and neofunctionalization are not only not mutually exclusive phenomena, but also are intimately related (Figure 1B). Subfunctionalization of highly regulated genes is a very common event. Subsequently, enhancers and other regulatory modules that were lost in the duplicated copies might well become raw material retaining structural enhancer properties, which by subtle mutations—supposedly neutral, as these regions became non-functional after the DDC—might thus acquire new spatial–temporal expression profiles, and this on its own might mean neofunctionalization, or even result in morphological innovation.

Gene duplication per se is not a real force to induce morphological changes, but rather, in our view, may facilitate them, although not precisely in the way that Susumu Ohno suggested. We now know that gene networks are essentially the same in most metazoans, and that they were present, more or less interconnected, at the origin of metazoans. What we suggest is that gene duplication by means of subfunctionalization diminishes gene pleiotropy, and that this reduced level of pleiotropy increases the evolvability of gene regulation. In the original DCC model [30], neofunctionalization by acquisition of new enhancers was included as a possible route after gene duplication. We go further and propose, more explicitly, that after duplication the duplicate genes became more specialized, so that they are expressed in fewer territories. Therefore, their regulatory regions became simpler, and as such, less constrained to change. We surmise that, as a consequence, subtle changes in these ‘fresher’ regulatory regions would allow co-option of these duplicated genes in new territories and times. These novel co-options might allow new interconnections amongst distinct gene networks, increasing their complexity and connectivity, hence providing the grounds for further increase in organism complexity, that is, paving the path for evolution (controversially assuming here that evolution may be linked to an increase in complexity). What we think is extremely exciting, is the apparent paradox within a paradox, namely, that the initial reduction of pleiotropy after gene duplication might eventually result in a huge leap in the complexity of gene networks interaction, that is, in an increase in organism complexity. We propose the term DDI to explain this increased evolvability of duplicated cis-regulatory regions. The apparent solution to this conundrum: ‘making it simple to make it complex’.

The ambition to the future: testing cis-regulatory evolution experimentally
Testing any scientific hypothesis in the lab by means of in vivo or in vitro experiments is a crucial part of any science project, and Evolution should become a discipline not out of this logical process. However, most evolutionary studies so far come from inferences on comparisons (anatomical, genetic, genomic). We are entering the so-called third wave of Evo-Devo [35], that is, after finding the surprising similarities (first wave) and subtle differences (second wave) between animal clades, it is time to test in the lab those genetic changes identified, and induce, if they are the true causal events, those morphological changes predicted.

If cis-regulation is the main driving force of evolutionary change, those experimental Evo-Devo studies should concentrate in analysing the function and the change in function of CRE. Two main types of experiments can be envisaged. First, the concrete function of a given identified CRE can be identified using transgenesis in which the putative CRE controls a reporter gene in the endogenous context.
By this, we can elucidate which component of the temporal and/or spatial expression is due to a given CRE [36–38]. Of course, bioinformatics has a very important role in identifying in silico those good CRE candidates for transgenesis analysis and excellent bibliography to do so is available [13, 34] and Ovcharenko, I., in this issue. Then, when a proper function has been identified, the extent of evolutionary conservation of function must be tested. For this purpose, heterologous transgenesis—CRE from species A in species B—will inform if the function of this cis-regulatory region is conserved or not. These sorts of experiments get highly exciting when working with key evolutionary novelties, for instance, the origin of tetrapod limb. Posterior Hox genes of mammalian HoxA and HoxD clusters are expressed in, and control the patterning of developing limbs [39]. Posterior HoxD genes are controlled by the Global Control Region (GCR), a region placed far upstream the cluster [40], which also drives expression of these genes in the genital bud and the central nervous system. The GCR is made up by a set of cis-regulatory regions that controls all the posterior genes of the cluster as a whole. This meta-cis-regulation was probably responsible for the consolidation of vertebrate clusters [41]. Spitz et al. [40] injected a β-lac reporter gene in mice under the control of a 7.8-kb GCR region from the fish Tetraodon and recovered reporter expression in the central nervous system (where the expression in the forebrain was enlarged) and the genital bud, but not in limb buds. Consequently, a plausible explanation for this observation was that the tetrapod GCR had gained CREs for the regulation of expression of posterior HoxD genes concomitantly with the origin of the tetrapod limbs. Heterologous injection of dissected tetrapod GCR in fishes and dissected teleost GCR in mouse will help to elucidate the proper syntaxis of CREs that was concomitant with the origin of the tetrapod limb.

The second sort of experiments involves the functional testing of cis-regulatory changes. The expression of an endogenous gene under the control of an enhancer that is responsible of an evolutionary novelty might allow knowing whether this enhancer is truly the main responsible to account for such change. In a precious series of experiments, cis-changes that allowed the generation of wings from ancestral forelegs in bats are being identified. The enlarged length of bats’ forelimbs bones with respect to others mammals is critical for the correct formation of wings, and consequently, for achieving powered flight. This evolutionary novelty is due in part to differences in Bmp2 expression in some digits (Figure 2A) [42] and in differences of regulation of Prxl, an important regulator of long bone elongation during limb development [43]. First, Sears et al. [42] proposed that the differential level of expression in digits between mice and bats is likely due to differences in cis-regulation, by acquisition of a strong enhancer in the bat lineage. And recently, Cretekos et al. [43] substituted the mice enhancer of the Prxl gene by the corresponding region of the bat gene [44]. Strikingly, these transgenic mice exhibited a significant increase of ~6% in forelimb bone length with respect to wild type (Figure 2B). Although it does not explain on its own the leg to wing transition, and surely many other genes/processes were involved, the experiment by Cretekos et al. is a breakthrough in the study of cis-regulatory evolution. In the coming years, this sort of experiments targeting major evolutionary innovations at key crossroads of evolution will, surely, shed light on the mysteries and mechanisms of evolution.

**CONCLUSIONS**

The answer to the initial question on the molecular mechanism that drove morphological evolution is far from being simple. We emphasized here the potential of CRE evolution, and highlighted an intriguing link between gene duplication and increasing in the evolvability of gene regulation. We introduced the DDI model, in which the differential loss and simplification of regulatory regions after gene duplication increases the chances to evolve new enhancers, increasing the interconnections within regulatory gene networks, a path that may well flow into morphological innovations. However, it does not escape our attention that other processes, like the generation of protein variants by changing alternative splicing, exon shuffling, subtle protein mutations, changes in regulatory RNAs, changes in gene editing and many other processes sure had relevance at distinct evolutionary times. Only a deeper understanding of the evolution of genome architecture and regulation, and in-depth analyses of complex gene networks in a comparative manner, together with the exciting new experimental approaches to test those changes, will allow us to understand how these multitude of ‘endless forms most beautiful’ appeared on Earth.
In a recent article, Kassahn and collaborators analysed changes in protein domains and regulatory control in the teleost lineage, and concluded that the most significant consequence of whole genome duplication was to enable more specialized regulatory control of development via de acquisition of novel spatio-temporal domains in duplicated gene copies, in a sharp example of the DDI model. (Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome Res*; E-pub ahead of print: 12 July 2009).

**Figure 2**: Regulatory evolution and experimental Evo-Devo. (A) Bats’ wing evolved from an ancestor with forelimbs similar to actual mice. The BMP2 gene is expressed at high levels in the digits of developing bat forelimb, probably due to a regulatory change in the regulatory region of the gene. The high level of BMP2 expression explains partly the increase in the size of the skeletal components of the bat forelimb. (B) Experimental Evo-Devo in vertebrates. Transgenic mice in which the endogenous enhancer of the limb-expressed gene Prx1 was replaced by the BatE enhancer of the bat Prx1 gene showed an increase of ~6% in the length of the forelimb, mimicking partly the increase that facilitated the origin of flights in the bat lineage. (A) Based in reference [42], (B) based in reference [43].

**NOTE ADDED IN PROOF**

In a recent article, Kassahn and collaborators analysed changes in protein domains and regulatory control in the teleost lineage, and concluded that the most significant consequence of whole genome duplication was to enable more specialized regulatory control of development via de acquisition of novel spatio-temporal domains in duplicated gene copies, in a sharp example of the DDI model. (Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome Res*; E-pub ahead of print: 12 July 2009).

**Key Points**

- Regulatory gene networks are highly conserved in Metazoans.
- Evolution was mostly driven by changes in CREs.
- After gene duplication, cis-regulatory regions became simpler by subfunctionalization, beget more evolvable, and smooth the progress to gene recruitment, facilitating innovation and novelty.
- We propose the DDI model to explain the increase of evolvability of duplicated regulatory regions.
- Experimental Evo-Devo will serve to test in the laboratory the regulatory changes responsible of morphological evolution.
Acknowledgements
The authors are indebted to Manuel Irinia, Ignacio Maeso, Salvatore D’Aniello and Beatriz Albuxch for long-lasting and stimulating discussions and to Hannah Katz Rotnitzky for checking the English. J.G.F. also thanks Gemma Marfany for a highly motivating discussion, Jaume Baguà for sharing passion and deep knowledge and Laura for an inspiring dialogue. J.G.F. dedicates this article to the memory of Laura’s father.

FUNDING
Grant BMC2008-03776 from the Spanish Ministerio de Educación y Ciencia to the lab; FI fellowship (Generalitat de Catalunya) to J.P.A.

References


