Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Biotechnology

How do regulatory networks evolve and expand throughout evolution?

Karin Voordeckers[1,2,3], Ksenia Pougach[1,2,3] and Kevin J Verstrepen[1,2]

Throughout evolution, regulatory networks need to expand and adapt to accommodate novel genes and gene functions. However, the molecular details explaining how gene networks evolve remain largely unknown. Recent studies demonstrate that changes in transcription factors contribute to the evolution of regulatory networks. In particular, duplication of transcription factors followed by specific mutations in their DNA-binding or interaction domains propels the divergence and emergence of new networks. The innate promiscuity and modularity of regulatory networks contributes to their evolvability: duplicated promiscuous regulators and their target promoters can acquire mutations that lead to gradual increases in specificity, allowing neofunctionalization or subfunctionalization.

Addresses
[1] CMPG Laboratory for Genetics and Genomics, KU Leuven, Gaston Geenslaan 1, B-3001 Leuven, Belgium
[2] VIB Laboratory for Systems Biology, Gaston Geenslaan 1, B-3001 Leuven, Belgium

Corresponding author: Verstrepen, Kevin J (kevin.verstrepen@biw.vib-kuleuven.be)

Current Opinion in Biotechnology 2015, **34**:180–188

This review comes from a themed issue on **Systems biology**

Edited by **Sarah Maria Fendt** and **Costas D Maranas**

For a complete overview see the Issue and the Editorial

Available online 24th February 2015

http://dx.doi.org/10.1016/j.copbio.2015.02.001

Introduction

Even in closely related species with highly similar genome sequences, gene expression patterns can be quite different [1,2]. This divergence in gene expression and regulation has been postulated to play a major role in evolution and is believed to be one of the primary sources of phenotypic variation between species [3–11].

Changes in transcriptional regulation can occur at different levels: through changes in DNA binding sites located around or inside target genes (so-called *cis* mutations) or by changes in *trans*, that is, differences in the abundance or activity of transcription factors (TF) — regulatory proteins that recognize and bind specific *cis*-regulatory sequences [12]. Comparative genomics studies have indicated a considerable amount of *cis*-regulatory sequence variation between species [13–15] and it has been argued for a long time that changes in *cis*-regulatory elements underlie most of the observed changes in transcriptional regulation [8,16,17]. Mutations in transcription factors were considered to be an unlikely source of variation, mostly because of the possible negative pleiotropic effects such mutations can evoke [16,18]. A mutation in a protein-coding region of a transcriptional regulator may simultaneously affect multiple target genes of this regulator (and thus can have widespread detrimental effects), whereas a mutation in a *cis*-regulatory element would only cause changes in the expression pattern of this particular gene and might thus be better tolerated by the cell [8].

Recent studies indicate that mutations in regulatory proteins may be more common than previously appreciated [19–21]. Moreover, these changes can play a prominent role in regulatory network evolution by altering expression, molecular interactions and post-translational modifications of the regulator [22–24,25••]. In keeping with this, it is well known that several transcription factors have DNA binding domains belonging to large paralogous families, although the transcription factors can differ extensively in sequence [26]. Hence, evolution through TFs appears to be a successful strategy for regulation of gene expression, although the exact nature and extent to which this mechanism has contributed to gene expression regulation has remained unclear [27••].

Duplication of a gene encoding a transcription factor was suggested to be the least complicated way for a transcription factor to evolve without significantly decreasing the fitness of an organism [28]. For example, one of the gene copies can retain the ancestral function (thus avoiding any negative pleiotropic effects), while the other is released from negative selective pressure, can mutate and in some cases evolve a different function [23]. Indeed, many transcription factors are known to arise by gene duplication, and a number of them have acquired a new function [29–31]. In addition, duplication of target genes — both small-scale and whole-genome — and subsequent diversification of the resulting duplicates have been shown to

[3] These authors contributed equally to this work.

be important contributors to the evolution of transcriptional networks [1,32•,33,34]. Gene duplication has been widely recognized as the prime sources of novel genes in genomes: 50% of the genes in prokaryotes and around 90% of the genes in eukaryotes are the result of duplication [31,35–39]. Since these new genes need to be regulated correctly, the adaptation of gene regulation (and thus of regulatory networks) is particularly important [1,2,31].

In this review, we discuss recent insights in how duplication of a transcription factor gene can propel the rewiring and expansion of regulatory networks. We specifically focus on how gene duplication and subsequent divergence allows circumventing the potential negative effects associated with pleiotropy of a single copy transcription factor that could lead to misregulation of target genes.

Gene duplication is an important driver of regulatory network evolution

Gene duplication is increasingly recognized as the chief mechanism underlying evolutionary innovation. Whereas the exact evolutionary pathways and forces are often complex, a simplified model explains how duplication of a gene allows one of the two copies to retain the ancestral function whereas the other copy is relieved from negative selection and is allowed to mutate and explore novel functions [35,40]. Such duplication events are often associated with genes encoding enzymes, but they may also occur for the regulatory genes [23,41,42••,43••,44]. Duplication of a transcriptional regulator, its target gene(s) or duplication of both may establish novel interactions in the regulatory network or even lead to the emergence of a novel regulatory cascade [31] (Figure 1).

Comparative genomics reveals that many transcription factors, as well as their target genes, arose by duplication [29–31]. After duplication of a regulatory gene, the two identical copies are likely redundant, recognizing the same binding sites, responding to the same signal and, therefore, regulating the same set of target genes as the ancestral pre-duplication regulator. During subsequent divergence, one or both of the duplicated transcription factor paralog genes may acquire mutations that change the DNA binding domain and switch to regulating different target genes [42••]. Alternatively, the two paralogs can continue to regulate the same target genes as their ancestor but respond to a different signal, or bind different protein partners (cofactors) [45••,46,47]. A seemingly frequently occurring scenario is that of subfunctionalization (or 'division of labor'), where each paralog evolves to regulate a subset of the target genes originally regulated by the single ancestral transcription factor [40,45••,48] (Figure 2). Such subfunctionalization might not seem to contribute much to evolution, but in reality, division of labor among paralog regulators followed by specific mutations may allow a more precise and specific regulation of target genes. Another possible fate for duplicated genes is neofunctionalization, where one of the duplicates acquires a novel function that was not present in the pre-duplication protein. Such neofunctionalization could explain the emergence of completely new pathways that regulate new gene functions (Figure 2).

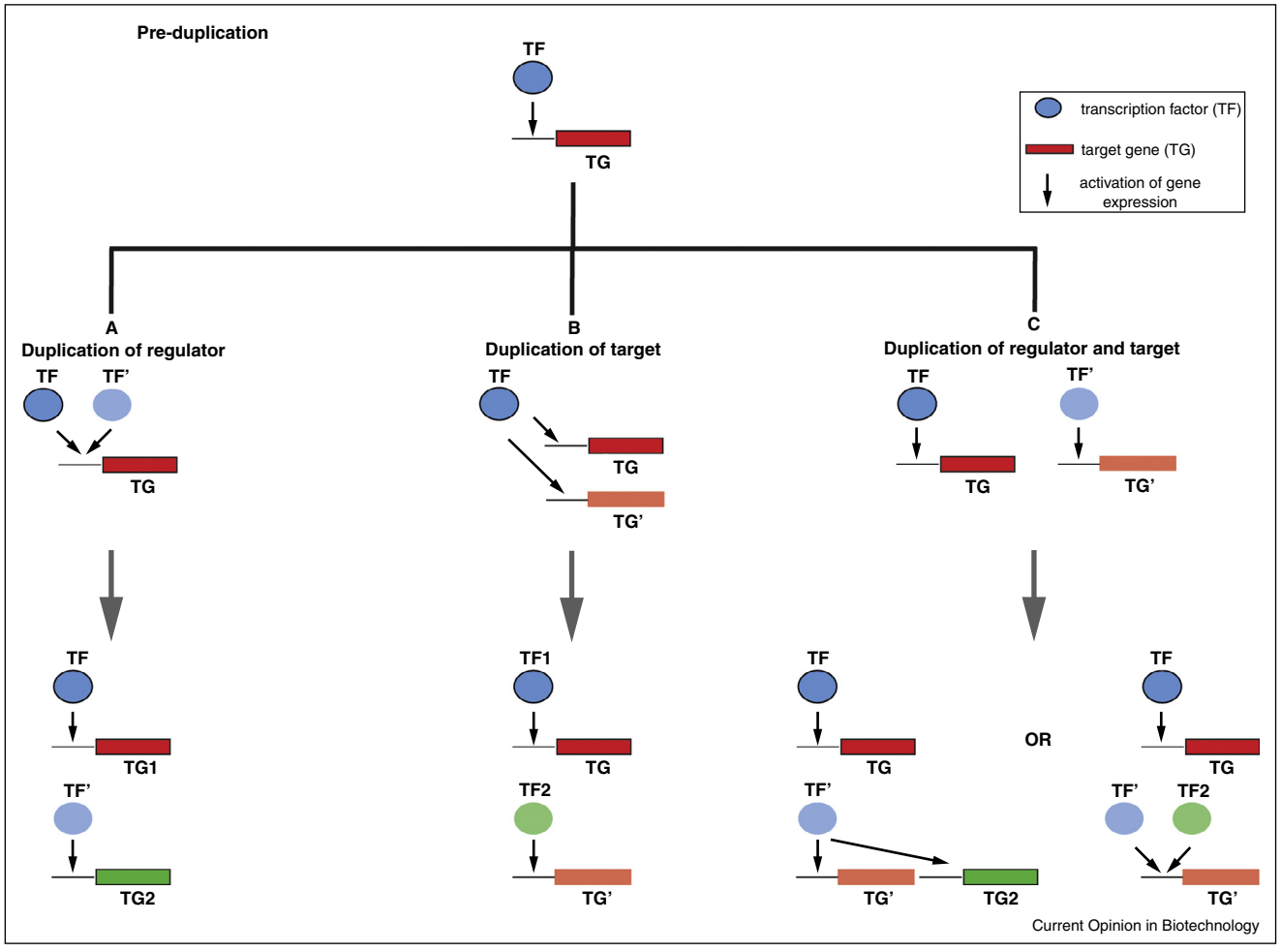
Interestingly, despite the multitude of examples of how gene and whole-genome duplications have contributed to the evolution and expansion of gene regulatory networks, the exact molecular details and mutational pathways are not yet well understood. How can two identical transcription factors gradually diverge into two distinct proteins, each responding to a specific input and each regulating a specific set of targets? It is important to note that this is a complex problem, because evolution generally happens gradually, and during the entire process, fitness valleys associated with misregulation of target genes should be avoided. In the following paragraphs, we describe the results of recent studies that have elucidated mutational pathways underlying the evolution of duplicated transcription networks. Together, these studies begin to shed light on how transcriptional regulation evolves.

Subfunctionalization of duplicated transcriptional networks

Many transcription factors interact with a multitude of other proteins and also with different DNA motifs. In case of subfunctionalization, loss of some of these ancestral interactions in the resulting paralogs can lead to competitive interference between the two paralogs, a situation also referred to as paralog interference [41,45••]. Imagine for example a transcription factor that needs to bind a specific cofactor as well as DNA. If, after duplication, one of the paralogs acquires mutations that impair cofactor binding but do not affect DNA binding, then this paralog will reduce transcriptional activity of the other copy by competing for DNA binding. Baker *et al.* demonstrated the negative effects of such paralog interference in the case of a fungal MADS-box transcriptional regulator [45••]. Duplication of the ancestral transcription factor resulted in two paralogs that each control expression of a specific subset of targets of the ancestral regulator [45••]. The two paralogs diverged by acquiring specific mutations that altered cofactor binding preference. In a clever set of ancestral gene reconstructions, the authors showed that closely after the duplication the regulatory network was indeed experiencing paralog interference. Several specific subsequent mutations that weakened the DNA-binding affinity of one of the paralogs were required to resolve paralog interference.

However, in some cases, paralog interference can also be an integral part of the emergence of new regulatory loops.

Figure 1



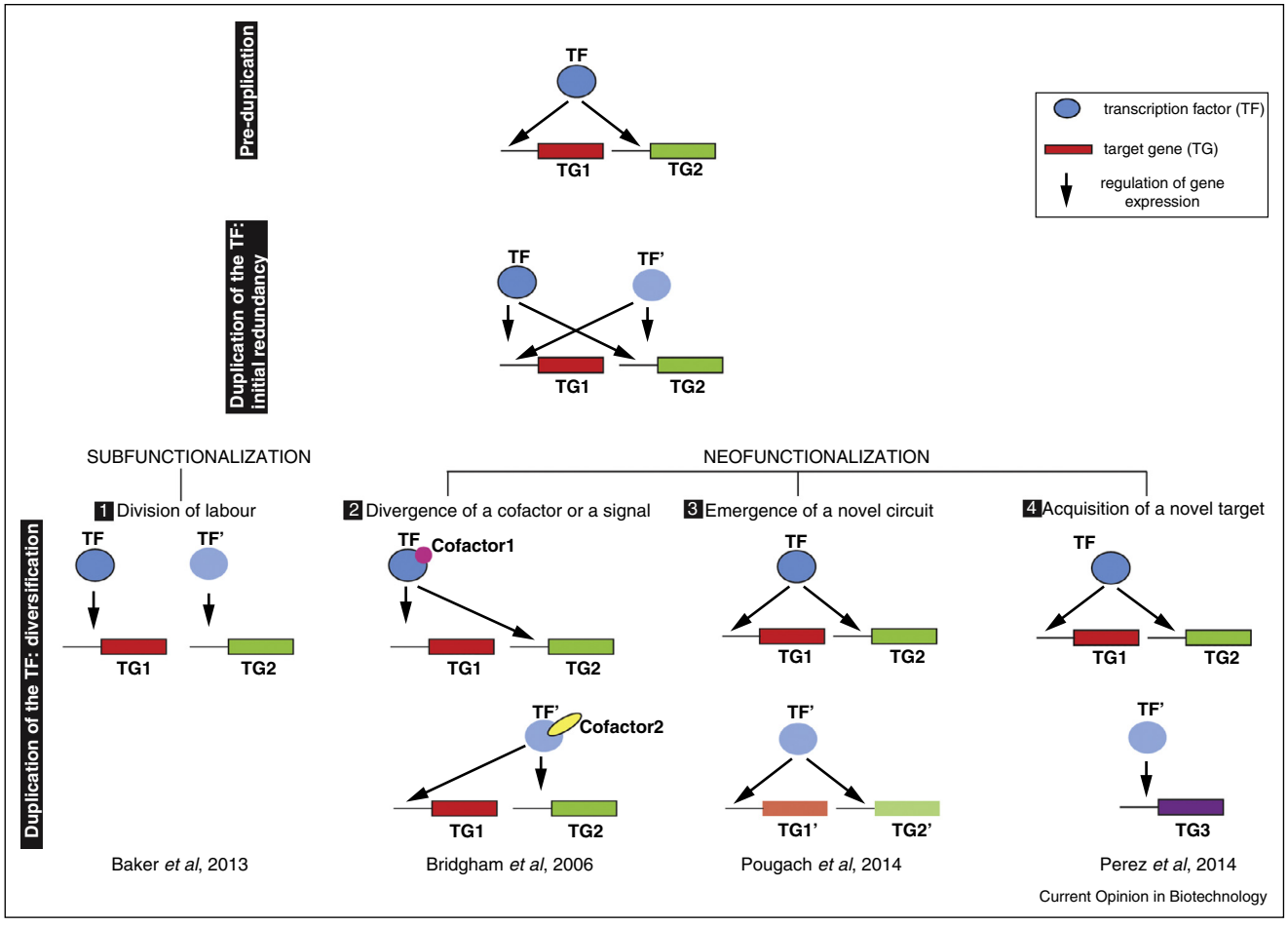
Duplication events are important drivers of regulatory network evolution. **(a)** Duplication of a transcription factor (TF) initially leads to both regulators (TF and TF′) controlling the same target gene (TG). Subsequent divergence can cause one of the paralogs (TF′) to acquire different target(s) (TG2) (see also Figure 2). **(b)** After duplication of a target gene (TG), both copies (TG and TG′) are initially regulated by the same transcription factor (TF). Divergence can cause the duplicated target to become regulated by a different transcription factor (TF2). It should be noted that both in scenario a and b, paralog divergence does not necessarily happen after duplication: the two copies of the regulator can regulate the same target genes (a); and the two copies of the target can be regulated by the same transcription factor (b) (not depicted). Alternatively, the duplicated copies can also be lost from the genome (not depicted). **(c)** Concerted duplication of a transcriptional regulator and its target (e.g., following whole-genome duplication events) can expand the regulatory network through divergence of target and/or regulator.

Bridgham *et al.* showed that in the evolution of steroid hormone receptors, mutations in one of the paralogs after duplication abolished its activation by a specific ligand [41]. Upon ligand binding, these receptors normally undergo a conformational change that allows them to recruit cofactors that facilitate transcription of their targets. Mutations affecting ligand binding caused one of the paralogs to evolve a new function: it acted as a repressor for the transcriptional activity of the other paralog by competing for the same DNA binding site (neofunctionalization through loss-of-function). This in turn allowed for fine-tuning of expression of the target genes of these transcriptional regulators.

Neofunctionalization of transcriptional networks

While subfunctionalization of duplicated networks clearly plays an important role in the evolution of gene regulation, other cases, such as the one described in the paragraph above, involve the emergence of novel functions (neofunctionalization). This is often the case when novel gene functions arise, since this likely requires the evolution of new regulatory networks to ensure proper regulation of the novel functions. A recent study of the evolutionary history of the *MAL* gene family in yeast reveals in detail how such 'neofunctionalization' of networks can occur. In this case, duplication and

Figure 2

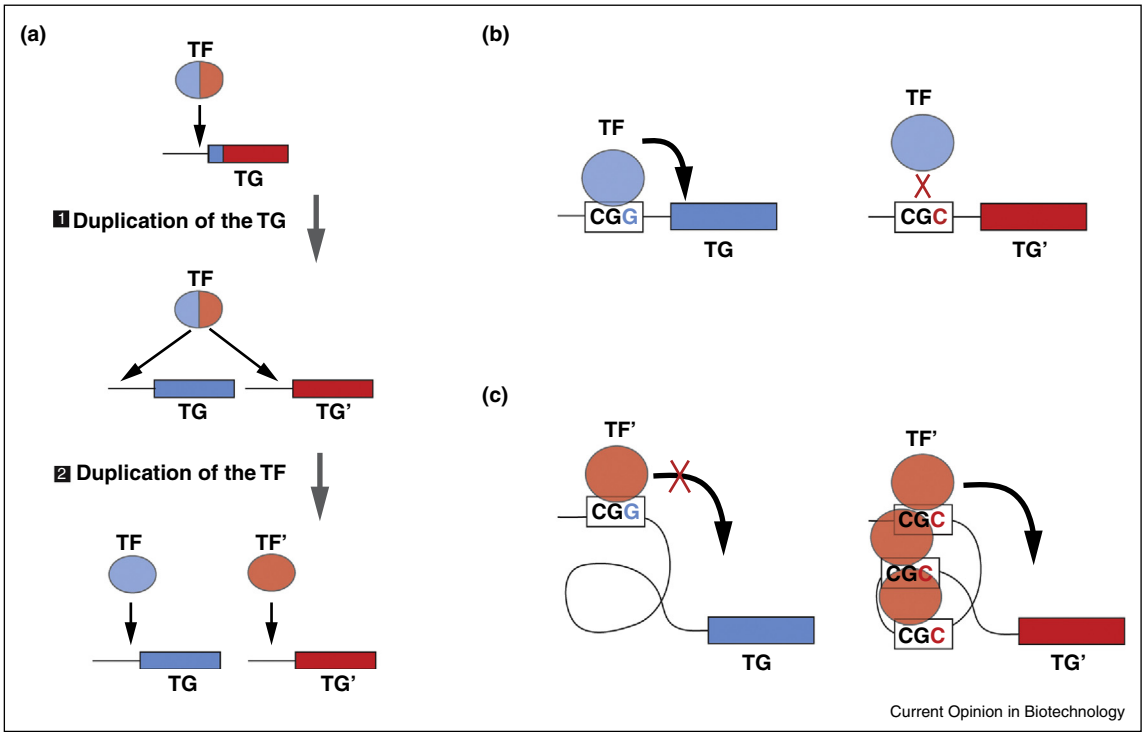


Transcription factor duplication can lead to expansion of the existing regulatory network or emergence of a novel network. Early after the duplication of the ancestral transcription factor (TF) the resulting paralogs are redundant and regulate the same set of target genes (TG1 and TG2). Subsequent diversification can occur through subfunctionalization (or division of labor, **1** [45]) or neofunctionalization (**2**–**4**). (2) Some transcription factor paralogs diversify at the level of signal or cofactor recognition, so that they still regulate the same set of target genes but are activated by different cofactors [46]. (3) Duplication of the target genes together with the regulator can lead to emergence of a novel network, where each transcription factor paralog activates its own set of target gene copies (TG1, TG2 and TG1′ and TG2′ are regulated by TF and TF′ respectively) [43]. (4) Some transcription factor paralogs can acquire regulatory control over new target genes that were not part of the ancestral network before duplication (TG3) [42].

neofunctionalization of a hydrolytic enzyme was followed by duplication of its transcription regulator gene, which eventually allowed the emergence of a novel regulatory circuit [43••]. The *MAL* gene family is involved in the uptake and metabolism of complex carbohydrates such as maltose, palatinose and other α-glycosides. It consists of three subfamilies: *MALS*, *MALT* and *MALR* [44,49]. The *MALT* subfamily encodes a set of different transporters that import a range of disaccharides, which are subsequently hydrolyzed by a set of specific MalS glycosidases. Some of the intracellular disaccharides likely bind to the so-called MalR regulator proteins, and this complex regulates expression of *MALS* and *MALT* genes [49].

The present-day MalS enzymes originated from a single-copy ancestral gene through multiple duplication events followed by functional diversification. These events created two main groups of MalS enzymes, each with its own specific substrate preference: one group can hydrolyze carbohydrates with an α 1-4 glycosidic bond, whereas the other group can break down disaccharides with an α 1-6 bond [44,49]. Two independent regulatory networks, each one dedicated to regulating one class of MalS enzymes, emerged through duplication and diversification of a single copy transcription factor [43••,44] (see Figure 3). This ancestral regulator displayed broad binding site specificity and was able to bind two different DNA motifs present in the promoters of the two substrate classes. After duplication, one paralog evolved increased binding specificity for one of the motifs, whereas the other paralog acquired mutations that reduced its affinity for specific sequences, so that it now needs multiple

Figure 3



Evolution of MAL gene regulatory network. **(a)** In case of the *MAL* genes, the ancestral transcription factor (TF) was regulating expression of the ancestral hydrolytic enzyme (target gene, TG). This ancestral transcription factor was promiscuous and could be activated in presence of two classes of disaccharides (each activity is depicted as either red or blue), while the ancestral enzyme could effectively hydrolyze only the red disaccharides (with α 1-4 glycosidic bond), and had only minor activity for the blue disaccharides (α 1-6 glycosidic bond). (**1**) Duplication of the target gene led to significant improvement of the blue hydrolytic activity and was followed by the duplication of the transcription factor gene (**2**). One of the emerging transcription factor paralogs took regulatory control over the target gene with the novel activity, thus leading to emergence of a novel regulatory network (blue). The effective disentanglement of the red and blue regulatory networks was possible because the blue regulator lost the ability to bind to the promoter regions of the red target genes and can only bind its own blue target genes. The opposite is true for the red regulator. The details of the mutations that lead to these changes are depicted in panels b and c. **(b)** The blue regulator acquired a specific mutation in its DNA binding domain that prevents it from binding CGC motifs present in the promoters of the red target genes, but allows binding CGG motifs from the promoters of the blue target genes. **(c)** The red regulator retained the ancestral ability to bind both types of motifs (CGG and CGC) but requires multiple binding sites to activate the expression. Blue target genes have only one binding site in their promoters and thus cannot be activated by the red regulator. For more information, see [43].

binding sites before it can bind (see also Figure 3). This example shows in detail how duplication and subsequent diversification of a promiscuous transcription factor allowed avoiding fitness valleys associated with paralog interference [43••]. In other words, the two networks were gradually disentangled so that in the end, each *MALS* gene is specifically activated by its own substrates. Interestingly, comparative genomics revealed that the changes in the amino acid sequence of the Mal target proteins preceded the changes on the regulatory level. Hence, in this case, functional divergence of the target genes probably acted as a driving force for the gradual disentanglement of the two regulatory networks. Moreover, the promiscuity of the ancestral, pre-duplication transcription factor, which was able to bind multiple binding motifs, was likely a key facilitator of the post-duplication divergence. Indeed, a gradual increase in the binding specificity of a duplicated promiscuous transcription factor may allow gradual disentanglement of the paralog networks.

A very recent study focused on the mechanism of binding site specificity changes in vertebrate steroid hormone receptor evolution [50••]. One of the duplicated transcription factors retained the ability to recognize the DNA motif preferred by the ancestral regulator, while the other acquired mutations and now specifically binds a different motif. In contrast to the *MAL* study discussed above, the ancestral regulator in this case was not promiscuous and could only bind one type of motif. Interestingly, negative protein–DNA interactions were intimately involved in the evolution of the new DNA binding specificity: after duplication, mutations in one copy compromised its binding to the original motif, whereas other mutations

caused a loss of unfavorable interactions with the new motif.

Another study of the divergence of a duplicated transcription factor elucidated the molecular details of how a duplicated transcription factor can acquire a set of completely novel target genes [42••]. By studying a group of *LYS* transcription factors in *Candida albicans* that arose through successive duplications, they demonstrated that these paralogs diverged through a combination of changes in DNA binding specificities (preferences for different DNA motifs, different spacing between motifs, preferences for direct versus inverted repeats), as well as different cofactor binding properties. Interestingly, *LYS14* of *Saccharomyces cerevisiae* is a key regulator of lysine biosynthesis, but in *C. albicans* none of the four *LYS14* homologs regulates the lysine pathway. Instead, they are involved in the white-opaque switching or proliferation in a mammalian host.

Interestingly, mutations that do not perturb the actual protein–DNA binding site can also affect binding affinity and specificity of transcriptional regulators. An elegant study on the transcriptional activator CAP demonstrated that mutations far away from the binding site altering protein internal dynamics can have a very strong effect on binding affinity [51]. Hence, many different types of mutations can lead to transcriptional regulator divergence.

Importance of promiscuity for overcoming the negative effects of pleiotropy

Initially considered to be mostly unwanted side-activities, recent studies are now highlighting the importance of promiscuity as crucial factor for driving evolution. The evolution of the Mal network is a good example (see above). Promiscuity seems to be a feature inherent to most proteins and enzymes: a large fraction of them have been reported to possess multiple activities [52,53]. Many transcription factors also show variation in DNA binding abilities: they can recognize both a (preferred) motif as well as an additional motif(s) [54,55,56•].

Recent studies demonstrate the ability of duplicated transcription factors to diversify and bind a number of different DNA motifs [55,56•]. A new study focuses on the family of fungal C_2H_2 Zn-finger transcription factors [56•]. These proteins can be divided into several 'specificity' groups, each group encompassing proteins with identical canonical DNA-recognition amino acid residues. Since these residues are the same for all proteins of the same group, these regulators are expected to bind the same DNA sequences. However, other non-canonical binding sites specific for each protein are also present. This way, each transcription factor recognizes not only a motif common to all members of the group but also its own preferred motif, which is not bound by other regulators from the same group. Such modularity of DNA binding specificity enables a protein to bind different sites while not affecting binding to the core DNA motif. This allows the transcription factor to overcome negative effects of pleiotropy. Moreover, this modularity is conserved since the divergence of *C. albicans* and *S. cerevisiae* and thus may be functionally important.

Another beautiful example of the importance of promiscuity as a promoter of evolvability of transcription factors is the LEAFY plant regulatory protein [25••], and see also recent discussion in [67,68]. LEAFY is a major regulator of flower development and cell division in land plants. It binds different types of DNA motifs in different plant species, and this specificity is determined by only a handful crucial residues. The radical shifts in DNA binding specificity of this regulator observed in the evolution from algae to land plants were enabled by passing through an intermediate, highly promiscuous form that was able to bind all three types of motifs present in modern-day plant species. Interestingly, this promiscuous state is still preserved in hornwort, which is taxonomically situated somewhere between algae and higher land plants. Together, these results illustrate how a promiscuous intermediate can also provide an evolutionary route to new functions without passing through a fitness valley.

Small-scale versus whole-genome duplication: effect of origin of duplication on regulatory network evolution

While gene duplicates clearly contribute significantly to regulatory divergence, their exact contribution has been suggested to depend on their origin. In small-scale duplications, where only one or a few genes are duplicated, transcription factors and their target genes are usually not duplicated together, resulting in immediate expression divergence [57]. After whole-genome duplications, paralogs (might) take longer to diversify, and studies have shown that their contribution to regulatory divergence is more pronounced and more prolonged [32•,48]. The ancestor of the yeast *S. cerevisiae* underwent at least one whole-genome duplication; and several studies have shown that this event could have driven the evolution of novel regulatory patterns through divergence of paralog promoters. Key examples include the evolution of the *GAL* regulatory network and the emergence of fast anaerobic growth [10,58]. Examples of the effect of *trans* mutations also exist. The regulation of ribosomal gene expression offers a key example of how whole-genome duplication created two paralogs of an ancestral transcription factor that subsequently diverged into an activator and a repressor of ribosomal genes, respectively [34].

Conclusions

Rewiring and expansion of regulatory networks is crucial throughout evolution since these events generate new expression patterns and allow proper regulation of novel

gene functions. Increasing evidence points to transposable elements as an important source of DNA-binding domains that can be recruited as transcription factors and contribute to regulatory network evolution [59,60]. Additionally, transposable elements can provide new regulatory sequences, thus introducing new regulatory interactions [61–65].

Apart from this, the examples discussed in this review clearly show that transcription factor duplication and subsequent divergence can generate new patterns of gene expression while avoiding negative pleiotropic effects inevitably associated with changes in regulatory proteins [43••,45••]. However, studies have shown that duplication is not a prerequisite for a regulator to acquire a novel function [25••,66]. Promiscuity (or bifunctionality) of the transcription factor seems to play a major role in regulatory network evolution; with or without gene duplication.

Acknowledgements

KP acknowledges financial support from TRIPLE I and a Belspo mobility grant from the Belgian Federal Science Policy Office co-funded by the Marie Curie Actions from the European Commission. KV acknowledges financial support from Fonds voor Wetenschappelijk Onderzoek (FWO) by a postdoctoral fellowship. Research in the lab of KJV is supported by European Research Council (ERC) Starting Grant 241426, Human Frontier Science (HFSP) program Grant RGP0050/2013, KU Leuven Program Financing NATAR, Vlaams Instituut voor Biotechnologie (VIB), European Molecular Biology Organization (EMBO) Young Investigator program, FWO, and Agentschap voor Innovatie door Wetenschap en Technology (IWT).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Gu X, Zhang Z, Huang W: **Rapid evolution of expression and regulatory divergences after yeast gene duplication**. *Proc Natl Acad Sci U S A* 2005, **102**:707-712.
2. Gu Z, Rifkin SA, White KP, Li WH: **Duplicate genes increase gene expression diversity within and between species**. *Nat Genet* 2004, **36**:577-579.
3. Arnoult L, Su KF, Manoel D, Minervino C, Magrina J, Gompel N, Prud'homme B: **Emergence and diversification of fly pigmentation through evolution of a gene regulatory module**. *Science* 2013, **339**:1423-1426.
4. King MC, Wilson AC: **Evolution at two levels in humans and chimpanzees**. *Science* 1975, **188**:107-116.
5. Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB: **Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene**. *Nature* 2006, **440**:1050-1053.
6. Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, Jiggins CD, Chamberlain NL, Kronforst MR, Chen R *et al.*: **optix drives the repeated convergent evolution of butterfly wing pattern mimicry**. *Science* 2011, **333**:1137-1141.
7. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, Kingsley DM: **Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks**. *Nature* 2004, **428**:717-723.
8. Wray GA: **The evolutionary significance of cis-regulatory mutations**. *Nat Rev Genet* 2007, **8**:206-216.
9. Tirosh I, Barkai N, Verstrepen KJ: **Promoter architecture and the evolvability of gene expression**. *J Biol* 2009, **8**:95.
10. Ihmels J, Bergmann S, Gerami-Nejad M, Yanai I, McClellan M, Berman J, Barkai N: **Rewiring of the yeast transcriptional network through the evolution of motif usage**. *Science* 2005, **309**:938-940.
11. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S *et al.*: **The genomic basis of adaptive evolution in threespine sticklebacks**. *Nature* 2012, **484**:55-61.
12. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA: **Structure and evolution of transcriptional regulatory networks**. *Curr Opin Struct Biol* 2004, **14**:283-291.
13. Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, Seringhaus MR, Wang LY, Gerstein M, Snyder M: **Divergence of transcription factor binding sites across related yeast species**. *Science* 2007, **317**:815-819.
14. Bradley RK, Li XY, Trapnell C, Davidson S, Pachter L, Chu HC, Tonkin LA, Biggin MD, Eisen MB: **Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related Drosophila species**. *PLoS Biol* 2010, **8**:e1000343.
15. Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD, Rolfe PA, Conboy CM, Gifford DK, Fraenkel E: **Tissue-specific transcriptional regulation has diverged significantly between human and mouse**. *Nat Genet* 2007, **39**:730-732.
16. Carroll SB: **Evolution at two levels: on genes and form**. *PLoS Biol* 2005, **3**:e245.
17. Prud'homme B, Gompel N, Carroll SB: **Emerging principles of regulatory evolution**. *Proc Natl Acad Sci U S A* 2007, **104(Suppl 1)**:8605-8612.
18. Hsia CC, McGinnis W: **Evolution of transcription factor function**. *Curr Opin Genet Dev* 2003, **13**:199-206.
19. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD *et al.*: **Natural selection on protein-coding genes in the human genome**. *Nature* 2005, **437**:1153-1157.
20. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L: **Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors**. *Nat Genet* 2003, **35**:57-64.
21. Brem RB, Yvert G, Clinton R, Kruglyak L: **Genetic dissection of transcriptional regulation in budding yeast**. *Science* 2002, **296**:752-755.
22. Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJ: **A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors**. *Cell* 2009, **138**:314-327.
23. Reece-Hoyes JS, Pons C, Diallo A, Mori A, Shrestha S, Kadreppa S, Nelson J, Diprima S, Dricot A, Lajoie BR *et al.*: **Extensive rewiring and complex evolutionary dynamics in a *C. elegans* multiparameter transcription factor network**. *Mol Cell* 2013, **51**:116-127.
24. Lynch VJ, May G, Wagner GP: **Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB**. *Nature* 2011, **480**:383-386.
25. •• Sayou C, Monniaux M, Nanao MH, Moyroud E, Brockington SF, Thevenon E, Chahtane H, Warthmann N, Melkonian M, Zhang Y *et al.*: **A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity**. *Science* 2014, **343**:645-648.
An elegant study demonstrating how promiscuity allows a single-copy transcription factor to evolve without experiencing negative effects of pleiotropy.
26. Madan Babu M, Teichmann SA, Aravind L: **Evolutionary dynamics of prokaryotic transcriptional regulatory networks**. *J Mol Biol* 2006, **358**:614-633.
27. •• Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K *et al.*:

Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 2014, **158**:1431-1443.
This exhaustive study determined the binding site preference of a plethora of transcription factors.

28. Hoekstra HE, Coyne JA: **The locus of evolution: evo devo and the genetics of adaptation**. *Evolution* 2007, **61**:995-1016.

29. Falciani F, Hausdorf B, Schroder R, Akam M, Tautz D, Denell R, Brown S: **Class 3 Hox genes in insects and the origin of zen**. *Proc Natl Acad Sci U S A* 1996, **93**:8479-8484.

30. Hernandez-Hernandez T, Martinez-Castilla LP, Alvarez-Buylla ER: **Functional diversification of B MADS-box homeotic regulators of flower development: adaptive evolution in protein–protein interaction domains after major gene duplication events**. *Mol Biol Evol* 2007, **24**:465-481.

31. Teichmann SA, Babu MM: **Gene regulatory network growth by duplication**. *Nat Genet* 2004, **36**:492-496.

32. • Thompson DA, Roy S, Chan M, Styczynsky MP, Pfiffner J, French C, Socha A, Thielke A, Napolitano S, Muller P *et al.*: **Evolutionary principles of modular gene regulation in yeasts**. *Elife (Cambridge)* 2013, **2**:e00603.
In this exhaustive study, the authors compare gene expression profiles from 15 different yeasts and show how paralogs have significantly contributed to divergence of gene expression.

33. Wapinski I, Pfeffer A, Friedman N, Regev A: **Natural history and evolutionary principles of gene duplication in fungi**. *Nature* 2007, **449**:54-61.

34. Wapinski I, Pfiffner J, French C, Socha A, Thompson DA, Regev A: **Gene duplication and the evolution of ribosomal protein gene regulation in yeast**. *Proc Natl Acad Sci U S A* 2010, **107**:5505-5510.

35. Conant GC, Wolfe KH: **Turning a hobby into a job: how duplicated genes find new functions**. *Nat Rev Genet* 2008, **9**:938-950.

36. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**:1151-1155.

37. Madan Babu M, Teichmann SA: **Evolution of transcription factors and the gene regulatory network in *Escherichia coli***. *Nucleic Acids Res* 2003, **31**:1234-1244.

38. Teichmann SA, Park J, Chothia C: **Structural assignments to the *Mycoplasma genitalium* proteins show extensive gene duplications and domain rearrangements**. *Proc Natl Acad Sci U S A* 1998, **95**:14658-14663.

39. Ohno S: *Evolution by Gene Duplication*. Berlin: Springer-Verlag; 1970, .

40. Lynch M, Force A: **The probability of duplicate gene preservation by subfunctionalization**. *Genetics* 2000, **154**:459-473.

41. Bridgham JT, Brown JE, Rodriguez-Mari A, Catchen JM, Thornton JW: **Evolution of a new function by degenerative mutation in cephalochordate steroid receptors**. *PLoS Genet* 2008, **4**:e1000191.

42. •• Perez JC, Fordyce PM, Lohse MB, Hanson-Smith V, DeRisi JL, Johnson AD: **How duplicated transcription regulators can diversify to govern the expression of nonoverlapping sets of genes**. *Genes Dev* 2014, **28**:1272-1277.
This elegant study provides insight into how a duplicated transcription factor can acquire novel targets.

43. •• Pougach K, Voet A, Kondrashov FA, Voordeckers K, Christiaens JF, Baying B, Benes V, Sakai R, Aerts J, Zhu B *et al.*: **Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network**. *Nat Commun* 2014, **5**:4868.
The authors show in molecular detail how duplication and subsequent divergence of a promiscuous transcriptional regulator allows to avoid fitness valleys associated with paralog interference and results in two dedicated regulatory networks.

44. Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ: **Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication**. *PLoS Biol* 2012, **10**:e1001446.

45. •• Baker CR, Hanson-Smith V, Johnson AD: **Following gene duplication, paralog interference constrains transcriptional circuit evolution**. *Science* 2013, **342**:104-108.
An important paper demonstrating the negative effects of paralog interference. The authors also show in molecular detail how the transcription factor paralogs subfunctionalize after the duplication.

46. Bridgham JT, Carroll SM, Thornton JW: **Evolution of hormone-receptor complexity by molecular exploitation**. *Science* 2006, **312**:97-101.

47. Carroll SM, Ortlund EA, Thornton JW: **Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor**. *PLoS Genet* 2011, **7**:e1002117.

48. Conant GC, Wolfe KH: **Functional partitioning of yeast co-expression networks after genome duplication**. *PLoS Biol* 2006, **4**:e109.

49. Brown CA, Murray AW, Verstrepen KJ: **Rapid expansion and functional divergence of subtelomeric gene families in yeasts**. *Curr Biol* 2010, **20**:895-903.

50. •• McKeown AN, Bridgham JT, Anderson DW, Murphy MN, Ortlund EA, Thornton JW: **Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module**. *Cell* 2014, **159**:58-68.
This work uses ancestral reconstruction to identify the genetic, biochemical and biophysical causes of transcription factor divergence and evolution and demonstrate the importance of negative DNA–protein interactions in this process.

51. Tzeng SR, Kalodimos CG: **Protein activity regulation by conformational entropy**. *Nature* 2012, **488**:236-240.

52. Aharoni A, Gaidukov L, Khersonsky O, Mc QGS, Roodveldt C, Tawfik DS: **The 'evolvability' of promiscuous protein functions**. *Nat Genet* 2005, **37**:73-76.

53. Khersonsky O, Tawfik DS: **Enzyme promiscuity: a mechanistic and evolutionary perspective**. *Annu Rev Biochem* 2010, **79**:471-505.

54. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X *et al.*: **Diversity and complexity in DNA recognition by transcription factors**. *Science* 2009, **324**:1720-1723.

55. Nakagawa S, Gisselbrecht SS, Rogers JM, Hartl DL, Bulyk ML: **DNA-binding specificity changes in the evolution of forkhead transcription factors**. *Proc Natl Acad Sci U S A* 2013, **110**:12349-12354.

56. • Siggers T, Reddy J, Barron B, Bulyk ML: **Diversification of transcription factor paralogs via noncanonical modularity in C_2H_2 zinc finger DNA binding**. *Mol Cell* 2014, **55**:640-648.
This recent study shows how transcription factor paralogs can overcome negative pleiotropic effect by modularly gaining novel binding sites.

57. Lynch M, Katju V: **The altered evolutionary trajectories of gene duplicates**. *Trends Genet* 2004, **20**:544-549.

58. Hittinger CT, Carroll SB: **Gene duplication and the adaptive evolution of a classic genetic switch**. *Nature* 2007, **449**:677-681.

59. Babu MM, Iyer LM, Balaji S, Aravind L: **The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons**. *Nucleic Acids Res* 2006, **34**:6505-6520.

60. Balaji S, Babu MM, Iyer LM, Aravind L: **Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains**. *Nucleic Acids Res* 2005, **33**:3994-4006.

61. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH *et al.*: **Evolution of the mammalian transcription factor binding repertoire via transposable elements**. *Genome Res* 2008, **18**:1752-1762.

62. Jacques PE, Jeyakani J, Bourque G: **The majority of primate-specific regulatory sequences are derived from transposable elements**. *PLoS Genet* 2013, **9**:e1003504.

63. Jordan IK, Rogozin IB, Glazko GV, Koonin EV: **Origin of a substantial fraction of human regulatory sequences from transposable elements**. *Trends Genet* 2003, **19**:68-72.

64. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G: **Transposable elements have rewired the core regulatory network of human embryonic stem cells**. *Nat Genet* 2010, **42**:631-634.

65. Xie D, Chen CC, Ptaszek LM, Xiao S, Cao X, Fang F, Ng HH, Lewin HA, Cowan C, Zhong S: **Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species**. *Genome Res* 2010, **20**:804-815.

66. Cheatle Jarvela AM, Brubaker L, Vedenko A, Gupta A, Armitage BA, Bulyk ML, Hinman VF: **Modular evolution of DNA-binding preference of a Tbrain transcription factor provides a mechanism for modifying gene regulatory networks**. *Mol Biol Evol* 2014, **31**:2672-2688.

67. Brockington SF, Moyroud E, Sayou C, Monniaux M, Nanao MH, Thevenon E, Chahtane H, Warthmann N, Melkonian M, Zhang Y *et al.*: **Evolution. Response to Comment on ''A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity''**. *Science* 2015, **347**:621.

68. Brunkard JO, Runkel AM, Zambryski PC: **Evolution. Comment on ''A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity''**. *Science* 2015, **347**:621.