From elements to modules: regulatory evolution in Ascomycota fungi
Dana J Wohlbach¹,*, Dawn Anne Thompson²,*, Audrey P Gasch¹ and Aviv Regev²,³

Regulatory divergence is likely a major driving force in evolution. Comparative transcriptomics provides a new glimpse into the evolution of gene regulation. Ascomycota fungi are uniquely suited among eukaryotes for studies of regulatory evolution, because of broad phylogenetic scope, many sequenced genomes, and facility of genomic analysis. Here we review the substantial divergence in gene expression in Ascomycota and how this is reconciled with the modular organization of transcriptional networks. We show that flexibility and redundancy in both cis-regulation and trans-regulation can lead to changes from altered expression of single genes to wholesale rewiring of regulatory modules. Redundancy thus emerges as a major driving force facilitating expression divergence while preserving the coherent functional organization of a transcriptional response.

Addresses
¹ Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706, United States
² Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 01242, United States
³ Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 01240, United States

* These authors contributed equally to this work.

Corresponding author: Gasch, Audrey P (agasch@wisc.edu) and Regev, Aviv (aregev@broad.mit.edu)

Introduction
The incredible diversity of living creatures defies their similarity in protein sequence and gene content. King and Wilson [1] first proposed that organismal diversity is likely driven by regulatory differences controlling when, where, and how genetic material is expressed. Nearly 35 years later, although examples of regulatory divergence are known in a wide range of species [2] including bacteria [3], fungi [4], flies [5], and mammals [6], the mechanisms through which regulatory systems evolve are still poorly understood. In recent years, comparative genomics approaches have allowed us to identify the functional components of genomes and to trace evolutionary events at different time scales [7,8]. These approaches are also being used to infer the evolution of gene-expression regulation through two general approaches: characterization of cis-regulatory elements in orthologous promoter sequences, and comparative analysis of mRNA profiles across organisms. While studies relying on sequence data are more prevalent, functional studies of comparative gene regulation are now starting to shed light on how genome evolution is linked to functional changes.

Among eukaryotes, the Ascomycota fungi (Figure 1a) are particularly suitable for studies of eukaryotic regulatory evolution. More than 100 genome sequences exist for individual strains within Saccharomyces species and across fungi spanning hundreds of millions of years of evolution. They include model organisms (S. cerevisiae and S. pombe) and important human and agricultural pathogens (e.g. C. albicans, Aspergilli, and F. graminarium). Furthermore, sequenced genomes include species that diverged before and after a whole genome duplication (WGD, Figure 1a), which occurred ~150 million years ago (mya) [9,10]. This provides a unique opportunity to explore the effect of gene duplication on regulatory divergence. Finally, the relatively small (ca. 9–100 Mb) genomes are computationally tractable, but still display many of the hallmarks of eukaryotic gene regulation.

In this review, we focus on recent advances made in understanding the evolution of gene regulation in Ascomycota, from micro-evolutionary scales (<5 million years and typically within species) to macro-evolutionary time frames over tens of millions of years involving extensive speciation. (It is estimated that S. cerevisiae can undergo ~2900 generations per year [11].) We examine conservation and divergence from two different perspectives: the regulation of mRNA expression at the level of single genes, and the coordinated expression of genes in regulons or ‘modules’ (i.e. sets of coregulated target genes) within a network. As we show, flexibility and apparent redundancy in gene copies, functional elements and molecular interactions may play a major role in driving the emergence of regulatory divergence, while conserving the functional backbone of transcriptional responses.
Functional and mechanistic corollaries of expression conservation and divergence

Available compendia of genome-wide mRNA profiles have allowed direct comparison of orthologous mRNA expression patterns across species. Large datasets exist for model *Ascomycota* (*S. cerevisiae*, *S. pombe*, and *C. albicans*), and smaller datasets are available for other fungi in the phylum (e.g. other *sensu stricto* Saccharomycoses [12] *C. glabrata* [13], *K. lactis* [14], and some *Euascomycota* [15,16]) as well as different strains within *S. cerevisiae* [12,17–20]. Together, these afford a global assessment of expression divergence (ED) [12] — a quantitative measure of the differences in the expression of a pair of orthologs between two species.

An emerging theme is that different classes of genes have different propensities for ED [20,21,22**]. Most notably, genes with high-ED versus low-ED are distinguished by a broad functional dichotomy: genes with conserved expression are typically involved in growth and general metabolism, whereas those with divergent expression are often subtelomeric, responsive to external and internal signals (e.g. stress) and are nonessential. This dichotomy in divergence between species [22**,23] is also mirrored when measuring variation between isogenic cells in a yeast culture [24] or between genetic variants of *S. cerevisiae* [18,20,25].

ED is also correlated to distinct promoter architectures. The promoters of genes with conserved expression (low-ED) lack upstream TATA elements, contain well-positioned nucleosomes, and display a substantial nucleosome-depleted region where most regulatory elements reside. Conversely, promoters of high-ED genes contain TATA elements and display distributed nucleosomes, which are regulated through dynamic chromatin remodeling [12,26**,27]. Promoters of high-ED genes are also associated with the presence of unstable tandem repeats proposed to drive changes in nucleosome organization and gene expression [28].

That specific genes tend to show large versus small expression differences both within clonal populations and across species suggests a higher tolerance for expression differences at particular gene classes, and thus a higher prevalence of evolved expression. Indeed, ED levels are correlated with the level of constraint measured on steady-state expression across *S. cerevisiae* strains (KH Eng *et al*., submitted for publication). It is intriguing that genes belonging to particular functional classes contain promoters linked to high or low ED. Indeed, variable expression of certain genes within a population (e.g. stress defense genes) may have been directly selected for through modification of promoter structure [26**].

**Gene duplication facilitates regulatory neo-functionalization**

Gene duplication may provide a unique opportunity for ED of at least one of the two paralogs [29,30,31**]. In a comprehensive study of *S. cerevisiae* paralogs whose origins range to the last common ancestor with *S. pombe*...
(Figure 1a, root) we found surprisingly little divergence of the molecular function of paralogs, but substantial (~70%) divergence in gene regulation, as reflected in the cis-regulatory elements, the transcription factors (TFs) bound to the genes’ promoters, and the gene regulon to which they belong [31]. This is consistent with the idea that regulatory divergence occurs at an elevated rate compared to divergence of the coding sequence of paralogs.

Two likely scenarios can underlie diverged expression of paralogs. Regulatory subfunctionalization occurs when multiple, distinct regulatory elements controlling expression of the single ancestral gene are ‘split’ between its two descendants, such that each paralog retains only some of the regulatory inputs. In contrast, regulatory neo-functionalization occurs when one paralog evolves a new control not used by the ancestral gene. A recent study [32] proposed neo-functionalization as the dominant mode of diverged expression of S. cerevisiae paralogs, since often the expression pattern of only one paralog was distinct from that of the single preduplication ortholog in C. albicans.

ED of paralogs may have an important adaptive role. First, regulatory divergence can effectively lead to sub-functionalization of paralogs, even if paralogs have the same biochemical function (but serve different roles because of expression differences). This may have occurred in the post-WGD regulatory divergence of glycolysis/gluconeogenesis enzymes [33], such as the hexose kinases, Hxk2 and Hxk1, that share hexokinase activity but display distinct expression patterns. Notably, sub-functionalization may be partial and remnants of the shared (joint) ancestral control may still allow paralogs to ‘backup’ each other [34]. Finally, constraints on expression levels could in turn influence the evolution of gene copy number [29,31], since low-ED genes also exhibit few duplication and loss events, whereas high-ED genes have substantial variation in copy number between species.

Flexibility in regulatory mechanisms can drive expression divergence

Genetic changes in both cis and trans elements can contribute to ED (Figure 1b). A genetic change can affect expression in cis, either directly by altering regulatory sequences controlling gene expression, or indirectly by modifying the activity of the gene’s product and consequently affecting expression through feedback [35]. Polymorphisms in cis appear to contribute most to ED in phylogenetically close species,

Alternative modes for the regulatory evolution of transcriptional modules. Each panel shows a distinct scenario of the inferred evolution of an ancestral program (at tree root) into programs observed in extant species (leaves). (a) Conservation of both cis-elements (boxes) and TFs (ovoids). (b) Elaboration of a program by the emergence of a new cis-element in addition to the ancestral one. (c) Divergence of a program by the loss of cis-element. (d–f) Divergence of a program through cis-redundancy or trans-redundancy. (d) In cis, a switch from one cis-element to another through an intermediate that has both sites, bound by distinct factors. (e) In trans, two paralogous TFs bind the same site following gene duplication. After sub-functionalization, only one of the factors controls the module. (f) In trans, switching of one TF for another that does not share ancestry, but can bind a similar site.
independent of environmental factors [22**]. Alternatively, a polymorphism distant to the affected gene can affect its expression in trans, by affecting its regulators directly (e.g. a polymorphism in a regulatory protein), or indirectly by altering physiology that provokes an expression response despite perfect conservation of the direct regulatory system [36,37].

Gain and loss of cis-regulatory motifs, and the potential for corresponding changes in TF binding, can occur on relatively short time scales (<5–20 million years), both within [38] and between species [22**;39,40**;41**]. Of the lineage-specific binding-site losses within sensu stricto Saccharomyces, over half may correspond to newly emerged binding sites in the same regulatory regions [39]. Such binding-site turnover — where the appearance of a new binding site allows loss of the other — explains how expression can be conserved despite plasticity in regulatory sequences.

In other cases the apparent loss of a binding site, without concomitant turnover, corresponds to a change in gene-expression pattern, and ultimately loss of TF control (Figures 2b and 3) [39,40**]. For example, ChIP–chip studies of two TFs, Ste12p and Tec1p, in three sensu stricto Saccharomyces species (~20 mya) [7,40**] show that only 20% of TF binding events were conserved across all three species. Similar conservation (15%) was observed for the binding of Mcm1 between more distant species (S. cerevisiae, K. lactis, and C. albicans) [41**]. In many cases the loss of TF binding correlated with the loss of the binding site, but in some cases TF binding occurred despite absence of a discernible DNA motif [41**]. Such ‘promiscuous’ binding to noncanonical sites followed by selection for optimal binding may facilitate the acquisition of new targets [42].

Expression divergence in transcriptional regulatory networks
It is challenging to reconcile the substantial evolutionary diversity in the expression of individual genes with the
functional organization of regulatory networks. In particular, it is well established that transcriptional modules (regulons of coregulated genes) play a central role in regulatory networks [43–45]. Various modules are conserved across organisms from *E. coli* to humans [46,47], including fungi [21]. However, if the regulation of individual genes is highly evolvable, how are multiple evolutionary events coordinated to sustain conservation of coexpression in regulons? One possibility is that changes occur predominantly in *trans* (e.g. divergent TF responsiveness but conserved *cis*-regulatory elements), thus conserving coexpression (module identity) while diverging module expression. However, several cases demonstrate dramatic differences in module composition or its associated *cis*-elements. As we show below, multiple forms of functional redundancy allow for complex divergence of regulatory mechanism while maintaining coregulation within modules.

Conservation of regulatory modules over long evolutionary timeframes

Many regulatory modules are conserved across *Ascomycota*, often associated with conserved *cis*-regulatory elements and TFs (Figure 2a). Conserved modules can be identified by a statistically significant overlap in orthologous genes between modules of genes with correlated expression within each species [21,48]. Alternatively, we can identify conserved regulation in gene modules [4] by comparing *cis*-regulatory elements enriched in the promoters of sets of orthologous coregulated genes. Three-quarters of the *cis*-regulatory elements associated with recognizable modules are conserved by the latter criterion between *S. cerevisiae* and species that emerged around the WGD (~150 mya). Nearly one-third of these modules are conserved out to *C. albicans* (~230 mya), and several are conserved in *S. pombe* (>300 mya) [4,21].

These computational results have been confirmed experimentally, showing conserved regulatory connections for a TF, its cognate *cis*-element, and a large fraction of orthologous targets. Examples include the Gcn4/Cpc1 TFs and amino acid responsive genes [16,49], AP-1 factors and oxidative-stress defense genes [13,50,51] and several TFs and their cell-cycle-regulated targets [41**,52,53]. Features of binding-site organization within the promoter are often conserved as well [4,21,41**,54], suggesting that regulatory mechanisms and combinatorial regulation can also be under purifying selection.

Reshaping modules by gain and loss of gene targets

That orthologous regulatory modules can be detected over long time frames does not preclude substantial regulatory evolution. First, coexpression of genes in a module can collectively evolve simply by altering TF activation in response to different environmental cues and upstream signals. Second, *cis*-regulatory changes can drive gain and loss of targets to affect module composition, as well as regulatory patterns. Although the underlying mechanism is the same, the effects of target gain and loss can appear very different depending on the time scale. Over short timescales (5–20 million years), this can appear as fine-tuning of module members (because of neutral or selective forces), as suggested for Ste12/Tec1 targets [39,40**] (Figure 3a). Selection over longer time may promote the gain or loss of entire functional sub-modules, thus splitting or merging modules to create higher order coordination [55] (Figure 3b). This may have occurred in the evolution of mitochondrial and cytoplasmic ribosomal-protein (mRP or cRP) gene expression [55]. mRPs and cRPs were proposed to be ancestrally regulated as a single module through the Rapid Growth Elements (RGE, AATTTT), but the module apparently split in post-WGD species when mRPs lost the RGE site, leading to decoupling of mRP and cRP expression in species that can decouple respiration from growth.

Conversely, the gain of new targets by an existing TF, whether by drift or selection, can lead to the cooption of an existing regulatory module under more elaborate control, to the ‘birth’ of a new module of coregulated genes, or to a wholesale change in a target regulon, such that there is little or no overlap in the targets of orthologous regulators (Figure 3c,d). Such dramatic changes in regulatory systems are typically observed over longer time scales (~150 to over 300 million years). For example, in *C. albicans*, Mcm1 control, along with combinatorial regulation by Wor1, emerged in a novel host-adaptation module [41**], whereas the Sko1 TF was coopted to regulate a cell-wall biogenesis module, in addition to ancestral targets involved in stress [56]. Such cooption may lead to higher order coordination between modules, and is likely related to adaptation of *C. albicans* to its human host. In more extreme cases, drift in TF targets can lead to a full switch in regulated genes, along with coevolved differences in environmental responsiveness. For example, sometime after the divergence of the lineages leading to *C. albicans* and *S. cerevisiae*, the Gal4 TF acquired galactose metabolism (GAL) genes under its control, lost the ancestral glycolysis and subtelomeric targets still seen in *C. albicans*, and acquired responsiveness to galactose [49,57*,58,59], resulting in a complete functional and mechanistic switch.

The role of redundancy in mediating regulatory rewiring of conserved modules

The formation of ‘seemingly redundant’ regulatory mechanisms may facilitate the dramatic rewiring of regulatory mechanisms while maintaining gene coregulation. In several cases, wholesale rewiring is observed between species diverged 200–300 mya, and appears to have been preceded by a period of redundant regulation on the order of 100–150 mya.
One prominent mode of redundancy is the formation of a module under the control of multiple regulatory systems, through distinct cis-regulatory sites (Figure 2d). In this ‘regulatory switching’ scenario, an ancient regulatory program is augmented by the appearance of a new set of distinct cis-regulatory elements, followed by the loss of one system’s sites from the target gene promoters. For example, the switch in the regulation of the RP genes from the Homol-D system in S. pombe to the Rap1 system in S. cerevisiae, may have occurred through the formation of an intermediate, where both sites appear in close proximity in RP promoters, as is still observed in several extant species from A. gossypii to C. glabrata (100–150 mya) [4,21,60**]. A similar redundancy is observed in the regulation of the GAL module [58] and may have contributed to a switch from one cis-regulatory element to another in the control of the mating module [54]. Notably, in species whose module harbors both regulatory controls, they may have distinct roles, for example in regulating distinct expression levels, while being redundant for coexpression.

From the trans side, the presence of multiple TFs capable of binding similar sites can lead to rewiring. For example, redundant binding by two paralogous TFs can result in subfunctionalization of one paralog and its subsequent loss of control of a regulatory system (Figure 2e). This has occurred for the paralogous TFs Mcm1 and Arg80, each of which regulates a subset of the ancestral protein’s targets [41**,61]. More generally, many TFs are derived from a few broad families, resulting in overlapping specificities and the propensity for ‘redundant’ regulation through the same elements (Figure 2f). This may have contributed to the replacement of Gcn4 by another bZIP family member as the regulator of the amino acid metabolism module in S. pombe [21]. Finally, relative promiscuity in binding similar sites can lead to switching from one TF to another even when they do not share ancestry, as has been suggested for the switch from the ancestral a2 activator (seen in C. albicans) to the derived alpha2 repressor (in S. cerevisiae) in the control of the mating module [62].

Flexibility and redundancy in cooperative and combinatorial interactions between distinct TFs can also facilitate rewiring [63]. For example, flexibility in the required components for a multi-TF complex may underlie the cooption of Hap4 TF to regulate iron uptake in S. pombe [64]. In certain cases, this flexibility involves an elaborate interplay of trans-factors and cis-factors. The transition from a2 to alpha2 regulation in the mating module may have been facilitated through interactions of both factors with the conserved cofactor Mcm1 [54,62]. Interestingly, evolved Mcm1–DNA interactions at elements with high flanking AT content in S. cerevisiae may have alleviated the need for a2-dependent activation, allowing for loss of a2 regulation at these genes [54]. Since regulation through combinatorial and cooperative control by a larger number of TFs may be associated with greater binding flexibility, combined cis–trans redundancy may be a major driving force for regulatory evolution.

**Future prospects**

While great advances have been made in the phenomenology and mechanistic understanding of the evolution of gene regulation, the relative roles of neutral drift and selective forces in promoting divergence remain largely unknown. For individual genes, purifying selection can be effectively invoked for the conservation of sites in closely related species, and for the low-ED of genes involved in growth processes. However, it is unclear how much of the increased divergence of high-ED genes is due to direct positive selection versus neutral drift and relaxed constraint. For gene modules, purifying selection may act to conserve coexpression and coregulation of certain regulons. However, the repeated divergence of the regulatory mechanisms controlling other conserved modules (e.g. RPs) suggests the potential adaptive importance of conserving coexpression while diverging how a module is coupled to its inputs. Identifying these selective forces is highly challenging. Understanding how changes in the direct control of modules are related to upstream changes is highly challenging. Understanding how changes in the direct control of modules are related to upstream changes in signal sensing [65] may shed light on this question.

The major promise of *Ascomycota* for studies of evolution of gene expression is in the facility of experiments in both model and nonmodel organisms in this phylogeny. In *S. cerevisiae*, experimental evolution and direct engineering [66,67] can highlight the effect of well-defined selective pressures on regulatory evolution. At a broader phylogenetic scope, recent studies comparing expression or *in vivo* TF binding profiles in two or three organisms have shown early promise. A critical next step is to expand the experimental scope to cover a broader phylogenetic range and density, thus allowing us to directly study the divergence of expression in individual genes and gene modules.

**Acknowledgements**

The authors thank Sigrid Hart for graphics included in the figures. DJW was supported by an NLM training grant 5T15LM007359. DAT was supported by a Human Frontiers Science Program Research Grant. APG was supported by an NSF CAREER award (#0447887) and NIGMS R01GM083989-01. AR was supported by HHMI, by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund, by an NIH Pioneer Award and by the Sloan Foundation.

**References and recommended reading**

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

- of special interest
- **of outstanding interest**


Using allele-specific gene-expression profiling in two yeast species and their hybrid, the authors confirm that the majority of ED between species is due to cis changes, but that the majority of hybrid-specific expression occurs through trans effects. Additionally, they show that parental genom-omes accumulate compensating cis-trans mutations, suggesting a mechanism for the emergence of novel hybrid phenotypes.


Using a novel approach to assigning gene orthology, the authors trace the evolutionary history of gene duplications and losses across 17 Ascomy-cota fungi. The study reveals new insights into the consequences of gene amplification and loss, demonstrating that the propensity for dynamic gene copy number varies by gene functional group, essentially, and transcription regulation. The authors also discuss the inferred conse-quences of gene duplication on the evolution of biochemical function and gene-expression regulation.


Using ChIP-chip, the authors examined the binding of two transcription factors in three *sensu stricto* Saccharomycetes and found a surprising amount of cis-regulatory diversity between closely related species. In only about 20% of cases did a transcription factor bind to the same site in all three species, supporting a major role for cis-regulatory mutations in evolution.


Using ChIP-chip and fungal comparative genomics, the authors identify gene-expression modules dependent on Mcm1 and its transcriptional cofactors in three hemiascomycetes. The work characterizes lineage-specific rates of target gene gain and loss and demonstrates species-specific differences in binding-site architecture.


In this study, Hittinger and Carroll examined the evolution of a pair of duplicated *S. cerevisiae* genes, GAL1 and GAL3, by testing fitness differences resulting from differing gene regulation in *K. lactis*, which has not undergone the gene duplication. They find evidence supporting subfunctionalization of the ancestral gene following duplication.


The authors examined the regulation of ribosomal-protein gene expression in *C. albicans*, using ChIP-chip to define the Tbf1/Cbf1 regulon and comparative genomics to chronicle the switch in RP regulation between *C. albicans* and other fungi.


