Final report on a postdoctoral project, funded by BELSPO.

Project name: "Emergence of a novel regulatory network through duplication of a transcription factor and its target gene"

Beneficiary: Ksenia Pugach

Promotors: Patrick Van Dijck

Host institute: KULeuven

State of the art.

It is generally accepted that gene duplication is one of the main sources of new genes in genomes (Conant and Wolfe, 2008; Zhang, 2003). Genes duplicate remarkably often and it is estimated that at least 50% of prokaryotic genes and over 90% of eukaryotic genes are products of gene duplication (Teichmann and Babu, 2004). Newly duplicated genes can acquire mutations and develop new functions(Chen et al., 2010). Emergence of the new gene copy with a different novel function might require reprogramming or expansion of an existing regulatory network to ensure that both paralogs are expressed at appropriate times and in response to appropriate signals(Gu et al., 2004). Gu and coworkers suggested a model of asymmetrical regulatory evolution of paralog genes after the duplication: regulation of one gene copy evolves rapidly, while the other copy retains the ancestral expression profile (Gu et al., 2005). In keep with this theory, many studies show that two copies of a duplicated gene often become differentially regulated within a short period after the duplication (Gu et al., 2005; Thompson et al., 2013; Tirosh and Barkai, 2007).

While the number of studies demonstrating divergent transcriptional regulation of paralog genes is increasing, only few studies have investigated the molecular details underlying this divergence. Three possible scenarios of gene regulation after duplication of the target gene were suggested (Teichmann and Babu, 2004). In first scenario (inheritance), both copies of the gene can stay under a regulation of the same transcription factor. Or alternatively, gene can become a part of another regulatory network and be regulated by a different transcription factor in response to a different signal (loss and gain). A third scenario therefore involves the generation of a new regulatory cascade by duplication and functional divergence of an existing transcription factor, so that each of the two paralog target genes becomes regulated by one of the two newly duplicated transcription factors.

To investigate these scenarios we focus on the *MAL* gene family of the eukaryotic model organism *Saccharomyces cerevisiae*. MAL genes encode proteins necessary to import and metabolize maltose and similar disaccharides (such as isomaltose, palatinose and other α-glicosides). Maltose enters the

cells via active transport through the MalT transporters. Once inside the cell, it is hydrolyzed into two glucose units by the MalS maltases. Some of the intracellular maltose presumably binds the MalR regulators. The MalR-maltose complex activates the expression of the *MALS* and *MALT* as long as maltose is present in the system (Chow et al., 1989).

We have previously shown that the *MALS* genes in *Saccharomyces cerevisiae* underwent several duplication events, with some of the paralogs gaining a novel activity towards α1->6 glycosidic bonds (found for example in isomaltose and palatinose), while other MalS paralogs retained the ancestral preference for α1->4 glycosidic bonds (found for example in maltose) (Brown et al., 2010; Voordeckers et al., 2012). Therefore, *MAL* gene regulatory network represents a perfect model system to study evolution of the regulatory network as a response to the duplication and neofunctionalization of its target genes.

Objectives.

Our main goal is to establish the changes in *MAL* gene regulatory network following the duplication and neofunctionalization of *MALS* target genes and discover the molecular detail underlying these changes.

Results.

First question we set out to answer was if the maltose- and isomaltose-specific *MALS* paralogs are differentially regulated or these genes are expressed in presence of both types of disaccharides. Using fluorescently tagged target genes we found that expression of different *MALS* genes is regulated specifically by the sugar they show activity for (Fig.1). This specific regulation of the two paralog gene groups became possible because the central transcription factor MalR that regulated the ancestral *MALS* gene was also repeatedly duplicated. One of the resulting MalR regulators (Yfl052w) activates the expression of MalS enzymes with novel isomaltase activity, while other *MALR* paralog (Malx3) retained control over maltose-specific target genes (Fig.1 and Fig.2).

The absence of crosstalk between the two regulatory networks suggests that the maltose- and palatinose-specific MalR regulators bind different DNA binding sites. To test this, we determined the DNA binding sites of the palatinose-specific regulator Yfl052w using the ChIP-exo technique, and compared these to the known binding sites of the maltose-specific regulator Malx3 (Fig.3). The ChIP-exo analysis supports the results reported in Fig. 1 and Fig.2 and indicates that both transcription factors bind different sites. Specifically, when the α 1-6 disaccharide palatinose is present, Yfl052w binds the promoter regions of palatinose-specific genes (*IMA1*, *IMA5*, *YFL052W*) and the *MAL11*

promiscuous transporter, but not the promoters of maltose–specific genes (Fig. 4). Instead, these maltose-specific genes are known to be bound by Malx3 in the presence of maltose.

We found that a surprisingly small number of single-nucleotide mutations are sufficient to ensure that the factor specifically activates its target promoters, without interfering with the regulation of non-target paralog promoters. Specifically, for Yfl052w to distinguish between promoters of its target genes and promoters of Malx3 target genes only two specific C to G substitutions in the binding site and one Arg12Cys substitution in the DNA-binding domain are required (Fig.5 and Fig.6).

Number of the MalR binding sites also plays a role in minimization of the cross-talk between two networks. Maltose-specific MalX3 requires presence of several binding sites to activate its target gene, thus preventing it from activating the palatinose-specific target genes, which bear only one MalR binding site in their promoter region.

The availability of many whole-genome sequences across the yeast lineage allowed us to investigate the evolution of the *MAL* gene family and establish a likely evolutionary path from the ancestral circuit to a present day state. We have previously shown that the common ancestor of today's yeast species likely only had one copy of each of the three types of *MAL* genes (*MALS*, *MALT* and *MALR*). In some species, including *S. cerevisiae*, the *MAL* genes underwent several duplication events (Brown et al., 2010; Voordeckers et al., 2012). In other species, like *L. elongisporus*, the *MAL* genes were not duplicated and the ancient, simple three-gene network seems to be preserved. Some species, like *S. bayanus*, represent an intermediate state, where the target genes have already duplicated and acquired a novel function, but the regulatory *MALR* gene is still present in a single copy.

Based on our results we suggested a model of *MAL* gene network evolution as depicted in Fig. 7. In this model, duplication and functional divergence of the *MALS* genes already happened in the common ancestor of *K. thermotolerans* and *S. cerevisiae* (Fig. 7, events 1and 2), but these genes were still controlled by one promiscuous MalR regulator that resembled today's *S. bayanus* Malx3 protein (which has an Arg residue in position 12). Similar to the present-day Malx3 regulator, the ancestral regulator was able to bind both CGG and CGC-motifs and induced the expression of both maltose- and palatinose-specific genes in presence of both types of sugars. The promoters of these genes probably did not yet diverge, with both CGC and CGG motifs present upstream of maltose as well as palatinose-specific genes similar to the genome of present day *S. bayanus*. Yfl052w-like regulators most probably first appeared later in the evolution as a result of duplication (Fig. 7, event 3) and subsequent mutations, including the key Arg12Cys mutation (Fig. 7, event 4). This mutated Yfl052w-like paralog can no longer bind the CGC motifs, which are selected for in the promoters of maltose-specific genes. By contrast, the Malx3-like regulator evolves a weaker activity, so that it loses the

ability to activate expression of palatinose-specific genes, which are selected to have only one CGG-containing binding site, while retaining the ability to activate maltose-specific promoters that contain three CGC binding sites (Fig. 7, events 5 and 6).

Conclusions.

Several studies have investigated the regulatory divergence between species on a genome-wide level. Together, these studies show that changes in gene regulation occur frequently and are important drivers of functional and morphological evolution. This is especially true for the evolution of the regulation of newly duplicated genes. Since paralogs often evolve different functions, these functionally diverged duplicates may need to be regulated independently. However, despite the importance of the evolution and divergence of gene regulation, the exact molecular mechanisms and mutational pathways that lead to the emergence of such novel regulatory networks remain largely unknown.

Our results show how duplication of a p romiscuous transcription factor and its target genes led to the development of two separate regulatory networks, with one paralog of the transcription factor regulating a set of target genes involved in maltose uptake and metabolism, and another regulating target genes responsible for palatinose consumption. Specifically, we find that only two point mutations in the promoter regions of the target genes, combined with two single-nucleotide mutations in the DNA-binding domain of the transcription factor paralogs are sufficient to ensure that each transcription factor paralog specifically activates its target promoters, without interfering with the regulation of the target genes of the other paralog. Our results provide, for the first time and in great detail, a definitive experimental evidence for the role of duplication of the ancestral transcription factor in emergence of the novel regulatory network.

While the predominant opinion in the field is that evolution on the regulatory level precedes the actual changes in the protein sequence of the target genes, our data indicate that the opposite is also possible. It seems likely that the pre-duplication ancestral *MAL* gene regulatory network was very simple and resembled the network in present-day *L. elongisporus* (Fig.7). The *L. elongisporus* MalR regulator is promiscuous and activates expression of a (bifunctional) MalS hydrolase and a transporter in response to either maltose or palatinose. Several duplication events of *MALS* genes followed by optimization of either maltase or palatinase activity in different paralogs led to emergence of two functional classes of MalS hydrolases in *S. cerevisiae*. Interestingly, our analyses suggest that the specialization of palatinose-specific MalR regulators and the separation of the two regulatory networks likely occurred after the neofunctionalization of *MALS* target genes, around the branching of the *S. bayanus* and *S. cerevisiae* clades (Fig.7). The functional divergence of the MalS

enzymes generated a situation where it became beneficial for the cells to regulate each of the MalS enzymes separately, so that each enzyme is only activated by its proper substrate and paralog interference is avoided. In keeping with this hypothesis, we have previously shown that activation of *MAL* genes in conditions where they are not required comes at a considerable fitness cost.

This study has been published in Nature Communications journal (a Nature open access journal IF=10,75)

References.

Brown, C.A., Murray, A.W., and Verstrepen, K.J. (2010). Rapid expansion and functional divergence of subtelomeric gene families in yeasts. Curr. Biol. *20*, 895–903.

Chen, S., Zhang, Y.E., and Long, M. (2010). New genes in Drosophila quickly become essential. Science *330*, 1682–1685.

Chow, T.H., Sollitti, P., and Marmur, J. (1989). Structure of the multigene family of MAL loci in Saccharomyces. Mol. Gen. Genet. *217*, 60–69.

Conant, G.C., and Wolfe, K.H. (2008). Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. *9*, 938–950.

Gu, X., Zhang, Z., and Huang, W. (2005). Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proc. Natl. Acad. Sci. U. S. A. *102*, 707–712.

Gu, Z., Rifkin, S.A., White, K.P., and Li, W.-H. (2004). Duplicate genes increase gene expression diversity within and between species. Nat. Genet. *36*, 577–579.

Teichmann, S. a, and Babu, M.M. (2004). Gene regulatory network growth by duplication. Nat. Genet. *36*, 492–496.

Thompson, D. a, Roy, S., Chan, M., Styczynsky, M.P., Pfiffner, J., French, C., Socha, A., Thielke, A., Napolitano, S., Muller, P., et al. (2013). Evolutionary principles of modular gene regulation in yeasts. Elife *2*, e00603.

Tirosh, I., and Barkai, N. (2007). Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. Genome Biol. *8*, R50.

Voordeckers, K., Brown, C. a, Vanneste, K., van der Zande, E., Voet, A., Maere, S., and Verstrepen, K.J. (2012). Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. PLoS Biol. *10*, e1001446.

Zhang, J. (2003). Evolution by gene duplication: an update. Trends Ecol. Evol. *18*, 292–298.

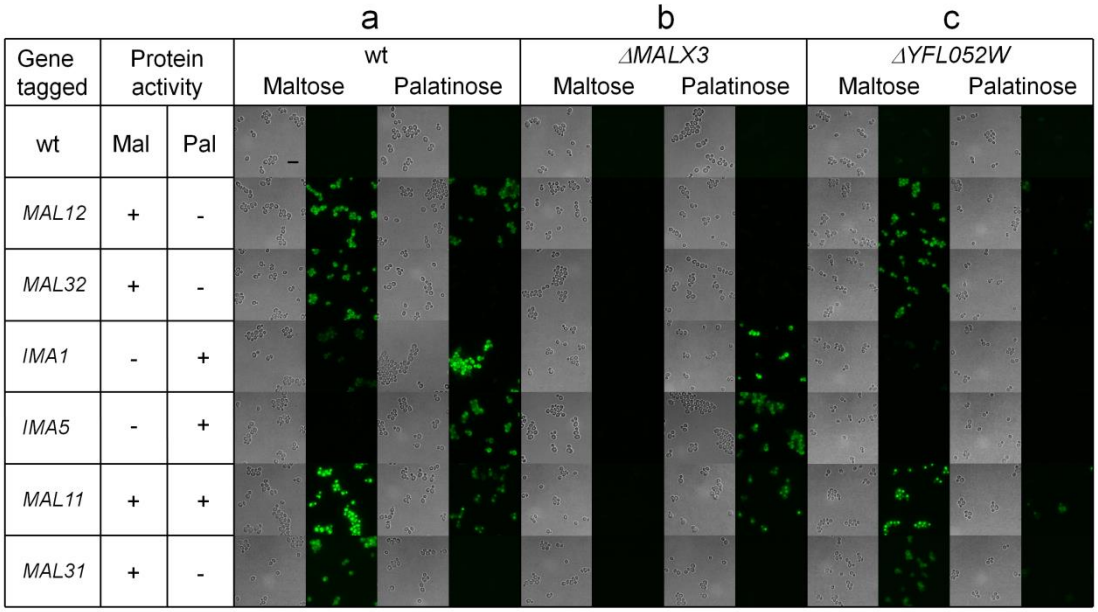

Figure 1. Maltose and isomaltose-specific genes are differentially regulated. Representative brightfield and fluorescence microscopy images of yeast cells with various *MALS* or *MALT* genes fluorescently tagged are shown for wt cells (**a**), and strains carrying deletions of genes encoding transcriptional regulators (panel **b:** *MALX3*, panel **c:** *YFL052W*). Cells were grown in presence of either palatinose (α 1-6 disaccharide) or maltose (α 1-4 disaccharide) as indicated above the pictures. Gene names are listed in the first column, and protein activities towards the two types of sugars (maltose or palatinose) are indicated in the second and third columns. Scale bar is included in the upper left image and equals 10 μm. The experiment was repeated at least three times.

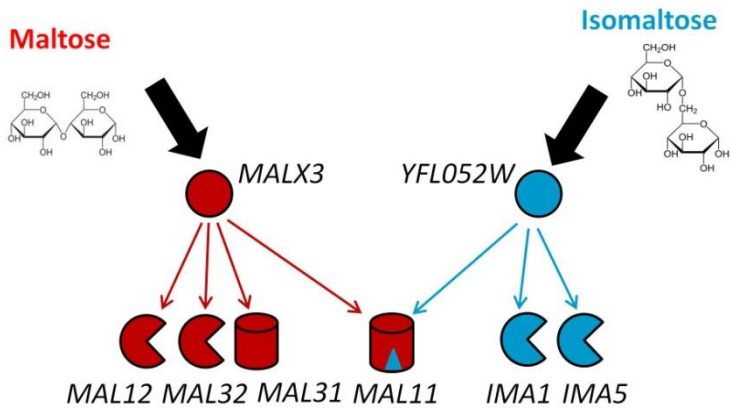

Figure 2. Regulatory networks for maltose and isomaltose uptake. *MALR* genes are depicted in a shape of a circle; *MALT* genes are represented as barrels, *MALS* genes – three-quarter pies. Blue color indicates specificity of the corresponding protein to the palatinose, red – to maltose.

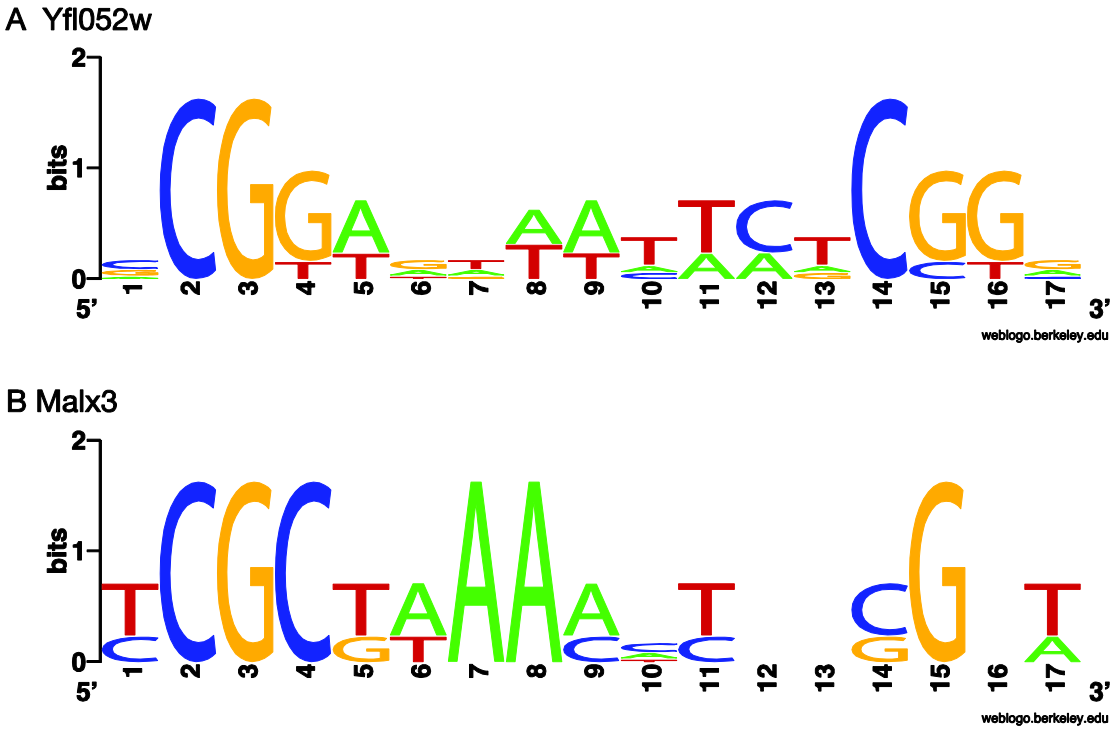

Figure 3. Different DNA-binding specificity of different MalR transcription factors. (**a**) Sequence logo of Yfl052w DNA binding site CGG(9N)CGG. (**b**) Sequence logo of Malx3 DNA binding site CGC(9N)CGN.

Figure 4. ChIP-exo analysis reveals Yfl052w binding sites. Raw ChIP-exo reads from one of the two duplicated experiments are shown in IGV viewer for each gene of interest, scaled [min=0, max=25000 reads]. Chromosomal coordinate of each Yfl052w binding site coming from the comprehensive bioinformatic analysis of peak pairs is indicated on top of each peak. The promoter region of each gene of interest is encircled in red. Note that this figure shows the mapped raw reads. Further data processing enables precise identification of the binding motifs.

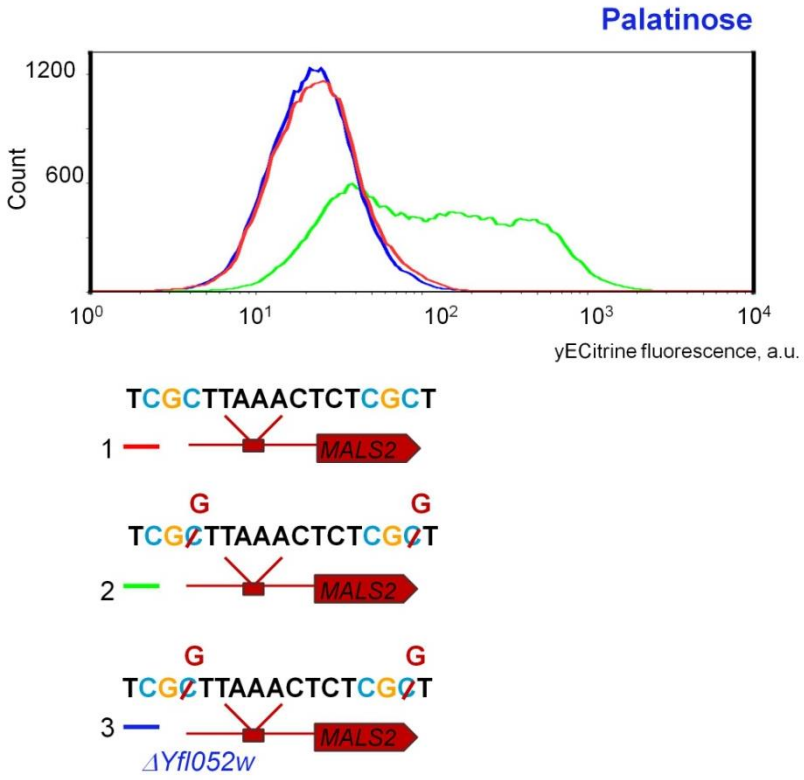

Figure 5. Yfl052w and Malx3 regulators show different DNA-binding specificity. Two point mutations in a maltose-inducible promoter yield a palatinose-inducible promoter. (1) Histogram of the fluorescence signal of a strain with a yECitrine-tagged *MAL32* gene. This maltose-specific reporter gene shows no expression in palatinose and can be used to estimate the background fluorescence levels. (2) Single nucleotide C to G substitution in both CGC motifs in the upstream Malx3 binding site of the of *MAL32* promoter leads to expression of this gene in palatinose. (3) Deletion of palatinose-specific regulator Yfl052w abolishes the expression of the mutant promoter.

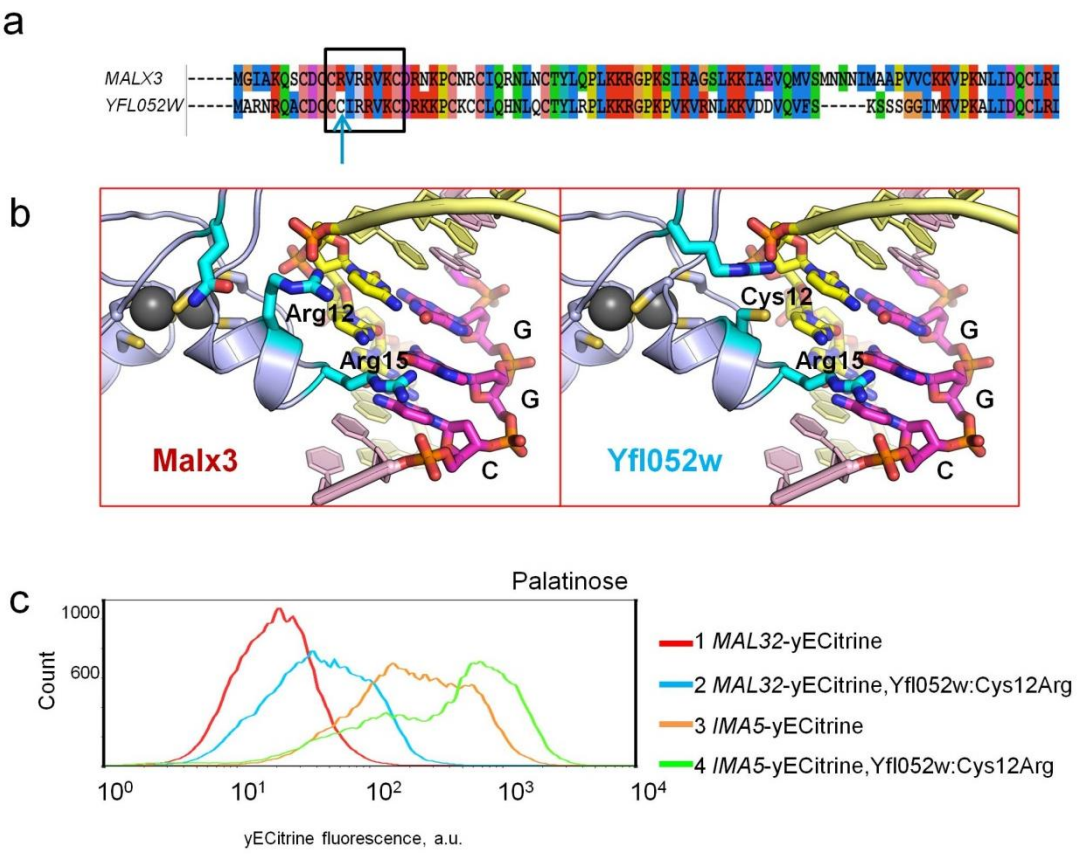

Figure 6. Differences in the DNA-binding domain of Malx3 and Yfl052w explain their different binding specificity. (**a**) Alignment of the Malx3 and Yfl052w DNA-binding domains. Amino acids predicted to interact with the DNA binding site are indicated with a black rectangle. The key position 12 that differs between Malx3 and Yfl052w is highlighted with a blue arrow. (**b**) Molecular modelling of the interaction between the Zn-finger domain and its DNA binding site. Important base pairs are represented as yellow and magenta sticks, important amino acids are represented as blue sticks. The Arg15 is shared between both transcription factors and is responsible for the recognition of the G in the middle of the CGG binding motif. Arg12 in Malx3 does not take part in recognition of the CGG motif, but Cys12 in Yfl052w does interact with the DNA and is responsible for the preference for a G nucleotide in the third position of the motif. (**c**) A mutated version of the palatinose-specific Yfl052 activator (Cys12Arg) is able to partly activate the *MAL32* promoter in response to palatinose.

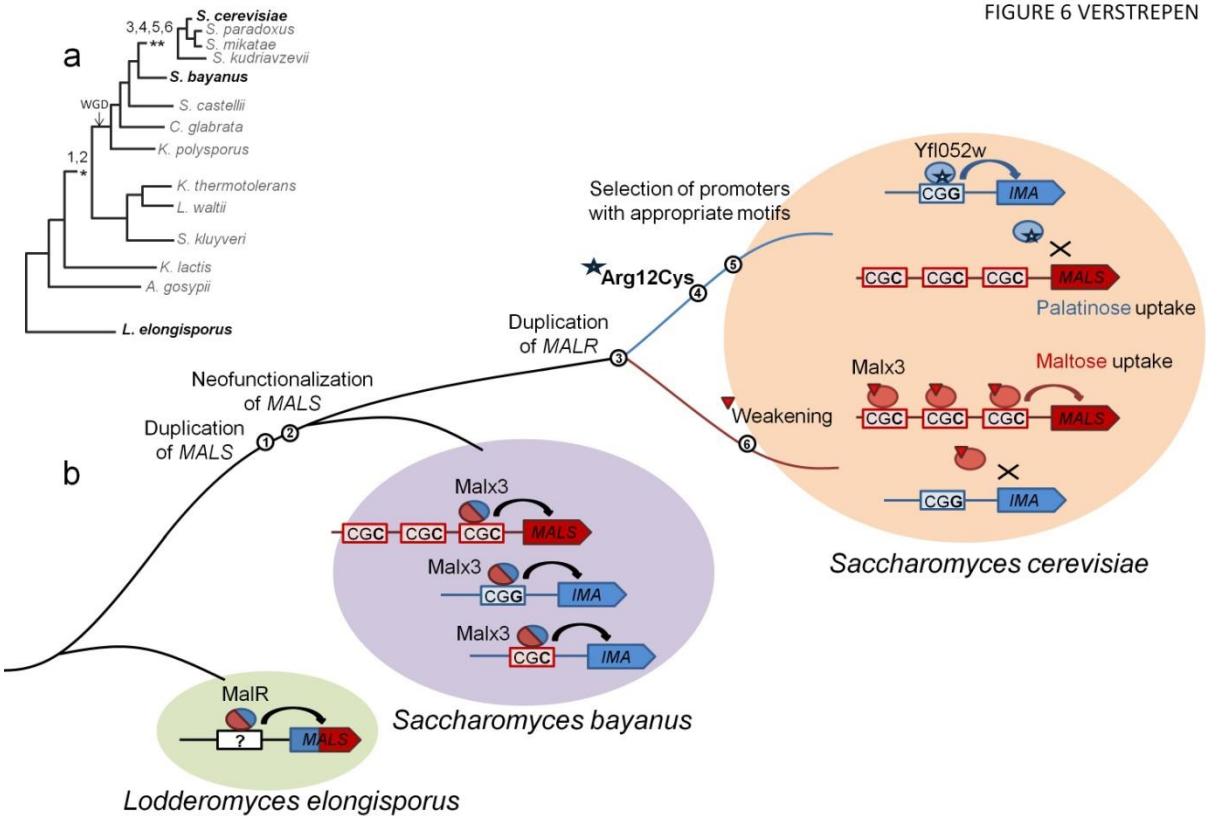

Figure 7. Possible evolutionary mutational path of *MAL* regulatory network diversification. (**a**) Simplified phylogenetic tree of the fungal lineage. The numbers correspond to the key evolutionary events listed in panel (b). WGD denotes the documented whole-genome duplication event in the fungal lineage. (**b**) Likely evolutionary path of the *MAL* regulatory network. The path starts from the common ancestor of *L. elongisporus*, *S. bayanus* and *S. cerevisiae* and ends at the modern day *S. cerevisiae*. In the common ancestor of *L. elongisporus*, *S. bayanus* and *S. cerevisiae*, maltose and isomaltose enzymatic activities are not separated and coexist in a single ancestral MalS enzyme, which is regulated by the single promiscuous MalR regulator. In the common ancestor of *S. cerevisiae* and *K. thermotolerans*, the *MALS* genes duplicated and neofunctionalized (1, 2), so that both types of target genes (maltose and palatinose specific) are present and are regulated by one promiscuous Malx3-like transcription factor that has an Arg residue at position 12 allowing it to bind both CGG and CGC motifs. The regulation is not specific at this point, that is, palatinose and maltose specific genes are equally expressed in presence of their respective substrate as well as a non-specific disaccharide (as it is in *S. bayanus*). Two separate regulatory circuits appear around the deviation of *S. bayanus* from the *Saccharomyces* tree. The *MALR* gene is duplicated (3) and this duplication event is followed by single nucleotide mutatios in the first position of the Arg12 codon, changing it to Cys in one of the paralogs (4), thus preventing it from binding CGC motifs in the promoters of maltose specific genes. Analysis of genomes that carry only one type *MALR* gene suggests that in the ancestral yeast CGG and CGC motifs were randomly distributed among maltose- and palatinose-specific genes. This implies that these binding sites needed to change in concert with the mutations

in the MALR paralogs, so that palatinose-specific genes only contain one CGG site, and maltose-specific genes contain three CGC motifs so that they can still be activated by the weakened Malx3 paralog (5, 6).