

YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts

Pedro T. Monteiro^{1,2,*}, Jorge Oliveira^{2,†}, Pedro Pais^{3,4,†}, Miguel Antunes^{3,4,†}, Margarida Palma^{3,4}, Mafalda Cavalheiro^{3,4}, Mónica Galocha^{3,4}, Cláudia P. Godinho^{3,4}, Luís C. Martins^{3,4}, Nuno Bourbon^{3,4}, Marta N. Mota^{3,4}, Ricardo A. Ribeiro^{3,4}, Romeu Viana^{3,4}, Isabel Sá-Correia^{3,4,*} and Miguel C. Teixeira^{3,4,*}

¹Department of Computer Science and Engineering, Instituto Superior Técnico (IST), Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisbon, Portugal, ²INESC-ID, R. Alves Redol, 9, 1000–029 Lisbon, Portugal, ³Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisbon, Portugal and ⁴iBB-Institute for BioEngineering and Biosciences, Biological Sciences Research Group, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

Received July 29, 2019; Revised September 05, 2019; Editorial Decision September 20, 2019; Accepted October 01, 2019

ABSTRACT

The YEASTRACT+ information system (<http://YEASTRACT-PLUS.org/>) is a wide-scope tool for the analysis and prediction of transcription regulatory associations at the gene and genomic levels in yeasts of biotechnological or human health relevance. YEASTRACT+ is a new portal that integrates the previously existing YEASTRACT (<http://www.yeastract.com/>) and PathoYeastract (<http://pathoyeastract.org/>) databases and introduces the NCYeastract (Non-Conventional Yeastract) database (<http://ncyeastract.org/>), focused on the so-called non-conventional yeasts. The information in the YEASTRACT database, focused on *Saccharomyces cerevisiae*, was updated. PathoYeastract was extended to include two additional pathogenic yeast species: *Candida parapsilosis* and *Candida tropicalis*. Furthermore, the NCYeastract database was created, including five biotechnologically relevant yeast species: *Zygosaccharomyces baillii*, *Kluyveromyces lactis*, *Kluyveromyces marxianus*, *Yarrowia lipolytica* and *Komagataella phaffii*. The YEASTRACT+ portal gathers 289 706 unique documented regulatory associations between transcription factors (TF) and target genes and 420 DNA binding sites, considering 247 TFs from 10 yeast species. YEASTRACT+ continues to make available tools for the prediction of the TFs involved in the regulation of gene/genomic expression. In this release, these

tools were upgraded to enable predictions based on orthologous regulatory associations described for other yeast species, including two new tools for cross-species transcription regulation comparison, based on multi-species promoter and TF regulatory network analyses.

INTRODUCTION

Yeasts are unicellular eukaryotic organisms with extraordinary diversity and properties. Many species have a strong impact in our way of life. For millennia, yeasts have played a crucial role in traditional biotechnology. In the past century, yeasts have become a crucial tool in modern biotechnology, serving as host for the production of added-value compounds, including recombinant proteins and primary and secondary metabolites. On the other hand, several yeast species have also been recognized as secondary pathogens in humans, being responsible for both superficial and deadly systemic infections.

Saccharomyces cerevisiae is a well-known eukaryotic model and was among the first to be used in biotechnological applications, that rendered it its street names ‘baker’s yeast’ or ‘brewer’s yeast’. Many other yeast species were later used, or are being considered for use, as cell factories, often being selected for their unique properties. For example, the methylotrophic yeast *Komagataella phaffii* (formerly *Pichia pastoris*) is a well-established host for recombinant protein production that is being subjected to metabolic engineering to produce a diversity of bioproducts (1,2); the food spoilage yeast *Zygosaccharomyces baillii* (3,4) exhibits strong resistance to stress induced by weak organic

*To whom correspondence should be addressed. Tel: +351 218417682; Email: mnpct@tecnico.ulisboa.pt

Correspondence may also be addressed to Isabel Sá-Correia. Tel: +351 218417682; Email: isacorreia@tecnico.ulisboa.pt

Correspondence may also be addressed to Pedro T. Monteiro. Tel: +351 213100320; Email: pedro.tiago.monteiro@tecnico.ulisboa.pt

†The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

acids food preservatives and, also for this reason, has been proposed as a cell factory for organic acids production; the Crabtree-negative species *Kluyveromyces lactis* is widely used in cheese production, given its ability to convert lactate into lactic acid and is attractive for the production of metabolites and heterologous proteins (5); the thermotolerant yeast *Kluyveromyces marxianus* is an established synthetic biology platform comparable to *S. cerevisiae* that hold potential for the industrial production of renewable chemicals (6); and *Yarrowia lipolytica* considered an oleaginous yeast, based on its ability to accumulate large amounts of lipids (7).

Candida species, on the other hand, are responsible for more than 400 000 life-threatening infections worldwide every year, being recognized as the fourth most common cause of nosocomial infections (8). More than 90% of all *Candida* infections are caused by *Candida albicans*, *Candida glabrata*, *Candida parapsilosis* and *Candida tropicalis*, with *C. albicans* being responsible for around half of all reported cases (9). These infections are typically associated to very high mortality rates (10), which are thought to be a consequence of the specific features associated to *Candida* pathogenesis.

The complete understanding of the molecular and regulatory mechanisms that control the productivity and stress resistance in biotechnologically relevant yeasts is key to guide the design of more effective cell factories. This includes, of course, understanding the transcriptional control of both biosynthetic and stress resistance pathways. For example, resistance to acetic acid is a crucial feature for many fermentative processes where acetic acid accumulates as by-product or is present as an inhibitor in the used raw materials, such as lignocellulosic hydrolysates. Moreover, acetic acid is a very common food preservative. Acetic acid tolerance in yeasts is tightly controlled at the transcriptional level by several transcription factors (TFs), particularly by Haa1, whose role in several yeast species, such as *S. cerevisiae* (11,12), *Z. baillii* (13) and *C. glabrata* (14), has been documented.

The complete understanding of the molecular mechanisms that control pathogenesis-related phenotypes in these human pathogens is also key to guide the design of more effective therapeutic options. For example, the clinical evolution of *Candida* strains toward improved fitness, virulence or drug resistance is often the consequence of changes in TF activities. The most well-known of such cases is the acquisition of drug resistance mediated by gain-of-function mutations in the TF(s) that control the expression of drug efflux pumps capable of extruding azole drugs (15), namely Pdr1 from *C. glabrata* (16) and Tac1 from *C. albicans*, and *C. parapsilosis* (17).

In this release, YEASTRACT+ is presented as a portal that unites and integrates the previously existing YEASTRACT (18–22) and PathoYeasttract (23) databases and the new NCYeasttract (Non-Conventional Yeasttract) database. Besides describing the updated versions of the YEASTRACT and PathoYeasttract databases, in terms of the regulatory information gathered for *S. cerevisiae*, *C. albicans* and *C. glabrata*, it also presents upgrades in three directions: (i) two other *Candida* species were included in the PathoYeasttract database, *C. parapsilosis* and *C. tropicalis*;

(ii) the NCYeasttract was developed, including five yeast species *Z. baillii*, *K. lactis*, *K. marxianus*, *Y. lipolytica* and *K. phaffii*; and (iii) further computational tools were added to the information system to increase the power of regulatory network visualizing tools and to improve the comparative genomics tools to enable cross-species comparison of regulatory networks and promoter sequences.

Data upgrade

In this paper, YEASTRACT+ is presented as a unified portal for the study of transcription regulation in yeasts. It includes updates on the YEASTRACT and PathoYeasttract databases and the launching of the NCYeasttract database, as detailed in the following paragraphs.

YEASTRACT was updated to include a total of 195 498 unique regulatory associations between TFs and target genes in *S. cerevisiae*, 20% more associations than in the previous version. It also includes 42 new TF binding sites, 16% more than before. Furthermore, the full data were revisited to populate the database with information that was missing on directionality (whether the TF is acting as repressor or activator in each association) or on environmental condition in which the transcriptional association was found to take place; this information was extended and now covers all TF–TG associations in YEASTRACT.

PathoYeasttract was updated to gather all available information on transcriptional associations published in peer-reviewed international journals for *C. albicans* and *C. glabrata*. It now includes 35 687 and 3508 unique regulatory associations for these species, respectively, an update that represents a 50% increment on the available data. At least one target gene was found for 40 and 118 TFs, out of the 117 and 163 TF predicted to be encoded by the *C. albicans* and *C. glabrata* genomes, respectively. At last, 39 and 76 consensus sequences recognized as binding sites for 12 and 45 TFs in *C. albicans* and *C. glabrata*, respectively, are now deposited in the database. Furthermore, the database was expanded to include two more pathogenic yeasts, *C. parapsilosis* and *C. tropicalis*. As before, data on predicted protein orthology for these new species were retrieved from CGD, Candida Gene Order Browser (CGOB—<http://cgob.ucd.ie/>) (24) and associated Gene Ontology (GO) terms, and their corresponding hierarchy, were retrieved from the GO consortium database (25). Available data on these new species includes 698 unique associations between TF and target genes in *C. tropicalis*, and 6986 unique associations between TF and target genes in *C. parapsilosis*. At least one target gene was found for 16 and 11 TFs, out of the 131 and 145 TF predicted to be encoded by the *C. tropicalis* and *C. parapsilosis* genomes, respectively. Moreover, six consensus sequences were identified for a single TF, Efg1, in *C. parapsilosis*, while only one binding site was identified for one TF, Wor1, in *C. tropicalis*, in both cases based on bioinformatics predictions alone.

NCYeasttract is the newest database in the YEASTRACT-PLUS portal. NCYeasttract gathers currently all published regulatory associations and TF binding sites for five of the most relevant non-conventional (non-*Saccharomyces*) yeasts of biotechnological potential: *Z. baillii*, *K. lactis*, *K. marxianus*, *Y. lipolytica* and *K. phaffii*.

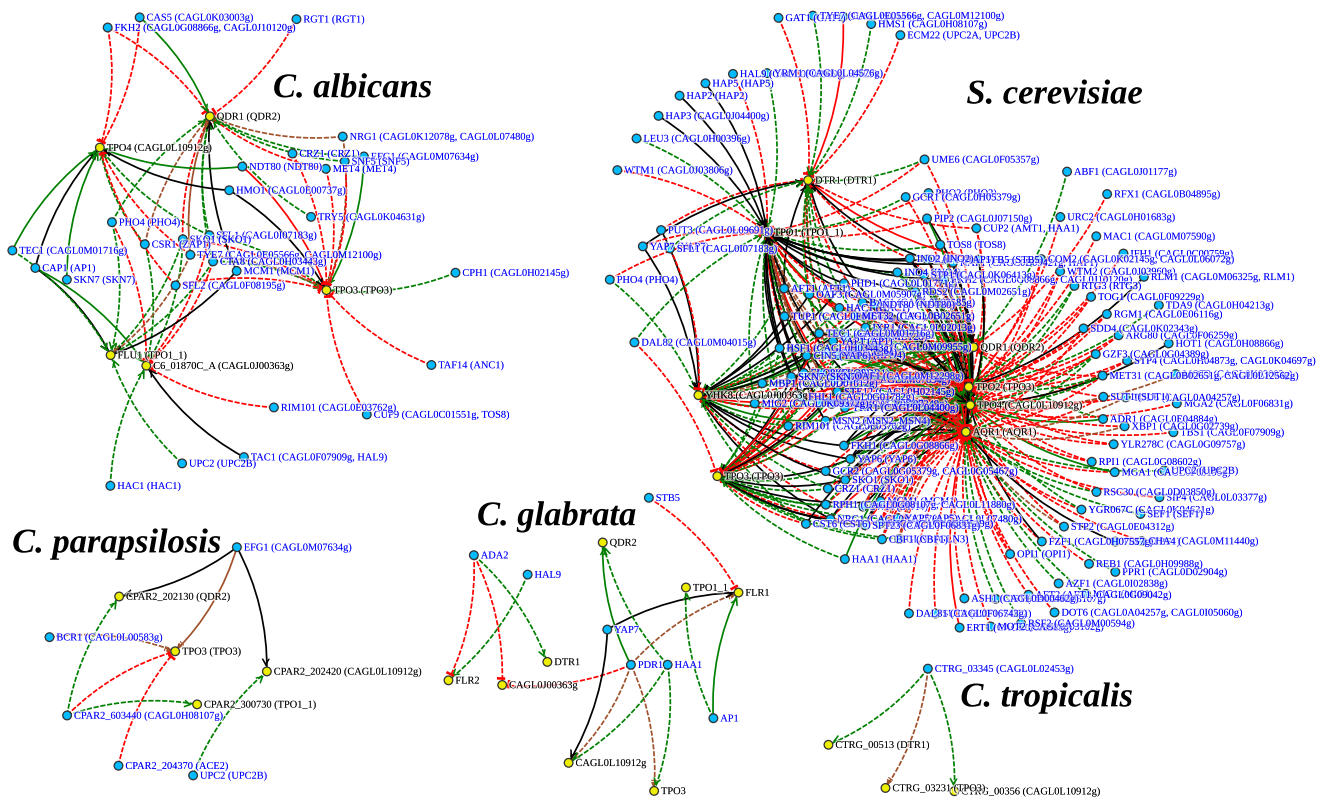


Figure 1. Regulation of the Drug:H⁺ Antiporter genes from *Candida glabrata*, considering their known regulatory associations in *C. glabrata* or the regulation of their orthologs in *Candida albicans*, *Candida parapsilosis*, *Candida tropicalis* and *Saccharomyces cerevisiae*. The images were obtained using the ‘Homologous network’ tool, offered among the ‘Cross species’ queries. The experimental evidences underlying each regulatory association (in full lines, in the case of DNA-binding evidence, or in dashed lines in the case of expression evidence), as well as the sign of the interaction—positive (green), negative (red), positive and negative (brown) or undefined (black) are highlighted.

31, 126, 3, 9238 and 581 unique regulatory associations are currently available at NCYeasttract, for each one of the referred species, respectively. Also 0, 2, 0, 1 and 0 TF binding sites, predicted for 1, 3, 0, 2 and 0 of the TFs in *Z. baillii*, *K. lactis*, *K. marxianus*, *Y. lipolytica* and *K. phaffii*, respectively, are included. Data on gene and promoter sequences and gene annotation for these five species were retrieved from the respective Genbank genome assembly, considering the following strains, respectively: IST302 (26), NRRL Y-1140, DMKU3-1042, CLIB122 and GS 115. GO was either annotated using the Blast2GO software (27) or retrieved from the fungal genomes browser Ensembl Fungi (28). Additionally, orthology associations between all the non-conventional yeast species added to the database and *S. cerevisiae* were inferred with the software OrthoFinder (29) and Proteinortho (30).

Altogether, YEASTRACT-PLUS gathers a total of 190 520 unique documented regulatory associations between TFs and target genes and 420 DNA binding sites, considering 247 TFs in the 10 yeast species. Also, more than 289 706 GO term associations to the genes gathered in the database.

All TF–target gene association or TF-consensus sequence association is automatically linked to specific information on the experimental setup used to identify it and to the environmental conditions in which that association was found to take place. The reference of the paper from which the information was retrieved, as well as the link to

the corresponding PubMed page is always provided. The experimental basis of the associations between TFs and target genes is also included in the database. The underlying experimental evidence was collected and classified as either *DNA Binding* or *Expression Evidence*. *DNA Binding Evidence* refers to experimental data obtained from chromatin immunoprecipitation (ChIP), ChIP-on-chip, ChIP-seq and electrophoretic mobility shift assay studies. *Expression Evidence* refers to data obtained from the comparative analysis of gene expression changes occurring in response to the deletion, mutation or over-expression of a given TF, based on reverse transcriptase-polymerase chain reaction, microarray analysis, RNA sequencing or expression proteomics. Based on the gathered *Expression Evidence* each regulatory association was classified as positive or negative, considering TFs that act as activators or repressors in these conditions, respectively. Based on this classification, YEASTRACT+ contains a total of 45 209 regulatory associations based on DNA-binding evidence and 161 783 on expression evidence. Again, in all cases the environmental condition in which the regulatory association was found to occur was included in the database, and clustered into 14 groups, including: biofilm formation, carbon source quality/availability, cell cycle/morphology, complex industrial media, human niche conditions, *in vitro*, ion/metal/phosphate/sulphur/vitamin availability, lipid supplementation, nitrogen source quality/availability,

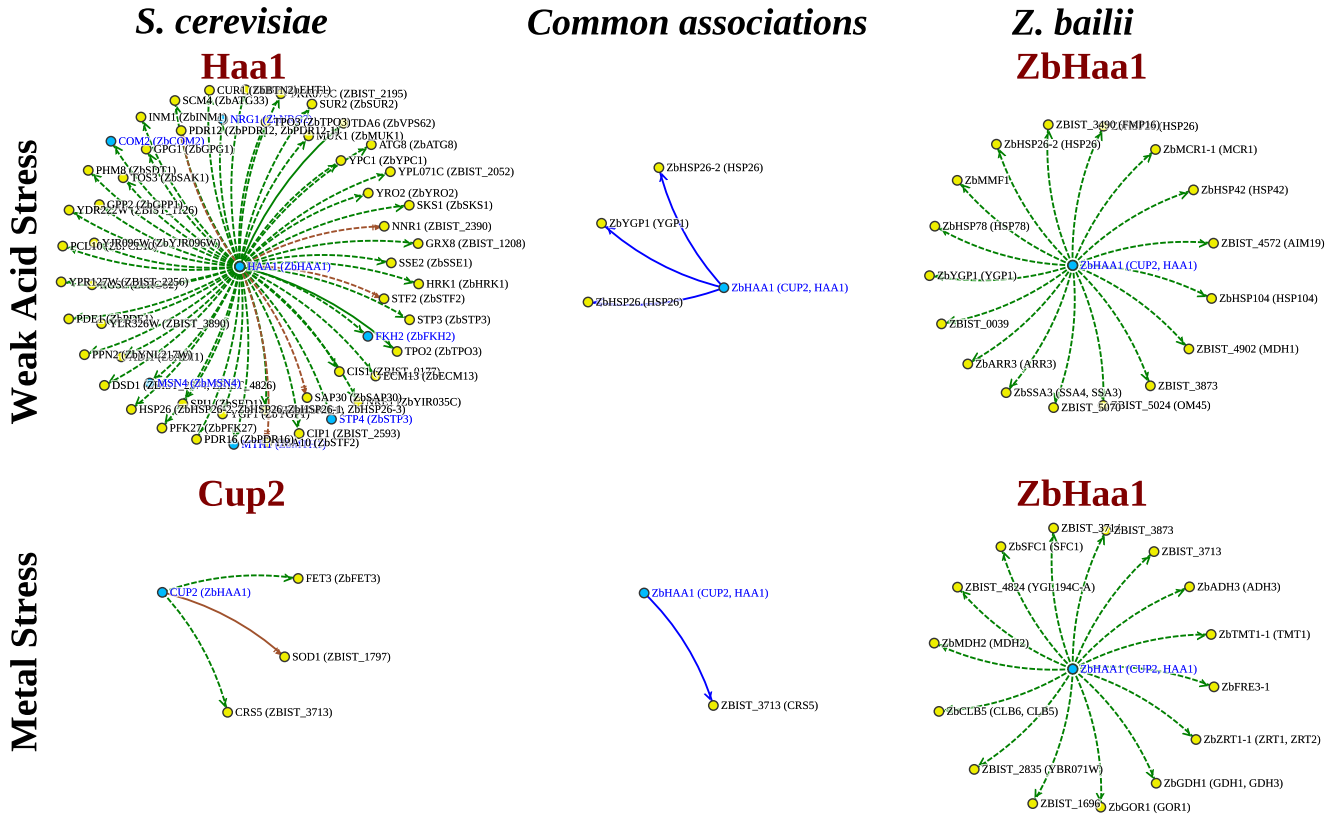


Figure 2. Cross-species comparison of the ZbHaa1 versus ScHaa1/ScCup2 orthologous TF regulons, active under weak acid or metal stress, as obtained in the ‘Network comparison’ tool. The experimental evidences underlying each regulatory association (in full lines, in the case of DNA-binding evidence, or in dashed lines in the case of expression evidence), as well as the sign of the interaction—positive (green), negative (red), positive and negative (brown) or undefined (black) are highlighted.

oxygen availability, pathogenicity, stress and unstressed log-phase growth (control). Each of these clusters is composed by sub-clusters to enable a finer filtering of the existing regulatory associations.

Improved tools for cross-species comparison

Homologous network prediction. There is a huge difference in the amount of available information for each of the yeast species currently hosted by YEASTRACT+. For example, considering the PathoYeast data, there are 5- to 10-fold more regulatory associations gathered for *C. albicans* than for *C. parapsilosis* and *C. glabrata*, respectively, and this number jumps to more than 50-fold in comparison to *C. tropicalis*. But even *C. albicans*, which might be considered a well-studied organism, within the group of pathogenic yeasts, is relatively poorly known when compared to the model eukaryote *S. cerevisiae*, for which the YEASTRACT database gathers 547% more regulatory associations.

The lack of available data on some of the pathogenic yeasts hampers dramatically the current predictive power of a database such as PathoYeast, unless the data gathered for better studied yeasts can be used to predict regulatory associations in less well studied species. With that aim, the possibility to predict the transcriptional control of a gene, or set of genes, based on the transcription of their orthologs in the remaining species is now offered in YEAS-

TRACT+. To illustrate the usefulness of this tool, Figure 1 represents the regulatory network that controls the 10 Drug:H+ Antiporters encoded in the *C. glabrata* genome, whose role in drug resistance has been well established in recent years (15,31–35). This family of transporters include Aqr1, Dtr1, Flr1, Flr2, Qdr2, Tpo1_1, Tpo1_2, Tpo3 and two uncharacterized predicted drug transporters encoded by ORFs *CAGL0L10912g* and *CAGL0J00363g*. When using the ‘Homologous network’ tool to study the regulatory network controlling these 10 genes in *C. glabrata*, the obtained result shows that there are seven TFs involved in the control of eight of these genes, whose function is mostly related to drug resistance (Pdr1, Stb5 and Ada2) and stress response (Ap1, Haa1, Yap7 and Hal9). Interestingly, two of the selected genes, *CgAQR1* and *CgTPO1_2*, have no known regulators in *C. glabrata*. However, if the ‘Network prediction’ tool is used considering ‘Search for homologous regulations in’, the network of TFs predicted to control the same genes can be expanded to many other TFs, depending on the selected organism.

For example, if the search is run against *C. albicans*, the network of TFs predicted to control the same genes increases to 29 TFs, some of which related to drug resistance (e.g. Tac1) and stress response (e.g. Cap1, Skn7 and Rim101), but others related to other interesting functions, such as carbon and nitrogen metabolism (Rgt1, Nrg1 and Met4) or biofilm formation (Tec1 and Snf5). These results

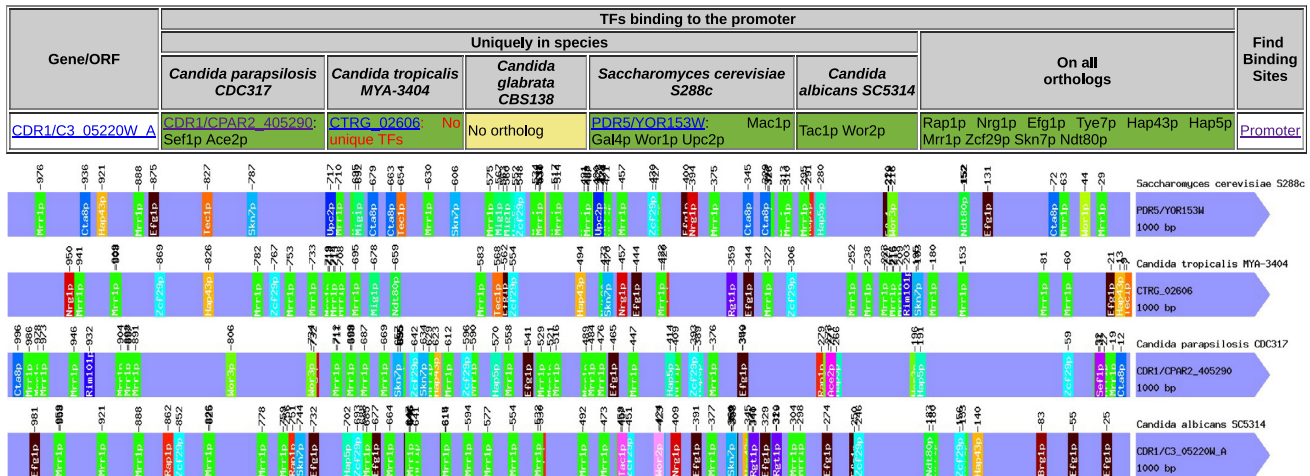


Figure 3. Cross-species comparison of the promoters of the *CDR1* orthologous genes in *Candida* species and *Saccharomyces cerevisiae*, as obtained in the ‘Promoter analysis’ tool.

are not just interesting in terms of their ability to expand the predicted network of TFs controlling a set of genes, but also suggest that these genes or their unexpected regulators may be involved in unforeseen biological processes. In this context, it is interesting to point out that at least one of the DHA transporters in *C. glabrata*, *CgTpo1_2*, is indeed required for biofilm formation (34), and others of the same family are suggested to do so as well, based on the TFs that control them.

Network comparison. Comparison of interspecies regulatory networks can be useful to identify functional elements without previous knowledge of function as well as identify significant evolutionary changes regarding transcriptional regulatory mechanisms. In YEASTRACT+, the user can now search for common and unique transcription regulatory associations of orthologous TFs in the 10 yeast species deposited therein, in a pairwise fashion, using the tool ‘Network comparison’. To exemplify the exploitation of this new tool, the documented *Z. baillii* Haa1-regulon (13) was compared to those controlled by its two orthologs in *S. cerevisiae*, Haa1 (12) and Cup2 (36). Interestingly, the bi-functional *Z. baillii* Haa1 TF was found to play a dual role in the response to copper and to low-chain weak acid stress, functions that are fulfilled separately by its ortholog TFs Cup2 and Haa1, respectively, in *S. cerevisiae* (13,37). The comparison between the ZbHaa1 and Haa1/Cup2 regulons, under acetic acid or copper stress, was carried out using this tool, leading to the definition of three regulatory networks for each stress condition, including the complete regulons of each TF and the core network found to be shared in the two species (Figure 2).

The results obtained show that the ZbHaa1- and Haa1-regulons in the response to acetic acid stress have two genes in common, *HSP26* and *YGPI*, which in *Z. baillii* correspond to three ORFs, *ZBIST_0079* (*HSP26*), *ZBIST_3442* (*HSP26.2*) and *ZBIST_0509* (*YGPI*). A similar approach can be used to compare the ZbHaa1 and Cup2 regulons under copper stress. In this case, the single shared target gene is *CRS5*, encoding a copper-binding metallothionein. These

observations suggest that these stress responsive genes appear to be in the core of acetic acid or copper stress response in yeasts.

Promoter analysis. Promoter sequences are essential to the control of the expression of the downstream genes. Interestingly, it has been seen that promoter sequences suffer much faster evolution rates than coding sequences, except in specific regulatory regions (38). TF binding sites appear to be conserved to some extent, within phylogenetically related organisms (38), enabling the careful prediction of TF target genes in closely related species, based on the existence of predicted TF binding sites.

In this release of the YEASTRACT+ portal, a new ‘Promoter Analysis’ tool, for ‘Cross species’ comparison, is offered. In this tool, it is possible to search for the consensus sequences recognized by the TFs of a selected species in the promoter sequences of a given gene in one species, and in the promoters of all its orthologs in the remaining species deposited in the database. As an example, the promoters of the orthologs of the *C. albicans* *CDR1* gene from the four *Candida* species were compared to that of the *S. cerevisiae*, considering the TF binding sites described for *C. albicans* (Figure 3).

First, it is highlighted that *C. glabrata* does not have a clear ortholog for this gene. As such, a ‘No ortholog’ message comes out. Second, this tool enables the search for TF consensus sequences that are present in all the promoters. The fact that nine TFs are predicted to bind to all these orthologous genes suggests that they are under a core conserved regulation. Additionally, there are specific TFs which are potential regulators of only one of these orthologous genes, for example, *Upc2* and *Sef1* in *C. parapsilosis*. This observation suggests that there might be some degree of specialization of this gene in each species, placing it under the control of alternative signaling pathways. Evidently, all hypotheses drawn through this analysis have to be considered only as indications that may guide further experimental work.

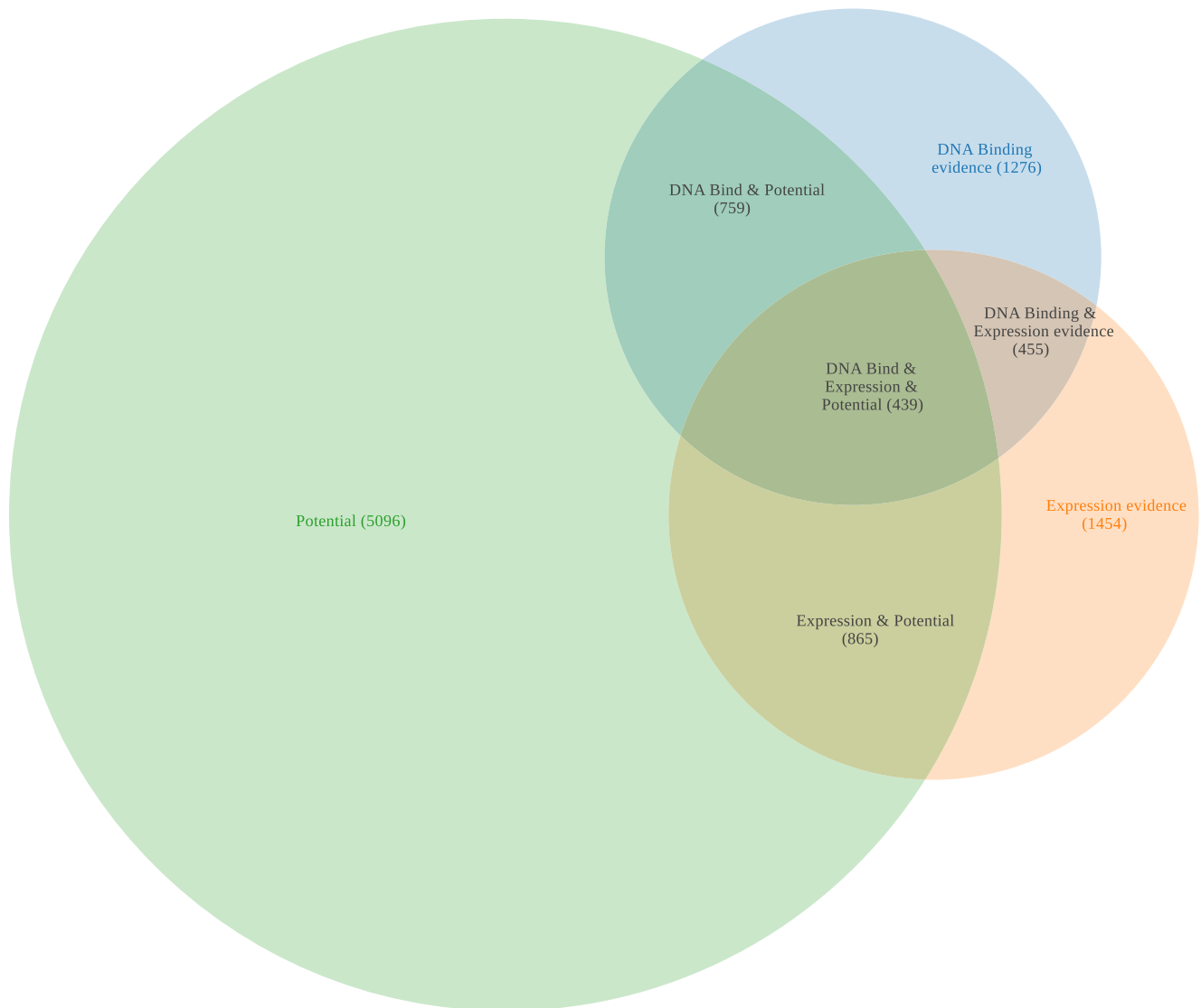


Figure 4. The Efg1 regulon in *Candida parapsilosis* as obtained in the PathoYeast database, considering different sources of experimental or predictive data. The image is obtained automatically, by accessing the Efg1 protein table, through the quick search tool, followed by selecting the ‘See Venn diagram regulon’ option.

Additional resources. YEASTRACT+ continues to offer specific tools to enable its user to predict gene and genomic regulation. These predictions can be made either considering documented regulatory associations, those experimentally validated and published in peer-reviewed international journals, or potential regulatory associations, those based on the occurrence of TF consensus sequences in the promoter regions of the considered target genes. Additional filtering of the data can be carried out, by restricting the search to specific experimental data and/or specific environmental conditions. In the first case, the user can select only DNA-binding evidence or only expression evidence, or eventually only the regulatory associations for which there is simultaneous DNA binding and expression evidence. The filtering choice can produce quite different outcomes, as there are many cases of TFs for which the regulatory information based on DNA-binding or expression evidence

can have little overlap, possibly due to lack of uniformity in the conducted experiments and to the fact that many genes whose expression is affected by a given TF, are not necessarily direct targets of this TF. For example, Figure 4 illustrates this idea with the *C. parapsilosis* TF Efg1, involved in the control of concentric-to-smooth colony morphology switch and biofilm formation.

CpEfg1 affects the expression of 3213 genes, as detected mostly through genome-wide expression studies, and binds the promoter region of 2929 genes, as detected mostly through ChIP-seq analysis. However, only for 894 genes there is evidence of Efg1 both controlling their expression and binding to their promoter regions. Among this subset of Efg1 target genes, only 439 display Efg1 consensus binding sequences in their promoter regions. Altogether this example shows that there are much diversified sources of information on regulatory associations that need to be taken

into account when using the tools made available at YEASTRACT+.

FUTURE DIRECTIONS

The YEASTRACT+ team is committed to continue to offer updated, reliable and complete information on the field of transcriptional regulation in yeasts to the international community working in the Biology and Biotechnology of yeasts, or on the molecular basis of candidaemia and its prophylaxis and treatment. Furthermore, the possibility to run a systematic inter-species comparison of transcription regulatory networks in different yeast will continue to be developed, especially through the development of more complex dedicated tools and the extension of the YEASTRACT+ databases to other relevant yeasts.

ACKNOWLEDGEMENTS

The YEASTRACT+ portal was constructed, in part, by joining the YEASTRACT (<http://yeastract.com>) and PathoYeast (<http://pathoyeast.org>) databases. All past and present colleagues and collaborators of the PathoYeast and YEASTRACT projects are deeply acknowledged. The information about *Candida* and *Saccharomyces* genes other than documented regulations, potential regulations and the TF binding sites, were gathered from: *Saccharomyces* Genome Database (SGD), *Candida* Genome Database (CGD), *Candida* Gene Order Browser (CJOB) and GO Consortium. We acknowledge these four platforms for making that information available to the public in easily downloadable files.

FUNDING

Lisboa Portugal Regional Operational Programme (Lisboa2020) [LISBOA-01-0145-FEDER-022231, including grants to J.O., N.M., R.V.]; Fundação para a Ciência e a Tecnologia (FCT), National Funds [PTDC/BBB-BIO/4004/2014, PTDC/BBB-BEP/0385/2014, ERA-IB-2/0003/2015, PTDC/BII-BIO/28216/2017]; Sabbatical Grant [SFRH/BSAB/143643/2019 to P.T.M.]; BIOTECnico-Biotechnology and Biosciences and AEM-Applied and Environmental Microbiology FCT PhD Grants (to P.P., M.A., M.C., M.G., L.M., M.M., R.R.); iBB-Institute for Bioengineering and Biosciences from Programa Operacional Regional de Lisboa 2020 (Project N. 007317); iBB and INESC-ID from FCT [UID/BIO/04565/2013, UID/CEC/50021/2019]. Funding for open access charge: FCT.

Conflict of interest statement. None declared.

REFERENCES

- Gasser, B. and Mattanovich, D. (2018) A yeast for all seasons—is *Pichia pastoris* a suitable chassis organism for future bioproduction? *FEMS Microbiol. Lett.*, **365**, doi:10.1093/femsle/fny181.
- Kurtzman, C.P. (2009) Biotechnological strains of *Komagataella* (*Pichia*) *pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *J. Ind. Microbiol. Biotechnol.*, **36**, 1435–1438.
- Palma, M. and Sá-Correia, I. (2019) Physiological genomics of the highly weak-acid-tolerant food spoilage yeasts of *Zygosaccharomyces bailii* sensu lato. *Prog. Mol. Subcell. Biol.*, **58**, 85–109.
- Palma, M., Guerreiro, J.F. and Sá-Correia, I. (2018) Adaptive response and tolerance to acetic acid in *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*: A physiological genomics perspective. *Front. Microbiol.*, **9**, 274.
- Spohner, S.C., Schaum, V., Quitmann, H. and Czermak, P. (2016) *Kluyveromyces lactis*: an emerging tool in biotechnology. *J. Biotechnol.*, **222**, 104–116.
- Cernak, P., Estrela, R., Poddar, S., Skerker, J.M., Cheng, Y.-F., Carlson, A.K., Chen, B., Glynn, V.M., Furlan, M., Ryan, O.W. *et al.* (2018) Engineering *Kluyveromyces marxianus* as a robust synthetic biology platform host. *Mbio.*, **9**, e01410-18.
- Beopoulos, A., Cescut, J., Haddouche, R., Uribealarea, J.-L., Molina-Jouve, C. and Nicaud, J.-M. (2009) *Yarrowia lipolytica* as a model for bio-oil production. *Prog. Lipid Res.*, **48**, 375–387.
- Wisplinghoff, H., Bischoff, T., Tallent, S.M., Seifert, H., Wenzel, R.P. and Edmond, M.B. (2004) Nosocomial bloodstream infections in US hospitals: Analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin. Infect. Dis.*, **39**, 309–317.
- Guinea, J. (2014) Global trends in the distribution of *Candida* species causing candidemia. *Clin. Microbiol. Infect.*, **20**, 5–10.
- Denning, D.W. and Bromley, M.J. (2015) How to bolster the antifungal pipeline. *Science*, **347**, 1414–1416.
- Mira, N.P., Henriques, S.F., Keller, G., Teixeira, M.C., Matos, R.G., Arraiano, C.M., Winge, D.R. and Sá-Correia, I. (2011) Identification of a DNA-binding site for the transcription factor Haa1, required for *Saccharomyces cerevisiae* response to acetic acid stress. *Nucleic Acids Res.*, **39**, 6896–6907.
- Mira, N.P., Becker, J.D. and Sá-Correia, I. (2010) Genomic expression program involving the Haa1p-Regulon in *Saccharomyces cerevisiae* response to acetic acid. *Omi. A J. Integr. Biol.*, **48**, 587–601.
- Antunes, M., Palma, M. and Sá-Correia, I. (2018) Transcriptional profiling of *Zygosaccharomyces bailii* early response to acetic acid or copper stress mediated by ZbHaa1. *Sci. Rep.*, **8**, 14122.
- Bernardo, R.T., Cunha, D. V., Wang, C., Pereira, L., Silva, S., Salazar, S.B., Schröder, M.S., Okamoto, M., Takahashi-Nakaguchi, A., Chibana, H. *et al.* (2017) The CgHaa1-Regulon mediates response and tolerance to acetic acid stress in the human pathogen *Candida glabrata*. *G3*, **7**, 1–18.
- Costa, C., Dias, P.J., Sá-Correia, I. and Teixeira, M.C. (2014) MFS multidrug transporters in pathogenic fungi: Do they have real clinical impact? *Front. Physiol.*, **5**, 197.
- Ferrari, S., Sanguinetti, M., Torelli, R., Posteraro, B. and Sanglard, D. (2011) Contribution of *CgPDR1*-regulated genes in enhanced virulence of Azole-resistant *Candida glabrata*. *PLoS One*, **6**, e17589.
- Berkow, E.L., Manigaba, K., Parker, J.E., Barker, K.S., Kelly, S.L. and Rogers, P.D. (2015) Multidrug transporters and alterations in sterol biosynthesis contribute to azole antifungal resistance in *Candida parapsilosis*. *Antimicrob. Agents Chemother.*, **59**, 5942–5950.
- Monteiro, P.T., Mendes, N.D., Teixeira, M.C., D'orey, S., Tenreiro, S., Mira, N.P., Pais, H., Francisco, A.P., Carvalho, A.M., Lourenço, A.B. *et al.* (2008) YEASTRACT-DISCOVERER: New tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **36**, 132–136.
- Teixeira, M.C., Monteiro, P.T., Guerreiro, J.F., Gonçalves, J.P., Mira, N.P., Dos Santos, S.C., Cabrito, T.R., Palma, M., Costa, C., Francisco, A.P. *et al.* (2014) The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **42**, 161–166.
- Teixeira, M.C., Monteiro, P.T., Jain, P., Tenreiro, S., Fernandes, A.R., Mira, N.P., Alenquer, A., Freitas, A.T., Oliveira, A.L. and Sá-Correia, I. (2006) The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **34**, D446–D451.
- Abdulrehman, D., Monteiro, P.T., Teixeira, M.C., Mira, N.P., Lourenço, A.B., Dos Santos, S.C., Cabrito, T.R., Francisco, A.P., Madeira, S.C., Aires, R.S. *et al.* (2011) YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in *Saccharomyces cerevisiae* through a web services interface. *Nucleic Acids Res.*, **39**, 136–140.
- Teixeira, M.C., Monteiro, P.T., Palma, M., Costa, C., Godinho, C.P., Pais, P., Cavalheiro, M., Antunes, M., Lemos, A., Pedreira, T. *et al.* (2018) YEASTRACT: an upgraded database for the analysis of

- transcription regulatory networks in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **46**, D348–D353.
23. Monteiro,P.T., Pais,P., Costa,C., Manna,S., Sá-Correia,I. and Teixeira,M.C. (2017) The PathoYeast database: an information system for the analysis of gene and genomic transcription regulation in pathogenic yeasts. *Nucleic Acids Res.*, **45**, D597–D603.
 24. Maguire,S.L., ÓhÉigeartaigh,S.S., Byrne,K.P., Schröder,M.S., O’Gaora,P., Wolfe,K.H. and Butler,G. (2013) Comparative genome analysis and gene finding in *Candida* species using CGOB. *Mol. Biol. Evol.*, **30**, 1281–1291.
 25. The Gene Ontology Consortium. (2019) The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
 26. Palma,M., Münsterkötter,M., Peça,J., Güldener,U. and Sá-Correia,I. (2017) Genome sequence of the highly weak-acid-tolerant *Zygosaccharomyces bailii* IST302, amenable to genetic manipulations and physiological studies. *FEMS Yeast Res.*, **17**, doi:10.1093/femsyr/fox025.
 27. Conesa,A., Gotz,S., Garcia-Gomez,J.M., Terol,J., Talon,M. and Robles,M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.
 28. Gene Ontology Consortium. (2012) The Gene Ontology: enhancements for 2011. *Nucleic Acids Res.*, **40**, D559–D564.
 29. Emms,D.M. and Kelly,S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, **16**, 157.
 30. Lechner,M., Findeiss,S., Steiner,L., Marz,M., Stadler,P.F. and Prohaska,S.J. (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*, **12**, 124.
 31. Costa,C., Pires,C., Cabrito,T.R., Renaudin,A., Ohno,M., Chibana,H., Sá-Correia,I. and Teixeira,M.C. (2013) *Candida glabrata* drug: H⁺ antiporter CgQdr2 confers imidazole drug resistance, being activated by transcription factor CgPdr1. *Antimicrob. Agents Chemother.*, **57**, 3159–3167.
 32. Costa,C., Nunes,J., Henriques,A., Mira,N.P., Nakayama,H., Chibana,H. and Teixeira,M.C. (2014) *Candida glabrata* drug:H⁺ antiporter CgTpo3 (ORF CAGL0I10384G): Role in azole drug resistance and polyamine homeostasis. *J. Antimicrob. Chemother.*, **69**, 1767–1776.
 33. Romão,D., Cavalheiro,M., Mil-Homens,D., Santos,R., Pais,P., Costa,C., Takahashi-Nakaguchi,A., Fialho,A.M., Chibana,H. and Teixeira,M.C. (2017) A new determinant of *Candida glabrata* virulence: the acetate exporter CgDtr1. *Front. Cell Infect. Microbiol.*, **7**, 473.
 34. Santos,R., Costa,C., Mil-Homens,D., Romão,D., de Carvalho,C.C.C.R., Pais,P., Mira,N.P., Fialho,A.M. and Teixeira,M.C. (2017) The multidrug resistance transporters CgTpo1.1 and CgTpo1.2 play a role in virulence and biofilm formation in the human pathogen *Candida glabrata*. *Cell Microbiol.*, **19**, 1–13.
 35. Pais,P., Costa,C., Pires,C., Shimizu,K., Chibana,H. and Teixeira,M.C. (2016) Membrane proteome-wide response to the antifungal drug clotrimazole in *Candida glabrata*: role of the transcription factor CgPdr1 and the drug:H⁺ antiporters CgTpo1.1 and CgTpo1.2. *Mol. Cell Proteomics*, **15**, 57–72.
 36. Keller,G., Bird,A. and Winge,D.R. (2005) Independent metalloregulation of Ace1 and Mac1 in *Saccharomyces cerevisiae*. *Eukaryot. Cell*, **4**, 1863–1871.
 37. Palma,M., Dias,P.J., de Roque,F.C., Luzia,L., Guerreiro,J.F. and Sá-Correia,I. (2017) The *Zygosaccharomyces bailii* transcription factor Haa1 is required for acetic acid and copper stress responses suggesting subfunctionalization of the ancestral bifunctional protein Haa1/Cup2. *BMC Genomics*, **18**, 75.
 38. Gasch,A.P., Moses,A.M., Chiang,D.Y., Fraser,H.B., Berardini,M. and Eisen,M.B. (2004) Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biol.*, **2**, e398.