

The BioData.pt Microbiology and Biotechnology Community
YEASTRACT+ portal

Pedro T. Monteiro

INESC-ID / IST - Universidade de Lisboa

Use *S.cerevisiae* as a **model organism**



Until 2005...

- Use SGD or RSAT to retrieve the promoter sequences
- Use TRANSFAC or SCPD to get a list of TF binding sites (TFBSs)
- Use RSAT to search TFBSs within the promoter sequences
- Use PUBMED to search for associated bibliographic references



Use *S.cerevisiae* as a **model organism**



Until 2005...

- Use SGD or RSAT to retrieve the promoter sequences
- Use TRANSFAC or SCPD to get a list of TF binding sites (TFBSs)
- Use RSAT to search TFBSs within the promoter sequences
- Use PUBMED to search for associated bibliographic references



But this presented only **descriptive** view...

How to explore regulatory information?

Search for regulatory associations for:

- Transcription factors (TFs) acting upon a given ORF/gene
- Target genes (TGs) being regulated by a given TF
- Regulatory [network visualization](#) of the results

This called for **mechanistic** view!

- Repository of regulatory associations in *S. cerevisiae*
- Longstanding collaboration (15y)
 - iBB / Dep Bioeng @ IST
 - INESC-ID / Dep Comp Sci @ IST
- Continuously updated
 - DB curation (> 1000 papers) @ BSRG
 - Server + new functionalities @ INESC-ID

Timeline

2006



YEASTRACT
1st RELEASE

- Repository of regulatory associations in *S. cerevisiae*
- Longstanding collaboration (15y)
 - iBB / Dep Bioeng @ IST
 - INESC-ID / Dep Comp Sci @ IST
- Continuously updated
 - DB curation (> 1000 papers) @ BSRG
 - Server + new functionalities @ INESC-ID

Timeline

2006

2008



YEASTRACT
1st RELEASE

YEASTRACT
2nd RELEASE



[Tutorial](#)

[Quick Search](#)

[DISCOVERER](#) new

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Mat

- [Search by I](#)
- [Find TF Bn](#)
- [Search Mot](#)

Utilities:

- [ORF list](#)

[Home](#) >> [Group Genes by TF](#) >> [Result](#)

[Tutorial](#) - [Help](#)



Genes grouped by TF, ordered by the percentage of genes regulated by TF, relative to the total number of genes in the list



Transcription Factor	%	ORF/Genes	
Abf1p	80.0 %	ACS1 BAP3 AMS1 ADE5,7 ARN2 ABF1 ALD3 ADH1	<input checked="" type="checkbox"/>
Gcn4p	20.0 %	ADE5,7 ADH1	<input checked="" type="checkbox"/>
Rap1p	20.0 %	ADE5,7 ADH1	<input checked="" type="checkbox"/>
Stp2p	20.0 %	BAP3 AMS1	<input checked="" type="checkbox"/>
Gcr1p	20.0 %	PDR3 ADH1	<input checked="" type="checkbox"/>
Msn4p	10.0 %	ALD3	<input checked="" type="checkbox"/>

Transcriptional Regulatory Network
Selected TFs count also as targets

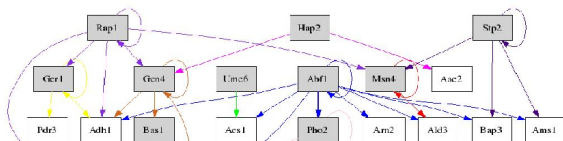
[Image](#)

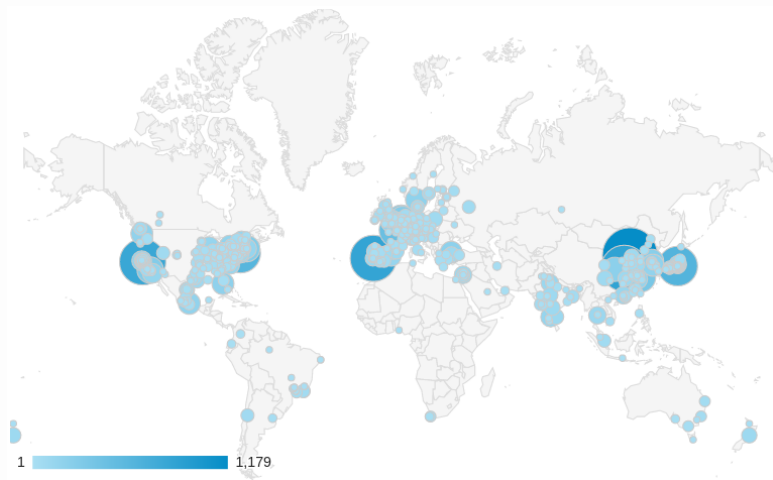
[Home](#) >> [Group Genes by TF](#) >> [Result](#) >> [Image](#)

[Tutorial](#) - [Help](#)



Transcriptional Regulatory Network





Pageviews/year

2006	86.568
2007	148.983
2008	170.870
2009	145.059
2010	161.072
2011	170.425
2012	167.460
2013	154.617
2014	115.847
2015	100.769
2016	97.177
2017	83.216
2018	99.173
2019	129.423
2020	116.532

~1.947.191

• Joint publications

- (Teixeira *et al.*, *Nucleic Acids Research* 2006)
- (Monteiro *et al.*, *Nucleic Acids Research* 2008)
- (Teixeira *et al.*, *Molecular BioSystems* 2010)
- (Abdulrehman *et al.*, *Nucleic Acids Research* 2011)
- (Monteiro *et al.*, *IET Systems Biology* 2011)
- (Teixeira, Monteiro *et al.*, *Nucleic Acids Research* 2014)
- (Monteiro *et al.*, *Nucleic Acids Research* 2017)
- (Teixeira, Monteiro *et al.*, *Nucleic Acids Research* 2018)
- (Monteiro, Oliveira *et al.*, *Nucleic Acids Research* 2020)
- ...

• Joint supervision of students


- MSc thesis
- Project in Systems Biology

• Joint research projects

- FCT project **CANTROL**
- FCT project **MIXED-UP**
- FCT infrastructure project **Biodata.pt**

[Web Services](#)
[DISCOVERER](#)
[Cross species **NEW**](#)
[Promoter analysis](#)
[Homologous network](#)
[Network comparison](#)
[Rank by Unique TFBS](#)
[Rank genes](#)
[Rank by TF](#)
[Rank by GO](#)
[Regulatory Associations](#)
[Search for TFs](#)
[Search for Genes](#)
[Search for Associations](#)
[Pattern Matching](#)
[Search by DNA Motif](#)
[Find TF Binding Site\(s\)](#)
[Utilities](#)
[Search for Orthologs **NEW**](#)
[IUPAC Code Generation](#)
[ORF List ↔ Gene List](#)
[Retrieve](#)
[TF-Consensus List](#)
[Regulation Matrix](#)
[Upstream Sequence](#)
[About](#)
[Contact/Credits](#)
[How to cite](#)
[Acknowledgments](#)
[Data submission form](#)

Locus Information

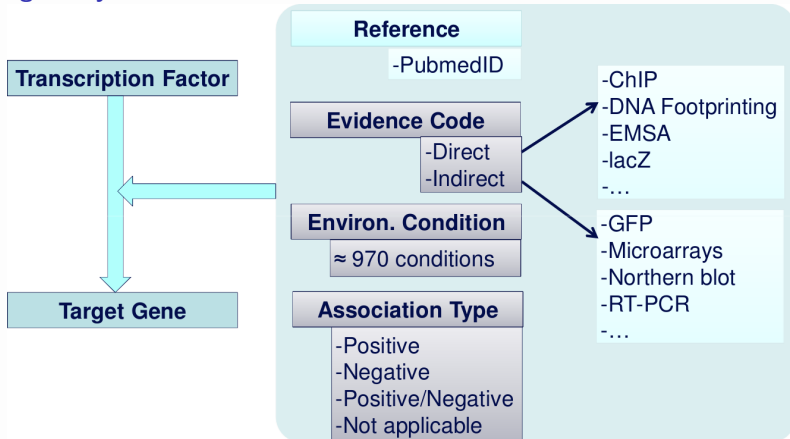
Locus Info	Protein Info	GO Info	Orthologs
Standard Name YAP1 Systematic Name YML007W Alias PAR1, PDR4, SNQ3 Strain S288c Description Basic leucine zipper (bZIP) transcription factor required for oxidative stress tolerance; activated by H2O2 through the multistep formation of disulfide bonds and transit from the cytoplasm to the nucleus; mediates resistance to cadmium External source Open in SGD  Gene Sequence Download FASTA View Seq			
Promoter Sequence Download FASTA			
<pre>>YAP1 YML007W CTCAAGTACTTCTGAACGGCTTCCTTTTGGGCAGAGTTGTTCTCGACATCAATGCACCT AAACCGTCTTTTTACCAATATCATCACCATGTAACCTCCCTAGTGAATTGGGCCGTGCTTT TTTCTCAATCTGTTTCACTCATCTTATGCTGCCACTATATATATTTTATTCCTGTT TTTTAAATATGTAAACTTAAATCGAAAGTAAAACAGATAAGGGAACCTTGAATGAAA ATTAACAGCTTATAAATACAAAGACGAATGGAAAGAAAAAGAGAGCAGCAGAGGATAGG AGAAAAACCCGAACATATATAAAAGCAATGAAGAAGTTATCGGCAACAGCAAGTATGG AGGTTTCTACCAATCTTCGCATAAAACCGCATGCTCGTAAGGCGATAAAAATGAAGTGT ACCATGAGACGAAGTGGAGCAGTTTACAGTGAATAATACATTTGCTGCTCCTCACTAC CTCGCTTCTCGAAAACCTTTTTTGATTTATCGGAAACGGCAGTAAACAGCAGCCAAAAT TGAATAAAGTACGGGAACGAGGTACCATATACGAAGATCGGGGCTTTGTTGGGCCCTAA ATTTTGTGATCTTACCGTTTGCATCTTTTCAAAAATAGTAACCGCCCTAGCTGTGTTT GTGATTTGACCTAGGTTACTCTTTTCTTTTCTGGGTGCGGGTAAACAATTTGGGCCCGC CCAAAGCGCCGCTTTTGTATGGGAACCGGAAACCCCTCCGATGAAGAGTAGGAGGGTGGC AACTGATGGATGCGTAAAGTCTTAAGAGATACATTTGCTTAATAGTCTCCGTTTACCGA TTAAGCACAGTACCTTTACGTTATATATAGGATGGTGTGTTAGCTTTTTTCTGAGCCC CTTGTGACTTGTGCATGAACACGAGCCATTTTTAGTTGGTTTAAAGGGAAGTTTTTGGC ACCCAAAACGTTTTAAAGAAGGAAAAGTTGTTCTTAAACC</pre>			
Chr coordinates Chr XIII from 253848-255800			

Go Back

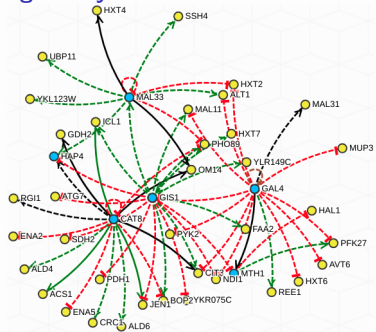
BioData.pt Talks - YEASTRACT+

8/20

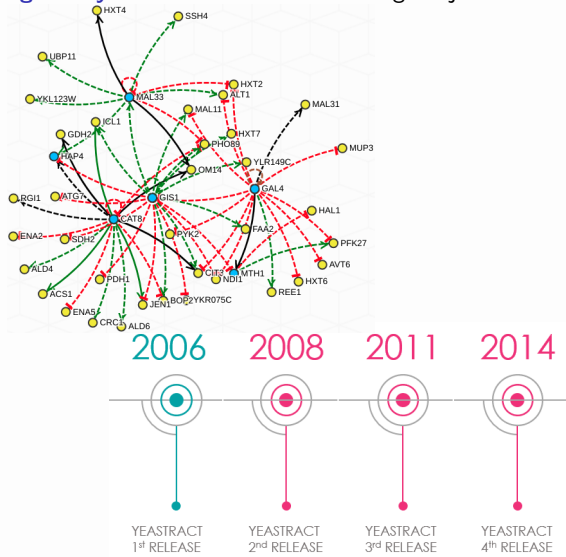
Regulatory associations data



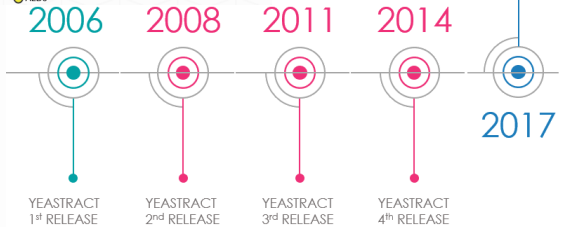
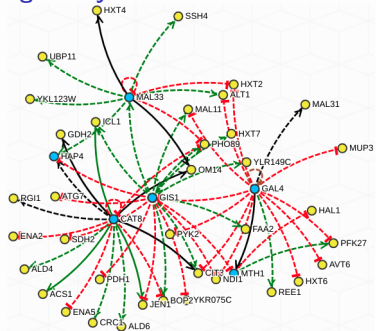
Regulatory network visualization using D3.js



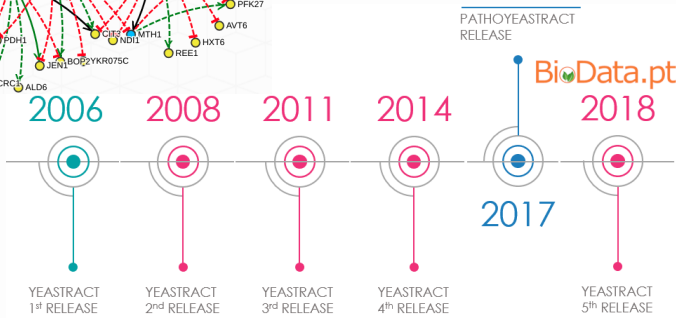
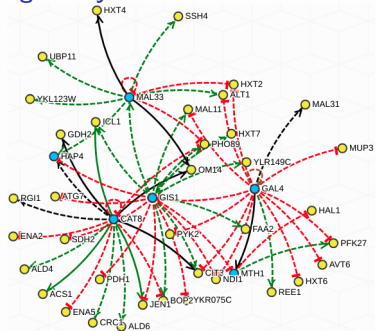
Regulatory network visualization using D3.js



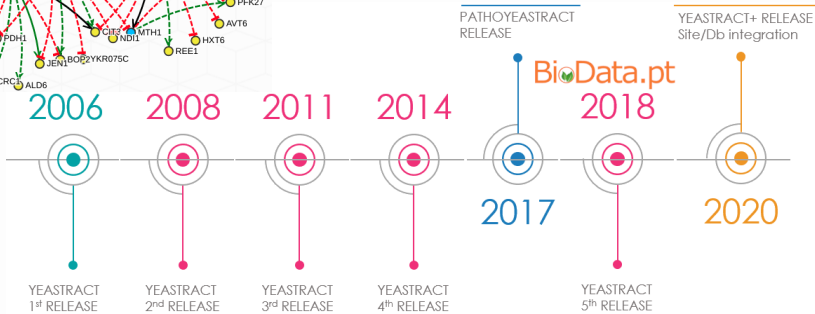
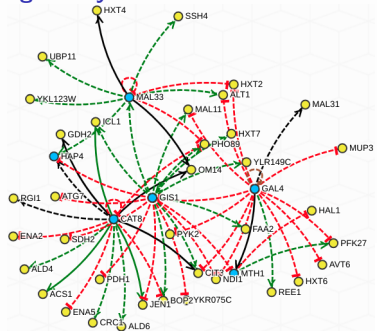
Regulatory network visualization using D3.js



Regulatory network visualization using D3.js



Regulatory network visualization using D3.js

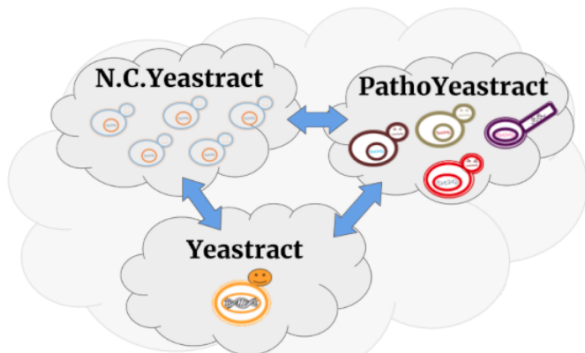


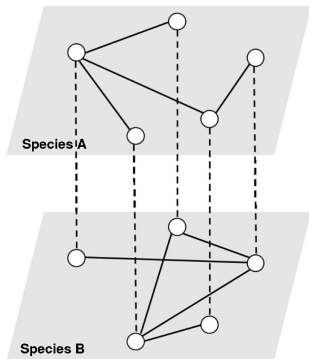
Welcome to YEASTRACT+

YEASTRACT+ (Yeast Search for Transcriptional Regulators And Consensus Tracking)+ is a portal that opens up to three distinct, but interconnected databases:

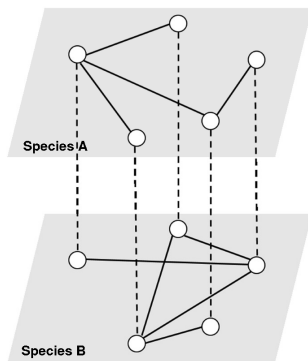
- **Yeastract**, which focus on the model yeast and cell factory *Saccharomyces cerevisiae*;
- **PathoYeastract**, created to provide a resource for clinicians and biomedical scientists working with four pathogenic yeasts responsible for more than 90% of all detected candidiasis: *Candida glabrata*, *Candida albicans*, *Candida parapsilosis* and *Candida tropicalis*;
- **N.C.Yeastract**, created to guide the analysis and application of the non-conventional biotechnologically-relevant yeasts: *Zygosaccharomyces bailii*, *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Komagataella pastoris* and *Yarrowia lipolytica*.

YEASTRACT+ combines the existing database information of all yeast species, providing bioinformatics tools for the prediction and visualization of gene and genomic regulation based on orthologous regulatory associations described for other yeast species, based on comparative genomics.





Taking advantage of [systems/DB integration](#)



Taking advantage of **systems/DB integration**

In-house homology all-against-all

- Run **BLASTp**
- Homology criteria:
 - p-value: 10^{-5}
 - bidirectional
 - tolerance of 10%
- Syntheny of at-least- $\{1,2,3\}$ neighbours

Cross species: network comparison

Show documented regulations in:

- All *Candida albicans* SC5314
- All *Saccharomyces cerevisiae* S288c
- Unique to *Saccharomyces cerevisiae* S288c
- Common (Homolog gene in parentheses)
- Unique to *Candida albicans* SC5314
- All *Candida albicans* SC5314

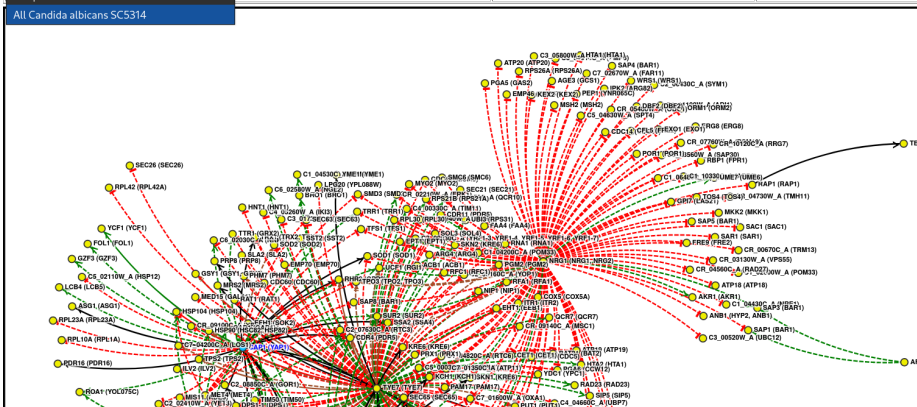
Show additional regulatory associations between TFs

Supporting evidence:

- DNA Binding only
- DNA Binding + Expression
- Expression only

Type:

- Positive
- Negative
- Unspecified
- Common



Homologous documented regulations in species:

Candida glabrata CBS138

Candida albicans SC5314

Candida glabrata CBS138

Zygosaccharomyces bailii IST302

Saccharomyces cerevisiae S288c

Candida parapsilosis CDC317

Candida tropicalis MYA-3404

Yarrowia lipolytica CLIB122

Komagataella phaffii GS115

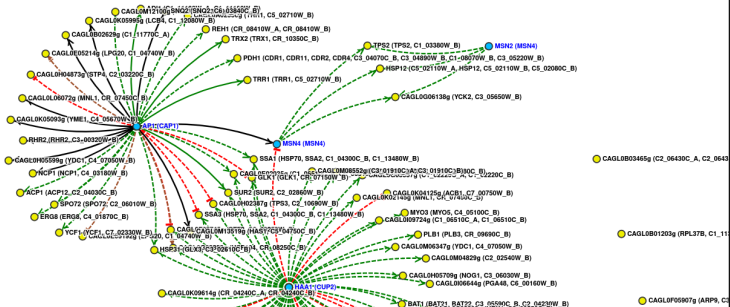
Show additional regulatory associations between TFs

Supporting evidence:

- DNA Binding only
- DNA Binding + Expression
- Expression only

Type:

- Positive
- Negative
- Unspecified



How can one analyse a newly sequenced genome?

How can one analyse a newly sequenced genome?

Rhodotorula toruloides

- Nonpathogenic, red-colored basidiomycetous fungus
- Can use a wide range of carbon sources for growth
- Good producer of carotenoids and some important enzymes
- Can accumulate lipids up to over 70% of its dry cell weight
- Genome assemblies from Genbank
- No previous database or tools available

- Open-source code: <https://gitlab.com/oliveira.jorge.88/web>
- Install YEASTRACT on your own webserver!
- Load genome assembly from Genbank (.gbff format)
- Compute homologs to a reference genome
- Compute TFBS matches on new promoters
- Infer/Visualize networks

- Open-source code: <https://gitlab.com/oliveira.jorge.88/web>
- Install YEASTRACT on your own webserver!
- Load genome assembly from Genbank (.gbff format)
- Compute homologs to a reference genome
- Compute TFBS matches on new promoters
- Infer/Visualize networks

1 From a genome assembly to full 2 regulatory network prediction: the case 3 study of *Rhodotorula toruloides* putative 4 Haa1-regulon

5 Jorge Oliveira^a, Miguel Antunes^{b,c}, Miguel C. Teixeira^{b,c}, Isabel Sá-Correia^{b,c}, Pedro T.
6 Monteiro^{a,d}

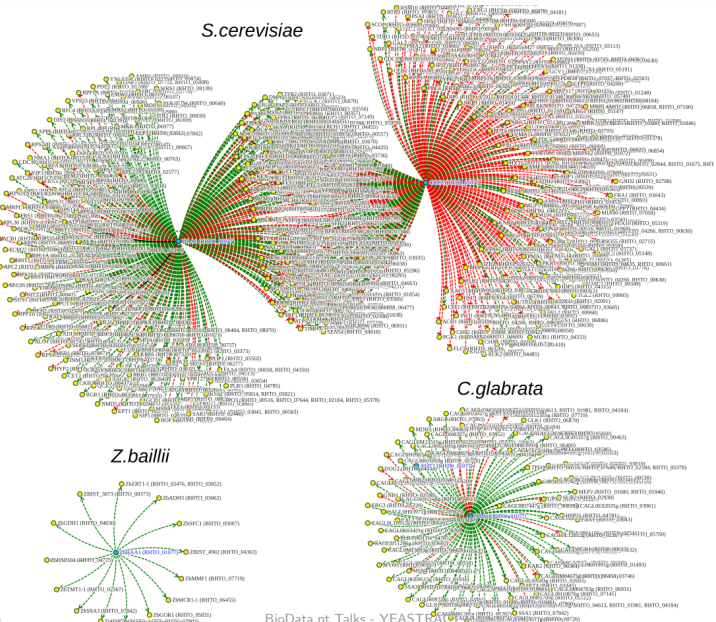
7

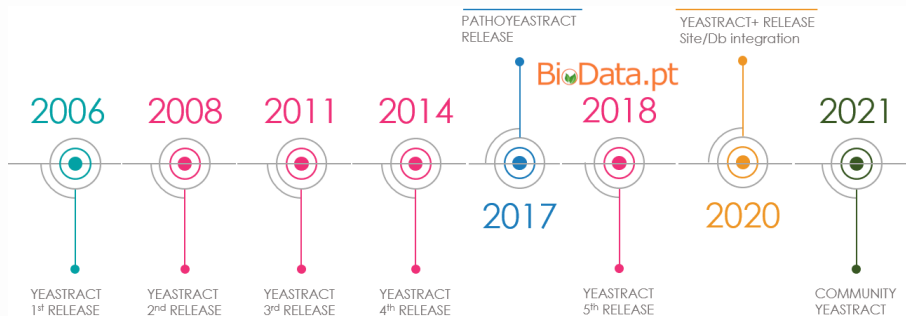
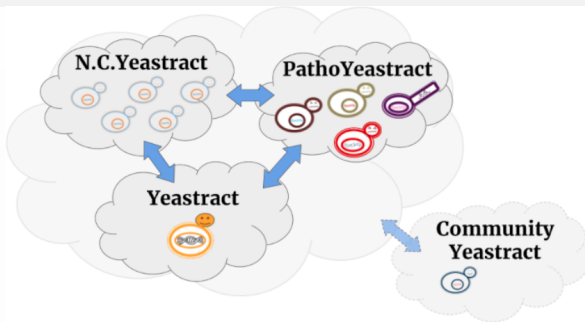
8 ^a INESC-ID, Lisbon, Portugal

9 ^b Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon,
10 Portugal

11 ^c iBB - Institute for Bioengineering and Biosciences, Biological Sciences Research Group,

12 Instituto Superior Técnico, Lisboa, Portugal







We want your (yeast) Genome!

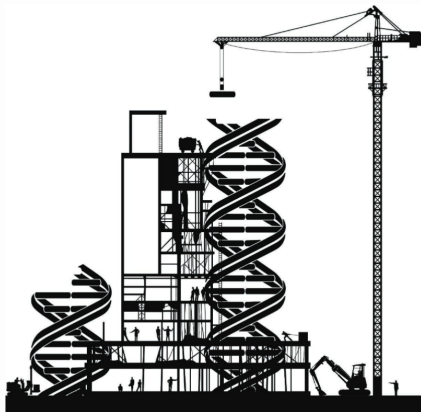


Image: E. Pennisi, Science 2014

<http://yeastract-plus.org/community/>

Thank you!

SW Algorithms and Tools for Constraint Solving Research Group	Biological Sciences Research Group
<p>Pedro T. Monteiro, PhD Assistant Professor IST Homepage</p>	<p>Miguel C. Teixeira, PhD Associate Professor IST Homepage</p>
<p>Jorge Oliveira, PhD Homepage</p>	<p>Margarida Palma, PhD Homepage</p> <p>Claudia P. Godinho, PhD Homepage</p>
<p>Information & Decision Support Systems Research Group</p>	<p>Pedro Pais, PhD student Homepage</p> <p>Mónica Galocha, PhD student Homepage</p> <p>Ricardo A. Ribeiro, PhD student Homepage</p>
<p>Arlindo L. Oliveira, PhD Full Professor IST Homepage</p>	<p>Mafalda Cavalheiro, PhD student Homepage</p> <p>Luis C. Martins, PhD student Homepage</p> <p>Romeu Viana, PhD student Homepage</p> <p>Isabel Sá-Correia Full Professor IST Homepage</p>
	<p>Miguel Antunes, PhD student Homepage</p> <p>Nuno B. Melo, PhD student Homepage</p>

Past contributors

Knowledge Discovery & Bioinformatics Research Group	Biological Sciences Research Group
<p>Ana T. Freitas Alexandre P. Francisco Sara C. Madeira</p>	<p>Sandra Tenreiro Alexandra Fernandes Marta Alenquer Artur Lourenço</p>
<p>Pooja Jain Dário Abdulrehman Sofia d'Orey Nuno D. Mendes</p>	<p>Nuno Mira Joana Guerreiro Sandra R. Santos Ténia R. Cabrito</p>
<p>Alexandra Carvalho Hélio Pais Ricardo Aires Joana Gonçalves</p>	<p>Catarina Costa</p>
<p>SW Algorithms and Tools for Constraint Solving Research Group</p>	
<p>Tiago Pedreira Alexandre Lemos</p>	