



HAL
open science

Genolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes.

David James Sherman, Tiphaine Martin, Macha Nikolski, Cyril Cayla,
Jean-Luc Souciet, Pascal Durrens

► To cite this version:

David James Sherman, Tiphaine Martin, Macha Nikolski, Cyril Cayla, Jean-Luc Souciet, et al.. Genolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes.. Nucleic Acids Research, 2009, 37, pp.D550-D554. 10.1093/nar/gkn859 . hal-04943449

HAL Id: hal-04943449

<https://hal.science/hal-04943449v1>

Submitted on 12 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Génolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes

David J. Sherman^{1,2,*}, Tiphaine Martin¹, Macha Nikolski¹, Cyril Cayla¹,
Jean-Luc Souciet³ and Pascal Durrens¹ for the Génolevures Consortium[†]

¹LaBRI, Laboratoire Bordelais de Recherche en Informatique, UMR CNRS 5800, ²INRIA, Institut National de Recherche en Informatique, Centre de Recherche Bordeaux Sud-Ouest, 351 Cours de la Libération, 33405 Talence cedex and ³Université Louis Pasteur, CNRS, UMR 7156, GDR 2354, Institut de Botanique, 28 rue Goethe, F-67000 Strasbourg, France

Received September 15, 2008; Revised October 15, 2008; Accepted October 16, 2008

ABSTRACT

The Génolevures online database (<http://cbi.labri.fr/Genolevures/> and <http://genolevures.org/>) provides exploratory tools and curated data sets relative to nine complete and seven partial genome sequences determined and manually annotated by the Génolevures Consortium, to facilitate comparative genomic studies of Hemiascomycete yeasts. The 2008 update to the Génolevures database provides four new genomes in complete (subtelomere to subtelomere) chromosome sequences, 50 000 protein-coding and tRNA genes, and *in silico* analyses for each gene element. A key element is a novel classification of conserved multi-species protein families and their use in detecting synteny, gene fusions and other aspects of genome remodeling in evolution. Our purpose is to release high-quality curated data from complete genomes, with a focus on the relations between genes, genomes and proteins.

INTRODUCTION

Since 1999, the Génolevures Consortium explores eukaryote genome evolution through the large-scale comparison of manually annotated yeast genomes. The Génolevures on-line database undergoes a major update with every significant data release: in 2004 with 13 partial genomes (1,2), in 2006 with four complete genomes (3,4) and in 2008 with four complete genomes (this work). Our purpose is to release high-quality data sets produced by the 13 partners in the Consortium, rather than to provide a site for integration of data available elsewhere, although we do integrate external data for comparison.

Yeasts are small eukaryotes covering an evolutionary range comparable to the entire Chordate phylum (5), and the unique combination of genetic and genomic tools available for yeasts make them ideal candidates for experimental study of metabolism, genetic engineering and molecular genetics. All of the many yeasts sequenced so far have small genomes (10–20 Mb) which allows for detailed comparative genomics for a relatively modest price. The economic impact of yeasts is widespread; different species are used for the production of beer, wine and bread and more recently of various metabolic products, such as vitamins, ethanol, citric acid, lipids, etc. Yeasts can degrade hydrocarbons (genera *Candida*, *Yarrowia* and *Debaryomyces*), metabolize xylose (*Pichia stipitis*), depolymerise tannin extracts (*Zygosaccharomyces rouxii*), and produce hormones and vaccines in industrial quantities through heterologous gene expression (6). Several human diseases are due to yeast species among them the Hemiascomycetes *Candida albicans*, *Candida glabrata*, *Candida tropicalis* and even *Saccharomyces cerevisiae* in immunocompromised patients (7). The biology of *S. cerevisiae* has been extensively studied for decades as a model organism for molecular genetics and cell biology studies, and as a cell factory. Its genome (8) is the most thoroughly annotated among eukaryotes, and is a common reference for the annotation of other species. Génolevures focuses on the Hemiascomycete yeasts, a homogeneous phylogenetic group which nonetheless covers a broad range of physiological and ecological lifestyles. Hemiascomycete yeast genes contain introns and alternative splicing is observed (9). Comparative genomics studies in this group have proven informative (10–16); see (5) for review.

The Génolevures Consortium sequences, annotates and analyzes complete genomes from various branches of the Hemiascomycete class, and subjects them to large-scale

*To whom correspondence should be addressed. Tel: +33 540 00 6922; Fax: +33 540 00 6669; Email: david.sherman@labri.fr

[†]The members of the Génolevures Consortium are listed in the Appendix.

in silico and experimental comparisons (Table 1). From these comparisons, we produce classifications of genes, proteins and sequences to address questions of molecular evolution, such as gene conservation, specific genes, function conservation and genome remodeling. We do not provide detailed annotations of individual genes and proteins of *S. cerevisiae*, which are already carefully maintained by MIPS and CYGD (<http://mips.gsf.de/projects/fungi>) (17) and SGD (<http://www.yeastgenome.org/>) (18) as well as in general purpose databases such as UniProt (19) and EMBL (20).

NEW GENOMES IN 2008

The Génolevures Consortium has released four new genomes in 2008, sequenced and assembled at high coverage and completely annotated by a team of 20 curators (Génolevures Consortium, submitted for publication). These genomes are those of *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, previously only sequenced to low or partial coverage, and *Debaryomyces hansenii*, partly resequenced to improve coverage. The former three, plus *K. lactis* (3) and *Ashbya gossypii* (21), are members of the *Saccharomycetaceae* clades that are not descended from the ancestor that is thought to have undergone a whole genome duplication (unlike *C. glabrata* and *Saccharomyces cerevisiae*). The availability in the Génolevures of complete genomes for these five unduplicated genomes will allow close study of the core repertoire of protein families and functions and of the dynamics of genome remodeling.

PROTEIN FAMILIES

Extensive map reshuffling and a wide range of GC compositions from one species to another present a challenge for genome comparison in the yeasts. An essential tool for addressing this challenge is provided by ‘protein families’, a classification of protein-coding gene sequences into phylogenetic groups; members of a family are homologous and in many cases this homology is suggestive of functional similarity. The 48 889 proteins in the predicted proteomes of the nine complete genomes are classified into 7927 families, of which 4369 are common to at least two species. Of these latter families, 2591 are common to the nine species. From these families, other sub-classifications are made, such as the identification of syntenic orthologs in the hemiascomycetous yeasts (Génolevures Consortium, submitted for publication).

Families are computed as follows: four complementary sets of all against all alignments are produced by the Blast (22) and Smith–Waterman (23) algorithms, with and without filtering for homeomorphy, common domain architecture along the full length of the proteins (24). Symmetric distance matrices derived from amino acid identities are constructed and submitted to algorithmic clustering using the MCL ‘Markov clustering’ method (25,26) with a range of statistical parameters. These competing partitions are reconciled using the consensus ensemble clustering

method of Ref. (27) and manually curated using both literature search and systematic comparisons with COG (28) and PIRSF (24) classifications. For a given yeast species, each family may be represented by one or several (paralogous) proteins.

Protein families are available in Génolevures at (<http://genolevures.org/fam/>) and tables describing all families are available in the Datasets area. Information on each family is accessible on individual pages presenting a graphical representation of the members’ relationships and links to individual members. (See URL construction rules, below.) Phyletic patterns and phylogenetic profiles are shown for each family; the former uses a simple one-letter code for each species as in Ref. (28) to indicate whether the family contains at least one member from that species, and the latter expands this to a numerical count of members for each species. Complementary information, such as GO terms, multiple sequence alignments and motif analyses are available for almost all families.

SYNTENY

Although synteny is in general poorly conserved in the Hemiascomycete yeasts (29,30), within phylogenetically delimited groups pairwise coverage is high and can be informative for studying genome dynamics. Pairs of homologous chromosomal regions between two species can be identified by comparing gene content and order, using protein families as an indication of gene-level homology within the regions. We identified these ‘pairwise syntenic blocks’ using the i-ADHoRE method (31) on gene homology relations identified using protein families. The approximately 19 000 pairwise synteny blocks obtained in the fashion can be interactively examined using the genome browser, as described in the Datasets area of the web site, and can be downloaded.

FUSIONS

Chromosomal rearrangements and segmental duplications may result in the creation or destruction of genes, when the breakpoint falls within the boundary of existing genes. ‘Gene fission’ occurs when a gene is broken in this way, and the 5’ (or even the 3’) fragment continues to be transcribed—creating a new, albeit truncated, gene. ‘Gene fusion’ occurs when the chromosomal rearrangement or duplication event leads to a fortuitous combination of previously existing genes, creating a new, longer gene. Fusions and fissions are mechanisms by which new genes can be formed in radical steps that do not obey the tree-like phylogenetic relations between species, and which can lead to significant changes in protein function. These events have to date been rarely taken into account in genome comparisons. Using a new method for detection of genes involved in gene fusion and fission events that makes its computations at the level of paralogous groups (32), we computed a catalogue of such events for complete genomes from the Hemiascomycete yeasts as well as other fungi. Both the paralogous groups of proteins used in the computation, and complete lists of identified fusion/fission events, are

Table 1. Summary of genomes available in Génolevures online, and for illustration, the number of protein-coding genes classified by a selection of the different analyses made available

Species	Assembly	Genome size (Mb)	Number of protein coding genes in:				
			Predicted proteome	Protein families	Synteny blocks	Gene fusions	Tandem arrays
<i>Candida glabrata</i>	Complete	12.3	5210	4869	4850	152	120
<i>Debaryomyces hansenii</i> var. <i>hansenii</i>	Complete	12.2	6286	5213	1155	194	344
<i>Kluyveromyces lactis</i> var. <i>lactis</i>	Complete	10.6	5088	4810	4844	150	96
<i>Kluyveromyces thermotolerans</i>	Complete	10.4	5105	4898	4937		117
<i>Saccharomyces kluyveri</i>	Complete	11.3	5308	5048	5177		133
<i>Yarrowia lipolytica</i>	Complete	20.5	6456	5036	699	162	148
<i>Zygosaccharomyces rouxii</i>	Complete	9.8	4999	4802	4884		112
<i>Candida tropicalis</i>	Partial	15.0	1130 ^b		18.1%		
<i>Kluyveromyces marxianus</i>	Partial	14.0	1546 ^b		49.8%		
<i>Pichia angusta</i>	Partial	9.0	2502 ^a		18.9%		
<i>Pichia sorbitophila</i>	Partial	13.9	1593 ^a		59.2%		
<i>Saccharomyces bayanus</i> var. <i>uvarum</i>	Partial	11.5	2887 ^a		97.9%		
<i>Saccharomyces exiguus</i>	Partial	18.0	1600 ^b		70.7%		
<i>Saccharomyces servazzii</i>	Partial	12.3	1535 ^b		70.2%		
<i>Saccharomyces cerevisiae</i>	Complete	12.1	5769	5381	5450	172	187
<i>Eremothecium (Ashbya) gossypii</i>	Complete	8.7	4718	4474	4584	146	79
<i>Pichia stipitis</i>	Complete	15.4	5816				
<i>Schizosaccharomyces pombe</i>	Complete	14.1	5004			152	

The four genomes at the bottom of the table are not produced by Génolevures but are integrated into the online database for comparative purposes.

^a0.4 × coverage.

^b0.2 × coverage.

now available in Génolevures and are downloadable in the Datasets area.

OTHER DATA SETS

Tandem gene repeats are another means of gene formation. Unlike gene copies resulting from segmental duplications or retrotransposons, truncated or chimeric genes are not observed at the boundaries of tandem gene arrays. Data on tandem gene repeats are available in Datasets area.

The Génosplicing database of spliceosomal introns and intron motifs in Hemiascomycete yeasts (C. Neuvéglise, unpublished data) is available from the Datasets area, and can be used for both for the study of splicing patterns and for the development of methods for predicting gene architectures from genomic data.

The YETI classification of yeast membrane and transporter proteins (33) defined using evolutionary relationships traced using non-ambiguous functional and phylogenetic criteria derived from the TCDB (34) classification system is available in the Datasets area.

Other analyses, such as (35) considering coverage of KEGG pathways (36), are available.

EXPLORING GÉNOLEVURES DATA

The design of the Génolevures on-line database has been revamped to provide improved tools for gaining insight into the mechanisms of eukaryotic molecular evolution. The key questions in the Génolevures use cases are:

(1) What genes exist, as orthologs for my favorite gene or as members of a functional class (keyword search, alignment, homologs)?

- (2) What is known about a given chromosomal element (chromosomal elements)?
- (3) What relations exist in a protein family (protein families)?
- (4) How are the individual genomes organized (maps and genome browser, synteny)?
- (5) How are genes and proteins classified (data sets for fusions, tandem repeats, introns, pathways, YETI)?

Additionally, the web site provides a query system that simultaneously searches for and can return: genes that have or may have a translation product, RNA and other genes that may have a transcription product only, *cis*-active elements and cross-genome protein families.

AVAILABILITY OF DATA AND URL CONSTRUCTION RULES

All data from the Génolevures web site are freely available, and instructions for proper citation are included in each section. The Génolevures web site is developed using a 'representational state transfer' architecture (37) and URLs for individual identified resources built from the database can be constructed systematically. Identified resources include chromosomal elements, such as genes (prefix/**elt**/*Abbrev*/*Element_identifier*), protein families (prefix/**fam**/*Family_identifier*), DNA sequences (prefix/**seq**/*Sequence_identifier*) and biomolecular pathways (prefix/**pathway**/*Pathway_identifier*). Nomenclatures for genome abbreviations, systematic chromosomal element names, and protein family names are described in the Documentation area of the web site.

Génolevures uses a bespoke object model mapped to a relational database, and uses the SOFA (38) and GO (39)

ontologies extensively. The web interface is developed in Mason. Key software components are provided by the NCBI Blast program suite (22) and the Stein Lab's Generic Genome Browser (40).

ONGOING DEVELOPMENTS

The Génolevures Consortium continues its effort to sequence and annotate other complete genomes of Hemiascomycete yeasts. These genomes as well as the incorporation of other genomes annotated by third parties will help to refine the classifications presented above.

SUPPLEMENTARY DATA

Supplementary Data are available Online at <http://genolevures.org/>.

FUNDING

French National center for Scientific Research (CNRS) (GDR 2354, partial); French National Research Agency (ANR) (ANR-05-BLAN-0331; GENARISE, partial); Région Aquitaine ('Pôle de Recherche en Informatique') (2005-1306001AB, partial) and ACI IMPBIO (IMPB114, partial) ('Génolevures En Ligne'). Funding for open access charge: CNRS GDR 2354.

Conflict of interest statement. Large-scale Smith-Waterman alignments used in constructing protein families were performed using GenCore 6 software licensed from Bioceleration Inc., 77 Milltown Road, Suite B4, East Brunswick, NJ 08816, USA.

REFERENCES

- Soucié, J.L. and Génolevures Consortium. (2000) Special issue: Génolevures. *FEBS Lett.*, **487**, 1–149.
- Sherman, D.J., Durrrens, P., Beyne, E., Nikolski, M. and Soucié, J.L. (2004) Génolevures: comparative genomics and molecular evolution of hemiascomycetous yeasts. *Nucleic Acids Res.*, **32**, D315–D318.
- Dujon, B., Sherman, D.J., Fischer, G., Durrrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E. *et al.* (2004) Genome evolution in yeasts. *Nature*, **430**, 35–44.
- Sherman, D., Durrrens, P., Iragne, F., Beyne, E., Nikolski, M. and Soucié, J.L. (2006) Génolevures complete genomes provide data and tools for comparative genomics of hemiascomycetous yeasts. *Nucleic Acids Res.*, **34**, D432–D435.
- Dujon, B. (2006) Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet.*, **22**, 375–387.
- Kurtzmann, C.P. and Fell, J.W. (eds). (1998) *The Yeasts, A Taxonomic Study*, 4th edn. Elsevier, Amsterdam.
- Tawfik, O.W., Papsian, C.J., Dixon, A.Y. and Potter, L.M. (1989) Saccharomyces cerevisiae pneumonia in a patient with acquired immune deficiency syndrome. *J. Clin. Microbiol.*, **27**, 1689–1691.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M. *et al.* (1996) Life with 6000 genes. *Science*, **274**, 546, 563–567.
- Neuvéglise, C., Chalvet, F., Wincker, P., Gaillardin, C. and Casaregola, S. (2005) Mutator-like element in the yeast *Yarrowia lipolytica* displays multiple alternative splicings. *Eukaryot. Cell*, **4**, 615–624.
- Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A. and Johnston, M. (2003) Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science*, **301**, 71–76.
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. and Lander, E.S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, **423**, 241–254.
- Richard, G.F., Kerrest, A., Lafontaine, I. and Dujon, B. (2005) *Mol. Biol. Evol.*, **22**, 1011–1023.
- Fabre, E., Muller, H., Therizols, P., Lafontaine, I., Dujon, B. and Fairhead, C. (2005) Comparative genomics of hemiascomycete yeasts: genes involved in DNA replication, repair, and recombination. *Mol. Biol. Evol.*, **22**, 856–873.
- Knop, M. (2006) Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding. *Bioessays*, **28**, 696–708.
- Rokas, A. and Carroll, S.B. (2005) More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Mol. Biol. Evol.*, **22**, 1337–1344.
- Blank, L.M., Lehmbeck, F. and Sauer, U. (2005) Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res.*, **5**, 545–558.
- Mewes, H.W., Dietmann, S., Frishman, D., Gregory, R., Mannhaupt, G., Mayer, K., Muensterkötter, M., Ruepp, A., Spannagl, M., Stuempfen, V. *et al.* (2008) MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Res.*, **36**, D196–D201.
- Nash, R., Weng, S., Hitz, B., Balakrishnan, R., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E. *et al.* (2007) Expanded protein information at SGD: new pages and proteome browser. *Nucleic Acids Res.*, **35**, D468–D471.
- Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. *et al.* (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.
- Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., Browne, P., van den Broek, A., Castro, M., Cochrane, G. *et al.* (2005) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **33**, D29–D33.
- Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S. *et al.* (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science*, **304**, 304–307.
- Altschul, F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
- Wu, C.H., Nikolskaya, A., Huang, H., Yeh, L.S., Natale, D.A., Vinayaka, C.R., Hu, Z.Z., Mazumder, R., Kumar, S., Kourtesis, P., Ledley, R.S. *et al.* (2004) PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Res.*, **32**, D201–D205.
- Van Dongen, S. (2000) Graph clustering by flow simulation. *PhD Thesis*, University of Utrecht, The Netherlands.
- Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, **30**, 1575–1584.
- Nikolski, M. and Sherman, D.J. (2007) Family relationships: should consensus reign?—consensus clustering for protein families. *Bioinformatics*, **23**, e71–e76.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N. *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, **4**, 41.
- Llorente, B., Malpertuy, A., Neuvéglise, C., de Montigny, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P. *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. *FEBS Lett.*, **487**, 101–112.
- Fischer, G., Rocha, E.P., Brunet, F., Vergassola, M. and Dujon, B. (2006) Highly variable rates of genome rearrangements between hemiascomycetous yeast lineages. *PLoS Genet.*, **2**, e32.
- Simillion, C., Vandepoel, K., Saeys, Y. and Van de Peer, Y. (2004) Building genomic profiles for uncovering segmental homology in the twilight zone. *Genome Res.*, **14**, 1095–1106.

32. Durrrens,P., Nikolski,M., Sherman,D.J. (2008) Fusion and fission of genes define a metric between fungal genomes. *PLOS Comput. Biol.*, **4**, e1000200.
33. De Hertogh,B., Hancy,F., Goffeau,A. and Baret,P.V. (2005) Emergence of species-specific transporters during evolution of the hemiascomycete phylum. *Genetics*, **172**, 771–781.
34. Saier,M.H. (2000) A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Bio. Rev.*, **64**, 354–411.
35. Iragne,F., Nikolski,M. and Sherman,D. (2008) Extrapolation of metabolic pathways as an aid to modelling completely sequenced nonSaccharomyces yeasts. *FEMS Yeast Res.*, **8**, 132–139.
36. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
37. Fielding,R. and Taylor,R.N. (2002) Principled design of the modern web architecture. *ACM Trans. Internet Techn.*, **2**, 115–150.
38. Eilbeck,K., Lewis,S.E., Mungall,C.J., Yandell,M., Stein,L., Durbin,R. and Ashburner,M. (2005) The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.*, **6**, R44.
39. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.*, **25**, 25–29.
40. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W., Arva,A. and Lewis,S. (2002) The Generic Genome Browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.

APPENDIX

The Génolevures Consortium is coordinated by J.L.Souciet and is composed of laboratories from Institut de Biologie et de Technologies de Saclay (iBiTec-S), CEA, F-91191 Gif-sur-Yvette CEDEX, France (C.Marck); Laboratoire de Chimie Bactérienne, CNRS-UPR9043, Université de la Méditerranée, 31 Chemin Joseph Aiguier 13402 Marseille Cedex 20

(E.Talla); Laboratoire Microbiologie, Adaptation et Pathogénie, Université Claude Bernard - Lyon 1, CNRS/UMR 5240, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne, France. (M. Lemaire); Génoscope (CEA), 2 rue Gaston Crémieux, BP 191, F-91057 Evry Cedex, France (J. Weissenbach); Laboratoire Bordelais de Recherche en Informatique, LaBRI, UMR 5800 CNRS, INRIA Bordeaux Sud-Ouest (MAGNOME), 351 cours de la Libération, 33405 Talence Cedex, France (D.Sherman); Laboratoire de Microbiologie et Genetique Moleculaire, UMR 1238 INRA/2585 CNRS/Institut National Agronomique Paris-Grignon, AgroParisTech, F-78850 Thiverval-Grignon, France (C.Gaillardin); Institut Pasteur/URA2171 CNRS and Université Pierre et Marie Curie UFR927, 25 rue du Docteur Roux, F-75724 Paris Cedex 15, France (B. Dujon); Laboratoire Génétique Moléculaire, Génomique, Micribiologie, UMR 7156, Université Louis Pasteur, Institut de Botanique, 28 rue Goethe, F-67000 Strasbourg, France (S. Potier, J.-L. Souciet); Washington University School of Medicine, Department of Genetics, Campus Box 8510 4566 Scott Ave., St Louis, MO 63110 USA (M. Johnston); Washington University School of Medicine, Genome Sequencing Center, Campus Box 8501, 4444 Forest Park Avenue, St Louis, MO 63108 USA; (R. Fulton); Pasteur Genopole: Intégration et Analyse Génomique, 28 rue du Docteur Roux, F-75724 Paris Cedex 15, France (L.Frangeul); Unité de Génétique, Université Catholique de Louvain, Croix du Sud 2 bte 14, 1348 Louvain-la-Neuve, Belgique (P.Baret); Institut de Biologie Moléculaire et Cellulaire, Architecture et Réactivité de L'ARN, IBMC (UPR 9002 CNRS), 15 rue René Descartes, 67084 Strasbourg, France (E.Westhof).