D348–D352 *Nucleic Acids Research, 2005, Vol. 33, Database issue*
doi:10.1093/nar/gki009

The Ashbya Genome Database (AGD)—a tool for the yeast community and genome biologists

Leandro Hermida[1], Sophie Brachat, Sylvia Voegeli, Peter Philippsen and Michael Primig[1,*]

Department of Applied Microbiology, Biozentrum and [1]Swiss Institute of Bioinformatics, Klingelbergstrasse 50-70, CH-4056 Basel, Switzerland

Received July 23, 2004; Revised and Accepted September 14, 2004

ABSTRACT

The Ashbya Genome Database (AGD) is a comprehensive online source of information covering genes from the filamentous fungus *Ashbya gossypii*. The database content is based upon comparative genome annotation between *A.gossypii* and the closely related budding yeast *Saccharomyces cerevisiae* taking both sequence similarity and synteny (conserved order and orientation) into account. Release 2 of AGD contains 4718 protein-encoding loci located across seven chromosomes. Information can be retrieved using systematic or standard locus names from *A.gossypii* as well as budding and fission yeast. Approximately 90% of the genes in the genome of *A.gossypii* are homologous and syntenic to loci of budding yeast. Therefore, AGD is a useful tool not only for the various yeast communities in general but also for biologists who are interested in evolutionary aspects of genome research and comparative genome annotation. The database provides scientists with a convenient graphical user interface that includes various locus search and genome browsing options, data download and export functionalities and numerous reciprocal links to external databases including SGD, MIPS, GeneDB, KEGG, GermOnline and Swiss-Prot/TrEMBL. AGD is accessible at http://agd.unibas.ch/.

INTRODUCTION

Since the first eukaryotic genome sequence of budding yeast *Saccharomyces cerevisiae* was published in 1996 (1), similar projects have been completed for other yeast species including *Schizosaccharomyces pombe* and *Ashbya gossypii* (2–7). The identification of open reading frames (ORFs) and loci that encode RNAs was substantially facilitated by comparative genomics because this approach exploits the conservation of gene order and orientation (synteny) in several related genomes. It is therefore possible to distinguish genuine ORFs from annotation artifacts and to assign start codons and intron/exon boundaries in a more reliable manner. Moreover, aligning promoter sequences from closely related species yields novel conserved elements that might be involved in transcriptional regulation (2,3).

The genome sequence of *A.gossypii* has proven to be extremely useful for refining the annotation of the budding yeast genome (8) and, most importantly, has provided the ultimate proof that *S.cerevisiae* has undergone a whole-genome duplication event (5). The sequence and annotation data are therefore highly interesting from an evolutionary standpoint because they permit reconstruction of genome rearrangements and gene deletions in budding yeast. Furthermore, *A.gossypii* is an excellent model system to study important biological problems like polar growth using genetic, biochemical and genomic experimental approaches (9–13). It should be emphasized that high-density oligonucleotide microarrays will become available for *A.gossypii* in the future, opening up an avenue to extensive profiling studies of the poorly understood transcriptional program underlying the life cycle of this interesting filamentous fungus (P. Philippsen, unpublished data). The annotation information stored in Ashbya Genome Database (AGD) is constantly experimentally verified and updated. Researchers using the database are encouraged to provide feedback on errors and inconsistencies to further improve the reliability of AGD's entries.

Here, we describe Release 2 of the AGD, a comprehensive and cross-referenced source of manually verified genome annotation data accessible at http://agd.unibas.ch/.

DATABASE DEVELOPMENT AND CROSS-CONNECTION

The database and web site were constructed using the Ensembl application programming interface (API) and base web code version 14 (14–16). Details of the hardware and software required to create an Ensembl development environment can be found at http://www.ensembl.org/Docs/wiki/

*To whom correspondence should be addressed. Tel: +41 61 267 2098; Fax: +41 61 267 3398; Email: michael.primig@unibas.ch

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use permissions, please contact journals.permissions@oupjournals.org.

© 2005, the authors

html/EnsemblDocs/InstallEnsemblWebsite.html. The MySQL database schema and structure was created to contain *A.gossypii* data in an Ensembl-compliant format. MySQL database creation and data population was accomplished with a series of Perl (http://www.perl.org/) scripts using the source GenBank sequence data and annotation files for *A.gossypii* provided by Dietrich *et al*. (5). The loading scripts also facilitate quick and seamless re-population of the database when updated GenBank files are made available. *A.gossypii* GenBank files for AGD Release 1 are presently accessible on the supplementary download page http://agd.unibas.ch/Ashbya_gossypii/download_ashbya. Customized web pages were created specifically for AGD using the Ensembl API and base web site structure. The database schema and scripts are available upon request.

Hyperlinks into AGD from external sources are possible via the uniform resource locator (URL) http://agd.unibas.ch/Ashbya_gossypii/unisearch?type=Gene&q=[KEYWORD] where [KEYWORD] can be a budding or fission yeast gene systematic name (e.g. *YPR175W*, *SPBP8B7.14C*) or standard name (*DPB2*) or an *A.gossypii* systematic name (e.g. *AEL267C*).

THE SCOPE AND QUALITY OF AGD

The goal of AGD Release 2 is to cover all 4718 currently annotated loci from *A.gossypii* that are based on the most upstream ATG start triplet of a given ORF. The database provides convenient access to annotation information available for homologous genes from budding and fission yeasts. Note that the annotation approach in rare cases leads to start codons being wrongly assigned. This is in part due to the GC-content being higher in *A.gossypii* than in budding yeast (53% versus 38%) (5). We are in the process of experimentally verifying the annotation of ORFs and start codons in a number of questionable cases including some reported by users. The results of this ongoing work comprising DNA re-sequencing and/or verification of transcriptional start sites by RACE–PCR are immediately incorporated into update files submitted to Genbank and AGD. Lists of loci for which updated ORFs and/or start codons are available between database releases are accessible via the supplementary download page of AGD.

INFORMATION AND DATA RETRIEVAL

Search options

Currently, it is possible to search AGD using *A.gossypii*'s systematic locus name (e.g. *ADR058C*) and the systematic or standard names from budding or fission yeast, respectively (Figure 1A). Note that fission yeast genes are linked to *A.gossypii* loci either directly by an automated BLAST similarity search (S. Brachat and P. Philippsen, unpublished data) or via their budding yeast homologs (V. Wood, personal communication). A popup menu includes several refined options like gene, mRNA, peptide or contig/clone. Moreover, a wildcard search using truncated gene names (e.g. *CDC**, *RAD**, *SRB** and *SPO**) allows the retrieval of several loci from *A.gossypii* that are shown together with their putative homologs from budding and fission yeast. It is also possible to display chromosomal regions by defining the chromosome number and the base coordinates (Figure 1B).

Genome browsing and various levels of views

The welcome page provides a whole-genome view via an interactive image of the seven *A.gossypii* chromosomes that enables scientists to home in on any region of interest. By clicking on a chromosome, the user gains access to the *MapView* page, which gives some general information and statistics (gene content and chromosome length) about the chromosome. From the *MapView* page, the user can click anywhere on the chromosome ideogram to jump to the contig-level view of *A.gossypii* features at that point. For the selection of specific chromosome regions, users can click on the link to the *AnchorView* page, which allows the selection of two features on a chromosome as anchor points to display the contig-level region between them. Both methods of browsing lead to the *ContigView* page, which comprises four sections: *Chromosome*, *Overview* (Figure 1C), *Detailed View* (Figure 1D) and *Basepair View*. Each of these sections can be hidden by clicking on the − tick box. The views are fully interactive and enable the user to click on any of the indicated genes, transcripts and legends to call up the respective locus report page. The *Detailed View* and *Basepair View* sections include zoom, image size and window jump functions allowing flexible genome navigation.

AGD report pages

The *AGD Gene Report* page begins with sections that indicate the gene name, database ID and the genomic location (Figure 2A). The *Description* section contains information on the homologous *S.cerevisiae* and *S.pombe* loci (indicating systematic and standard names) and the *External Links* section provides access to annotation databases for budding yeast [SGD (17), MIPS (18)], fission yeast [GeneDB (19)] and the relevant sections of databases covering multiple species [KEGG (20), GermOnline (21–23), Swiss-Prot/TrEMBL (24)]. These links enable users to access information obtained about homologous genes in related species that may be relevant for the locus in *A.gossypii* as well. The *Predicted Transcripts* section contains a graphical display of the target gene's chromosomal localization and links to the transcript, exon and protein report pages. The *Transcript/Translation Summary* section also gives access to transcript, exon and protein information, and contains a graphical display of the mRNA and protein structures (Figure 2B–D). The [*View transcript information*] link leads the user to the *AGD Transcript Report* page, which contains additional information regarding the mRNA sequence. Various markups can be added to the mRNA sequence display, which show the user the transcript's underlying exons, protein codons and peptide sequence. The *AGD Exon Report* page adds detailed information about the exon/intron structure of the gene and displays flanking 5′ and 3′ untranslated regions. The *AGD Protein Report* page can be called up via the [*View protein information*] link and shows peptide sequence and property information.

Data export

From the *Download AGD Release 2* link in the welcome page and the *export data* link in the AGD *Gene Report* pages, the

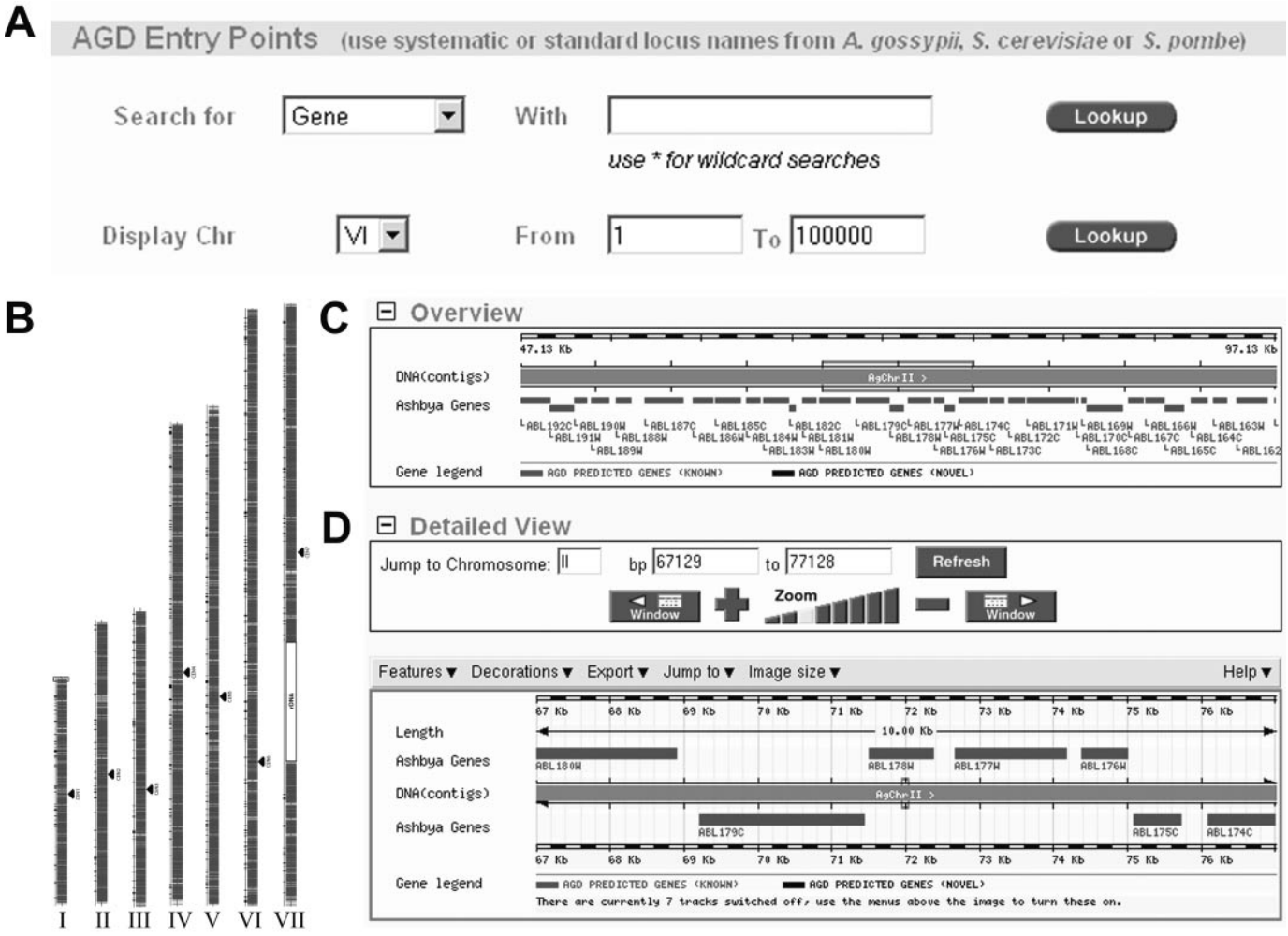

Figure 1. Search and browsing options in AGD. (**A**) The search form and (**B**) the chromosomal browsing function. (**C** and **D**) The *Overview* and the *Detailed View* of the *ContigView* page, respectively.

individual mRNA or peptide sequences can be retrieved in FASTA format via an appropriate form (Figure 2E). A user can also download entire regions of the *A.gossypii* genome complete with annotations in EMBL, GenBank or FASTA format, or region feature lists for uploading into other databases in GFF, tab-delimited and comma-delimited formats. Finally, it is possible to retrieve the complete source AGD database content as GenBank files, the complete set of ORF translations in FASTA format and a correspondence list of *A.gossypii* and *S.cerevisiae* locus names by following the link to the supplementary download page.

FUTURE DEVELOPMENT

We are currently implementing a BLAST functionality that will enable users to query *A.gossypii* genes or gene products against the *S.cerevisiae* genome. At a later stage, this feature will include the genomes of other species as well. In addition, we plan to provide a two-way interactive synteny viewer that displays the chromosomal organization of loci from *A.gossypii* together with their counterparts from budding yeast. Future releases of AGD will contain a number of additional features including non-coding RNAs and GeneOntology assignments to improve annotation and query options. High-density oligonucleotide microarrays covering the entire genome will become available in the foreseeable future, and it is therefore our intention to extend the database's functionalities to cover high-throughput expression data. Finally, results from large-scale gene deletion studies produced by functional genomics will also be incorporated and information on BAC clones and plasmids containing *A.gossypii* genes will be made available.

CONCLUSION

AGD is a highly useful online source of information for the extensive budding and fission yeast communities and also targets scientists interested in the evolution of genomes. AGD's content is based upon comparative genome annotation between *S.cerevisiae* and its evolutionary ancestor *A.gossypii* whereby every single locus identified by automatic annotation was manually inspected. Gene annotation is further improved by a sustained effort to experimentally

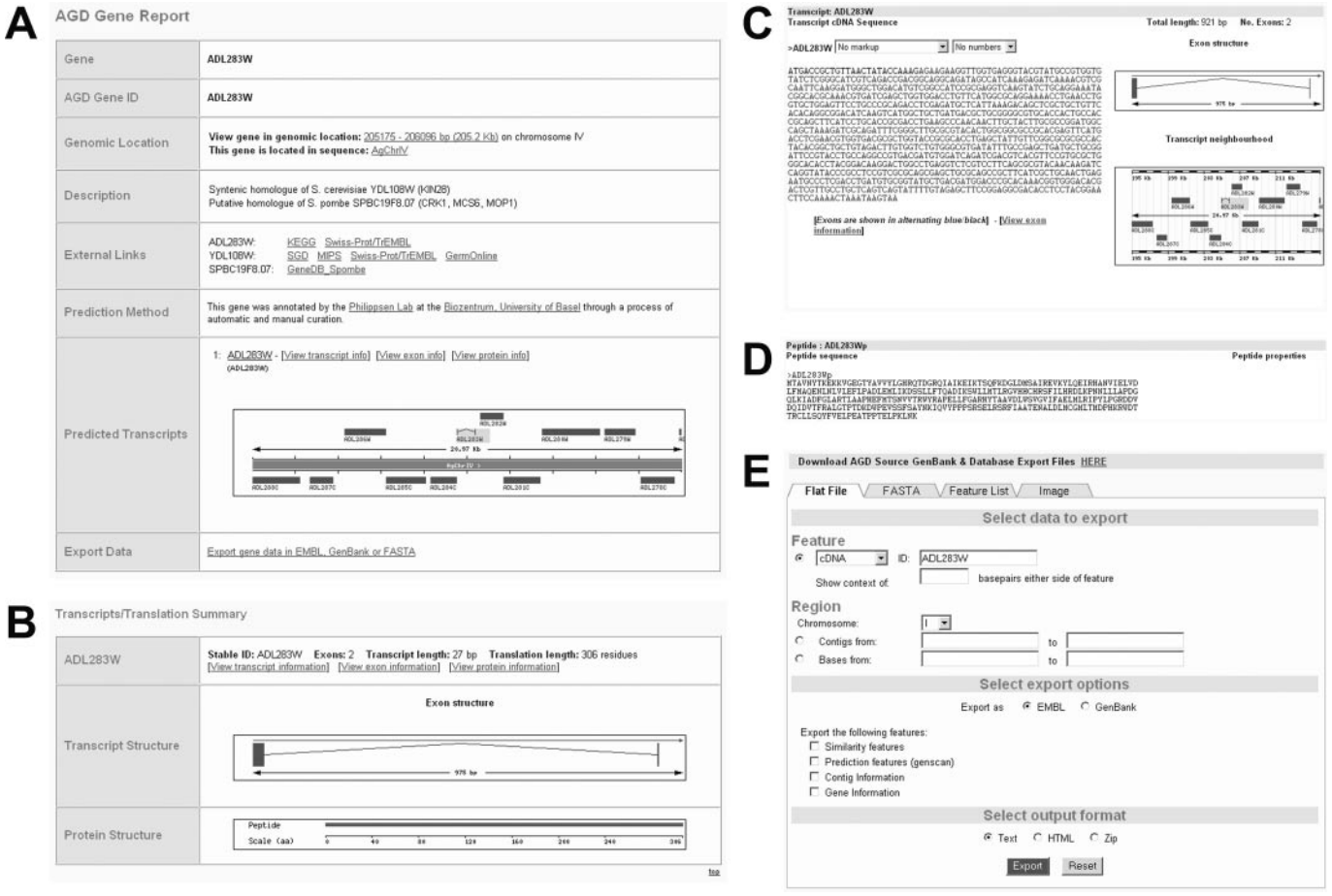

Figure 2. The locus report page of AGD. (**A** and **B**) The *Gene Report* and *Transcript*/*Translation* sections of the *Gene Report* page for the *A.gossypii* homolog of *KIN28* (*S.cerevisiae*) and *SPBC19F8.07* (*S.pombe*). (**C** and **D**) Examples of the *Transcript* and *Peptide* view, and (**E**) the data *ExportView* interface that is accessible via links in the welcome page and each *Gene Report* page.

verify all questionable cases especially those reported by researchers who use AGD.

ACKNOWLEDGEMENTS

We thank R. Jenni, R. Poehlmann (BioPhIT), N. Brown ([BC]^2) and D. Flanders (FMI) for excellent IT support, and M. Aslett and V. Wood (Sanger Institute) for critical reading of the manuscript. L.H. is supported by the Swiss Institute of Bioinformatics.

REFERENCES

1. 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, 563–547.
2. 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.
3. 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.
4. Wood,V., Gwilliam,R., Rajandream,M.A., Lyne,M., Lyne,R., Stewart,A., Sgouros,J., Peat,N., Hayles,J., Baker,S. *et al.* (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature*, **415**, 871–880.
5. 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.
6. Dujon,B., Sherman,D., Fischer,G., Durrens,P., Casaregola,S., Lafontaine,I., De Montigny,J., Marck,C., Neuveglise,C., Talla,E. *et al.* (2004) Genome evolution in yeasts. *Nature*, **430**, 35–44.
7. Kellis,M., Birren,B.W. and Lander,E.S. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, **428**, 617–624.
8. Brachat,S., Dietrich,F.S., Voegeli,S., Zhang,Z., Stuart,L., Lerch,A., Gates,K., Gaffney,T. and Philippsen,P. (2003) Reinvestigation of the *Saccharomyces cerevisiae* genome annotation by comparison to the genome of a related fungus: *Ashbya gossypii*. *Genome Biol.*, **4**, R45.
9. Wright,M.C. and Philippsen,P. (1991) Replicative transformation of the filamentous fungus *Ashbya gossypii* with plasmids containing *Saccharomyces cerevisiae* ARS elements. *Gene*, **109**, 99–105.
10. Steiner,S., Wendland,J., Wright,M.C. and Philippsen,P. (1995) Homologous recombination as the main mechanism for DNA integration and cause of rearrangements in the filamentous ascomycete *Ashbya gossypii*. *Genetics*, **140**, 973–987.
11. Wendland,J., Ayad-Durieux,Y., Knechtle,P., Rebischung,C. and Philippsen,P. (2000) PCR-based gene targeting in the filamentous fungus *Ashbya gossypii*. *Gene*, **242**, 381–391.
12. Knechtle,P., Dietrich,F. and Philippsen,P. (2003) Maximal polar growth potential depends on the polarisome component AgSpa2 in the filamentous fungus *Ashbya gossypii*. *Mol. Biol. Cell*, **14**, 4140–4154.

13. Bauer,Y., Knechtle,P., Helfer,H., Wendland,J. and Philippsen,P. (2004) A Ras-like GTPase is involved in hyphal growth guidance in the filamentous fungus *Ashbya gossypii*. *Mol. Biol. Cell*, **15**, 4622–4632.
14. Stabenau,A., McVicker,G., Melsopp,C., Proctor,G., Clamp,M. and Birney,E. (2004) The Ensembl core software libraries. *Genome Res.*, **14**, 929–933.
15. Curwen,V., Eyras,E., Andrews,T.D., Clarke,L., Mongin,E., Searle,S.M. and Clamp,M. (2004) The Ensembl automatic gene annotation system. *Genome Res.*, **14**, 942–950.
16. Birney,E., Andrews,T.D., Bevan,P., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cuff,J., Curwen,V., Cutts,T. *et al.* (2004) An overview of Ensembl. *Genome Res.*, **14**, 925–928.
17. Christie,K.R., Weng,S., Balakrishnan,R., Costanzo,M.C., Dolinski,K., Dwight,S.S., Engel,S.R., Feierbach,B., Fisk,D.G., Hirschman,J.E. *et al.* (2004) Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.*, **32**, D311–D314.
18. Mewes,H.W., Amid,C., Arnold,R., Frishman,D., Guldener,U., Mannhaupt,G., Munsterkotter,M., Pagel,P., Strack,N., Stumpflen,V. *et al.* (2004) MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.*, **32**, D41–D44.
19. Hertz-Fowler,C., Peacock,C.S., Wood,V., Aslett,M., Kerhornou,A., Mooney,P., Tivey,A., Berriman,M., Hall,N., Rutherford,K. *et al.* (2004) GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res.*, **32**, D339–D343.
20. Kanehisa,M., Goto,S., Kawashima,S., Okuno,Y. and Hattori,M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
21. Wiederkehr,C., Basavaraj,R., Sarrauste de Menthiere,C., Koch,R., Schlecht,U., Hermida,L., Masdoua,B., Ishii,R., Cassen,V., Yamamoto,M. *et al.* (2004) Database model and specification of GermOnline Release 2.0, a cross-species community annotation knowledgebase on germ cell differentiation. *Bioinformatics*, **20**, 808–811.
22. Primig,M., Wiederkehr,C., Basavaraj,R., Sarrauste de Menthiere,C., Hermida,L., Koch,R., Schlecht,U., Dickinson,H.G., Fellous,M., Grootegoed,J.A. *et al.* (2003) GermOnline, a new cross-species community annotation database on germ-line development and gametogenesis. *Nature Genet.*, **35**, 291–292.
23. Wiederkehr,C., Basavaraj,R., Sarrauste de Menthiere,C., Hermida,L., Koch,R., Schlecht,U., Amon,A., Brachat,S., Breitenbach,M., Briza,P. *et al.* (2004) GermOnline, a cross-species community knowledgebase on germ cell differentiation. *Nucleic Acids Res.*, **32**, D560–D567.
24. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.C., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C., Phan,I. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.