D560–D567 *Nucleic Acids Research, 2004, Vol. 32, Database issue*
DOI: 10.1093/nar/gkh055

GermOnline, a cross-species community knowledgebase on germ cell differentiation

C. Wiederkehr, R. Basavaraj, C. Sarrauste de Menthière[1], L. Hermida, R. Koch, U. Schlecht, A. Amon[2], S. Brachat, M. Breitenbach[3], P. Briza[3], S. Caburet[4], M. Cherry[5], R. Davis[5], A. Deutschbauer[5], H. G. Dickinson[6], T. Dumitrescu[7], M. Fellous[4], A. Goldman[8], J. A. Grootegoed[9], R. Hawley[10], R. Ishii[11], B. Jégou[12], R. J. Kaufman[13], F. Klein[14], N. Lamb[1], B. Maro[15], K. Nasmyth[16], A. Nicolas[17], T. Orr-Weaver[18], P. Philippsen, C. Pineau[12], K. P. Rabitsch[16], V. Reinke[19], H. Roest[9], W. Saunders[7], M. Schröder[13], T. Schedl[20], M. Siep[9], A. Villeneuve[5], D. J. Wolgemuth[21], M. Yamamoto[11], D. Zickler[22], R. E. Esposito[23] and M. Primig*

Biozentrum and Swiss Institute of Bioinformatics, Basel, Switzerland, [1]Institut de Génétique Humaine, Montpellier, France, [2]Howard Hughes Medical Institute, MIT, Cambridge, MA, USA, [3]University of Salzburg, Austria, [4]Hôpital Cochin, Paris, France, [5]Stanford University, Palo Alto, CA, USA, [6]Oxford University, UK, [7]University of Pittsburgh, PA, USA, [8]University of Sheffield, UK, [9]Erasmus University, Rotterdam, The Netherlands, [10]Stowers Institute, Kansas City, KS, USA, [11]University of Tokyo, Japan, [12]University of Rennes I, Bretagne, France, [13]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, MI, USA, [14]University of Vienna, Austria, [15]Université Paris VI, France, [16]Institute of Molecular Pathology, Vienna, Austria, [17]Institut Curie, Paris, France, [18]Whitehead Institute, MIT, Cambridge, MA, USA, [19]Yale University, New Haven, CT, USA, [20]Washington University, St Louis, MO, USA, [21]Columbia University, New York, NY, USA, [22]Université Paris Sud, Orsay, France and [23]University of Chicago, IL, USA

Received August 15, 2003; Revised and Accepted September 23, 2003

ABSTRACT

GermOnline provides information and microarray expression data for genes involved in mitosis and meiosis, gamete formation and germ line development across species. The database has been developed, and is being curated and updated, by life scientists in cooperation with bioinformaticists. Information is contributed through an online form using free text, images and the controlled vocabulary developed by the GeneOntology Consortium. Authors provide up to three references in support of their contribution. The database is governed by an international board of scientists to ensure a standardized data format and the highest quality of GermOnline's information content. Release 2.0 provides exclusive access to microarray expression data from *Saccharomyces cerevisiae* and *Rattus norvegicus*, as well as curated information on ~700 genes from various organisms. The locus report pages include links to external databases that contain relevant annotation, microarray expression and proteome data. Conversely, the *Saccharomyces* Genome Database (SGD), *S.cerevisiae* GeneDB and Swiss-Prot link to the budding yeast section of GermOnline from their respective locus pages. GermOnline, a fully operational prototype subject-oriented knowledgebase designed for community annotation and array data visualization, is accessible at http://www.germonline.org. The target audience includes researchers who work on mitotic cell division, meiosis, gametogenesis, germ line development, human reproductive health and comparative genomics.

INTRODUCTION

The recent developments of large-scale DNA sequencing techniques, microarray technology and bioinformatics enable scientists to identify potential genes, study their patterns of expression and analyse their promoters at the genomic level (1). Novel genetic and biochemical high-throughput approaches to determining the functions and interactions of gene products further contribute to the rapidly growing body of data that biologists and bioinformaticists need to process and interpret (2–5). To meet this challenge, a common biological language is being developed by the GeneOntology (GO) Consortium. This project is a collaborative effort of many major genome databases e.g. the *Saccharomyces* Genome Database (SGD), The Arabidopsis Information Resource (TAIR), Flybase, Wormbase and the Mouse

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

 © Oxford University Press 2004; all rights reserved

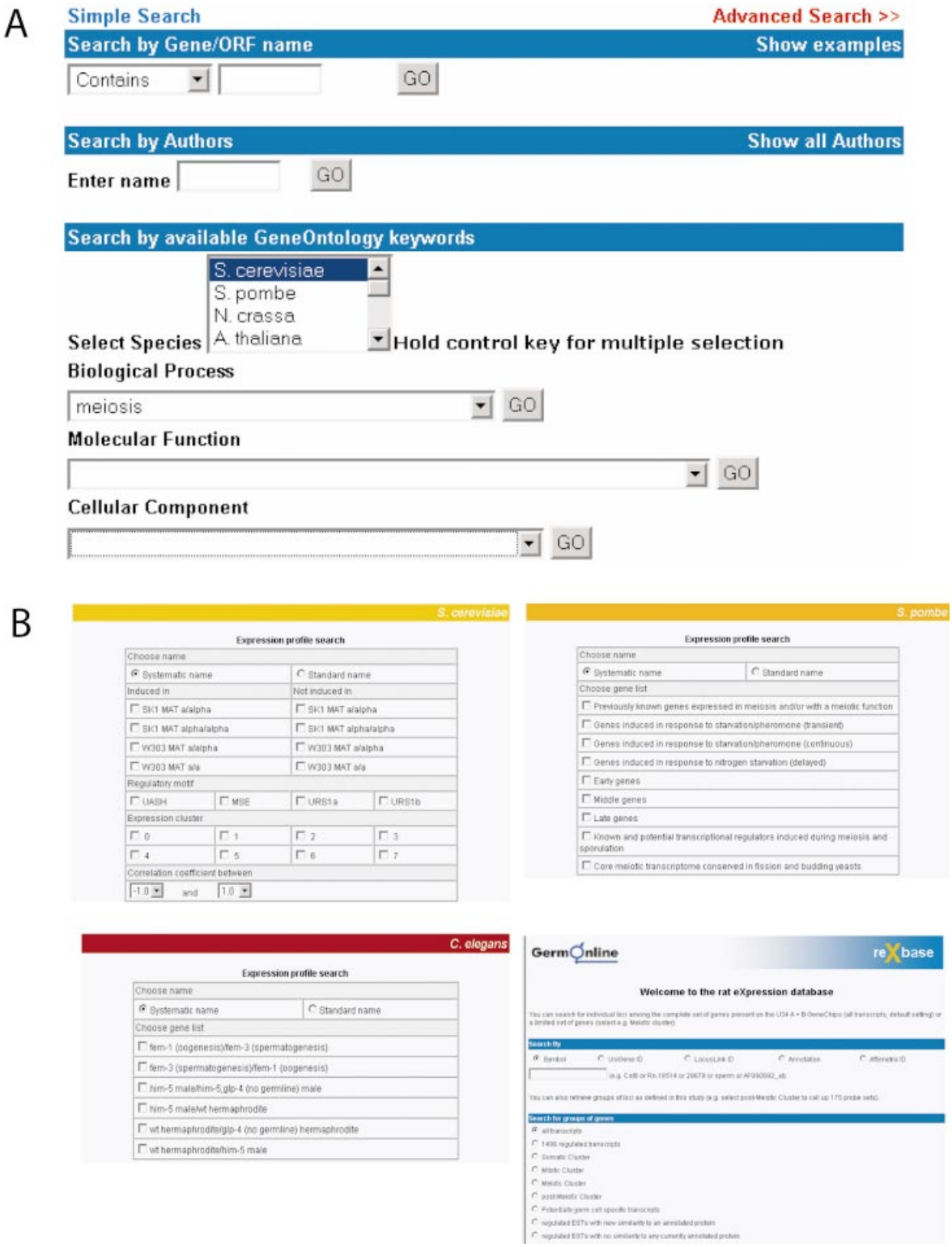


Figure 1. Simple and advanced search options. Users can search for individual genes and authors who contributed to the database as well as GO keywords that were associated with the gene in question during curation as shown in (**A**). Advanced search forms that enable the user to retrieve lists of genes that display a certain expression pattern or have other properties (e.g. the presence of known regulatory sites in their promoters) are summarized for several species as indicated in (**B**).

Genome Database (MGD) (6). The aim is to develop a coherent semantic framework for gene nomenclature and description across species. The controlled vocabulary describes the biological process genes are involved in, the molecular functions their products fulfil and the cellular component to which gene products localize (7).

A common feature of genome databases is that each contains general information on a particular species (or genus) and it is often laborious to compare evolutionarily conserved loci across organisms from within their locus pages. Genome databases are developed and maintained by curators and bioinformaticists who are responsible for defining and naming a locus through a unique identification code and gene annotation in cooperation with the members of sequencing consortia. While researchers provide the raw data they are usually not directly involved in the formulation of gene descriptions, the choice of GO keywords or the decision as to how much emphasis is given to models and theories discussed in the literature. Some databases have recently begun to solicit contributions from users for information content, e.g. SGD (8); however, this remains supplementary information and does not constitute the core knowledge provided by the database.

GermOnline is a new, unique comprehensive subject-oriented database containing information from diverse species specifically focused on mitotic growth and meiotic development. These important biological processes have been extensively studied in a variety of model organisms including *Homo sapiens* and a substantial amount of knowledge and genomics data is available that needs to be organized; the field is therefore particularly suitable for community-based annotation presented in the context of high-throughput data. Information about gene function in experimental systems used to study sexual reproduction will ultimately help to improve human reproductive health.

Scientists who publish their findings in peer-reviewed journals contribute the essence of their results using free text, images and/or figures (including legends), automatically updated GO keywords and original references. This information is presented within the context of relevant microarray and proteomics data as well as other external sources of gene annotation. Multiple contributions on a given locus from different groups are encouraged to ensure that gene function(s) are broadly discussed and different perspectives covered in depth. The present paper describes how scientists can retrieve data from and contribute information to GermOnline and outlines the curation procedure. The latter is overseen by the members of an international board of scientists who work in the fields of meiosis, germ line development and gametogenesis using a variety of model organisms.

Programmatic aspects of the database model, database interconnectivity and mirroring as well as automatic update of locus lists and GO keywords will be published elsewhere. The database is accessible at http://www.germonline.org. To ensure round-the-clock availability and convenient access, a network of mirrors in Europe (http://germonline.igh.cnrs.fr), Japan (http://germonline.biochem.s.u-tokyo.ac.jp/) and the US (http://germonline.yeastgenome.org) has been installed.

THE SCOPE OF GermOnline

It is the aim of GermOnline to provide the most up-to-date expert information on all genes involved in germ cell growth and development across species. GermOnline currently covers 11 model organisms including yeasts (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Neurospora crassa*), plants (*Aribidopsis thaliana*, *Zea mays*), a nematode (*Caenorhabditis elegans*), an arthropod (*Drosophila melanogaster*), vertebrates (*Danio rerio*, *Xenopus laevis*), mammals (*Mus musculus*, *Rattus norvegicus*) and *Homo sapiens.* Approximately 190 000 locus pages and 475 000 locus names (including aliases) are available for queries. In over 700 cases of highly relevant genes, the locus reports contain curated information from one or several scientists in addition to links to useful internal and external information sources. *S.cerevisiae* is the first organism for which genes involved in meiosis, spore formation and post-germination vegetative growth are practically all represented in the database. Work on this species is therefore considered proof of principle since GermOnline provides microarray expression data for nearly all the 6000 known yeast genes and curated information for ~680 genes (many with multiple entries). About 30 genes are available as prototype contributions for other species at the time of writing this paper. Scientists will now be encouraged to participate in this global project and contribute their data and knowledge. Contact with researchers who work on the species represented in GermOnline will be established through a sustained information campaign that includes talks and poster presentations at scientific meetings, a 'call for contributions' advertisement in various scientific journals as well as solicitation by electronic and regular mail. Since the yeast community responded very positively to the project, we fully expect the same reaction from scientists who use higher eukaryotic model systems.

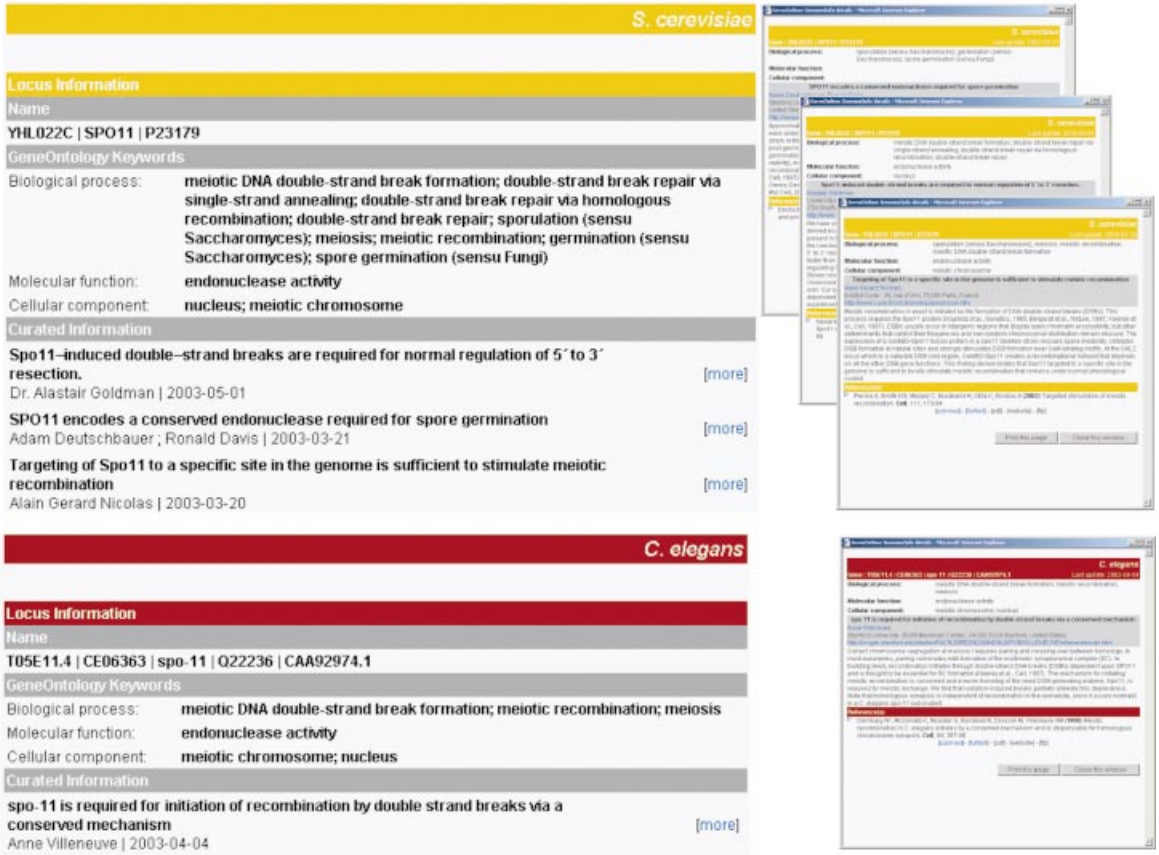
HOW TO RETRIEVE INFORMATION

Search options

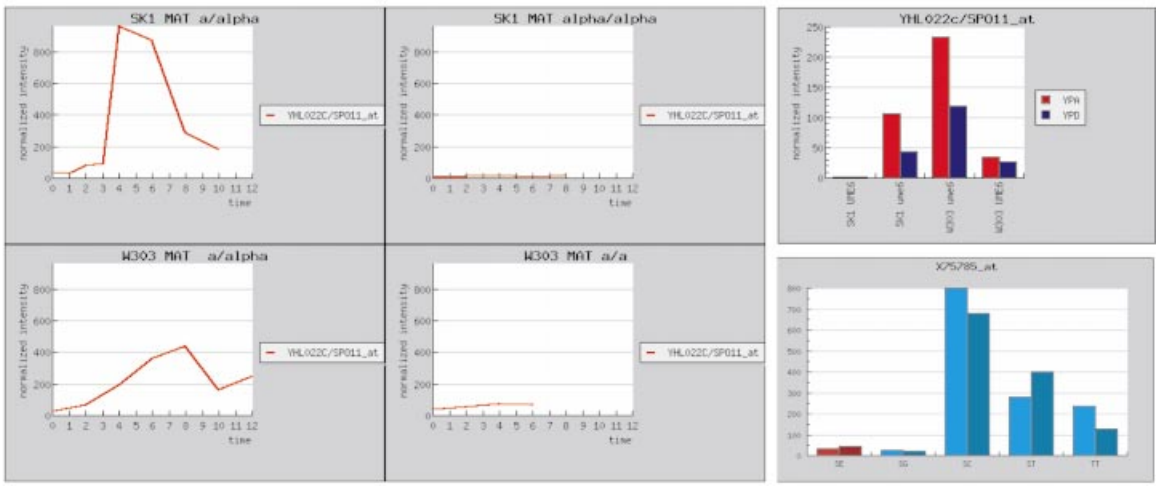
There are several entry points into GermOnline. The welcome page contains simple search options to access the locus page of a particular gene using either the systematic or genetic name, gene symbol or any alias provided by the relevant species-specific databases. Users can also search for authors who have

Figure 2. The locus page and expression data samples. (**A**) The section of the locus report page that summarizes the GO keywords and the title of curated contributions from different authors on the SPO11 genes from yeast and worm. Also shown is the gene-info box, which contains the author's contact information, the respective gene descriptions and the original reference relevant for the contribution. (**B**) Examples of the graphical display of various expression data sets on yeast sporulation (left), transcriptional deregulation in the absence of the Ume6 transcription factor (top right) and gene expression in purified rat Sertoli cells and testicular germ cells (bottom right). (**C**) An example of how data sources in the locus page are referenced: the title of the most recent publication as well as links to PubMed and (whenever available) relevant websites, search forms and specific deep links to locus pages (more) are given.

A


S. cerevisiae
Locus Information
Name
YHL022C | SPO11 | P23179
GeneOntology Keywords
Biological process: meiotic DNA double-strand break formation; double-strand break repair via single-strand annealing; double-strand break repair via homologous recombination; double-strand break repair; sporulation (sensu Saccharomyces); meiosis; meiotic recombination; germination (sensu Saccharomyces); spore germination (sensu Fungi)
Molecular function: endonuclease activity
Cellular component: nucleus; meiotic chromosome
Curated Information
Spo11–induced double–strand breaks are required for normal regulation of 5´ to 3´ resection.
Dr. Alastair Goldman | 2003-05-01
[more]
SPO11 encodes a conserved endonuclease required for spore germination
Adam Deutschbauer ; Ronald Davis | 2003-03-21
[more]
Targeting of Spo11 to a specific site in the genome is sufficient to stimulate meiotic recombination
Alain Gerard Nicolas | 2003-03-20
[more]
C. elegans
Locus Information
Name
T05E11.4 | CE06363 | spo-11 | Q22236 | CAA92974.1
GeneOntology Keywords
Biological process: meiotic DNA double-strand break formation; meiotic recombination; meiosis
Molecular function: endonuclease activity
Cellular component: meiotic chromosome; nucleus
Curated Information
spo-11 is required for initiation of recombination by double strand breaks via a conserved mechanism
Anne Villeneuve | 2003-04-04
[more]

B


SK1 MAT a/alpha
SK1 MAT alpha/alpha
YHL022c/SPO11_at
W303 MAT a/alpha
W303 MAT a/a
normalized intensity
time
YPA
YPD
X75785_at

C

Reference database
Saccharomyces Genome Database (SGD) provides biochemical and structural information for budding yeast proteins.
Pubmed | website | search | more
The Wellcome Trust Sanger Institute S. serevisiae GeneDB.
website | search | more
MIPS: a database for genomes and protein sequences.
Pubmed | search | more
Genome annotation
Reinvestigation of the Saccharomyces cerevisiae genome annotation by comparison to the genome of a related fungus: Ashbya gossypii.
Pubmed | website
A Re-annotation of the Saccharomyces cerevisiae Genome.
Pubmed | website

contributed information. Moreover, it is possible to choose one or more species and call up genes associated with particular GO keywords from the three categories, Biological Process, Molecular Function and Cellular Component. It should be emphasized that not all existing GO keywords can be selected to search the database but only those used during the curation process (this avoids a 'no entry' report); the available keywords are listed in alphabetical order in a pop-up menu (Fig. 1A). An advanced search option enables users to retrieve groups of genes according to their expression pattern or phenotype as determined in various high-throughput studies. Currently such query forms are available for *S.cerevisiae*, *S.pombe*, *C.elegans* and *R.norvegicus* (Fig. 1B). It is also possible to enter the section of a given species via the link in the navigation bar and directly access all information provided for the given organism.

The locus report

The locus report is separated into four sections, which contain links to internal and external data sources. A colour code marks out different species to facilitate navigation and to visually distinguish different sections of the database. The first section, *Locus Information*, displays the names, aliases and GO keywords chosen by one or more authors. Curated Information displays the title of a contribution followed by a link that leads to the complete description of a gene provided by the authors. Their names, contact information and their original references are shown in a separate window called the gene info box (Fig. 2A). If several contributions are available they are displayed in the locus page in chronological order based upon their date of publication in GermOnline with the most recent appearing at the top of the list. Furthermore, this section contains deep links to species-specific annotation databases (6), high-throughput loss-of-function studies (9–13) and, importantly, a direct connection to the locus report page of a potential homologue in another species represented in GermOnline. Note that deep links are convenient because they do not lead to the welcome or search page but directly to the locus report or data page of an external database.

Expression Information covers microarray expression data and is divided into links to relevant studies, consolidated expression databases and the US, European and Asian public array data repositories. The latter are displayed in the locus pages of all species. When following the links users can call up a graphical display of expression data externally to GermOnline, for e.g. *S.cerevisiae* undergoing the mitotic cell cycle and sporulation, sporulation in *S. pombe*, germline development in *C.elegans* and spermatogenesis in *R.norvegicus* [some examples of which are shown in Fig. 2B (14–19)]. External data sources are those that are either maintained by other laboratories or that contain expression data for poorly characterized loci that are not yet annotated and therefore not yet represented in GermOnline, such as numerous ESTs present on the U34 rat GeneChip (20).

Protein/Proteome Information is split into Protein/Proteome and Interactome sections, which include references with deep links to Swiss-Prot and data from biochemical and genetic high-throughput studies as well as deep links to GRID, a comprehensive database on protein–protein interaction data (21).

The *Toolbox* provides links to various web servers including some that enable researchers to predict the three-dimensional structure of proteins that are similar to those for which the crystal structure is known (22), make multiple sequence alignments (23) or cluster their microarray expression data (24).

All sources of information accessible through the locus report are referenced by the title of the most recent relevant publication as well as links to PubMed (or a journal's web page), and, whenever possible, websites, search forms and deep links to the respective locus report pages (Fig. 2C). When information is available for only a subset of the genes in a genome, links are displayed only in those cases. For example, the annotation studies in *S.cerevisiae* mark out several hundred ORFs as spurious, hypothetical or very hypothetical (25,26).

Cross-references to other databases

GermOnline is highly cross-referenced and provides constantly updated deep links to more than 50 published external data sources including all relevant annotation and microarray expression databases. Conversely, SGD (8), *S.cerevisiae* and *S.pombe* GeneDB (Database Cross-Reference) and Swiss-Prot (27) connect to GermOnline from within their locus report pages ('Literature and Functional Analysis' in the case of SGD and 'Cross-References' for Swiss-Prot and GeneDB, respectively). Furthermore, reciprocal deep links are provided that lead to the Swiss-Model Repository, which contains automatically predicted putative protein structures that are similar to those for which the crystal structure is known (22) (28) and Mammalian Reproductive Genetics (MRG), which provides annotation and microarray expression data for mouse genes involved in spermatogenesis (B. Braun and V. Cassen; unpublished).

THE SUBMISSION AND CURATION PROCESS

Registration and curation

Scientists who wish to contribute information are asked to register. This simple and brief process ensures efficient communication between authors, curators and the GermOnline staff. Contact information is displayed in the personal gene info box (see Fig. 2A). While an author is generally considered to be the Principal Investigator (PI), other members of the laboratory who provide the email address of the registered PI are also encouraged to make contributions. An automatic email notifies authors (and PIs) of the state of curation once information is submitted. The flow of events is depicted in Figure 3A. An incoming submission is allocated to a curator who accesses the form and either publishes the content or asks for revisions. The curation process may be aborted and the contribution rejected should a submission be without merit or a hoax. When revisions are requested, authors are notified by email and in a text box visible below the submission form. This process may be repeated until the final revised version is published. Authors can make only one contribution to a given gene at a time. Multiple entries from one scientist on one gene are not possible. However, database entries on several different genes from one author are highly encouraged.

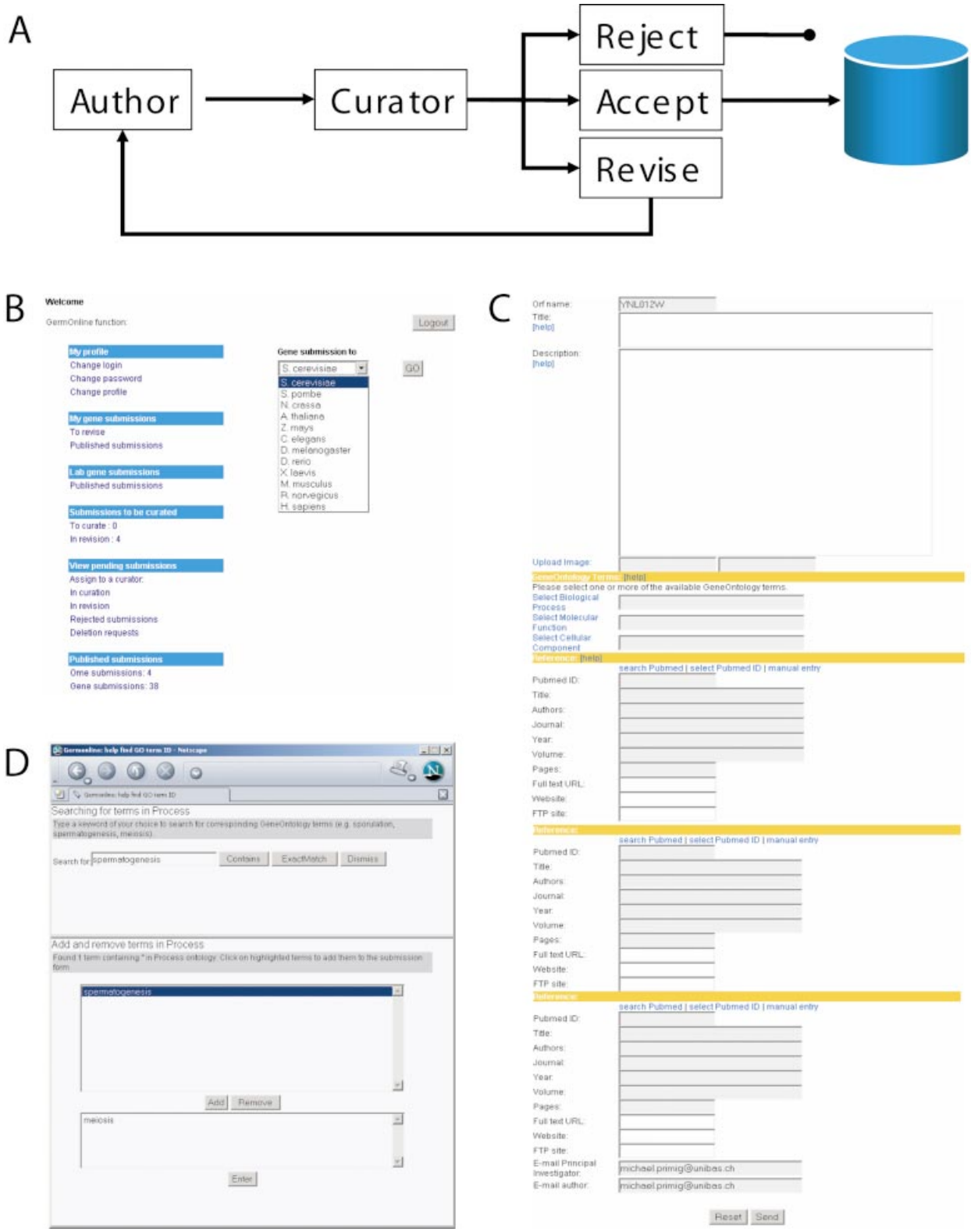


Figure 3. The submission and curation process. (**A**) Summary of the flow of events after an author has made a contribution. (**B**) A screen shot of the personal home page, which contains many functionalities (not all of which are available to all users) including a pop-up menu to choose the species. Once an organism and the gene/ORF of interest are selected, authors can call up the submission form shown in (**C**). To select GO keywords users follow the appropriate links in the submission form and then type a query in the search form displayed in (**D**). A list of hits is provided and the author can select and forward the keyword into the submission form by pressing the 'enter' button.

Making the initial contribution

Once an author has registered, immediate access to the submission form is given via the user's individualized home page after login (Fig. 3B). After a species and a gene/ORF are selected from the list of available organisms and loci, the submission form is displayed (Fig. 3C). It is recommended that the official locus ID or gene name/symbol as defined by the respective species databases be used as a particular alias may not always be available.

The submission form provides text fields for title and description. The title should be a concise summary of an author's findings on a gene function. The description briefly outlines the experimental approach and then summarizes facts concerning the biological process, the molecular function and the localization of the gene product in question. Authors are encouraged to include a brief comment on the next step(s) of their research. Important work from other groups is mentioned and cited in the text. To support the description, images and/or figures including titles and legends can also be uploaded. GO keywords can be selected by following the 'Select Process' link and typing a search term (e.g. cell cycle, meiosis, spermatogenesis, oogenesis) and options can be chosen to refine the search as indicated ('Contains' or 'Exact Match', Fig. 3D). Authors are able to select as many keywords as they deem appropriate. If a keyword is missing, it is recommended that it is submitted directly to GO where the updated keyword lists are retrieved regularly by GermOnline (7). Note that the names of keywords can change, they can become obsolete or even get deleted as the project evolves; any changes implemented by GO are valid for GermOnline as well. Finally, it is possible to provide up to three of the most recent PubMed references, which are automatically retrieved after entering the PubMed ID; publications that are not in PubMed can be entered manually (29).

Updating an existing database entry

Any update based upon the most recent publication replaces the previous database entry, thereby ensuring that the content does not become a historical overview of published work (like PubMed), or become outdated, but remains a concise, precise summary of the state-of-the-art views held by the author. Authors follow the link 'Published Submissions' in their home page and select one of their previous gene entries to update to directly access the current information in the submission form (Fig. 3B). After appropriate changes are made the submission is curated again. Updated information is not visible in the database during the curation process.

THE NEXT STEP

GermOnline is an ongoing project and therefore future directions and the development of functionalities will largely depend on community feedback. Improved services and options currently planned for releases higher than 2.0 include integration with the Ashbya Genome Database (AGD; S. Brachat, L. Hermida, R. Basavaraj, T. Schwede, P. Philippsen and M. Primig, manuscript in preparation), Flybase, Wormbase, MGD and Rat Genome Database (RGD). In addition, more data and links to potential homologues and experimentally verified orthologues will be integrated and tools for similarity searches and array data mining will be provided. More elaborate search options will include Boolean connection of the keywords from the GO categories. Finally, an online submission form for microarray data is envisaged.

CONCLUSION

GermOnline is a novel approach to biological information management and gene annotation. The idea is to meet the challenge of large-scale genome annotation by dividing the task into biological subjects and relevant species rather than having to cope with thousands of genes involved in different functions in all known species (30). This may help the global community of scientists who work on specific problems, often using only one or a few model organisms, to define groups of genes they can annotate in depth, collaborating with database developers and professional curators. Such expert information is all the more useful when placed within the context of relevant high-throughput data that provide insight into gene conservation and gene expression as well as protein localization and protein–protein interaction. This community-based approach suggested for data management in the field of germ cell growth and differentiation is applicable to all conserved biological processes studied in model organisms of various complexity and *H.sapiens*.

ACKNOWLEDGEMENTS

We thank A. Mizeracki for critical reading of the manuscript and B. Braun, B. Masdoua and V. Cassen for stimulating discussions and contributions to database programming. We acknowledge initial support of the project by N. Lamb and support of our IT infrastructure by Roger Jenni, F. Roesel and M. Jacquot. Funding for GermOnline Workshops was provided by the National Science Foundation (US), INSERM and CNRS. This project is funded by a grant from the Swiss Institute of Bioinformatics.

REFERENCES

1. Lockhart,D.J. and Winzeler,E.A. (2000) Genomics, gene expression and DNA arrays. *Nature*, **405**, 827–836.
2. Ito,T., Chiba,T., Ozawa,R., Yoshida,M., Hattori,M. and Sakaki,Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA*, **98**, 4569–4574.
3. Kamath,R.S., Fraser,A.G., Dong,Y., Poulin,G., Durbin,R., Gotta,M., Kanapin,A., Le Bot,N., Moreno,S., Sohrmann,M. *et al.* (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, **421**, 231–237.
4. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Veronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., Andre,B. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.
5. Uetz,P., Giot,L., Cagney,G., Mansfield,T.A., Judson,R.S., Knight,J.R., Lockshon,D., Narayan,V., Srinivasan,M., Pochart,P. *et al.* (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623–627.
6. Baxevanis,A.D. (2003) The Molecular Biology Database Collection: 2003 update. *Nucleic Acids Res.*, **31**, 1–12.
7. 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.
8. Weng,S., Dong,Q., Balakrishnan,R., Christie,K., Costanzo,M., Dolinski,K., Dwight,S.S., Engel,S., Fisk,D.G., Hong,E. *et al.* (2003)

Saccharomyces Genome Database (SGD) provides biochemical and structural information for budding yeast proteins. *Nucleic Acids Res.*, **31**, 216–218.
9. Colaiacovo,M.P., Stanfield,G.M., Reddy,K.C., Reinke,V., Kim,S.K. and Villeneuve,A.M. (2002) A targeted RNAi screen for genes involved in chromosome morphogenesis and nuclear organization in the *Caenorhabditis elegans* germline. *Genetics*, **162**, 113–128.
10. Briza,P., Bogengruber,E., Thur,A., Rutzler,M., Munsterkotter,M., Dawes,I.W. and Breitenbach,M. (2002) Systematic analysis of sporulation phenotypes in 624 non-lethal homozygous deletion strains of *Saccharomyces cerevisiae*. *Yeast*, **19**, 403–422.
11. Enyenihi,A.H. and Saunders,W.S. (2003) Large-scale functional genomic analysis of sporulation and meiosis in *Saccharomyces cerevisiae*. *Genetics*, **163**, 47–54.
12. Deutschbauer,A.M., Williams,R.M., Chu,A.M. and Davis,R.W. (2002) Parallel phenotypic analysis of sporulation and postgermination growth in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **99**, 15530–15535.
13. Rabitsch,K.P., Toth,A., Galova,M., Schleiffer,A., Schaffner,G., Aigner,E., Rupp,C., Penkner,A.M., Moreno-Borchart,A.C., Primig,M. *et al.* (2001) A screen for genes required for meiosis and spore formation based on whole-genome expression. *Curr. Biol.*, **11**, 1001–1009.
14. Mata,J., Lyne,R., Burns,G. and Bahler,J. (2002) The transcriptional program of meiosis and sporulation in fission yeast. *Nature Genet.*, **32**, 143–147.
15. Chu,S., DeRisi,J., Eisen,M., Mulholland,J., Botstein,D., Brown,P.O. and Herskowitz,I. (1998) The transcriptional program of sporulation in budding yeast. *Science*, **282**, 699–705.
16. Primig,M., Williams,R.M., Winzeler,E.A., Tevzadze,G.G., Conway,A.R., Hwang,S.Y., Davis,R.W. and Esposito,R.E. (2000) The core meiotic transcriptome in budding yeasts. *Nature Genet.*, **26**, 415–423.
17. Williams,R.M., Primig,M., Washburn,B.K., Winzeler,E.A., Bellis,M., Sarrauste de Menthiere,C., Davis,R.W. and Esposito,R.E. (2002) The Ume6 regulon coordinates metabolic and meiotic gene expression in yeast. *Proc. Natl Acad. Sci. USA*, **99**, 13431–13436.
18. Cho,R.J., Campbell,M.J., Winzeler,E.A., Steinmetz,L., Conway,A., Wodicka,L., Wolfsberg,T.G., Gabrielian,A.E., Landsman,D., Lockhart,D.J. *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, **2**, 65–73.
19. Reinke,V., Smith,H.E., Nance,J., Wang,J., Van Doren,C., Begley,R., Jones,S.J., Davis,E.B., Scherer,S., Ward,S. *et al.* (2000) A global profile of germline gene expression in *C. elegans*. *Mol. Cell*, **6**, 605–616.
20. Liu,G., Loraine,A.E., Shigeta,R., Cline,M., Cheng,J., Valmeekam,V., Sun,S., Kulp,D. and Siani-Rose,M.A. (2003) NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Res.*, **31**, 82–86.
21. Breitkreutz,B.J., Stark,C. and Tyers,M. (2003) The GRID: the General Repository for Interaction Datasets. *Genome Biol.*, **4**, R23.
22. Schwede,T., Kopp,J., Guex,N. and Peitsch,M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.*, **31**, 3381–3385.
23. Poirot,O., O'Toole,E. and Notredame,C. (2003) Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res.*, **31**, 3503–3506.
24. Getz,G. and Domany,E. (2003) Coupled two-way clustering server. *Bioinformatics*, **19**, 1153–1154.
25. Wood,V., Rutherford,K., Ivens,A., Rajandream,M. and Barrell,B. (2001) A re-annotation of the *Saccharomyces cerevisiae* genome. *Comp. Funct. Genet.*, **2**, 143–154.
26. 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.
27. 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.
28. Kopp,J. and Schwede,T. (2004) The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res.*, **32**, D230–D234.
29. Wheeler,D.L., Church,D.M., Federhen,S., Lash,A.E., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Sequeira,E., Tatusova,T.A. *et al.* (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res.*, **31**, 28–33.
30. Stein,L. (2001) Genome annotation: from sequence to biology. *Nature Rev. Genet.*, **2**, 493–503.