GermOnline: a cross-species database

• **Abstract:**

    – GermOnline is a web-accessible relational data-base that enables life scientists to make a significant and sustained contribution to the annotation of genes relevant for the fields of mitosis, meiosis, germ line development and gam-etogenesis across species.

    – This novel approach to genome annotation includes a platform for knowledge submission and curation as well as microarray data storage and visualization hosted by a global network of servers.

GermOnline: a cross-species database

- **what?** germ cell differentiation
- **information?** microarray expression data
  - for genes involved in mitoses and meiosis
  - gamete formation and germ line development
- **vocabulary?** used by GeneOntology consortium
- **controlled?** by a panel of international scientists (for standardisation of data and the content)
- **locus report pages?** links to external databases for relevant annotations
  - microarray expression data
  - proteome data

GernOnline: Where?

- http://www.germonline.org/
- Servers:
  - Europe:
    - http://germonline.unibas.ch/
    - http://germonline.igh.cnrs.fr/
  - US:
    - http://germonline.yeastgenome.org/
  - Japan:
    - http://germonline.biochem.s.u-tokyo.ac.jp/

GermOnline: what species?

- Saccharomyces cerevisiae SGD (yeast)
- Schizosaccharomyces pombe (yeast)
- Nourospora crassa (yeast)
- Aribidopsis thaliana (plant)
- Zea mays (plant)
- Caenorhabditis elegans (worm, nematode)
- Drosophila melanogaster (fruitfly arthropod)
- Danio rerio (vertebrates)
- Xenopus laevis (vertebrates)
- Mus musculus (mouse, mammal)
- Rattus norvegicus (rat , mammal)
- Homo sapiens (human , mammal)

GermOnline: why?

- Study the same genes and their patters in different species

- solve growing data problems

- solve nomenclatura problems (GO)

- use existent different specialized databases (fluitfly, mouse, worm, human)

GermOnline: what? more specific

- Mitotic growth and meiotic development

- info on gene function about sexual reproduction

- 190,000 locus pages, 475,000 locus names
  - there are on in 700 genes extendend links
  - microarray on all aprox. Known 6000 yeast genes

GermOnline database

- The database model is designed to fulfil three major functions:

  - first, organizes information about species and loci

  - second, the model covers information about the submission and curation process.

  - third, the model organizes data from microarray studies displayed as curves or bar diagrams and provides complex queries for genes that display a particular transcriptional pattern

1: organizes information about species and loci.

- Tables for species/loci are in red
- The *Species* table contains a numerical **species_id**
  - that identifies each organism as well as information about the genome, transcriptome and proteome, when available
- The *Orf* table contains the locus-specific numerical **orf_id**
  - associated with several external **locus identifiers** and **genetic names** as well as their **aliases**. These names are stored in the **orf_name** attribute field.
  - The **externalId** establishes a link between the unique idertifier of the external database and the **orf_id**. To connect the locus pages of homologues within GermOnline, their corresponding **orf_ids** are stored in the *Ortholog* table.

- *The Nomenclature* table is being used to identify the source of locus lists, For this a **nomenclature_id** field is used
  - This table holds information about the species-specific reference databases, NCBIs Locuslink and Swiss-Prot as well as the respective source uniform resource locators (URLs).

- The *Link*, *LinkAssign* and *Layout* tables comprise a collection of links and deep links to individual locus report pages that point to external databases and to relevant studies about meiosis.

2: info. about the submission and curation process..

- The tables for submission/curation are in blue
- GermOnline staff mediates the interaction between curators and authors.
- Authors submitting articles are registered in the tables:
  - *User*, *Lab*, *User_role*, *Title* and *Country* tables
  - Only one entry for each locus from a research group represented by the Principal Investigator is allowed.
- The content of a contribution is stored in tables called *Submission*, *Reference*, *Image*, *Term* and *Termassign*.

- A submission on a given locus consists of a title, a description, images, controlled vocabulary from the GeneOntology consortium and up to three of the most recent and relevant references (more are possible upon request by the authors).

- The table *Submitstate* describes the status of a contribution during curation (e.g. **submitted, in revision, rejected or published**).
- Request for revisions or other remarks are stored in the **Comments** table allowing authors and curators to communicate

2: info. about the submission and curation process..

- Invited submissions about the genome, transcriptome or proteome of a species are contributed by authors or GeO staff.

- Since these contributions are not curated like submissions on gene, a special authorization is necessary; this information is stored in the *Permission* table.

3: data from microarray studies displayed as curves or bar diagrams

- expression data is shown in colour green.

- The expression data is kept in the *Sample* and *ExpressionValue* tables.

- The lists of loci identified in various expression profiling experiments are stored in the *AnalysisValue table* and can be called up through advanced search forms.

GermOnline search: how?

- Access by: Mus musculus
- Search by genes: SPO11 -> locus

- Access by: Mus musculus
- Gene ontology keywords:
  - biological process:
    - **cell growth** -> 1 orf: MGI:103222 -> locus
  - molecular function:
    - DNA binding -> 1 orf: MGI:1890476 -> locus
  - cellular component:
    - chromosome -> 1 orf MGI:109542 -> locus

GermOnline search: how?

» **Locus information:**

– name: SPO11|P231

– GeneOntology Keywords

- Biological process:meiotic DNA double-strand break formation;...
- Molecular function: endonuclease activity
- Cellular component: nucleus; meiotic chromosome

– Curated info:

- Spo11–induced double–strand breaks are required for normal regulation of 5´ to 3´ resection. Dr. Alastair Goldman | 2003-05-01

– Reference database:

- **Saccharomyces Genome Database (SGD)**

– Gene annotation

– Gene insertion/deletion/RNAi projects

– Go to the putative homolog of

GermOnline search: how?

» **Locus information:**

– Gene annotation

- **Reinvestigation of the Saccharomyces cerevisiae genome annotation by comparison to the genome of a related fungus: Ashbya gossypii.**
- A Re-annotation of the Saccharomyces cerevisiae Genome.
- ……

– Gene insertion/deletion/RNAi projects

- SCMD: Saccharomyces cerevisiae Morphological Database.
- Large-Scale Functional Genomic Analysis of Sporulation and Meiosis in Saccharomyces cerevisiae.
- …...

– Go to the putative homolog of

- The Ashbya gossypii genome as a tool for mapping the ancient Saccharomyces cerevisiae genome.
- The genome sequence of Schizosaccharomyces pombe.

- ….

GermOnline search by genes: clgn

» **Expression information:**

- Major molecular differences between mammalian sexes are involved in drug metabolism and renal function

- Essential role of Plz in maintenance of spermatogonial stem cells

- Analysis of Crem-dependent gene expression during mouse spermatogenesis

- Identification of testosterone regulated genes in testes of hypogonadal mice using oligonucleotide microarray

- A gene atlas of the mouse and human protein-encoding transcriptomes.

- The Murine Testicular Transcriptome: Characterizing Gene Expression in the Testis During the Progression of Spermatogenesis.

- A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets.

GermOnline search by genes: clgn

» **Protein/Proteome information:**

- **Proteome:**
  – The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.

  – The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models.

- **Interactome:**

GermOnline search by genes: clgn

» **Toolbox:**

- Multipe sequence alignment, protein modeling & cluster analysis

ExPASy: the proteomics server for in-depth protein knowledge and analysis.
Pubmed | website

SWISS-MODEL: an automated protein homology-modeling server.
Pubmed | website

Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments.
Pubmed | website

Swiss EMBnet node web server.
Pubmed | website Europe | website Switzerland

Signal search analysis server.
Pubmed | website

Coupled two-way clustering server.
Pubmed | website Weizman | website Biozentrum

The NAR web server issue 2004.
website

GermOnline search by: author

- **List of authors:**

Amon Angelika	Bestor Timothy H	Borde Valerie
bourc'his deborah	Brachat Sophie	Breitenbach Michael
Briza Peter	Caburet Sandrine	Chung Sanny S. W.
Davis Ronald	Deutschbauer Adam	Dickinson Hugh Gordon
Dumitrescu Teodora	Esposito Rochelle Easton	Fellous Marc
Goldman Alastair	Grootegoed J. Anton	Gruissem Wilhelm
Hawley Scott	Kaufmann Randal	Klein Franz
Maro Bernard	Mercier Raphaël	Mitchell Aaron
Nasmyth Kim	Nicolas Alain Gerard	Okabe Masaru
OrrWeaver Terry L	Pandolfi Pier Paolo	Pawlowski WojtekP
Philippsen Peter	Pineau Charles Guenole	Poss Kenneth D
Primig Michael	Rabitsch Kirsten Peter	Reinke Valerie
Sasaki Hiroyuki	Saunders William	Schedl Tim B
Scherthan Harry	Schlecht Ulrich	Schroeder Martin
Siep Michel	Villeneuve Anne	Wolgemuth Debra Joanne
Yamamoto Masayuki		

GermOnline search by: Prof. Angelika Amon

- S. cerevisiae

    – YHR014W | SPO13 | P23624   -> locus

    – YMR001C | CDC5 | P32562 | MSD2 | PKX2   -> locus

GermOnline a cross-species database

- **CONCLUSION:**

    – We describe the **first functional web-based platform for community annotation** and **microarray data visualization** designed to organize data by biological subjects and **across species**.

    – The project is based upon **open source** software and all scientists and computer programmers working in the field of database development are **invited to contribute** their ideas.