Published online 16 November 2006 *Nucleic Acids Research, 2007, Vol. 35, Database issue* ***D439–D444***
doi:10.1093/nar/gkl777

ParameciumDB: a community resource that integrates the *Paramecium tetraurelia* genome sequence with genetic data

Olivier Arnaiz, Scott Cain[1], Jean Cohen and Linda Sperling*

Centre de Génétique Moléculaire, CNRS, 91198 Gif-sur-Yvette, France and [1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

Received August 10, 2006; Revised September 27, 2006; Accepted September 29, 2006

ABSTRACT

ParameciumDB (http://paramecium.cgm.cnrs-gif.fr) is a new model organism database associated with the genome sequencing project of the unicellular eukaryote *Paramecium tetraurelia*. Built with the core components of the Generic Model Organism Database (GMOD) project, ParameciumDB currently contains the genome sequence and annotations, linked to available genetic data including the Gif Paramecium stock collection. It is thus possible to navigate between sequences and stocks via the genes and alleles. Phenotypes, of mutant strains and of knockdowns obtained by RNA interference, are captured using controlled vocabularies according to the Entity-Attribute-Value model. ParameciumDB currently supports browsing of phenotypes, alleles and stocks as well as querying of sequence features (genes, UniProt matches, InterPro domains, Gene Ontology terms) and of genetic data (phenotypes, stocks, RNA interference experiments). Forms allow submission of RNA interference data and some bioinformatics services are available. Future ParameciumDB development plans include coordination of human curation of the near 40 000 gene models by members of the research community.

A MODEL ORGANISM FOR THE 21ST CENTURY

Paramecium is a unicellular eukaryote that belongs to the ciliate phylum. Ciliates are the only unicellular organisms that separate germinal and somatic functions. Diploid but silent micronuclei undergo meiosis and transmit the genetic information to the next sexual generation. Highly polyploid macronuclei express the genetic information but develop anew at each sexual generation, through extensive programmed rearrangements of the genome.

Paramecium has long served as a model for the study of complex functions characteristic of multi-cellular organisms, including the somatic differentiation process (1), the biogenesis of cortical structures such as the ciliary basal bodies (2), regulated secretion (3), and receptor- and ion channel-mediated cell signaling in response to environmental stimuli (4,5). Standardized methods for Paramecium genetics have been available for half a century (6) and genes can be cloned by functional complementation (7). Over the past decade, gene silencing by somatic transformation (8) and RNA interference (RNAi) by feeding with bacteria that produce double-stranded RNA (9) have become routine laboratory procedures.

Paramecium is a privileged model for investigation of non-Mendelian heredity and the underlying epigenetic mechanisms. Sonneborn (10) was the first to document cytoplasmic heredity of mating type and other traits in Paramecium. It is just now becoming clear that these examples of cytoplasmic heredity can be explained by homology-dependent mechanisms that involve non-coding RNA. The mechanisms, related to RNA interference and largely conserved among eukaryotes, enable comparison of the maternal somatic genome with the zygotic genome by base pairing at the time of sexual processes. This comparison allows the maternal rearrangement pattern to be transmitted to the new somatic nucleus (11). A second phenomenon, the cytoplasmic heredity of cortical pattern (12), is related to prion heredity. Cortical heredity also turns out to involve mechanisms that probably operate in all cells, to relay memory of cell shape through template-guided assembly of new structures in a pre-existing cellular space (13).

The *Paramecium tetraurelia* macronuclear genome was sequenced by a whole genome shotgun approach at the Genoscope French National Sequencing Center, allowing discovery of nearly 40 000 protein-coding genes in the 72 Mb assembly. This unusually large number of genes, especially for a unicellular organism, is the result of at least three successive whole genome duplications in the Paramecium lineage (14). The exceptional conservation of synteny between duplicated chromosomes, the high level of retention

*To whom correspondence should be addressed. Tel: +33169823209; Fax: +33169823150; Email: sperling@cgm.cnrs-gif.fr of genes and the large number of paralogs that could consequently be identified, make Paramecium an outstanding model for studying the evolutionary consequences of whole genome duplication.

PARAMECIUMDB ARCHITECTURE

ParameciumDB was built using components of the Generic Model Organism Database (GMOD, http://www.gmod.org) toolkit. GMOD is an open source project initiated in 2002 with the objective of providing generic software so that a small community with limited informatics infrastructure can build a new model organism database. We provide a brief description of the GMOD core components used to build ParameciumDB.

Chado is a modular relational database schema developed at FlyBase. ParameciumDB implements the Chado schema using the PostgreSQL open source relational database management system. The Chado schema includes a genetic module, and we have added a stock sub-module to the genetic module. Chado implements data classes using controlled, structured vocabularies known as ontologies. Ontologies capture knowledge in a way that can be understood by humans and processed by machines, and represent an important step toward bioinformatics data integration. The GMOD standard implementation of Chado integrates the widely used Sequence (15), Gene (16) and Relationship (17) Ontologies. We have begun to develop Paramecium Anatomy Ontology, largely orthogonal to Gene Ontology (GO), for use in modeling phenotypes.

Turnkey is a generic framework that autogenerates a web interface from a database schema (http://turnkey.sourceforge.net). Turnkey relies on the Perl module SQL::Translator to do this and the only input is an SQL file describing the database tables and relationships. Turnkey builds the interface code into an Apache/mod_perl web server. For each data table in the database, Turnkey generates a template for a web page that contains all the values in the table and shows all the relationships to other tables in the database. Customization of the web pages is achieved by modification of the templates.

The Generic Genome Browser (GBrowse) is a mature, widely used standard for viewing genome annotations via the web (18). GBrowse is an integral part of the web interface generated by Turnkey. Software that allows the GBrowse CGI script to communicate with a Chado database is provided by the GMOD project.

PARAMECIUMDB CONTENTS

ParameciumDB contains genome sequence data and annotations from the *Paramecium* genome sequencing project (EMBL/GenBank/DDBJ accession nos CT867985–CT868681) and genetic data including the Gif Paramecium stock collection and RNA interference experiments (Table 1).

Table 1. Data in ParameciumDB (August 2006)

39 642	Gene models
85 212	UniProt protein matches (19 035 gene models have at least 1 match)
45 072	InterPro domains (20 767 predicted proteins have at least 1 domain)
4 978	Best Reciprocal Hits to *Tetrahymena thermophila* gene predictions
982	Stocks
185	Mutant alleles (35 alleles are linked to sequences)
29	Genotypes characterized by non-Mendelian heredity
59	RNAi experiments
57	Phenotypes

We have begun to model phenotypes in ParameciumDB using a schema proposed for the description of mouse phenotypes (19), involving five classes of ontology: organism, entity, attribute, value and assay. For example, the *Paramecium* mutant sm19-1 (20) has a phenotype: entity *cell* with attribute *size* returning value *small* by a *visual inspection* assay. The entity in this case can be found in GO. The attributes and values we have used to describe *Paramecium* phenotypes come from the Phenotype Attribute Ontology (PATO, http://obo.sourceforge.net/). We have begun to develop a Paramecium Anatomy Ontology, since we need more granular 'cellular component' and 'biological process' terms than presently available in GO to describe some species- or phylum-specific traits (nuclear dimorphism; cytological features such as a cell cortex that is organized as several thousand repeating unit territories around the ciliary basal bodies; regulated secretion of very elaborate defensive organelles; swimming behavior in response to external stimuli mediated by ion channels and receptors, etc.). We have also begun to develop a Paramecium Assay Ontology that could perhaps be integrated into a more general ontology of assays (20).

USING PARAMECIUMDB

Overview

Every page of ParameciumDB contains a top row of navigation tabs (Home, Search, Gbrowse, Blast, Tools, Help) and a sidebar. The sidebar on the home page (and some information pages) contains internal and external links for community news, downloads and information about specific topics such as the genome sequencing project, the stock collection, *Paramecium* mitochondrial and ribosomal DNA. The Help page provides some explanation of how data is organized in ParameciumDB and strategies for finding data using the Search page. The Search page contains help buttons that bring up windows with additional tips and/or examples.

The Search page features three boxes. The first two allow the user to query Sequence and Genetic data (an example, showing integration of sequence and genetic data, is provided in Gene page and Allele page sections). The third box makes it possible to browse some of the tables in the database.

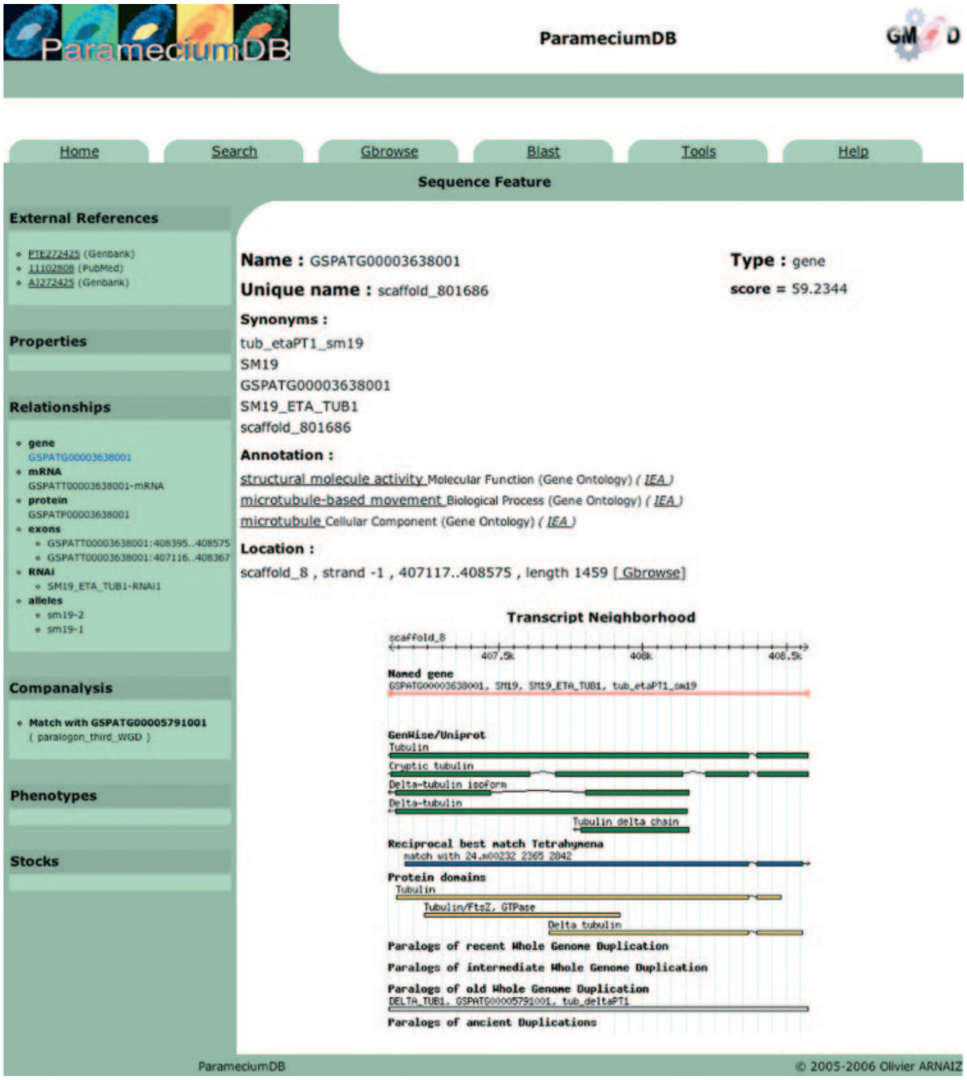
Browsing tables using the third box on the Search page is particularly useful for genetic data. For example, each row in the Phenotype table is linked to a page for the Phenotype, showing how it has been modeled with ontology terms and providing links to all related mutant alleles and RNAi reagents. Stock and allele tables can also be browsed, although a more direct approach would be to use the genetic query box to search for a stock, or to use the sequence query box to search for an allele.

Database searches are achieved by querying the appropriate data category. Thus queries of different sequence features (Named genes, UniProt match descriptions, InterPro domains, GO terms) or genetic data (Phenotypes, Stocks, RNAi experiments) involve selecting the data category from the pull-down menu of the search box and filling in a search term. All searches are case insensitive and are surrounded by wild cards. A query will return a database page if there is only one result, but more often it will return a table with multiple results. For example, for a sequence feature search, the table will contain the name of each sequence feature linked to the corresponding database page as well as its location on the genome sequence.

Gene page

Suppose that we want to find information about the gene SM19 encoding eta-tubulin (21). We can type 'sm19' as Gene name in the Sequence box and launch a search. The search will return a table with three results, the wild-type SM19 gene and two alleles, sm19-1 and sm19-2. The gene page obtained by clicking on SM19 is shown in Figure 1. The gene page contains three regions, a central panel, a GBrowse image and the left sidebar with links to all related data, presented by category.

The central panel shows that we are looking at a Sequence Feature of Type 'gene'. The score in this case is that provided by Genoscope's automated annotation (14). For sequence features of type 'match', i.e. for UniProt matches, it is the match score. All names and synonyms are provided, and these are all targets of Gene name searches. The 'Annotation' section provides GO terms and their evidence codes. Clicking on the GO term brings up a page with all the genes in ParameciumDB associated with the term and an inset frame with the AmiGO browser page for that term (http://www.godatabase.org/). ParameciumDB GO terms (August 2006) are electronically inferred from the InterPro matches. This particular case is a good example of the limits of electronically inferred annotation. Eta-tubulin, a recently discovered member of the tubulin superfamily, also has a new function. Eta-tubulin is necessary for the process of basal body duplication and is not involved in the process of microtubule-based movement,



External References

- PTE272425 (Genbank)
- 11102808 (PubMed)
- AJ272425 (Genbank)

Properties

Relationships

- gene
  GSPATG00003638001
- mRNA
  GSPATT00003638001-mRNA
- protein
  GSPATP00003638001
- exons
  - GSPATT00003638001:408395..408575
  - GSPATT00003638001:407116..408367
- RNAi
  - SM19_ETA_TUB1-RNAI1
- alleles
  - sm19-2
  - sm19-1

Companalysis

- Match with GSPATG00005791001
  (paralogon_third_WGD)

Phenotypes

Stocks

Name : GSPATG00003638001

Unique name : scaffold_801686

Type : gene

score = 59.2344

Synonyms :

tub_etaPT1_sm19
SM19
GSPATG00003638001
SM19_ETA_TUB1
scaffold_801686

Annotation :

structural molecule activity Molecular Function (Gene Ontology) (IEA)
microtubule-based movement Biological Process (Gene Ontology) (IEA)
microtubule Cellular Component (Gene Ontology) (IEA)

Location :

scaffold_8 , strand -1 , 407117..408575 , length 1459 [Gbrowse]

Transcript Neighborhood

scaffold_8
407.5k
408k
408.5k
Named gene
GSPATG00003638001, SM19, SM19_ETA_TUB1, tub_etaPT1_sm19
GenWise/Uniprot
Tubulin
Cryptic tubulin
Delta-tubulin isoform
Delta-tubulin
Tubulin delta chain
Reciprocal best match Tetrahymena
match with 24.m00232 2365 2842
Protein domains
Tubulin
Tubulin/FtsZ, GTPase
Delta tubulin
Paralogs of recent Whole Genome Duplication
Paralogs of intermediate Whole Genome Duplication
Paralogs of old Whole Genome Duplication
DELTA_TUB1, GSPATG00005791001, tub_deltaPT1
Paralogs of ancient Duplications

ParameciumDB © 2005-2006 Olivier ARNAIZ

Figure 1. A ParameciumDB gene page. This is the gene page for the gene SM19.

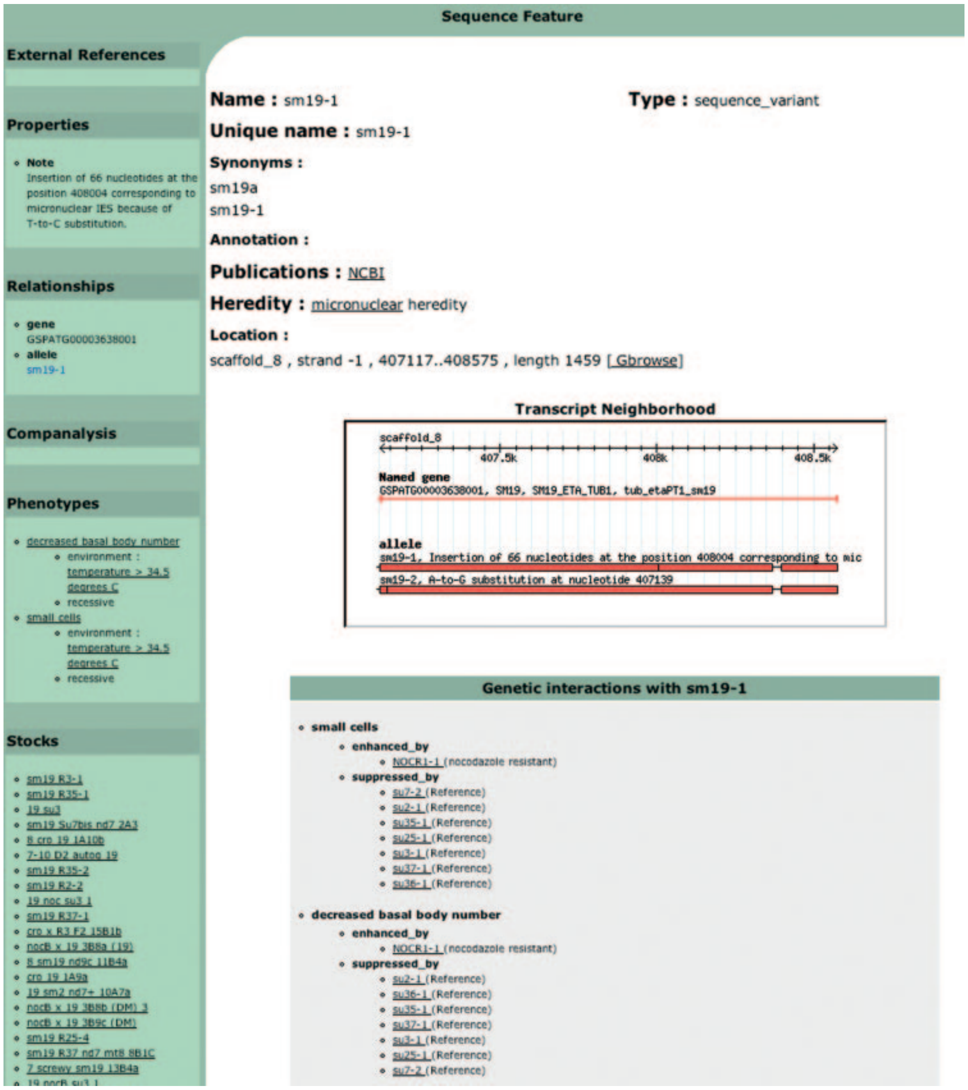
as electronically inferred from InterPro domains that are signatures of the tubulin superfamily.

The inset GBrowse frame shows some of the annotation information and results of some computational analyses. Each of these, when clicked, brings up the corresponding sequence feature page. Match scores appear when the mouse hovers over the feature. The match page for Tetrahymena best reciprocal matches contains a link (at the top of the left sidebar) to the corresponding gene page in the Tetrahymena Genome Database (22). *Tetrahymena thermophila*, another ciliate, is the closest relative to *Paramecium* with a sequenced genome (23); however, the two organisms are separated by ~500 MY and orthologous proteins share only ~40% amino acid identity (14).

The sidebar shows all related data in ParameciumDB, including external references to GenBank and PubMed as well as relationships to other sequence features including alleles and RNAi reagents. There is a computational analysis result for this gene, a paralog related by one of the series of whole genome duplications in the Paramecium lineage (14): eta-tubulin arose through duplication of an ancestral delta-tubulin gene.

Allele page

The allele page for the sm19-1 allele is reached by clicking on that allele's name in the sidebar. An allele page (Figure 2) is similar to a gene page. The sidebar has the same layout; however, there is now data regarding phenotypes and stocks. The central panel shows that we are looking at a Sequence Feature of type 'sequence_variant', the Sequence Ontology term for allele. Three new categories now feature in the central panel. 'Publications' provides a link to NCBI with a PubMed query that brings up all of the relevant literature. 'Heredity' is used to distinguish Mendelian (micronuclear) heredity from non-Mendelian (macronuclear) heredity. A box shows 'genetic interactions' and in this case both phenotypes observed for this allele at the non-permissive temperature are enhanced or suppressed by other alleles. Note that each phenotype is observed under



Sequence Feature

External References

Properties

- Note
  Insertion of 66 nucleotides at the position 408004 corresponding to micronuclear IES because of T-to-C substitution.

Relationships

- gene
  GSPATG00003638001
- allele
  sm19-1

Companalysis

Phenotypes

- decreased basal body number
    - environment : temperature > 34.5 degrees C
    - recessive
- small cells
    - environment : temperature > 34.5 degrees C
    - recessive

Stocks

- sm19 R3-1
- sm19 R35-1
- 19 su3
- sm19 Su7bis nd7 2A3
- 8 cro 19 1A10b
- 7-10 D2 autog 19
- sm19 R35-2
- sm19 R2-2
- 19 noc su3 1
- sm19 R37-1
- cro x R3 F2 15B1b
- nocB x 19 3B8a (19)
- 8 sm19 nd9c 11B4a
- cro 19 1A9a
- 19 sm2 nd7+ 10A7a
- nocB x 19 3B8b (DM) 3
- nocB x 19 3B9c (DM)
- sm19 R25-4
- sm19 R37 nd7 mt8 8B1C
- 7 screwy sm19 13B4a
- 19 nocB su3 1

Name : sm19-1 Type : sequence_variant

Unique name : sm19-1

Synonyms :

sm19a

sm19-1

Annotation :

Publications : NCBI

Heredity : micronuclear heredity

Location :

scaffold_8 , strand -1 , 407117..408575 , length 1459 [Gbrowse]

Transcript Neighborhood

scaffold_8

407.5k 408k 408.5k

Named gene

GSPATG00003638001, SM19, SM19_ETA_TUB1, tub_etaPT1_sm19

allele

sm19-1, Insertion of 66 nucleotides at the position 408004 corresponding to mic

sm19-2, A-to-G substitution at nucleotide 407139

Genetic interactions with sm19-1

- small cells
    - enhanced_by
        - NOCR1-1 (nocodazole resistant)
    - suppressed_by
        - su7-2 (Reference)
        - su2-1 (Reference)
        - su35-1 (Reference)
        - su25-1 (Reference)
        - su3-1 (Reference)
        - su37-1 (Reference)
        - su36-1 (Reference)

- decreased basal body number
    - enhanced_by
        - NOCR1-1 (nocodazole resistant)
    - suppressed_by
        - su2-1 (Reference)
        - su36-1 (Reference)
        - su35-1 (Reference)
        - su37-1 (Reference)
        - su3-1 (Reference)
        - su25-1 (Reference)
        - su7-2 (Reference)

Figure 2. A ParameciumDB allele page. This is the allele page for the allele sm19-1.

a given environment, as shown in the Phenotypes section of the sidebar. Finally, the inset GBrowse frame shows the nature of the mutation for all alleles of the gene under consideration.

In this example, we can navigate back and forth from phenotypes to stocks to alleles to sequences; thanks to the fact that the SM19 gene has been cloned by functional complementation (21). However many of the *Paramecium* genes that were identified by a genetic approach have not yet been cloned. The corresponding allele pages are not linked to sequences, and do not contain a GBrowse image.

Finding genes by sequence homology

The Blast navigation tab brings up a Blast server (currently, our implementation of NCBI's wwwblast server) allowing the user to paste in a nucleotide or protein sequence and run a Blast search against the proteins predicted by the automated annotation of the genome sequence (14). The search results are linked to the corresponding ParameciumDB gene pages.

Other features

RNA interference data can be submitted to ParameciumDB using a form that can be found from the Tool page. The Tool page also provides access to bioinformatics services, e.g. a Smith Waterman alignment of two nucleotide sequences provided by the user. The alignment is presented along with a histogram of the lengths of stretches of identical nucleotides and shows their positions on the alignment. This tool is useful in designing RNAi reagents and interpreting RNAi experiments.

PERSPECTIVES

In the absence of a dedicated curator, many simple improvements to ParameciumDB are still low on our priority list, such as images for gene and allele pages to illustrate, respectively, the localization of gene products or the phenotypes of mutant cells. We made the decision not to attempt to curate the *Paramecium* literature, at least for the moment, though many cross-references are provided to other databases, and each allele page has a well-crafted query that successfully retrieves the relevant literature from PubMed. We prefer to concentrate our efforts on exploiting the full power of the Chado schema design by implementing the module that can handle MIAME-compliant transcriptome data and by interfacing other tools with our Chado database, such as the SynView synteny viewer (24) and a complex query interface using BioMart (25).

Our immediate plans for ParameciumDB development are outlined below.

Re-annotation project

The value of the *Paramecium* genome sequence for research and teaching is critically dependent on the quality of the genome annotations. Although the currently available automated annotations are of very good quality, thanks to the combined use of many resources including a large cDNA collection (14), we are already aware of many gene models that will require human curation to resolve contradictory evidence. We have therefore begun adapting the Apollo Genome Editor (26) to read and write directly to ParameciumDB. Our objective is to train interested members of the Paramecium research community so that they can use Apollo, from anywhere in the world, to edit gene models and save their edits back to an instance of ParameciumDB. We will use the open source BioPipe workflow management software (27) to build computational pipelines to ensure that our community curators are working with up-to-date annotation evidence.

Improved phenotype descriptions

The use of ontologies for phenotype description is a new and fast-moving field. ParameciumDB's use of ontologies requires further development for at least three reasons. First, we are not using the full power of the Relationship Ontology, which is critical for computations based on the phenotypes. Second, Paramecium Anatomy Ontology requires development in order to meet the Open Biomedical Ontologies (http://obo.sourceforge.net/) standards, e.g. most terms do not have definitions. Moreover, it may turn out to be preferable in the long run to integrate this ontology with the GO Cellular Component and Biological Process ontologies. Finally, inclusion in ParameciumDB of the exponentially growing corpus of RNAi data is forcing us to devise a way for database users to propose new phenotypes. The PheNote tool, currently under development (Mark Gibson, personal communication), may provide the means toward this end.

ACKNOWLEDGEMENTS

We wish to thank the GMOD project, FlyBase and other associated MODs for making ParameciumDB possible, and Lincoln Stein in particular for his support and encouragement. We gratefully acknowledge Brian O'Connor, Allen Day and Mark Gibson for designing software features and/or patches specifically for ParameciumDB. We are indebted to Genoscope in more ways than can be enumerated here, and in particular to Jean Weissenbach, Patrick Wincker and Vincent Schächter for advice and support and Jean-Marc Aury, Olivier Jaillon and Benjamin Noël for access to data before release of the genome sequence. We wish to thank all of our colleagues in the Paramecium GDRE consortium for making requests and providing criticism that have shaped the user interface. The Paramecium stock collection and genetic data were curated with the help of Janine Beisson and Françoise Ruiz (CGM, Gif-sur-Yvette). We acknowledge financial support from the CNRS, the ACI IMPBio2004 Contract 14 to L.S. and the ANR Contract NT05-2_41522 to J.C. Funding to pay the Open Access publication charges for this article was provided by the 'Action Concertée Incitative Informatique, Mathématique et Physique en Biologie Moléculaire (IMPBio)'.

Conflict of interest statement. None declared.

REFERENCES

1. Bétermier,M. (2004) Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res. Microbiol.*, **155**, 399–408.

2. Beisson,J. and Wright,M. (2003) Basal body/centriole assembly and continuity. *Curr. Opin. Cell Biol.*, **15**, 96–104.
3. Vayssié,L., Skouri,F., Sperling,L. and Cohen,J. (2000) Molecular genetics of regulated secretion in *Paramecium. Biochimie*, **82**, 269–288.
4. Van Houten,J.L., Yang,W.Q. and Bergeron,A. (2000) Chemosensory signal transduction in *Paramecium. J. Nutr.*, **130**, 946S–949S.
5. Saimi,Y. and Kung,C. (2002) Calmodulin as an ion channel subunit. *Annu. Rev. Physiol.*, **64**, 289–311.
6. Sonneborn,T.M. (1950) Methods in the general biology and genetics of *Paramecium aurelia*. *J. Exp. Zool.*, **113**, 87–148.
7. Keller,A.M. and Cohen,J. (2000) An indexed genomic library for *Paramecium* complementation cloning. *J. Eukaryot. Microbiol.*, **47**, 1–6.
8. Ruiz,F., Vayssie,L., Klotz,C., Sperling,L. and Madeddu,L. (1998) Homology-dependent gene silencing in *Paramecium. Mol. Biol. Cell*, **9**, 931–943.
9. Galvani,A. and Sperling,L. (2002) RNA interference by feeding in *Paramecium*. *Trends Genet.*, **18**, 11–12.
10. Sonneborn,T.M. (1937) Sex, sex inheritance and sex determination in *Paramecium aurelia*. *Proc. Natl Acad. Sci. USA*, **23**, 378–385.
11. Meyer,E. and Chalker,D. (2007) Epigenetics of ciliates. In Allis,D., Jenuwein,T. and Reinberg,D. (eds), *Epigenetics*. Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 127–150.
12. Beisson,J. and Sonneborn,T.M. (1965) Cytoplasmic inheritance of the organization of the cell cortex in *Paramecium aurelia*. *Proc. Natl Acad. Sci. USA*, **53**, 275–282.
13. Beisson,J. (2006) Preformed cell structure and cell heredity. In Chernoff,Y. (ed.), *Protein-Based Inheritance*. Landes Bioscience, Georgetown, TX, in press.
14. Aury,J., Jaillon,O., Duret,L., Noël,B., Jubin,C., Porcel,B., Ségurens,B., Daubin,V., Anthouard,V., Aiach,N. *et al.* (2006) Global trends of whole genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, , **444**, 171–178.
15. 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.
16. The Gene Ontology Consortium (2001) Creating the gene ontology resource: design and implementation. *Genome Res.*, **11**, 1425–1433.
17. Smith,B., Ceusters,W., Klagges,B., Kohler,J., Kumar,A., Lomax,J., Mungall,C., Neuhaus,F., Rector,A.L. and Rosse,C. (2005) Relations in biomedical ontologies. *Genome Biol.*, **6**, R46.
18. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W., Arva,A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
19. Gkoutos,G.V., Green,E.C., Mallon,A.M., Hancock,J.M. and Davidson,D. (2005) Using ontologies to describe mouse phenotypes. *Genome Biol.*, **6**, R8.
20. Ruiz,F., Garreau de Loubresse,N. and Beisson,J. (1987) A mutation affecting basal body duplication and cell shape in *Paramecium. J. Cell. Biol.*, **104**, 417–430.
21. Ruiz,F., Krzywicka,A., Klotz,C., Keller,A., Cohen,J., Koll,F., Balavoine,G. and Beisson,J. (2000) The SM19 gene, required for duplication of basal bodies in *Paramecium*, encodes a novel tubulin, eta-tubulin. *Curr. Biol.*, **10**, 1451–1454.
22. Stover,N.A., Krieger,C.J., Binkley,G., Dong,Q., Fisk,D.G., Nash,R., Sethuraman,A., Weng,S. and Cherry,J.M. (2006) Tetrahymena Genome Database (TGD): a new genomic resource for *Tetrahymena thermophila* research. *Nucleic Acids Res.*, **34**, D500–D503.
23. Eisen,J.A., Coyne,R.S., Wu,M., Wu,D., Thiagarajan,M., Wortman,J.R., Badger,J.H., Ren,Q., Amedeo,P., Jones,K.M. *et al.* (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.,* **4**, e286. 10.1371/journal.pbio.0040286.
24. Wang,H., Su,Y., Mackey,A.J., Kraemer,E.T. and Kissinger,J.C. (2006) SynView: a GBrowse-compatible approach to visualizing comparative genome data. *Bioinformatics*, **22**, 2308–2309.
25. Kasprzyk,A., Keefe,D., Smedley,D., London,D., Spooner,W., Melsopp,C., Hammond,M., Rocca-Serra,P., Cox,T. and Birney,E. (2004) EnsMart: a generic system for fast and flexible access to biological data. *Genome Res.*, **14**, 160–169.
26. Lewis,S.E., Searle,S.M., Harris,N., Gibson,M., Lyer,V., Richter,J., Wiel,C., Bayraktaroglir,L., Birney,E., Crosby,M.A. *et al.* (2002) Apollo: a sequence annotation editor. *Genome Biol.*, **3**, RESEARCH0082.
27. Hoon,S., Ratnapu,K.K., Chia,J.M., Kumarasamy,B., Juguang,X., Clamp,M., Stabenau,A., Potter,S., Clarke,L. and Stupka,E. (2003) Biopipe: a flexible framework for protocol-based bioinformatics analysis. *Genome Res.*, **13**, 1904–1915.