Protist, Vol. ▮, ▮▮▮—▮▮▮, ▮▮ ▮▮▮▮
http://www.elsevier.de/protis
Published online date 12 April 2006

Protist

PROTIST TOOLS

Protists on the Web — Redoux Originally Protist 150,109—112 (1999)

Pardon my French, but I decided to update the "Protists on the Web" article I wrote for Protist News in the last century since many of the links have succumbed to the increasing entropy of the web. I again stress that my list is illustrative and critical rather than all-inclusive, so please forgive me if I fail to include your favorite site or even your own site (but of course with author's egoistical prerogative, I won't forget to mention my own sites!).

History of Protist Research

There are references in the ancient Chinese and Greek writings to malaria, but of course the etiology was not known. The study of protists began with the discovery of the microscope and its utilization by Robert Hooke and Antony van Leeuwenhoek. There are excellent sites on both Hooke (http://www.ucmp.berkeley.edu/history/hooke.html) and Leeuwenhoek (http://www.ucmp.berkeley.edu/history/leeuwenhoek.html). Leeuwenhoek was a true pioneer and is credited for making the first observations of bacteria, protozoa and cells from metazoans. He used single-lens microscopes with extremely short focal lengths. Functional reconstructions of his microscopes are described at http://www.mindspring.com/~alshinn/ and detailed plans for construction of a working replica are also provided (http://www.mindspring.com/~alshinn/Leeuwenhoekplans.html). Leeuwenhoek's studies had essentially no impact on science at that time. This was not due to a lack of communication since he wrote many letters to the Royal Society in London for almost 30 years and he entertained many important visitors to his home with glimpses of his "animalcules". A modern reexamination by Brian Ford of several long lost specimens sent to the Royal Society by Leeuwenhoek is described at http://www.sciences.demon.co.uk/wavintr.htm. The main reason for the lack of a major impact was probably the absence of a proper conceptual framework into which his observations would fit, perhaps caused partially, but not entirely, by his proclivity towards keeping the construction of his single-lens microscope methods a personal secret. The total lack of any meaningful influence of Leeuwenhoek on science in the 1700s meant that the true functional discovery of protists awaited the development of compound light microscopes corrected for both spherical and chromatic aberrations, microscopes that anyone could build or purchase. This fascinating history is documented in the interesting web site on the "History of the Light Microsope" (http://www.cas.muohio.edu/~mbi-ws/microscopes/history.html). A true paradigm shift — the articulation of the theory by Schleiden and Schwann (http://home.tiscalinet.ch/biografien/biografien/schwann.htm) that all organisms are composed of cells — finally occurred in 1839, soon after development of the achromatic lens in 1827.

The realization that many diseases are caused by unicellular microorganisms was another major conceptual change that greatly stimulated research on protists. In fact, further advances in the study of protists were closely tied with the development of tropical medicine. Some major advances in knowledge were the discovery of the *Leishmania* flagellates as the causal agents of kala-azar by William Leishman (http://www.lshtm.ac.uk/library/archives/leishman.html) and Charles Donovan in 1900—1903, the discovery of the trypanosome as the causal agent of animal and human sleeping sickness by Bruce in 1903, the discoveries of the malaria parasite in blood and the malaria parasite in the mosquito vector by Laveran in 1880 and Ross in 1897, and the discovery that the *Trypanozoma cruzi* trypanosome present in reduviid insects was the causal agent of a chronic human disease by Chagas in 1910. These are described and literature references are provided in my Online Course in

doi:10.1016/j.protis.2006.03.001

Molecular Parasitology (http://dna.kdna.ucla.edu/parasite_course-old/default.htm).

On-Line Images of Protists

Images of protists are available in several web sites. A site at http://megasun.bch.umontreal.ca/protists/gallery.html shows a variety of protists selected by their inclusion in the mitochondrial genome sequencing project. Information about the taxonomy, culture and classification is also presented. An excellent site with more than 5000 images of protists is the "Protist Information Server" at http://mtlab.biol.tsukuba.ac.jp/WWW/.

The comprehensive web site on the euglenoid protists by Triemer at http://www.plantbiology.msu.edu/triemer/Euglena/Index.htm is also an excellent resource of images and information. Other web sites which include images and life cycles of protist parasites are http://www.biosci.ohio-state.edu/~parasite/home.html, and http://www.cdfound.to.it/HTML/atlas.htm.

The medically important protozoal parasites, *Entamoeba* (http://homepages.lshtm.ac.uk/entamoeba/), *Giardia* (http://vm.cfsan.fda.gov/~mow/chap22.html), *Cryptosporidium* (http://www.cellsalive.com/parasit.htm, http://www.cdc.gov/ncidod/dpd/parasites/cryptosporidiosis/default.htm, http://vm.cfsan.fda.gov/~mow/chap24.html), *Trypanosoma cruzi* (http://www.who.int/tdr/diseases/chagas/default.htm), and *Leishmania* (http://www.who.int/tdr/diseases/leish/lifecycle.htm) have their own web sites. And of course there are several valuable web sites on the *Plasmodium* protists and malaria (http://www.cdc.gov/malaria/, http://www.malaria.org/, http://www-micro.msb.le.ac.uk/224/Malaria.html, http://www.who.int/topics/malaria/en/, http://www.wehi.edu.au/MalDB-www/who.html. A recent site describes the Malaria Vaccine Initiative (http://www.malariavaccine.org/). This is where the study of protists merges with the study of the diseases caused by these organisms.

Obviously, due to a variety of reasons such as biomedical funding and interest in ameliorating the human diseases, there are more researchers currently working on the protozoal parasites than on free-living protists. There are, however, some excellent web sites for free-living protists. For example the Clamy Center contains genomic, genetic and bibliographic information, the *Chlamydomonas* culture collection, and other resources for the *Chlamydomonas* community. A FAQ on *Clamydomonas* can be found at http://www.yale.edu/rosenbaum/green_yeast.html. And an extensive collection of images on protists in general is at the Microbial Digital Specimen Archives at http://protist.i.hosei.ac.jp/PDB/Images/menuE.html.

Protist Culture Collections

A list of the major protist culture collections can be found at http://megasun.bch.umontreal.ca/protists/pcco.html. This includes the major ATCC collection at http://www.atcc.org/.

Phylogeny of Protists

An interesting phylogenetic systematics site which has information on protists is http://www.ucmp.berkeley.edu/alllife/eukaryotasy.html. The "Tree of Life Web Project" (http://tolweb.org/tree/) is a "phylogenetic navigator" of over 4000 World Wide Web pages containing systematic and phylogenetic information about the diversity of life.

A site for the International Society for Evolutionary Protistology can be found at http://megasun.bch.umontreal.ca/isep/isep.html. The NCBI Taxonomy Homepage, which is operationally very important for those that enter protist sequences into Genbank or wish to recover them, can be found at http://www3.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html.

Protistologists

An attempt to provide a current and comprehensive list of email addresses for many researchers in the field of protistology can be found at: http://www.uga.edu/protozoa/names.htm (my own name is not included, so it is not really comprehensive!).

Course Sites which cover Protists

There are a variety of courses which have to varying extent put some or all of the course material online. Most of the courses I am familiar with deal with the parasitic protists. A course on Parasitology by Fred Opperdoes which includes much information on the protists is at http://www.icp.ucl.ac.be/~opperd/parasites/. A site by Titus Bradley with specific information on "Malaria

and Drug resistance" is at http://www-micro.msb.le.ac.uk/224/Bradley/Bradley.html. My own Online Course in Molecular Parasitology was mentioned above.

Protist Genome Sequence Databases

Probably the most useful type of Protist web sites are those covering the various genome sequencing projects. This includes the sequence databases and organism-specific blast servers. The African trypanosome Genome Project site at the Sanger Center is at http://www.sanger.ac.uk/Projects/T_brucei. The related *T. brucei* Genome Project site at TIGR is at http://www.tigr.org/tdb/mdb/tbdb/. The TIGR site has a useful Blast server at http://www.tigr.org/tdb/mdb/tbdb/seq_search.html where one can search sequences against the BAC, P1 and Sheared DNA Sequence Database. Another Blast server can be found at: http://www.ebi.ac.uk/blast2/parasites.html.

The UK *Leishmania* Genome Network web site is at http://www.sanger.ac.uk/Projects/L_major//. Sites with *Leishmania* chromosome maps can be found at http://www.ebi.ac.uk/parasites/LGN/chromsum.html and http://www.dbbm.fiocruz.br/genome/LGN/lshkary.html, and *Leishmania* EST sequences are at http://www.ebi.ac.uk/parasites/LGN/leishest.html. The literature reference for the final *Leishmania* genome results is http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=;Retrieve&db=PubMed&list_uids=16020728&dopt=Citation

Genome mapping information for *Plasmodium falciparum* is available at http://www.wehi.edu.au/MalDB-www/genomeInfo/MapData/MapData.htmll. The NCBI malaria genome site at http://www.ncbi.nlm.nih.gov/Malaria/ is a useful link to both the sequences obtained at NCBI and at the Sanger laboratory. The direct site for chromosomes 1, 3, 4, 9 and 13 from the Sanger lab is at http://www.sanger.ac.uk/Projects/P_falciparum/. A link to the complete *Plasmodium* chromosome 3 sequence is at http://www.sanger.ac.uk/Projects/P_falciparum/mal3/. A *P. falciparum* home page from the Sanger lab is at http://www.sanger.ac.uk/Projects/P_falciparum/. The TIGR *P. falciparum* Genome Database site for chromosomes 2, 10, 11 and 14 is at http://www.tigr.org/tdb/edb/pfdb/pfdb.html.

A genome web site for *Toxoplasma* is available at http://www.ebi.ac.uk/parasites/toxo/toxpage.html. The extremely useful *Toxoplasma* ToxoDB web site is at http://toxodb.org/ToxoDB.shtml.

A site for EST sequences from *Cryptosporidium parvum* is found at http://medsfgh.ucsf.edu/id/CpTags/home.html.

A useful web site that has updated codon bias tables for many protist parasites is at http://www.ebi.ac.uk/parasites/cutg.html. Sequence database of organelles also exist. MitBase, a "Comprehensive and Integrated Mitochondrial DNA Data base", is described at http://nar.oxfordjournals.org/cgi/content/full/27/1/128#hd10. It will include the mitochondrial genomes of several protists in addition to higher eukaryotes. GoBase at http://megasun.bch.umontreal.ca/gobase/gobase.html is a relational sequence database of mitochondrial genomes. It is a "taxonomically broad organelle genome database that organizes and integrates diverse data related to organelles". It will in the future include chloroplast sequences and sequences of prokaryotes thought to be related to the bacterial ancestors of mitochondria and chloroplasts.

The Goringer trypanosome mitochondria guide RNA database is also available at http://biosun.bio.tu-darmstadt.de/goringer/gRNA/gRNA.html, and a kinetoplast minicircle sequence database (it still has a somewhat distorted picture of Douglas Barker!) is available at http://www.ebi.ac.uk/parasites/kDNA/Source.html.

And last but not least, my own U-insertion/deletion web site at http://164.67.39.27/trypanosome/index.html and the related U-insertion/deletion edited sequence database at http://dna.kdna.ucla.edu/trypanosome/database.html were developed to allow ready access to information on this novel protist phenomenon. I created the database as a result of the reluctance of Genbank to have a mechanism to archive edited sequence information. For example, to indicate U insertions in the fully edited mRNA sequence in Genbank, one must add a separate feature for each U-insertion and any U-deletions are not indicated at all. Furthermore, there is no way to show the alignment of the guide RNA sequence with the pre-edited sequence and the edited sequence. Our database contains sequences in GCG format of pre-edited and unedited maxicircle genes from trypanosomatid protists. The data is provided both in the form of clickable maps of the maxicircle genomes and a table of all the genes organized by species and gene. A useful "map" format is employed to show the alignment of the pre-edited genomic sequences with the mature edited RNA sequences and the corresponding amino acid translations of the edited sequences. This preserves the U-deletion information which is

otherwise lost. U-insertion information is preserved as lower case u's. Alignments of the overlapping guide RNAs and edited mRNAs are also provided. The database also contains several published multiple ribosomal RNA alignments of kinetoplastids used to construct phylogenetic trees (http://dna.kdna.ucla.edu/trypanosome/files/alignments.html). This latter type of information is often very difficult to obtain.

Similar archival problems exist with the C-insertion edited *Physarum* mitochondrial RNA sequences (http://nsm1.utdallas.edu/bio/miller/physarum/physedit.htm), and to a lesser extent, with all other types of substitution editing (A to I, C to U) (http://164.67.39.27/RNA/index.aspx) or specific base modifications as occur in tRNAs (http://medlib.med.utah.edu/RNAmods/) and rRNAs.

The rRNA modification story is similar to the trypanosome editing story in that specific SnoRNAs encode the sequence localization information for methylations and pseudouridylations by forming at least two types of base-paired structures. Several databases exist for these sequencecs: http://lowelab.ucsc.edu/snoRNAdb/, http://www.bio.umass.edu/biochem/rna-sequence/Yeast_snoRNA_Database/snoRNA_DataBase.html, http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/home/, http://www-snorna.biotoul.fr/.

Of course there is a strong argument that specific databases should be developed for specific purposes, but to me that defeats the idea of a centralized archive monitored for accuracy and maintained by experts. Go to it, Genbank!

Summary

The World Wide Web is a marvelously chaotic landscape of valuable and also not so valuable information. A serious problem is the authentication of information provided on any specific site. But as in any complex informational ecosystem, Darwinian selection is at play and eventually results in the culling of bad information and the highlighting of good information. The most serious problem is that web-based digital information is ephemeral in that servers break down and URLs and web sites themselves change and often are not maintained. The long-term maintenance of information has been a problem for all civilizations and has been partially but inadequately solved in various ways — stone tablets with cuneiforms, papyrus and eventually paper with inscribed symbols, magnetic, CD and DVD discs with digital information, etc. But the novelty of web-based information is that information transfer is ubiquitous, instantaneous and world wide, and this resultant redundancy may actually be the mechanism for maintaining the information in the long run. A web site may disappear or be modified but the original site you may be sure is archived somewhere and probably in multiple locations. The "bots", "crawlers" and "spiders" of the many commercial search engines are actively gathering web information as we speak. In fact the pack rat mentality of "Do no evil" Google in storing forever all Google searches and the nefarious digital vacuum cleaning activities of the United States National Security Agency may paradoxically prove to be the saviors of civilization's web-based knowledge.

Larry Simpson
UCLA Microbiology, Immunology and Molecular Genetics
David Geffen School of Medicine at UCLA
1602 Molecular Science Building
P.O. Box 951489
Los Angeles, CA 90095-1489, USA

fax 1 310 825 4215
e-mail simpson@kdna.ucla.edu