

MINI-REVIEW:

# A small directory of World Wide Web sites for human molecular genetics

Glória R. Franco and Sérgio D.J. Pena

Our knowledge of human genetics in general, and human molecular genetics in particular, has virtually exploded in the past decade. As a consequence conventional printed books and even journals have not been able to keep pace with the hurricane of new data. Probably the best example of this is McKusick's *Mendelian Inheritance in Man (MIM)*, the "bible" of human genetics. *MIM* started in the early 1960s as a catalog of X-linked traits, already in response to a need for systematization of expanding data sets. The first edition of *MIM* was printed in 1966 with little more than 1000 entries. The 11th edition came out in 1994, already with two volumes and containing more than 6000 entries. By this time it was clear that the effort to continuously update *MIM* was Herculean and the "on line" version of *MIM (OMIM)* appeared as part of the Genome Database (GDB). Although a CD-ROM version appeared in 1995 it is clear now that the most effective way to make use of *OMIM* is through the Internet at the address <http://www3.ncbi.nlm.nih.gov/omim>. A roughly similar development can be observed for *Entrez*, which is a graphical, integrated retrieval system for DNA and protein sequence data, annotation, and associated bibliographic information, including literature abstracts. During the past few years, *Entrez* was available as a stand-alone application on CD-ROM, but this has been now discontinued after the last version reached the very cumbersome size of 5 disks! *Entrez* can be accessed by Internet at the address

<http://www3.ncbi.nlm.nih.gov/Entrez/>. But what do these addresses mean and how does one reach them?

Internet is a global computer network consisting of more than 21,000 connected local and regional networks in over 100 countries, that had its origins about 20 years ago as an experimental U.S. Defense Department network for military research, but has evolved considerably since that time (Boguski and Ouellette, 1995). Internet is neither owned nor operated by any single organization, being supported by the U.S. and other governments, and by private and international organizations. The logistic aspects of distributing and obtaining information through Internet were enormously simplified with the advent of the World Wide Web (WWW), a user-friendly interface composed of text information in the form of Hypertext and images, which after its development at CERN (the European Laboratory for Particle Physics), has grown rapidly in popularity and proliferated to thousands of sites worldwide (Boguski and Ouellette, 1995). To gain access to the WWW a software program or 'browser' is needed. Two browsers, Mosaic and Netscape, are widely available. Mosaic was developed by the National Center of Supercomputing Applications (NCSA) in the USA and is freely available from a number of sources. Netscape is a commercial program and its widespread use has introduced many features into the hypertext markup language (HTML), that do not work with Mosaic. WWW documents written in HTML contain links to other documents in the form of uniform resource locators (URLs). These links are displayed as highlighted (or colored) text or graphics and a mouse click will either display the linked file or establish connection to the WWW server that hosts the file or can make it available. The first characters

(xxxx://) of each URL address specify the type of protocol used (e.g. http or hypertext transfer protocol). The next part of the address specifies the host computer and resembles the host name included in Internet e-mail addresses (e.g. ebi.ac.uk). This address will normally be sufficient to get to the Home Page (main directory) of a WWW site from where you can select the particular resource you wish to access from the menus displayed by simply pointing and clicking. You may need to go through several levels before you reach the document you require. Alternatively, you can reach this point directly by typing in the full URL address, specifying the subdirectory and file you require, e.g. <http://www.ebi.ac.uk/subs/allsubs.html> will get you the file 'allsubs.html' from the subdirectory 'subs' (*Genome Digest*, 1995).

The purpose of this minireview is to provide molecular geneticists with a small directory of useful sites to visit on the Web. We use a table format throughout and give first the address of some major directories where further addresses can be retrieved. We then list some sites where recent knowledge on the genome project and information on medical molecular genetics can be obtained. Finally, give directions to reach the major molecular databases of human genetic interest available on WWW, where nucleotide, peptide or protein structural data can be obtained and homology searches performed. We also give a list of recommended readings. Our directory does not have by any means the intention of being complete. We welcome suggestions and new addresses, which will then be included in future editions.

**Table I - Major directories and information service.**

Name	Description	Location
Baylor Biologist's Control Panel	Biological database research tools, etc.	<a href="http://gc.bcm.tmc.edu:8088/bio/bio_home.html">http://gc.bcm.tmc.edu:8088/bio/bio_home.html</a>
Bioinformatics and Biology Resources	A good introduction to the WWW for researchers, including an excellent list of biology WWW resources	<a href="http://aeivos.wa.com/sciences/biology.html">http://aeivos.wa.com/sciences/biology.html</a>
ExPASy	Comprehensive library of documents describing e-mail servers, databases, and software for molecular biology; 2-D PAGE	<a href="http://expasy.hcuge.ch/www/">http://expasy.hcuge.ch/www/</a>
HUM-MOLGEN	Communications forum for human molecular genetics	<a href="http://www.informatik.uni-rostok.de/HUM-MOLGEN/">http://www.informatik.uni-rostok.de/HUM-MOLGEN/</a>
Human Genome Most Used Links	Mapping data sites, access to YAC search pages, Baylor College of Medicine services, etc.	<a href="http://www-ls.lanl.gov/HGhotlist.html">http://www-ls.lanl.gov/HGhotlist.html</a>
Human Genome Project	Major gateway to news and research data about the Human Genome Project	<a href="http://www.nchgr.nih.gov">http://www.nchgr.nih.gov</a>
MedWeb: Genetics and Molecular Biology	Tools, sites, databases, electronic newsletters and journals, etc.	<a href="http://www.cc.emory.edu/WHSC/medweb.gen.html">http://www.cc.emory.edu/WHSC/medweb.gen.html</a>
Molecular Biology and Bioinformatics	This sampler allows quick access to interesting WWW servers	<a href="http://www.utu.fi/~jarnie/biolinks.html">http://www.utu.fi/~jarnie/biolinks.html</a>
Molecular Biology on the Net	Sequence analysis facilities and information; molecular biology on-line publications, etc.	<a href="http://www.ch.ic.ac.uk/medbact/molbiol.html">http://www.ch.ic.ac.uk/medbact/molbiol.html</a>
Molecular Biology on the Web	WWW information service run by the Pasteur Institute	<a href="http://www.pasteur.fr/other/biology/english/rubriques-uk.html">http://www.pasteur.fr/other/biology/english/rubriques-uk.html</a>
Molecular Biology Tools	WWW information service run by CINVESTAV in Mexico City	<a href="http://www.gene.cinvestav.mx/tools.html">http://www.gene.cinvestav.mx/tools.html</a>
PEDRO's BioMolecular Research Tools	Survey of Molecular Biology Databases and Servers	<a href="http://www.public.iastate.edu/~pedro/research_tools.html">http://www.public.iastate.edu/~pedro/research_tools.html</a>
Randal Smith database	Survey of Molecular Biology Databases and Servers	<a href="http://www.ai.sri.com/people/pkarp/mimbd/rsmith.html">http://www.ai.sri.com/people/pkarp/mimbd/rsmith.html</a>
Reference Library Data Base (RLDB)	Hans Lehrach's Reference Library System	<a href="http://gea.lif.icnet.uk/">http://gea.lif.icnet.uk/</a>
UMKC Molecular Biology, Biochemistry and Neuroscience WebPage	Medline, nucleotide and protein database searches; the UMKC Molecular Biology Database, etc. Site still under construction	<a href="http://cctr.umkc.edu/user/drosenthal.mbb.html">http://cctr.umkc.edu/user/drosenthal.mbb.html</a>
World-Wide Web Virtual Library: Biochemistry, Biophysics and Molecular Biology	Information categorized first by provider, then by subject	<a href="http://golgi.harvard.edu/biopages/biochem.html">http://golgi.harvard.edu/biopages/biochem.html</a>

**Table II - Information on genome projects.**

Name	Description	Location
<b>Human Genome</b>		
Human Genome Project	Major gateway to news and research data about the Human Genome Project	<a href="http://www.nchgr.nih.gov">http://www.nchgr.nih.gov</a>
USA Department of Energy (DOE) Human Genome Program	General information about the Human Genome Program databases and bioinformatics resources and the DOE Human Genome Centers research projects	<a href="http://www.er.doe.gov/production/oher/hug_top.html">http://www.er.doe.gov/production/oher/hug_top.html</a>
UK Medical Research Council	Sponsors of UK Human Genome Mapping Project	<a href="http://www.nimr.mrc.ac.uk/MRC">http://www.nimr.mrc.ac.uk/MRC</a>
GenomeNet	Information on the Japanese Human Genome Project	<a href="http://www.genome.ad.jp/">http://www.genome.ad.jp/</a>
Location Database (ldb)	Integrated genetic and physical maps of the human genome	<a href="http://cedar.genetics.soton.ac.uk/public_html">http://cedar.genetics.soton.ac.uk/public_html</a>
CEPH-Généthon-map	Genetic and physical maps of the human genome	<a href="http://www.cephb.fr/bio/ceph-genethon-map.html">http://www.cephb.fr/bio/ceph-genethon-map.html</a>
Généthon Physical Map Database (HugeMap)		<a href="http://www.genethon.fr/genethon_en.html">http://www.genethon.fr/genethon_en.html</a>
Integrated Human Genetic Maps		<a href="http://www.chlc.org/">http://www.chlc.org/</a>
Radiation Hybrid Database		<a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a>
<b>Other Genomes</b>		
Lawrence Berkeley National Laboratory Human Genome Center	Information on HGC projects, Syndb Online LBNL Human/Mouse ACEDB-style Database. <i>Drosophila</i> physical mapping, etc.	<a href="http://www-hgc.lbl.gov/GenomeHome.html">http://www-hgc.lbl.gov/GenomeHome.html</a>
Mouse Genome Database	Maintained at the Jackson Laboratory	<a href="http://www.informatics.jax.org/">http://www.informatics.jax.org/</a>
Flybase ( <i>Drosophila</i> )	<i>Drosophila</i> Genome Project	<a href="http://morgan.harvard.edu">http://morgan.harvard.edu</a>
<i>Caenorhabditis elegans</i> cosmid sequences	<i>Caenorhabditis elegans</i> Genome Project	<a href="http://www.sanger.ac.uk/ftpsite.html">http://www.sanger.ac.uk/ftpsite.html</a>
<i>Saccharomyces</i> Genome Database	Available using ACEDB software for Macintosh and a variety of UNIX systems	<a href="http://genome-www.stanford.edu">http://genome-www.stanford.edu</a>
Japan Rice Genome Research Project		<a href="http://www.staff.or.jp">http://www.staff.or.jp</a>

**Table III - Major WWW molecular database host centers.**

Abbreviation	WWW Site/Server	Location
NCBI	National Center for Biotechnology Information	<a href="http://www3.ncbi.nlm.nih.gov/">http://www3.ncbi.nlm.nih.gov/</a>
NIH	National Institutes of Health	<a href="http://www.nih.gov/">http://www.nih.gov/</a>
EBI	European Bioinformatics Institute	<a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a>
EMBL	European Molecular Biology Laboratory	<a href="http://www.embl-heidelberg.de/">http://www.embl-heidelberg.de/</a>
NIG	National Institute of Genetics	<a href="http://www.nig.ac.jp/">http://www.nig.ac.jp/</a>
NCGR	National Center for Genome Resources	<a href="http://www.ncgr.org/">http://www.ncgr.org/</a>
JHU	Johns Hopkins University Bioinformatics Web Server	<a href="http://www.gdb.org/hopkins.html">http://www.gdb.org/hopkins.html</a>
ExpASy	ExpASy Molecular Biology Server, Geneva University Hospital and the University of Geneva	<a href="http://expasy.hcuge.ch/www/">http://expasy.hcuge.ch/www/</a>
TIGR	The Institute for Genomic Research	<a href="http://www.tigr.org/">http://www.tigr.org/</a>
HGC/ BCM	Human Genome Center/Baylor College of Medicine	<a href="http://www.bcm.tmc.edu/">http://www.bcm.tmc.edu/</a>
BNL	Brookhaven National Laboratory	<a href="http://suntid.bnl.gov/bnl.html">http://suntid.bnl.gov/bnl.html</a>
ICGEB	International Center for Genetic Engineering and Biotechnology	<a href="http://base.icgeb.trieste.it/">http://base.icgeb.trieste.it/</a>
FHCRC	Fred Hutchinson Cancer Research Center	<a href="http://blocks.fhcrc.org/">http://blocks.fhcrc.org/</a>
UCL	University College London - Department of Biochemistry and Molecular Biology	<a href="http://www.biochem.ucl.ac.uk/">http://www.biochem.ucl.ac.uk/</a>
ABC	Agricultural Biotechnology Center	<a href="http://www.abc.hu/">http://www.abc.hu/</a>
CEPH	Centre d'Etudes du Polymorphisme Humain Foundation Jean Dausset	<a href="http://www.cephb.fr/">http://www.cephb.fr/</a>
CHLC	Cooperative Human Linkage Center	<a href="http://www.chlc.org/">http://www.chlc.org/</a>
HGG	Human Genetics Group	<a href="http://cedar.genetics.soton.ac.uk/public_html">http://cedar.genetics.soton.ac.uk/public_html</a>
	Sanger Center	<a href="http://www.sanger.ac.uk/">http://www.sanger.ac.uk/</a>
	Généthon	<a href="http://www.genethon.fr/genethon_en.html">http://www.genethon.fr/genethon_en.html</a>

**Table IV - Databases of interest for medical molecular genetics.**

Disease/Mutation Databases:	Address
Breast Cancer Mutation Data Base	<a href="http://www.nchgr.nih.gov/Intramural-research/Lab-transfer/Bic">http://www.nchgr.nih.gov/Intramural-research/Lab-transfer/Bic</a>
Cystic Fibrosis Mutation Data Base	<a href="http://www.genet.sickkids.on.ca/cftr/">http://www.genet.sickkids.on.ca/cftr/</a>
MITOMAP (Human Mitochondrial Genome Database)	<a href="http://www.gen.emory.edu/mitomap.html">http://www.gen.emory.edu/mitomap.html</a>
Neurogenetics homepage of the MGH: Information on a variety of neurogenetic diseases including neurofibromatoses, tuberous sclerosis and Von Hippel-Lindau disease	<a href="http://neurosurgery.mgh.harvard.edu/ngenethp.htm">http://neurosurgery.mgh.harvard.edu/ngenethp.htm</a>
OMIM (Online Mendelian Inheritance in Man)	<a href="http://www3.ncbi.nlm.nih.gov/omim">http://www3.ncbi.nlm.nih.gov/omim</a>
Phenylalanine hydroxylase (PAH) Mutation Database	<a href="http://www.mcgill.ca/pahdb/">http://www.mcgill.ca/pahdb/</a>
Search the Human X Chromosome	<a href="http://ibc.wustl.edu:70/Oh/CGM/human_x">http://ibc.wustl.edu:70/Oh/CGM/human_x</a>
von Willebrand Factor Database	<a href="http://mmg2.im.med.umich.edu/vWF/">http://mmg2.im.med.umich.edu/vWF/</a>

Table V - Major molecular databases.

Database	Description	Address
<b>Nucleotide Databases</b>		
Human Genome Database (GDB)	Genomic database composed of all major categories of genetic and physical maps	<a href="http://www.gdb.gdb.org/">http://www.gdb.gdb.org/</a>
The Genetic Location Database (LDB)	Analytical database for constructing integrated genetic and physical maps	<a href="http://cedar.genetics.soton.ac.uk/public_html">http://cedar.genetics.soton.ac.uk/public_html</a>
GenBank	Nucleotide sequence database	<a href="http://www.ncbi.nlm.nih.gov/Search/index.html">http://www.ncbi.nlm.nih.gov/Search/index.html</a>
European Molecular Biology Laboratory Database (EMBL)	Nucleotide sequence database	<a href="http://www.ebi.ac.uk/queries/queries.html">http://www.ebi.ac.uk/queries/queries.html</a>
DNA Data Bank of Japan (DDBJ)	Nucleotide sequence database	<a href="http://www.ddbj.nig.ac.jp/">http://www.ddbj.nig.ac.jp/</a>
The Expressed Sequence Tag Database (dbEST)	Database containing information on single-pass cDNA sequences	<a href="http://www.ncbi.nlm.nih.gov/dbEST/index.html">http://www.ncbi.nlm.nih.gov/dbEST/index.html</a>
The Sequence Tagged Sites Database (dbSTS)	Database of sequence and mapping data on STSs	<a href="http://www.ncbi.nlm.nih.gov/dbSTS/index.html">http://www.ncbi.nlm.nih.gov/dbSTS/index.html</a>
Integrated Molecular Analysis of Gene Expression (IMAGE)	Public database of cDNA/EST sequences and related data	<a href="http://www-bio.lnl.gov/bbrp/genome/genome.html">http://www-bio.lnl.gov/bbrp/genome/genome.html</a>
The Human cDNA Database (HCD)	Database of cDNA/EST sequences and related data. Access restricted to registered users	<a href="http://www.tigr.org/tdb/hcd/hcd.html">http://www.tigr.org/tdb/hcd/hcd.html</a>
Genome Sequence Data Base (GSDB)	Relational database of DNA sequences and annotation	<a href="http://www.ncgr.org/gsdb/gsdb.html">http://www.ncgr.org/gsdb/gsdb.html</a>
Human CpG Island database		<a href="http://biomaster.uio.no/cpgisle.html">http://biomaster.uio.no/cpgisle.html</a>
<b>Protein Databases</b>		
SWISS-PROT	Protein sequence database with a high level of annotation and a minimal level of redundancy	<a href="http://expasy.hcuge.ch/sprot/sprot-top.html">http://expasy.hcuge.ch/sprot/sprot-top.html</a>
Protein Identification Resource (PIR)	International protein sequence database	<a href="http://www.gdb.org/Dan/proteins/pir.html">http://www.gdb.org/Dan/proteins/pir.html</a>
OWL	Non-redundant protein sequence database produced from SWISS-PROT, PIR, GenPept and NRL-3D	<a href="http://www.gdb.org/Dan/proteins/owl.html">http://www.gdb.org/Dan/proteins/owl.html</a>
PROSITE	Database of biologically significant protein sites and patterns derived from multiple alignments of proteins	<a href="http://expasy.hcuge.ch/sprot/prosite.html">http://expasy.hcuge.ch/sprot/prosite.html</a>
BLOCKS	Database of multiple alignments of most conserved ungapped regions in protein families	<a href="http://www.blocks.fhrc.org/">http://www.blocks.fhrc.org/</a>
PRINTS	Database of protein fingerprints derived from multiple alignments of conserved motifs that can be in distinct regions of the protein	<a href="http://www.biochem.ucl.ac.uk/bms/dbbrowser/PRINTS/PRINTS.html">http://www.biochem.ucl.ac.uk/bms/dbbrowser/PRINTS/PRINTS.html</a>
DOMAIN	Database of protein domains	<a href="http://www.abc.hu/blast.html">http://www.abc.hu/blast.html</a>
SBASE	Database of annotated protein domains	<a href="http://base.icgeb.trieste.it/sbase/">http://base.icgeb.trieste.it/sbase/</a>
ProDom	Database of families of protein domains generated from all-against-all comparison of proteins in the SWISS-PROT database	<a href="http://www.sanger.ac.uk/~esr/prodom.html">http://www.sanger.ac.uk/~esr/prodom.html</a>
Protein Data Bank (PDB)	Database of experimentally determined 3D structures of biological macromolecules	<a href="http://www.pdb.bnl.gov/">http://www.pdb.bnl.gov/</a>
Molecular Modeling Database (MMDB)	3D structure database derived from PDB	<a href="http://www.ncbi.nlm.nih.gov/structure">http://www.ncbi.nlm.nih.gov/structure</a>
NRL-3D	Sequence-structure database derived from PDB and searchable within the PIR environment	<a href="http://www.gdb.org/Dan/proteins/nrl3d.html">http://www.gdb.org/Dan/proteins/nrl3d.html</a>
SWISS-3D IMAGE	Database of 3D images of proteins and other macromolecules	<a href="http://expasy.hcuge.ch/sw3d/sw3d-top.html">http://expasy.hcuge.ch/sw3d/sw3d-top.html</a>
Molecules R US	Database of 3D images	<a href="http://molbio.info.nih.gov/cgi-bin/pdb">http://molbio.info.nih.gov/cgi-bin/pdb</a>
FSSP	Database for 3D protein fold classification based on all-against-all comparison of structures present in PDB	<a href="http://www.embl-heidelberg.de/dali/fssp">http://www.embl-heidelberg.de/dali/fssp</a>

## RECOMMENDED READING

**Boguski, M.S. and Ouellette, B.F.F.** (1995). Internet basics for biologists. In: *Current Protocols in Human Genetics*. Vol. 2. (Dracopoli, N.C. *et al.*, eds). John Wiley & Sons, New York, pp. 11.1.1-11.3.44.

**Fischer, C., Schweigert, S., Spreckelsen, C. and Vogel, F.** (1996). Programs, databases, and expert systems for human geneticists - a survey. *Hum Genet.* 97: 129-137.

**Genome Digest** (1995). Genome information on the World Wide Web. *Genome Digest* 2: 10-13. (Genome Digest is a free publication of HUGO Europe, One Park Square West, London NW1 4LJ, England).

**Rashbass, J. and Walsh, L.** (1996). *The Trends Guide to Internet*. Elsevier Trend Journals, Cambridge (obtainable at <http://www.elsevier.com/locate/trendsguide>).