

Jon K. Lærdahl,
Structural Bioinformatics

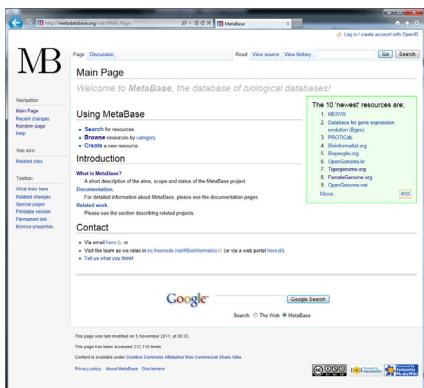
Databases/Resources on the web

Jon K. Lærdahl
jonkl@medisin.uio.no

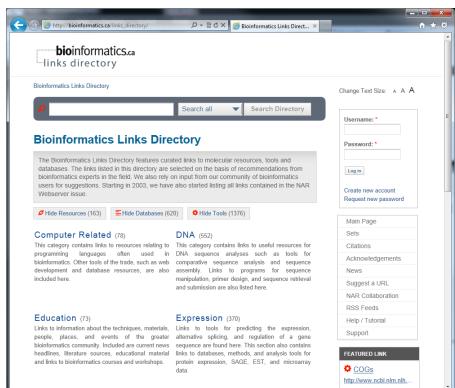
 

Jon K. Lærdahl,
Structural Bioinformatics

A lot of biological databases available on the web...



MetaBase, the database of biological databases (1801 entries)
 - <http://metadatabase.org>

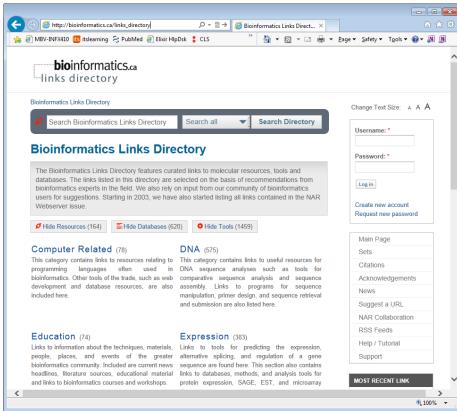


bioinformatics.ca – links directory (620 databases)
 - http://bioinformatics.ca/links_directory

Jon K. Lærdahl,
Structural Bioinformatics

btw, the **bioinformatics.ca** links directory is an excellent resource



bioinformatics.ca – links directory

- http://bioinformatics.ca/links_directory
- Currently
 - 1459 tools
 - 620 databases
 - 164 “resources”
- The problem is not to find a tool or database, but to know what is “gold” and what is “junk”

UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

Some important centres for bioinformatics

- National Center for Biotechnology Information (NCBI)
 - part of the US National Library of Medicine (NLM), a branch of the National Institutes of Health
 - located in Bethesda, Maryland
- European Bioinformatics Institute (EMBL-EBI)
 - part of European Molecular Biology Laboratory (EMBL)
 - located in Hinxton, Cambridgeshire, UK

UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

NCBI databases

- Provided the GenBank DNA sequence database since 1992
- Online Mendelian Inheritance in Man (OMIM) - known diseases with a genetic component and links to genes
 - started early 1960s as a book
 - online version, OMIM, since 1987
 - on the WWW by NCBI in 1995
 - currently >22,000 entries (14,400 genes)
- EST - nucleotide database subset that contains only Expressed Sequence Tag records
- Gene - genes and associated information for a number of organisms in addition to and including human
- Protein sequence database - collection of protein sequence entries compiled from a variety of sources including Swiss-Prot, PIR, PRF, PDB, and translations from annotated coding regions in GenBank and RefSeq
- PubMed - access to over 15 million citations from MEDLINE and additional life sciences journals
- SNP - repository for both single nucleotide substitutions and short deletion and insertion polymorphisms

All data is publicly available


Oslo
universitetssykehus
UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

NCBI databases

Table 1. The Entrez Databases (as of September 1, 2012)

Database	Section within this article	Records	Data source
Site search	Introduction	10 886	N
Assembly	Recent developments	9597	D, C, N
PubMed	Literature	23 076 32	C
PubMed central	Literature	2 532 184	D, C
NLM catalog	Literature	1 461 835	C, N
MeSH	Literature	236 253	N
Books	Literature	186 112	C, N
Taxonomy	Taxonomy	932 345	C, N
Protein	DNA and RNA	75 666 069	D (GenBank)
Nucleotide	DNA and RNA	66 319 706	D (GenBank), C, N
GSS	DNA and RNA	34 533 114	D (GenBank)
BioSample	DNA and RNA	970 304	N
SRP	DNA and RNA	228 739	D
ProjSet	DNA and RNA	159 145	D (GenBank)
Protein	Proteins	56 394 380	C, N
Protein clusters	Proteins	794 663	N
GEO profiles	Genes and expression	63 811 486	D
Probe	Genes and expression	14 248 527	D
Gene	Genes and expression	11 290 772	C, N
UniGene	Genes and expression	5 831 327	N
GEO data sets	Genes and expression	841 518	N
Biosystems	Genes and expression	396 029	C
Homologene	Genes and expression	133 012	N
Gene	Genomes	29 597 231	D, N
UnisTS	Genomes	545 553	D (dbSTS)
BioProject	Genomes	58 227	D
Genome	Genomes	8276	C, N
Epigenomics	Genomes	5484	D
SNP	Genetics and medicine	162 674 447	D (dbSNP), N
dbVar	Genetics and medicine	2 729 416	D
dbGaP	Genetics and medicine	143 624	D
Online mendelian inheritance in animals	Genetics and medicine	2810	C
PubChem substance	Chemicals and bioassays	100 157 112	D
PubChem compound	Chemicals and bioassays	35 545 66	N
PubChem bioassay	Chemicals and bioassays	621 642	D
Structure	Domains and structures	83 913	C, N
CDD	Domains and structures	46 389	C, N

D, direct submission; C, collaboration/agreement; N, internal NCBI/NLM curation.


Oslo
universitetssykehus
UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

EMBL-EBI databases

- European Nucleotide Archive (ENA) nucleotide sequence database
- Ensembl - automatic and manually curated annotation on selected eukaryotic (vertebrate) genomes
- Ensembl Genomes – Ensembl for “all other organisms”
- UniProt – protein sequence and functional information
- ChEMBL – database of bioactive compounds
- IntAct - repository of molecular interactions, including protein-protein, protein-small molecule and protein-nucleic acid interactions
- CiteXplore – 25 million literature abstracts including PubMed, Agricola & patents
- Gene Ontology (GO) - controlled vocabulary to describe gene and gene product attributes in any organism
- Gene Ontology Annotation (GOA) – GO annotations for proteins in UniProt

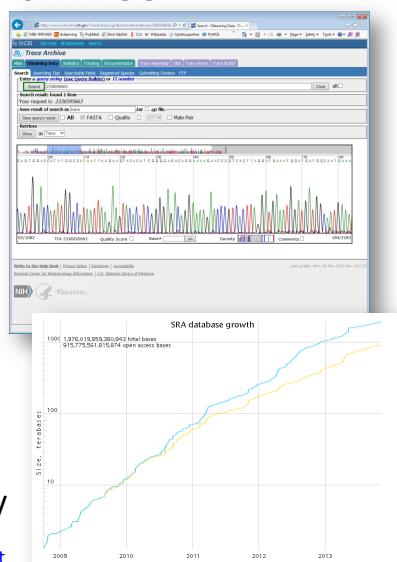
 Oslo universitetssykehus All data is publicly available UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

NCBI «Trace Archives»

- Trace Archive
 - Repository of raw data sequencing traces from gel and capillary electrophoresis sequencers
 - >2 billion traces
- Sequence Read Archive (SRA)
 - Data from high-throughput sequencing (454, Illumina, IonTorrent, SOLiD, etc.)
 - 2300 Tbases (2.3×10^{15}) open access sequences
 - At present >1 Tbase added daily

<http://nar.oxfordjournals.org/content/40/D1/D54.abstract>



 Oslo universitetssykehus 1 Pbp ≈ 100,000 human genomes UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

UniProt 

- Database of protein sequences and functional annotations – “a single worldwide database of protein sequence and function” (2002)
- UniProt consortium
 - EMBL-EBI
 - Swiss Institute of Bioinformatics (SIB)
 - Swiss-Prot (Amos Bairoch, 1986)
 - TrEMBL (Translated EMBL Nucleotide Sequence Data Library, 1996)
 - Protein Information Resource (PIR)
 - roots in Margaret Dayhoff's *Atlas of Protein Sequence and Structure* (1965)
- <http://www.uniprot.org>

 Oslo universitetssykehus

UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

An even better place to look for good biological databases -

Nucleic Acids Research

Contents

Volume 41, Database issue, January 1, 2013

The 2013 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection	X.M.Fernández-Sáez and M.Y.Galperin	D1-D7
Database resources of the National Center for Biotechnology Information	NCBI Resource Coordinators	D8-D20
The International Nucleotide Sequence Database Collaboration	Y.Nakamura, G.Cochrane and I.Karsch-Mizrachi on behalf of the International Nucleotide Sequence Database Collaboration	D21-D24
DDBJ new system and service refactoring	O.Ogawara, J.Mashima, Y.Kodama, E.Kanamitsu, Y.Nakamura, K.Okubo and T.Takamiya	D25-D29
Facing growth in the European Nucleotide Archive	G.Chadronne, B.Alabio, C.Amit, J.Bower, A.Cerdeño-Tarugo, I.Cleland, R.Gibson, N.Goodliffe, M.Jang, S.Kay, R.Lemonen, X.Liu, P.McLaren, D.McNeil, J.McNeil, N.Paksereshti, S.Paliogianni, Y.Park, S.Plaister, R.Radhakrishnan, S.Rivière, M.Rossetto, A.Santoni, A.Schäfer, A.Schirmer, A.Schoen, A.Torbio, D.Vaughn and V.Zalinin	D30-D35
GenBank	D.A.Benson, M.Cavanaugh, K.Clark, I.Karsch-Mizrachi, D.J.Lipman, J.Ostell and E.W.Sayers	D36-D42
Update on activities at the Universal Protein Resource (UniProt) in 2013	The UniProt Consortium	D43-D47
Easenol 2013	P.Bill, I.Björk, M.R.Amde, D.Borelli, K.Böhl, S.Brown, D.Carrasco-Silva, P.Clapham, G.Couper, S.Furley, S.Fitzgerald, J.Gold, C.García-Gómez, L.Gordon, T.Hourlier, S.Hunt, T.Ishii, S.Ishii, A.Kelkar, S.Kelkar, M.Konowalczyk, E.Kubischa, S.Lingden, T.Maurer, W.M McLaren, M.Muffato, R.Nag, B.O'Donnell, A.Ogihara, S.Ogihara, A.Otton, E.Pritchard, H.S.Riat, G.R.S.Ritchie, M.Ruffier, M.Schuster, D.SherpaRi, S.Sobhak, K.Taylor, A.Tan, A.Tan, A.Tan, A.Tan, A.Tan, A.Tan, B.L.Aken, E.Birney, F.Cunningham, I.Dunham, J.Harrow, J.Heger, T.J.P.Hubbard, N.Ianoukhov, K.Ianoukhov, J.Kel, G.Spudich, A.Yates, A.Zadisa and S.M.J.Scott	D48-D55

<http://nar.oxfordjournals.org/content/41/D1.toc>

 Oslo universitetssykehus

UiO : Department of Informatics
University of Oslo

While we are visiting NAR: a good place to look for bioinformatics tools

Jon K. Lærdahl,
Structural Bioinformatics

Nucleic Acids Research

Contents

Volume 41, Web Server issue, July 1, 2013

Editorial: Nucleic Acids Research Annual Web Server Issue in 2013	G.Rosenau	W1-W2
DALIGN in GORBIICS—multiple sequence alignment using various sources of external information	LAI Ait, Z.Yanuk and B.Morgensen	W3-W7
MISTIC: mutual information server to infer interaction sites	J.L.Simoneit, E.Tripa, A.Chromoszorc, M.Nelson and C.Mario Burla	W8-W14
R3D Align web server for global nucleotide to nucleotide alignments of RNA 3D structures across families and superfamilies of metagenome new family from the MFATP sequence alignment server with enhanced mobility	R.R.Kalirig, A.J.Petrov, N.R.Lentini and C.Z.Cribel	W15-W21
BLAST+: more efficient report with mobility improvements	S.Kunkus, C.M.Zmasek, O.Nishimura and I.Zaslavsky	W22-W28
iBLLUST: an immunoglobulin variable domain sequence analysis tool	J.Ye, N.MA, T.L.Madden and J.M.Ostlund	W29-W33
Genome Maps, a new generation genome browser	I.Gómez, F.Silvestre, R.Sánchez, A.d.Maria, R.Alcolea, J.A.Gómez, M.Bleas and J.Díez-Pazos	W41-W46
NARflex: a web server for the study of nucleic acid flexibility	A.Karplus, L.Wang, C.Gómez-González, C.Gómez, J.I.Gelpí and M.Oreco	W47-W55
DNShape: a method for the high-throughput analysis of DNA sequence features on a genomic scale	T.Zhou, L.Yang, Y.Liu, J.Dear, A.C.Dantas-Mello, J.Costa, R.J.Fred and R.Kroth	W56-W62
INSTEIN: an ensemble-based tool for integrative analysis of expression data	J.Xia, C.D.Fritz, M.Kroger, O.M.Penna, D.Hall, S.J.Williams, J.W.Holland	W63-W70
User-friendly solution for microarray quality control and preprocessing on ArrayAnalyst.org	L.M.Tijssen, M.Jallard, M.L.Adriaens, S.Gaj, P.J.de Groot, M.Müller and C.T.Evelo	W71-W76
WEB-based GENE Set Enrichment Toolkit	J.Wang, G.Dunigan, Z.Shi and B.Zhang	W77-W83
FIDEAC: a server for the functional interpretation of differential expression analysis	D.D'Andrea, I.Grossi, M.Mazzoppi and A.Tremonti	W84-W88

Nucleic Acids Res. Web server issues

- released once every year, in July
- 11th issue (2013)
- 95 web servers

<http://nar.oxfordjournals.org/content/41/W1.toc>

If you need an article or a citation for a bioinformatics tool or database, the NAR web server or databases issues are often good places to look

UiO : Department of Informatics
University of Oslo

Huge number of databases!

Jon K. Lærdahl,
Structural Bioinformatics

• In bioinformatics, the number of databases, tools, algorithms, and papers is enormous

- impossible to have an overview, especially if bioinformatics is not your main research area
- instead of trying to do everything yourself:

Get yourself a bioinformatics expert colleague or collaborator!

<http://www.oxfordjournals.org/nar/database/a>

NAR online Molecular Biology Database Collection, currently contains 1512 databases

UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

Good and bad databases

- Some are exceptionally good, well maintained and often updated
 - EMBL-EBI, NCBI, Ensembl,...
 - <http://string.embl.de>
 - <http://www.pdb.org>
 - Maintained by 10s and 100s of experts...
- Species specific
 - [\(Schizosaccharomyces pombe\)](http://www.pombase.org)
 - [\(Drosophila\)](http://flybase.org)
 - [\(Escherichia coli K-12 MG1655\)](http://ecocyc.org)
- Unique content
 - <http://www.proteinatlas.org>
- Also many have poor quality, are never updated, are unreliable

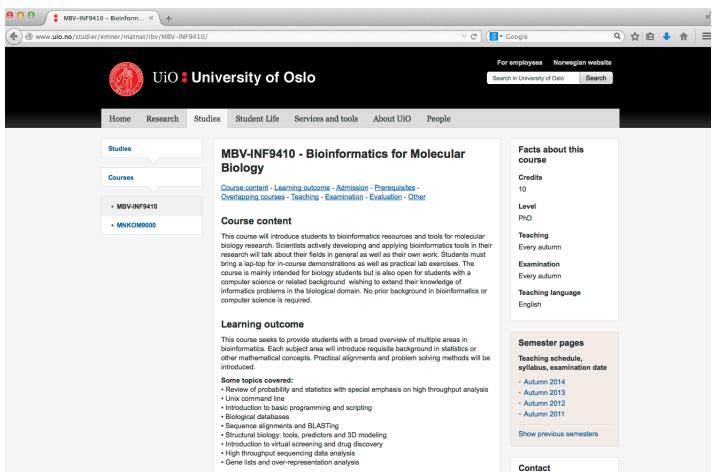
Trick is to know what is good and what is bad...

 Oslo
universitetssykehus

Let your favourite bioinformatician follow the field!

Jon K. Lærdahl,
Structural Bioinformatics

Much more in MBV-INFx410



The screenshot shows the course page for MBV-INFx410. The page includes a navigation bar with links for Home, Research, Studies, Student Life, Services and tools, About UIO, and People. Under the Studies menu, 'Courses' is selected, showing options for MBV-INFx410 and MNKOM9000. The main content area displays information about the course, including its name, credits (10), level (P/D), teaching (every autumn), examination (every autumn), and teaching language (English). It also lists facts about the course such as prerequisites, learning outcomes, and some topics covered. A sidebar on the right provides links to semester pages, teaching schedule, syllabus, and examination date, listing autumn semesters from 2011 to 2013.

 Oslo
universitetssykehus

UIO : Department of Informatics
University of Oslo

**CLS Wednesday seminars –
Bioinformatics/CLS seminars every 14
days**

Jon K. Lærdahl,
Structural Bioinformatics

UiO Department of Informatics
The Faculty of Mathematics and Natural Sciences

Wednesday seminars

Wednesday seminars - Computational Life Science (CLS) seminar series

Time: Wednesdays at 11:15 Place: Kravitz Syntex Inc.

Seminars for and by researchers from the computational biology community in the Oslo area, both presentations from local groups and invited speakers.

Please contact Jon K. Lærdahl or Ingeborg Bjørnstad Engås.

CLS seminar series 2011-12

Seminars held at the Joint Centre of Bioinformatics Oslo, 2006-2010

Upcoming events

Hans A. Eide
Time and place: Nov. 7, 2012 11:15 AM - 12:00 PM, Seminar room Java (2423), On-Jean Johansen Døttes hus | Add to calendar

Bastien Chervaux
Time and place: Dec 5, 2012 11:15 AM - 12:00 PM, Seminar room Java (room 2423), On-Jean Johansen Døttes hus | Add to calendar

All Upcoming events
Previous events

Earlier seminars

Abnormal flows in the brain and the spine: how to better understand diseases by computation
Last modified Feb 25, 2011 04:21 PM by Ingeborg Bjørnstad Engås

Analysis of genomic copy number alterations in cancer
Last modified Jan 26, 2012 12:30 AM by Ingeborg Bjørnstad Engås

Challenges of fusion gene detection (HTS series)
Last modified Feb 3, 2011 02:59 AM by Sigrid Væra Lie

CHP-Ses reveals changes in chromatin landscapes during osteogenic differentiation (HTS series)
Last modified Mar 30, 2011 10:54 AM by Ingeborg Bjørnstad Engås

CMG-Biotools, a platform for Comparative Microbial Genomics
Last modified Apr 12, 2011 11:41 PM by Ingeborg Bjørnstad Engås

Coremine - text mining for organizing information
Last modified Nov 9, 2011 12:15 PM by Ingeborg Bjørnstad Engås

Darwinsky_2012.pdf
Last modified Jan 27, 2012 09:53 AM by Ingeborg Bjørnstad Engås

Oslo universitetssykehus

UiO Department of Informatics
University of Oslo

**cbo-all@usit.uio.no – the mailing list for
bioinformatics and computational biology in the
Oslo region**

Jon K. Lærdahl,
Structural Bioinformatics

E-postlister ved UIO - Universitetets Senter for Informasjonsteknologi

cbo-all@usit.uio.no
Bioinformatics - Oslo region

News for people interested in bioinformatics in the Oslo region

Subscribers: 338
Owners: ehong@ifi.uio.no
ln.angen@ifi.uio.no
jostein@ifi.uio.no
torbjorn@ifi.uio.no

Contact owners

Subscribe
Unsubscribe
Info
Archive
Post
RSS

Language Selection: English
Powered by Sopra 5.4.7

- News about
 - Seminars
 - Courses
 - Jobs
 - Conferences

Relevant mainly for people in the Oslo region

Anyone can send an e-mail to the list

Curators check that the message is relevant (to avoid spam) and releases the message

Currently >400 subscribers

UiO Department of Informatics
University of Oslo

Oslo universitetssykehus

Jon K. Lærdahl,
Structural Bioinformatics

Ensembl genome browser and database

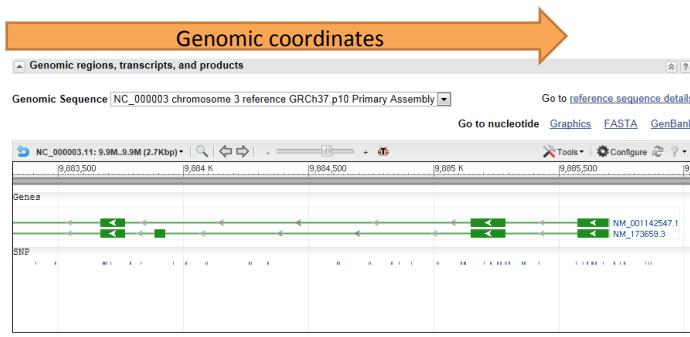
 Oslo
universitetssykehus

UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

Genome browsers

- Graphical interface for genomic data
- Shows information from biological databases mapped onto genomic sequence



Genomic coordinates →

Various annotations = "tracks"

 Oslo
universitetssykehus NCBI Gene database UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

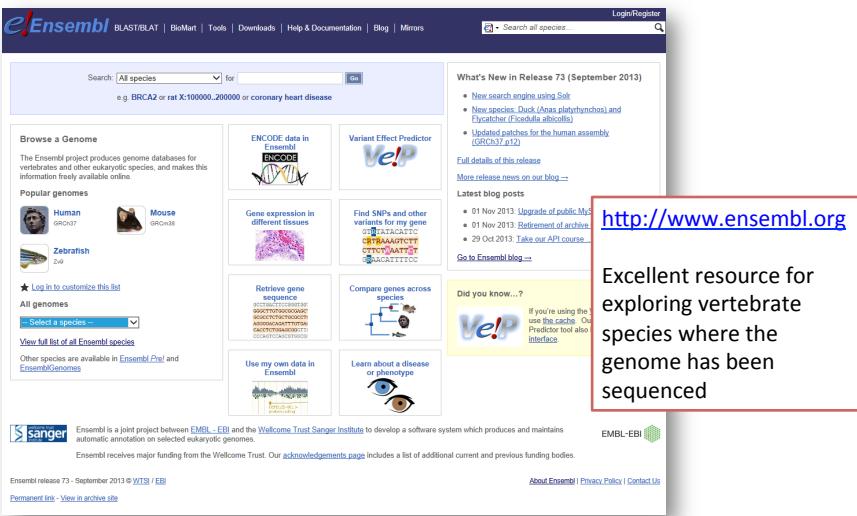
Ensembl Genome Browser

- Joint project between EMBL-EBI and the Wellcome Trust Sanger Institute
- Central resource for studying genomes of vertebrates
 - Mainly chordates, but some few extra (e.g. *C. elegans* and *S. cerevisiae*)
 - Updated several times a year with new genome assemblies and new species
 - Annotations of genomes (e.g. genes and their splice variant, SNPs) added by the Ensembl pipeline
 - Automatic gene prediction (with or without experimental evidence) & some curator input

 Oslo universitetssykehus UIO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

Ensembl Genome Browser



<http://www.ensembl.org>

Excellent resource for exploring vertebrate species where the genome has been sequenced

 Oslo universitetssykehus UIO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

Ensembl Genome Browser

Currently >70 species

Oslo

New

UIO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

EnsemblGenomes

EnsemblGenomes

About | Species | Working with communities | FAQs

Agricultural-Omics Training Course

Register now for an upcoming EBI training course in Agricultural-Omics. For more details, please go to <http://www.ebi.ac.uk/training/course/agricultural-omics>

posted 2013-10-03

What's New in Release 20 (September 2013)

The twentieth release of Ensembl Genomes features updates to version 73 of the Ensembl software across all divisions, and a number of new genomes have been added bringing the total number of genomes to 9295 ([full list](#)). Detailed notes can be found here. See the individual homepages for more information.

Ensembl Bacteria

Ensembl Bacteria has been updated to include the latest versions of 9,039 genomes (9,842 eubacteria and 247 archaea) from the INSDC databases. Cross-references to Rhea and MetaCyc have also been added, as have Enzyme Commission classifications. In addition, data from RegulonDB have been used to add operon and other regulatory features to *E. coli* K-12 MG1655.

Ensembl Fungi

Two new plant pathogen genomes, *Microbotryum violaceum* and *Blumeria graminis*. Cross-references to Phyllobase were added for plant pathogens.

Ensembl Metazoa

Three metazoan species have updated assemblies and gene models in release 20 of Ensembl Metazoa: the pea aphid, the western honey bee, and the purple sea urchin. The variation data for *Anopheles gambiae* has been updated to include ~7.5 million additional variants.

Ensembl Plants

The first assembly of the bread wheat genome, *Triticum aestivum*, from the IWGSC has been added in this release. In addition we have loaded the latest assembly for *Oryza sativa* from IRGSP.

- Bacteria, protists, fungi, plants and other metazoa

UIO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

Ensembl 2013

D48–D55 *Nucleic Acids Research*, 2013, Vol. 41, Database issue
doi:10.1093/nar/gks1236

Published online 30 November 2012

Ensembl 2013

Paul Flicek^{1,2,*}, Ikhlaq Ahmed¹, M. Ridwan Armode², Daniel Barrell², Kathryn Beal¹,
Simon Brent¹, Denise Carvalho-Silva¹, Peter Clapham², Guy Coates², Susan Fairley²,
Stephen Fitzgerald¹, Laurent Gil¹, Carlos Garcia-Giron², Leo Gordon¹, Thibaut Hourlier²,
Sarah Hunt¹, Thomas Juettemann¹, Andreas K. Kähäri², Stephen Keenan¹,
Monika Komorowska¹, Eugenia Kulesha¹, Ian Longden¹, Thoman Maurel¹,
William M. McLaren¹, Matthieu Muffato¹, Rishi Nag², Bert Overduin¹, Miguel Pignatelli¹,
Bethan Pritchard², Emily Pritchard¹, Harpreet Singh Riat², Graham R. S. Ritchie¹,
Magali Ruffier¹, Michael Schuster¹, Daniel Sheppard¹, Daniel Sobral¹, Kieron Taylor¹,
Anja Thormann¹, Stephen Trevanion¹, Simon White², Steven Wilder¹,
Bronwen L. Aken², Ewan Birney¹, Fiona Cunningham¹, Ian Dunham¹, Jennifer Harrow²,
Javier Herreno¹, Tim J. P. Hubbard², Nathan Johnson¹, Rhoda Kinsella¹, Anne Parker²,
Giulietta Spudich¹, Andy Yates¹, Amonida Zadissa² and Stephen M. J. Searle²

¹European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton Cambridge CB10 1SD, UK and
²Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Received October 11, 2012; Revised October 31, 2012; Accepted November 1, 2012

ABSTRACT
The Ensembl project (<http://www.ensembl.org>) provides genome information for sequenced chordate genomes with a particular focus on human, mouse, zebrafish and rat. Our resources include evidence-based annotations for all supported species, large-scale whole genome alignments, gene alignments across vertebrates and clade-specific alignments for eutherian mammals, primates, birds and fish; variation data resources for 17 species and regulation annotations based on ENCODE and other data sets. Ensembl data are accessible through the genome browser at <http://www.ensembl.org> and through other tools and programmatic interfaces.

Read the article yourself!

 Oslo universitetssykehus

UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

UCSC Genome Browser

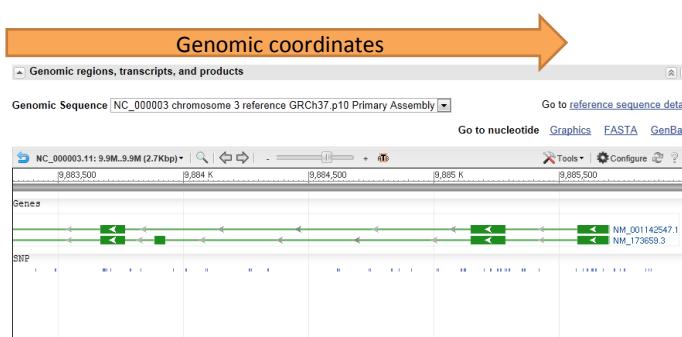
 Oslo universitetssykehus

UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

Genome browsers

- Graphical interface for genomic data
- Shows information from biological databases mapped onto genomic sequence



The screenshot shows a genome browser interface for chromosome 3. At the top, there's a header with 'Genomic regions, transcripts, and products' and a search bar. Below the header, the genome sequence is shown with coordinates from 8,883,500 to 8,895,500. The main area displays 'Genes' and 'SNP' tracks. An orange arrow points from the text 'Genomic coordinates' to the top of the browser window. Another orange arrow points from the text 'Various annotations = "tracks"' to the bottom of the browser window.

Genomic coordinates →

Various annotations = "tracks"

Genomic regions, transcripts, and products

Genomic Sequence | NC_000003 chromosome 3 reference GRCh37.p10 Primary Assembly

Go to reference sequence details

Go to nucleotide Graphics FASTA GenBank

NC_000003.11: 9.9M..9.9M (2.7Kbp)

8,883,500 8,884 K 8,884,500 8,885 K 8,885,500

Genes

SNP

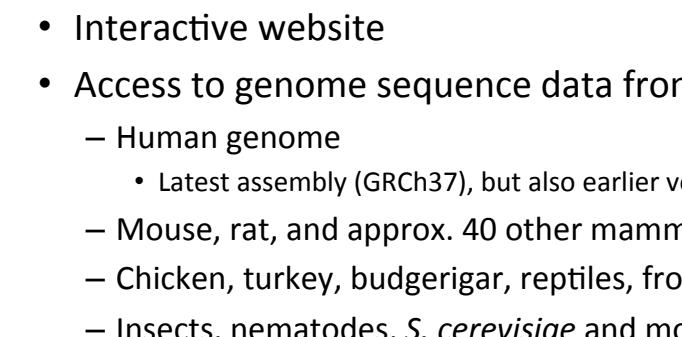
NCBI Gene database UiO : Department of Informatics
University of Oslo

Oslo universitetssykehus

Jon K. Lærdahl,
Structural Bioinformatics

UCSC Genome Browser

- Developed and maintained at the University of California, Santa Cruz (UCSC)
- Interactive website
- Access to genome sequence data from
 - Human genome
 - Latest assembly (GRCh37), but also earlier versions
 - Mouse, rat, and approx. 40 other mammals
 - Chicken, turkey, budgerigar, reptiles, frogs, and fishes
 - Insects, nematodes, *S. cerevisiae* and more



The screenshot shows the UCSC Genome Browser interface. It features a search bar at the top and a main panel displaying genomic tracks for a specific region. The tracks include genes and SNPs. The UI includes navigation buttons like 'Zoom In' and 'Zoom Out'. An orange arrow points from the text 'Various annotations = "tracks"' to the bottom of the browser window.

Various annotations = "tracks"

Genomic regions, transcripts, and products

Genomic Sequence | NC_000003 chromosome 3 reference GRCh37.p10 Primary Assembly

Go to reference sequence details

Go to nucleotide Graphics FASTA GenBank

NC_000003.11: 9.9M..9.9M (2.7Kbp)

8,883,500 8,884 K 8,884,500 8,885 K 8,885,500

Genes

SNP

NCBI Gene database UiO : Department of Informatics
University of Oslo

Oslo universitetssykehus

Jon K. Lærdahl,
Structural Bioinformatics

UCSC Genome Browser

BRIEFINGS IN BIOINFORMATICS VOL. 14, NO. 2, 144–161
Advance Access published on 29 August 2012
doi:10.1093/bib/bbs038

The UCSC genome browser and associated tools

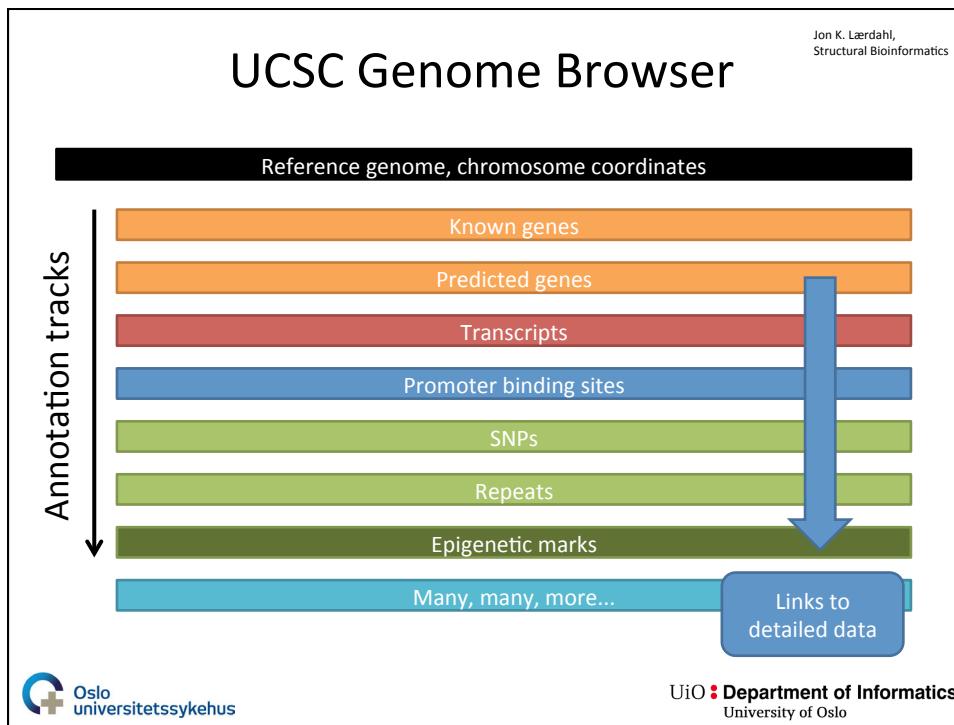
Robert M. Kuhn, David Haussler and W. James Kent
Submitted 8th February 2012; Received (in revised form) : 9th June 2012

<http://genome.ucsc.edu>

Abstract
 The UCSC Genome Browser (<http://genome.ucsc.edu>) is a graphical viewer for genomic data now in its 13th year. Since the early days of the Human Genome Project, it has presented an integrated view of genomic data of many kinds. Now home to assemblies for 58 organisms, the Browser presents visualization of annotations mapped to genomic coordinates. The ability to juxtapose annotations of many types facilitates inquiry-driven data mining. Gene predictions, mRNA alignments, epigenomic data from the ENCODE project, conservation scores from vertebrate whole-genome alignments and variation data may be viewed at any scale from a single base to an entire chromosome. The Browser also includes many other widely used tools, including BLAT, which is useful for alignments from high-throughput sequencing experiments. Private data uploaded as Custom Tracks and Data Hubs in many formats may be displayed alongside the rich compendium of precomputed data in the UCSC database. The Table Browser is a full-featured graphical interface, which allows querying, filtering and intersection of data tables. The Saved Session feature allows users to store and share customized views, enhancing the utility of the system for organizing multiple trains of thought. Binary Alignment/Map (BAM), Variant Call Format and the Personal Genome Single Nucleotide Polymorphisms (SNPs) data formats are useful for visualizing a large sequencing experiment (whole-genome or whole-exome), where the differences between the data set and the reference assembly may be displayed graphically. Support for high-throughput sequencing extends to compact, indexed data formats, such as BAM, bigBed and bigWig, allowing rapid visualization of large datasets from RNA-seq and ChIP-seq experiments via local hosting.

Kuhn *et al.* Brief. Bioinform. **14**, 144 (2012)

UiO Department of Informatics
University of Oslo



Jon K. Lærdahl,
Structural Bioinformatics

UCSC Genome Browser

Cow, chromosome 22, around position 17,380,000

Annotation tracks

Uo : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

UCSC Genome Browser

<http://genome.ucsc.edu>

Access to the databases and tools

Start here

General information

News, updates, announcements

Uo : Department of Informatics
University of Oslo

UCSC Genome Browser

Jon K. Lærdahl,
Structural Bioinformatics

Human (Homo sapiens) Genome Browser Gateway

The UCSC Genome Browser was created by the Genome Biostatistics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade: Mammal genome: Human assembly: hg19 position: chr21 33,031,597-33,041,570 search term:

[Click here to reset](#) the browser user interface settings to their defaults.
[track search](#) | [add custom tracks](#) | [track hubs](#) | [configure tracks and display](#)

Human Genome Browser – hg19 assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

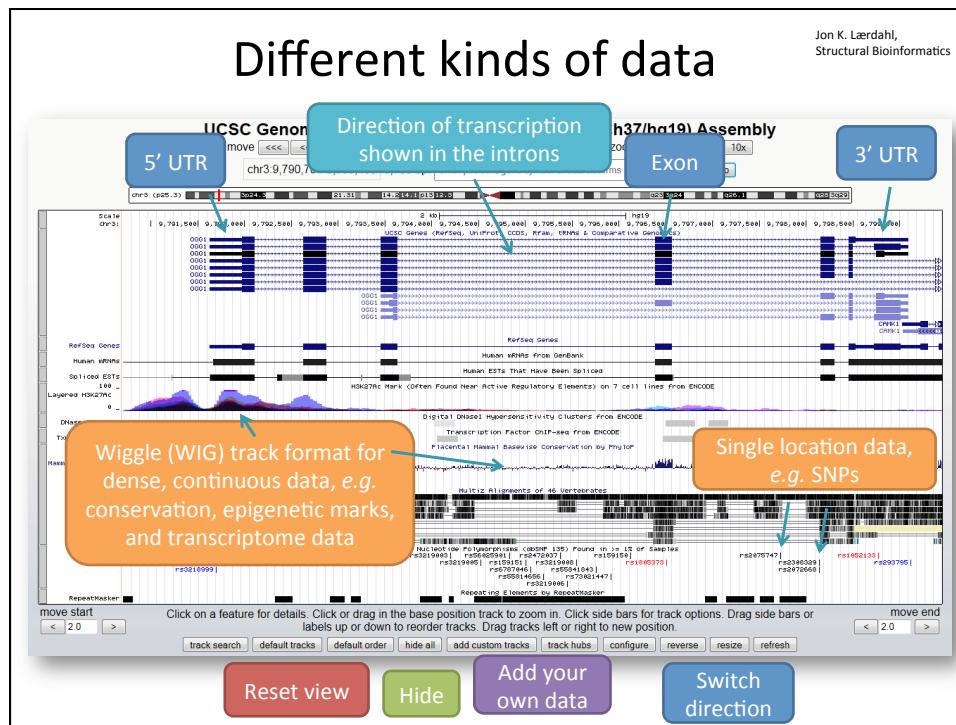
Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following lists shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chr1Un_g1000212	Displays all of the unplaced contig g1000212
20p13	Displays region for band p13 on chr 20
chr2 1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3 1000000-2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061 RH180175 15g11,15q13	Displays region between genome landmarks, such as the STS markers RH18061 and RH180175, or chromosome bands 15g11 to 15q13, or SNPs rs1042522 and rs180370. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.
rs1042522,rs180370	
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marsfield maps. Includes 100 kb upstream and downstream of the marker.
AA205474	Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC009101	Displays region of clone with GenBank accession AC009101
AF083811	Displays region of mRNA with GenBank accession number AF083811
PRNP	Displays region of gene with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414	Displays the region of genome with RefSeq identifier NM_017414
NP_059110	Displays the region of genome with protein accession number NP_059110

Examples of searching options – correct query format

UiO Department of Informatics
University of Oslo



Published online 30 October 2010
Nucleic Acids Research, 2011, Vol. 39, Database issue D871–D875
doi:10.1093/nar/gkq1017

ENCODE data in UCSC

ENCODE whole-genome data in the UCSC genome browser (2011 update)

Brian J. Raney^{1,*}, Melissa S. Cline¹, Kate R. Rosenbloom¹, Timothy R. Dreszer¹, Katrina Learned¹, Galt P. Barber¹, Laurence R. Meyer¹, Cricket A. Sloan¹, Venkat S. Malladi¹, Krishna M. Roskin¹, Bernard B. Suh¹, Angie S. Hinrichs¹, Hiram Clawson¹, Ann S. Zweig¹, Vanessa Kirkup¹, Pauline A. Fujita¹, Brooke Rhead¹, Kayla E. Smith¹, Andy Pohl¹, Robert M. Kuhn¹, Donna Karolchik¹, David Haussler^{1,2} and W. James Kent¹

¹Center for Biomolecular Science and Engineering, School of Engineering and ²Howard Hughes Medical Institute, University of California Santa Cruz (UCSC), Santa Cruz, CA 95064, USA

Received September 15, 2010; Accepted October 9, 2010

ABSTRACT
The ENCODE project is an international consortium with a goal of cataloguing all the functional elements in the human genome. The ENCODE Data Coordination Center (DCC) at the University of California, Santa Cruz serves as the central repository for ENCODE data. Recent data include a collection of high-throughput genome-wide data generated with technologies such as ChIP-Seq, RNA-Seq, DNA digestion and others. This data helps illuminate transcription factor-binding sites, histone marks, chromatin accessibility, DNA methylation, nucleosome positioning, RNA binding sites and cell-state indicators. It includes sequences with quality scores, alignments, signals calculated from the alignments, and in most cases, element or peak calls calculated from the signal data. Each data set is available for visualization and download via the UCSC Genome Browser. In addition, the ENCODE data can also be retrieved using a metadata system that captures the experimental parameters of each assay. The ENCODE web portal at UCSC (<http://encodeproject.org/>) provides information about the ENCODE data and links for access.

The goal of the ENCODE project is to identify all functional elements in the genome. This involves generating high-throughput genome-wide data using various technologies, such as ChIP-Seq, RNA-Seq, and DNase-seq. These data help to map transcription factor-binding sites, histone marks, chromatin accessibility, DNA methylation, nucleosome positioning, and RNA binding sites. The data is used to create a metadata system that captures experimental parameters for each assay. The ENCODE web portal at UCSC (<http://encodeproject.org/>) provides information about the ENCODE data and links for access.

Figure: A diagram illustrating the ENCODE pipeline. It shows a genomic DNA strand with various regulatory elements: Hypersensitive Sites, SC (super-enhancers), DNase-seq/FAR-seq, ChIP-seq, Computational predictions and RT-PCR, and RNA-seq. These elements are mapped onto a gene structure with a promoter, transcription factor binding sites, and a transcript. Below the gene, a legend identifies Long-range regulatory elements (enhancers, repressors, silencers, insulators) and cis-regulatory elements (promoters, transcription factor binding sites).

<http://genome.ucsc.edu/ENCODE/aboutScaleup.html>

Oslo universitetssykehus

UiO : Department of Informatics
University of Oslo