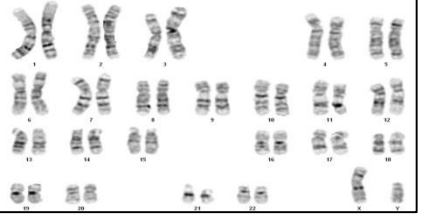
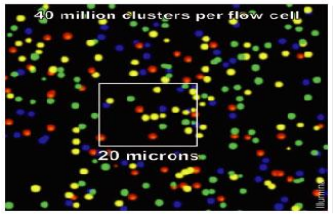
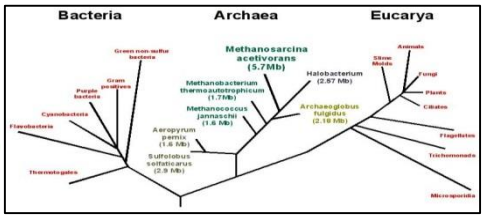




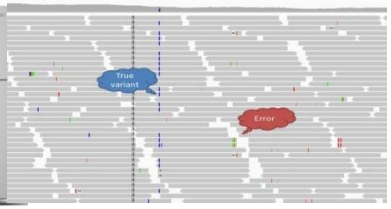


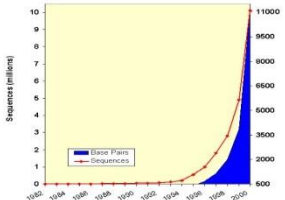
Bioinformatic Resources

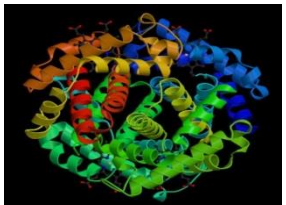
北京大学生物信息学中心 魏丽萍

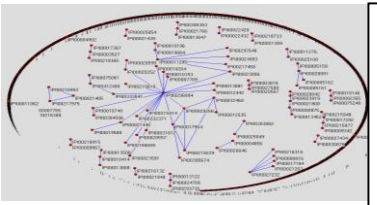
Liping Wei, Ph.D.

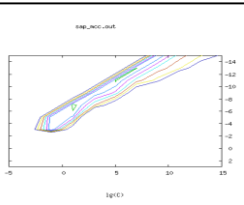
Center for Bioinformatics, Peking University

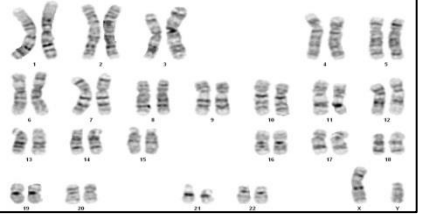


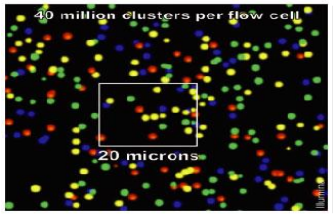


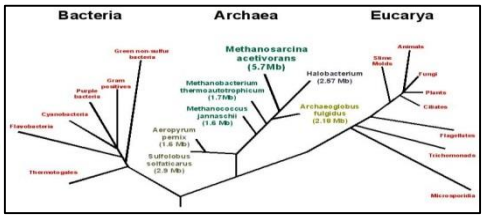










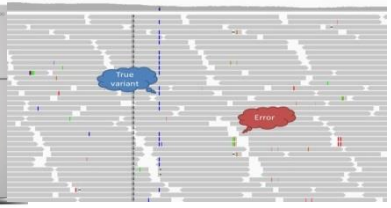


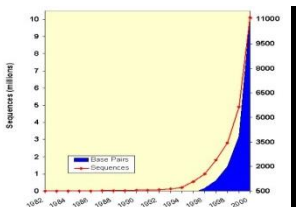
Unit 5:
Individual Resources

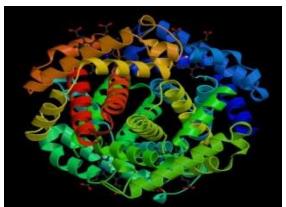
北京大学生物信息学中心 魏丽萍

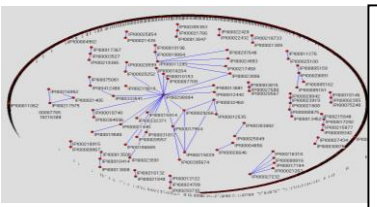
Liping Wei, Ph.D.

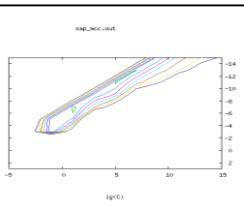
Center for Bioinformatics, Peking University



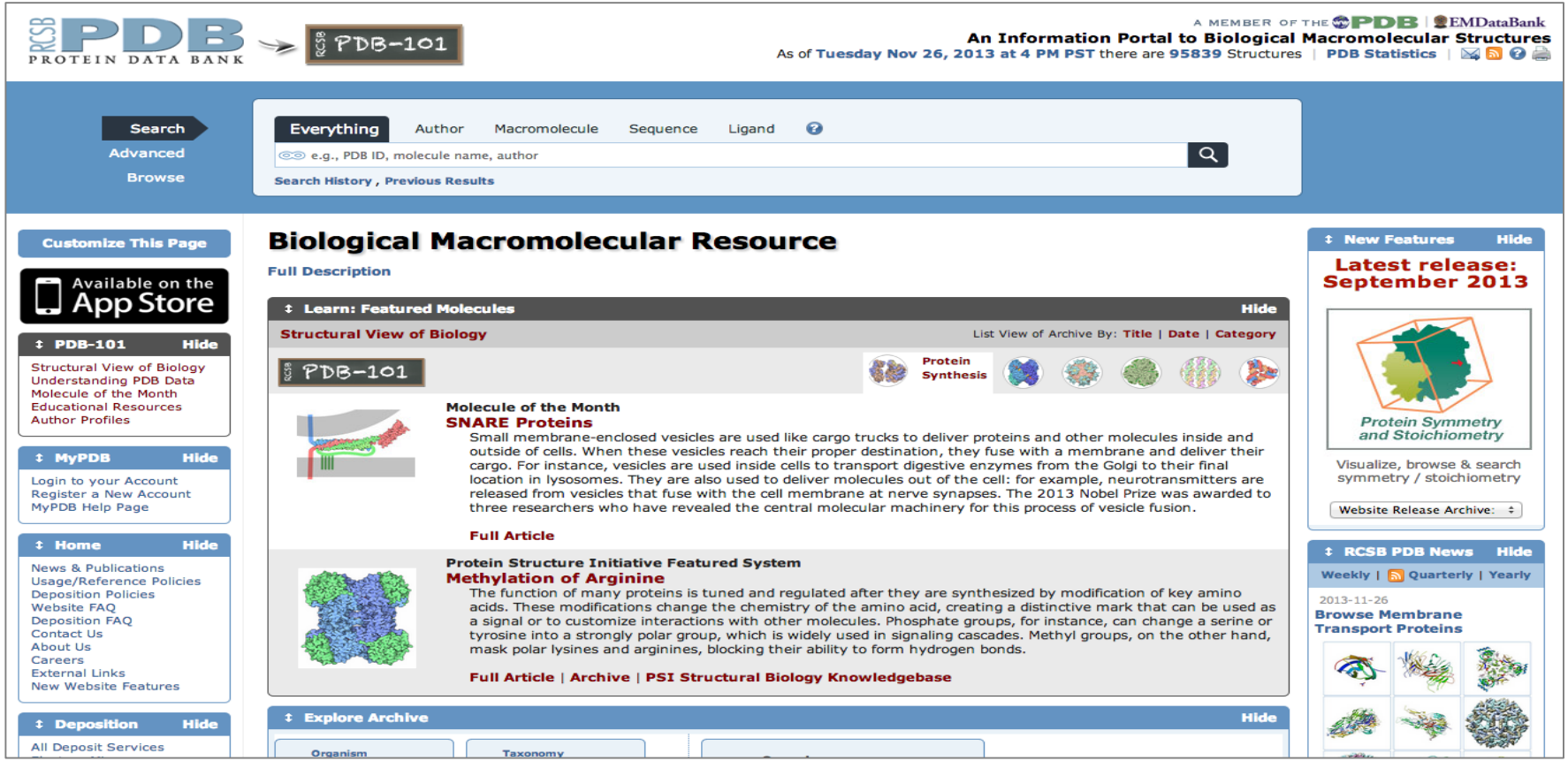








PDB (http://www.rcsb.org/)



Screenshot made at http://www.rcsb.org/pdb/home/home.do/ on December 1st, 2013 (UTC+0800)

PDB statistics (December 2013)

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	79083	1496	4125	4	84708
NMR	8937	1054	197	7	10195
ELECTRON MICROSCOPY	489	51	162	0	702
HYBRID	53	3	2	1	59
other	152	4	6	13	175
Total	88714	2608	4492	25	95839

Data collected from http://www.rcsb.org/pdb/statistics/holdings.do on December 1st, 2013 (UTC+0800)

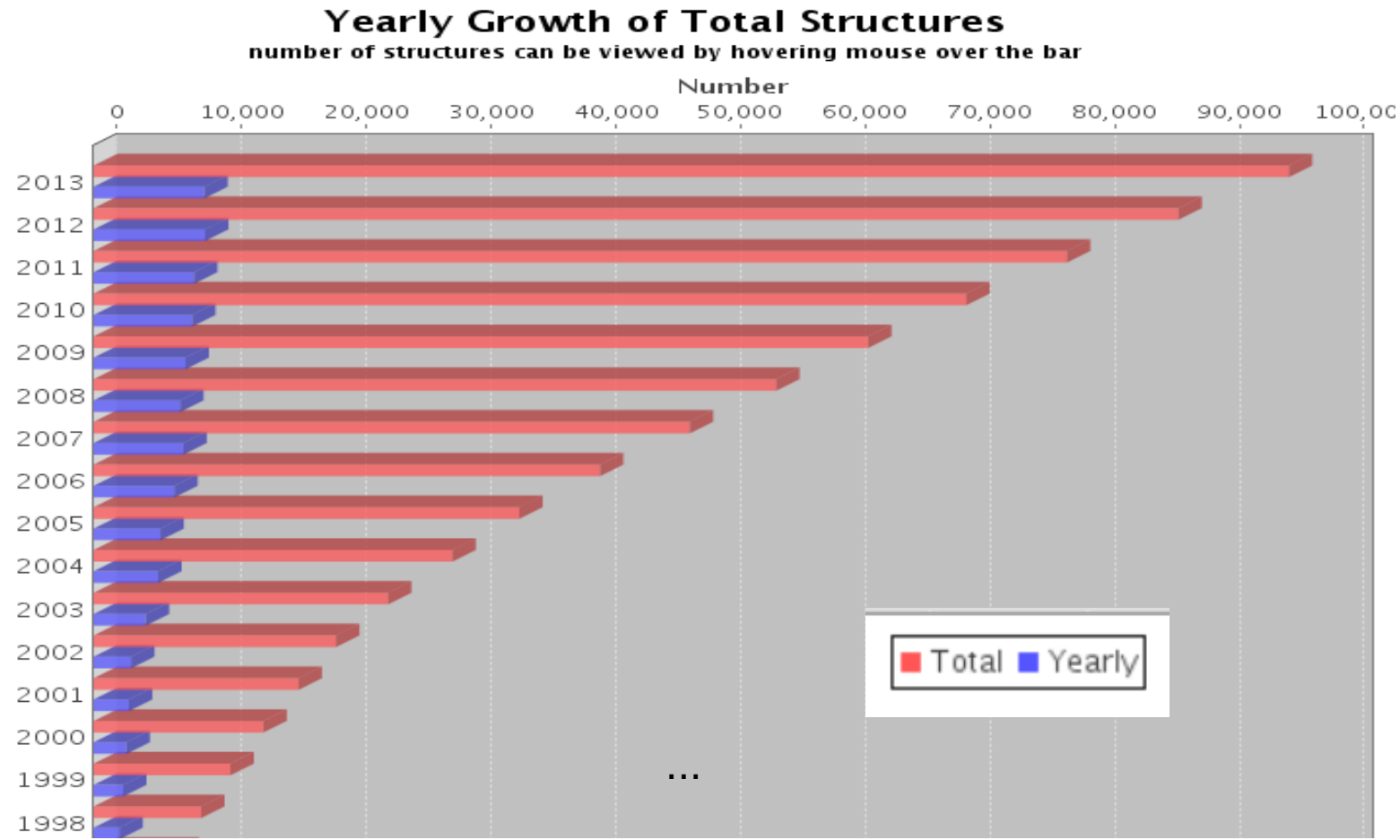
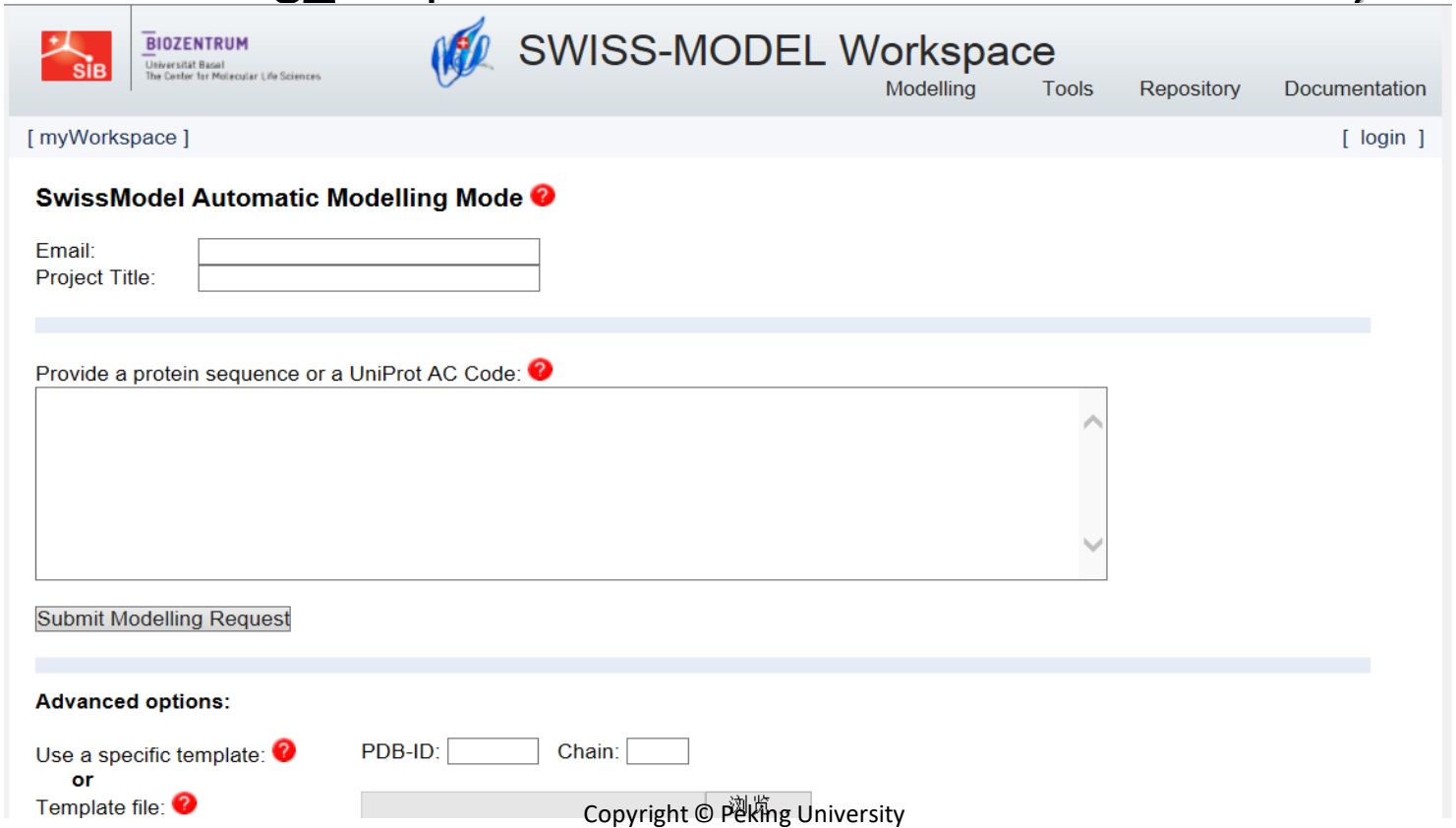


Image collected from http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100

SWISS-MODEL Workspace

(http://swissmodel.expasy.org/workspace/index.php?func=modelling_simple1&userid=USERID&token=TOKEN)



SWISS-MODEL Repository

(http://swissmodel.expasy.org/repository/?pid=smr01&zid=async)

BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences

SWISS-MODEL Repository

Modelling Tools Repository Documentation

[Repository Query] [Full Text Query]

Welcome to the SWISS-MODEL Repository

The SWISS-MODEL Repository is a database of annotated three-dimensional comparative protein structure models generated by the fully automated homology-modelling pipeline SWISS-MODEL.

Example Queries:
[P23298] [GLDA_ECOLI] [IPI00743503] [NP_416402] [GI:28872740] [ENTREZ:54401] [Sequence]

SEARCH

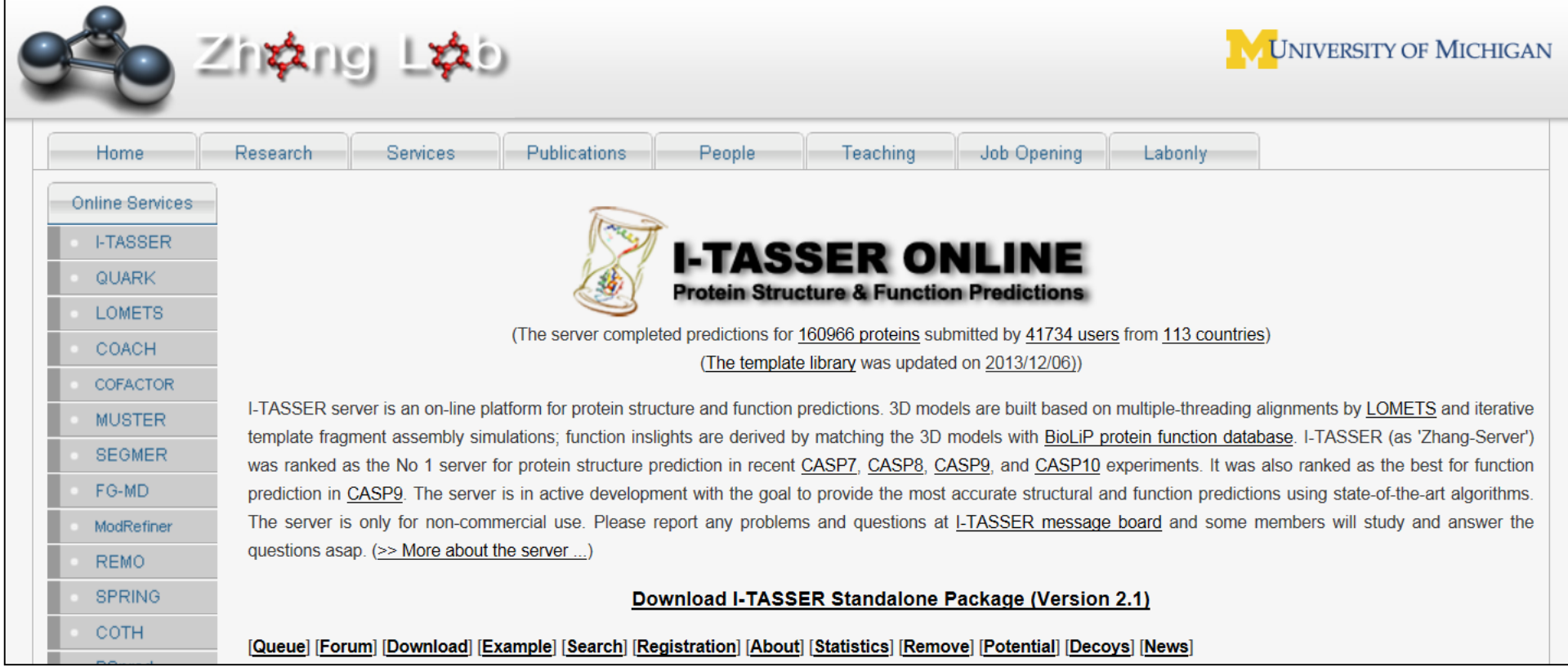
List of “model organisms” regularly updated in Repository

The proteomes of the following organisms are regularly updated in SWISS-MODEL Repository. A reference proteome is the complete proteome of a representative, well-studied model organism or an organism of interest for biomedical research. **Reference proteomes** are retrieved from the UniProt database.

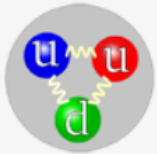
Organism	Taxonomy	Number of models
Human	**Homo Sapiens**	144435
Mouse	**Mus musculus**	129327
Caenorhabditis elegans	**C. elegans**	49205
Escherichia coli (strain K12)	**E.coli K12**	6814
Mouse-ear cress	**A. thaliana**	85066
Fruit Fly	**D. melanogaster**	53483
Bakers Yeast	**S. cerevisiae**	11199
Caulobacter crescentus	**C. crescentus**	6394
Mycobacterium tuberculosis	**M. tuberculosis**	7913
Pseudomonas aeruginosa	**P. aeruginosa**	9964
Staphylococcus aureus	**S. aureus**	3755
Plasmodium falciparum	**P. falciparum**	8109

Statistics based on UniProt release **2013_11**

I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/)

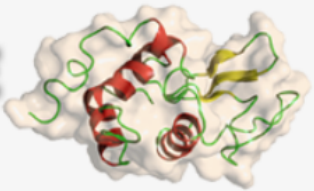


QUARK (http://zhanglab.ccmb.med.umich.edu/QUARK/)



QUARK ONLINE

Ab Initio Protein Structure Prediction



QUARK is a computer algorithm for ab initio protein folding and protein structure prediction, which aims to construct the correct protein 3D model from amino acid sequence only. QUARK models are built from small fragments (1-20 residues long) by replica-exchange Monte Carlo simulation under the guide of an atomic-level knowledge-based force field. QUARK was ranked as the No 1 server in Free-modeling (FM) in CASP9 and CASP10 experiments. Since no global template information is used in QUARK simulation, the server is suitable for proteins which are considered without homologous templates.

Go to Job Q24927 to view an example of QUARK output. The description of predicted feature files can be seen in readme.txt. Questions about the QUARK server can be posted at the Service System Discussion Board.

Cut and paste your sequence (in FASTA format, less than 200 AA. Please submit bigger proteins to I-TASSER Server):

Or upload the sequence from your local computer:

浏览...

Email: (mandatory, where results will be sent to. Only academic email accounts are acceptable.)

ID: (optional, your given name of the protein)

Optional: You can assign additional distance restraints for modeling (example is in distrestraint.txt; format is described in readme.txt).

Protein Model Portal (http://www.proteinmodelportal.org/)

ModWeb

Server Policy: By checking this box, I assert that I am part of an academic institution (not a government research lab such a
Modeller license .

I have a MODELLER access key:

M4T

Server Policy: I am a non-profit/academic user and this server will be used solely for educational purposes or for basic rese

SWISS-MODEL

Server Policy: Usage of SWISS-MODEL Server and Workspace are free of charge.

I-TASSER

Server Policy: Usage of I-TASSER is free of charge.
However, there is a limitation of one job per email address and only academic email addresses are allowed.

HHpred

Server Policy: Usage of HHpred is free of charge for academic use.

Phyre2

Server Policy: Usage of Phyre2 is restricted to academic users.

CASP (http://www.predictioncenter.org/)

Protein Structure Prediction Center

Menu

- Home
- FORCASP Forum
- PC Login
- PC Registration
- CASP Experiments
  - CASP ROLL
    - Home
    - My CASP ROLL profile
    - Targets
      - Target List
      - Target Submission
  - CASP11 (2014)
  - CASP10 (2012)
  - CASP9 (2010)
  - CASP8 (2008)
  - CASP7 (2006)
  - CASP6 (2004)
  - CASP5 (2002)
  - CASP4 (2000)
  - CASP3 (1998)

Welcome to the Protein Structure Prediction Center!

Our goal is to help advance the methods of identifying protein structure from sequence. The Center has been organized to provide the means of objective testing of these methods via the process of blind prediction. The Critical Assessment of protein Structure Prediction (CASP) experiments aim at establishing the current state of the art in protein structure prediction, identifying what progress has been made, and highlighting where future effort may be most productively focused.

There have been ten previous CASP experiments. The eleventh experiment will start in May 2014. Description of these experiments and the full data (targets, predictions, interactive tables with numerical evaluation results, dynamic graphs and prediction visualization tools) can be accessed following the links:

CASP1 (1994) | **CASP2 (1996)** | **CASP3 (1998)** | **CASP4 (2000)** | **CASP5 (2002)** | **CASP6 (2004)** | **CASP7 (2006)** | **CASP8 (2008)** | **CASP9 (2010)** | **CASP10 (2012)** | **CASP11 (2014)**

Raw data for the experiments held so far are archived and stored in our data archive.

In November 2011 we have opened a new rolling CASP experiment for all-year-round testing of ab initio modeling methods:

CASP ROLL

Details of the experiments have been published in a scientific journal *Proteins: Structure, Function and Bioinformatics*. **CASP proceedings** include papers describing the structure and conduct of the experiments, the numerical evaluation measures, reports from the assessment teams highlighting state of the art in different prediction categories, methods from some of the most successful prediction teams, and progress in various aspects of the modeling.

Prediction methods are assessed on the basis of the analysis of a large number of blind predictions of protein structure. Summary of numerical evaluation of the methods tested in the latest CASP experiment can be found on this web page. The main numerical measures used in evaluations are described in the papers [1], [2]. The latter paper also contains explanations of data handling procedures and guidelines for navigating the data presented on this website.

Message Board

Resuming CASP ROLL

Dear CASPers, Best regards for all of you in the New Year! Hoping that you had good rest after the CASP10 experiment and meeting, we are resuming CASP ROLL with two new targets later this week. ...

Predictors meeting in Gaeta

Dear CASP10 Participants, On the last day of the Meeting we will have our regular Predictors get-together. In advance, I would like to ask you to send in any comments regarding the CASP process in ...

Release of CASP10 results

Dear CASP10 Predictors, We have released results of the CASP10 and CASP ROLL experiments. You can check now interactive results tables

CASP10: Top servers for template-based modeling

035s	Zhang Server
114s	QUARK
108s	PMS
370s	HHpred-thread
122s	RaptorX-ZY
330s	BAKER-ROSETTASERVER
430s	HHpredA
223s	HHPredAO
486s	RaptorX
424s	MULTICOM-NOVEL

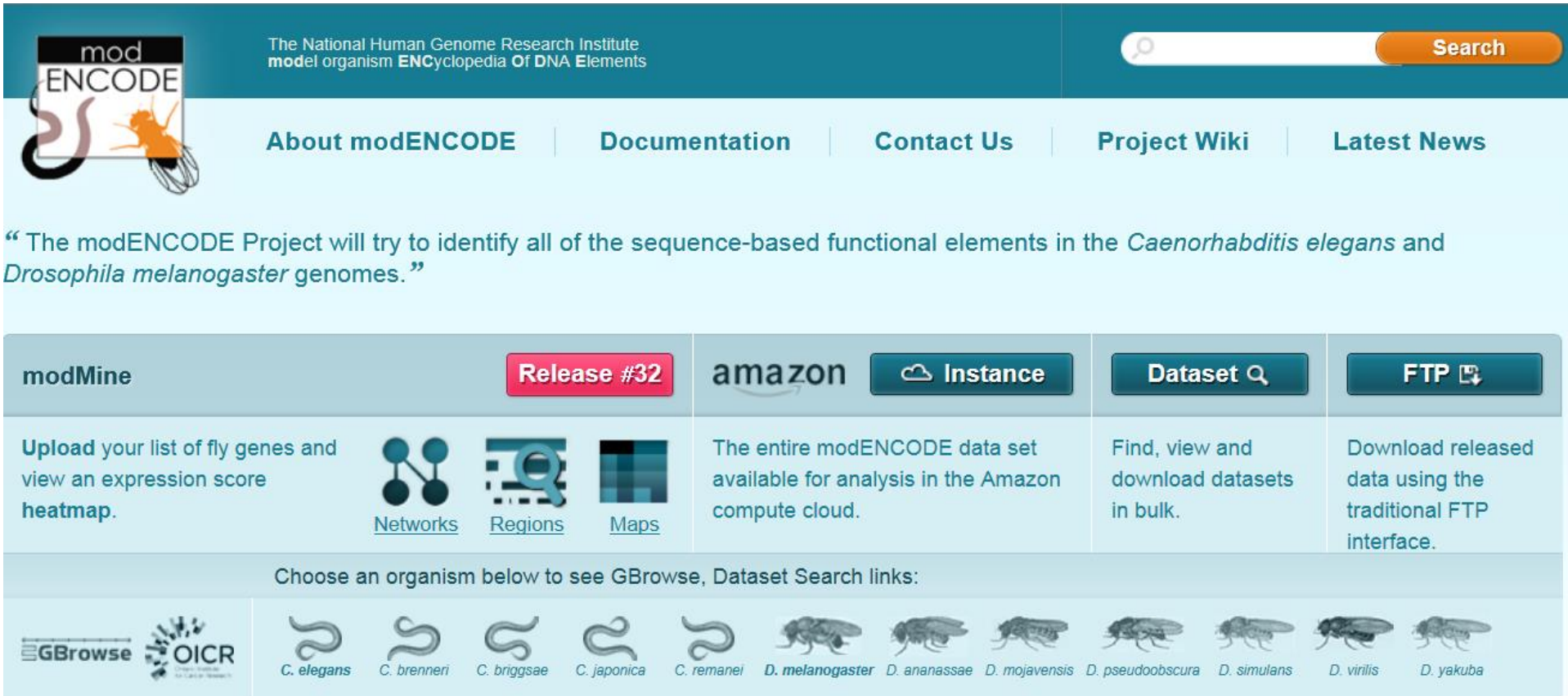
CASP10: Top servers for free modeling

Group name	Group ID	#Top Hits	Target: T0651-D0	T0653-D1	T0658-D1	T0663-D0	T0666-D1	T0684-D2	T0690-D0	T0693-D1	T0695-D1	T0713-D0	T0719-D6	T0726-D3	T0734-D1	T0735-D2	T0737-D1	T0739-D1	T0739-D2	T0740-D1	T0741-D1
keasar	315	4	1	1	1	1	1	1	1	2			2	2	1	2	1	1	1	1	1
QUARK	114 s	3	2	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1
Pcons	267	3	2	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1
Zhang-Server	035 s	2	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1
Zhang_Ab_Initio	45	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1
TASSER	79	2	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1
Pcomb	130	2	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1
TsaiLab	201	2	1	2	1	1	1	1	1	1	2	1	1	1		1	1			1	1
ProQ2	388	2	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1
Mufold2	405	2	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
PconsQ	428	2	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1
CNIO	475	2	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1

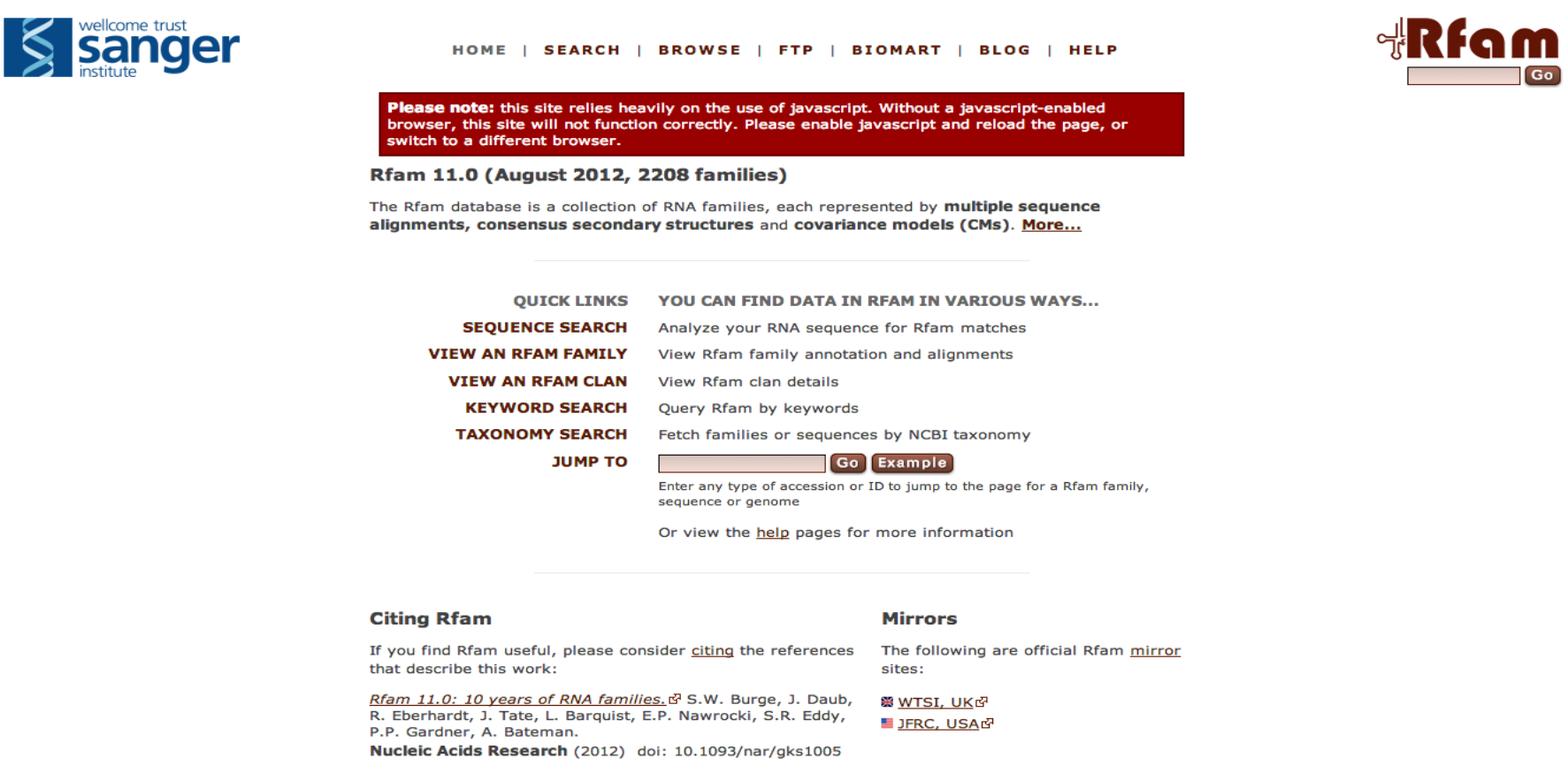
- Group name and ID in blue are servers.
- Target name in gray are their best models from one of the servers.

best model
submitted model
no submission

modENCODE (http://www.modencode.org/)



Rfam (http://rfam.sanger.ac.uk)



Screenshot made at http://rfam.sanger.ac.uk/ on December 1st, 2013 (UTC+0800)

PlantTFDB (http://planttfdb.cbi.pku.edu.cn/)

Plant Transcription Factor Database

v3.0

Center for Bioinformatics, Peking University, China Previous versions:v1.0 v2.0

Home | Blast | Search | Download | Prediction | Help | About | Links Search (eg: LFY)

Browse by Species

open all | close all

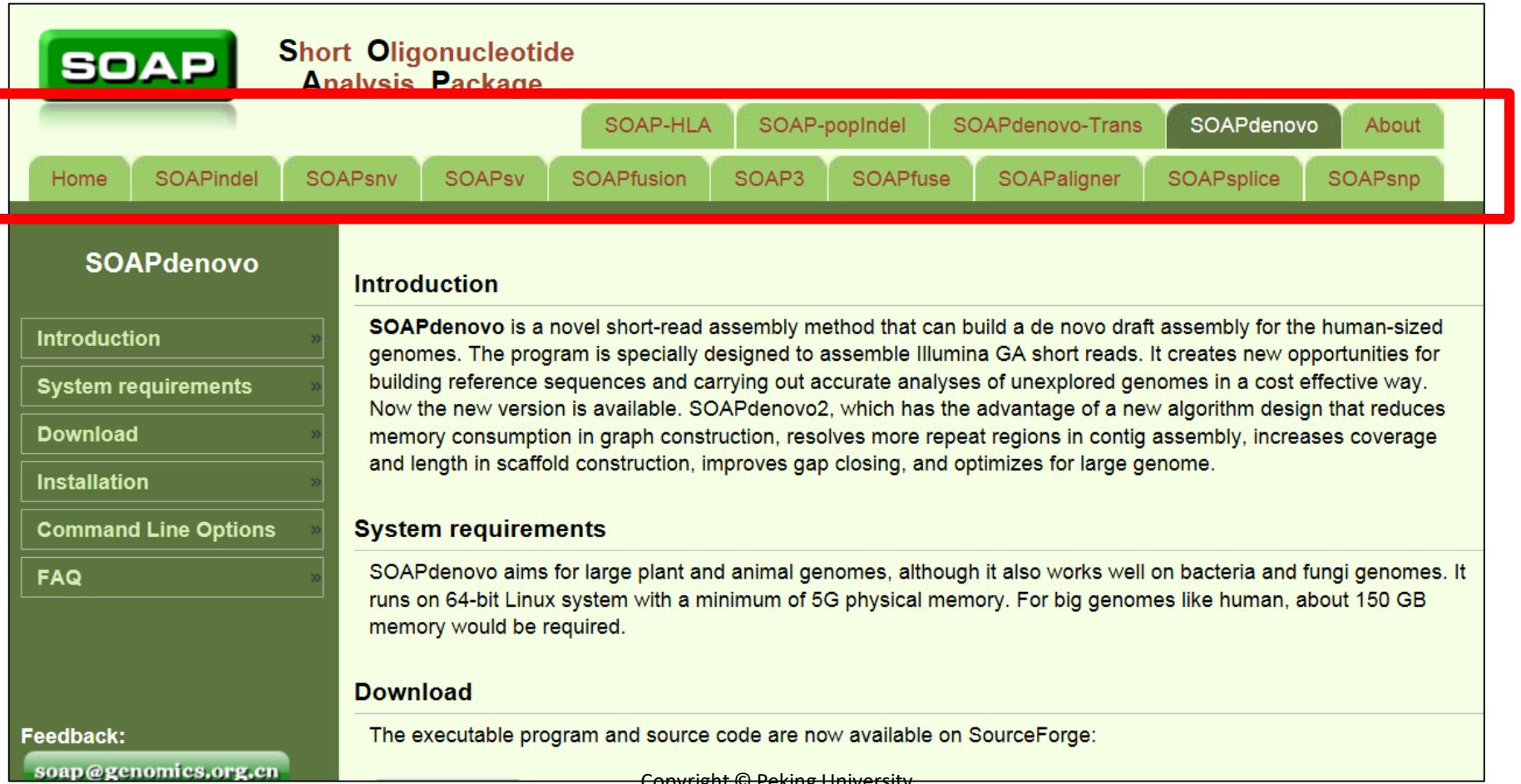
Taxonomic Group (83 species) (G)-species with genome sequence

- Chlorophyta (10 species)
- Bryophyta (1 species)
- Lycopodiophyta (1 species)
- Coniferopsida (4 species)
- Basal Magnoliophyta (1 species)
- Monocot (17 species)
- Eudicot (49 species)

Browse by Family

AP2 (1776)	ARF (1914)	ARR-B (914)	B3 (4051)	BBR-BPC (492)	BES1 (651)

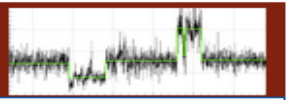
SOAPdenovo (http://soap.genomics.org.cn/soapdenovo.html)



CNVnator (http://sv.gersteinlab.org/cnvnator/)

Genome Structural Variations

portal for genome SV research in the Gerstein Lab



Structural Variations (SVs) and Copy Number Variations (CNVs) are a major source of genomic variation. However, compared to SNPs, accurate detection, genotyping and understanding of CNVs is lagging behind due to much greater analytical challenges related to SV/CNV detection and analysis. In our lab we analyse SVs/CNVs using high-throughput sequencing and different analytical approaches. Related tools, databases and publications are listed below.

TOOLS

		PAPERS
vcf2diploid	personal genome constructor, it can be used to construct a personal diploid genome sequence by including personal variants into reference genome.	2011
CNVnator	a tool for CNV discovery and genotyping from depth of read mapping.	2011a,2011b
AGE	a tools that implements an algorithm for optimal alignment of sequences with SVs.	2011
BreakSeq	a pipeline for annotation, classification and analysis of SVs at single nucleotide resolution.	2010
PEMer	a computational and simulation framework for discovering SVs by paired-end read mapping.	2009,2007

DATABASES AND DATASETS

		PAPERS
BreakSeq	The database contains information about breakpoints of SVs at single nucleotide level. The information has been gathered from literature.	2010
Break-DB	The database contains information about SVs and associated breakpoints detected by PEMer.	2009

WEB SUPPLEMENTS

Supplement to **Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library.** Nat Biotechnol. 2010 Jan;28(1):47-55. Epub 2009 Dec 27.

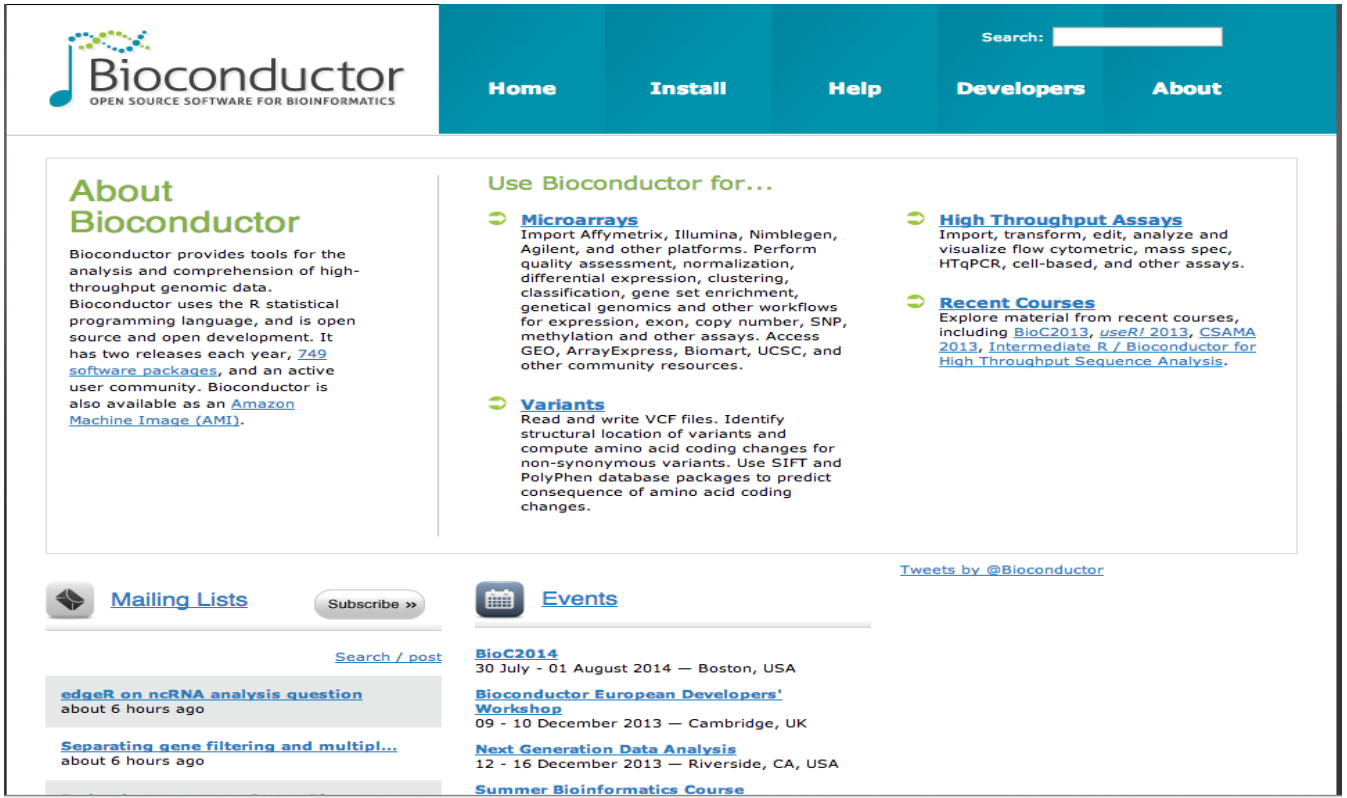
Supplement to **Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome.** Science. 2007 Oct 19;318(5849):420-6. Epub 2007 Sep 27.

PAPERS

Click here for a complete list of SV-related papers published in our group. Individual references to some of these have also been provided above.

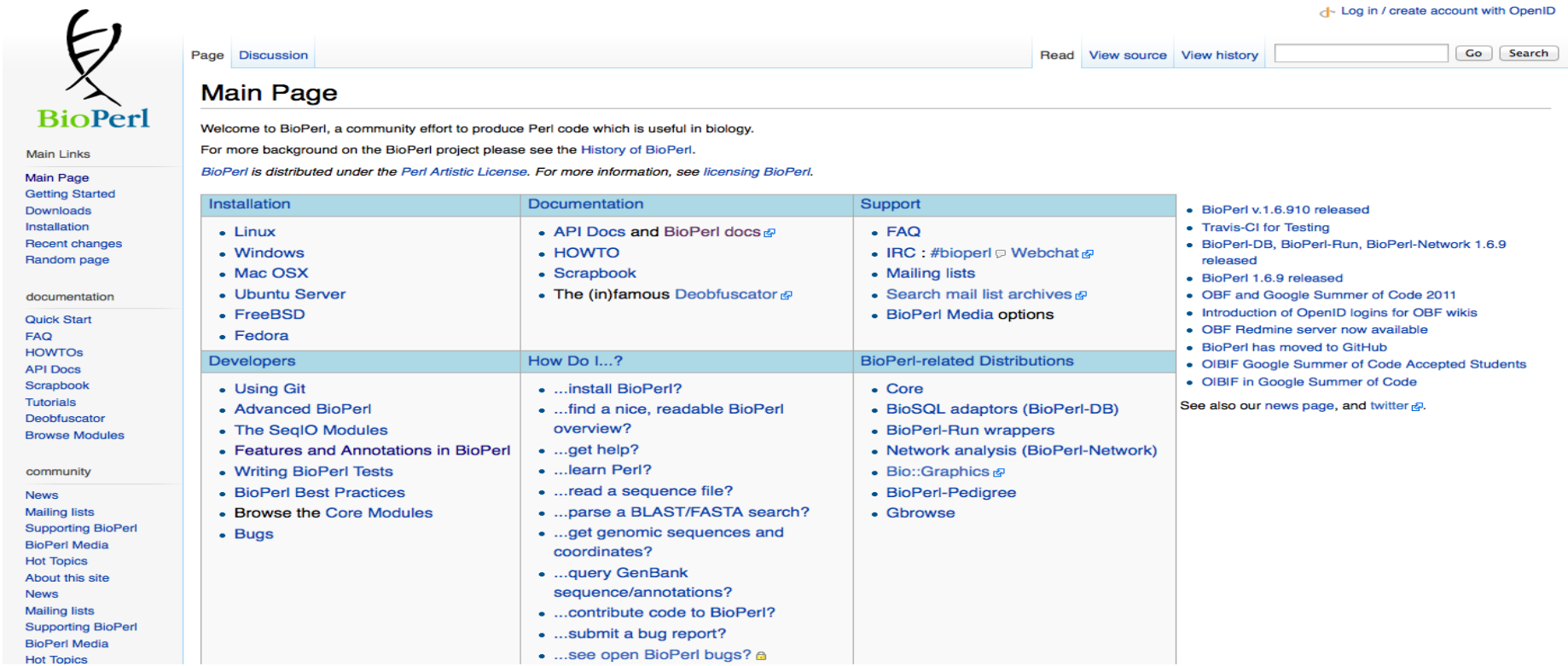
Screenshot made at http://sv.gersteinlab.org/ (upper figure) on December 1st, 2013 (UTC+0800)

Bioconductor (http://bioconductor.org/)



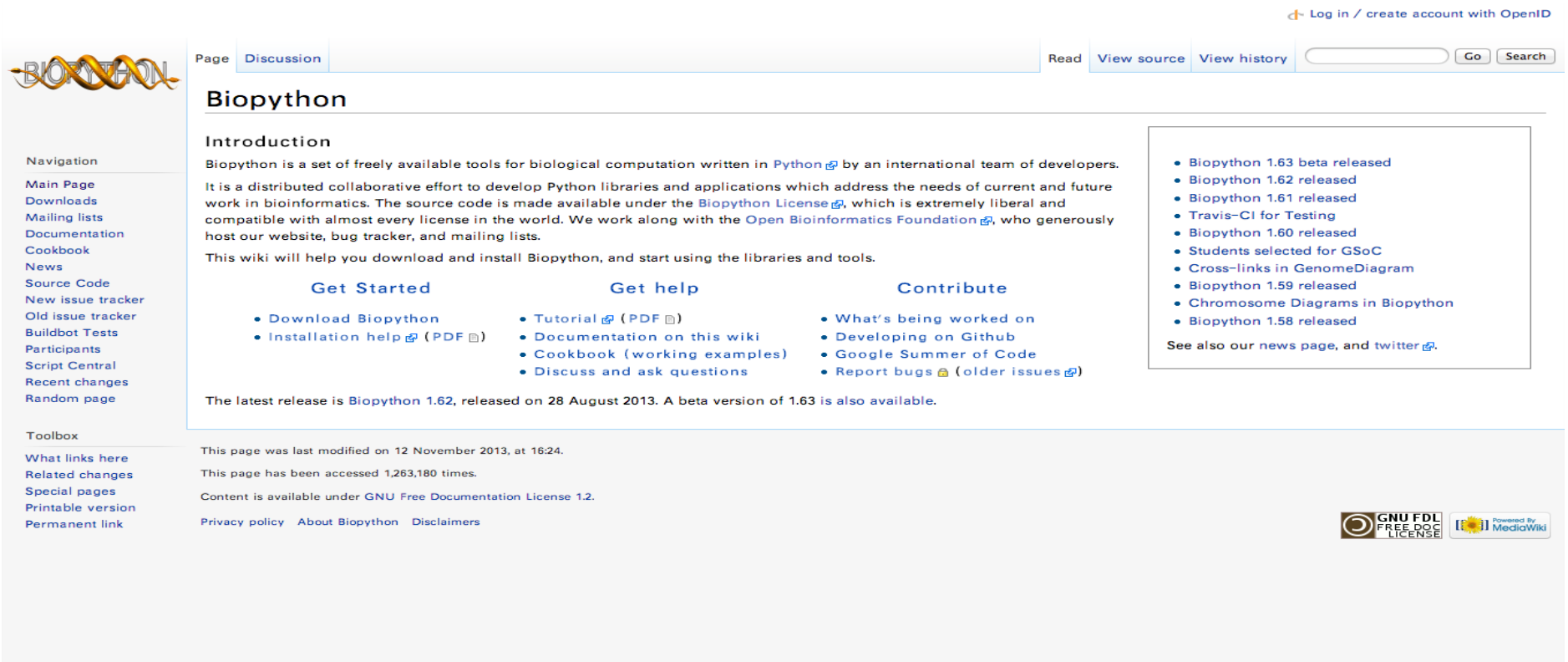
Screenshot made at http://bioconductor.org/ on December 1st, 2013 (UTC+0800)

BioPerl (http://bioperl.org/)



Screenshot made at http://bioperl.org/wiki/Main_Page on December 1st, 2013 (UTC+0800)

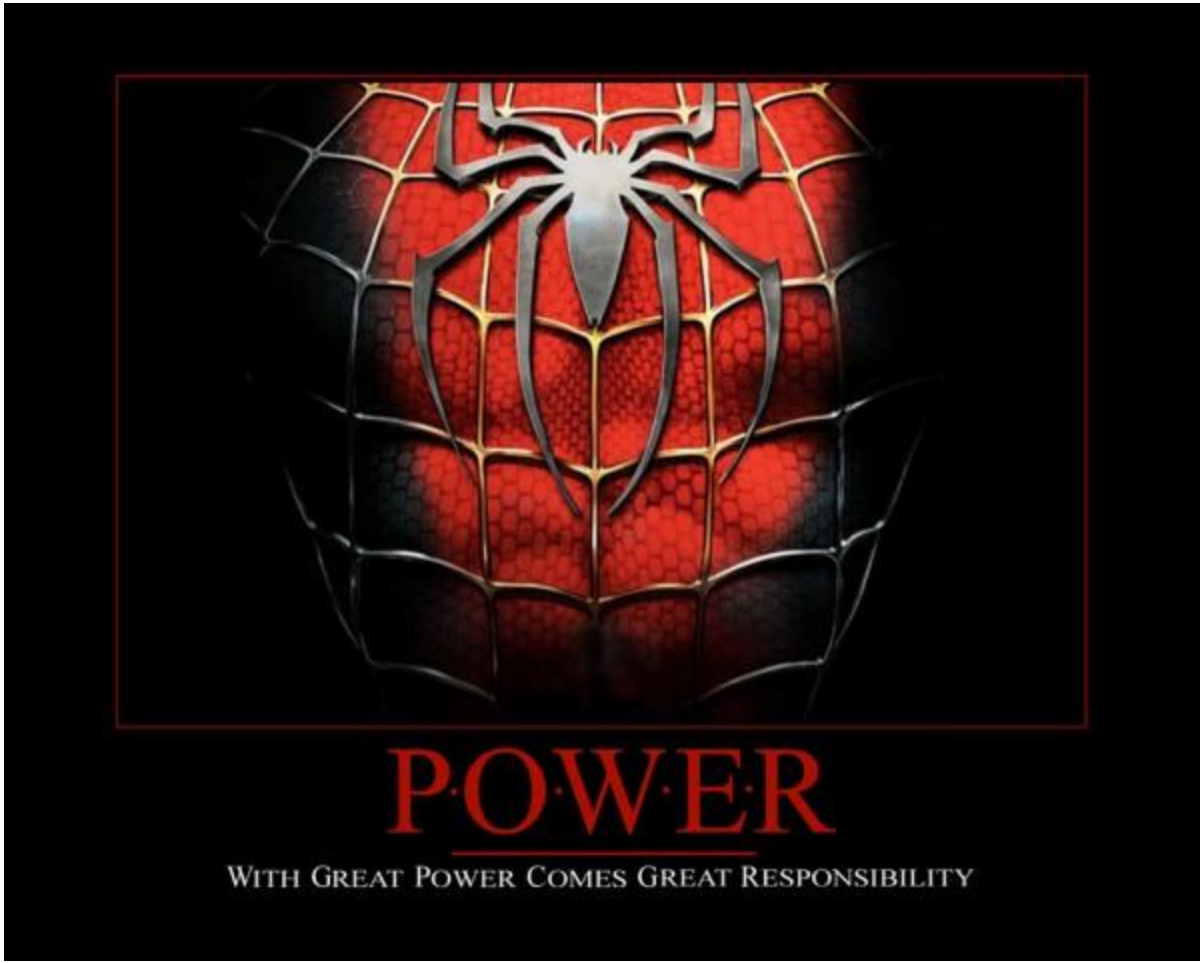
BioPython (http://biopython.org)



Screenshot made at http://biopython.org/wiki/Biopython on December 1st, 2013 (UTC+0800)

Look forward to teaching something YOU develop!

Use the resources and enjoy the power!



http://www.notorious-rob.com/2011/03/22/social-media-taught/

生物信息学：导论与方法
Bioinformatics: Introduction and Methods

Ge Gao 高歌 & Liping Wei 魏丽萍
Center for Bioinformatics, Peking University

https://www.coursera.org/course/pkubioinfo