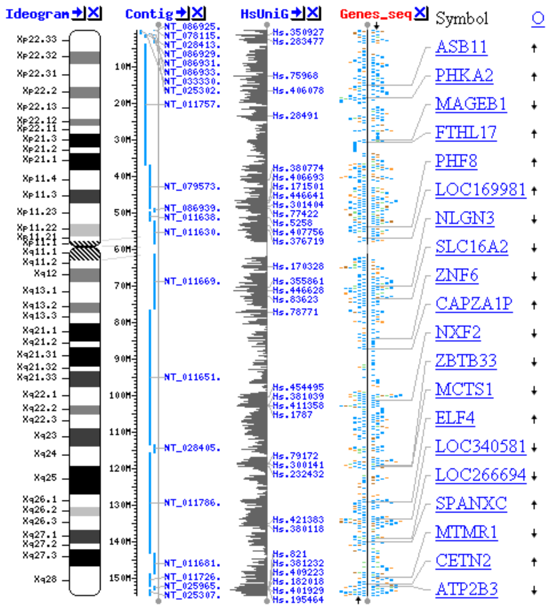
Bioinformatics

For the journal, see Bioinformatics (journal).

Bioinformatics /ˌbaɪ.oʊˌɪnfərˈmætɪks/ is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, bioinformatics combines computer science, statistics, mathematics, and engineering to study and process biological data.



Map of the human X chromosome (from the NCBI website). Assembly of the human genome is one of the greatest achievements of bioinformatics.

Bioinformatics is both an umbrella term for the body of biological studies that use computer programming as part of their methodology, as well as a reference to specific analysis "pipelines" that are repeatedly used, particularly in the fields of genetics and genomics. Common uses of bioinformatics include the identification of candidate genes and nucleotides (SNPs). Often, such identification is made with the aim of better understanding the genetic basis of disease, unique adaptations, desirable properties (esp. in agricultural species), or differences between populations. In a less formal way, bioinformatics also tries to understand the organisational principles within nucleic acid and protein sequences.

1 Introduction

Bioinformatics has become an important part of many areas of biology. In experimental molecular biology, bioinformatics techniques such as image and signal processing allow extraction of useful results from large amounts of raw data. In the field of genetics and genomics, it aids in sequencing and annotating genomes and their observed mutations. It plays a role in the text mining of biological literature and the development of biological and gene ontologies to organize and query biological data. It also plays a role in the analysis of gene and protein expression and regulation. Bioinformatics tools aid in the comparison of genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology. At a more integrative level, it helps analyze and catalogue the biological pathways and networks that are an important part of systems biology. In structural biology, it aids in the simulation and modeling of DNA, RNA, and protein structures as well as molecular interactions.

1.1 History

Paulien Hogeweg and Ben Hesper coined the term *bioinformatics* in 1970 to refer to the study of information processes in biotic systems.[1][2][3] This definition placed bioinformatics as a field parallel to biophysics (the study of physical processes in biological systems) or biochemistry (the study of chemical processes in biological systems).[1]

1.1.1 Sequences

Computers became essential in molecular biology when protein sequences became available after Frederick Sanger determined the sequence of insulin in the early 1950s. Comparing multiple sequences manually turned out to be impractical. A pioneer in the field was Margaret Oakley Dayhoff, who has been hailed by David Lipman, director of the National Center for Biotechnology Information, as the "mother and father of bioinformatics."[4] Dayhoff compiled one of the first protein sequence databases, initially published as books[5] and pioneered methods of sequence alignment and molecular evolution.[6] Another early contributor to bioinformatics was Elvin A. Kabat, who pioneered biological sequence analysis in 1970 with his comprehensive volumes of antibody sequences released with Tai Te Wu between 1980 and 1991.[7]

1.1.2 Genomes

As whole genome sequences became available, again with the pioneering work of Frederick Sanger,[8] it became evident that computer-assisted analysis would be insightful. The first analysis of this type, which had important input from cryptologists at the National Security Agency, was applied to the nucleotide sequences of the bacteriophages MS2 and PhiX174. As a proof of principle, this work showed that standard methods of cryptology could reveal intrinsic features of the genetic code such as the codon length and the reading frame. This work seems to have been ahead of its time—it was rejected for publication by numerous standard journals and finally found a home in the Journal of Theoretical Biology.[9] The term bioinformatics was re-discovered and used to refer to the creation of databases such as GenBank in 1982. With public availability of data, tools for their analysis were quickly developed and described in journals, such as *Nucleic Acids Research*, which published specialized issues on bioinformatics tools as early as 1982.

1.2 Goals

To study how normal cellular activities are altered in different disease states, the biological data must be combined to form a comprehensive picture of these activities. Therefore, the field of bioinformatics has evolved such that the most pressing task now involves the analysis and interpretation of various types of data. This includes nucleotide and amino acid sequences, protein domains, and protein structures.[10] The actual process of analyzing and interpreting data is referred to as computational biology. Important sub-disciplines within bioinformatics and computational biology include:

- Development and implementation of computer programs that enable efficient access to, use and management of, various types of information
- Development of new algorithms (mathematical formulas) and statistical measures that assess relationships among members of large data sets. For example, there are methods to locate a gene within a sequence, to predict protein structure and/or function, and to cluster protein sequences into families of related sequences.

The primary goal of bioinformatics is to increase the understanding of biological processes. What sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques to achieve this goal. Examples include: pattern recognition, data mining, machine learning algorithms, and visualization. Major research efforts in the field include sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and protein–protein interactions, genome-wide association studies, and the modeling of evolution.

Bioinformatics now entails the creation and advancement of databases, algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the management and analysis of biological data.

Over the past few decades rapid developments in genomic and other molecular research technologies and developments in information technologies have combined to produce a tremendous amount of information related to molecular biology. Bioinformatics is the name given to these mathematical and computing approaches used to glean understanding of biological processes.

1.3 Approaches

Common activities in bioinformatics include mapping and analyzing DNA and protein sequences, aligning DNA and protein sequences to compare them, and creating and viewing 3-D models of protein structures.

There are two fundamental ways of modelling a Biological system (e.g., living cell) both coming under Bioinformatic approaches.

- Static
  - Sequences – Proteins, Nucleic acids and Peptides
  - Interaction data among the above entities including microarray data and Networks of proteins, metabolites
- Dynamic
  - Structures – Proteins, Nucleic acids, Ligands (including metabolites and drugs) and Peptides (structures studied with bioinformatics tools are not considered static anymore and their dynamics is often the core of the structural studies)
  - Systems Biology comes under this category including reaction fluxes and variable concentrations of metabolites
  - Multi-Agent Based modelling approaches capturing cellular events such as signalling, transcription and reaction dynamics

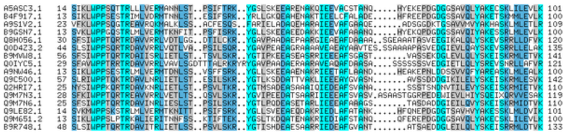
A broad sub-category under bioinformatics is structural bioinformatics.

1.4 Relation to other fields

Bioinformatics is a science field that is similar to but distinct from biological computation and computational biology. Biological computation uses bioengineering and biology to build biological computers, whereas bioinformatics uses computation to better understand biology. Bioinformatics and computational biology have similar aims and approaches, but they differ in scale: bioinformatics organizes and analyzes basic biological data, whereas computational biology builds theoretical models of biological systems, just as mathematical biology does with mathematical models.

Analyzing biological data to produce meaningful information involves writing and running software programs that use algorithms from graph theory, artificial intelligence, soft computing, data mining, image processing, and computer simulation. The algorithms in turn depend on theoretical foundations such as discrete mathematics, control theory, system theory, information theory, and statistics.

2 Sequence analysis



```
A5ASC3.1  14 SIKLWPPSQTTRLLLVERMANNLST..PSIFTRK..YGSLSKEEARENAKQIEEVACSTANQ.....HYEKEPDGDGGSAVQLYAKECSKLILEVLK 101
B4F917.1  13 SIKLWPPSESTRIMLVDRMTNNLST..ESIFSRK..YRLLGKQEAHENAKTIEELCFALADE.....HFREEPDGDGSSAVQLYAKETSKMMLEVLK 100
A9S1V2.1  23 VFKLWPPSQGTREAVRQKMALKLSS..ACFESQS..FARIELADAQEHARAIEEVAFGAAQE......ADSGGDKTGSAVVMVYAKHASKLMLETLR 109
B9GSN7.1  13 SVKLWPPGQSTRLMLVERMTKNFIT..PSFISRK..YGLLSKEEAEEDAKKIEEVAFAAANQ.....HYEKQPDGDGSSAVQIYAKESSRLMLEVLK 100
Q8H056.1  30 SFSIWPPTQRTRDAVVRRLVDTLGG..DTILCKR..YGAVPAADAEPAARGIEAEAFDAAAA..SGEAAATASVEEGIKALQLYSKEVSRRLLDFVK 120
Q0D4Z3.2  44 SLSIWPPSQRTRDAVVRRLVQTLVA..PSILSQR..YGAVPEAEAGRAAAAVEAEAYAAVTES.SSAAAAPASVEDGIEVLQAYSKEVSRRLLELAK 135
B9MVW8.1  56 SFSIWPPTQRTRDAIISRLIETLST..TSVLSKR..YGTIPKEEASEASRRIEEEAFSGAST.......VASSEKDGLEVLQLYSKEISKRMLETVK 141
Q0IYC5.1  29 SFAVWPPTRRTRDAVVRRLVAVLSGDTTTALRKRYRYGAVPAADAERAARAVEAQAFDAASA....SSSSSSSVEDGIETLQLYSREVSNRLLAFVR 121
A9NW46.1  13 SIKLWPPSESTRLMLVERMTDNLSS..VSFFSRK..YGLLSKEEAAENAKRIEETAFLAAND.....HEAKEPNLDDSSVVQFYAREASKLMLEALK 100
Q9C500.1  57 SLRIWPPTQKTRDAVLNRLIETLST..ESILSKR..YGTLKSDDATTVAKLIEEEAYGVASN.......AVSSDDDGIKILELYSKEISKRMLESVK 142
Q2HRI7.1  25 NYSIWPPKQRTRDAVKNRLIETLST..PSVLTKR..YGTMSADEASAAAIQIEDEAFSVANA.......SSSTSNDNVTILEVYSKEISKRMIETVK 110
Q9M7N3.1  28 SFKIWPPTQRTREAVVRRLVETLTS..QSVLSKR..YGVIPEEDATSAARIIEEEAFSVASV.ASAASTGGRPEDEWIEVLHIYSQEIXQRVVESAK 119
Q9M7N6.1  25 SFSIWPPTQRTRDAVINRLIESLST..PSILSKR..YGTLPQDEASETARLIEEEAFAAAGS......TASDADDGIEILQVYSKEISKRMIDTVK 110
Q9LE82.1  14 SVKMWPPSKSTRLMLVERMTKNITT..PSIFSRK..YGLLSVEEAEQDAKRIEDLAFATANK.....HFQNEPDGDGTSAVHVYAKESSKLMLDVIK 101
Q9M651.2  13 SIKLWPPSLPTRKALIERITNNFSS..KTIFTEK..YGSLTKDQATENAKRIEDIAFSTANQ.....QFEREPDGDGGSAVQLYAKECSKLILEVLK 100
B9R748.1  48 SLSIWPPTQRTRDAVITRLIETLSS..PSVLSKR..YGTISHDEAESAARRIEDEAFGVANT.......ATSAEDDGLEILQLYSKEISRRMLDTVK 133
```

The sequences of different genes or proteins may be aligned side-by-side to measure their similarity. This alignment compares protein sequences containing WPP domains.

Main articles: Sequence alignment and Sequence database

Since the Phage Φ-X174 was sequenced in 1977,[11] the DNA sequences of thousands of organisms have been decoded and stored in databases. This sequence information is analyzed to determine genes that encode proteins, RNA genes, regulatory sequences, structural motifs, and repetitive sequences. A comparison of genes within a species or between different species can show similarities between protein functions, or relations between species (the use of molecular systematics to construct phylogenetic trees). With the growing amount of data, it long ago became impractical to analyze DNA sequences manually. Today, computer programs such as BLAST are used daily to search sequences from more than 260 000 organisms, containing over 190 billion nucleotides.[12] These programs can compensate for mutations (exchanged, deleted or inserted bases) in the DNA sequence, to identify sequences that are related, but not identical. A variant of this sequence alignment is used in the sequencing process itself. The so-called shotgun sequencing technique (which was used, for example, by The Institute for Genomic Research to sequence the first bacterial genome, *Haemophilus influenzae*)[13] does not produce entire chromosomes. Instead it generates the sequences of many thousands of small DNA fragments (ranging from 35 to 900 nucleotides long, depending on the sequencing technology). The ends of these fragments overlap and, when aligned properly by a genome assembly program, can be used to reconstruct the complete genome. Shotgun sequencing yields sequence data quickly, but the task of assembling the fragments can be quite complicated for larger genomes. For a genome as large as the human genome, it may take many days of CPU time on large-memory, multiprocessor computers to assemble the fragments, and the resulting assembly usually contains numerous gaps that must be filled in later. Shotgun sequencing is the method of choice for virtually all genomes sequenced today, and genome assembly algorithms are a critical area of bioinformatics research.

Another aspect of bioinformatics in sequence analysis is annotation. This involves computational gene finding to search for protein-coding genes, RNA genes, and other functional sequences within a genome. Not all of the nucleotides within a genome are part of genes. Within the genomes of higher organisms, large parts of the DNA do not serve any obvious purpose. This so-called junk DNA may, however, contain unrecognized functional elements. Bioinformatics helps to bridge the gap between genome and proteome projects — for example, in the use of DNA sequences for protein identification.

See also: sequence analysis, sequence mining, sequence profiling tool and sequence motif

2.1 Genome annotation

Main article: Gene prediction

In the context of genomics, annotation is the process of marking the genes and other biological features in a DNA sequence. This process needs to be automated because most genomes are too large to annotate by hand, not to mention the desire to annotate as many genomes as possible, as the rate of sequencing has ceased to pose a bottleneck. Annotation is made possible by the fact that genes have recognisable start and stop regions, although the exact sequence found in these regions can vary between genes.

The first genome annotation software system was designed in 1995 by Owen White, who was part of the team at The Institute for Genomic Research that sequenced and analyzed the first genome of a free-living organism to be decoded, the bacterium *Haemophilus influenzae*. White built a software system to find the genes (fragments of genomic sequence that encode proteins), the transfer RNAs, and to make initial assignments of function to those genes. Most current genome annotation systems work similarly, but the programs available for analysis of genomic DNA, such as the GeneMark program trained and used to find protein-coding genes in *Haemophilus influenzae*, are constantly changing and improving.

2.2 Computational evolutionary biology

Evolutionary biology is the study of the origin and descent of species, as well as their change over time. Informatics has assisted evolutionary biologists by enabling researchers to:

- trace the evolution of a large number of organisms by measuring changes in their DNA, rather than through physical taxonomy or physiological observations alone,
- more recently, compare entire genomes, which permits the study of more complex evolutionary events, such as gene duplication, horizontal gene transfer, and the prediction of factors important in bacterial speciation,
- build complex computational models of populations to predict the outcome of the system over time[14]
- track and share information on an increasingly large number of species and organisms

Future work endeavours to reconstruct the now more complex tree of life.

The area of research within computer science that uses genetic algorithms is sometimes confused with computational evolutionary biology, but the two areas are not necessarily related.

2.3 Comparative genomics

Main article: Comparative genomics

The core of comparative genome analysis is the establishment of the correspondence between genes (orthology analysis) or other genomic features in different organisms. It is these intergenomic maps that make it possible to trace the evolutionary processes responsible for the divergence of two genomes. A multitude of evolutionary events acting at various organizational levels shape genome evolution. At the lowest level, point mutations affect individual nucleotides. At a higher level, large chromosomal segments undergo duplication, lateral transfer, inversion, transposition, deletion and insertion. Ultimately, whole genomes are involved in processes of hybridization, polyploidization and endosymbiosis, often leading to rapid speciation. The complexity of genome evolution poses many exciting challenges to developers of mathematical models and algorithms, who have recourse to a spectra of algorithmic, statistical and mathematical techniques, ranging from exact, heuristics, fixed parameter and approximation algorithms for problems based on parsimony models to Markov Chain Monte Carlo algorithms for Bayesian analysis of problems based on probabilistic models.

Many of these studies are based on the homology detection and protein families computation.

2.4 Genetics of disease

Main article: Genome-wide association studies

With the advent of next-generation sequencing we are obtaining enough sequence data to map the genes of complex diseases such as infertility,[15] breast cancer [16] or Alzheimer's Disease.[17] Genome-wide association studies are essential to pinpoint the mutations for such complex diseases.[18]

2.5 Analysis of mutations in cancer

Main article: Oncogenomics

In cancer, the genomes of affected cells are rearranged in complex or even unpredictable ways. Massive sequencing efforts are used to identify previously unknown point mutations in a variety of genes in cancer. Bioinformaticians continue to produce specialized automated systems to manage the sheer volume of sequence data produced, and they create new algorithms and software to compare the sequencing results to the growing collection of human genome sequences and germline polymorphisms. New physical detection technologies are employed, such as oligonucleotide microarrays to identify chromosomal gains and losses (called comparative genomic hybridization), and single-nucleotide polymorphism arrays to detect known *point mutations*. These detection methods simultaneously measure several hundred thousand sites throughout the genome, and when used in high-throughput to measure thousands of samples, generate terabytes of data per experiment. Again the massive amounts and new types of data generate new opportunities for bioinformaticians. The data is often found to contain considerable variability, or noise, and thus Hidden Markov model and change-point analysis methods are being developed to infer real copy number changes.

Another type of data that requires novel informatics development is the analysis of lesions found to be recurrent among many tumors.

3 Gene and protein expression

3.1 Analysis of gene expression

The expression of many genes can be determined by measuring mRNA levels with multiple techniques including microarrays, expressed cDNA sequence tag (EST) sequencing, serial analysis of gene expression (SAGE) tag sequencing, massively parallel signature sequencing

(MPSS), RNA-Seq, also known as "Whole Transcriptome Shotgun Sequencing" (WTSS), or various applications of multiplexed in-situ hybridization. All of these techniques are extremely noise-prone and/or subject to bias in the biological measurement, and a major research area in computational biology involves developing statistical tools to separate signal from noise in high-throughput gene expression studies. Such studies are often used to determine the genes implicated in a disorder: one might compare microarray data from cancerous epithelial cells to data from non-cancerous cells to determine the transcripts that are up-regulated and down-regulated in a particular population of cancer cells.

3.2 Analysis of protein expression

Protein microarrays and high throughput (HT) mass spectrometry (MS) can provide a snapshot of the proteins present in a biological sample. Bioinformatics is very much involved in making sense of protein microarray and HT MS data; the former approach faces similar problems as with microarrays targeted at mRNA, the latter involves the problem of matching large amounts of mass data against predicted masses from protein sequence databases, and the complicated statistical analysis of samples where multiple, but incomplete peptides from each protein are detected.

3.3 Analysis of regulation

Regulation is the complex orchestration of events starting with an extracellular signal such as a hormone and leading to an increase or decrease in the activity of one or more proteins. Bioinformatics techniques have been applied to explore various steps in this process. For example, promoter analysis involves the identification and study of sequence motifs in the DNA surrounding the coding region of a gene. These motifs influence the extent to which that region is transcribed into mRNA. Expression data can be used to infer gene regulation: one might compare microarray data from a wide variety of states of an organism to form hypotheses about the genes involved in each state. In a single-cell organism, one might compare stages of the cell cycle, along with various stress conditions (heat shock, starvation, etc.). One can then apply clustering algorithms to that expression data to determine which genes are co-expressed. For example, the upstream regions (promoters) of co-expressed genes can be searched for over-represented regulatory elements. Examples of clustering algorithms applied in gene clustering are k-means clustering, self-organizing maps (SOMs), hierarchical clustering, and consensus clustering methods such as the Bi-CoPaM. The later, namely Bi-CoPaM, has been actually proposed to address various issues specific to gene discovery problems such as consistent co-expression of genes over multiple microarray datasets.[19][20]

4 Structural bioinformatics

Main articles: Structural bioinformatics and Protein structure prediction
See also: Structural motif and Structural domain

Protein structure prediction is another important application of bioinformatics. The amino acid sequence of a protein, the so-called primary structure, can be easily determined from the sequence on the gene that codes for it. In the vast majority of cases, this primary structure uniquely determines a structure in its native environment. (Of course, there are exceptions, such as the bovine spongiform encephalopathy – a.k.a. Mad Cow Disease – prion.) Knowledge of this structure is vital in understanding the function of the protein. Structural information is usually classified as one of *secondary*, *tertiary* and *quaternary* structure. A viable general solution to such predictions remains an open problem. Most efforts have so far been directed towards heuristics that work most of the time.

One of the key ideas in bioinformatics is the notion of homology. In the genomic branch of bioinformatics, homology is used to predict the function of a gene: if the sequence of gene *A*, whose function is known, is homologous to the sequence of gene *B,* whose function is unknown, one could infer that B may share A's function. In the structural branch of bioinformatics, homology is used to determine which parts of a protein are important in structure formation and interaction with other proteins. In a technique called homology modeling, this information is used to predict the structure of a protein once the structure of a homologous protein is known. This currently remains the only way to predict protein structures reliably.

One example of this is the similar protein homology between hemoglobin in humans and the hemoglobin in legumes (leghemoglobin). Both serve the same purpose of transporting oxygen in the organism. Though both of these proteins have completely different amino acid sequences, their protein structures are virtually identical, which reflects their near identical purposes.

Other techniques for predicting protein structure include protein threading and *de novo* (from scratch) physics-based modeling.

5 Network and systems biology

Main articles: Computational systems biology, Biological network and Interactome

Network analysis seeks to understand the relationships within biological networks such as metabolic or protein-protein interaction networks. Although biological networks can be constructed from a single type of molecule or entity (such as genes), network biology often attempts to integrate many different data types, such as proteins, small molecules, gene expression data, and others, which are all connected physically, functionally, or both.

Systems biology involves the use of computer simulations of cellular subsystems (such as the networks of metabolites and enzymes that comprise metabolism, signal transduction pathways and gene regulatory networks) to both analyze and visualize the complex connections of these cellular processes. Artificial life or virtual evolution attempts to understand evolutionary processes via the computer simulation of simple (artificial) life forms.

5.1 Molecular interaction networks

Interactions between proteins are frequently visualized and analyzed using networks. This network is made up of protein-protein interactions from Treponema pallidum*, the causative agent of syphilis and other diseases.*

Main articles: Protein–protein interaction prediction and interactome

Tens of thousands of three-dimensional protein structures have been determined by X-ray crystallography and protein nuclear magnetic resonance spectroscopy (protein NMR) and a central question in structural bioinformatics is whether it is practical to predict possible protein–protein interactions only based on these 3D shapes, without performing protein–protein interaction experiments. A variety of methods have been developed to tackle the protein–protein docking problem, though it seems that there is still much work to be done in this field.

Other interactions encountered in the field include Protein–ligand (including drug) and protein–peptide. Molecular dynamic simulation of movement of atoms about rotatable bonds is the fundamental principle behind computational algorithms, termed docking algorithms, for studying molecular interactions.

6 Others

6.1 Literature analysis

Main articles: Text mining and Biomedical text mining

The growth in the number of published literature makes it virtually impossible to read every paper, resulting in disjointed sub-fields of research. Literature analysis aims to employ computational and statistical linguistics to mine this growing library of text resources. For example:

- Abbreviation recognition – identify the long-form and abbreviation of biological terms
- Named entity recognition – recognizing biological terms such as gene names
- Protein-protein interaction – identify which proteins interact with which proteins from text

The area of research draws from statistics and computational linguistics.

6.2 High-throughput image analysis

Computational technologies are used to accelerate or fully automate the processing, quantification and analysis of large amounts of high-information-content biomedical imagery. Modern image analysis systems augment an observer's ability to make measurements from a large or complex set of images, by improving accuracy, objectivity, or speed. A fully developed analysis system may completely replace the observer. Although these systems are not unique to biomedical imagery, biomedical imaging is becoming more important for both diagnostics and research. Some examples are:

- high-throughput and high-fidelity quantification and sub-cellular localization (high-content screening, cytohistopathology, Bioimage informatics)
- morphometrics
- clinical image analysis and visualization
- determining the real-time air-flow patterns in breathing lungs of living animals
- quantifying occlusion size in real-time imagery from the development of and recovery during arterial injury
- making behavioral observations from extended video recordings of laboratory animals

- infrared measurements for metabolic activity determination
- inferring clone overlaps in DNA mapping, e.g. the Sulston score

6.3 High-throughput single cell data analysis

Main article: Flow cytometry bioinformatics

Computational techniques are used to analyse high-throughput, low-measurement single cell data, such as that obtained from flow cytometry. These methods typically involve finding populations of cells that are relevant to a particular disease state or experimental condition.

6.4 Biodiversity informatics

Main article: Biodiversity informatics

Biodiversity informatics deals with the collection and analysis of biodiversity data, such as taxonomic databases, or microbiome data. Examples of such analyses include phylogenetics, niche modelling, species richness mapping, or species identification tools.

7 Databases

Main articles: List of biological databases and Biological database

Databases are essential for bioinformatics research and applications. There is a huge number of available databases covering almost everything from DNA and protein sequences, molecular structures, to phenotypes and biodiversity. Databases generally fall into one of three types. Some contain data resulting directly from empirical methods such as gene knockouts. Others consist of predicted data, and most contain data from both sources. There are meta-databases that incorporate data compiled from multiple other databases. Some others are specialized, such as those specific to an organism. These databases vary in their format, way of accession and whether they are public or not. Some of the most commonly used databases are listed below. For a more comprehensive list, please check the link at the beginning of the subsection.

- Used in Motif Finding: GenomeNet MOTIF Search
- Used in Gene Ontology: DAVID, FuncAssociate, GATHER
- Used in Gene Finding: Hidden Markov Model
- Used in finding Protein Structures/Family: PFAM
- Used for Next Generation Sequencing: (Not database but data format), FASTQ Format
- Used in Gene Expression Analysis: GEO
- Used in Network Analysis: Interaction Analysis Databases(BioGRID, MINT, HPRD, Curated Human Signaling Network), Functional Networks (STRING, KEGG)

Please keep in mind that this is a quick sampling and generally most computation data is supported by wet lab data as well.

8 Software and tools

Software tools for bioinformatics range from simple command-line tools, to more complex graphical programs and standalone web-services available from various bioinformatics companies or public institutions.

8.1 Open-source bioinformatics software

Many free and open-source software tools have existed and continued to grow since the 1980s.[21] The combination of a continued need for new algorithms for the analysis of emerging types of biological readouts, the potential for innovative *in silico* experiments, and freely available open code bases have helped to create opportunities for all research groups to contribute to both bioinformatics and the range of open-source software available, regardless of their funding arrangements. The open source tools often act as incubators of ideas, or community-supported plug-ins in commercial applications. They may also provide *de facto* standards and shared object models for assisting with the challenge of bioinformation integration.

The range of open-source software packages includes titles such as Bioconductor, BioPerl, Biopython, BioJava, BioJS, BioRuby, Bioclipse, EMBOSS, .NET Bio, Taverna workbench, and UGENE. To maintain this tradition and create further opportunities, the non-profit Open Bioinformatics Foundation[21] have supported the annual Bioinformatics Open Source Conference (BOSC) since 2000.[22]

8.2 Web services in bioinformatics

SOAP- and REST-based interfaces have been developed for a wide variety of bioinformatics applications allowing an application running on one computer in one part of the world to use algorithms, data and computing resources on servers in other parts of the world. The main advantages derive from the fact that end users do not have to deal with software and database maintenance overheads.

Basic bioinformatics services are classified by the EBI into three categories: SSS (Sequence Search Services), MSA (Multiple Sequence Alignment), and BSA (Biological Sequence Analysis).[23] The availability of these service-oriented bioinformatics resources demonstrate the applicability of web-based bioinformatics solutions, and range from a collection of standalone tools with a common data format under a single, standalone or web-based interface, to integrative, distributed and extensible bioinformatics workflow management systems.

8.3 Bioinformatics workflow management systems

Main article: Bioinformatics workflow management systems

A Bioinformatics workflow management system is a specialized form of a workflow management system designed specifically to compose and execute a series of computational or data manipulation steps, or a workflow, in a Bioinformatics application. Such systems are designed to

- provide an easy-to-use environment for individual application scientists themselves to create their own workflows
- provide interactive tools for the scientists enabling them to execute their workflows and view their results in real-time
- simplify the process of sharing and reusing workflows between the scientists.
- enable scientists to track the provenance of the workflow execution results and the workflow creation steps.

Some of the platforms giving this service: Galaxy, Kepler, Taverna, UGENE, Anduril.

9 Education platforms

Software platforms designed to teach bioinformatics concepts and methods include Rosalind and online courses offered through the Swiss Institute of Bioinformatics Training Portal. The Canadian Bioinformatics Workshops provides videos and slides from training workshops on their website under a Creative Commons license.

10 Conferences

There are several large conferences that are concerned with bioinformatics. Some of the most notable examples are Intelligent Systems for Molecular Biology (ISMB), European Conference on Computational Biology (ECCB), Research in Computational Molecular Biology (RECOMB) and American Society of Mass Spectrometry (ASMS).

11 See also

- Biodiversity informatics
- Bioinformatics companies
- Computational biology
- Computational biomodeling
- Computational genomics
- Functional genomics
- Health informatics
- International Society for Computational Biology
- Jumping library
- List of Master of Science in Bioinformatics
- List of free online bioinformatics courses
- List of open-source bioinformatics software
- List of scientific journals in bioinformatics
- Margaret Oakley Dayhoff
- Metabolomics
- Phylogenetics
- Proteomics
- Structural bioinformatics
- Gene Disease Database

12 References

[1] Hogeweg, P. (2011). Searls, David B., ed. "The Roots of Bioinformatics in Theoretical Biology". *PLoS Computational Biology* **7** (3): e1002021. Bibcode:2011PLSCB...7E0020H. doi:10.1371/journal.pcbi.1002021. PMC 3068925. PMID 21483479.

[2] Hesper B, Hogeweg P (1970). "Bioinformatica: een werkconcept" **1** (6). Kameleon. pp. 28–29.

[3] Hogeweg, P. (1978). "Simulating the growth of cellular forms". *Simulation* **31** (3): 90–96. doi:10.1177/003754977803100305.

[4] Moody, Glyn (2004). *Digital Code of Life: How Bioinformatics is Revolutionizing Science, Medicine, and Business.* ISBN 978-0-471-32788-2.

[5] Dayhoff, M.O. (1966) Atlas of protein sequence and structure. National Biomedical Research Foundation, 215 pp.

[6] Eck RV, Dayhoff MO. (1966) Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences. Science. 1966 Apr 15;152(3720):363-366

[7] Johnson, George; Tai Te Wua (January 2000). "Kabat Database and its applications: 30 years after the first variability plot". *Nucleic Acids Res* **28** (1): 214–218. doi:10.1093/nar/28.1.214. PMC 102431. PMID 10592229.

[8] Sanger, F.; Air, G. M.; Barrell, B. G.; Brown, N. L.; Coulson, A. R.; Fiddes, J. C.; Hutchison, C. A.; Slocombe, P. M.; Smith, M. (1977). "Nucleotide sequence of bacteriophage φX174 DNA". *Nature* **265** (5596): 687–95. Bibcode:1977Natur.265..687S. doi:10.1038/265687a0. PMID 870828.

[9] "The coding function of nucleotide sequences can be discerned by statistical analysis.". *Journal of Theoretical Biology.* **88**: 409–420. 1981. doi:10.1016/0022-5193(81)90274-5.

[10] Attwood TK, Gisel A, Eriksson N-E, Bongcam-Rudloff E (2011). "Concepts, Historical Milestones and the Central Place of Bioinformatics in Modern Biology: A European Perspective". *Bioinformatics – Trends and Methodologies*. InTech. Retrieved 8 Jan 2012.

[11] Sanger, F.; Air, G.M.; Barrell, B.G.; Brown, N.L.; Coulson, A.R.; Fiddes, J.C.; Hutchison, C.A.; Slocombe; Smith (February 1977). "Nucleotide sequence of bacteriophage phi X174 DNA". *Nature* **265** (5596): 687–95. Bibcode:1977Natur.265..687S. doi:10.1038/265687a0. PMID 870828.

[12] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (January 2008). "GenBank". *Nucleic Acids Res.* **36** (Database issue): D25–30. doi:10.1093/nar/gkm929. PMC 2238942. PMID 18073190.

[13] Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM; Adams; White; Clayton; Kirkness; Kerlavage; Bult; Tomb; Dougherty; Merrick; McKenney; Sutton; Fitzhugh; Fields; Gocyne; Scott; Shirley; Liu; Glodek; Kelley; Weidman; Phillips; Spriggs; Hedblom; Cotton; Utterback; Hanna; Nguyen; Saudek et al. (July 1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd". *Science* **269** (5223): 496–512. Bibcode:1995Sci...269..496F. doi:10.1126/science.7542800. PMID 7542800.

[14] Antonio Carvajal-Rodríguez (2012). "Simulation of Genes and Genomes Forward in Time". *Current Genomics* (Bentham Science Publishers Ltd.) **11** (1): 58–61. doi:10.2174/138920210790218007. PMC 2851118. PMID 20808525.

[15] Aston, K. I. (2014). "Genetic susceptibility to male infertility: News from genome-wide association studies". *Andrology* **2** (3): 315–21. doi:10.1111/j.2047-2927.2014.00188.x. PMID 24574159.

[16] Véron, A; Blein, S; Cox, D. G. (2014). "Genome-wide association studies and the clinic: A focus on breast cancer". *Biomarkers in Medicine* **8** (2): 287–96. doi:10.2217/bmm.13.121. PMID 24521025.

[17] Tosto, G; Reitz, C (2013). "Genome-wide association studies in Alzheimer's disease: A review". *Current Neurology and Neuroscience Reports* **13** (10): 381. doi:10.1007/s11910-013-0381-0. PMC 3809844. PMID 23954969.

[18] Londin, E; Yadav, P; Surrey, S; Kricka, L. J.; Fortina, P (2013). "Use of Linkage Analysis, Genome-Wide Association Studies, and Next-Generation Sequencing in the Identification of Disease-Causing Mutations". *Pharmacogenomics*. Methods in Molecular Biology **1015**. pp. 127–46. doi:10.1007/978-1-62703-435-7_8. ISBN 978-1-62703-434-0. PMID 23824853.

[19] Abu-Jamous B, Fa R, Roberts DJ, Nandi AK, Peddada SD; Fa; Roberts; Nandi (11 February 2013). Peddada, Shyamal D, ed. "Paradigm of Tunable Clustering Using Binarization of Consensus Partition Matrices (Bi-CoPaM) for Gene Discovery". *PLoS ONE* **8** (2): e56432. Bibcode:2013PLoSO...856432A. doi:10.1371/journal.pone.0056432. PMC 3569426. PMID 23409186.

[20] Abu-Jamous B, Fa R, Roberts DJ, Nandi AK (24 January 2013). "Yeast gene CMR1/YDL156W is consistently co-expressed with genes participating in DNA-metabolic processes in a variety of stringent clustering experiments". *Journal of the Royal Society Interface* **10** (81): 20120990–20120990. doi:10.1098/rsif.2012.0990.

[21] "Open Bioinformatics Foundation: About us". *Official website*. Open Bioinformatics Foundation. Retrieved 10 May 2011.

[22] "Open Bioinformatics Foundation: BOSC". *Official website*. Open Bioinformatics Foundation. Retrieved 10 May 2011.

[23] Nisbet, Robert (14 May 2009). "BIOINFORMATICS". *Handbook of Statistical Analysis and Data Mining Applications*. John Elder IV, Gary Miner. Academic Press. p. 328. Retrieved 9 May 2014.

13 Further reading

- Achuthsankar S Nair Computational Biology & Bioinformatics – A gentle Overview, Communications of Computer Society of India, January 2007
- Aluru, Srinivas, ed. *Handbook of Computational Molecular Biology*. Chapman & Hall/Crc, 2006. ISBN 1-58488-406-1 (Chapman & Hall/Crc Computer and Information Science Series)
- Baldi, P and Brunak, S, *Bioinformatics: The Machine Learning Approach*, 2nd edition. MIT Press, 2001. ISBN 0-262-02506-X

- Barnes, M.R. and Gray, I.C., eds., *Bioinformatics for Geneticists*, first edition. Wiley, 2003. ISBN 0-470-84394-2
- Baxevanis, A.D. and Ouellette, B.F.F., eds., *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, third edition. Wiley, 2005. ISBN 0-471-47878-4
- Baxevanis, A.D., Petsko, G.A., Stein, L.D., and Stormo, G.D., eds., *Current Protocols in Bioinformatics*. Wiley, 2007. ISBN 0-471-25093-7
- Cristianini, N. and Hahn, M. *Introduction to Computational Genomics*, Cambridge University Press, 2006. (ISBN 9780521671910 | ISBN 0-521-67191-4)
- Durbin, R., S. Eddy, A. Krogh and G. Mitchison, *Biological sequence analysis*. Cambridge University Press, 1998. ISBN 0-521-62971-3
- Gilbert D (2004). "Bioinformatics software resources". *Briefings in Bioinformatics* **5** (3): 300–304. doi:10.1093/bib/5.3.300. PMID 15383216.
- Keedwell, E., *Intelligent Bioinformatics: The Application of Artificial Intelligence Techniques to Bioinformatics Problems*. Wiley, 2005. ISBN 0-470-02175-6
- Kohane, et al. *Microarrays for an Integrative Genomics.* The MIT Press, 2002. ISBN 0-262-11271-X
- Lund, O. et al. *Immunological Bioinformatics.* The MIT Press, 2005. ISBN 0-262-12280-4
- Pachter, Lior and Sturmfels, Bernd. "Algebraic Statistics for Computational Biology" Cambridge University Press, 2005. ISBN 0-521-85700-7
- Pevzner, Pavel A. *Computational Molecular Biology: An Algorithmic Approach* The MIT Press, 2000. ISBN 0-262-16197-4
- Soinov, L. Bioinformatics and Pattern Recognition Come Together Journal of Pattern Recognition Research (JPRR), Vol 1 (1) 2006 p. 37–41
- Stevens, Hallam, *Life Out of Sequence: A Data-Driven History of Bioinformatics*, Chicago: The University of Chicago Press, 2013, ISBN 9780226080208
- Tisdall, James. "Beginning Perl for Bioinformatics" O'Reilly, 2001. ISBN 0-596-00080-4
- Dedicated issue of *Philosophical Transactions B* on Bioinformatics freely available
- Catalyzing Inquiry at the Interface of Computing and Biology (2005) CSTB report
- Calculating the Secrets of Life: Contributions of the Mathematical Sciences and computing to Molecular Biology (1995)
- Foundations of Computational and Systems Biology MIT Course
- Computational Biology: Genomes, Networks, Evolution Free MIT Course

14 External links

- Bioinformatics Resource Portal (SIB)

15 Text and image sources, contributors, and licenses

15.1 Text

- **Bioinformatics** *Source:* http://en.wikipedia.org/wiki/Bioinformatics?oldid=640826431 *Contributors:* Carey Evans, Mav, Malcolm Farmer, Youssefsan, Edward, Michael Hardy, Zashaw, Lexor, DIG, Shyamal, Kku, Gaurav, Cyde, Pde, 168..., Ahoerstemeier, Darkwind, Azazello, Susurrus, Dod1, Mxn, Hike395, Sjoerd de Vries, EALacey, Dcoetzee, Dysprosia, Wik, Zoicon5, Steinsky, Samsara, Raul654, Fcrozat, Rhys, Donarreiskoffer, Robbot, Schutz, Peak, Burningsquid, Seglea, JosephBarillari, Chopchopwhitey, Stewartadcock, Spin2cool, Cholling, Fuelbottle, Tobias Bergemann, David Gerard, Giftlite, Counsell, Tdhoufek, Brona, Dmb000006, Chameleon, Bobblewik, Alan Au, Christopherlin, Wmahan, Adenosine, Dullhunk, Bact, CryptoDerk, Quadell, AhmedMoustafa, Vanished user 1234567890, Biovini, PDH, Vina, Karol Langner, MacGyverMagic, APH, Gene s, Bornslippy, Kevin B12, Gscshoyru, Imjustmatthew, Muijz, Thorwald, Gazpacho, Metahacker, Discospinster, Rich Farmbrough, Guanabot, Mazi, Bender235, Tompw, El C, Kwamikagami, Alex Kosorukoff, Cyc, Reinyday, Viriditas, ZayZayEM, Malafaya, Cavrdg, Fotinakis, 3mta3, Kierano, Jjron, Angelsh, HasharBot, Jumbuck, Alansohn, Andkaha, Arthena, Jengeldk, Avenue, Gzur, Eramesan, Andreas C, Ensignyu, HenkvD, Alai, Ringbang, Don G., Bookandcoffee, Walshga, Dismas, Oleg Alexandrov, Natalya, Natarajanganesan, Woohookitty, Mindmatrix, RHaworth, Bonus Onus, Ruud Koot, Acerperi, Vasundhar, Tincup, Joerg Kurt Wegner, Turnstep, Asidhu, Qwertyus, Grammarbot, Porcher, Rjwilmsi, Mayumashu, Smoe, Bill37212, DonSiano, FlaBot, Vietbio, Mathbot, Scottzed, Nihiltres, MicroBio Hawk, Nivix, Otets, Banazir, Rgonzaga, AndriuZ, FreeKill, Tedder, Terrace4, Spencerk, Wavelength, RussBot, RobHutten, Muchness, Ansell, Pseudomonas, Joelrex, Martin.jambon, Cquan, Jchusid, Larry laptop, Rmky87, Tony1, Supten, JHCaufield, Kkmurray, Pawyilee, Leptictidium, Zzuuzz, Lt-wiki-bot, Mateo LeFou, Perfectlover, GraemeL, CWenger, Shawnc, Banus, □□□□ robot, Macdorman, Blastwizard, Vanka5, SmackBot, Malkinann, TestPilot, Dblobaum, Joconnol, CommodiCast, Warren.cheung, Edgar181, Tim@, Gilliam, Ohnoitsjamie, Nervexmachina, Ashcroft, Bluebot, Jethero, Wieghardt, Thumperward, Martin Jambon, Dan198792, Huji, Iwaterpolo, Can't sleep, clown will eat me, Frap, JonHarder, Postdoc, Jedgold, Bffo, Akriasas, G716, Phismith, EdGl, Bradenripple, Jcuticchia, Akpakp, Madeleine Price Ball, Ymichel, ArglebargleIV, Thenothing, Ben Moore, 16@r, Tarcieri, JHunterJ, Beetstra, Macha, Ehheh, Kpengboy, Parakkum, Jameslyonsweiler, Dave Messina, Pselvakumar, Mattigatti, Bioinformin, Kiwi2795, Mstrangwick, Foscoe, CRGreathouse, Thermochap, CmdrObot, MattWBradbury, Bill.albing, Ternto333, Cydebot, Ppgardne, VashiDonsk, Rifleman 82, Chasingsol, Marco.caminati, Lennonr2, Tawkerbot4, Msnicki, Narayanese, Girlwithglasses, Rintintin, Zhuozhuo, Epbr123, Barticus88, Opabinia regalis, ConceptExp, Smjc, Peter Znamenskiy, Tapir Terrific, Tellyaddict, EdJohnston, Renji143, S177, Thomaswgc, Wiki1forall, Neksa, KP Botany, TimVickers, PJY, Mbadri, Danger, Minimice, Kevin.cohen, Qwerty Binary, Vawter, Gökhan, FCAlive, JAnDbot, Deflective, MER-C, Ph.eyes, Bmunro, Andersduck, Henriettaminge, Efbfweborg, Magioladitis, VoABot II, David Ardell, Tupeliano, Winhide, Leofer, WhatamIdoing, Bmeguru, User A1, Stinkbeard, AuGold, Genometer, Glen, Bm richard, Yoni, Hawksj, Colin gravill, Gwern, MartinBot, AstarothCY, HoopyFrood, Anaxial, Nono64, Ashalatha.jangala, Boghog, Aetkin, Keesiewonder, Vegasprof, Bot-Schafter, Artgen, Jorfer, Bob, Whiteandnerdy52, Kamleong, GLHamilton, Pdcook, Remi0o, Jamiejoseph, VolkovBot, Joeoettinger, Redgecko, AlnoktaBOT, Jimmaths, TXiKiBoT, Ajkarloss, Rvencio, Scilit, Guillaume2303, Josephholsten, Minho Bio Lee, Agricola44, Cbergman, Littlealien182, Praveen pillay, Tmccrae, Duncan.Hull, Jamelan, Googed, Alexbateman, Bcheng23, EmxBot, Marashie, SieBot, Mikemoral, Graham Beards, Amandadawnbesemer, Jjwilkerson, Venus Victorious, Angusmca, Gordon014, Strife911, Mimihitam, Nicksh, Denisarona, Kayvan45622, Kotsiantis, Mike Yang, ClueBot, Shortliffe, PipepBot, Meekywiki, Basel1988, Peteruetz, Nabeelbasheer, DragonBot, Kjramesh, Jotterbot, Jkbioinfo, Jgrethe, Mobashirgenome, Marcoacostareyes, Burner0718, Tombadog, Qwfp, Cbock, Tpvipin, XLinkBot, Rror, Motorious, Avoided, Ariconte, SilvonenBot, MystBot, FireBrandon, Subhashis.behera, WilliamBonfieldFRS, Willking1979, LeeWatts, Mootros, Lynx8, Pascal.hingamp, MrOllie, Protonk, LinkFA-Bot, Bioinformaticsguru, CosmiCarl, Numbo3-bot, Veterinarian, Asasia, Zorrobot, Sankalpjain202, Bonnarj, W09110900, Legobot, आशीष भटनागर, Luckas-bot, Yobot, Bunnyhop11, Senator Palpatine, Joychen2010, Oleginger, Imtechchd, AnomieBOT, 1exec1, Jim1138, Ambertk, Piano non troppo, Bluerasberry, Materialscientist, Citation bot, Ganeshbio1, Gulan722, Perada, Xqbot, TheAMmollusc, Navigatorwiki, Surajbodi, Bio-ITWorld, Gilo1969, P99am, J04n, Prunesqualer, Shubinator, Mangst, Shadowjams, Methcub, Gonfus, Mauriceling, SexyGod, My walker 88, Bonio05, Dr02115, Hillarivallen, DrilBot, Colonialdirt, Tintenfischlein, Zorozorozoro123, LukeGoodsell, 124Nick, Vizbi, Agemoi, MertyWiki, Yves.lussier, Genypholy, TobeBot, Trappist the monk, SchreyP, Grammarxxx, Aoidh, Vishnugaikwad, Amkilpatrick, Minimac, FTasc, Generalboss3, Peccoud, Wojcz, P-O limhamn, EmausBot, John of Reading, Orphan Wiki, RaoInWiki, Garvind95, Cogiati, U 06111976, Vilietha, Sherell.jones, Malky132, L Kensington, ChuispastonBot, ClueBot NG, Yeturukalidas, Dforsdyke, Cntras, Shire Reeve, Helpful Pixie Bot, Amoghb, Xandrox, Xlnc1706, Bibcode Bot, BG19bot, Maxmans, Ckuanglim, AvocatoBot, Lpantano, Edward Gordon Gey, No snow, Gupta.udatha, Whoknew.dat, Ymei, Eidenberger, Nizamibilal1064, Harizotoh9, TheProfessor, Rlouhimo, BattyBot, Ashyel leephen, Pratyya Ghosh, ChrisGualtieri, Aloctavodia, Vivek bioinformatics, WikimatCS, Intessan, Wimblecf, Phcompeau, Joeinwiki, Mark viking, LPmore, SomeFreakOnTheInternet, SharptoothX, Mattsabe, TwilightMirror, Medwardm, Naf312, Paulorapazote, Codycann, Star767, Michaellevitt99, Imehedi, Cynthia MEDG, FreeBird541, DavidLeighEllis, NameAday, Zelmah, Ginsuloft, Sarmstrong123, Danneks, Btorcaso, Sirisindhu, TrystynAlxander, Markjulmar, Sanyk28, Éffièdaligrh, FASSMAN, Monkbot, Luongdl, RaihaT, Demi lion, Pranjali jadhav, Mohan kumar yadav, Morgantaschuk, Jorge Guerra Pires, Apaf1, The Scientific Gadfly, Razhielin and Anonymous: 500

15.2 Images

- **File:Commons-logo.svg** *Source:* http://upload.wikimedia.org/wikipedia/en/4/4a/Commons-logo.svg *License:* ? *Contributors:* ? *Original artist:* ?
- **File:En-Bioinformatics.ogg** *Source:* http://upload.wikimedia.org/wikipedia/commons/4/43/En-Bioinformatics.ogg *License:* CC BY-SA 3.0 *Contributors:* Derivative of Bioinformatics at Wikipedia *Original artist:* Mangst
- **File:Folder_Hexagonal_Icon.svg** *Source:* http://upload.wikimedia.org/wikipedia/en/4/48/Folder_Hexagonal_Icon.svg *License:* Cc-by-sa-3.0 *Contributors:* ? *Original artist:* ?
- **File:Genome_viewer_screenshot_small.png** *Source:* http://upload.wikimedia.org/wikipedia/commons/4/43/Genome_viewer_screenshot_small.png *License:* Public domain *Contributors:* ? *Original artist:* ?
- **File:Issoria_lathonia.jpg** *Source:* http://upload.wikimedia.org/wikipedia/commons/2/2d/Issoria_lathonia.jpg *License:* CC-BY-SA-3.0 *Contributors:* ? *Original artist:* ?
- **File:Open_book_nae_02.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/9/92/Open_book_nae_02.svg *License:* CC0 *Contributors:* OpenClipart *Original artist:* nae

- **File:Portal-puzzle.svg** *Source:* http://upload.wikimedia.org/wikipedia/en/f/fd/Portal-puzzle.svg *License:* Public domain *Contributors:* ? *Original artist:* ?
- **File:Sound-icon.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/4/47/Sound-icon.svg *License:* LGPL *Contributors:* Derivative work from Silsor's versio *Original artist:* Crystal SVG icon set
- **File:Speakerlink-new.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/3/3b/Speakerlink-new.svg *License:* CC0 *Contributors:* Own work *Original artist:* Kelvinsong
- **File:Symbol_book_class2.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/8/89/Symbol_book_class2.svg *License:* CC BY-SA 2.5 *Contributors:* Mad by Lokal_Profil by combining: *Original artist:* Lokal_Profil
- **File:The_protein_interaction_network_of_Treponema_pallidum.png** *Source:* http://upload.wikimedia.org/wikipedia/commons/b/b4/The_protein_interaction_network_of_Treponema_pallidum.png *License:* CC BY 1.0 *Contributors:* Titz B, Rajagopala SV, Goll J, Ha ̈user R, McKevitt MT, et al. (2008) The Binary Protein Interactome of Treponema pallidum – The Syphilis Spirochete. PLoS ONE 3(5): e2292. doi:10.1371/journal.pone.0002292 *Original artist:* Häuser et al.
- **File:Tree_of_life.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/0/09/Tree_of_life.svg *License:* CC-BY-SA-3.0 *Contributors:* ? *Original artist:* ?
- **File:WPP_domain_alignment.PNG** *Source:* http://upload.wikimedia.org/wikipedia/commons/a/a7/WPP_domain_alignment.PNG *License:* CC0 *Contributors:* Own work *Original artist:* Alexbateman
- **File:Wikiquote-logo.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/f/fa/Wikiquote-logo.svg *License:* Public domain *Contributors:* ? *Original artist:* ?
- **File:Wikiversity-logo.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/9/91/Wikiversity-logo.svg *License:* CC BY-SA 3.0 *Contributors:* Snorky (optimized and cleaned up by verdy_p) *Original artist:* Snorky (optimized and cleaned up by verdy_p)
- **File:Wiktionary-logo-en.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/f/f8/Wiktionary-logo-en.svg *License:* Public domain *Contributors:* Vector version of Image:Wiktionary-logo-en.png. *Original artist:* Vectorized by Fvasconcellos (talk · contribs), based on original logo tossed together by Brion Vibber

15.3 Content license

- Creative Commons Attribution-Share Alike 3.0