Genomics

Genomics is an interdisciplinary field of biology focusing on the structure, function, evolution, mapping, and editing of genomes. A genome is an organism's complete set of DNA, including all of its genes. In contrast to genetics, which refers to the study of *individual* genes and their roles in inheritance, genomics aims at the collective characterization and quantification of *all* of an organism's genes, their interrelations and influence on the organism.[1] Genes may direct the production of proteins with the assistance of enzymes and messenger molecules. In turn, proteins make up body structures such as organs and tissues as well as control chemical reactions and carry signals between cells. Genomics also involves the sequencing and analysis of genomes through uses of high throughput DNA sequencing and bioinformatics to assemble and analyze the function and structure of entire genomes.[2][3] Advances in genomics have triggered a revolution in discovery-based research and systems biology to facilitate understanding of even the most complex biological systems such as the brain.[4]

The field also includes studies of intragenomic (within the genome) phenomena such as epistasis (effect of one gene on another), pleiotropy (one gene affecting more than one trait), heterosis (hybrid vigour), and other interactions between loci and alleles within the genome.[5]

Contents

History

Etymology

From the Greek ΓΕΝ[6] *gen*, "gene" (gamma, epsilon, nu, epsilon) meaning "become, create, creation, birth", and subsequent variants: genealogy, genesis, genetics, genic, genomere, genotype, genus etc. While the word *genome* (from the German *Genom*, attributed to Hans Winkler) was in use in English as early as 1926,[7] the term *genomics* was coined by Tom Roderick, a geneticist at the Jackson Laboratory (Bar Harbor, Maine), over beer at a meeting held in Maryland on the mapping of the human genome in 1986.[8]

Early sequencing efforts

Following Rosalind Franklin's confirmation of the helical structure of DNA, James D. Watson and Francis Crick's publication of the structure of DNA in 1953 and Fred Sanger's publication of the Amino acid sequence of insulin in 1955, nucleic acid sequencing became a major target of early molecular biologists.[9] In 1964, Robert W. Holley and colleagues published the first nucleic acid sequence ever determined, the ribonucleotide sequence of alanine transfer RNA.[10][11] Extending this work, Marshall Nirenberg and Philip Leder revealed the triplet nature of the genetic code and were able to determine the sequences of 54 out of 64 codons in their experiments.[12] In 1972, Walter Fiers and his team at the Laboratory of Molecular Biology of the University of Ghent (Ghent, Belgium) were the first to determine the sequence of a gene: the gene for Bacteriophage MS2 coat protein.[13] Fiers' group expanded on their MS2 coat protein work, determining the complete nucleotide-sequence of bacteriophage MS2-RNA (whose genome encodes just four genes in 3569 base pairs [bp]) and Simian virus 40 in 1976 and 1978, respectively.[14][15]

DNA-sequencing technology developed

In addition to his seminal work on the amino acid sequence of insulin, Frederick Sanger and his colleagues played a key role in the development of DNA sequencing techniques that enabled the establishment of comprehensive genome sequencing projects.[5] In 1975, he and Alan Coulson published a sequencing procedure using DNA polymerase with radiolabelled nucleotides that he called the *Plus and Minus technique*.[16][17] This involved two closely related methods that generated short oligonucleotides with defined 3' termini. These could be fractionated by electrophoresis on a polyacrylamide gel (called polyacrylamide gel electrophoresis) and visualised using autoradiography. The procedure could sequence up to 80 nucleotides in one go and was a big improvement, but was still very laborious. Nevertheless, in 1977 his group was able to sequence most of the 5,386 nucleotides of the single-stranded bacteriophage φX174, completing the first fully sequenced DNA-based genome.[18] The refinement of the *Plus and Minus* method resulted in the chain-termination, or Sanger method (see below), which formed the basis of the techniques of DNA sequencing,
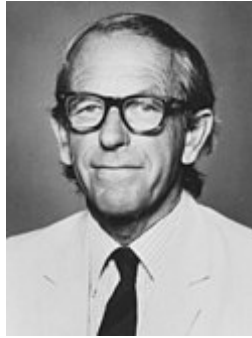
Frederick Sanger

Walter Gilbert

Frederick Sanger and Walter Gilbert shared half of the 1980 Nobel Prize in Chemistry for Independently developing methods for the sequencing of DNA.

genome mapping, data storage, and bioinformatic analysis most widely used in the following quarter-century of research.[19][20] In the same year Walter Gilbert and Allan Maxam of Harvard University independently developed the Maxam-Gilbert method (also known as the *chemical method*) of DNA sequencing, involving the preferential cleavage of DNA at known bases, a less efficient method.[21][22] For their groundbreaking work in the sequencing of nucleic acids, Gilbert and Sanger shared half the 1980 Nobel Prize in chemistry with Paul Berg (recombinant DNA).

Complete genomes

The advent of these technologies resulted in a rapid intensification in the scope and speed of completion of genome sequencing projects. The first complete genome sequence of a eukaryotic organelle, the human mitochondrion (16,568 bp, about 16.6 kb [kilobase]), was reported in 1981,[23] and the first chloroplast genomes followed in 1986.[24][25] In 1992, the first eukaryotic chromosome, chromosome III of brewer's yeast *Saccharomyces cerevisiae* (315 kb) was sequenced.[26] The first free-living organism to be sequenced was that of *Haemophilus influenzae* (1.8 Mb [megabase]) in 1995.[27] The following year a consortium of researchers from laboratories across North America, Europe, and Japan announced the completion of the first complete genome sequence of a eukaryote, *S. cerevisiae* (12.1 Mb), and since then genomes have continued being sequenced at an exponentially growing pace.[28] As of October 2011, the complete sequences are available for: 2,719 viruses, 1,115 archaea and bacteria, and 36 eukaryotes, of which about half are fungi.[29][30]

Most of the microorganisms whose genomes have been completely sequenced are problematic pathogens, such as *Haemophilus influenzae*, which has resulted in a pronounced bias in their phylogenetic distribution compared to the breadth of microbial diversity.[31][32] Of the other sequenced species, most were chosen because they were well-studied model organisms or promised to become good models. Yeast (*Saccharomyces cerevisiae*) has long been an important model organism for the eukaryotic cell, while the fruit fly *Drosophila melanogaster* has been a very important tool (notably in early pre-molecular genetics). The worm *Caenorhabditis elegans* is an often used simple model for multicellular organisms. The zebrafish *Brachydanio rerio* is used for many developmental studies on the molecular level, and the plant *Arabidopsis thaliana* is a model organism for flowering plants. The Japanese pufferfish (*Takifugu rubripes*) and the spotted green pufferfish (*Tetraodon nigroviridis*) are interesting because of their small and compact genomes, which contain very little noncoding DNA compared to most species.[33][34] The mammals dog (*Canis familiaris*),[35] brown rat (*Rattus norvegicus*), mouse (*Mus musculus*), and chimpanzee (*Pan troglodytes*) are all important model animals in medical research.[22]

A rough draft of the human genome was completed by the Human Genome Project in early 2001, creating much fanfare.[36] This project, completed in 2003, sequenced the entire genome for one specific person, and by 2007 this sequence was declared "finished" (less than one error in 20,000 bases and all chromosomes
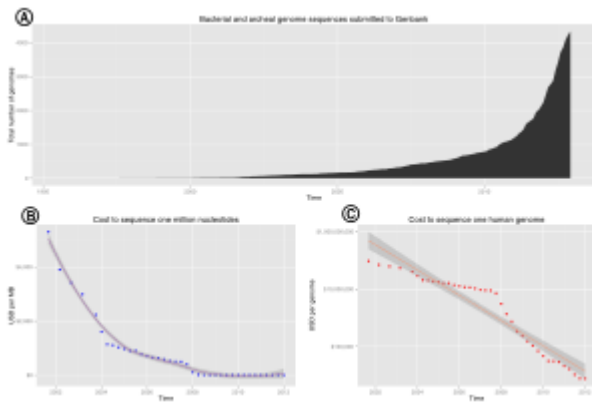
The number of genome projects has increased as technological improvements continue to lower the cost of sequencing. **(A)** Exponential growth of genome sequence databases since 1995. **(B)** The cost in US Dollars (USD) to sequence one million bases. **(C)** The cost in USD to sequence a 3,000 Mb (human-sized) genome on a log-transformed scale.

assembled).[36] In the years since then, the genomes of many other individuals have been sequenced, partly under the auspices of the 1000 Genomes Project, which announced the sequencing of 1,092 genomes in October 2012.[37] Completion of this project was made possible by the development of dramatically more efficient sequencing technologies and required the commitment of significant bioinformatics resources from a large international collaboration.[38] The continued analysis of human genomic data has profound political and social repercussions for human societies.[39]

The "omics" revolution

General schema showing the relationships of the genome, transcriptome, proteome, and metabolome (lipidome).

The English-language neologism **omics** informally refers to a field of study in biology ending in *-omics*, such as genomics, proteomics or metabolomics. The related suffix **-ome** is used to address the objects of study of such fields, such as the genome, proteome or metabolome respectively. The suffix *-ome* as used in molecular biology refers to a *totality* of some sort; similarly **omics** has come to refer generally to the study of large, comprehensive biological data sets. While the growth in the use of the term has led some scientists (Jonathan Eisen, among others[40]) to claim that it has been oversold,[41] it reflects the change in orientation towards the quantitative analysis of complete or near-complete assortment of all the constituents of a system.[42] In the study of symbioses, for example, researchers which were once limited to the study of a single gene product can now simultaneously compare the total complement of several types of biological molecules.[43][44]
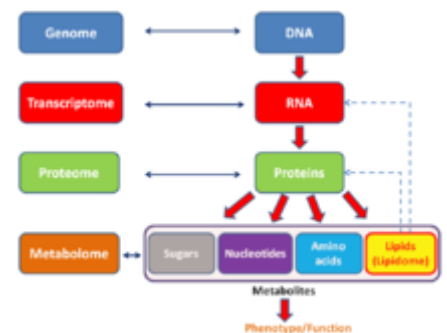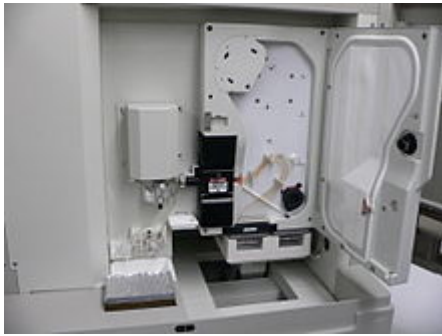
Genome analysis

After an organism has been selected, genome projects involve three components: the sequencing of DNA, the assembly of that sequence to create a representation of the original chromosome, and the annotation and analysis of that representation.[5]

Sequencing

Historically, sequencing was done in *sequencing centers*, centralized facilities (ranging from large independent institutions such as Joint Genome Institute which sequence dozens of terabases a year, to local molecular biology core facilities) which contain research laboratories with the costly instrumentation and technical support necessary. As sequencing technology continues to improve, however, a new generation of effective fast turnaround benchtop sequencers has come within reach of the average academic laboratory.[45][46] On the whole, genome sequencing approaches fall into two broad categories, *shotgun* and *high-throughput* (or *next-generation*) sequencing.[5]

Shotgun sequencing


An ABI PRISM 3100 Genetic Analyzer. Such capillary sequencers automated early large-scale genome sequencing efforts.

Overview of a genome project. First, the genome must be selected, which involves several factors including cost and relevance. Second, the sequence is generated and assembled at a given sequencing center (such as BGI or DOE JGI). Third, the genome sequence is annotated at several levels: DNA, protein, gene pathways, or comparatively.

Shotgun sequencing is a sequencing method designed for analysis of DNA sequences longer than 1000 base pairs, up to and including entire chromosomes.[47] It is named by analogy with the rapidly expanding, quasi-random firing pattern of a shotgun. Since gel electrophoresis sequencing can only be used for fairly short sequences (100 to 1000 base pairs), longer DNA sequences must be broken into random small segments which are then sequenced to obtain *reads*. Multiple overlapping reads for the target DNA are obtained by performing several rounds of this fragmentation and sequencing. Computer programs then use the overlapping ends of different reads to assemble them into a continuous sequence.[47][48] Shotgun sequencing is a random sampling process, requiring over-sampling to ensure a given nucleotide is represented in the reconstructed sequence; the average number of reads by which a genome is over-sampled is referred to as coverage.[49]

For much of its history, the technology underlying shotgun sequencing was the classical chain-termination method or 'Sanger method', which is based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication.[18][50] Recently, shotgun sequencing has been supplanted by high-throughput sequencing methods, especially for large-scale, automated genome analyses. However, the Sanger method remains in wide use, primarily for smaller-scale projects and for obtaining especially long contiguous DNA sequence reads (>500 nucleotides).[51] Chain-termination methods require a single-stranded DNA template, a DNA primer, a DNA polymerase, normal deoxynucleosidetriphosphates (dNTPs), and modified nucleotides (dideoxyNTPs) that terminate DNA strand elongation. These chain-terminating nucleotides lack a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, causing DNA polymerase to cease extension of DNA when a ddNTP is incorporated. The ddNTPs may be radioactively or fluorescently labelled for detection in DNA sequencers.[5] Typically, these machines can sequence up to 96 DNA samples in a single batch (run) in up to 48 runs a day.[52]
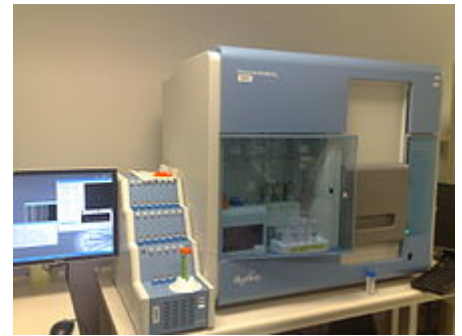
High-throughput sequencing

The high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences at once.[53][54] High-throughput sequencing is intended to lower the cost of DNA sequencing beyond what is
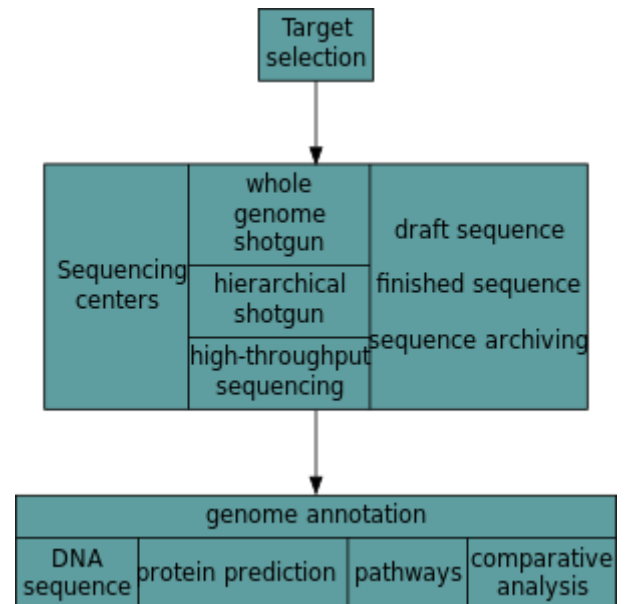
possible with standard dye-terminator methods. In ultra-high-throughput sequencing, as many as 500,000 sequencing-by-synthesis operations may be run in parallel.[55][56]

The Illumina dye sequencing method is based on reversible dye-terminators and was developed in 1996 at the Geneva Biomedical Research Institute, by Pascal Mayer and Laurent Farinelli.[57] In this method, DNA molecules and primers are first attached on a slide and amplified with polymerase so that local clonal colonies, initially coined "DNA colonies", are formed. To determine the sequence, four types of reversible terminator bases (RT-bases) are added and non-incorporated nucleotides are washed away. Unlike pyrosequencing, the DNA chains are extended one nucleotide at a time and image acquisition can be performed at a delayed moment, allowing for very large arrays of DNA colonies to be captured by sequential images taken from a single camera. Decoupling the enzymatic reaction and the image capture allows for optimal throughput and theoretically unlimited sequencing capacity; with an optimal configuration, the ultimate throughput of the instrument depends only on the A/D conversion rate of the camera. The camera takes images of the fluorescently labeled nucleotides, then the dye along with the terminal 3' blocker is chemically removed from the DNA, allowing the next cycle.[58]

Illumina Genome Analyzer II System. Illumina technologies have set the standard for high-throughput massively parallel sequencing.[45]

An alternative approach, ion semiconductor sequencing, is based on standard DNA replication chemistry. This technology measures the release of a hydrogen ion each time a base is incorporated. A microwell containing template DNA is flooded with a single nucleotide, if the nucleotide is complementary to the template strand it will be incorporated and a hydrogen ion will be released. This release triggers an ISFET ion sensor. If a homopolymer is present in the template sequence multiple nucleotides will be incorporated in a single flood cycle, and the detected electrical signal will be proportionally higher.[59]
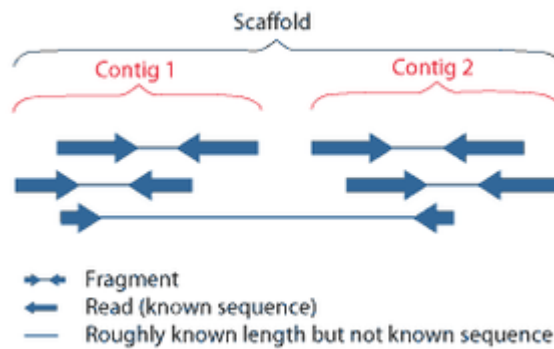
Assembly

Sequence assembly refers to aligning and merging fragments of a much longer DNA sequence in order to reconstruct the original sequence.[5] This is needed as current DNA sequencing technology cannot read whole genomes as a continuous sequence, but rather reads small pieces of between 20 and 1000 bases, depending on the technology used. Third generation sequencing technologies such as PacBio or Oxford Nanopore routinely generate sequencing reads >10 kb in length; however, they have a high error rate at approximately 15 percent.[60][61] Typically the short fragments, called reads, result from shotgun sequencing genomic DNA, or gene transcripts (ESTs).[5]

Assembly approaches

Assembly can be broadly categorized into two approaches: *de novo* assembly, for genomes which are not similar to any sequenced in the past, and comparative assembly, which uses the existing sequence of a closely related organism as a reference during assembly.[49] Relative to comparative assembly, *de novo* assembly is computationally difficult (NP-hard), making it less favourable for short-read NGS technologies. Within the *de novo* assembly paradigm there are two primary strategies for assembly, Eulerian path strategies, and overlap-layout-consensus (OLC) strategies. OLC strategies ultimately try to create a Hamiltonian path through an overlap graph which is an NP-hard problem. Eulerian path strategies are computationally more tractable because they try to find a Eulerian path through a deBruijn graph.[49]

Finishing

Overlapping reads form contigs; contigs and gaps of known length form scaffolds.
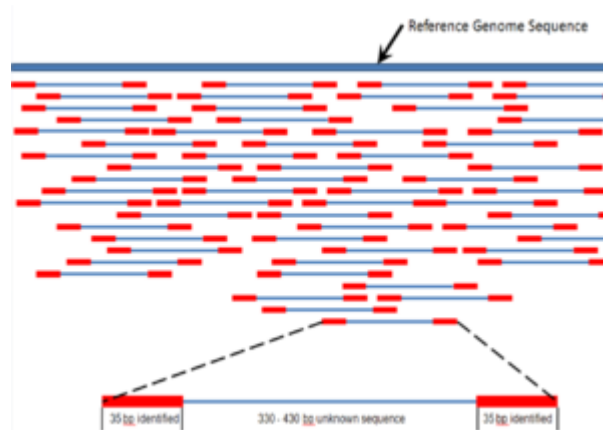
Paired end reads of next generation sequencing data mapped to a reference genome.

Multiple, fragmented sequence reads must be assembled together on the basis of their overlapping areas.

Finished genomes are defined as having a single contiguous sequence with no ambiguities representing each replicon.[62]

Annotation

The DNA sequence assembly alone is of little value without additional analysis.[5] Genome annotation is the process of attaching biological information to sequences, and consists of three main steps:[63]

1. identifying portions of the genome that do not code for proteins
2. identifying elements on the genome, a process called gene prediction, and
3. attaching biological information to these elements.

Automatic annotation tools try to perform these steps *in silico*, as opposed to manual annotation (a.k.a. curation) which involves human expertise and potential experimental verification.[64] Ideally, these approaches co-exist and complement each other in the same annotation pipeline (also see below).

Traditionally, the basic level of annotation is using BLAST for finding similarities, and then annotating genomes based on homologues.[5] More recently, additional information is added to the annotation platform. The additional information allows manual annotators to deconvolute discrepancies between genes that are given the same annotation. Some databases use genome context information, similarity scores, experimental data, and integrations of other resources to provide genome annotations through their Subsystems approach. Other databases (e.g. Ensembl) rely on both curated data sources as well as a range of software tools in their

automated genome annotation pipeline.[65] *Structural annotation* consists of the identification of genomic elements, primarily ORFs and their localisation, or gene structure. *Functional annotation* consists of attaching biological information to genomic elements.

Sequencing pipelines and databases

The need for reproducibility and efficient management of the large amount of data associated with genome projects mean that computational pipelines have important applications in genomics.[66]
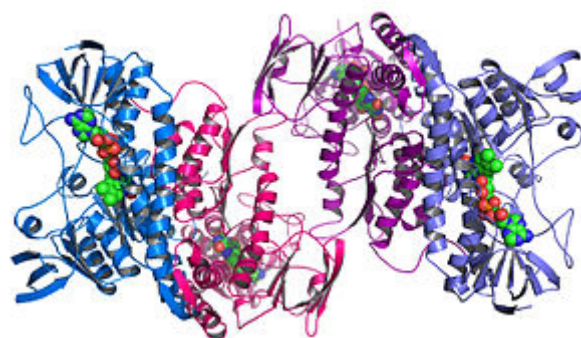
Research areas

Functional genomics

Functional genomics is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (and protein) functions and interactions. Functional genomics focuses on the dynamic aspects such as gene transcription, translation, and protein–protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures. Functional genomics attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts, and protein products. A key characteristic of functional genomics studies is their genome-wide approach to these questions, generally involving high-throughput methods rather than a more traditional "gene-by-gene" approach.

A major branch of genomics is still concerned with sequencing the genomes of various organisms, but the knowledge of full genomes has created the possibility for the field of functional genomics, mainly concerned with patterns of gene expression during various conditions. The most important tools here are microarrays and bioinformatics.

Structural genomics

Structural genomics seeks to describe the 3-dimensional structure of every protein encoded by a given genome.[67][68] This genome-based approach allows for a high-throughput method of structure determination by a combination of experimental and modeling approaches. The principal difference between structural genomics and traditional structural prediction is that structural genomics attempts to determine the structure of every protein encoded by the genome, rather than focusing on one particular protein. With full-genome sequences available, structure prediction can be done more quickly through a combination of experimental and modeling approaches, especially because the availability of large numbers of

An example of a protein structure determined by the Midwest Center for Structural Genomics.

sequenced genomes and previously solved protein structures allow scientists to model protein structure on the structures of previously solved homologs. Structural genomics involves taking a large number of approaches to structure determination, including experimental methods using genomic sequences or modeling-based approaches based on sequence or structural homology to a protein of known structure or based on chemical and physical principles for a protein with no homology to any known structure. As opposed to traditional

structural biology, the determination of a protein structure through a structural genomics effort often (but not always) comes before anything is known regarding the protein function. This raises new challenges in structural bioinformatics, i.e. determining protein function from its 3D structure.[69]

Epigenomics

Epigenomics is the study of the complete set of epigenetic modifications on the genetic material of a cell, known as the epigenome.[70] Epigenetic modifications are reversible modifications on a cell's DNA or histones that affect gene expression without altering the DNA sequence (Russell 2010 p. 475). Two of the most characterized epigenetic modifications are DNA methylation and histone modification. Epigenetic modifications play an important role in gene expression and regulation, and are involved in numerous cellular processes such as in differentiation/development and tumorigenesis.[70] The study of epigenetics on a global level has been made possible only recently through the adaptation of genomic high-throughput assays.[71]
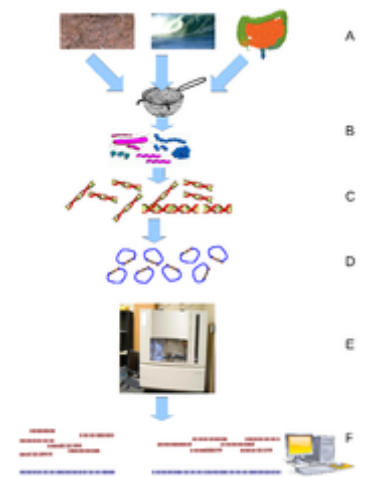
Metagenomics

Metagenomics is the study of *metagenomes*, genetic material recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, ecogenomics or community genomics. While traditional microbiology and microbial genome sequencing rely upon cultivated clonal cultures, early environmental gene sequencing cloned specific genes (often the 16S rRNA gene) to produce a profile of diversity in a natural sample. Such work revealed that the vast majority of microbial biodiversity had been missed by cultivation-based methods.[72] Recent studies use "shotgun" Sanger sequencing or massively parallel pyrosequencing to get largely unbiased samples of all genes from all the members of the sampled communities.[73] Because of its power to reveal the previously hidden diversity of microscopic life, metagenomics offers a powerful lens for viewing the microbial world that has the potential to revolutionize understanding of the entire living world.[74][75]

Environmental Shotgun Sequencing (ESS) is a key technique in metagenomics. (A) Sampling from habitat; (B) filtering particles, typically by size; (C) Lysis and DNA extraction; (D) cloning and library construction; (E) sequencing the clones; (F) sequence assembly into contigs and scaffolds.

Model systems

Viruses and bacteriophages

Bacteriophages have played and continue to play a key role in bacterial genetics and molecular biology. Historically, they were used to define gene structure and gene regulation. Also the first genome to be sequenced was a bacteriophage. However, bacteriophage research did not lead the genomics revolution, which is clearly dominated by bacterial genomics. Only very recently has the study of bacteriophage genomes become prominent, thereby enabling researchers to understand the mechanisms underlying phage evolution. Bacteriophage genome sequences can be obtained through direct sequencing of isolated bacteriophages, but can also be derived as part of microbial genomes. Analysis of bacterial genomes has shown that a substantial amount of microbial DNA consists of prophage sequences and prophage-like elements.[76] A detailed database mining of these sequences offers insights into the role of prophages in shaping the bacterial genome: Overall, this method verified many known bacteriophage groups, making this a useful tool for predicting the relationships of prophages from bacterial genomes.[77][78]

Cyanobacteria

At present there are 24 cyanobacteria for which a total genome sequence is available. 15 of these cyanobacteria come from the marine environment. These are six *Prochlorococcus* strains, seven marine *Synechococcus* strains, *Trichodesmium erythraeum* IMS101 and *Crocosphaera watsonii* WH8501. Several studies have demonstrated how these sequences could be used very successfully to infer important ecological and physiological characteristics of marine cyanobacteria. However, there are many more genome projects currently in progress, amongst those there are further *Prochlorococcus* and marine *Synechococcus* isolates, *Acaryochloris* and *Prochloron*, the N$_2$-fixing filamentous cyanobacteria *Nodularia spumigena*, *Lyngbya aestuarii* and *Lyngbya majuscula*, as well as bacteriophages infecting marine cyanobaceria. Thus, the growing body of genome information can also be tapped in a more general way to address global problems by applying a comparative approach. Some new and exciting examples of progress in this field are the identification of genes for regulatory RNAs, insights into the evolutionary origin of photosynthesis, or estimation of the contribution of horizontal gene transfer to the genomes that have been analyzed.[79]

Applications of genomics

Genomics has provided applications in many fields, including medicine, biotechnology, anthropology and other social sciences.[39]

Genomic medicine

Next-generation genomic technologies allow clinicians and biomedical researchers to drastically increase the amount of genomic data collected on large study populations.[80] When combined with new informatics approaches that integrate many kinds of data with genomic data in disease research, this allows researchers to better understand the genetic bases of drug response and disease.[81][82] Early efforts to apply the genome to medicine included those by a Stanford team led by Euan Ashley who developed the first tools for the medical interpretation of a human genome.[83][84][85] For example, the *All of Us* research program aims to collect genome sequence data from 1 million participants to become a critical component of the precision medicine research platform.[86]

Synthetic biology and bioengineering

The growth of genomic knowledge has enabled increasingly sophisticated applications of synthetic biology.[87] In 2010 researchers at the J. Craig Venter Institute announced the creation of a partially synthetic species of bacterium, *Mycoplasma laboratorium*, derived from the genome of *Mycoplasma genitalium*.[88]

Population and conservation genomics

Population genomics has developed as a popular field of research, where genomic sequencing methods are used to conduct large-scale comparisons of DNA sequences among populations - beyond the limits of genetic markers such as short-range PCR products or microsatellites traditionally used in population genetics. Population genomics studies genome-wide effects to improve our understanding of microevolution so that we may learn the phylogenetic history and demography of a population.[89] Population genomic methods are used for many different fields including evolutionary biology, ecology, biogeography, conservation biology and fisheries management. Similarly, **landscape genomics** has developed from landscape genetics to use genomic methods to identify relationships between patterns of environmental and genetic variation.

Conservationists can use the information gathered by genomic sequencing in order to better evaluate genetic factors key to species conservation, such as the genetic diversity of a population or whether an individual is heterozygous for a recessive inherited genetic disorder.[90] By using genomic data to evaluate the effects of evolutionary processes and to detect patterns in variation throughout a given population, conservationists can formulate plans to aid a given species without as many variables left unknown as those unaddressed by standard genetic approaches.[91]

See also

- Cognitive genomics
- Computational genomics
- Epigenomics
- Functional genomics
- GeneCalling, an mRNA profiling technology
- Genomics of domestication
- Genetics in fiction
- Glycomics
- Immunomics
- Metagenomics
- Pathogenomics
- Personal genomics
- Proteomics
- Transcriptomics
- Psychogenomics
- Whole genome sequencing
- Thomas Roderick

References

1. "WHO definitions of genetics and genomics" (https://www.who.int/genomics/geneticsVSgenomics/en/). World Health Organization.
2. *Concepts of genetics* (10th ed.). San Francisco: Pearson Education. 2012. ISBN 978-0-321-72412-0.
3. Culver KW, Labow MA (8 November 2002). "Genomics" (https://archive.org/details/genetics0000unse). In Robinson R (ed.). *Genetics*. Macmillan Science Library. Macmillan Reference USA. ISBN 978-0-02-865606-9.
4. Kadakkuzha BM, Puthanveettil SV (July 2013). "Genomics and proteomics in solving brain complexity" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6425491). *Molecular BioSystems*. **9** (7): 1807–21. doi:10.1039/C3MB25391K (https://doi.org/10.1039%2FC3MB25391K). PMC 6425491 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6425491). PMID 23615871 (https://pubmed.ncbi.nlm.nih.gov/23615871).
5. Pevsner J (2009). *Bioinformatics and functional genomics* (https://archive.org/details/bioinformaticsfu00pevs_0) (2nd ed.). Hoboken, NJ: Wiley-Blackwell. ISBN 978-0-470-08585-1.
6. Liddell HG, Scott R (1889). *Intermediate Greek-English Lexicon γίγνομαι* (https://web.archive.org/web/20180620231907/http://openlitreview.com/index.php/Genetic_engineering). Oxford: Clarendon Press. ISBN 978-1-61427-397-4. Archived from the original (http://openlitreview.com/index.php/Genetic_engineering) on 2018-06-20. Retrieved 2015-05-13.
7. "Genome, n" (http://www.oed.com/view/Entry/7761). *Oxford English Dictionary* (Third ed.). Oxford University Press. 2008. Retrieved 2012-12-01.(subscription required)

8. Yadav SP (December 2007). "The wholeness in suffix -omics, -omes, and the word om" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2392988). *Journal of Biomolecular Techniques*. **18** (5): 277. PMC 2392988 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2392988). PMID 18166670 (https://pubmed.ncbi.nlm.nih.gov/18166670).

9. Ankeny RA (June 2003). "Sequencing the genome from nematode to human: changing methods, changing science". *Endeavour*. **27** (2): 87–92. doi:10.1016/S0160-9327(03)00061-9 (https://doi.org/10.1016%2FS0160-9327%2803%2900061-9). PMID 12798815 (https://pubmed.ncbi.nlm.nih.gov/12798815).

10. Holley RW, Everett GA, Madison JT, Zamir A (May 1965). "Nucleotide sequences in the yeast alanine transfer ribonucleic acid" (http://www.jbc.org/content/240/5/2122.full.pdf) (PDF). *The Journal of Biological Chemistry*. **240** (5): 2122–8. doi:10.1016/S0021-9258(18)97435-1 (https://doi.org/10.1016%2FS0021-9258%2818%2997435-1). PMID 14299636 (https://pubmed.ncbi.nlm.nih.gov/14299636).

11. Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR, Zamir A (March 1965). "Structure of a ribonucleic acid". *Science*. **147** (3664): 1462–5. Bibcode:1965Sci...147.1462H (https://ui.adsabs.harvard.edu/abs/1965Sci...147.1462H). doi:10.1126/science.147.3664.1462 (https://doi.org/10.1126%2Fscience.147.3664.1462). PMID 14263761 (https://pubmed.ncbi.nlm.nih.gov/14263761). S2CID 40989800 (https://api.semanticscholar.org/CorpusID:40989800).

12. Nirenberg M, Leder P, Bernfield M, Brimacombe R, Trupin J, Rottman F, O'Neal C (May 1965). "RNA codewords and protein synthesis, VII. On the general nature of the RNA code" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC301388). *Proceedings of the National Academy of Sciences of the United States of America*. **53** (5): 1161–8. Bibcode:1965PNAS...53.1161N (https://ui.adsabs.harvard.edu/abs/1965PNAS...53.1161N). doi:10.1073/pnas.53.5.1161 (https://doi.org/10.1073%2Fpnas.53.5.1161). PMC 301388 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC301388). PMID 5330357 (https://pubmed.ncbi.nlm.nih.gov/5330357).

13. Min Jou W, Haegeman G, Ysebaert M, Fiers W (May 1972). "Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein". *Nature*. **237** (5350): 82–8. Bibcode:1972Natur.237...82J (https://ui.adsabs.harvard.edu/abs/1972Natur.237...82J). doi:10.1038/237082a0 (https://doi.org/10.1038%2F237082a0). PMID 4555447 (https://pubmed.ncbi.nlm.nih.gov/4555447). S2CID 4153893 (https://api.semanticscholar.org/CorpusID:4153893).

14. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, et al. (April 1976). "Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene". *Nature*. **260** (5551): 500–7. Bibcode:1976Natur.260..500F (https://ui.adsabs.harvard.edu/abs/1976Natur.260..500F). doi:10.1038/260500a0 (https://doi.org/10.1038%2F260500a0). PMID 1264203 (https://pubmed.ncbi.nlm.nih.gov/1264203). S2CID 4289674 (https://api.semanticscholar.org/CorpusID:4289674).

15. Fiers W, Contreras R, Haegemann G, Rogiers R, Van de Voorde A, Van Heuverswyn H, Van Herreweghe J, Volckaert G, Ysebaert M (May 1978). "Complete nucleotide sequence of SV40 DNA". *Nature*. **273** (5658): 113–20. Bibcode:1978Natur.273..113F (https://ui.adsabs.harvard.edu/abs/1978Natur.273..113F). doi:10.1038/273113a0 (https://doi.org/10.1038%2F273113a0). PMID 205802 (https://pubmed.ncbi.nlm.nih.gov/205802). S2CID 1634424 (https://api.semanticscholar.org/CorpusID:1634424).

16. Tamarin RH (2004). *Principles of genetics* (7 ed.). London: McGraw Hill. ISBN 978-0-07-124320-9.

17. Sanger F (1980). "Nobel lecture: Determination of nucleotide sequences in DNA" (http://nobelprize.org/nobel_prizes/chemistry/laureates/1980/sanger-lecture.pdf) (PDF). Nobelprize.org. Retrieved 2010-10-18.

18. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M (February 1977). "Nucleotide sequence of bacteriophage phi X174 DNA". *Nature*. **265** (5596): 687–95. Bibcode:1977Natur.265..687S (https://ui.adsabs.harvard.edu/abs/1977Natur.265..687S). doi:10.1038/265687a0 (https://doi.org/10.1038%2F265687a0). PMID 870828 (https://pubmed.ncbi.nlm.nih.gov/870828). S2CID 4206886 (https://api.semanticscholar.org/CorpusID:4206886).

19. Kaiser O, Bartels D, Bekel T, Goesmann A, Kespohl S, Pühler A, Meyer F (December 2003). "Whole genome shotgun sequencing guided by bioinformatics pipelines--an optimized approach for an established technique". *Journal of Biotechnology*. **106** (2–3): 121–33. doi:10.1016/j.jbiotec.2003.08.008 (https://doi.org/10.1016%2Fj.jbiotec.2003.08.008). PMID 14651855 (https://pubmed.ncbi.nlm.nih.gov/14651855).

20. Sanger F, Nicklen S, Coulson AR (December 1977). "DNA sequencing with chain-terminating inhibitors" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC431765). *Proceedings of the National Academy of Sciences of the United States of America*. **74** (12): 5463–7. Bibcode:1977PNAS...74.5463S (https://ui.adsabs.harvard.edu/abs/1977PNAS...74.5463S). doi:10.1073/pnas.74.12.5463 (https://doi.org/10.1073%2Fpnas.74.12.5463). PMC 431765 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC431765). PMID 271968 (https://pubmed.ncbi.nlm.nih.gov/271968).

21. Maxam AM, Gilbert W (February 1977). "A new method for sequencing DNA" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC392330). *Proceedings of the National Academy of Sciences of the United States of America*. **74** (2): 560–4. Bibcode:1977PNAS...74..560M (https://ui.adsabs.harvard.edu/abs/1977PNAS...74..560M). doi:10.1073/pnas.74.2.560 (https://doi.org/10.1073%2Fpnas.74.2.560). PMC 392330 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC392330). PMID 265521 (https://pubmed.ncbi.nlm.nih.gov/265521).

22. Darden L, James Tabery (2010). "Molecular Biology" (http://plato.stanford.edu/archives/fall2010/entries/molecular-biology/). In Zalta EN (ed.). *The Stanford Encyclopedia of Philosophy* (Fall 2010 ed.).

23. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. (April 1981). "Sequence and organization of the human mitochondrial genome". *Nature*. **290** (5806): 457–65. Bibcode:1981Natur.290..457A (https://ui.adsabs.harvard.edu/abs/1981Natur.290..457A). doi:10.1038/290457a0 (https://doi.org/10.1038%2F290457a0). PMID 7219534 (https://pubmed.ncbi.nlm.nih.gov/7219534). S2CID 4355527 (https://api.semanticscholar.org/CorpusID:4355527).(subscription required)

24. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, et al. (September 1986). "The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1167080). *The EMBO Journal*. **5** (9): 2043–2049. doi:10.1002/j.1460-2075.1986.tb04464.x (https://doi.org/10.1002%2Fj.1460-2075.1986.tb04464.x). PMC 1167080 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1167080). PMID 16453699 (https://pubmed.ncbi.nlm.nih.gov/16453699).

25. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, et al. (1986). "Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA". *Nature*. **322** (6079): 572–574. Bibcode:1986Natur.322..572O (https://ui.adsabs.harvard.edu/abs/1986Natur.322..572O). doi:10.1038/322572a0 (https://doi.org/10.1038%2F322572a0). S2CID 4311952 (https://api.semanticscholar.org/CorpusID:4311952).

26. Oliver SG, van der Aart QJ, Agostoni-Carbone ML, Aigle M, Alberghina L, Alexandraki D, Antoine G, Anwar R, Ballesta JP, Benit P (May 1992). "The complete DNA sequence of yeast chromosome III". *Nature*. **357** (6373): 38–46. Bibcode:1992Natur.357...38O (https://ui.adsabs.harvard.edu/abs/1992Natur.357...38O). doi:10.1038/357038a0 (https://doi.org/10.1038%2F357038a0). PMID 1574125 (https://pubmed.ncbi.nlm.nih.gov/1574125). S2CID 4271784 (https://api.semanticscholar.org/CorpusID:4271784).

27. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. (July 1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd". *Science*. **269** (5223): 496–512. Bibcode:1995Sci...269..496F (https://ui.adsabs.harvard.edu/abs/1995Sci...269..496F). doi:10.1126/science.7542800 (https://doi.org/10.1126%2Fscience.7542800). PMID 7542800 (https://pubmed.ncbi.nlm.nih.gov/7542800). S2CID 10423613 (https://api.semanticscholar.org/CorpusID:10423613).

28. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (October 1996). "Life with 6000 genes". *Science*. **274** (5287): 546, 563–7. Bibcode:1996Sci...274..546G (https://ui.adsabs.harvard.edu/abs/1996Sci...274..546G). doi:10.1126/science.274.5287.546 (https://doi.org/10.1126%2Fscience.274.5287.546). PMID 8849441 (https://pubmed.ncbi.nlm.nih.gov/8849441). S2CID 211123134 (https://api.semanticscholar.org/CorpusID:211123134).(subscription required)

29. "Complete genomes: Viruses" (https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239). *NCBI*. 17 November 2011. Retrieved 2011-11-18.

30. "Genome Project Statistics" (https://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html). *Entrez Genome Project*. 7 October 2011. Retrieved 2011-11-18.

31. Zimmer C (29 December 2009). "Scientists Start a Genomic Catalog of Earth's Abundant Microbes" (https://www.nytimes.com/2009/12/29/science/29microbes.html). *The New York Times*. ISSN 0362-4331 (https://www.worldcat.org/issn/0362-4331). Retrieved 2012-12-21.

32. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, et al. (December 2009). "A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3073058). *Nature*. **462** (7276): 1056–60. Bibcode:2009Natur.462.1056W (https://ui.adsabs.harvard.edu/abs/2009Natur.462.1056W). doi:10.1038/nature08656 (https://doi.org/10.1038%2Fnature08656). PMC 3073058 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3073058). PMID 20033048 (https://pubmed.ncbi.nlm.nih.gov/20033048).

33. "Human gene number slashed" (http://news.bbc.co.uk/2/hi/science/nature/3760766.stm). *BBC*. 20 October 2004. Retrieved 2012-12-21.

34. Yue GH, Lo LC, Zhu ZY, Lin G, Feng F (April 2006). "The complete nucleotide sequence of the mitochondrial genome of Tetraodon nigroviridis". *DNA Sequence*. **17** (2): 115–21. doi:10.1080/10425170600700378 (https://doi.org/10.1080%2F10425170600700378). PMID 17076253 (https://pubmed.ncbi.nlm.nih.gov/17076253). S2CID 21797344 (https://api.semanticscholar.org/CorpusID:21797344).

35. National Human Genome Research Institute (14 July 2004). "Dog Genome Assembled: Canine Genome Now Available to Research Community Worldwide" (http://www.genome.gov/12511476). *Genome.gov*. Retrieved 2012-01-20.

36. McElheny V (2010). *Drawing the map of life : inside the Human Genome Project* (https://archive.org/details/drawingmapoflife0000mcel). New York NY: Basic Books. ISBN 978-0-465-04333-0.

37. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA (November 2012). "An integrated map of genetic variation from 1,092 human genomes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3498066). *Nature*. **491** (7422): 56–65. Bibcode:2012Natur.491...56T (https://ui.adsabs.harvard.edu/abs/2012Natur.491...56T). doi:10.1038/nature11632 (https://doi.org/10.1038%2Fnature11632). PMC 3498066 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3498066). PMID 23128226 (https://pubmed.ncbi.nlm.nih.gov/23128226).

38. Nielsen R (October 2010). "Genomics: In search of rare human variants" (https://doi.org/10.1038%2F4671050a). *Nature*. **467** (7319): 1050–1. Bibcode:2010Natur.467.1050N (https://ui.adsabs.harvard.edu/abs/2010Natur.467.1050N). doi:10.1038/4671050a (https://doi.org/10.1038%2F4671050a). PMID 20981085 (https://pubmed.ncbi.nlm.nih.gov/20981085).

39. Barnes B, Dupré J (2008). *Genomes and what to make of them*. Chicago: University of Chicago Press. ISBN 978-0-226-17295-8.

40. Eisen JA (July 2012). "Badomics words and the power and peril of the ome-meme" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3617454). *GigaScience*. **1** (1): 6. doi:10.1186/2047-217X-1-6 (https://doi.org/10.1186%2F2047-217X-1-6). PMC 3617454 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3617454). PMID 23587201 (https://pubmed.ncbi.nlm.nih.gov/23587201).

41. Hotz RL (13 August 2012). "Here"s an Omical Tale: Scientists Discover Spreading Suffix" (https://www.wsj.com/articles/SB10000872396390444840104577551433143153716). *Wall Street Journal*. ISSN 0099-9660 (https://www.worldcat.org/issn/0099-9660). Retrieved 2013-01-04.

42. Scudellari M (1 October 2011). "Data Deluge" (http://www.the-scientist.com/?articles.view/articleNo/31212/title/Data-Deluge/). *The Scientist*. Retrieved 2013-01-04.

43. Chaston J, Douglas AE (August 2012). "Making the most of "omics" for symbiosis research" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3491573). *The Biological Bulletin*. **223** (1): 21–9. doi:10.1086/BBLv223n1p21 (https://doi.org/10.1086%2FBBLv223n1p21). PMC 3491573 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3491573). PMID 22983030 (https://pubmed.ncbi.nlm.nih.gov/22983030).

44. McCutcheon JP, von Dohlen CD (August 2011). "An interdependent metabolic patchwork in the nested symbiosis of mealybugs" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169327). *Current Biology*. **21** (16): 1366–72. doi:10.1016/j.cub.2011.06.051 (https://doi.org/10.1016%2Fj.cub.2011.06.051). PMC 3169327 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169327). PMID 21835622 (https://pubmed.ncbi.nlm.nih.gov/21835622).

45. Baker M (14 September 2012). "Benchtop sequencers ship off" (http://blogs.nature.com/news/2012/09/benchtop-sequencers-ship-off.html) (Blog). *Nature News Blog*. Retrieved 2012-12-22.

46. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (July 2012). "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3431227). *BMC Genomics*. **13**: 341. doi:10.1186/1471-2164-13-341 (https://doi.org/10.1186%2F1471-2164-13-341). PMC 3431227 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3431227). PMID 22827831 (https://pubmed.ncbi.nlm.nih.gov/22827831).

47. Staden R (June 1979). "A strategy of DNA sequencing employing computer programs" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC327874). *Nucleic Acids Research*. **6** (7): 2601–10. doi:10.1093/nar/6.7.2601 (https://doi.org/10.1093%2Fnar%2F6.7.2601). PMC 327874 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC327874). PMID 461197 (https://pubmed.ncbi.nlm.nih.gov/461197).

48. Anderson S (July 1981). "Shotgun DNA sequencing using cloned DNase I-generated fragments" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC327328). *Nucleic Acids Research*. **9** (13): 3015–27. doi:10.1093/nar/9.13.3015 (https://doi.org/10.1093%2Fnar%2F9.13.3015). PMC 327328 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC327328). PMID 6269069 (https://pubmed.ncbi.nlm.nih.gov/6269069).

49. Pop M (July 2009). "Genome assembly reborn: recent computational challenges" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2691937). *Briefings in Bioinformatics*. **10** (4): 354–66. doi:10.1093/bib/bbp026 (https://doi.org/10.1093%2Fbib%2Fbbp026). PMC 2691937 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2691937). PMID 19482960 (https://pubmed.ncbi.nlm.nih.gov/19482960).

50. Sanger F, Coulson AR (May 1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase". *Journal of Molecular Biology*. **94** (3): 441–8. doi:10.1016/0022-2836(75)90213-2 (https://doi.org/10.1016%2F0022-2836%2875%2990213-2). PMID 1100841 (https://pubmed.ncbi.nlm.nih.gov/1100841).

51. Mavromatis K, Land ML, Brettin TS, Quest DJ, Copeland A, Clum A, et al. (2012). Liu Z (ed.). "The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3520994). *PLOS ONE*. **7** (12): e48837. Bibcode:2012PLoSO...748837M (https://ui.adsabs.harvard.edu/abs/2012PLoSO...748837M). doi:10.1371/journal.pone.0048837 (https://doi.org/10.1371%2Fjournal.pone.0048837). PMC 3520994 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3520994). PMID 23251337 (https://pubmed.ncbi.nlm.nih.gov/23251337).

52. Illumina, Inc. (28 February 2012). *An Introduction to Next-Generation Sequencing Technology* (http://www.illumina.com/Documents/products/Illumina_Sequencing_Introduction.pdf) (PDF). San Diego, California, USA: Illumina, Inc. p. 12. Retrieved 2012-12-28.

53. Hall N (May 2007). "Advanced sequencing technologies and their wider impact in microbiology" (https://doi.org/10.1242%2Fjeb.001370). *The Journal of Experimental Biology*. **210** (Pt 9): 1518–25. doi:10.1242/jeb.001370 (https://doi.org/10.1242%2Fjeb.001370). PMID 17449817 (https://pubmed.ncbi.nlm.nih.gov/17449817).

54. Church GM (January 2006). "Genomes for all". *Scientific American*. **294** (1): 46–54. Bibcode:2006SciAm.294a..46C (https://ui.adsabs.harvard.edu/abs/2006SciAm.294a..46C). doi:10.1038/scientificamerican0106-46 (https://doi.org/10.1038%2Fscientificamerican0106-46). PMID 16468433 (https://pubmed.ncbi.nlm.nih.gov/16468433).

55. ten Bosch JR, Grody WW (November 2008). "Keeping up with the next generation: massively parallel sequencing in clinical diagnostics" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2570630). *The Journal of Molecular Diagnostics*. **10** (6): 484–92. doi:10.2353/jmoldx.2008.080027 (https://doi.org/10.2353%2Fjmoldx.2008.080027). PMC 2570630 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2570630). PMID 18832462 (https://pubmed.ncbi.nlm.nih.gov/18832462).

56. Tucker T, Marra M, Friedman JM (August 2009). "Massively parallel sequencing: the next big thing in genetic medicine" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2725244). *American Journal of Human Genetics*. **85** (2): 142–54. doi:10.1016/j.ajhg.2009.06.022 (https://doi.org/10.1016%2Fj.ajhg.2009.06.022). PMC 2725244 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2725244). PMID 19679224 (https://pubmed.ncbi.nlm.nih.gov/19679224).

57. Kawashima EH, Farinelli L, Mayer P (12 May 2005). "Method of nucleic acid amplification" (http://www.google.com/patents/US20050100900). Retrieved 2012-12-22.

58. Mardis ER (2008). "Next-generation DNA sequencing methods" (https://web.archive.org/web/20130518193940/http://www.lcg.unam.mx/frontiers/files/frontiers/annurev.genom.9.081307.164359.pdf) (PDF). *Annual Review of Genomics and Human Genetics*. **9**: 387–402. doi:10.1146/annurev.genom.9.081307.164359 (https://doi.org/10.1146%2Fannurev.genom.9.081307.164359). PMID 18576944 (https://pubmed.ncbi.nlm.nih.gov/18576944). Archived from the original (http://www.lcg.unam.mx/frontiers/files/frontiers/annurev.genom.9.081307.164359.pdf) (PDF) on 2013-05-18. Retrieved 2013-01-04.

59. Davies K (2011). "Powering Preventative Medicine" (http://www.bio-itworld.com/issues/2011/sept-oct/powering-preventative-medicine.html). *Bio-IT World* (September–October).

60. https://www.pacb.com/

61. "Oxford Nanopore Technologies" (https://nanoporetech.com/).

62. Chain PS, Grafham DV, Fulton RS, Fitzgerald MG, Hostetler J, Muzny D, et al. (October 2009). "Genomics. Genome project standards in a new era of sequencing" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3854948). *Science*. **326** (5950): 236–7. Bibcode:2009Sci...326..236C (https://ui.adsabs.harvard.edu/abs/2009Sci...326..236C). doi:10.1126/science.1180614 (https://doi.org/10.1126%2Fscience.1180614). PMC 3854948 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3854948). PMID 19815760 (https://pubmed.ncbi.nlm.nih.gov/19815760).

63. Stein L (July 2001). "Genome annotation: from sequence to biology". *Nature Reviews. Genetics*. **2** (7): 493–503. doi:10.1038/35080529 (https://doi.org/10.1038%2F35080529). PMID 11433356 (https://pubmed.ncbi.nlm.nih.gov/11433356). S2CID 12044602 (https://api.semanticscholar.org/CorpusID:12044602).

64. Brent MR (January 2008). "Steady progress and recent breakthroughs in the accuracy of automated genome annotation" (https://web.archive.org/web/20130529013309/http://www.genomics.arizona.edu/553/Readings/2012/Brent_2008.pdf) (PDF). *Nature Reviews. Genetics*. **9** (1): 62–73. doi:10.1038/nrg2220 (https://doi.org/10.1038%2Fnrg2220). PMID 18087260 (https://pubmed.ncbi.nlm.nih.gov/18087260). S2CID 20412451 (https://api.semanticscholar.org/CorpusID:20412451). Archived from the original (http://www.genomics.arizona.edu/553/Readings/2012/Brent_2008.pdf) (PDF) on 2013-05-29. Retrieved 2013-01-04.

65. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, et al. (January 2013). "Ensembl 2013" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531136). *Nucleic Acids Research*. **41** (Database issue): D48–55. doi:10.1093/nar/gks1236 (https://doi.org/10.1093%2Fnar%2Fgks1236). PMC 3531136 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531136). PMID 23203987 (https://pubmed.ncbi.nlm.nih.gov/23203987).

66. Keith JM (2008). Keith JM (ed.). *Bioinformatics*. Methods in Molecular Biology. **453**. pp. v–vi. doi:10.1007/978-1-60327-429-6 (https://doi.org/10.1007%2F978-1-60327-429-6). ISBN 978-1-60327-428-9. PMID 18720577 (https://pubmed.ncbi.nlm.nih.gov/18720577).

67. Marsden RL, Lewis TA, Orengo CA (March 2007). "Towards a comprehensive structural coverage of completed genomes: a structural genomics viewpoint" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1829165). *BMC Bioinformatics*. **8**: 86. doi:10.1186/1471-2105-8-86 (https://doi.org/10.1186%2F1471-2105-8-86). PMC 1829165 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1829165). PMID 17349043 (https://pubmed.ncbi.nlm.nih.gov/17349043).

68. Brenner SE, Levitt M (January 2000). "Expectations from structural genomics" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2144435). *Protein Science*. **9** (1): 197–200. doi:10.1110/ps.9.1.197 (https://doi.org/10.1110%2Fps.9.1.197). PMC 2144435 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2144435). PMID 10739263 (https://pubmed.ncbi.nlm.nih.gov/10739263).

69. Brenner SE (October 2001). "A tour of structural genomics" (http://compbio.berkeley.edu/people/brenner/pubs/brenner-2001-nrg-strgen.pdf) (PDF). *Nature Reviews. Genetics*. **2** (10): 801–9. doi:10.1038/35093574 (https://doi.org/10.1038%2F35093574). PMID 11584296 (https://pubmed.ncbi.nlm.nih.gov/11584296). S2CID 5656447 (https://api.semanticscholar.org/CorpusID:5656447).

70. Francis RC (2011). *Epigenetics : the ultimate mystery of inheritance*. New York: WW Norton. ISBN 978-0-393-07005-7.

71. Laird PW (March 2010). "Principles and challenges of genomewide DNA methylation analysis". *Nature Reviews. Genetics*. **11** (3): 191–203. doi:10.1038/nrg2732 (https://doi.org/10.1038%2Fnrg2732). PMID 20125086 (https://pubmed.ncbi.nlm.nih.gov/20125086). S2CID 6780101 (https://api.semanticscholar.org/CorpusID:6780101).

72. Hugenholtz P, Goebel BM, Pace NR (September 1998). "Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107498). *Journal of Bacteriology*. **180** (18): 4765–74. doi:10.1128/JB.180.18.4765-4774.1998 (https://doi.org/10.1128%2FJB.180.18.4765-4774.1998). PMC 107498 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107498). PMID 9733676 (https://pubmed.ncbi.nlm.nih.gov/9733676).

73. Eisen JA (March 2007). "Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1821061). *PLOS Biology*. **5** (3): e82. doi:10.1371/journal.pbio.0050082 (https://doi.org/10.1371%2Fjournal.pbio.0050082). PMC 1821061 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1821061). PMID 17355177 (https://pubmed.ncbi.nlm.nih.gov/17355177).

74. Marco D, ed. (2010). *Metagenomics: Theory, Methods and Applications*. Caister Academic Press. ISBN 978-1-904455-54-7.

75. Marco D, ed. (2011). *Metagenomics: Current Innovations and Future Trends*. Caister Academic Press. ISBN 978-1-904455-87-5.

76. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H (June 2003). "Prophage genomics" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC156470). *Microbiology and Molecular Biology Reviews*. **67** (2): 238–76, table of contents. doi:10.1128/MMBR.67.2.238-276.2003 (https://doi.org/10.1128%2FMMBR.67.2.238-276.2003). PMC 156470 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC156470). PMID 12794192 (https://pubmed.ncbi.nlm.nih.gov/12794192).

77. McGrath S, van Sinderen D, eds. (2007). *Bacteriophage: Genetics and Molecular Biology* (http://www.horizonpress.com/phage) (1st ed.). Caister Academic Press. ISBN 978-1-904455-14-1.

78. Fouts DE (November 2006). "Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1635311). *Nucleic Acids Research*. **34** (20): 5839–51. doi:10.1093/nar/gkl732 (https://doi.org/10.1093%2Fnar%2Fgkl732). PMC 1635311 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1635311). PMID 17062630 (https://pubmed.ncbi.nlm.nih.gov/17062630).

79. Herrero A, Flores E, eds. (2008). *The Cyanobacteria: Molecular Biology, Genomics and Evolution* (http://www.horizonpress.com/cyan) (1st ed.). Caister Academic Press. ISBN 978-1-904455-15-8.

80. Hudson KL (September 2011). "Genomics, health care, and society". *The New England Journal of Medicine*. **365** (11): 1033–41. doi:10.1056/NEJMra1010517 (https://doi.org/10.1056%2FNEJMra1010517). PMID 21916641 (https://pubmed.ncbi.nlm.nih.gov/21916641).

81. O'Donnell CJ, Nabel EG (December 2011). "Genomics of cardiovascular disease". *The New England Journal of Medicine*. **365** (22): 2098–109. doi:10.1056/NEJMra1105239 (https://doi.org/10.1056%2FNEJMra1105239). PMID 22129254 (https://pubmed.ncbi.nlm.nih.gov/22129254).

82. Lu YF, Goldstein DB, Angrist M, Cavalleri G (July 2014). "Personalized medicine and human genetic diversity" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4143101). *Cold Spring Harbor Perspectives in Medicine*. **4** (9): a008581. doi:10.1101/cshperspect.a008581 (https://doi.org/10.1101%2Fcshperspect.a008581). PMC 4143101 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4143101). PMID 25059740 (https://pubmed.ncbi.nlm.nih.gov/25059740).

83. Ashley, Euan A; Butte, Atul J; Wheeler, Matthew T; Chen, Rong; Klein, Teri E; Dewey, Frederick E; Dudley, Joel T; Ormond, Kelly E; Pavlovic, Aleksandra; Morgan, Alexander A; Pushkarev, Dmitry; Neff, Norma F; Hudgins, Louanne; Gong, Li; Hodges, Laura M; Berlin, Dorit S; Thorn, Caroline F; Sangkuhl, Katrin; Hebert, Joan M; Woon, Mark; Sagreiya, Hersh; Whaley, Ryan; Knowles, Joshua W; Chou, Michael F; Thakuria, Joseph V; Rosenbaum, Abraham M; Zaranek, Alexander Wait; Church, George M; Greely, Henry T; Quake, Stephen R; Altman, Russ B (May 2010). "Clinical assessment incorporating a personal genome" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2937184). *The Lancet*. **375** (9725): 1525–1535. doi:10.1016/S0140-6736(10)60452-7 (https://doi.org/10.1016%2FS0140-6736%2810%2960452-7). PMC 2937184 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2937184). PMID 20435227 (https://pubmed.ncbi.nlm.nih.gov/20435227).

84. Dewey, Frederick E.; Chen, Rong; Cordero, Sergio P.; Ormond, Kelly E.; Caleshu, Colleen; Karczewski, Konrad J.; Whirl-Carrillo, Michelle; Wheeler, Matthew T.; Dudley, Joel T.; Byrnes, Jake K.; Cornejo, Omar E.; Knowles, Joshua W.; Woon, Mark; Sangkuhl, Katrin; Gong, Li; Thorn, Caroline F.; Hebert, Joan M.; Capriotti, Emidio; David, Sean P.; Pavlovic, Aleksandra; West, Anne; Thakuria, Joseph V.; Ball, Madeleine P.; Zaranek, Alexander W.; Rehm, Heidi L.; Church, George M.; West, John S.; Bustamante, Carlos D.; Snyder, Michael; Altman, Russ B.; Klein, Teri E.; Butte, Atul J.; Ashley, Euan A. (15 September 2011). "Phased Whole-Genome Genetic Risk in a Family Quartet Using a Major Allele Reference Sequence" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3174201). *PLOS Genetics*. **7** (9): e1002280. doi:10.1371/journal.pgen.1002280 (https://doi.org/10.1371%2Fjournal.pgen.1002280). PMC 3174201 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3174201). PMID 21935354 (https://pubmed.ncbi.nlm.nih.gov/21935354).

85. Dewey, Frederick E.; Grove, Megan E.; Pan, Cuiping; Goldstein, Benjamin A.; Bernstein, Jonathan A.; Chaib, Hassan; Merker, Jason D.; Goldfeder, Rachel L.; Enns, Gregory M.; David, Sean P.; Pakdaman, Neda; Ormond, Kelly E.; Caleshu, Colleen; Kingham, Kerry; Klein, Teri E.; Whirl-Carrillo, Michelle; Sakamoto, Kenneth; Wheeler, Matthew T.; Butte, Atul J.; Ford, James M.; Boxer, Linda; Ioannidis, John P. A.; Yeung, Alan C.; Altman, Russ B.; Assimes, Themistocles L.; Snyder, Michael; Ashley, Euan A.; Quertermous, Thomas (12 March 2014). "Clinical Interpretation and Implications of Whole-Genome Sequencing" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4119063). *JAMA*. **311** (10): 1035–45. doi:10.1001/jama.2014.1717 (https://doi.org/10.1001%2Fjama.2014.1717). PMC 4119063 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4119063). PMID 24618965 (https://pubmed.ncbi.nlm.nih.gov/24618965).

86. "NIH-funded genome centers to accelerate precision medicine discoveries" (https://www.nih.gov/news-events/news-releases/nih-funded-genome-centers-accelerate-precision-medicine-discoveries). *National Institutes of Health: All of Us Research Program*. National Institutes of Health.

87. Church GM, Regis E (2012). *Regenesis : how synthetic biology will reinvent nature and ourselves*. New York: Basic Books. ISBN 978-0-465-02175-8.

88. Baker M (May 2011). "Synthetic genomes: The next step for the synthetic genome". *Nature*. **473** (7347): 403, 405–8. Bibcode:2011Natur.473..403B (https://ui.adsabs.harvard.edu/abs/2011Natur.473..403B). doi:10.1038/473403a (https://doi.org/10.1038%2F473403a). PMID 21593873 (https://pubmed.ncbi.nlm.nih.gov/21593873). S2CID 205064528 (https://api.semanticscholar.org/CorpusID:205064528).

89. Luikart, G.; England, P. R.; Tallmon, D.; Jordan S.; Taberlet P. (2003). "The Power and Promise of Population Genomics: From Genotyping to Genome Typing". *Nature Reviews* (4): 981-994

90. Frankham R (1 September 2010). "Challenges and opportunities of genetic approaches to biological conservation". *Biological Conservation*. **143** (9): 1922–1923. doi:10.1016/j.biocon.2010.05.011 (https://doi.org/10.1016%2Fj.biocon.2010.05.011).

91. Allendorf FW, Hohenlohe PA, Luikart G (October 2010). "Genomics and the future of conservation genetics". *Nature Reviews. Genetics*. **11** (10): 697–709. doi:10.1038/nrg2844 (https://doi.org/10.1038%2Fnrg2844). PMID 20847747 (https://pubmed.ncbi.nlm.nih.gov/20847747). S2CID 10811958 (https://api.semanticscholar.org/CorpusID:10811958).

Further reading

- Lesk AM (2017). *Introduction to Genomics* (3rd ed.). New York: Oxford University Press. p. 544. ISBN 978-0-19-107085-3. ASIN 0198754833 (https://www.amazon.co.uk/dp/0198754833).

- Stunnenberg HG, Hubner NC (2014). "Genomics meets proteomics: identifying the culprits in disease" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4021166). *Human Genetics*. **133** (6): 689–700. doi:10.1007/s00439-013-1376-2 (https://doi.org/10.1007%2Fs00439-013-1376-2). PMC 4021166 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4021166). PMID 24135908 (https://pubmed.ncbi.nlm.nih.gov/24135908).

- Shibata T (2012). "Cancer genomics and pathology: all together now" (https://doi.org/10.1111%2Fj.1440-1827.2012.02855.x). *Pathology International*. **62** (10): 647–59. doi:10.1111/j.1440-1827.2012.02855.x (https://doi.org/10.1111%2Fj.1440-1827.2012.02855.x). PMID 23005591 (https://pubmed.ncbi.nlm.nih.gov/23005591). S2CID 27886018 (https://api.semanticscholar.org/CorpusID:27886018).

- Roychowdhury S, Chinnaiyan AM (2016). "Translating cancer genomes and transcriptomes for precision oncology" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4713245). *CA: A Cancer Journal for Clinicians*. **66** (1): 75–88. doi:10.3322/caac.21329 (https://doi.org/10.3322%2Fcaac.21329). PMC 4713245 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4713245). PMID 26528881 (https://pubmed.ncbi.nlm.nih.gov/26528881).

- Vadim N G, Zhang Y (2013). "Chapter 16 Comparative Genomics Analysis of the Metallomes". In Banci L (ed.). *Metallomics and the Cell*. Metal Ions in Life Sciences. **12**. Springer. doi:10.1007/978-94-007-5561-10_16 (https://doi.org/10.1007%2F978-94-007-5561-10_16) (inactive 2021-01-14). ISBN 978-94-007-5560-4. electronic-book ISBN 978-94-007-5561-1 ISSN 1559-0836 (https://www.worldcat.org/search?fq=x0:jrnl&q=n2:1559-0836) electronic-ISSN 1868-0402 (https://www.worldcat.org/search?fq=x0:jrnl&q=n2:1868-0402)

External links

- Annual Review of Genomics and Human Genetics (http://arjournals.annualreviews.org/loi/genom/)
- BMC Genomics (http://www.biomedcentral.com/bmcgenomics/): A BMC journal on Genomics

- Genomics journal (http://www.elsevier.com/wps/find/journaldescription.cws_home/622838/description#description)
- Genomics.org (http://genomics.org): An openfree genomics portal.
- NHGRI (http://www.genome.gov/): US government's genome institute
- JCVI Comprehensive Microbial Resource (http://arquivo.pt/wayback/20160516145036/http://cmr.jcvi.org/)
- KoreaGenome.org (https://www.webcitation.org/5hOBP0LmN?url=http://koreagenome.org/): The first Korean Genome published and the sequence is available freely.
- GenomicsNetwork (https://web.archive.org/web/20171228195616/http://www.genomicsnetwork.ac.uk/): Looks at the development and use of the science and technologies of genomics.
- Institute for Genome Sciences (https://web.archive.org/web/20091115214635/http://www.igs.umaryland.edu/research_topics.php): Genomics research.
- MIT OpenCourseWare HST.512 Genomic Medicine (http://ocw.mit.edu/courses/health-sciences-and-technology/hst-512-genomic-medicine-spring-2004/) A free, self-study course in genomic medicine. Resources include audio lectures and selected lecture notes.
- ENCODE threads explorer (http://www.nature.com/encode/#/threads/machine-learning-approaches-to-genomics) Machine learning approaches to genomics. Nature (journal)
- Global map of genomics laboratories (https://web.archive.org/web/20131005093407/https://mapsengine.google.com/map/viewer?mid=zeHBUH4JmjxQ.kB0PwL4lW7VM)
- Genomics: Scitable by nature education (http://www.nature.com/scitable/ebooks/genomics-4)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Genomics&oldid=1010467608"

This page was last edited on 5 March 2021, at 16:43 (UTC).