Lecture

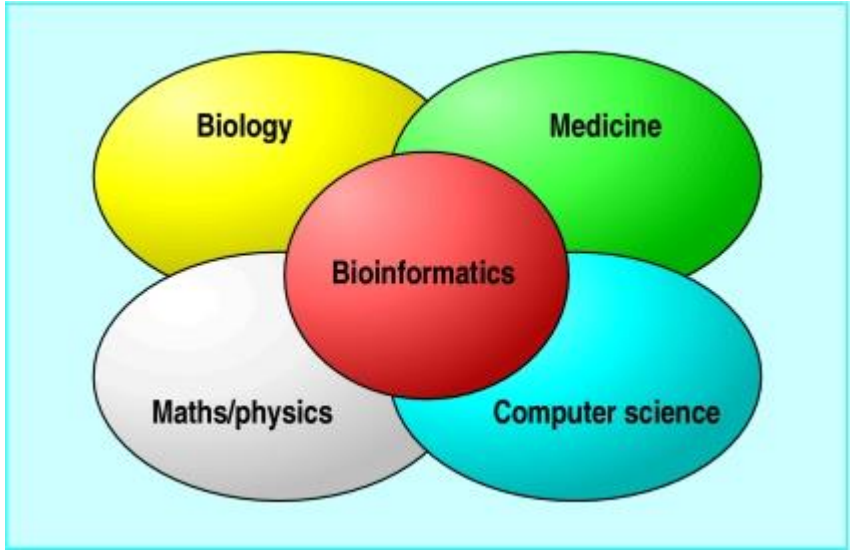
Bioinformatics

Several viewpoints

Definitions vary widely

Bio + informatics ➡ Processing of biological data using information technology.

- Bioinformatics is an interdisciplinary research area at the interface between computer science and biological science.
- A variety of definitions exist in the literature and on the world wide web; some are more inclusive than others.
- Bioinformatics is a union of biology and informatics: bioinformatics involves the technology that uses computers for storage, retrieval, manipulation, and distribution of information related to biological macromolecules such as DNA, RNA, and proteins.



Source: BMJ. 2002 Apr 27; 324(7344): 1018–1022. doi: 10.1136/bmj.324.7344.1018

Bioinformatics is the application of computational technology to handle the rapidly growing repository of information related to molecular biology.

Source:

Essential Bioinformatics by Jin Xiong

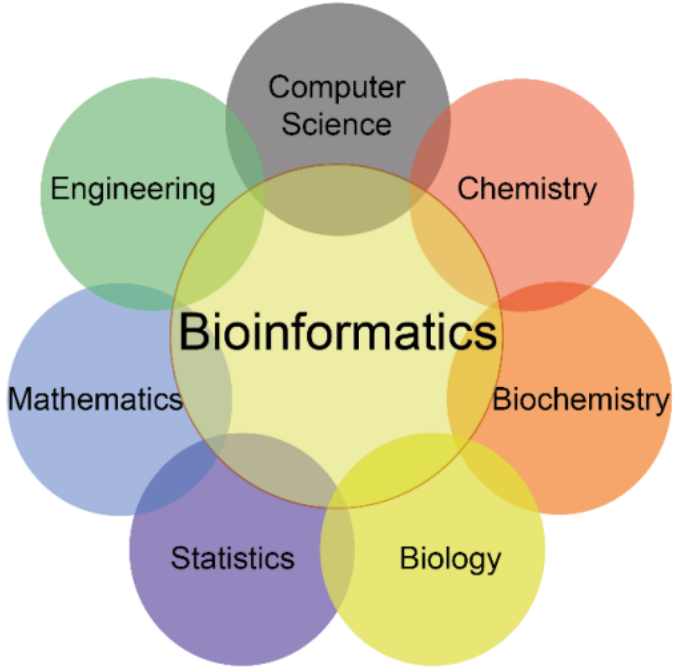
https://www.investopedia.com/terms/b/bioinformatics.asp#:~:text=Bioinformatics%20is%20the%20application%20of,%2C%20biotechnology%2C%20statistics%20and%20engineering.

Bioinformatics

Several viewpoints

Bio + informatics ➡ Processing of biological data using information technology.

- Bioinformatics combines different fields of study, including computer sciences, molecular biology, biotechnology, statistics and engineering.
- It is particularly useful for managing and analyzing large sets of data, such as those generated by the fields of genomics and proteomics.



Source: https://www.cleanpng.com/png-bioinformatics-computer-science-computational-biol-2602217/preview.html

Takeaways

- ✓ Bioinformatics employs computers and information technology to large molecular biology data sets.
- ✓ Bioinformatics is seen as a cutting-edge branch of the biotechnology sector, used for novel drug discovery and personalized medicines.
- ✓ The field closely combines computer science and artificial intelligence with microbiology and genomics.

Source:

Essential Bioinformatics by Jin Xiong

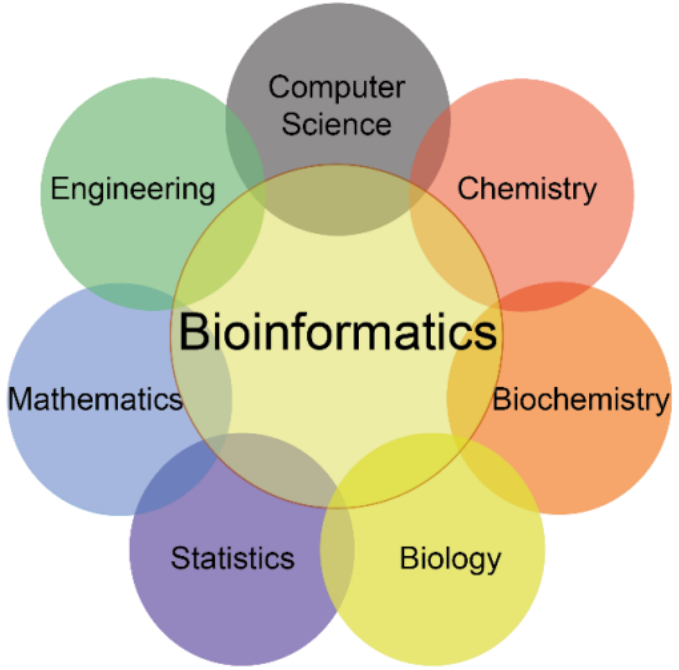
https://www.investopedia.com/terms/b/bioinformatics.asp#:~:text=Bioinformatics%20is%20the%20application%20of,%2C%20biotechnology%2C%20statistics%20and%20engineering.

Bioinformatics

Several viewpoints

Understanding Bioinformatics ➡ Processing of biological data using information technology.

- While the field of bioinformatics has existed for decades, the catalyst for its rapid growth in the current millennium came from the Human Genome Project, a landmark international scientific research project completed in April 2003 that made available for the first time the complete genetic blueprint of a human being.



Source: https://www.cleanpng.com/png-bioinformatics-computer-science-computational-biol-2602217/preview.html

- ✓ Bioinformatics finds application in a growing number of areas, such as gene sequencing, gene expression studies and drug discovery.
- ✓ For example, in medicine, bioinformatics can be used to identify links between specific diseases and the gene sequences that cause them.
- ✓ The field of pharmacogenomics uses bioinformatics data to tailor medical treatments to the patients based on their DNA.

Source:

Essential Bioinformatics by Jin Xiong

https://www.investopedia.com/terms/b/bioinformatics.asp#:~:text=Bioinformatics%20is%20the%20application%20of,%2C%20biotechnology%2C%20statistics%20and%20engineering.

Bioinformatics

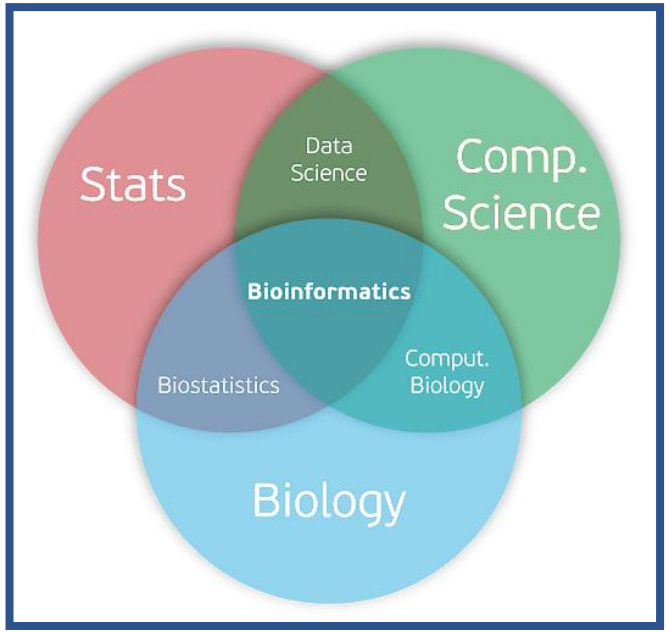
Bioinformatics versus Computational Biology

Bioinformatics

Mainly involve macromolecules

Computational Biology

All computational studies using biological data.



Source:
https://www.researchgate.net/post/What_are_the_differences_between_Bioinformatics_and_Computational_Biology

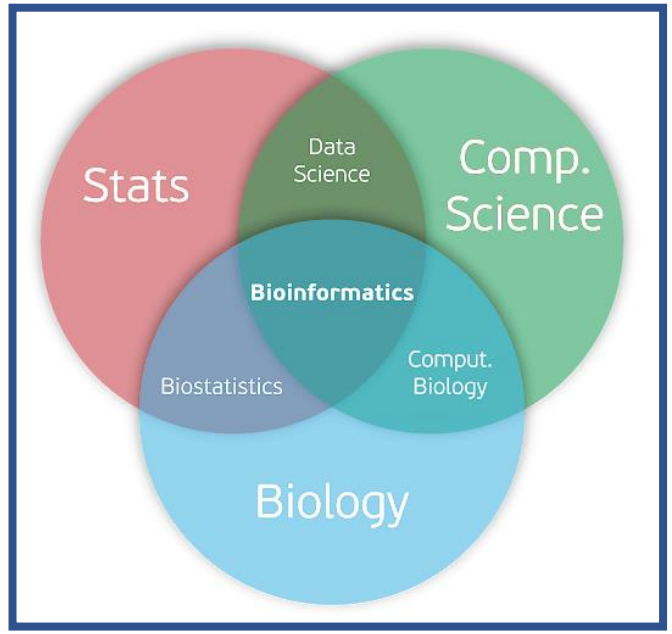
Source:

Essential Bioinformatics by Jin Xiong

Bioinformatics

Bioinformatics versus Computational Biology

- Bioinformatics overlaps with other areas of research that are designated computational biology.
- Bioinformatics is involved in organizing biological data related to genomes, proteomes etc. with a view to implement this information to agriculture, pharmacology, medicine and other commercial applications.
- Bioinformatics involve sequence, structural, and functional analysis of genes and genomes and their corresponding products and is often considered computational molecular biology.
- However, computational biology encompasses all biological areas that involve computation. For example, mathematical modeling of ecosystems, population dynamics, application of the game theory in behavioral studies, and phylogenetic construction using fossil records all employ computational tools, but do not necessarily involve biological macromolecules.



Source:
https://www.researchgate.net/post/What_are_the_differences_between_Bioinformatics_and_Computational_Biology

Source:
Essential Bioinformatics by Jin Xiong

Bioinformatics

Broad Objectives of Bioinformatics

Processing of biological data using information technology.

- ✓ To organize vast amount of molecular biology data in an efficient manner
- ✓ To develop tools that aid in the analysis of such data
- ✓ To interpret the data accurately and meaningfully

Source: https://mangalmay.org/blog/what-is-bioinformatics/

- The advent and rapid rise of bioinformatics have been due to the massive increases in computing power and laboratory technology.
- These advances have made it possible to process and analyze the digital information—DNA, genes, proteins—at the heart of life itself.
- As bioinformatics can be used in any system where information can be represented digitally, it can be applied across the entire spectrum of living organisms, from single cells to complex ecosystems.

Source:

Essential Bioinformatics by Jin Xiong

https://www.investopedia.com/terms/b/bioinformatics.asp#:~:text=Bioinformatics%20is%20the%20application%20of,%2C%20biotechnology%2C%20statistics%20and%20engineering.

Bioinformatics

Brief History of Bioinformatics

- Modern bioinformatics emerged recently to assist next-generation sequencing data analysis.
- However, the very beginnings of bioinformatics occurred more than 50 years ago, when desktop computers were still a hypothesis and DNA could not yet be sequenced.
- The foundations of bioinformatics were laid in the early 1960s with the application of computational methods to protein sequence analysis (notably, de novo sequence assembly, biological sequence databases and substitution models).
- Later on, DNA analysis also emerged due to parallel advances in (i) molecular biology methods, which allowed easier manipulation of DNA, as well as its sequencing, and (ii) computer science, which saw the rise of increasingly miniaturized and more powerful computers, as well as novel software better suited to handle bioinformatics tasks.
- In the 1990s through the 2000s, major improvements in sequencing technology, along with reduced costs, gave rise to an exponential increase of data.
- The arrival of 'Big Data' has laid out new challenges in terms of data mining and management, calling for more expertise from computer science into the field.
- Coupled with an ever-increasing amount of bioinformatics tools, biological Big Data had (and continues to have) profound implications on the predictive power and reproducibility of bioinformatics results.

Source:

Gauthier et al., Briefings in Bioinformatics, Volume 20, Issue 6, November 2019, Pages 1981–1996, https://doi.org/10.1093/bib/bby063

Bioinformatics

Brief History of Bioinformatics

1950–1970: The origins

It did not start with DNA analysis

Protein analysis was the starting point

Dayhoff: the first bioinformatician

A mathematical framework for amino acid substitutions

In 1978, Dayhoff et al developed the first probabilistic model of amino acid substitutions - point accepted mutations (PAMs).

Major events in Bioinformatics

Margaret Dayhoff (1925-1983)

- ✓ Dayhoff was an American physical chemist who pioneered the application of computational methods to the field of biochemistry.
- ✓ Dayhoff's contribution to this field is so important that David J. Lipman, former director of the National Center for Biotechnology Information (NCBI), called her 'the mother and father of bioinformatics'

Source: Gauthier et al., Briefings in Bioinformatics, Volume 20, Issue 6, November 2019, Pages 1981–1996, https://doi.org/10.1093/bib/bby063

Bioinformatics

Brief History of Bioinformatics

Major events in Bioinformatics

1965	Margaret Dayhoff's Atlas of Protein Sequences
1970	Needleman-Wunsch algorithm
1977	DNA sequencing and software to analyze it (Staden)
1981	Smith-Waterman algorithm developed
1981	The concept of a sequence motif (Doolittle)
1982	GenBank Release 3 made public
1982	Phage lambda genome sequenced
1983	Sequence database searching algorithm (Wilbur-Lipman)
1985	FASTP/FASTN: fast sequence similarity searching
1988	National Center for Biotechnology Information (NCBI) created at NIH/NLM
1988	EMBnet network for database distribution
1990	BLAST: fast sequence similarity searching
1991	EST: expressed sequence tag sequencing
1993	Sanger Centre, Hinxton, UK
1994	EMBL European Bioinformatics Institute, Hinxton, UK
1995	First bacterial genomes completely sequenced
1996	Yeast genome completely sequenced
1997	PSI-BLAST
1998	Worm (multicellular) genome completely sequenced
1999	Fly genome completely sequenced
2000	Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. Nature 2000 Oct 5;407(6804):651-4, PubMed
2000	The genome for Pseudomonas aeruginosa (6.3 Mbp) is published.
2000	The A. thaliana genome (100 Mb) is secquenced.
2001	The human genome (3 Giga base pairs) is published.

Source: https://www.roseindia.net/bioinformatics/history_of_bioinformatics.shtml

https://chagall.med.cornell.edu/BioinfoCourse/presentations2010/Lecture1_2010.pdf

Bioinformatics

Goals

- The ultimate goal of bioinformatics is to better understand a living cell and how it functions at the molecular level.
- By analyzing raw molecular sequence and structural data, bioinformatics research can generate new insights and provide a "global" perspective of the cell.
- The reason that the functions of a cell can be better understood by analyzing sequence data is ultimately because the flow of genetic information is dictated by the "central dogma" of biology in which DNA is transcribed to RNA, which is translated to proteins.
- Cellular functions are mainly performed by proteins whose capabilities are ultimately determined by their sequences.
- Therefore, solving functional problems using sequence and sometimes structural approaches has proved to be a fruitful endeavor.

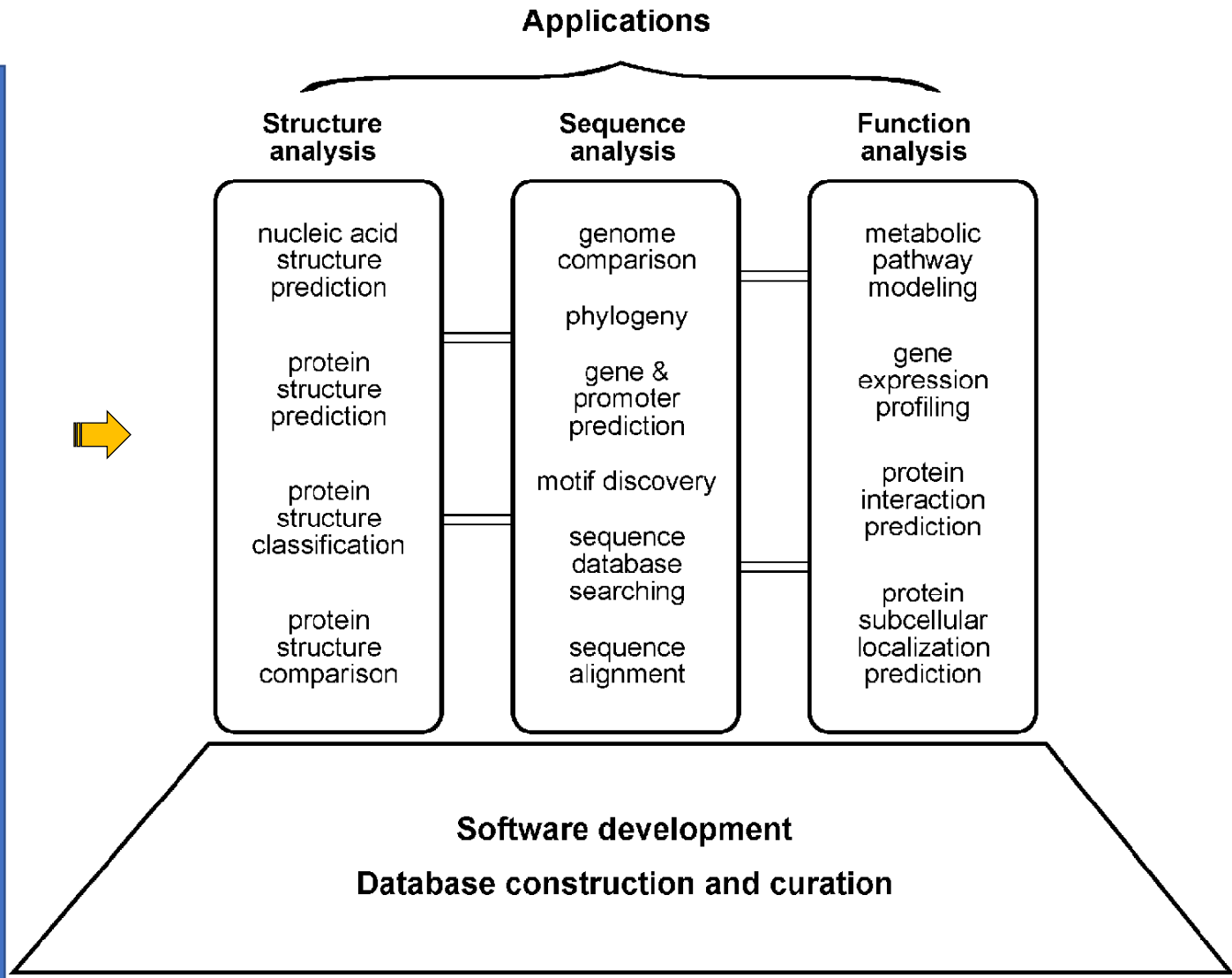
Source:

Essential Bioinformatics by Jin Xiong

Bioinformatics

Scope

- Bioinformatics consists of two subfields: the development of computational tools and databases and the application of these tools and databases in generating biological knowledge to better understand living systems.
- These two subfields are complementary to each other.
- The tool development includes writing software for sequence, structural, and functional analysis, as well as the construction and curating of biological databases.
- These tools are used in three areas of genomic and molecular biological research: molecular sequence analysis, molecular structural analysis, and molecular functional analysis.
- The analyses of biological data often generate new problems and challenges that in turn spur the development of new and better computational tools.



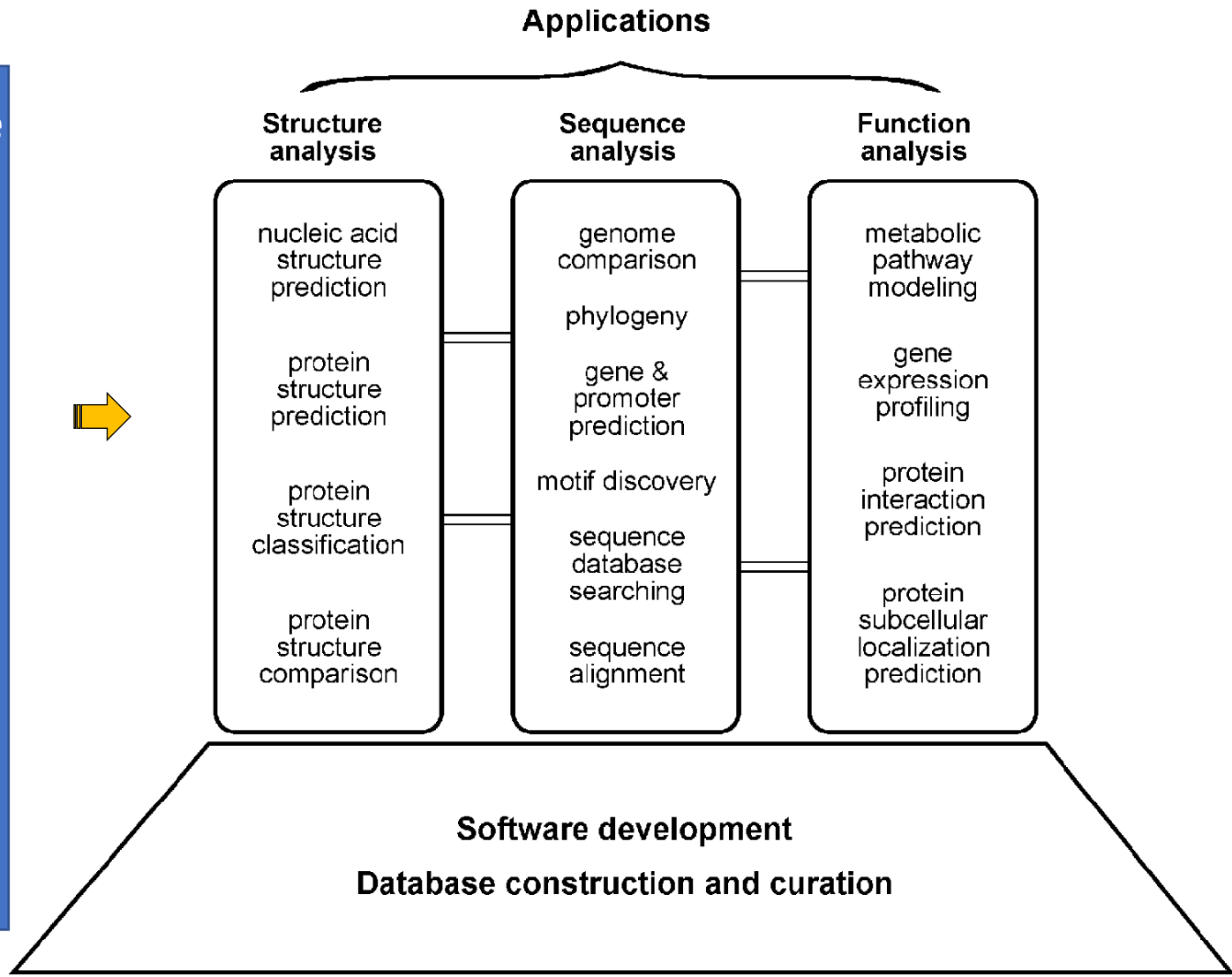
Source:

Essential Bioinformatics by Jin Xiong

Bioinformatics

Scope

- The areas of sequence analysis include sequence alignment, sequence database searching, motif and pattern discovery, gene and promoter finding, reconstruction of evolutionary relationships, and genome assembly and comparison.
- Structural analyses include protein and nucleic acid structure analysis, comparison, classification, and prediction.
- The functional analyses include gene expression profiling, protein–protein interaction prediction, protein subcellular localization prediction, metabolic pathway reconstruction, and simulation.



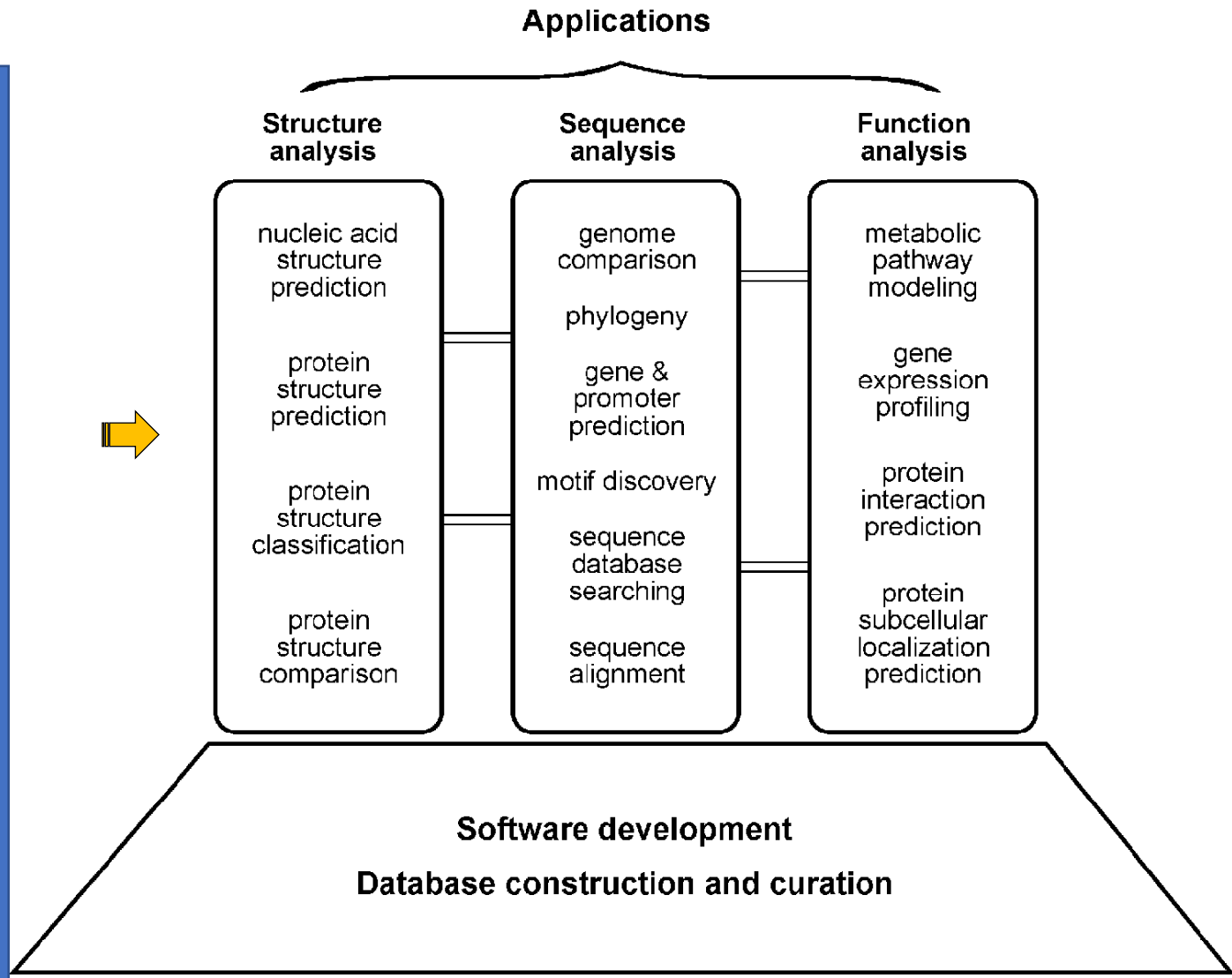
Source:

Essential Bioinformatics by Jin Xiong

Bioinformatics

Scope

- The three aspects of bioinformatics analysis are not isolated but often interact to produce integrated results.
- For example, protein structure prediction depends on sequence alignment data.
- Clustering of gene expression profiles requires the use of phylogenetic tree construction methods derived in sequence analysis.
- Sequence-based promoter prediction is related to functional analysis of co-expressed genes.
- Gene annotation involves a number of activities, which include distinction between coding and noncoding sequences, identification of translated protein sequences, and determination of the gene's evolutionary relationship with other known genes; prediction of its cellular functions employs tools from all three groups of the analyses.



Source:

Essential Bioinformatics by Jin Xiong

Bioinformatics

Broad applications

Medicine

- ✓ Molecular medicine
- ✓ Personalised medicine
- ✓ Gene therapy
- ✓ Drug development
- ✓ Antibiotic resistance

Agriculture

- ✓ Crop improvement
- ✓ Improve nutritional quality
- ✓ Development of drought resistant varieties
- ✓ Insect resistance

Phylogenetic

- ✓ Evolutionary studies

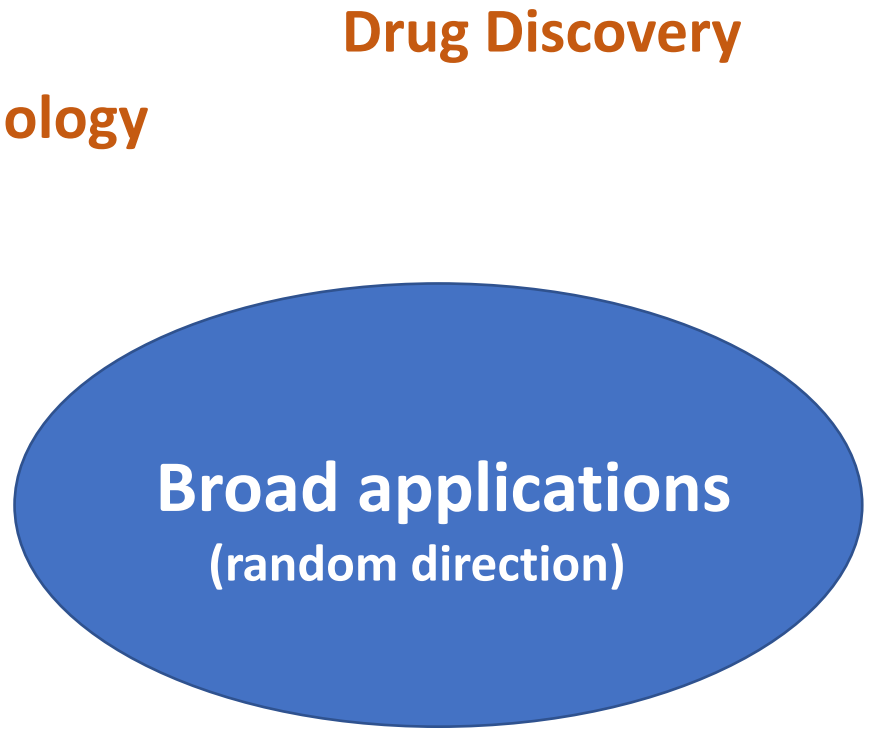
Microbial genome applications

- ✓ Waste cleanup
- ✓ Biotechnology
- ✓ Microbial genome sequencing

Source:

Essential Bioinformatics by Jin Xiong

Bioinformatics



Drug Discovery

System Biology

Microbiology

Omics

Biotechnology

Network Pharmacology

Phylogenetics

Drug repurposing

Agriculture

Source:

Essential Bioinformatics by Jin Xiong

Bioinformatics

Applications
(another direction)

- Essential for basic genomic and molecular biology research.
- Major impact on many areas of biotechnology and biomedical sciences.

- ✓ Bioinformatics plays a vital role in the areas of structural genomics, functional genomics, and nutritional genomics.
- ✓ It covers emerging scientific research and the exploration of proteomes from the overall level of intracellular protein composition (protein profiles), protein structure, protein-protein interaction, and unique activity patterns (e.g. post-translational modifications).
- ✓ Bioinformatics is used for transcriptome analysis where mRNA expression levels can be determined.
- ✓ Bioinformatics is used to identify and structurally modify a natural product, to design a compound with the desired properties and to assess its therapeutic effects, theoretically.
- ✓ Cheminformatics analysis includes analyses such as similarity searching, clustering, QSAR modeling, virtual screening, etc.
- ✓ Bioinformatics plays an increasingly important role in almost all aspects of drug discovery and drug development.
- ✓ Bioinformatics tools are very effective in prediction, analysis and interpretation of clinical and preclinical findings.

Source:
Essential Bioinformatics by Jin Xiong

https://microbenotes.com/bioinformatics-introduction-and-applications/

Bioinformatics

Applications

- Bioinformatics has not only become essential for basic genomic and molecular biology research, but is having a major impact on many areas of biotechnology and biomedical sciences.
- It has applications, for example, in knowledge-based drug design, forensic DNA analysis, and agricultural biotechnology.
- Computational studies of protein–drug interactions provide a rational basis for the rapid identification of novel leads for synthetic drugs.
- Knowledge of the three-dimensional structures of proteins allows molecules to be designed that are capable of binding to the receptor site of a target protein with great affinity and specificity.
- This informatics-based approach significantly reduces the time and cost necessary to develop drugs with higher potency, fewer side effects, and less toxicity than using the traditional trial-and-error approach.

Source:

Essential Bioinformatics by Jin Xiong

Bioinformatics

Applications

- It is worth mentioning that genomics and bioinformatics are now poised to revolutionize our healthcare system by developing personalized and customized medicine.
- The high speed genomic sequencing coupled with sophisticated informatics technology will allow a doctor in a clinic to quickly sequence a patient's genome and easily detect potential harmful mutations and to engage in early diagnosis and effective treatment of diseases.
- Bioinformatics tools are being used in agriculture. Plant genome databases and gene expression profile analyses have played an important role in the development of new crop varieties that have higher productivity and more resistance to disease.

Source:

Essential Bioinformatics by Jin Xiong

Bioinformatics

Types of data available

Huge data available 

- DNA/RNA sequence
- single-nucleotide polymorphisms (SNPs)
- protein sequence
- protein structure
- protein function
- organism-specific databases
- genomes
- gene expression
- biomolecular interactions
- molecular pathways
- scientific literature
- disease information

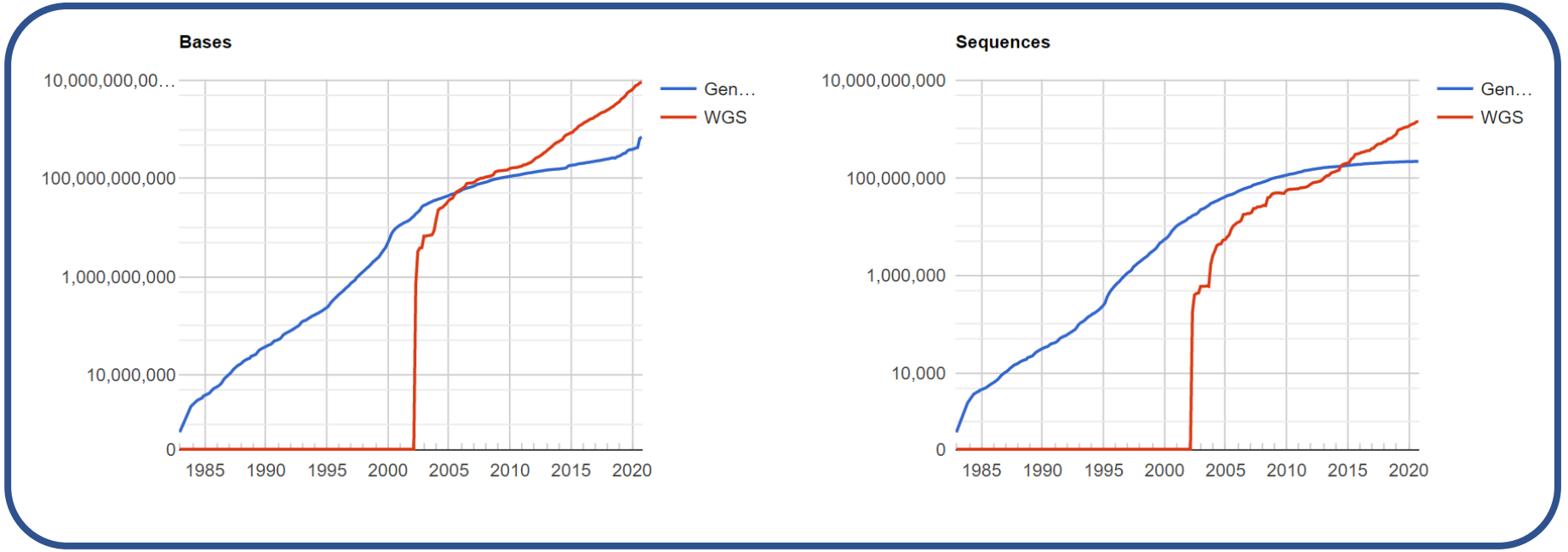
Source:

Essential Bioinformatics by Jin Xiong https://chagall.med.cornell.edu/BioinfoCourse/presentations2010/Lecture1_2010.pdf

Bioinformatics

Types of data available

Growth of GenBank (Gen…) and whole genome sequencing (WGS)



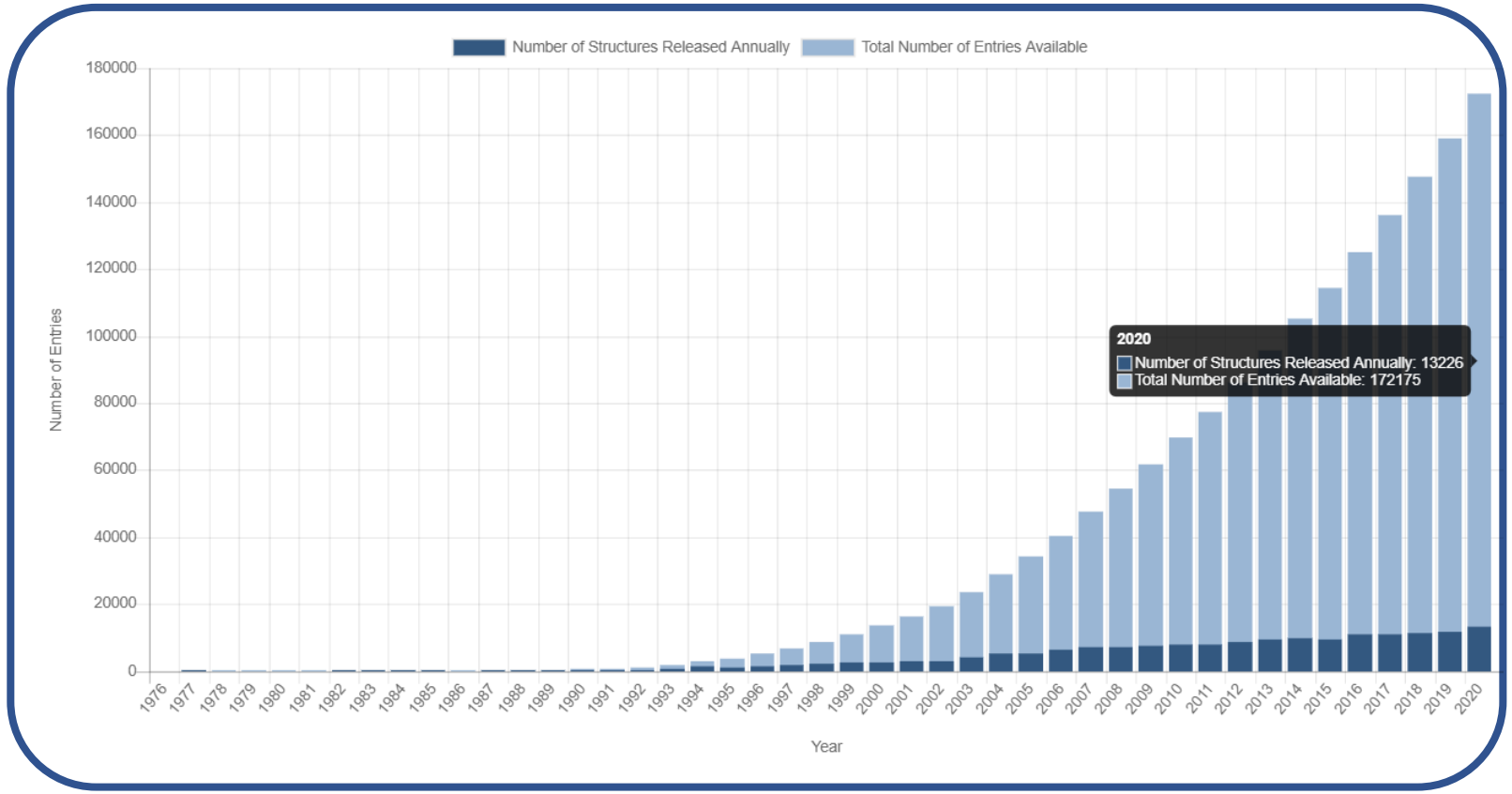
- ✓ **GenBank, beginning with Release 3 in 1982.**
- ✓ **From 1982 to the present, the number of bases in GenBank has doubled approximately every 18 months.**

Source: https://www.ncbi.nlm.nih.gov/genbank/statistics/#:~:text=Notes%20on%20GenBank%20statistics&text=From%201982%20to%20the%20present,doubled%20approximately%20every%2018%20months.

Bioinformatics

Types of data available

Growth of
Protein Data Bank (PDB)



Source: https://www.rcsb.org/stats