Βιοπληροφορική Ι

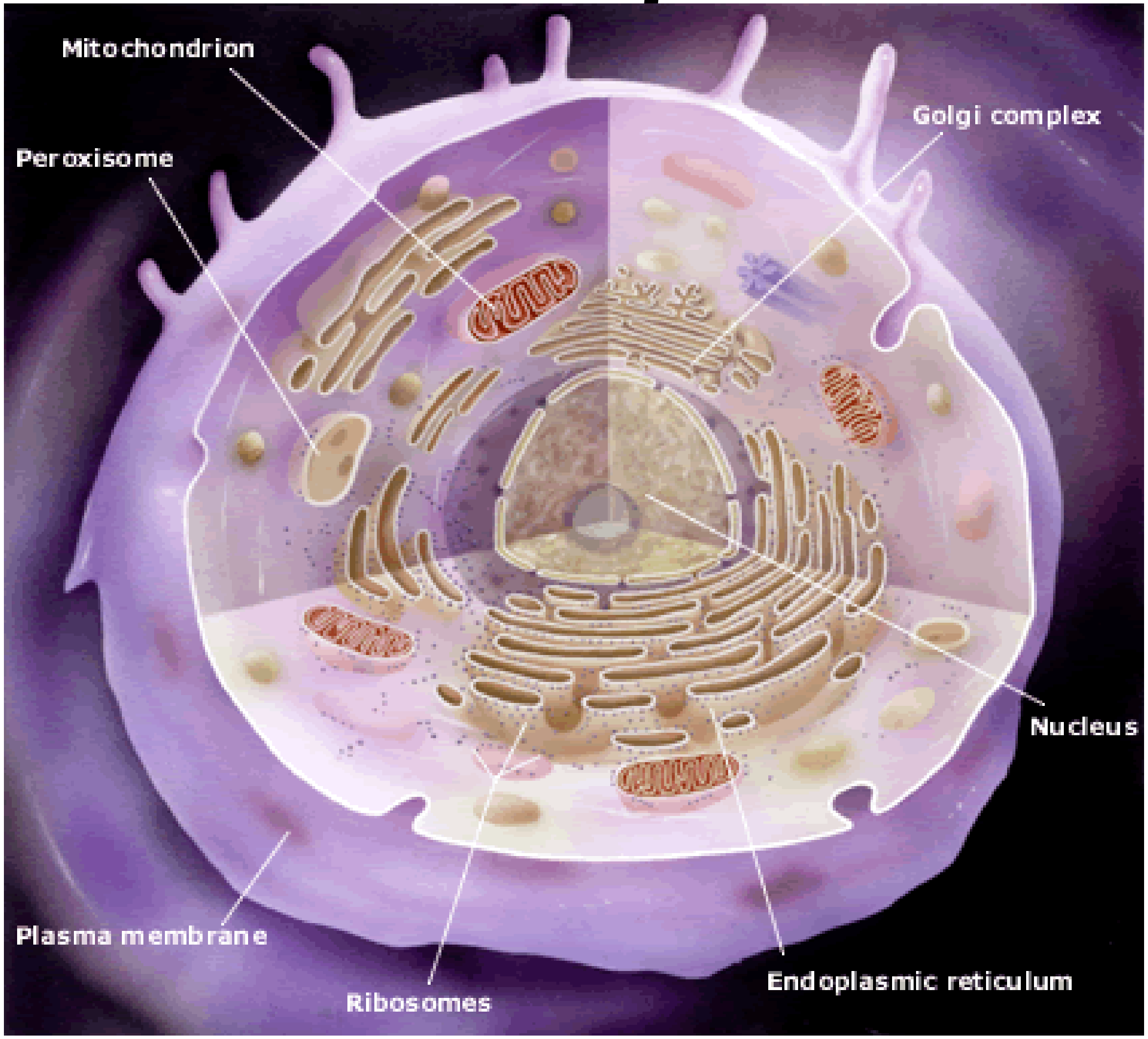
Παντελής Μπάγκος
Αναπληρωτής Καθηγητής

Πανεπιστήμιο Θεσσαλίας
Λαμία 2015

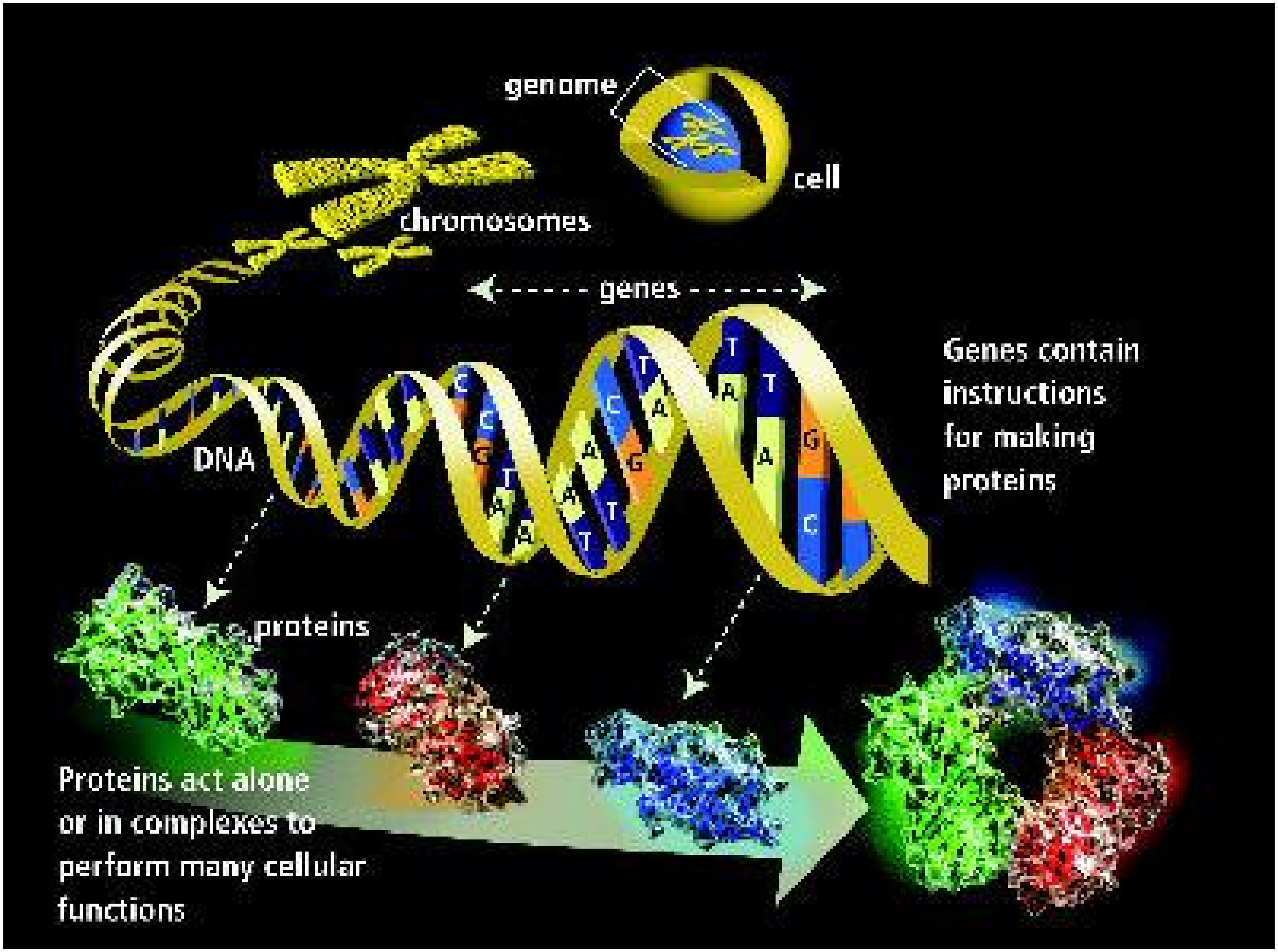
Βιοπληροφορική Ι

- **Εισαγωγή:** Ορισμός της Βιοπληροφορικής, Υποδιαιρέσεις της Βιοπληροφορικής, Τα είδη των δεδομένων στη Βιοπληροφορική.
- **Βάσεις δεδομένων:** Βάσεις δεδομένων βιβλιογραφίας, Βάσεις δεδομένων ακολουθιών πρωτεϊνών και DNA, Βάσεις δεδομένων δομών, Βάσεις δεδομένων διπλωμάτων και οικογενειών, Εξειδικευμένες Βάσεις δεδομένων, Εργαλεία ανάλυσης της πληροφορίας που είναι αποθηκευμένη στις βάσεις δεδομένων (Entrez, SRS).
- **Στοίχιση ακολουθιών:** Μέθοδοι εύρεσης ομοιοτήτων σε ακολουθίες, Ομολογία και ομοιότητα ακολουθιών και η σημασία τους, Οι αλγόριθμοι δυναμικού προγραμματισμού, Ολική στοίχιση (Global Alignment) και ο αλγόριθμος των Needleman και Wunsch, τοπική στοίχιση (Local Alignment) και ο αλγόριθμος των Smith και Waterman, Υπολογισμός της στατιστικής σημαντικότητας της στοίχισης, Οι πίνακες ομοιότητας και η σημασία τους, Οι ποινές για τα κενά, Ευριστικές μέθοδοι για αναζητήσεις ομοιοτήτων σε βάσεις δεδομένων (BLAST, FASTA κτλ).
- **Πολλαπλή στοίχιση ακολουθιών:** Πολυδιάστατοι αλγόριθμοι δυναμικού προγραμματισμού, Ευριστικές μέθοδοι πολλαπλής στοίχισης ακολουθιών (CLUSTAL, DIALIGN, MULTALIN κτλ), φυλογενετικά δέντρα και πολλαπλές στοιχίσεις.
- **Αλγόριθμοι πρόγνωσης στηριζόμενοι στην ακολουθία πρωτεϊνών και DNA**: Πρόγνωση δευτεροταγούς δομής πρωτεϊνών και RNA, Πρόγνωση διαμεμβρανικών τμημάτων πρωτεϊνών και προσανατολισμού τους, Εύρεση πιθανών γονιδίων σε ακολουθίες DNA, Πολλαπλές στοιχίσεις ακολουθιών με χρήση Hidden Markov Models, Κατάταξη ακολουθιών σε οικογένειες.
- **Δομική Βιοπληροφορική:** Αναπαράσταση βιολογικών δομών, Αναγνώριση πρωτεϊνικού διπλώματος, Προσαρμογή και υπέρθεση δομών στο χώρο, Συγκριτική προτυποποίηση με βάση την ομολογία

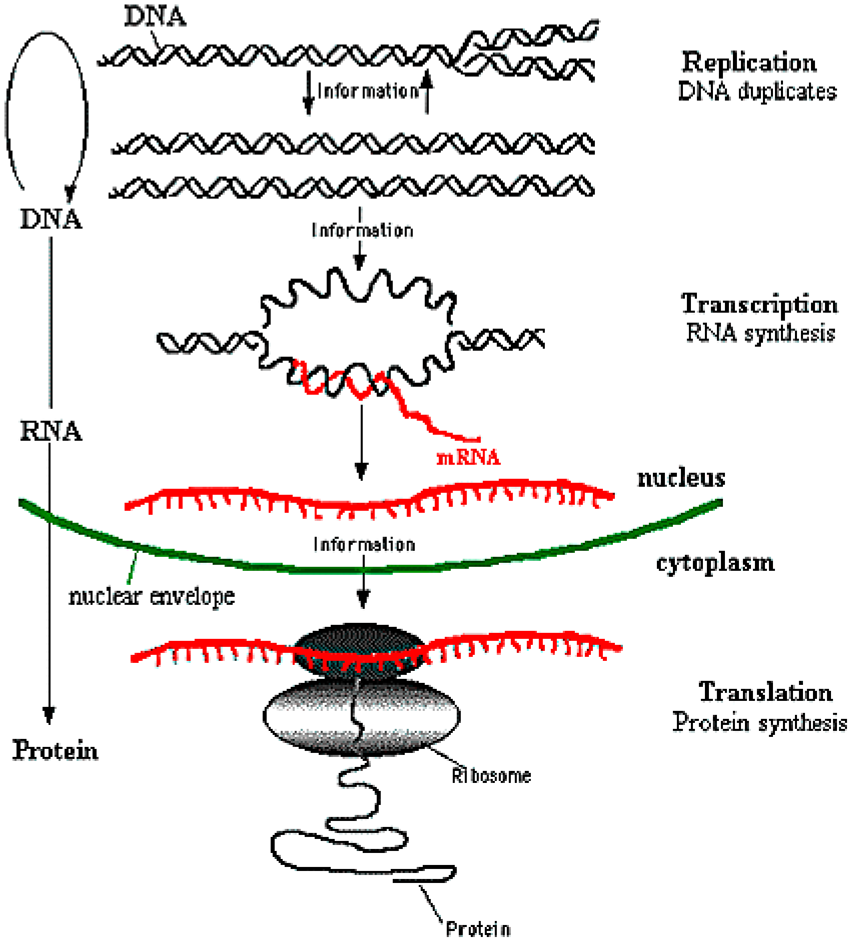
Οργάνωση Τυπικού Ευκαρυωτικού Κυττάρου



Οργάνωση της Πληροφορίας


genome
cell
chromosomes
genes
DNA
Genes contain instructions for making proteins
proteins
Proteins act alone or in complexes to perform many cellular functions

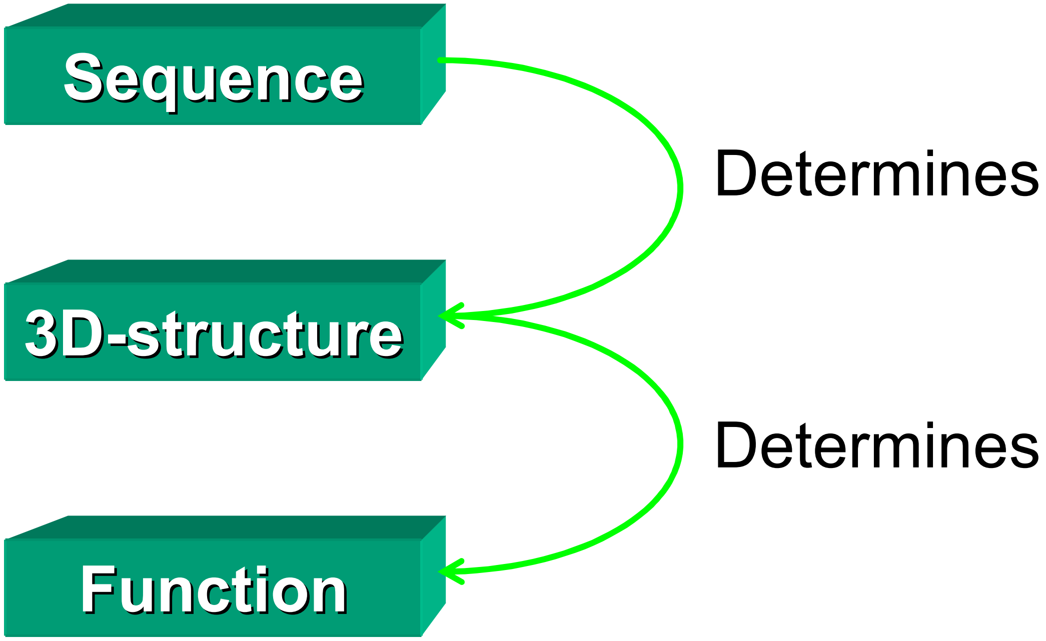
Το Κεντρικό Δόγμα της Μοριακής Βιολογίας ...



The Central Dogma of Molecular Biology

http://www.accessexcellence.org/AB/GG/central.html

... και μια φυσική προέκτασή του ...



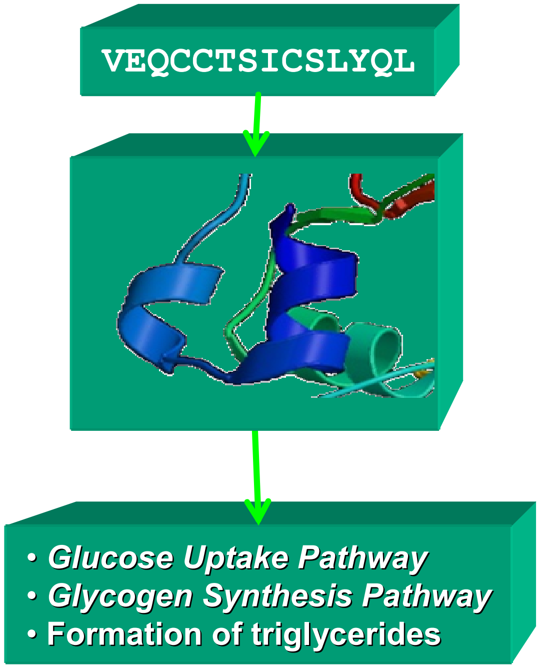
Sequence

Determines

3D-structure

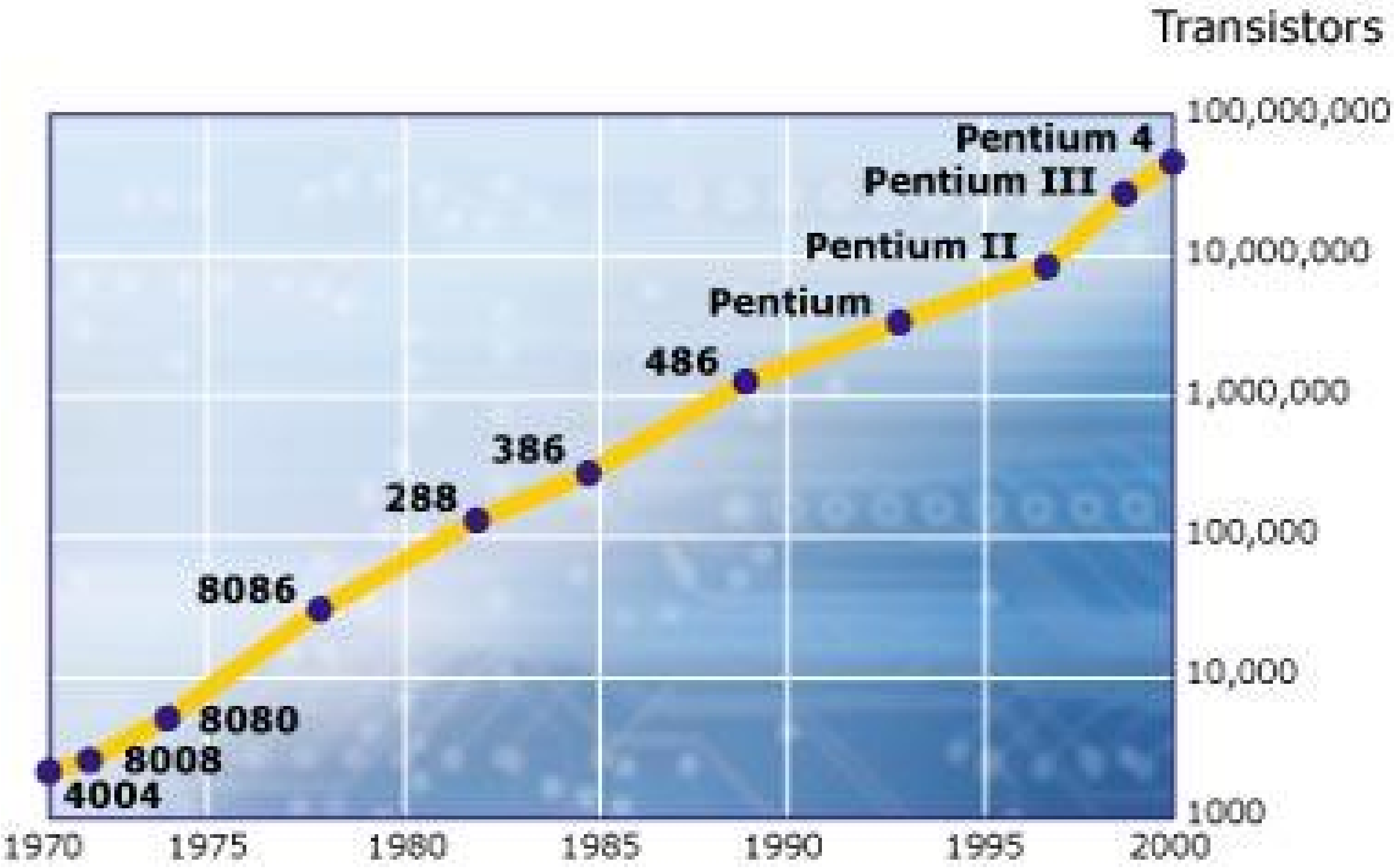
Determines

Function



VEQCCTSICSLYQL

- *Glucose Uptake Pathway*
- *Glycogen Synthesis Pathway*
- Formation of triglycerides



Ο νόμος του Moore (αύξηση της υπολογιστικής ισχύος)

Συνέπεια…

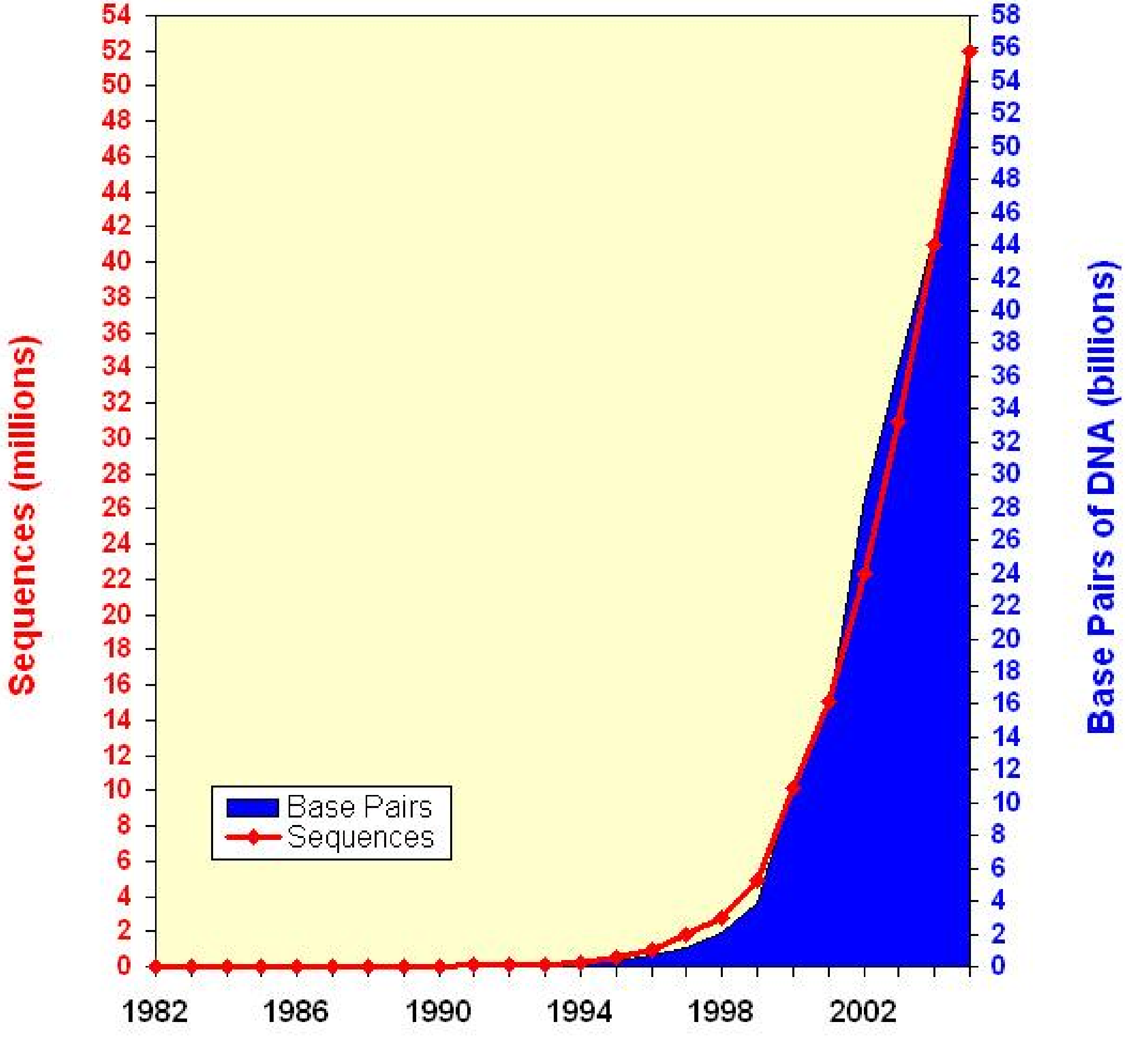
Ένας μέσος βιολόγος (όπως και κάθε μέσος χρήστης) το 2006, έχει στη διάθεσή του πολύ ισχυρότερους Η/Υ, με μεγαλύτερη αποθηκευτική ισχύ και μνήμη, συνδεδεμένους στο διαδίκτυο με μεγαλύτερες ταχύτητες απ’ ότι για παράδειγμα ήταν δυνατόν το 1996.

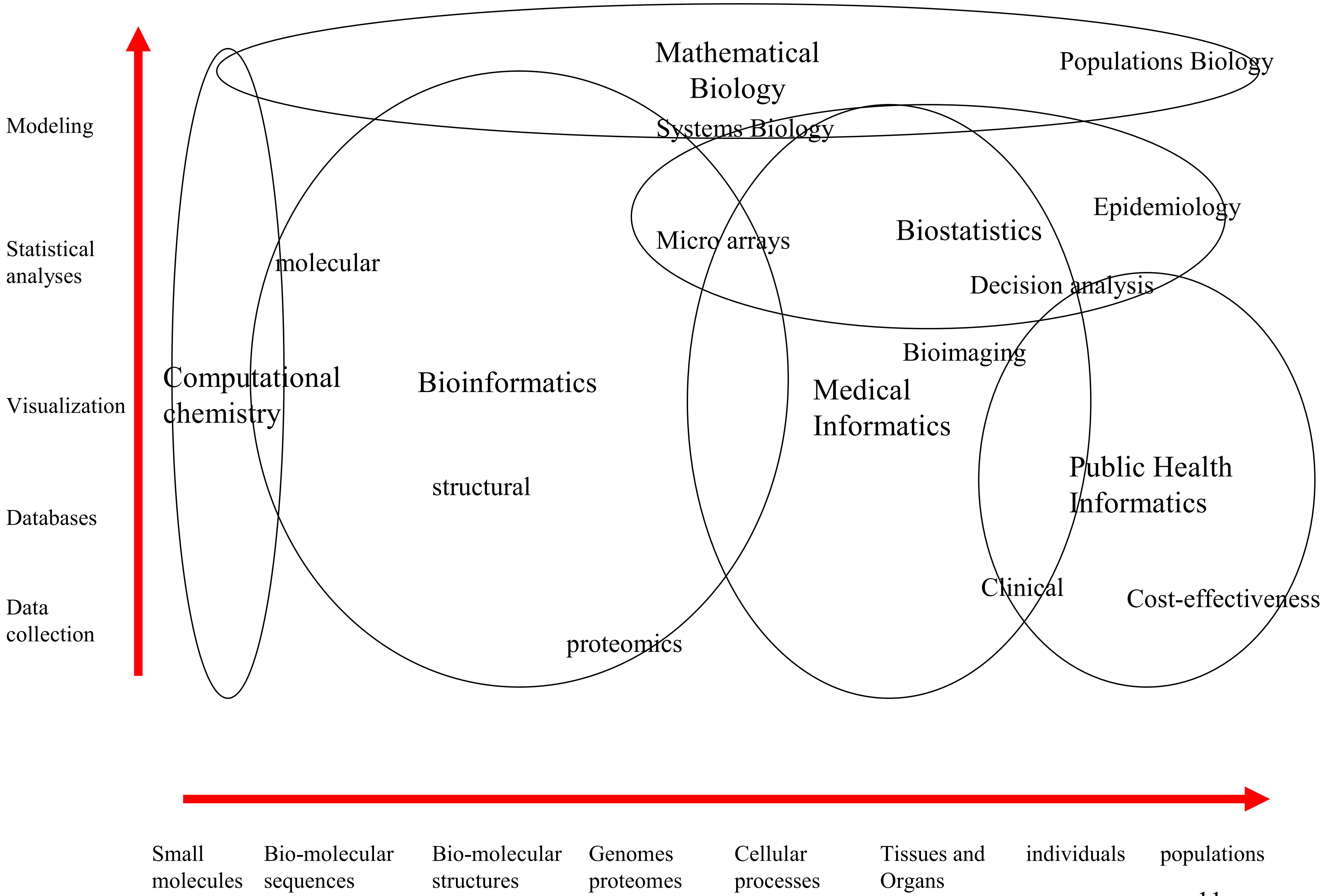
Επιπλέον…

Η παραγωγή βιολογικών δεδομένων αυξάνεται επίσης με ραγδαίο ρυθμό (προσδιορισμός αλληλουχίας γονιδιωμάτων, δομών πρωτεϊνών, δεδομένα γονιδιακής έκφρασης, λειτουργική γονιδιωματική κλπ)

Growth of GenBank

(1982 - 2005)





Modeling

Statistical analyses

Visualization

Databases

Data collection

Mathematical Biology

Populations Biology

Systems Biology

Epidemiology

Micro arrays

Biostatistics

molecular

Decision analysis

Bioimaging

Computational chemistry

Bioinformatics

Medical Informatics

structural

Public Health Informatics

Clinical

Cost-effectiveness

proteomics

Small molecules

Bio-molecular sequences

Bio-molecular structures

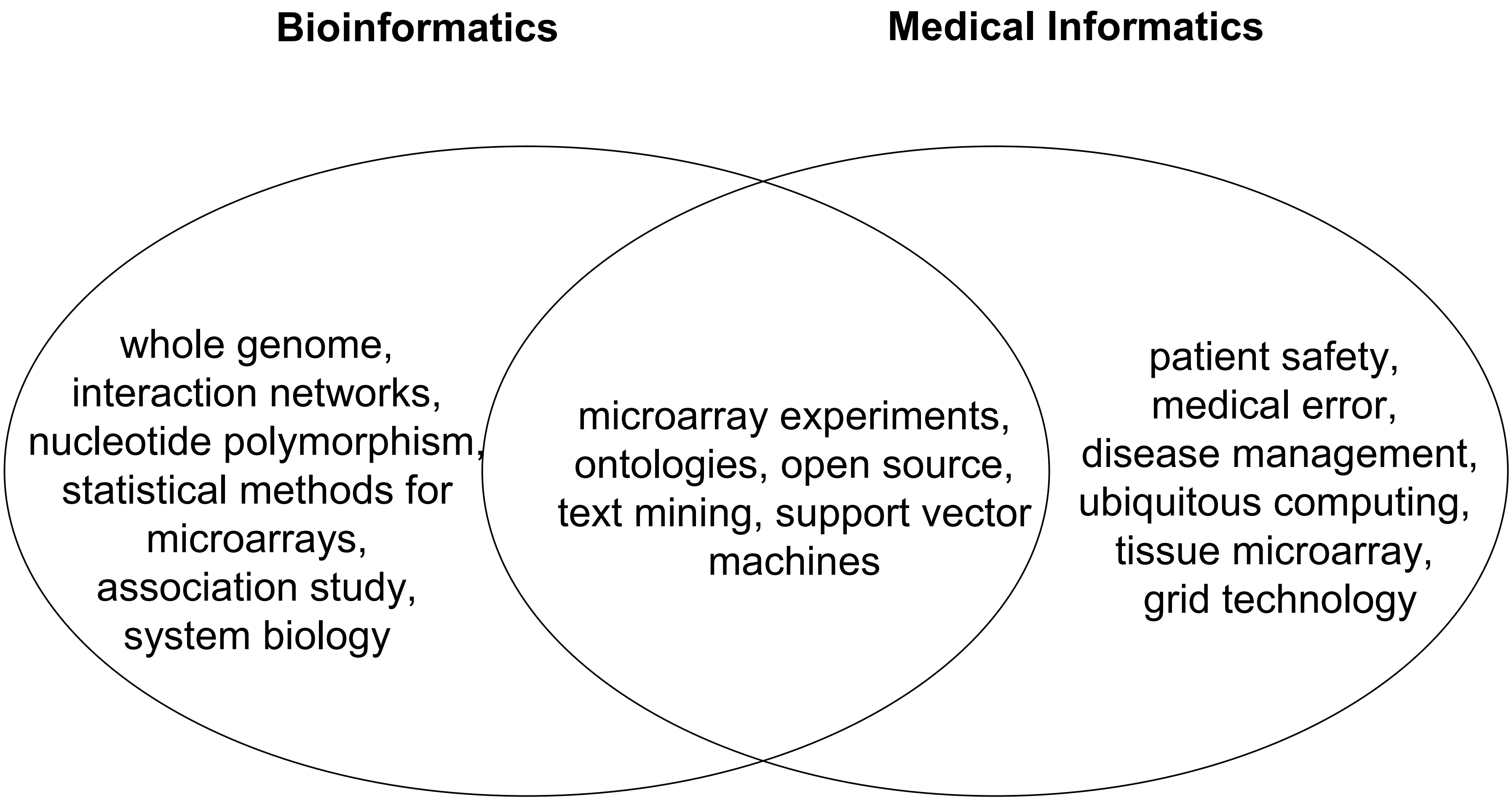
Genomes proteomes

Cellular processes

Tissues and Organs

individuals

populations



Bioinformatics

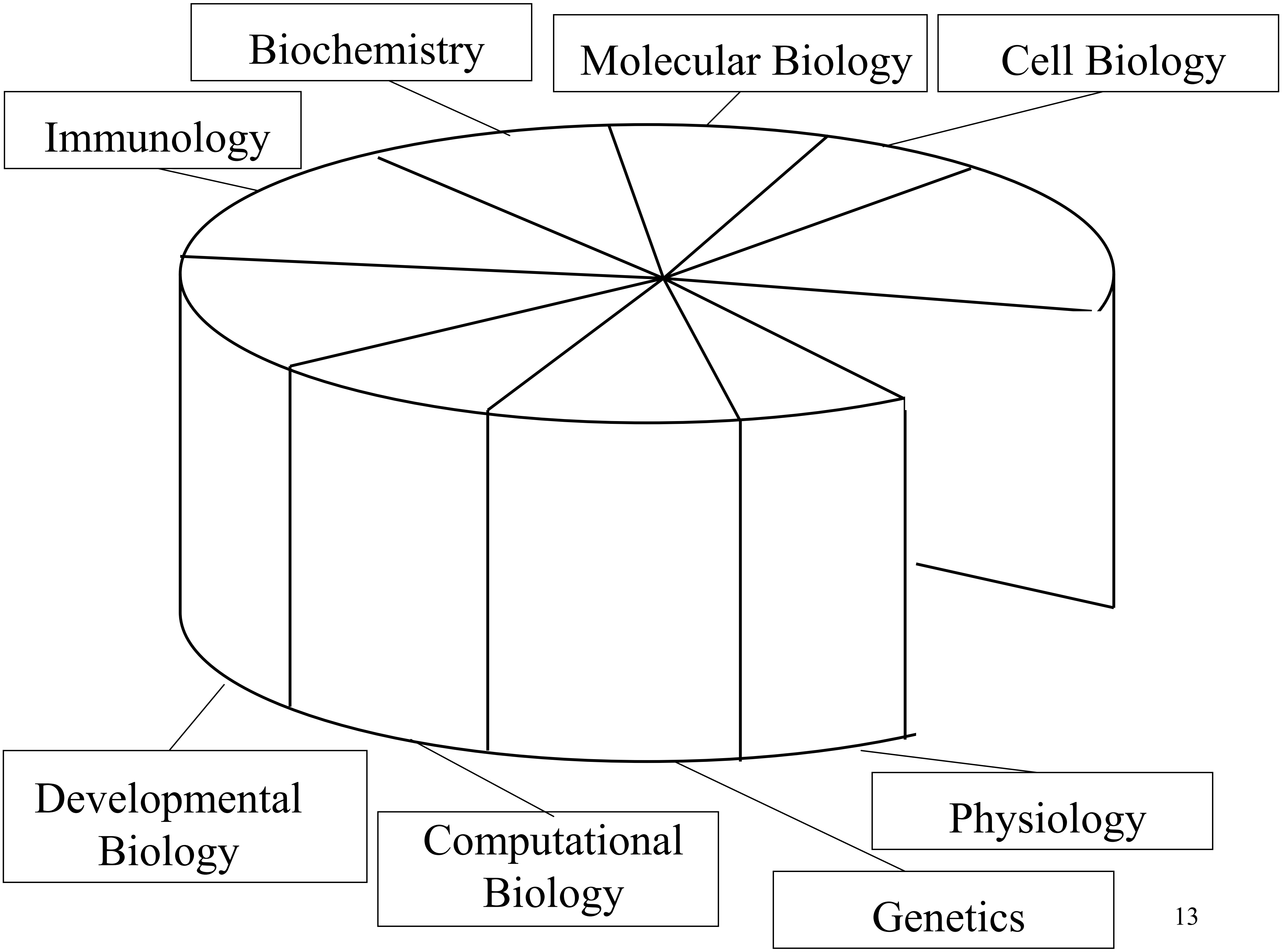
Medical Informatics

whole genome,
interaction networks,
nucleotide polymorphism,
statistical methods for
microarrays,
association study,
system biology

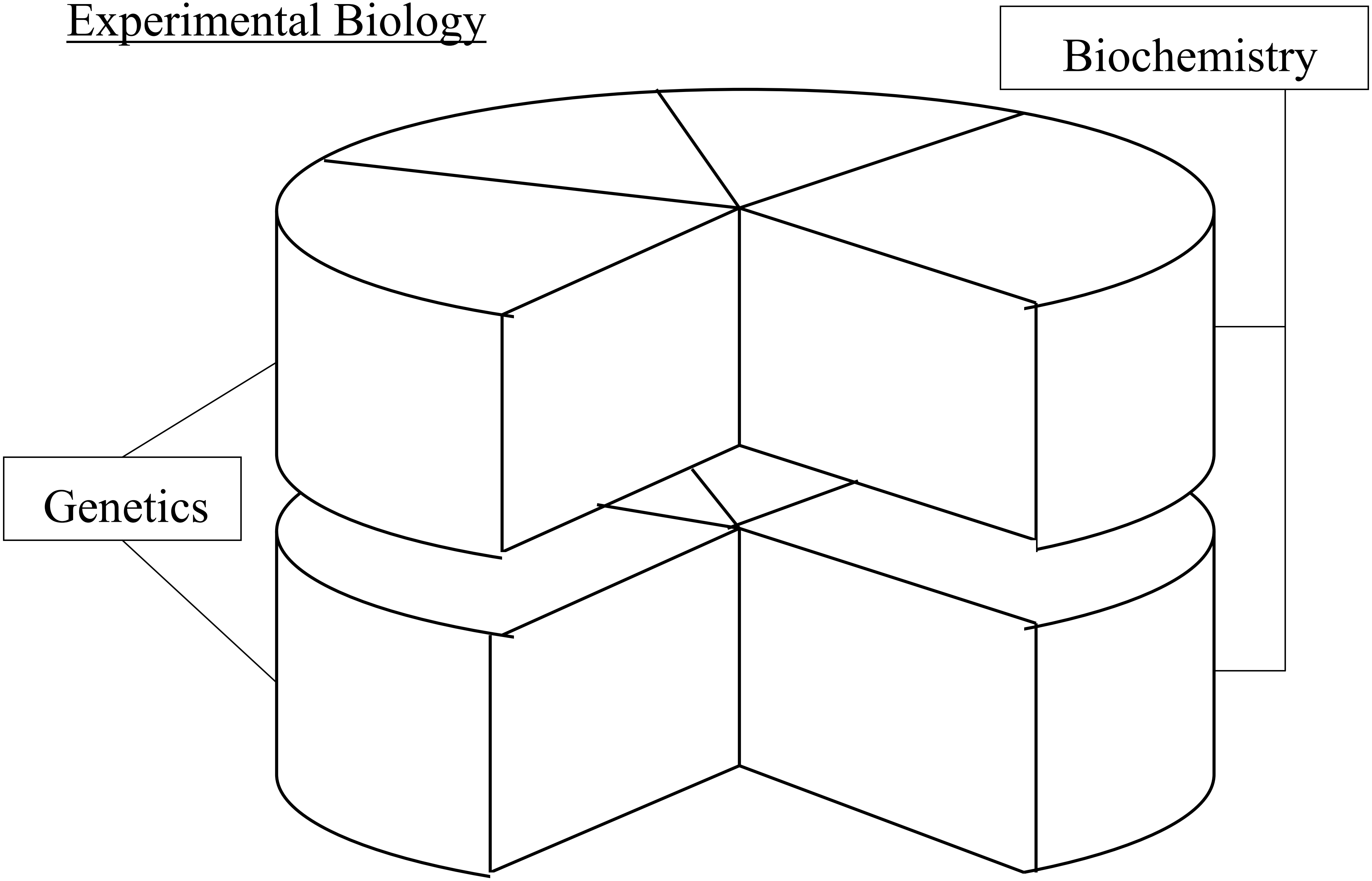
microarray experiments,
ontologies, open source,
text mining, support vector
machines

patient safety,
medical error,
disease management,
ubiquitous computing,
tissue microarray,
grid technology

Rebholz-Schuhman, D., Cameron, G., Clark, D., van Mulligen, E., Coatrieux, J. L., Del Hoyo Barbolla, E., . . . Van der Lei, J. (2007). SYMBIOmatics: synergies in Medical Informatics and Bioinformatics--exploring current scientific literature for emerging topics. *BMC Bioinformatics, 8 Suppl 1*, S18


Biochemistry
Molecular Biology
Cell Biology
Immunology
Developmental Biology
Computational Biology
Genetics
Physiology

Experimental Biology



Biochemistry

Genetics

Theoretical/Computational Biology

Τι είναι Βιοπληροφορική

- Βιοπληροφορική είναι ο επιστημονικός χώρος όπου η σύμπραξη της Βιολογίας με την Πληροφορική, την Στατιστική και τα Μαθηματικά εξερευνά νέους τρόπους για την προσέγγιση των βιολογικών προβλημάτων, καθώς και την αντίληψη βασικών αρχών της Βιολογίας.
- Πρόκειται για γνωστικό χώρο με συγκεκριμένο όσο και ευρύ πεδίο εφαρμογών και αλληλεπίδρασης με τη σύγχρονη δομική, μοριακή, πληθυσμιακή και περιβαλλοντική βιολογία.
- Ο κλάδος της Βιοπληροφορικής σήμερα θεωρείται, παγκόσμια, ένας από τους πλέον αναπτυσσόμενους, ενώ έχει ήδη επιδείξει σημαντικά επιτεύγματα και έχει συγκεντρώσει ιδιαίτερα σημαντικές επενδύσεις. Ουσιαστικά, κατέχει κεντρική θέση στις σύγχρονες εξελίξεις των Επιστημών της Ζωής, με πιο χαρακτηριστικό παράδειγμα τα προγράμματα "Αποκωδικοποίησης" των Γονιδιωμάτων, περιλαμβανομένου και αυτού του Ανθρώπου.

Definition of NCBI:

- «*Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline. There are three important sub-disciplines within bioinformatics: the development of new algorithms and statistics with which to assess relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information*»

Luscombe, Greenbaum, & Gerstein, 2001

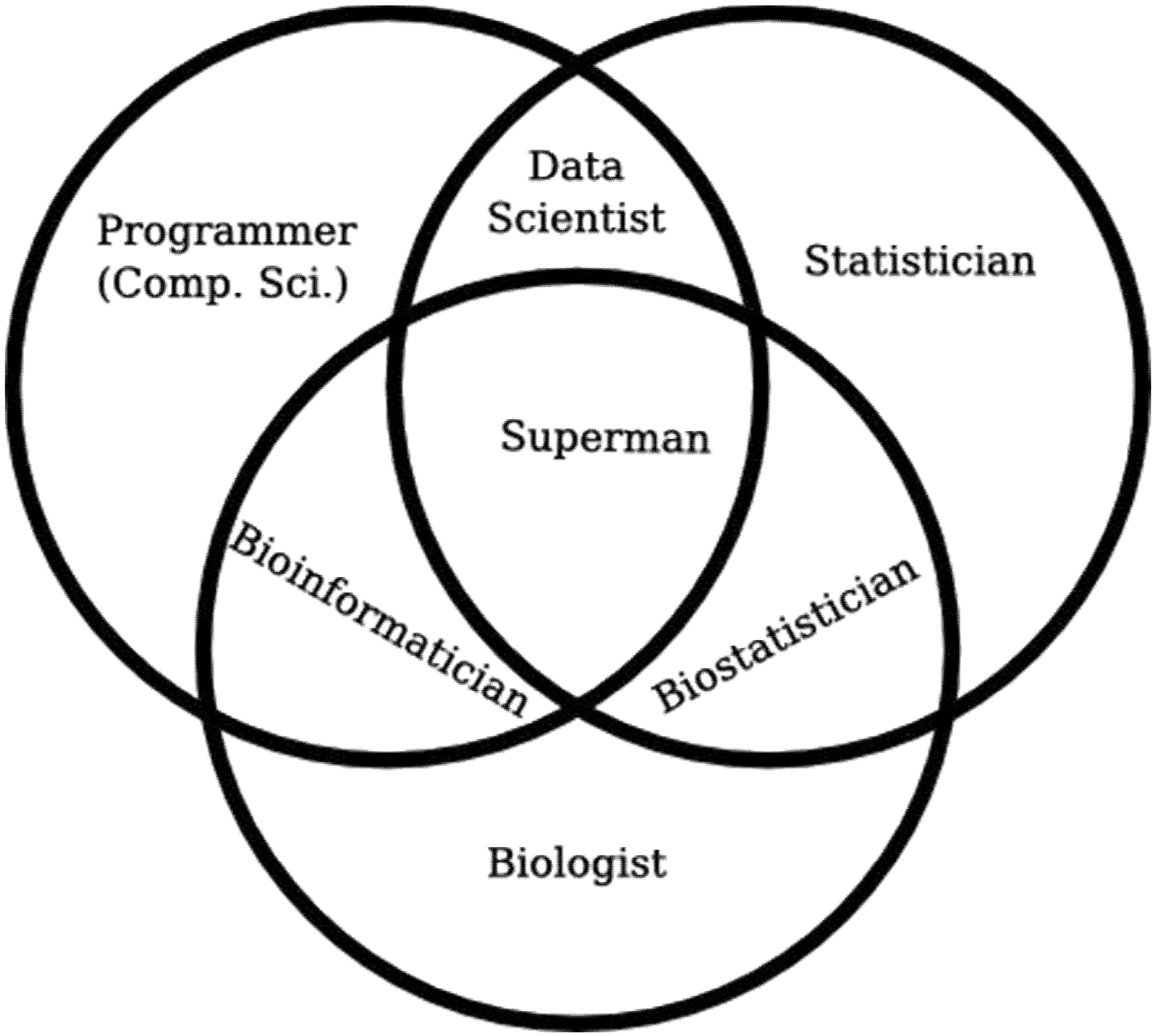
- *«Bioinformatics is conceptualizing biology in terms of macromolecules (in the sense of physical-chemistry) and then applying "informatics" techniques (derived from disciplines such as applied maths, computer science, and statistics) to understand and organize the information associated with these molecules, on a large-scale»*

Fredj Tekaia

- *«The mathematical, statistical and computing methods that aim to solve biological problems using DNA and amino acid sequences and related information»*

ISCB

- *«a scholarly society dedicated to advancing the scientific understanding of living systems through computation».*



http://blog.fejes.ca/?p=2418

How to train individuals with scientists of other fields, in order to simultaneously and efficiently communicate?

- A lot of theoretical work on how a Bioinformatics curriculum should be built
- Combine elements of the «mother» disciplines and several approaches have been described
- HSCBB encompasses all areas of computation applied to living systems

- The interdisciplinary nature of the field raises important questions regarding the training of young scientists
- The pioneers in the field come from a different discipline, most notably from

  Computer Science,

  Biology,

  Mathematics,

  Physics or

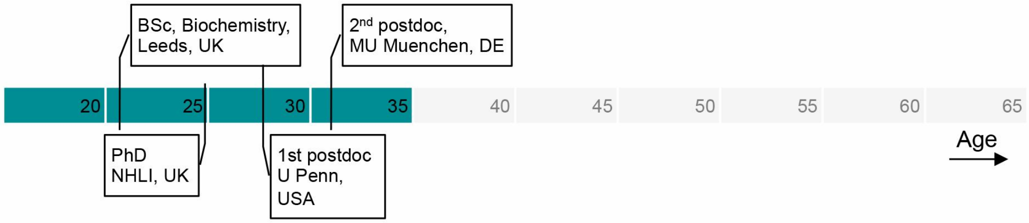
  Chemistry

- The field is clearly an interdisciplinary one, but there is also a need to train interdisciplinary scientists
  - Eddy, S. R. (2005). "Antedisciplinary" science. PLoS Comput Biol, 1(1), e6
- We need to distinguish «Bioinformatics users», «Bioinformatics scientists», and «Bioinformatics engineers»
  - Welch L, Lewitter F, Schwartz R, Brooksbank C, Radivojac P, Gaeta B, et al. (2014) Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies. PLoS Comput Biol 10(3): e1003496.

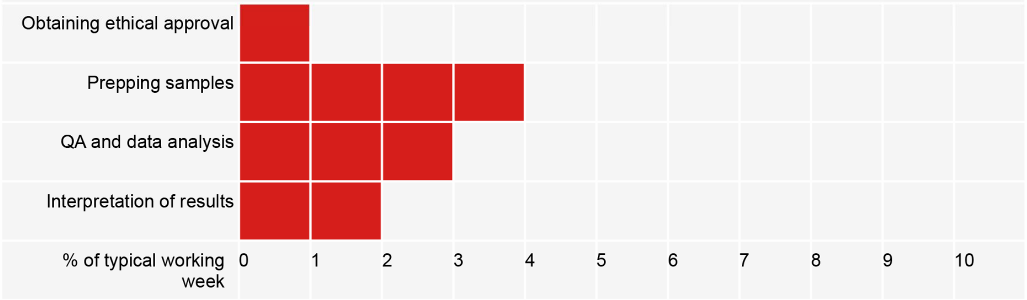
Leon (bioinformatics user)

Leon is on his second postdoctoral fellowship, working on quorum sensing in bacteria. “I’m using a combination of transcriptomics, proteomics and metabolomics to understand these pathogenic changes better” he explains. “I end up with big spreadsheets of protein or gene IDs and I’m trying to piece together which signaling pathways are involved in flipping to the pathogenic state”. He has been on an introductory Unix course but is much more comfortable with GUIs than with the command line. “I just have a visual brain”, he says.

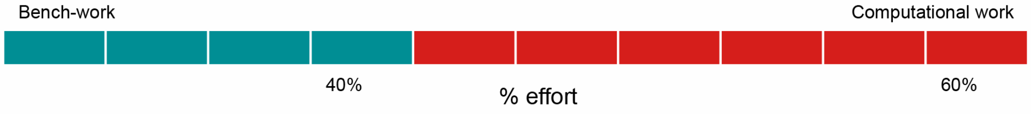
Career timeline



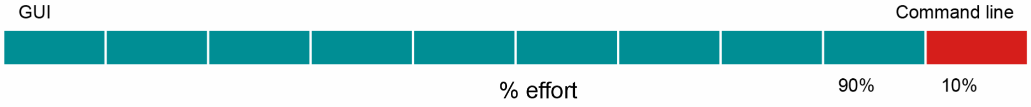
Typical activities



Distribution of time between bench-work and computational work



Preference for using GUI vs command line



Drivers

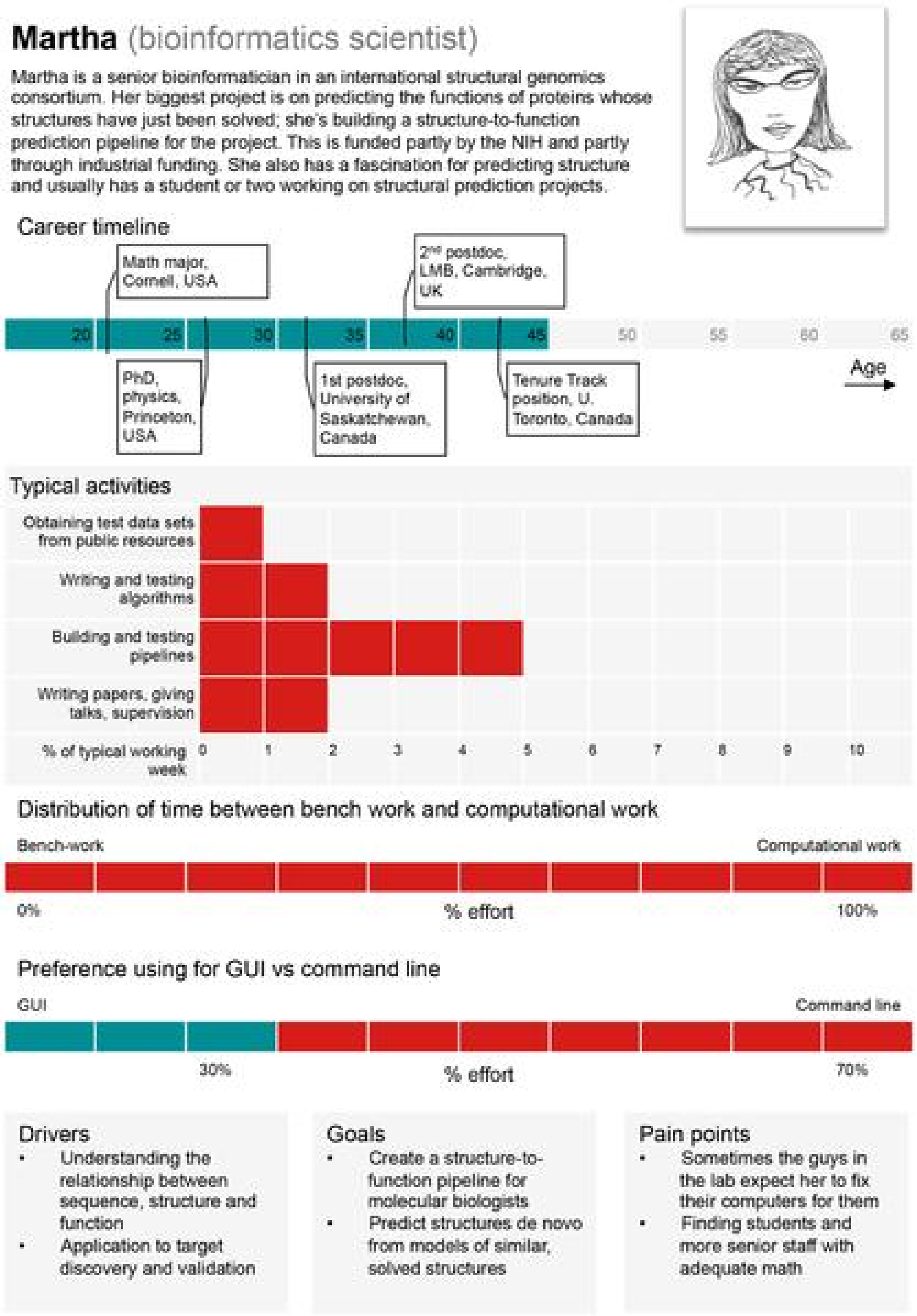
- Understanding what makes a usually harmless bacterium pathogenic in the lungs of people with cystic fibrosis

Goals

- QA of -omics data
- Statistical analysis of data
- Data integration and pathway analysis

Pain points

- Lack of access to departmental compute farm
- Sporadic to non-existent access to bioinformatics support



Martha (bioinformatics scientist)

Martha is a senior bioinformatician in an international structural genomics consortium. Her biggest project is on predicting the functions of proteins whose structures have just been solved; she's building a structure-to-function prediction pipeline for the project. This is funded partly by the NIH and partly through industrial funding. She also has a fascination for predicting structure and usually has a student or two working on structural prediction projects.

Career timeline

Math major, Cornell, USA
2nd postdoc, LMB, Cambridge, UK
20
25
30
35
40
45
50
55
60
65
Age
PhD, physics, Princeton, USA
1st postdoc, University of Saskatchewan, Canada
Tenure Track position, U. Toronto, Canada

Typical activities

Obtaining test data sets from public resources
Writing and testing algorithms
Building and testing pipelines
Writing papers, giving talks, supervision
% of typical working week
0
1
2
3
4
5
6
7
8
9
10

Distribution of time between bench work and computational work

Preference using for GUI vs command line

Drivers

- Understanding the relationship between sequence, structure and function
- Application to target discovery and validation

Goals

- Create a structure-to-function pipeline for molecular biologists
- Predict structures de novo from models of similar, solved structures

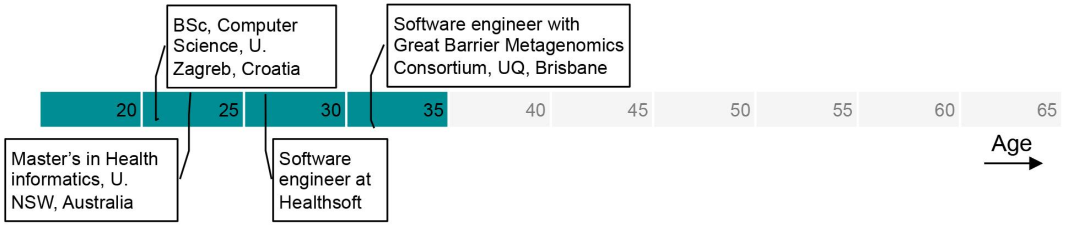
Pain points

- Sometimes the guys in the lab expect her to fix their computers for them
- Finding students and more senior staff with adequate math

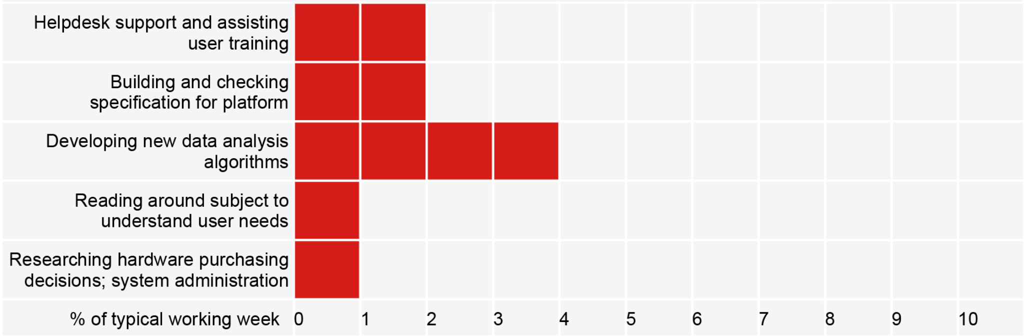
Ivan (bioinformatics engineer)

Ivan has just started a new support role in a bioinformatics core facility after working for an electronic health records company for four years. His main project is to develop a major new data integration platform for metagenomics data from coral reefs, but he also has to take his share of helpdesk queries on other projects. “I come from a computer science background, so talking the same language as the guys analysing the data is a bit of a challenge,” he says. “I also didn’t really figure that I’d be working on the GUI as well as the code – in my last job we had design folks to take care of that”.

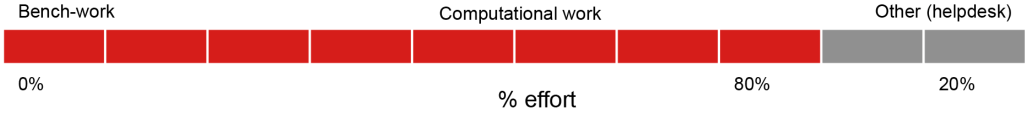
Career timeline



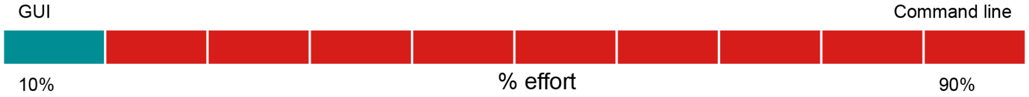
Typical activities



Distribution of time between bench-work and computational work



Preference for using GUI vs command line



Drivers

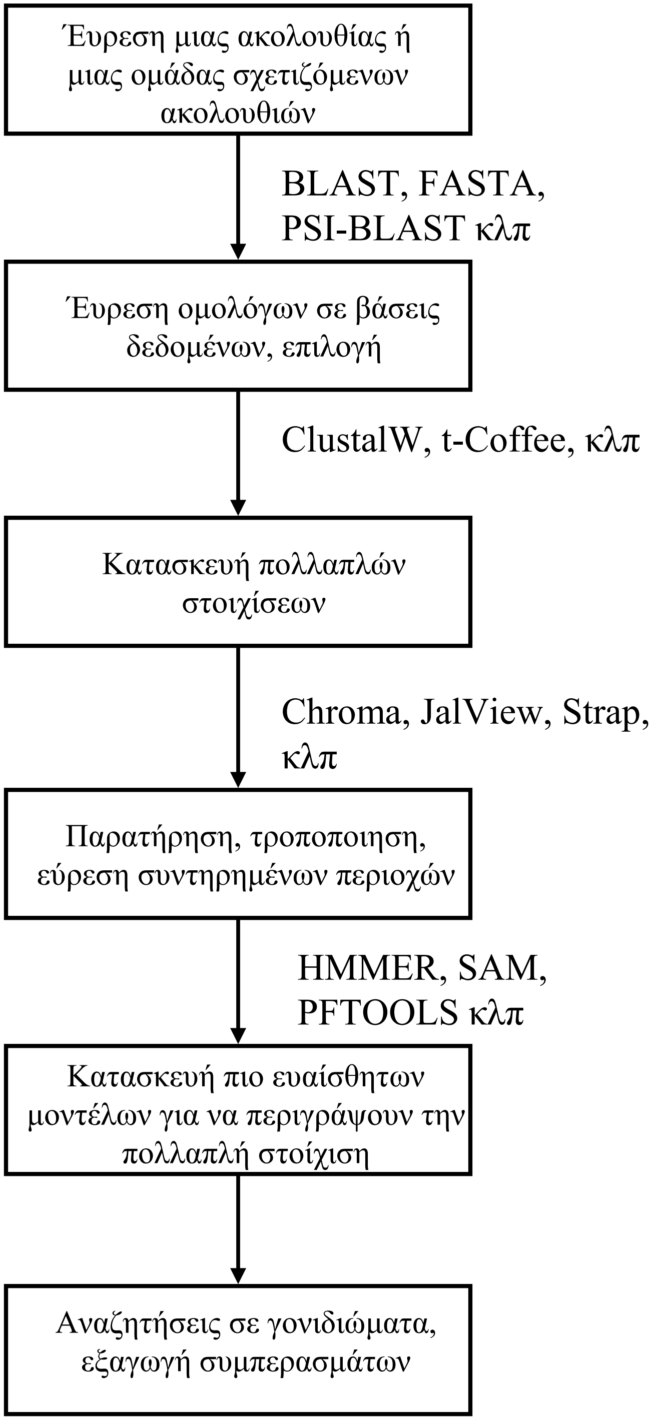
- Writing algorithms and developing a platform to support novel research
- Supporting other research projects in a busy academic department

Goals

- Define a spec that meets the needs of his users
- Prototype and build part of the platform
- Make sure his part of the project complements others

Pain points

- Has to work with another software engineer who isn’t a team player
- Sometimes struggles to interpret what his users want



Έυρεση μιας ακολουθίας ή μιας ομάδας σχετιζόμενων ακολουθιών

BLAST, FASTA, PSI-BLAST κλπ

Έυρεση ομολόγων σε βάσεις δεδομένων, επιλογή

ClustalW, t-Coffee, κλπ

Κατασκευή πολλαπλών στοιχίσεων

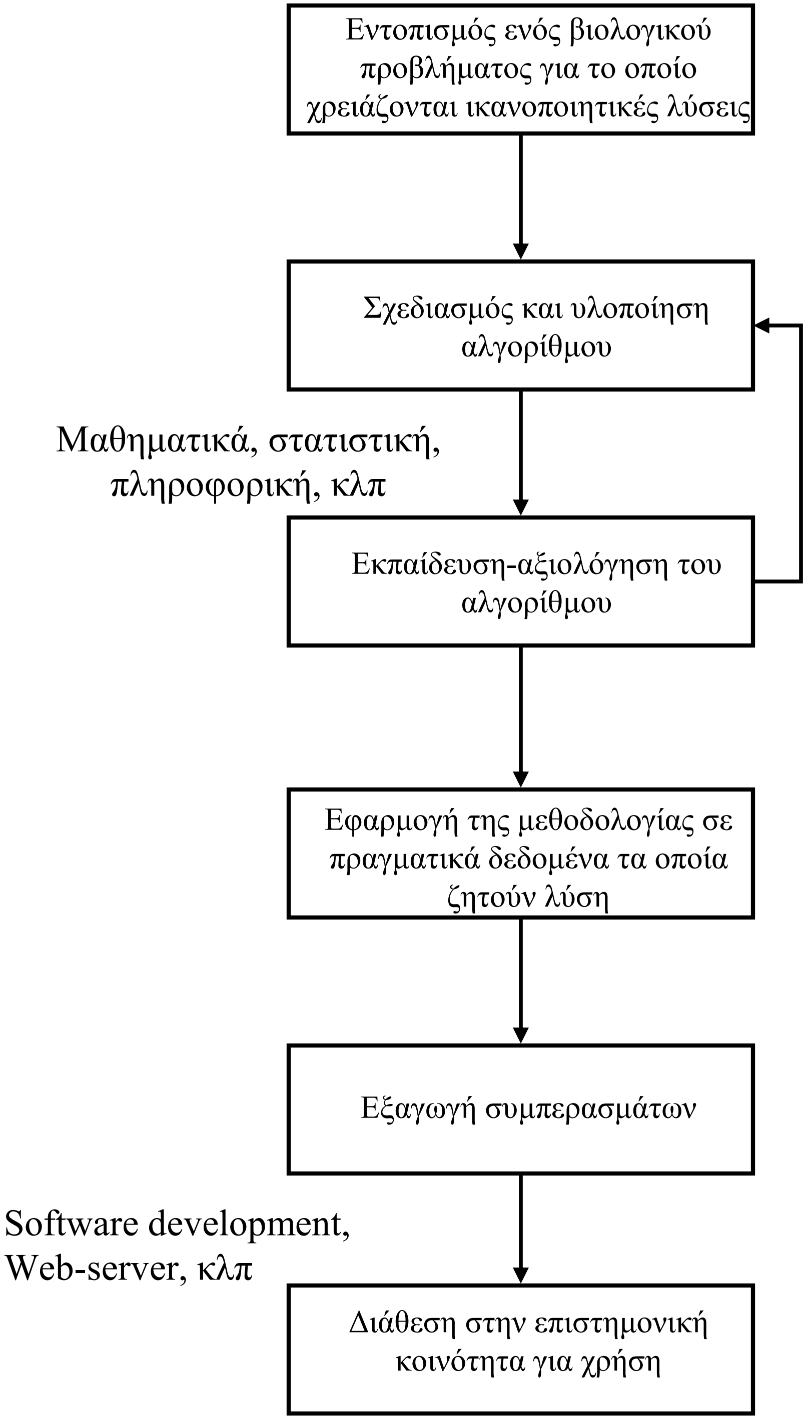
Chroma, JalView, Strap, κλπ

Παρατήρηση, τροποποιηση, εύρεση συντηρημένων περιοχών

HMMER, SAM, PFTOOLS κλπ

Κατασκευή πιο ευαίσθητων μοντέλων για να περιγράψουν την πολλαπλή στοίχιση

Αναζητήσεις σε γονιδιώματα, εξαγωγή συμπερασμάτων



Εντοπισμός ενός βιολογικού προβλήματος για το οποίο χρειάζονται ικανοποιητικές λύσεις

Σχεδιασμός και υλοποίηση αλγορίθμου

Μαθηματικά, στατιστική, πληροφορική, κλπ

Εκπαίδευση-αξιολόγηση του αλγορίθμου

Εφαρμογή της μεθοδολογίας σε πραγματικά δεδομένα τα οποία ζητούν λύση

Εξαγωγή συμπερασμάτων

Software development, Web-server, κλπ

Διάθεση στην επιστημονική κοινότητα για χρήση

Published data related to Bioinformatics community and research in Greece

- Little (valuable efforts made by HSCBB)
- A previous bibliometric study analyzed the scientific activity in Bioinformatics in Greece identified 405 published research conducted from 1976 until 2010
- This research showed that:

  the oldest and largest Universities

  seem dominant *and*

  some of the newer Universities have strong presence

- Research and teaching activity regarding Bioinformatics the last 15 years in Greece is booming
- A sufficient number of scientists who will advance the field has started to accumulate
- The origin of these scientists appears to be interdisciplinary
- Bioinformatics researchers come from some different scientific discipline and are not exclusively devoted to a given scientific field

Διαιρέσεις της Βιοπληροφορικής

- Ανάλυση ακολουθιών
- Ανάλυση δομών (δομική Βιοπληροφορική)
- Βάσεις βιολογικών δεδομένων
- Ανάλυση γενετικών δεδομένων
- Ανάλυση δεδομένων έκφρασης
- Ανάλυση πολύπλοκων δικτύων

Ανάλυση ακολουθιών

- Στοίχιση ακολουθιών (κατά ζεύγη και πολλαπλή)
- Εύρεση προτύπων (patterns)
- Πρόβλεψη λειτουργικών και δομικών χαρακτηριστικών
- Ανάλυση γονιδιωμάτων, κλπ

Ανάλυση δομών

- Αναπαράσταση δομών
- Υπέρθεση δομών - στοίχιση
- Αναγνώριση πρωτεϊνικού διπλώματος
- Συγκριτική προτυποποίηση με βάση την ομολογία

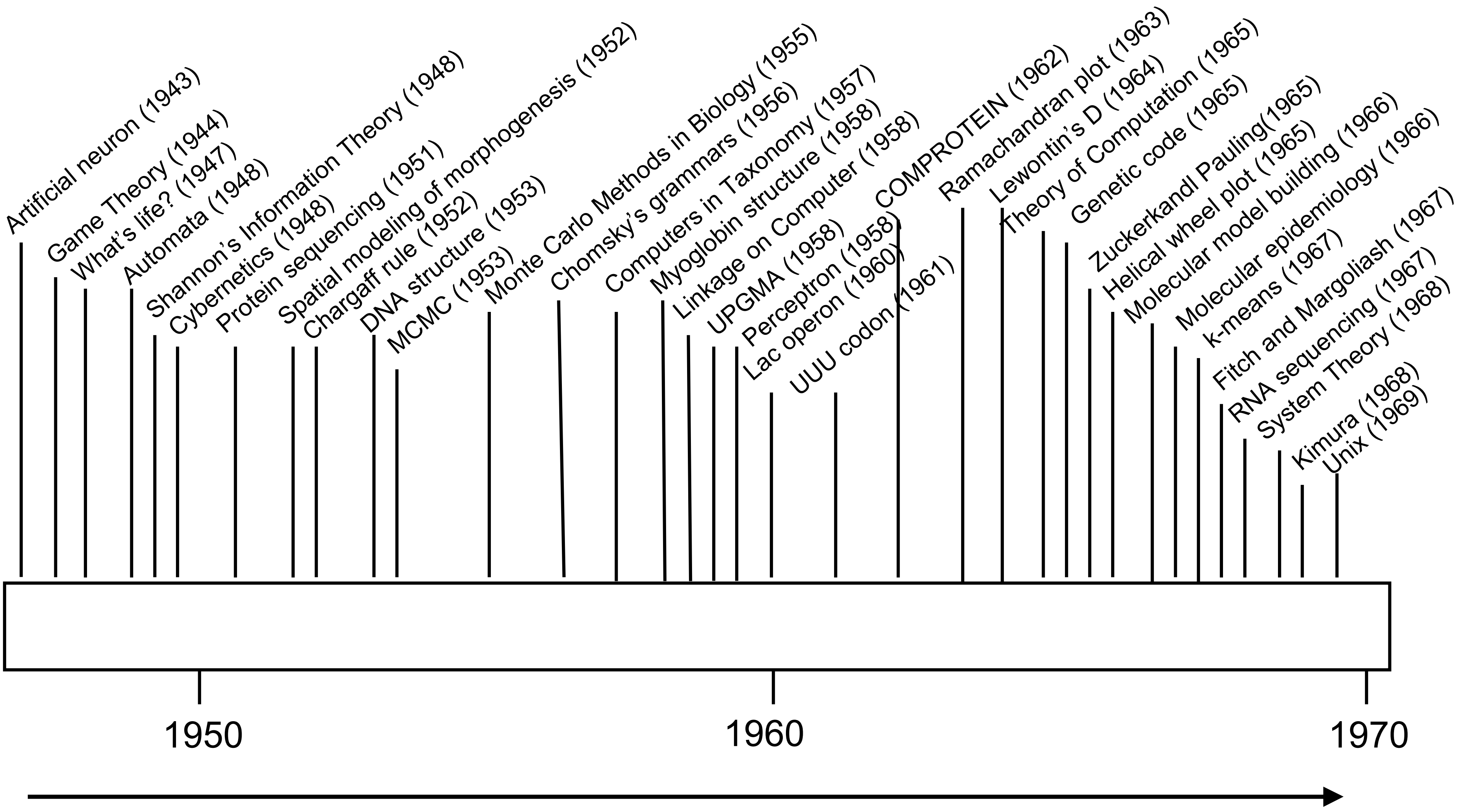
Βάσεις δεδομένων

- Συλλογή και σχολιασμός δεδομένων
- Σχεδιασμός βάσεων δεδομένων

Τα είδη των δεδομένων

- Ακολουθίες
- Δομές
- Άλλα

Ιστορια της Βιοπληροφορικής



Artificial neuron (1943)

Game Theory (1944)

What’s life? (1947)

Automata (1948)

Shannon’s Information Theory (1948)

Cybernetics (1948)

Protein sequencing (1951)

Spatial modeling of morphogenesis (1952)

Chargaff rule (1952)

DNA structure (1953)

MCMC (1953)

Monte Carlo Methods in Biology (1955)

Chomsky’s grammars (1956)

Computers in Taxonomy (1957)

Myoglobin structure (1958)

Linkage on Computer (1958)

UPGMA (1958)

Perceptron (1958)

Lac operon (1960)

UUU codon (1961)

COMPROTEIN (1962)

Ramachandran plot (1963)

Lewontin’s D (1964)

Theory of Computation (1965)

Genetic code (1965)

Zuckerkandl Pauling(1965)

Helical wheel plot (1965)

Molecular model building (1966)

Molecular epidemiology (1966)

k-means (1967)

Fitch and Margoliash (1967)

RNA sequencing (1967)

System Theory (1968)

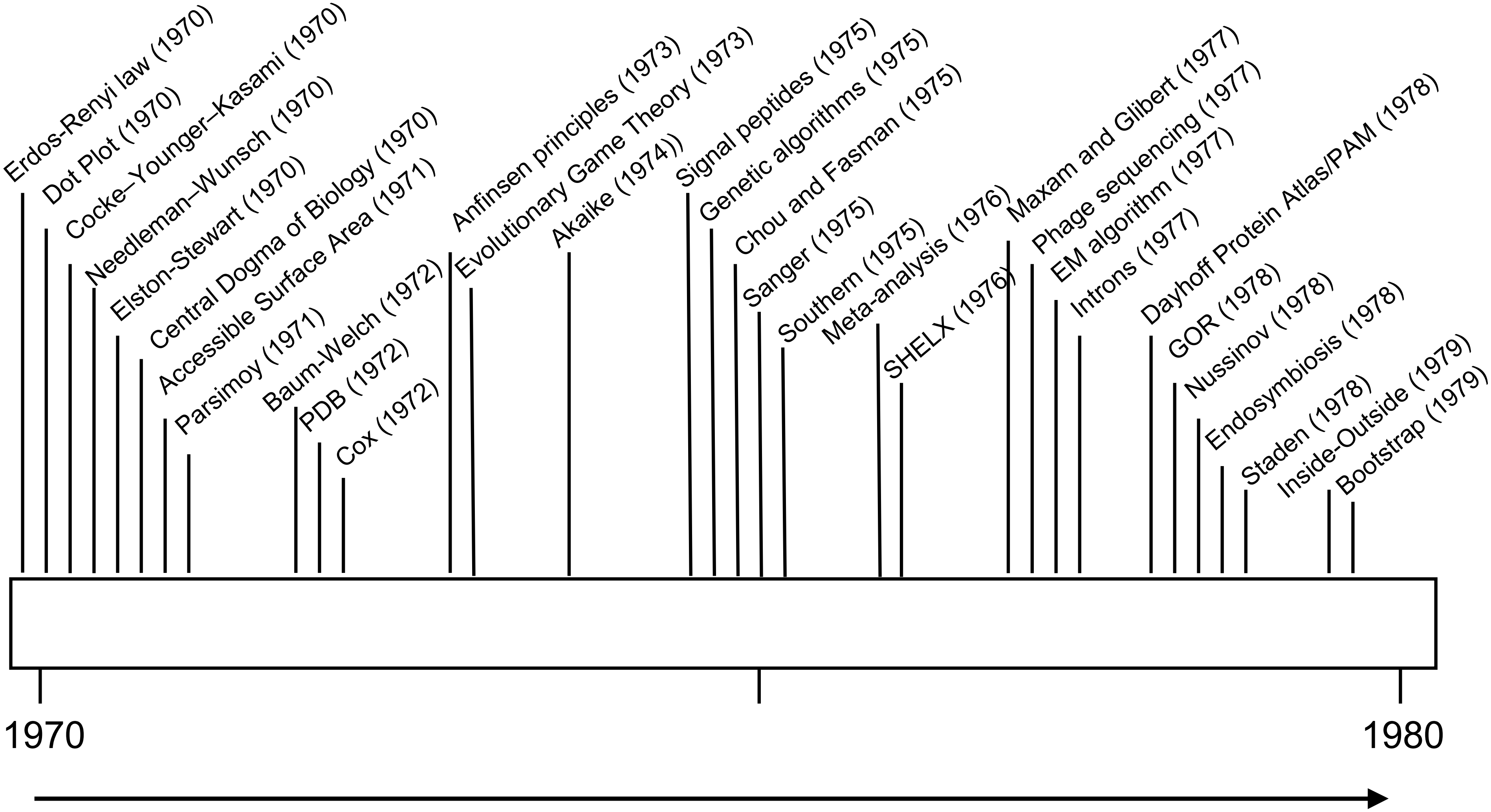
Kimura (1968)

Unix (1969)

1950

1960

1970



Erdos-Renyi law (1970)

Dot Plot (1970)

Cocke–Younger–Kasami (1970)

Needleman–Wunsch (1970)

Elston-Stewart (1970)

Central Dogma of Biology (1970)

Accessible Surface Area (1971)

Parsimoy (1971)

Baum-Welch (1972)

PDB (1972)

Cox (1972)

Anfinsen principles (1973)

Evolutionary Game Theory (1973)

Akaike (1974))

Signal peptides (1975)

Genetic algorithms (1975)

Chou and Fasman (1975)

Sanger (1975)

Southern (1975)

Meta-analysis (1976)

SHELX (1976)

Maxam and Glibert (1977)

Phage sequencing (1977)

EM algorithm (1977)

Introns (1977)

Dayhoff Protein Atlas/PAM (1978)

GOR (1978)

Nussinov (1978)

Endosymbiosis (1978)

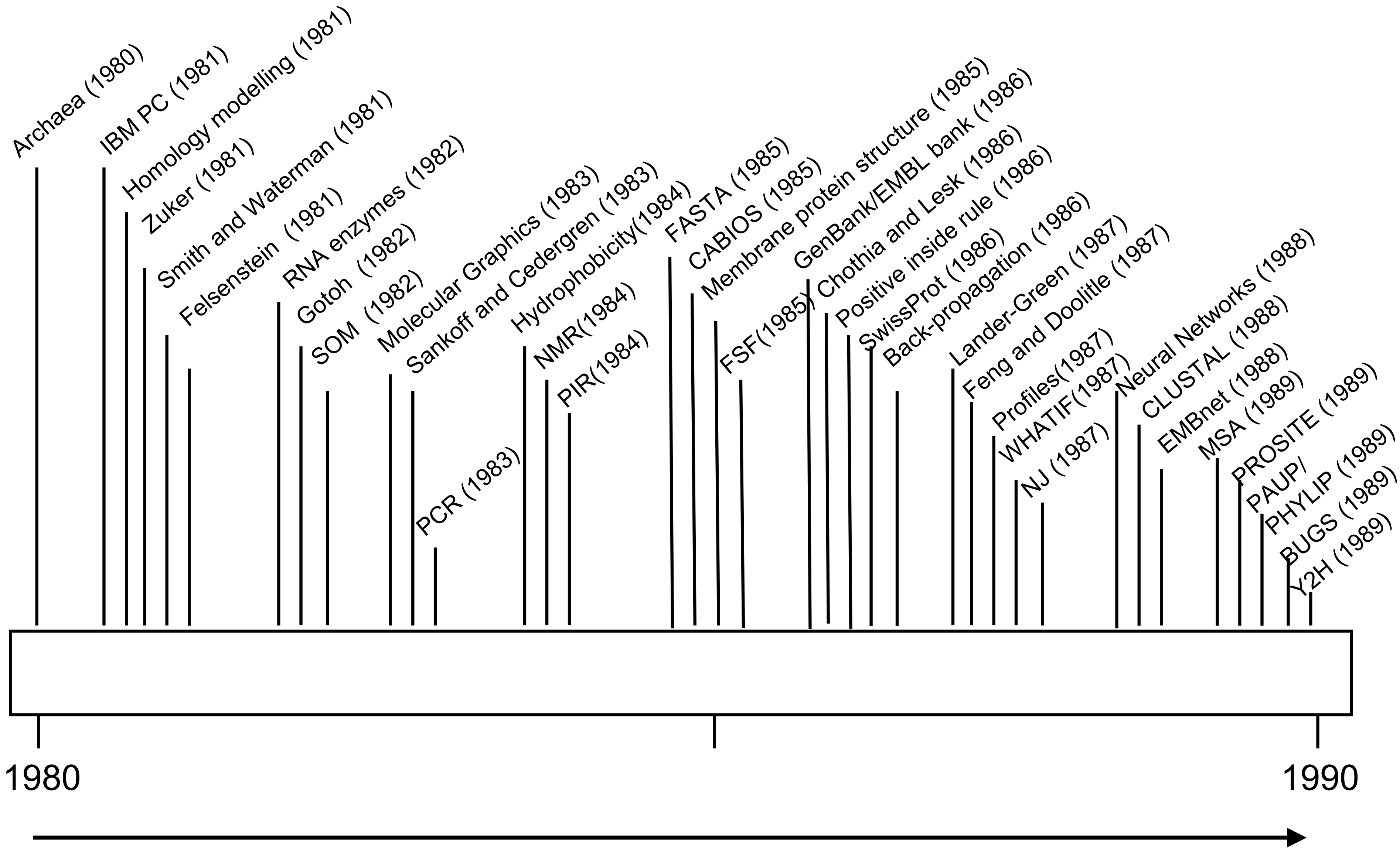
Staden (1978)

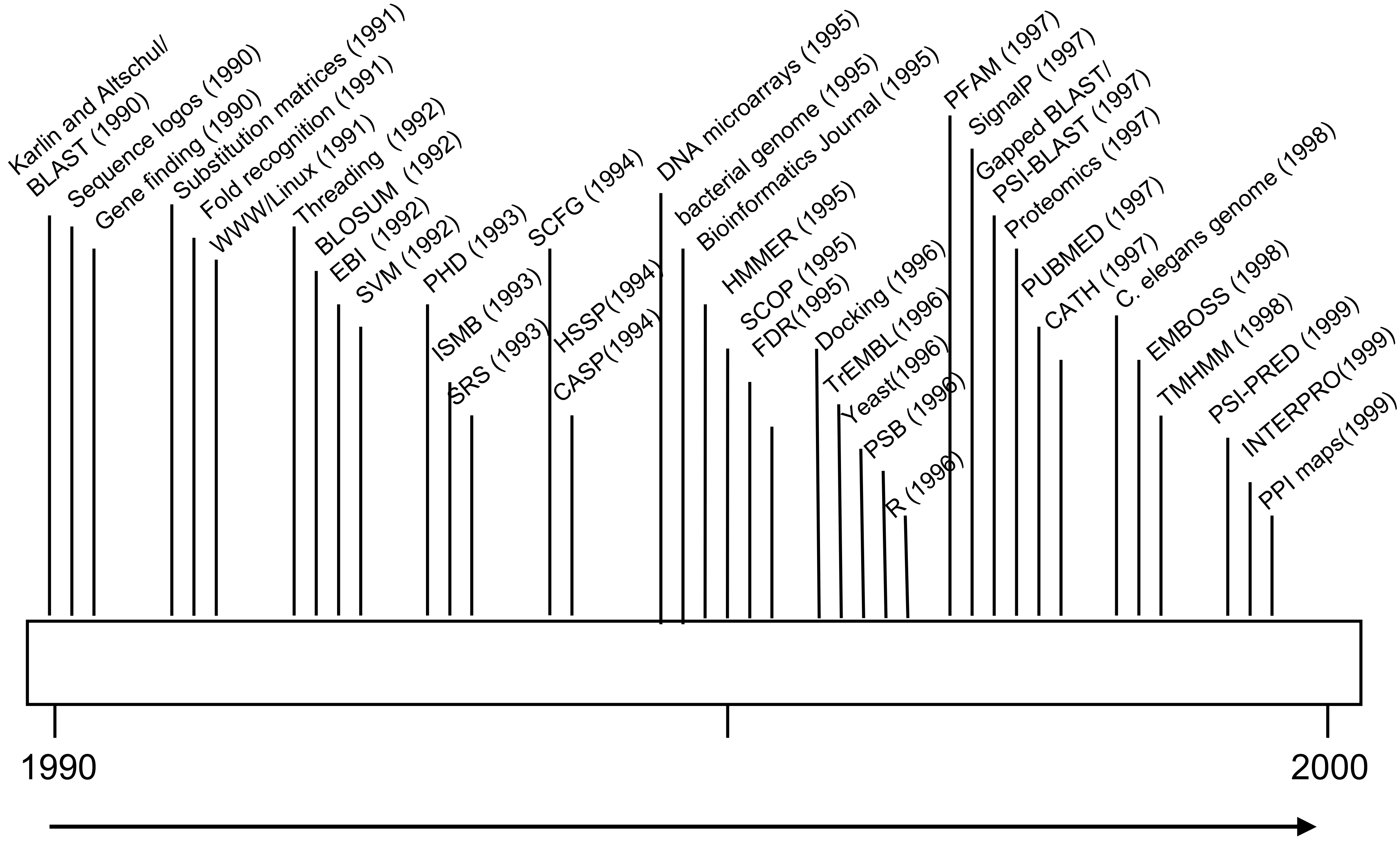
Inside-Outside (1979)

Bootstrap (1979)

1970

1980

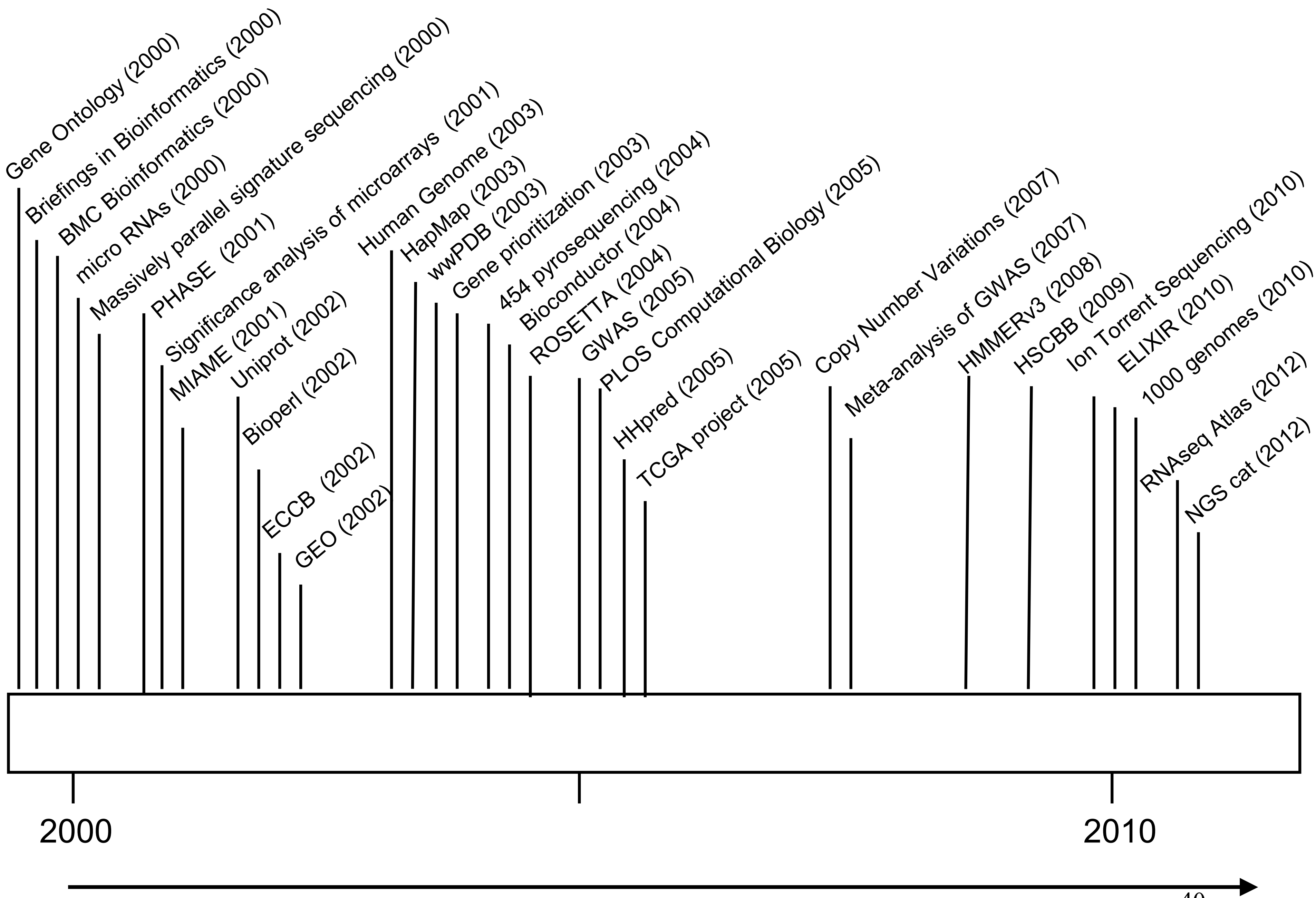

Archaea (1980)
IBM PC (1981)
Homology modelling (1981)
Zuker (1981)
Smith and Waterman (1981)
Felsenstein (1981)
RNA enzymes (1982)
Gotoh (1982)
SOM (1982)
Molecular Graphics (1983)
Sankoff and Cedergren (1983)
PCR (1983)
Hydrophobicity(1984)
NMR(1984)
PIR(1984)
FASTA (1985)
CABIOS (1985)
Membrane protein structure (1985)
FSF(1985)
GenBank/EMBL bank (1986)
Chothia and Lesk (1986)
Positive inside rule (1986)
SwissProt (1986)
Back-propagation (1986)
Lander-Green (1987)
Feng and Doolitle (1987)
Profiles(1987)
WHATIF(1987)
NJ (1987)
Neural Networks (1988)
CLUSTAL (1988)
EMBnet (1988)
MSA (1989)
PROSITE (1989)
PAUP/
PHYLIP (1989)
BUGS (1989)
Y2H (1989)
1980
1990

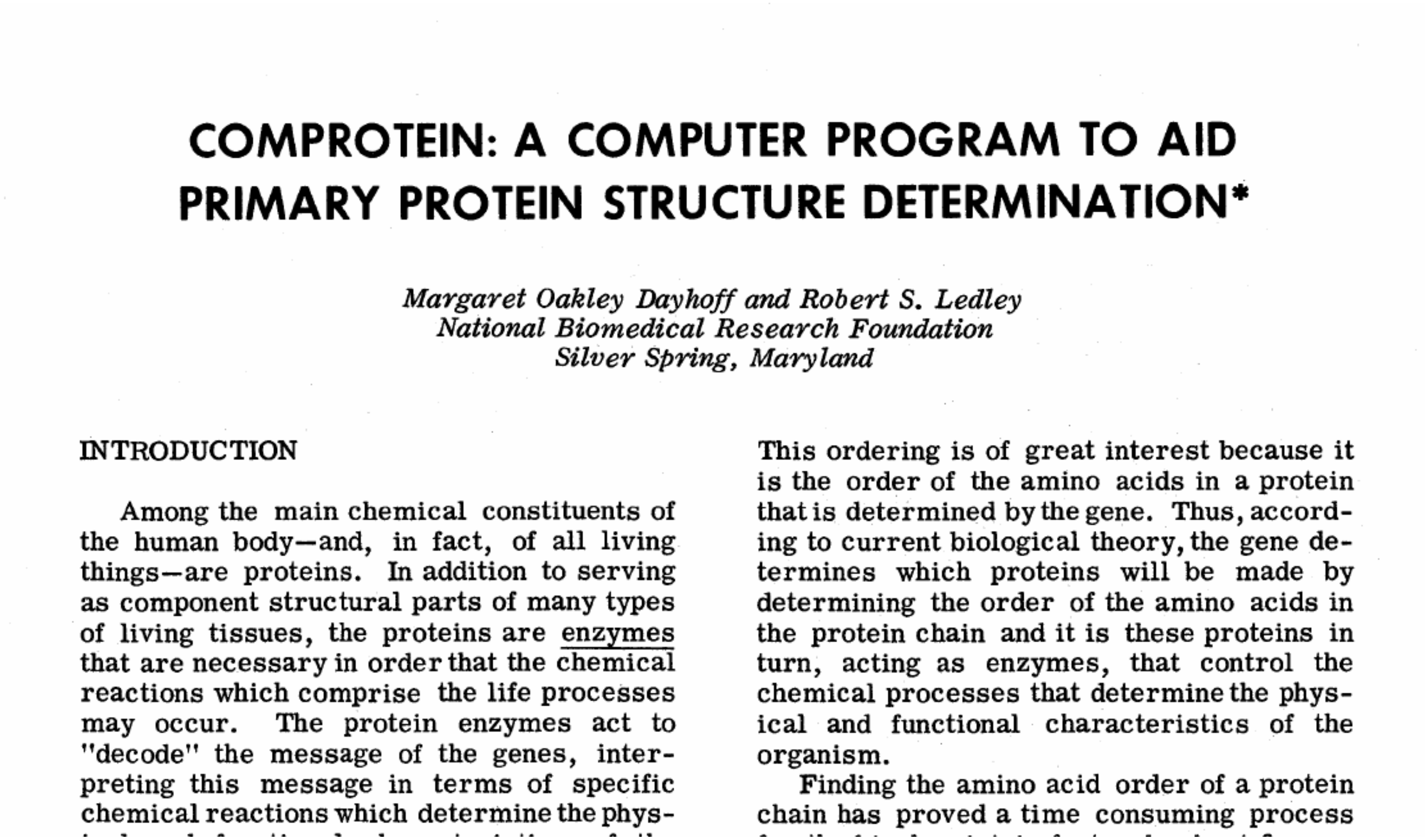


Karlin and Altschul/
BLAST (1990)
Sequence logos (1990)
Gene finding (1990)
Substitution matrices (1991)
Fold recognition (1991)
WWW/Linux (1991)
Threading (1992)
BLOSUM (1992)
EBI (1992)
SVM (1992)
PHD (1993)
ISMB (1993)
SRS (1993)
SCFG (1994)
HSSP(1994)
CASP(1994)
DNA microarrays (1995)
bacterial genome (1995)
Bioinformatics Journal (1995)
HMMER (1995)
SCOP (1995)
FDR(1995)
Docking (1996)
TrEMBL(1996)
Yeast(1996)
PSB (1996)
R (1996)
PFAM (1997)
SignalP (1997)
Gapped BLAST/
PSI-BLAST (1997)
Proteomics (1997)
PUBMED (1997)
CATH (1997)
C. elegans genome (1998)
EMBOSS (1998)
TMHMM (1998)
PSI-PRED (1999)
INTERPRO(1999)
PPI maps(1999)

1990

2000


Gene Ontology (2000)
Briefings in Bioinformatics (2000)
BMC Bioinformatics (2000)
micro RNAs (2000)
Massively parallel signature sequencing (2000)
PHASE (2001)
Significance analysis of microarrays (2001)
MIAME (2001)
Uniprot (2002)
Bioperl (2002)
ECCB (2002)
GEO (2002)
Human Genome (2003)
HapMap (2003)
wwPDB (2003)
Gene prioritization (2003)
454 pyrosequencing (2004)
Bioconductor (2004)
ROSETTA (2004)
GWAS (2005)
PLOS Computational Biology (2005)
HHpred (2005)
TCGA project (2005)
Copy Number Variations (2007)
Meta-analysis of GWAS (2007)
HMMERv3 (2008)
HSCBB (2009)
Ion Torrent Sequencing (2010)
ELIXIR (2010)
1000 genomes (2010)
RNAseq Atlas (2012)
NGS cat (2012)
2000
2010



COMPROTEIN: A COMPUTER PROGRAM TO AID PRIMARY PROTEIN STRUCTURE DETERMINATION*

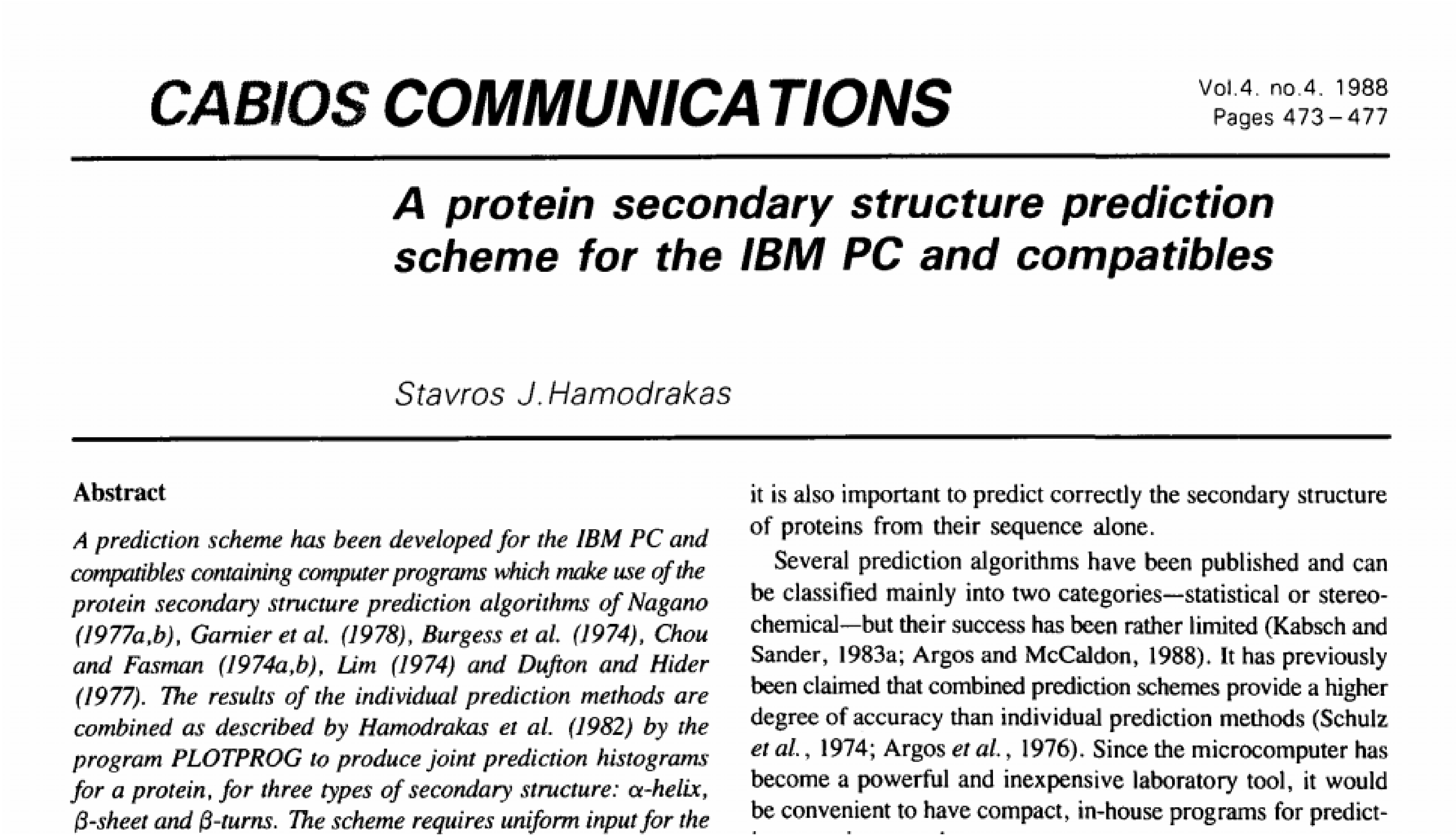
Margaret Oakley Dayhoff and Robert S. Ledley
National Biomedical Research Foundation
Silver Spring, Maryland

INTRODUCTION

Among the main chemical constituents of the human body—and, in fact, of all living things—are proteins. In addition to serving as component structural parts of many types of living tissues, the proteins are enzymes that are necessary in order that the chemical reactions which comprise the life processes may occur. The protein enzymes act to "decode" the message of the genes, interpreting this message in terms of specific chemical reactions which determine the phys-

This ordering is of great interest because it is the order of the amino acids in a protein that is determined by the gene. Thus, according to current biological theory, the gene determines which proteins will be made by determining the order of the amino acids in the protein chain and it is these proteins in turn, acting as enzymes, that control the chemical processes that determine the physical and functional characteristics of the organism.

Finding the amino acid order of a protein chain has proved a time consuming process



CABIOS COMMUNICATIONS

Vol.4. no.4. 1988
Pages 473 – 477

A protein secondary structure prediction scheme for the IBM PC and compatibles

Stavros J.Hamodrakas

Abstract

A prediction scheme has been developed for the IBM PC and compatibles containing computer programs which make use of the protein secondary structure prediction algorithms of Nagano (1977a,b), Garnier et al. (1978), Burgess et al. (1974), Chou and Fasman (1974a,b), Lim (1974) and Dufton and Hider (1977). The results of the individual prediction methods are combined as described by Hamodrakas et al. (1982) by the program PLOTPROG to produce joint prediction histograms for a protein, for three types of secondary structure: α-helix, β-sheet and β-turns. The scheme requires uniform input for the

it is also important to predict correctly the secondary structure of proteins from their sequence alone.

Several prediction algorithms have been published and can be classified mainly into two categories—statistical or stereochemical—but their success has been rather limited (Kabsch and Sander, 1983a; Argos and McCaldon, 1988). It has previously been claimed that combined prediction schemes provide a higher degree of accuracy than individual prediction methods (Schulz *et al.*, 1974; Argos *et al.*, 1976). Since the microcomputer has become a powerful and inexpensive laboratory tool, it would be convenient to have compact, in-house programs for predict-

Education in Greece

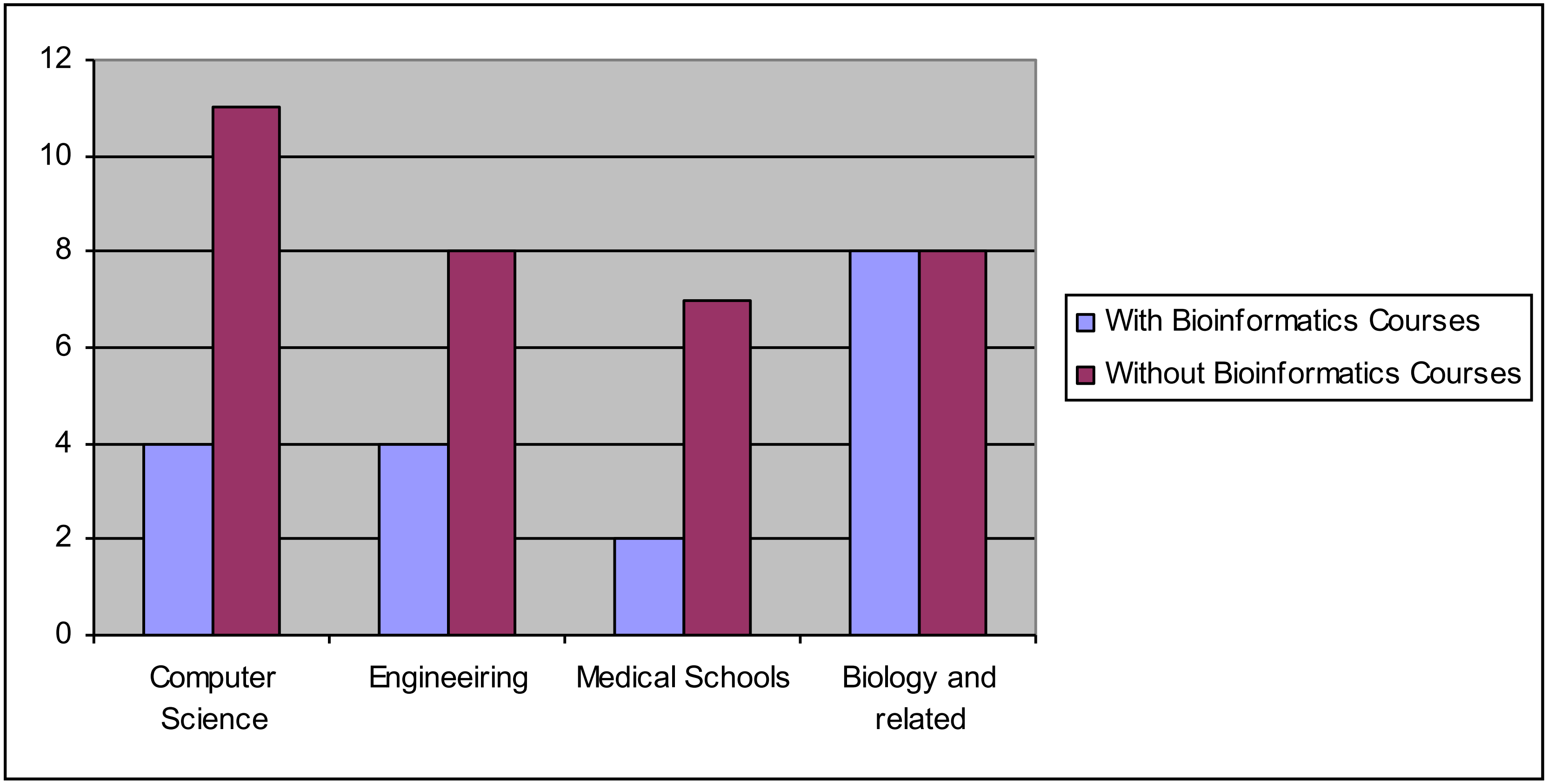
- Bioinformatics is taught at undergraduate level in 18 Departments
  - 8 Departments of Biological Sciences
  - 4 Departments of Computer Engineering
  - 4 Computer Science Departments
  - 2 Medical Schools
- In total there are 23 different courses
  - 3 of them are currently not offered, mostly in CS departments
- 22 faculty members involved
  - some courses are taught by more than one faculty member, whereas some others teach more than one courses

Faculty members

Degree	N	%
Biology	10	45.45
Physics	3	13.64
Chemistry	3	13.64
Computer Science	2	9.09
Engineer	2	9.09
Mathematics	2	9.09

Bioinformatics in curricula

- University of Thessaly
  - Department of Computer Science and Biomedical Informatics (3 courses*),
  - Department of Biochemistry and Biotechnology (1 course*),
  - Department of Electrical and Computer Engineering (1 course),
  - School of Medicine (1 course*)
- University of Thrace
  - Department of Molecular Biology and Genetics(4 courses*),
  - Department of Electrical and Computer Engineering (1 course)
- University of Crete
  - Department of Biology (1 course*),
  - School of Medicine (1 course*),
  - Department of Computer Science (2 courses*)
- University of Athens
  - Department of Biology (1 course*),
  - Department of Informatics and Telecommunications (1 course)
- University of Patras
  - Department of Biology (1 course)
  - Department of Computer Engineering and Informatics (1 course)
- University of Ioannina
  - Department of Biological Applications and Technologies (2 courses*)
- University of Piraeus
  - Department of Informatics (1 course)
- University of Western Macedonia
  - Department of Engineering Informatics and Telecommunications (1 course)
- University of Thessaloniki
  - Department of Biology (1 course)
- Agricultural University of Athens
  - Department of Biotechnology (1 course*)


12
10
8
6
4
2
0
Computer Science
Engineeiring
Medical Schools
Biology and related
With Bioinformatics Courses
Without Bioinformatics Courses

Structure of the curriculum

- Departments of Biology and Medical Schools:
  - Computing/ Programming
  - Biomathematics/ Biostatistics
  - Bioinformatics

- Departments of Computer Science and Engineering:
  - Probability and/or Statistics
  - Molecular Biology/Genetics
  - Bioinformatics

- This work’s long-term goal → draw conclusions also via interviews – presented in print at a later time- taken from Bioinformatics courses’ lecturers

Conclusions

- The Departments of Biological Sciences seem to have adapted to the new era of Bioinformatics
- All of them (8/8) include one or more relevant courses in their curricula
- The Departments of Computer Science, Computer Engineering, and the Medical Schools follow to a lesser extent
- Moreover, the Departments of Biology have achieved a smooth introduction of the students to Bioinformatics, including courses on Biostatistics and Biomathematics, and some of them (but not all) courses on Programming and Informatics
- The same is not the case for CS and Engineering departments, with the notable exception of the Department of the University of Crete which offers Biology Courses and the Department of CS and Biomedical Informatics which offers also courses in Biology, Biochemistry, Genetics, Physiology and Biostatistics.

Conclusions

- Older Universities played a crucial role in establishing Bioinformatics research and education in Greece
- Universities of the Periphery being newer and more flexible, invested in the field, by including more relevant courses in the curricula and by hiring faculty members trained in the field
- Biological Sciences Departments that offer the larger number of relevant courses:
  - Department of Molecular Biology and Genetics of the University of Thrace (4 courses +Biostatistics)
  - Department of Biological Applications and Technologies of the University of Ioannina (2 courses +Programming+Biostatistics)
  - Department of Biochemistry and Biotecnology of the University of Thessaly (1 course+ Programming+Biostatistics)
  - Department of Biology of the University of Crete (1 course+ Programming+Biostatistics)
- Departments of Computer Science that managed to include several courses of Bioinformatics and introduce Computer Science students better in this field :
  - Department of Computer Science and Biomedical Informatics of the University of Thessaly (3 courses+Biostatistics+Biochemistry+Biology I+Biology II+Genetics)
  - Department of Computer Science of the University of Crete (2 courses+2 courses on Biology)

What about Postgraduate studies?

- University of Athens
  - MSc Programme in «Bioinformatics» (Department of Biology)
  - MSc Programme in «Information Technologies in Medicine and Biology» (Specialization in «Bioinformatics»)
- University of Thessaly
  - MSc Programme in «Informatics and Computational Biomedicine» (Specialization in «Computational Medicine and Biology»)
  - MSc Programme in «Methodology of Biomedical Research, Biostatistics and Clinical Bioinformatics»
- University of Patras
  - MSc Programme in «Life Sciences Informatics» (Specialization in «Bioinformatics»)
- Agricultural University
  - MSc Programme in «Systems Biology»