



ROSETTA
INPHARMATICS

What is Bioinformatics?

Mark Boguski, M.D., Ph.D.
Bio 2001



ROSETTA
INPHARMATICS

www.rii.com

Origin of the term “Bioinformatics”

- “Informatics” coined in the early 1990s
- Arose in the context of early discussions of electronic publishing
- Definition: Informatics = Information Science & Technology
- Variations:
 - Medical Informatics
 - Bioinformatics (circa 1992)



ROSETTA
INPHARMATICS

www.rii.com

“Bioinformatics” extends back to the 1960s

J Theor Biol 1965 Jan;8(1):97-112

Computer aids to protein sequence determination.

Dayhoff MO.

Sci Am 1969 Jul;221(1):86-95

Computer analysis of protein evolution.

Dayhoff MO.



ROSETTA
INPHARMATICS

www.rii.com

Bioinformatics in the 1970s

Ann N Y Acad Sci 1974 Nov 29;241(0):439-48

A comparison between evolutionary substitutions and variants in human hemoglobins.

Fitch WM.

J Mol Evol 1975 Jun 9;5(1):1-24

Phylogenies from amino acid sequences aligned with gaps: the problem of gap weighting.

Fitch WM, Yasunobu KT.



ROSETTA
INPHARMATICS

www.rii.com

A watershed event in the early 1980s

Science 1983 Jul 15;221(4607):275-7

Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor.

Doolittle RF, Hunkapiller MW, Hood LE, Devare SG, Robbins KC, Aaronson SA, Antoniades HN.

Nature 1983 Jul 7-13;304(5921):35-9

Platelet-derived growth factor is structurally related to the putative transforming protein p28sis of simian sarcoma virus.

Waterfield MD, Scrace GT, Whittle N, Stroobant P, Johnsson A, Wasteson A, Westermark B, Heldin CH, Huang JS, Deuel TF.



ROSETTA
INPHARMATICS

www.rii.com

But what is Bioinformatics?



- Practical definitions
- Value first recognized by industry
- Universities slow to respond with training programs
 - Initially not viewed as an academic discipline

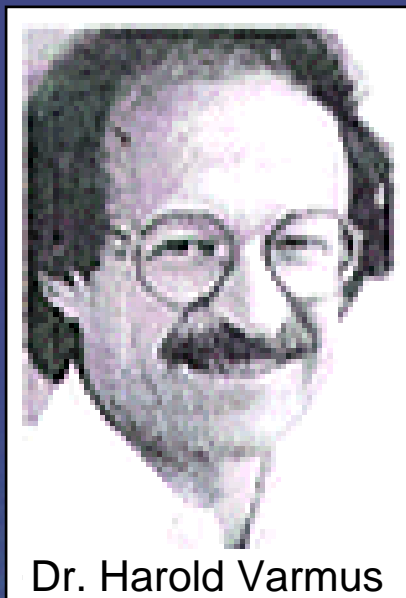


ROSETTA
INPHARMATICS

www.rii.com

“NIH Urged to Train Biologists on Computers”

Headline in The Washington Post, Monday June 7 1999



Recommendation of Federal Advisory Panel to NIH Director Varmus:

Establish 20 new U.S. centers to teach computer-based biomedical research at a cost of US\$8M per center per year.

Why?



ROSETTA
INPHARMATICS

www.rii.com

“It’s sink or swim as a tidal wave of data approaches”

Nature 399:517 10 June 1999



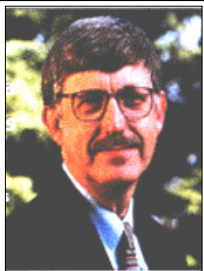
ROSETTA
INPHARMATICS

www.rii.com

The Accelerating Human Genome Project

Nature (September, 1998)

Human genome deadline cut by two years



Collins

POLICY FORUM: GENOMICS

Science (October, 1998)

The Human Genome Project: Reaching the Finish Line

R. Waterston and J. E. Sulston



Waterston

Nature (March, 1999)

Human Genome Project aims to finish 'working draft' next year



Gibbs

TECHVIEW: DNA SEQUENCING

Science (March, 1999)

Sequencing the Genome, Fast

James C. Mullikin and Amanda A. McMurray

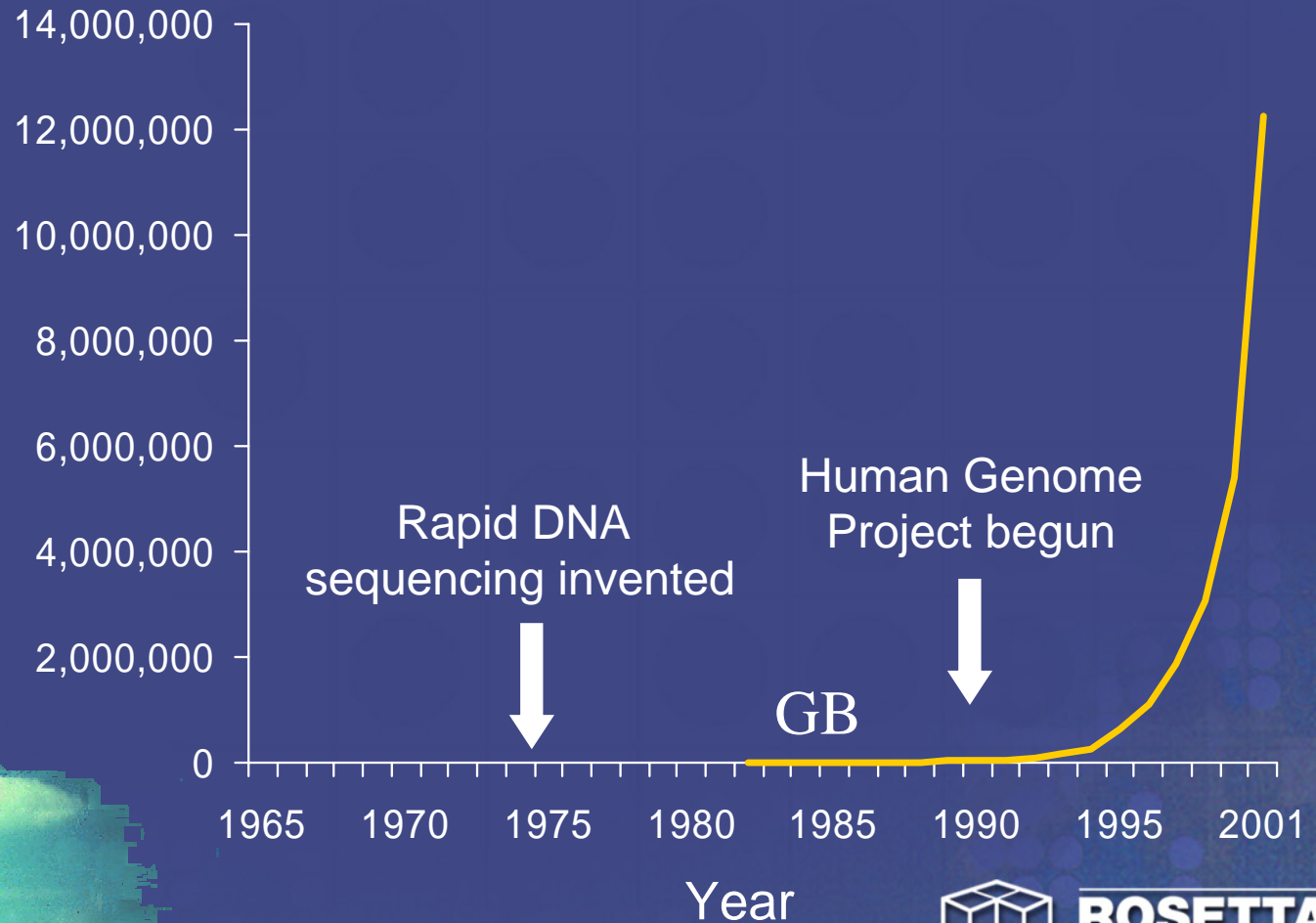


Lander

The rate at which DNA sequences began accumulating was exponential

Over 12 million
sequence entries
in GenBank

Nearly 13 billion
bases from
~50,000 species



National Library of Medicine



ROSETTA
INPHARMATICS

TRENDS GUIDE TO BIOINFORMATICS

Database searching
Sequence alignment
Gene finding
Functional genomics
Protein classification
Phylogenies



Trends Supplement 1998

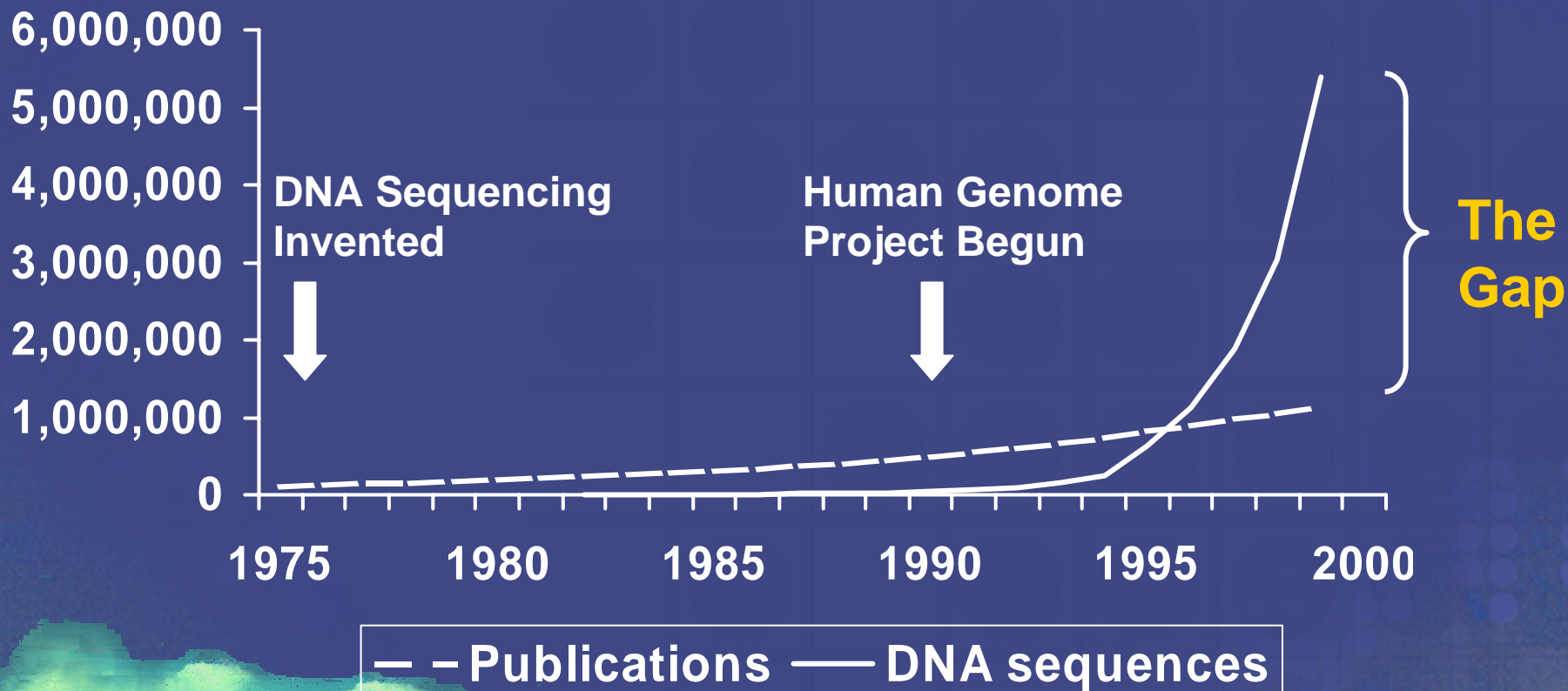
Survey of the field in 1998



ROSETTA
INPHARMATICS

www.rii.com

How do we bridge the gap between sequence and function?



Science (Genome Issue)
15 Oct. 1999

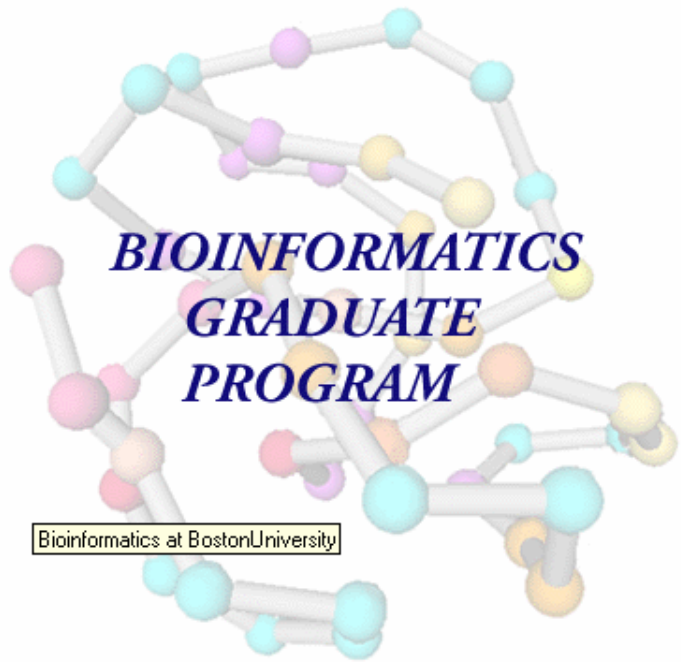


ROSETTA
INPHARMATICS



**BOSTON
University**

- About Bioinformatics
- People
- Admission
- Curriculum
- Opportunities
- Research Areas
- Announcements
- Contact Information



Bioinformatics at BostonUniversity



[BIOINFORMATICS RESOURCES LINKS](#)

Please send all comments to: bioinfo@bu.edu

Universities finally respond to the demand



**ROSETTA
INPHARMATICS**

www.rii.com

BU Bioinformatics: Curriculum - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Print

Address <http://bioinformatics.bu.edu/bioinformatics/curriculum.html> Go Links

ENG BE561: Protein and DNA Sequence Analysis
Presents fundamental concepts from molecular biology and molecular genetics. Teaches how to make biological inferences from DNA and protein sequences using mathematical and computer science techniques. Pair-wise sequence comparison; extension to multiple sequence alignment and conserved sequence pattern recognition; phylogenetic trees identification of coding regions; fragment assembly. Mathematical models and computational algorithms for genetic regulation. An introduction to protein 3-dimensional structure prediction.

ENG BE768: Biological Database Systems
Describes relational data models and database management systems; teaches the theories and techniques of constructing relational databases to store various biological data, including sequences, structures, genetic linkages and maps, and signal pathways. Introduces relational database query language SQL and the ORACLE database management system, with an emphasis on answering biologically important questions. Summarizes currently existing biological databases. Describes Web based programming tools to make databases accessible. Addresses questions in data integration and security. The future directions for biological database development are also discussed.

ENG BE700: Computational Genomics
This course "BE700/777 Computational Genomics" is a sequel of the other two core computational courses - "BE561 Protein and DNA sequence Analysis" and "BE768 Biological Database Analysis" (taught by Bob Berwick). The course number is BE700 since it is offered the first time and will become BE777 in the future. BE777 is a "hands on" course, and the goal is to apply theories and algorithms taught in BE561 and BE768 to real-life data sets, such as entire genomes.

Done Internet

Three of the
seven core
courses at B.U.



ROSETTA
INPHARMATICS

www.rii.com

Directions in Post-Genome Biology

- Genetic variation and human disease
- Comparative genomics
- Proteomics
- Microdevices and microsystems
- Innovative microarray applications
- Informatics
- Modeling and simulation



ROSETTA
INPHARMATICS

www.rii.com

Informatics

- Service functions
 - Periodic re-assembly of genome sequence and updating of annotation
- Infrastructure development
 - New tools for visualization & comprehension
- New theoretical formulations of complex units of genetic information
- Do we have enough compute power?



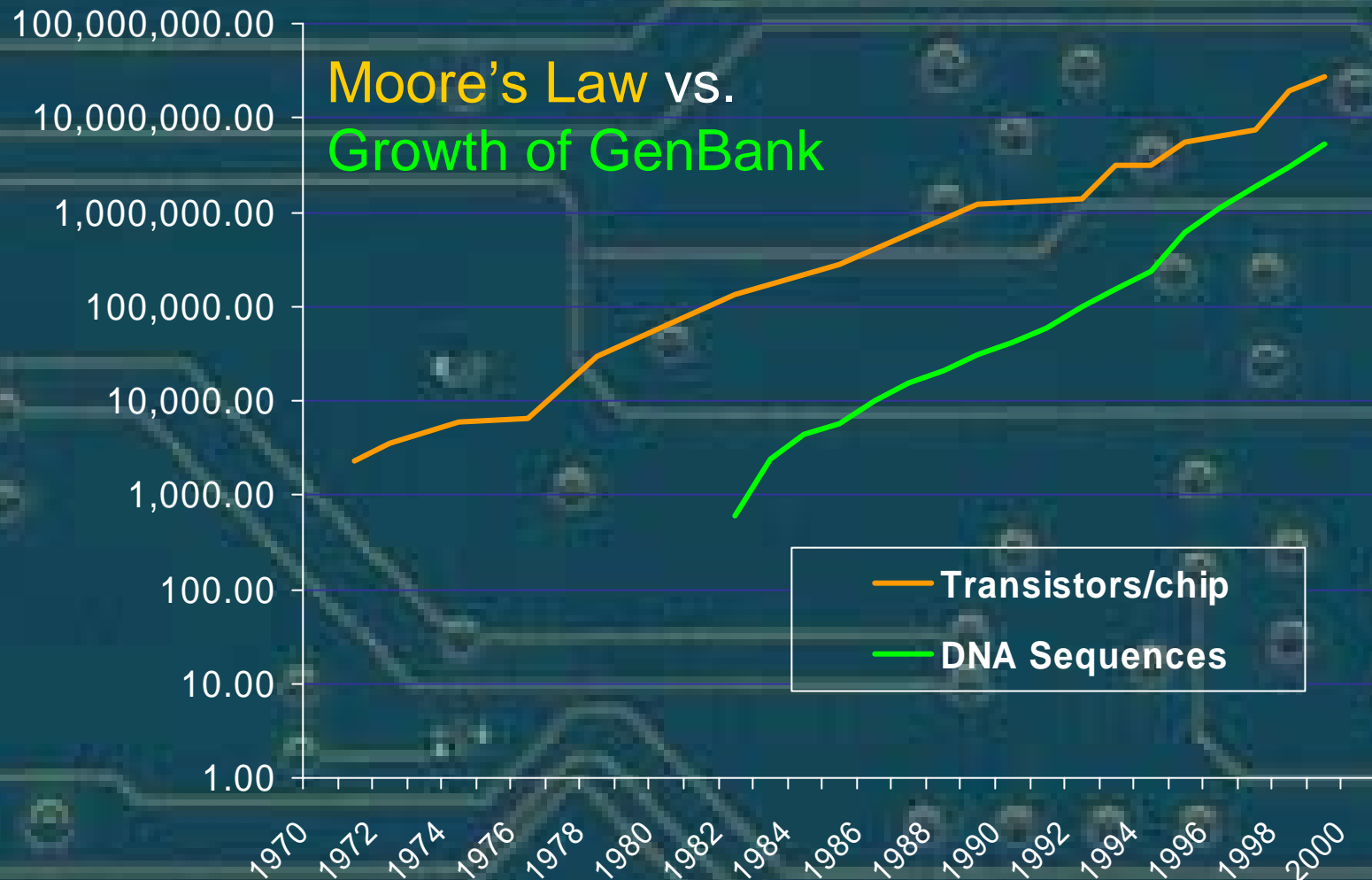
ROSETTA
INPHARMATICS

www.rii.com

“Anticipated advances in computer speed will be unable to keep up with the growing [DNA] sequence databases and the demand for homology searches of the data.”

**Charles DeLisi, 1988
U.S. Department of Energy**

Luckily, DeLisi's dire prediction has not (yet) come true



New Problems: Modeling and Simulation

- Single pathways & genetic networks
- Whole organisms
- Whole organs

New Skills Required

- Probability and statistics
- Applied math & computer science
- Biomedical engineering



ROSETTA
INPHARMATICS

www.rii.com

Computational model of heart failure

Model based on aberrant behaviour of cardiac ion transporter genes

Computation requires days of time on a large, multi-processor computer



$$\frac{\partial v(\underline{x}, t)}{\partial t} = \frac{1}{C_m} \left[\underbrace{-I_{ion}(v(\underline{x}, t)) - I_{app}(\underline{x}, t)}_{\text{Total Membrane Current}} + \frac{1}{\beta} \left(\frac{\kappa}{\kappa + 1} \right) \nabla \cdot \left(\underbrace{M_i(\underline{x}) \nabla v(\underline{x}, t)}_{\text{Coupling Current}} \right) \right], \dots \forall \underline{x} \in H$$



ROSETTA
INPHARMATICS