

21st Century Biology: Informatics in the Post-Genome Era

Robert J. Robbins
Fred Hutchinson Cancer Research Center
1124 Columbia Street, LV-101
Seattle, Washington 98104

rrobbins@fhcrc.org
(206) 667 4778

Topics

- Biotechnology will be the “magic” technology of the 21st Century.

Topics

- Biotechnology will be the “magic” technology of the 21st Century.
- Information Technology (IT) has a special relationship with biology.

Topics

- Biotechnology will be the “magic” technology of the 21st Century.
- Information Technology (IT) has a special relationship with biology.
- Moore’s Law constantly transforms IT (and everything else).

Topics

- Biotechnology will be the “magic” technology of the 21st Century.
- Information Technology (IT) has a special relationship with biology.
- Moore’s Law constantly transforms IT (and everything else).
- Current funding mechanisms for bio-information infrastructure are hopelessly inadequate to meet future needs and must be radically reformed.

Introduction

Magical Technology

Magic

To a person from 1897, much current technology would seem like magic.

Magic

To a person from 1897, much current technology would seem like magic.

What technology of 2097 would seem magical to a person from 1997?

Magic

To a person from 1897, much current technology would seem like magic.

What technology of 2097 would seem magical to a person from 1997?

Candidate: Biotechnology so advanced that the distinction between living and non-living is blurred.

IT-Biology Synergism

IT is Special

Information Technology:

- *affects the performance **and** the management of tasks*

IT is Special

Information Technology:

- *affects the performance and the management of tasks*
- *allows the manipulation of huge amounts of highly complex data*

IT is Special

Information Technology:

- *affects the performance and the management of tasks*
- *allows the manipulation of huge amounts of highly complex data*
- *is incredibly plastic*
(programming and poetry are both exercises in pure thought)

IT is Special

Information Technology:

- *affects the performance and the management of tasks*
- *allows the manipulation of huge amounts of highly complex data*
- *is incredibly plastic*
(programming and poetry are both exercises in pure thought)
- *improves exponentially* *(Moore's Law)*

Biology is Special

Life is Characterized by:

- *individuality*

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*
- *contingency*

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*
- *contingency*
- *high (digital) information content*

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*
- *contingency*
- *high (digital) information content*

No law of large numbers...

Biology is Special

Life is Characterized by:

- *individuality*
- *historicity*
- *contingency*
- *high (digital) information content*

No law of large numbers, since every living thing is genuinely unique.

IT-Biology Synergism

- *Physics needs calculus, the method for manipulating information about statistically large numbers of vanishingly small, independent, equivalent things.*

IT-Biology Synergism

- *Physics needs calculus, the method for manipulating information about statistically large numbers of vanishingly small, independent, equivalent things.*
- *Biology needs information technology, the method for manipulating information about large numbers of dependent, historically contingent, individual things.*

Biology is Special

For it is in relation to the statistical point of view that the structure of the vital parts of living organisms differs so entirely from that of any piece of matter that we physicists and chemists have ever handled in our laboratories or mentally at our writing desks.

Erwin Schrödinger. 1944. *What is Life*.

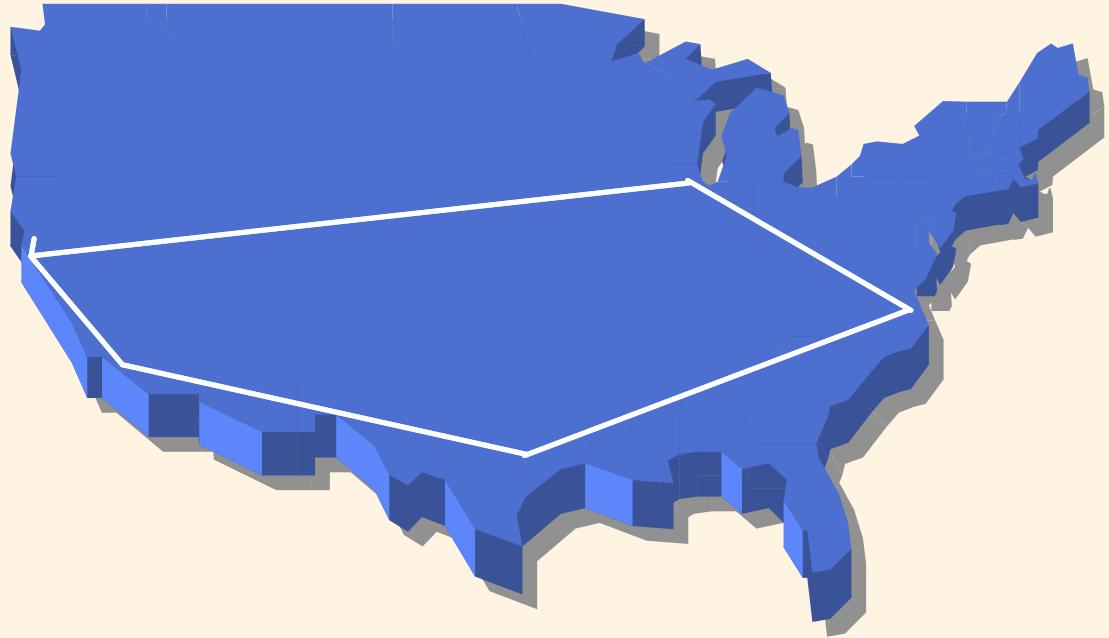
Genetics as Code

[The] chromosomes ... contain in some kind of code-script the entire pattern of the individual's future development and of its functioning in the mature state. ... [By] code-script we mean that the all-penetrating mind, once conceived by Laplace, to which every causal connection lay immediately open, could tell from their structure whether [an egg carrying them] would develop, under suitable conditions, into a black cock or into a speckled hen, into a fly or a maize plant, a rhodo-dendron, a beetle, a mouse, or a woman.

Erwin Schrödinger. 1944. *What is Life*.

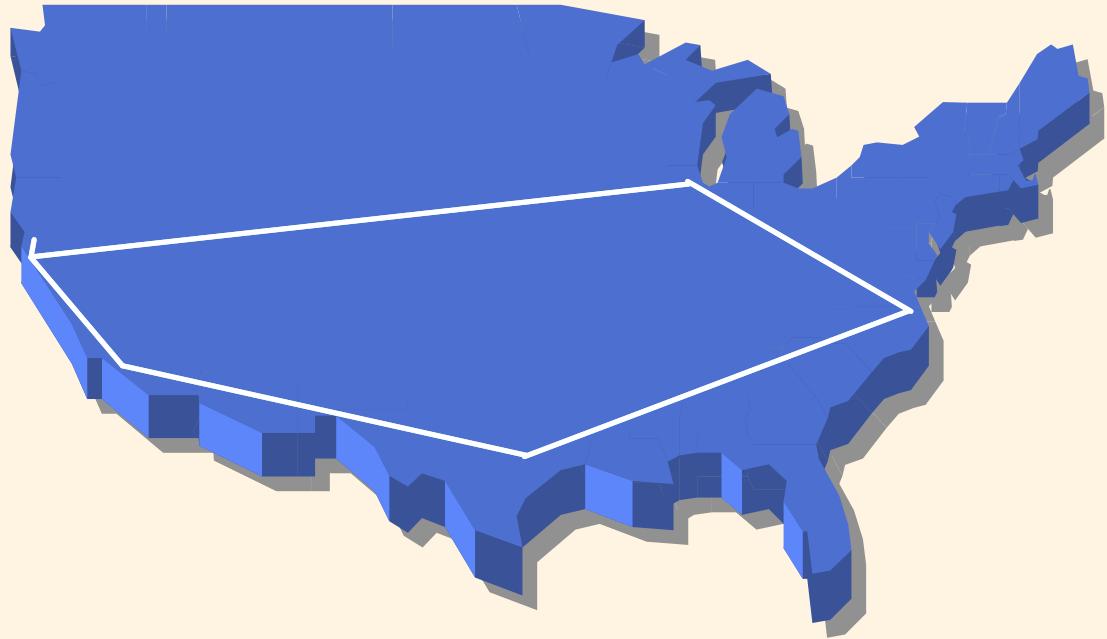
One Human Sequence

We now know that Schrödinger's mysterious human "code-script" consists of 3.3 billion base pairs of DNA.



One Human Sequence

We now know that Schrödinger's mysterious human "code-script" consists of 3.3 billion base pairs of DNA.



Typed in 10-pitch font, one human sequence would stretch for more than 5,000 miles. Digitally formatted, it could be stored on one CD-ROM. Biologically encoded, it fits easily within a single cell.

Bio-digital Information

DNA is a highly efficient digital storage device:

- There is more mass-storage capacity in the DNA of a side of beef than in all the hard drives of all the world's computers.

Bio-digital Information

DNA is a highly efficient digital storage device:

- There is more mass-storage capacity in the DNA of a side of beef than in all the hard drives of all the world's computers.
- Storing all of the (redundant) information in all of the world's DNA on computer hard disks would require that the entire surface of the Earth be covered to a depth of three miles in Conner 1.0 gB drives.

Genomics: An Example

Computers as Instruments

Computers are not just tools for cataloging existing knowledge. They are instruments that change the way we can see the biological world. Computers allow us to see genomes, just as radio telescopes let us see quasars and microscopes let us see cells.

Human Genome Project - Goals

- construction of a high-resolution genetic map of the human genome;

USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Human Genome Project - Goals

- construction of a high-resolution genetic map of the human genome;
- production of a variety of physical maps of all human chromosomes and of the DNA of selected model organisms;

USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Human Genome Project - Goals

- construction of a high-resolution genetic map of the human genome;
- production of a variety of physical maps of all human chromosomes and of the DNA of selected model organisms;
- determination of the complete sequence of human DNA and of the DNA of selected model organisms;

USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Human Genome Project - Goals

- construction of a high-resolution genetic map of the human genome;
- production of a variety of physical maps of all human chromosomes and of the DNA of selected model organisms;
- determination of the complete sequence of human DNA and of the DNA of selected model organisms;
- development of capabilities for collecting, storing, distributing, and analyzing the data produced;

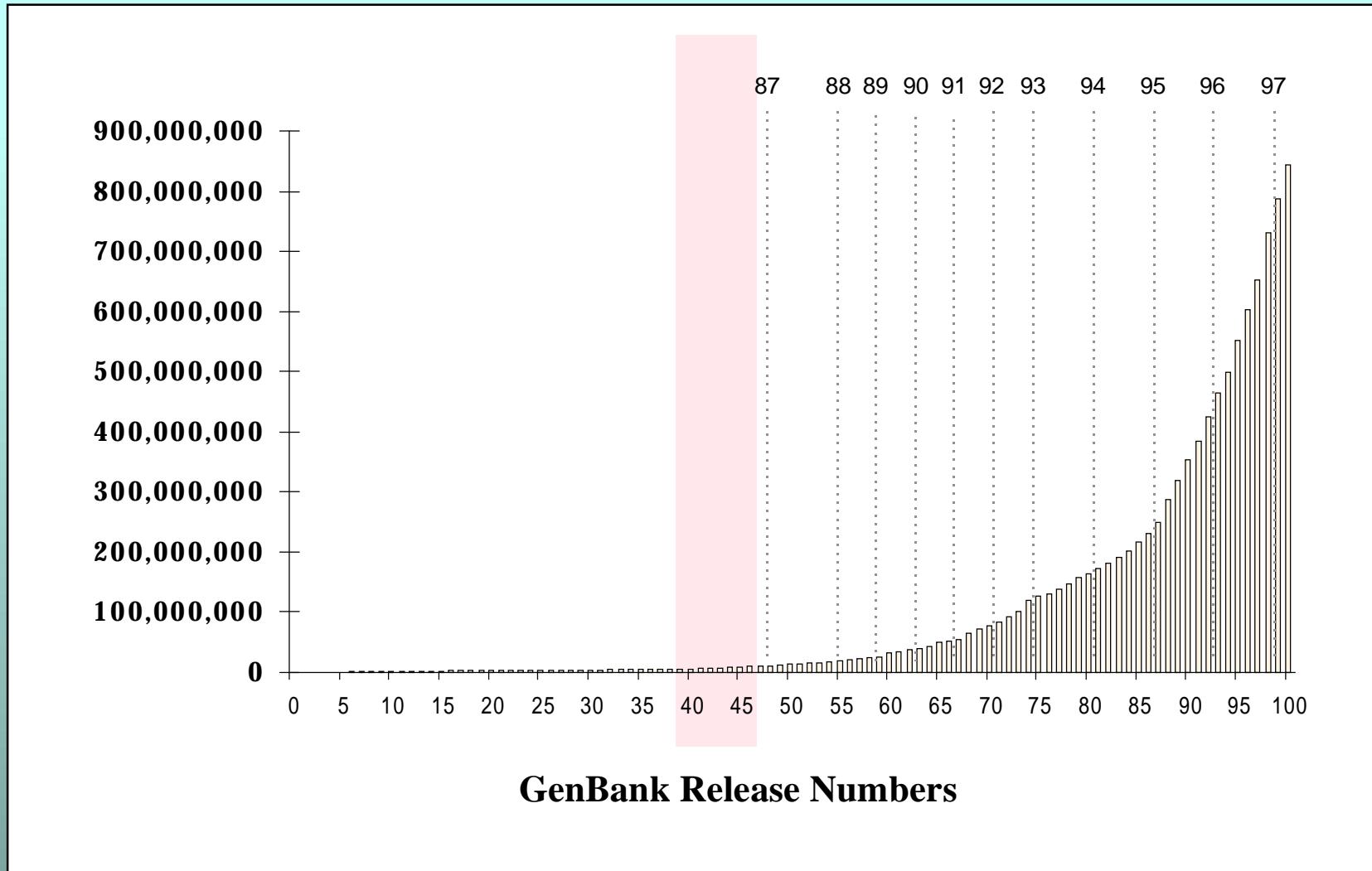
USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Human Genome Project - Goals

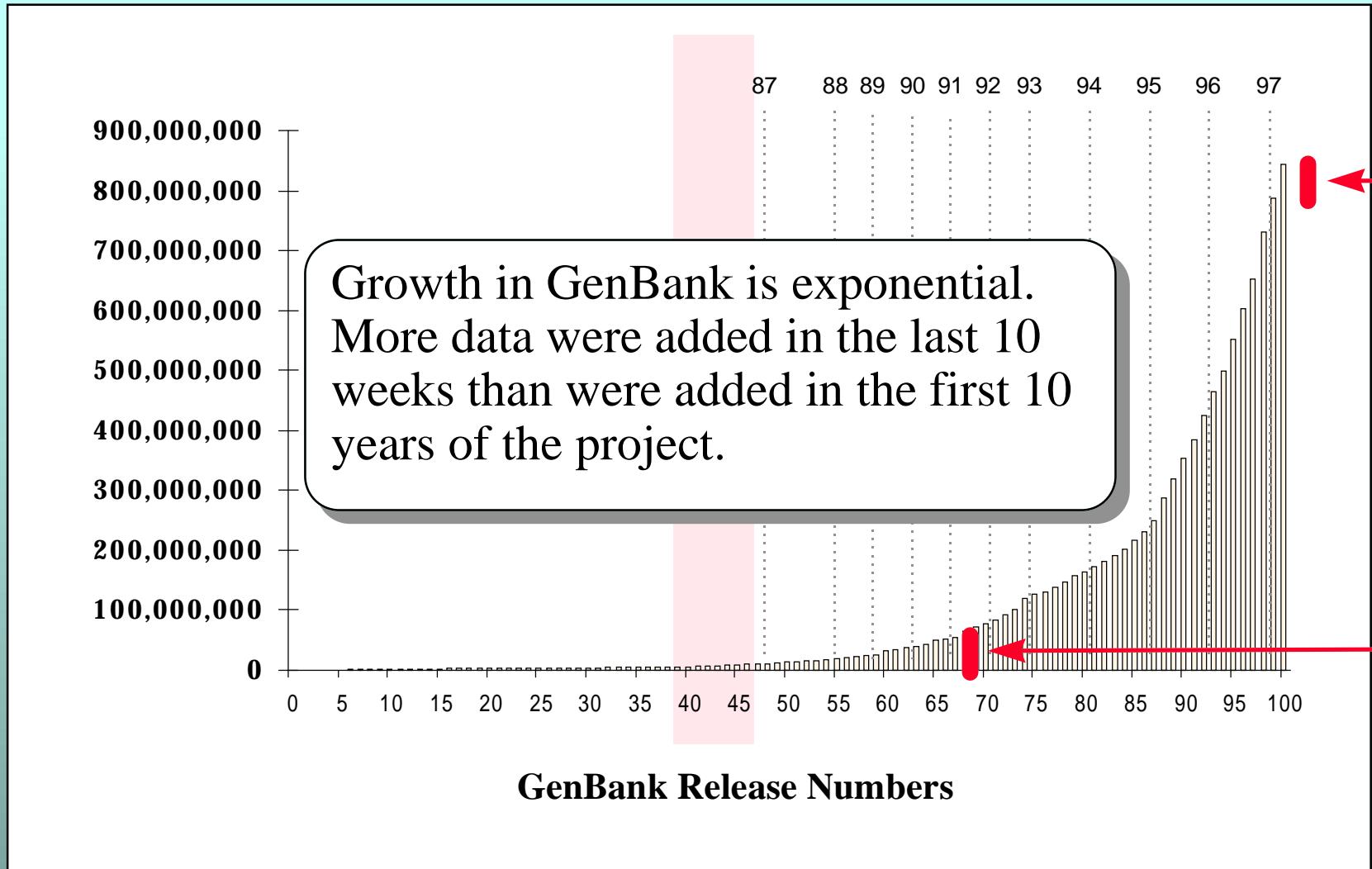
- construction of a high-resolution genetic map of the human genome;
- production of a variety of physical maps of all human chromosomes and of the DNA of selected model organisms;
- determination of the complete sequence of human DNA and of the DNA of selected model organisms;
- development of capabilities for collecting, storing, distributing, and analyzing the data produced;
- creation of appropriate technologies necessary to achieve these objectives.

USDOE. 1990. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years.*

Base Pairs in GenBank



Base Pairs in GenBank



Infrastructure and the HGP

Progress towards all of the [Genome Project] goals will require the establishment of well-funded centralized facilities, including a stock center for the cloned DNA fragments generated in the mapping and sequencing effort and a data center for the computer-based collection and distribution of large amounts of DNA sequence information.

National Research Council. 1988. *Mapping and Sequencing the Human Genome*. Washington, DC: National Academy Press. p. 3

Databases and the Genome Project

[The] database developer should provide, in some real sense, an intellectual focus for the interpretation of genomic data.

NIH-DOE Ad Hoc Committee on Genome Databases

21st Century Biology

The Approach

Paradigm Shift in Biology

The new paradigm, now emerging, is that all the 'genes' will be known (in the sense of being resident in databases available electronically), and that the starting point of a biological investigation will be theoretical. An individual scientist will begin with a theoretical conjecture, only then turning to experiment to follow or test that hypothesis.

Walter Gilbert. 1991. Towards a paradigm shift in biology. *Nature*, 349:99.

Paradigm Shift in Biology

To use [the] flood of knowledge, which will pour across the computer networks of the world, biologists not only must become computer literate, but also change their approach to the problem of understanding life.

Walter Gilbert. 1991. Towards a paradigm shift in biology. *Nature*, 349:99.

21st Century Biology

The People

Human Resources Issues

- Reduction in need for non-IT staff

Human Resources Issues

- Reduction in need for non-IT staff
- Increase in need for IT staff, especially “information engineers”

Human Resources Issues

- Reduction in need for non-IT staff
- Increase in need for IT staff, especially “information engineers”

In modern biology, a general trend is to convert expert work into staff work and finally into computation. New expertise is required to design, carry out, and interpret continuing work.

Human Resources Issues

Elbert Branscomb: “You must recognize that some day you may need as many computer scientists as biologists in your labs.”

Human Resources Issues

Elbert Branscomb: “You must recognize that some day you may need as many computer scientists as biologists in your labs.”

Craig Venter: “At TIGR, we already have twice as many computer scientists on our staff.”

Exchange at DOE workshop on high-throughput sequencing.

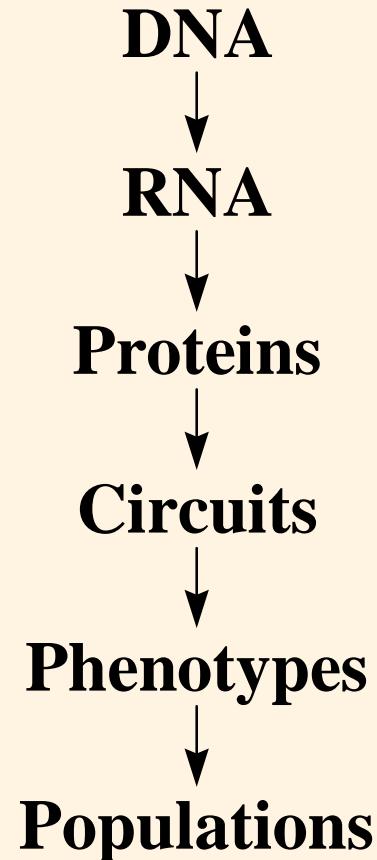
21st Century Biology

The Science

Fundamental Dogma

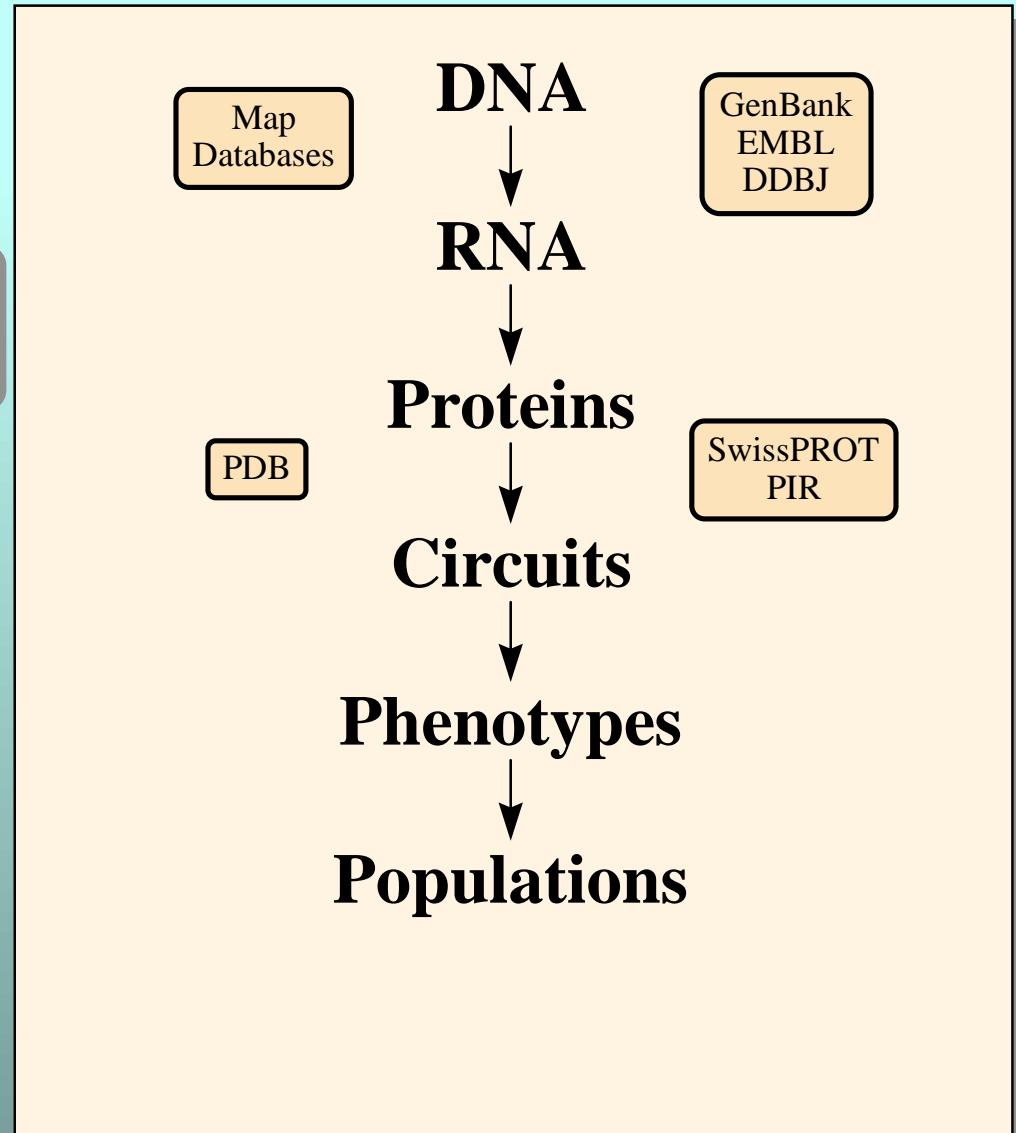
The fundamental dogma of molecular biology is that genes act to create phenotypes through a flow of information from DNA to RNA to proteins, to interactions among proteins, and ultimately to phenotypes.

Collections of individual phenotypes, of course, constitute a population.



Fundamental Dogma

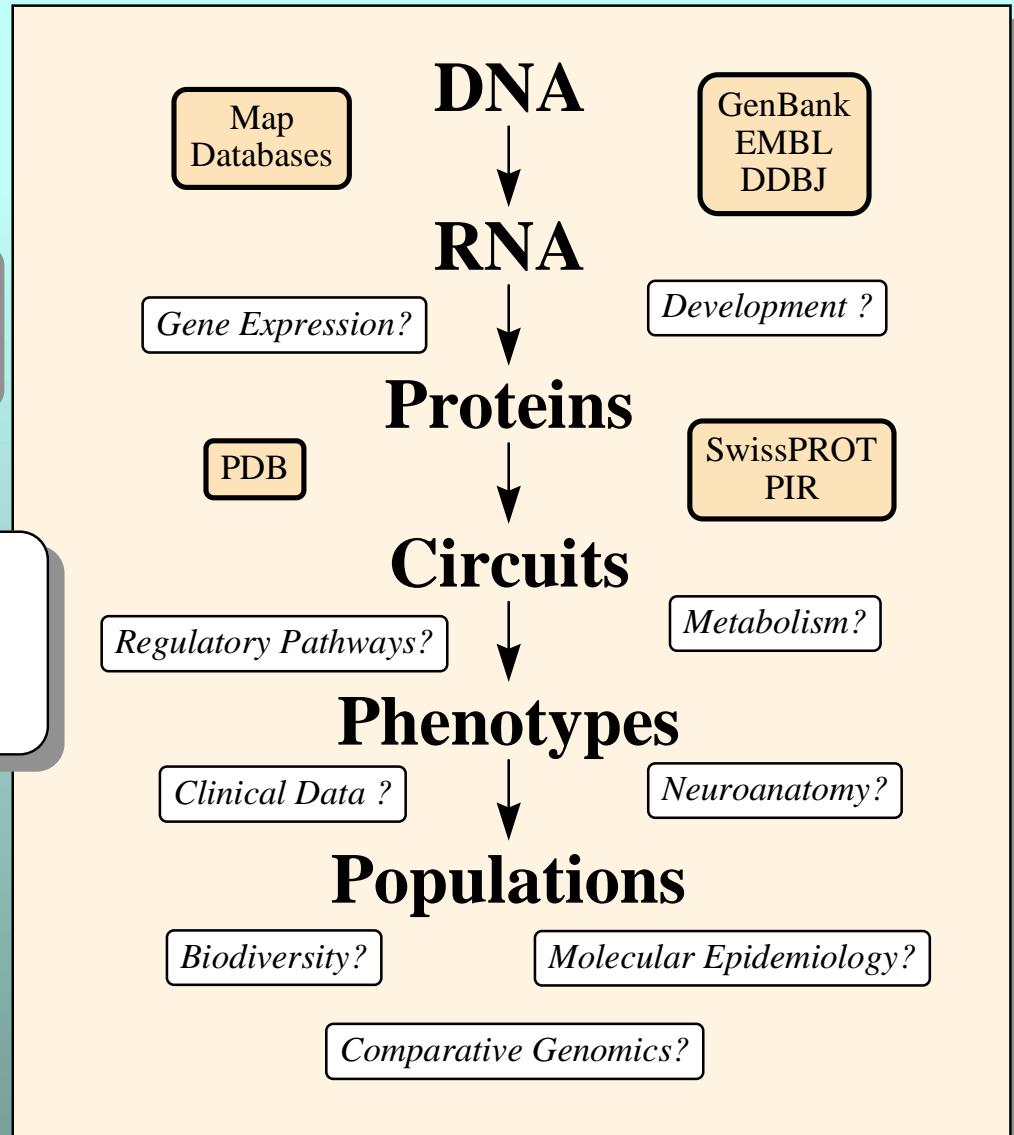
Although a few databases already exist to distribute molecular information,



Fundamental Dogma

Although a few databases already exist to distribute molecular information,

the post-genomic era will need many more to collect, manage, and publish the coming flood of new findings.



21st Century Biology

The Literature

Electronic Data Publishing

G D B -- Beta Hemoglobin

**** Locus Detail View ****

Symbol: HBB
Name: hemoglobin, beta
MIM Num: 141900
Location: 11p15.5
Created: 01 Jan 86 00:00

**** Polymorphism Table ****

Probe	Enzyme
beta-globin cDNA	RsaI
beta-globin cDNA,JW10+	Avall
Pstbeta,JW102,BD23,pB+	BamHI
pRK29,Unknown	HindII
beta-IVS2 probe	HphI
IVS-2 normal	HphI
Unknown	AvrII
beta-IVS2 probe	AsuI

Electronic Data Publishing

O M I M -- Beta Hemoglobin

G D B -- Beta Hemoglobin

** Locus Detail View **

Symbol: HBB
Name: hemoglobin, beta
MIM Num: 141900
Location: 11p15.5
Created: 01 Jan 86 00:00

** Polymorphism Table **

Probe	Enzyme
beta-globin cDNA	Rsal
beta-globin cDNA,JW10+	Avall
Pstbeta,JW102,BD23,pB+	BamHI
pRK29,Unknown	HindII
beta-IVS2 probe	HphI
IVS-2 normal	HphI
Unknown	AvrII
beta-IVS2 probe	AsuI

BIN--BETA LOCUS
L ANEMIA, INCLUDED;
IAS, INCLUDED;
IAS, BETA-GLOBIN

loci determine the
types of polypeptide
oglobin, Hb A. By
ing heavy-labeled
c messenger RNA,
ound labeling of a
a group B chromo-
ded, incorrectly as it
beta-gamma-delta
on a group B
the zone of labeling
chromosome than on
ich by this reasoning

Electronic Data Publishing

GenBank -- Beta Hemoglobin

O M I M -- Beta Hemoglobin

G D B -- Beta Hemoglobin

** Locus Detail View **

Symbol: HBB
 Name: hemoglobin, beta
 MIM Num: 141900
 Location: 11p15.5
 Created: 01 Jan 86 00:00

** Polymorphism Table **

Probe	Enzyme
beta-globin cDNA	Rsal
beta-globin cDNA,JW10+	Avall
Pstbeta,JW102,BD23,pB+	BamHI
pRK29,Unknown	HindII
beta-IVS2 probe	HphI
IVS-2 normal	HphI
Unknown	AvrII
beta-IVS2 probe	AsuI

BIN--BETA LOCUS
 L ANEMIA, INCLUDED;
 IAS, INCLUDED;
 IAS, BETA-GLOBIN

loci determine the
 types of polypeptide
 globin, Hb A. By
 using heavy-labeled
 c messenger RNA,
 found labeling of a
 a group B chromo-
 ded, incorrectly as it
 beta-gamma-delta
 on a group B
 the zone of labeling
 chromosome than on
 ich by this reasoning

beta
 gion

00096
 00161

PFH;
 e; RNA

site;

actactgtctagt
 ctcatgtcttgag
 aaaaaattagcca
 tgagcgcaactcca

Electronic Data Publishing

PIR -- Beta Hemoglobin

GenBank -- Beta Hemoglobin

O M I M -- Beta Hemoglobin

G D B -- Beta Hemoglobin

** Locus Detail View **

Symbol: HBB
 Name: hemoglobin, beta
 MIM Num: 141900
 Location: 11p15.5
 Created: 01 Jan 86 00:00

** Polymorphism Table **

Probe	Enzyme
beta-globin cDNA	Rsal
beta-globin cDNA,JW10+	Avall
Pstbeta,JW102,BD23,pB+	BamHI
pRK29,Unknown	HindII
beta-IVS2 probe	HphI
IVS-2 normal	HphI
Unknown	AvrII
beta-IVS2 probe	AsuI

BIN--BETA LOCUS
 L ANEMIA, INCLUDED;
 IAS, INCLUDED;
 IAS, BETA-GLOBIN

loci determine the
 types of polypeptide
 globin, Hb A. By
 using heavy-labeled
 c messenger RNA,
 found labeling of a
 a group B chromo-
 somed, incorrectly as it
 beta-gamma-delta
 on a group B
 the zone of labeling
 chromosome than on
 which by this reasoning

beta
 globin

00096
 00161

PFH;
 e; RNA

site;

tactactgtctagt
 ctcatgtcttgag
 aaaaaattagcca
 tgagcgcaactcca

in beta chain
 ee, pygmy
 gorilla

t 15867

S A V T A L W G
 G E A L G R L L
 F F E S F G D L
 N P K V K A H G
 D G L A H L D N
 S E L H C D K L
 L L G N V L V C

Electronic Scholarly Publishing

Back Forward Stop Refresh Home Search Favorit... Print Font Mail

Address <http://www.esp.org/> Links

ESP
ELECTRONIC
SCHOLARLY
PUBLISHING

about
help
CREDITS

ESP

ELECTRONIC SCHOLARLY PUBLISHING

Publication Collections

[Classical Genetics](#) reprints of key papers from the early history of genetics (TEXT ONLY version).

[Classical Genetics](#) reprints of key papers from the early history of genetics (HIGH GRAPHICS version). NOTE: the high graphics version is under

although all
trieved from the
ponents are
al.

at material from a
S.

rints of
e presentations,
instructional materials, and data

Shortcut to browse1.htm

HTTP://WWW.ESP.ORG

The ESP site is dedicated to the electronic publishing of scientific and other scholarly materials. Of particular interest are the history of science, genetics, computational biology, and genome research.

Electronic Scholarly Publishing

The *Classical Genetics: Foundations* series provides ready access to typeset-quality, electronic editions of important publications that can otherwise be very difficult to find.

ESP
ELECTRONIC
SCHOLARLY
PUBLISHING

about
help
search
CREDITS

Classical Genetics: Foundations

Early Mendelism

(395,699 bytes; 41 pages, no figures)

Mendel, Gregor. 1865. Experiments in plant hybridization. *Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865, Abhandlungen*, 3-47.

In February and March of 1865, the Brünn Natural History Society in Brünn, Czechoslovakia, heard Gregor Mendel present the results of his investigations into the mechanisms governing inheritance in pea plants. The next year, the work was published as Mendel, Gregor. 1866. "Versuche über Pflanzen Hybriden." *Verhandlungen des naturforschenden Vereines in Brünn*, 4:3-47.

In this remarkable work, Mendel established the foundation for what later became the science of genetics. However, his work was largely ignored when it appeared and Mendel moved on to other things. He died in 1884, never knowing that his experiments in plant hybridization were of any consequence.

His work was rediscovered at the turn of the century and its significance immediately recognized. Genetics, as a formal scientific discipline, exploded into activity in 1900.

Gregor Mendel, 1862

Copyright
1996, 1997
ESP

Electronic Scholarly Publishing

“Hardy” (of Hardy-Weinberg) is a name well known to most students of biology.

Back Forward Stop Refresh Home Search Favorit... Print Font Mail

Address <http://www.esp.org/graphics/browse1.htm> Links

ESP
ELECTRONIC
SCHOLARLY
PUBLISHING

Classical Genetics: Foundations *Early Mendelism*

(19,473 bytes; 1 page, no figures)  [▶ Hardy, G. H. 1908. Mendelian Proportions in a Mixed Population. *Science, NS. XXVIII: 49-50*](#)

Every geneticist has heard of the Hardy-Weinberg Law and of Hardy-Weinberg Equilibrium, and nearly all basic biology texts teach that G. H. Hardy played a seminal role in founding population genetics. But, what most biologists don't realize is that Hardy's **total** contribution to biology consisted of a **single** letter to the editor in *Science*. The letter began,

I am reluctant to intrude in a discussion concerning matters of which I have no expert knowledge, and I should have expected the very simple point which I wish to make to have been familiar to biologists. However, some remarks of Mr. Udny Yule, to which Mr. R. C. Punnett has called my attention, suggest that it may still be worth making

With that, Hardy offered his "simple point" and then washed his hands of biology. His autobiography, *A Mathematician's Apology*, makes no mention of population genetics.

Copyright
1996, 1997
E S P

Electronic Scholarly Publishing

But how many have read, or even seen, **all** of Hardy's biological writings?

This is it: A single, one-page letter to the editor of *Science*.

Back Forward Stop Refresh Home Search Favorit... Print Font Mail

Address <http://www.esp.org/history/hardy.pdf> Links

JULY 10, 1908 SCIENCE N. S. Vol. XXVIII:49-50

DISCUSSION AND CORRESPONDENCE

Mendelian Proportions in a Mixed Population

To The Editor of Science: I am reluctant to intrude in a discussion concerning matters of which I have no expert knowledge, and I should have expected the very simple point which I wish to make to have been familiar to biologists. However, some remarks of Mr. Udry Yule, to which Mr. R. C. Punnett has called my attention, suggest that it may still be worth making.

In the *Proceedings of the Royal Society of Medicine* (Vol I., p. 165) Mr. Yule is reported to have suggested, as a criticism of the Mendelian position, that if brachydactyly is dominant "in the course of time one would expect, in the absence of counteracting factors, to get three brachydactylous persons to one normal."

It is not difficult to prove, however, that such an expectation would be quite groundless. Suppose that Aa is a pair of Mendelian characters, A being dominant, and that in any given generation the numbers of pure dominants (AA), heterozygotes (Aa), and pure recessives (aa) are as $p:2q:r$. Finally, suppose that the numbers are fairly large, so that the mating may be regarded as random, that the sexes are evenly distributed among the three varieties, and that all are equally fertile. A little mathematics of the multiplication-table type is enough to show that in the next generation the numbers will be as

this proportion would afterwards have no tendency to decrease.

In a word, there is not the slightest foundation for the idea that a dominant character should show a tendency to spread over a whole population, or that a recessive should tend to die out.

I ought perhaps to add a few words on the effect of the small deviations from the theoretical proportions which will, of course, occur in every generation. Such a distribution as $p_1:2q_1:r_1$, which satisfies the condition $q = p_1r_1$, we may call a *stable* distribution. In actual fact we shall obtain in the second generation not $p_1:2q_1:r_1$ but a slightly different distribution $p_2:2q_2:r_2$, which is not "stable." This should, according to theory, give us in the third generation a "stable" distribution $p_3:2q_3:r_3$, also differing from $p_1:2q_1:r_1$; and so on. The sense in which the distribution $p_1:2q_1:r_1$ is "stable" is this, that if we allow for the effects of casual deviations in any subsequent generation, we should, according to theory, obtain at the next generation a new "stable" distribution differing but slightly from the original distribution.

I have, of course, considered only the very simplest hypotheses possible. Hypotheses other than [*sic*] that of purely random mating will give different results, and, of course, if, as appears to be the case sometimes, the character is not independent of that of sex, or has an influence on fertility, the whole question may be greatly complicated. But such complications seem to be irrelevant to the simple issue

Page 1 of 1 | 128% | 8.5 x 11 in

Electronic Scholarly Publishing

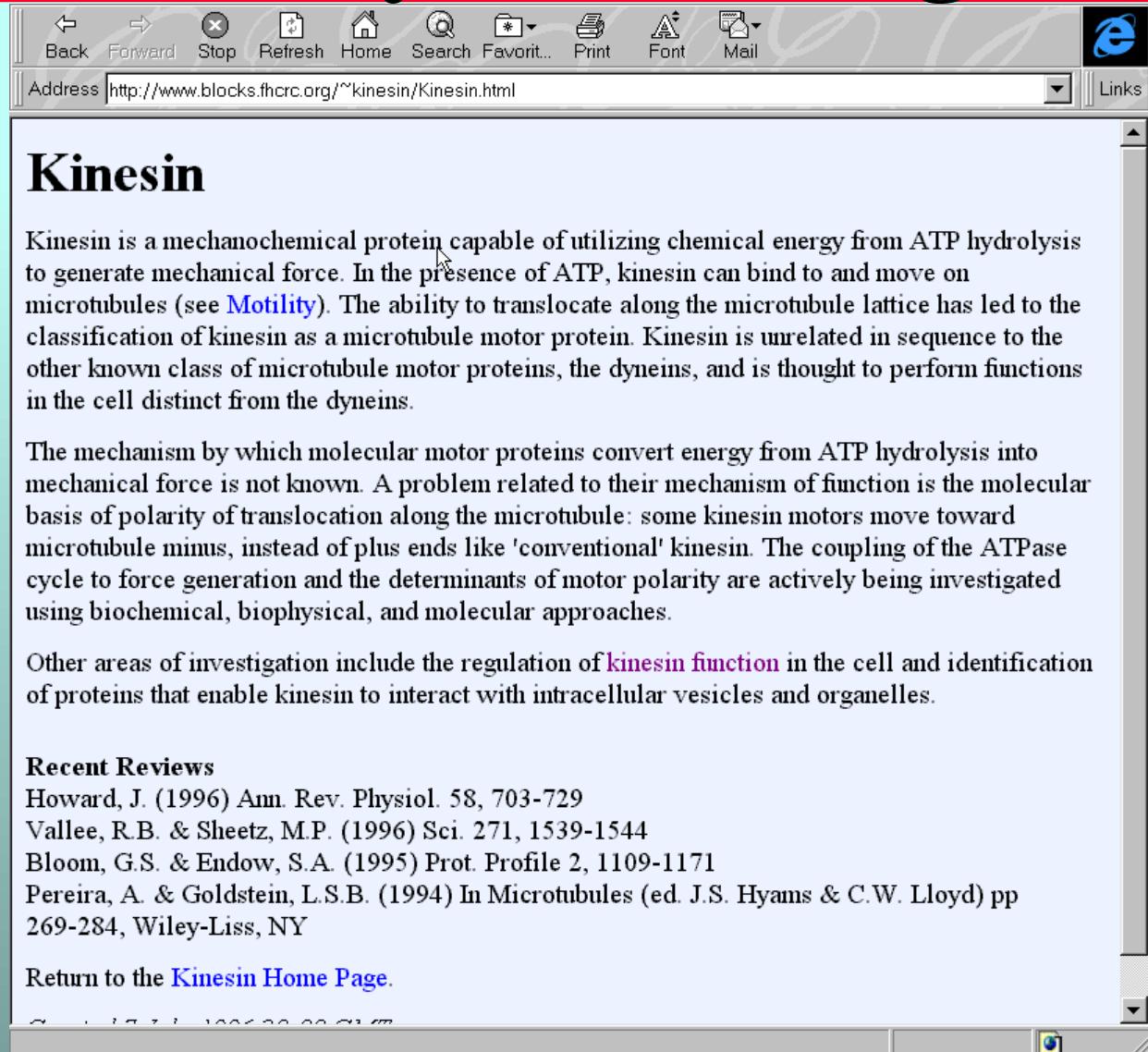
The screenshot shows a web browser window with the address bar containing <http://www.blocks.fhcrc.org/~kinesin/>. The page title is "The Kinesin Home Page". The navigation menu includes links for [Kinesin](#), [Kinesin Related Proteins](#), [Kinesin Tree](#), [Kinesin Sequences](#), and [Kinesin People](#). The main content area features several panels: "Vesicle Transport" with a fluorescence micrograph of a neuron, "Crystal Structure" with a 3D ribbon diagram of a kinesin motor, "Kinesin Motor Proteins" with a phylogenetic tree, and "Microtubule Motility" with a grayscale micrograph showing microtubule tracks. A "Spindle Movement" panel is partially visible at the bottom. The browser's status bar at the bottom shows "Done".

<http://www.blocks.fhcrc.org/~kinesin>

Electronic publishing is especially appropriate for some kinds of dynamic review papers.

Electronic Scholarly Publishing

Reviews may be revised
and maintained in real
time...



The screenshot shows a web browser window with the address bar containing <http://www.blocks.fhcrc.org/~kinesin/Kinesin.html>. The page content is as follows:

Kinesin

Kinesin is a mechanochemical protein capable of utilizing chemical energy from ATP hydrolysis to generate mechanical force. In the presence of ATP, kinesin can bind to and move on microtubules (see [Motility](#)). The ability to translocate along the microtubule lattice has led to the classification of kinesin as a microtubule motor protein. Kinesin is unrelated in sequence to the other known class of microtubule motor proteins, the dyneins, and is thought to perform functions in the cell distinct from the dyneins.

The mechanism by which molecular motor proteins convert energy from ATP hydrolysis into mechanical force is not known. A problem related to their mechanism of function is the molecular basis of polarity of translocation along the microtubule: some kinesin motors move toward microtubule minus, instead of plus ends like 'conventional' kinesin. The coupling of the ATPase cycle to force generation and the determinants of motor polarity are actively being investigated using biochemical, biophysical, and molecular approaches.

Other areas of investigation include the regulation of [kinesin function](#) in the cell and identification of proteins that enable kinesin to interact with intracellular vesicles and organelles.

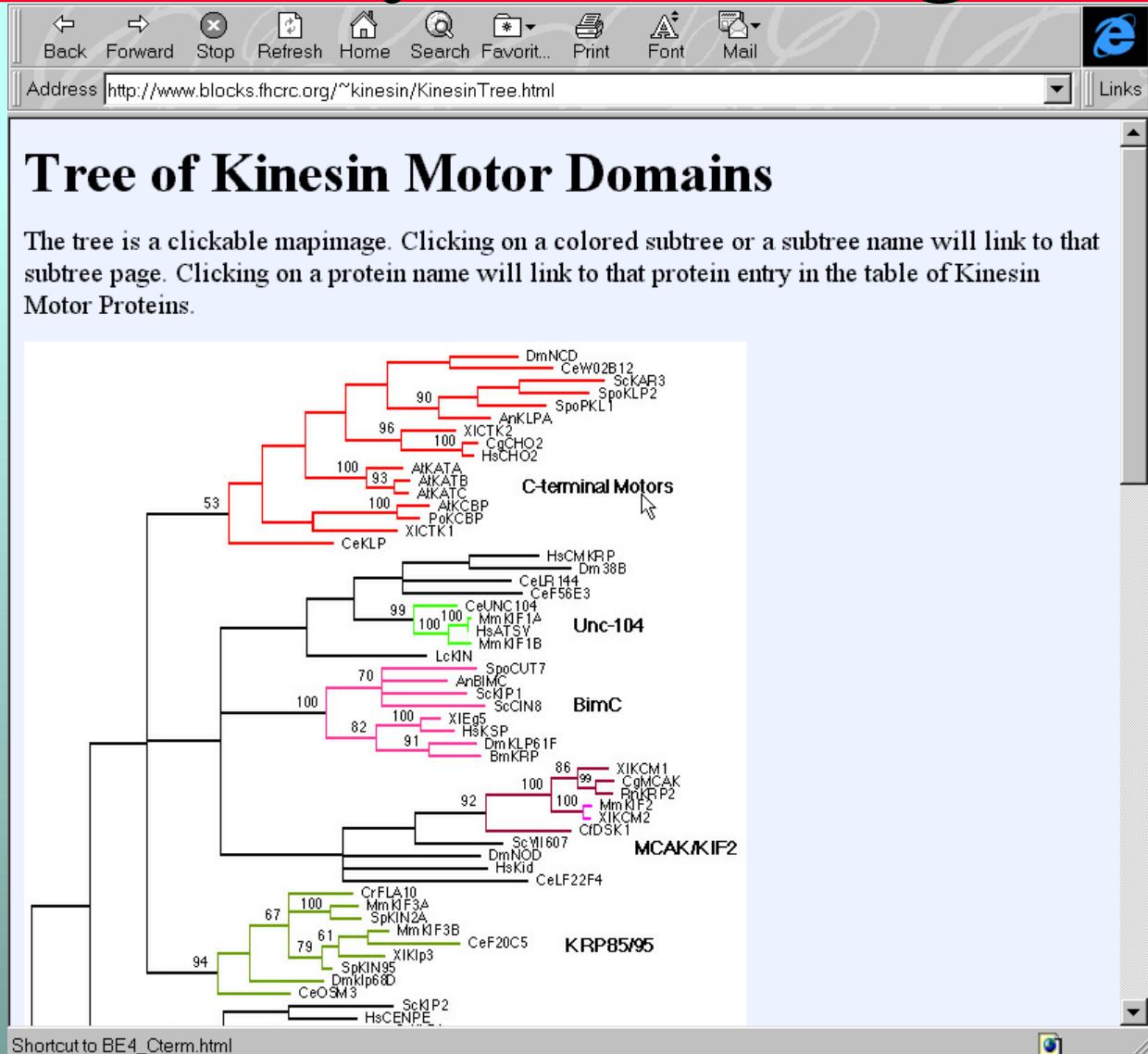
Recent Reviews

Howard, J. (1996) *Ann. Rev. Physiol.* 58, 703-729
Vallee, R.B. & Sheetz, M.P. (1996) *Sci.* 271, 1539-1544
Bloom, G.S. & Endow, S.A. (1995) *Prot. Profile* 2, 1109-1171
Pereira, A. & Goldstein, L.S.B. (1994) In *Microtubules* (ed. J.S. Hyams & C.W. Lloyd) pp 269-284, Wiley-Liss, NY

Return to the [Kinesin Home Page](#).

Electronic Scholarly Publishing

and it is easy to provide large amounts of in-depth supporting and related data.



Electronic Scholarly Publishing

Back Forward Stop Refresh Home Search Favorit... Print Font Mail

Address http://www.blocks.fhcrc.org/~kinesin/BE4_Cterm.html Links

The C-terminal Motor Subfamily

The C-terminal motor subfamily in the most recent analysis is no longer supported by a high bootstrap value. This is an indication of divergence within the group due to newly discovered members of the subfamily. The tree shown is taken from a tree built in a search of 73 kinesin motor domains. The kinesin proteins in this group have in common a C-terminal motor domain, and 4 members of the group (DmNcd, ScKAR3, CgCHO2, AtKCBP) have now been demonstrated to be minus-end directed motors.

D. melanogaster Ncd

NH₂ terminus COOH terminus

18nm

Basic/proline rich region +200 a.a.: binds MTs in DmNcd

α-helical coiled-coil ~150 a.a.

Globular motor domain ~350 a.a., binds MTs, ATPase activity ~ 6x7nm

Phylogenetic tree showing relationships between various kinesin motor domains with bootstrap values: DmNCD, CeW02B12, ScKAR3, SpoKLP2, SpoKLP1, AnKLP, XICTK2, CgCHO2, HsCHO2, AtKATA, AtKATB, AtKATC, AtKCBP, PoKCBP, XICTK1, CeKLP.

Species/protein	Molecular mass (kDa)	Motor polarity & velocity	Subcellular localization	Comments
<i>A. thaliana</i> KatA	89	ND	ND	
<i>A. thaliana</i> KatB	82	ND	ND	
<i>A. thaliana</i> KatC	84	ND	ND	
<i>A. thaliana</i> KCBP	143	Minus, 8-10	ND	Calmodulin binding

Done

Electronic Scholarly Publishing

The screenshot shows a web browser window with the address bar containing <http://www.esp.org/books/darwin/beagle/>. The browser's toolbar includes buttons for Back, Forward, Stop, Refresh, Home, Search, Favorites, Print, Font, and Mail. The website content is displayed in a frame with a black background on the left and a white background for the main text. The left sidebar features the ESP logo (ELECTRONIC SCHOLARLY PUBLISHING) and a vertical menu with buttons for 'about', 'help', 'search', and 'CREDITS'. The main content area displays the title 'Charles Darwin' and 'Voyage of the Beagle' at the top. Below this, it lists 'Electronic Scholarly Publishing' and 'HTML Edition'. The main title is 'THE VOYAGE OF THE BEAGLE 2nd Edition' by 'BY CHARLES DARWIN'. It also mentions 'From The Harvard Classics Volume 29', 'Copyright, 1909', and 'P. F. Collier & Son, New York'. At the bottom, there is a link for 'Table of Contents'.

<http://www.esp.org/books/darwin/beagle>

Entire monographs can be made instantly available to readers world-wide..

Electronic Scholarly Publishing

Today's computer technology was nearly unimaginable just ten years ago. The technology of ten years from now will also bring many surprises.

How is it that IT can maintain such an amazing rate of sustained change?

And what, if any, are the implications of that rate of change for biology?

The screenshot shows a web browser window with the address bar containing <http://www.esp.org/books/darwin/beagle/>. The browser's toolbar includes icons for Back, Forward, Stop, Refresh, Home, Search, Favorites, Print, Font, and Mail. The main content area displays the title "Charles Darwin" and "Voyage of the Beagle" at the top. Below this, the title "THE VOYAGE OF THE BEAGLE" is centered, followed by the word "Contents". A list of chapters is displayed, with the first chapter, "I. ST. JAGO -- CAPE DE VERD ISLANDS", highlighted in purple. The list includes chapters I through XIV. On the left side of the browser window, there is a vertical navigation menu with buttons for "about", "help", "search", and "CREDITS". The ESP logo is visible in the top left corner of the page content. At the bottom left of the page, there is a copyright notice: "Copyright 1996, 1997 E S P". The browser's status bar at the bottom shows "Done".

ESP
ELECTRONIC
SCHOLARLY
PUBLISHING

Charles Darwin *Voyage of the Beagle*

THE VOYAGE OF THE BEAGLE
Contents

Preface

- I. ST. JAGO -- CAPE DE VERD ISLANDS
- II. RIO DE JANEIRO
- III. MALDONADO
- IV. RIO NEGRO TO BAHIA BLANCA
- V. BAHIA BLANCA
- VI. BAHIA BLANCA TO BUENOS AYRES
- VII. BUENOS AYRES AND ST. FE
- VIII. BANDA ORIENTAL AND PATAGONIA
- IX. SANTA CRUZ, PATAGONIA, AND THE FALKLAND ISLANDS
- X. TIERRA DEL FUEGO
- XI. STRAIT OF MAGELLAN. -- CLIMATE OF THE SOUTHERN COASTS
- XII. CENTRAL CHILE
- XIII. CHILOE AND CHONOS ISLANDS
- XIV. CHILOE AND CONCEPCION: GREAT EARTHQUAKE

about
help
search
CREDITS

Copyright
1996, 1997
E S P

Done

Moore's Law

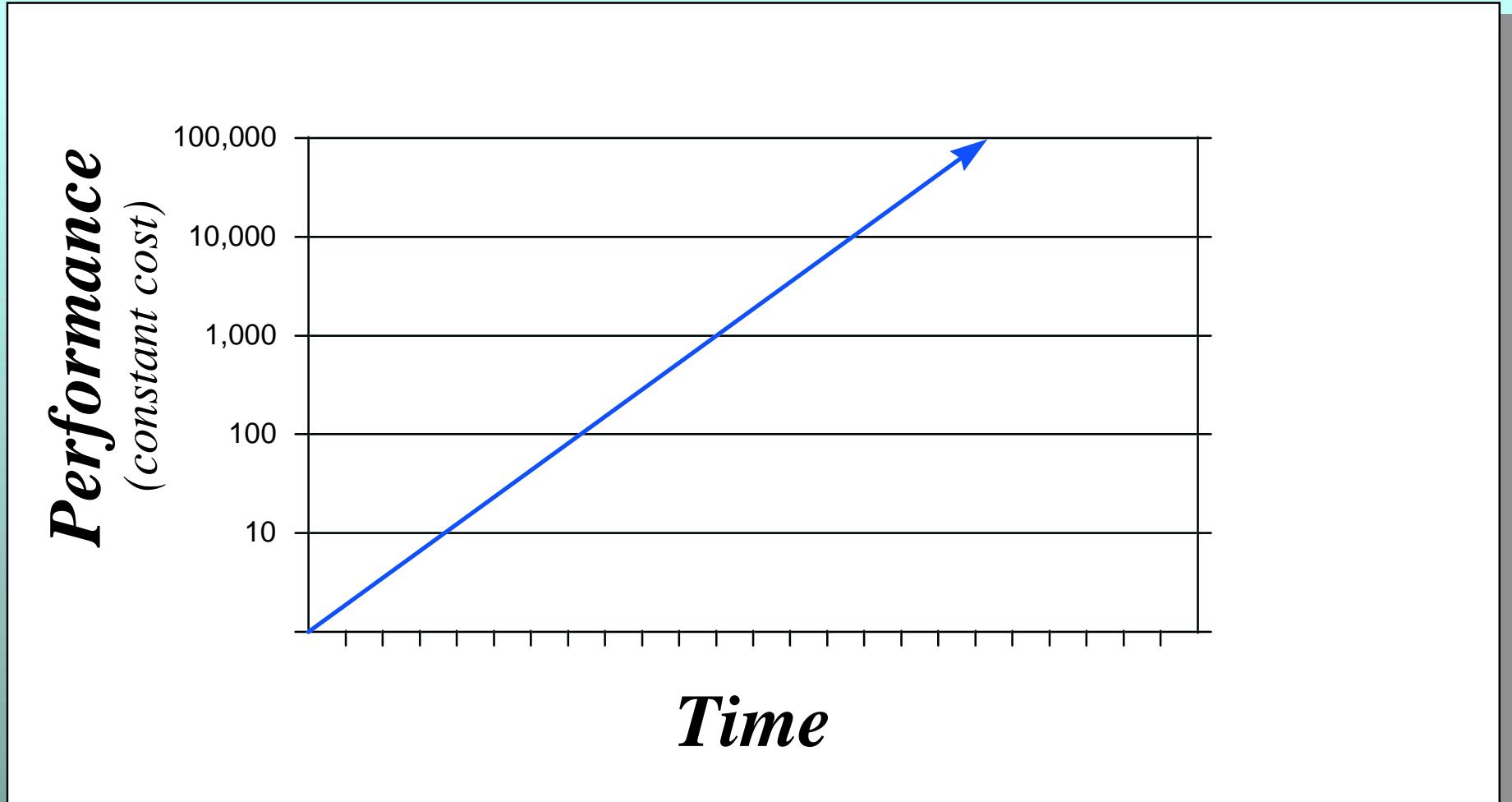
*Transforms InfoTech
(and everything else)*

Moore's Law: *The Statement*

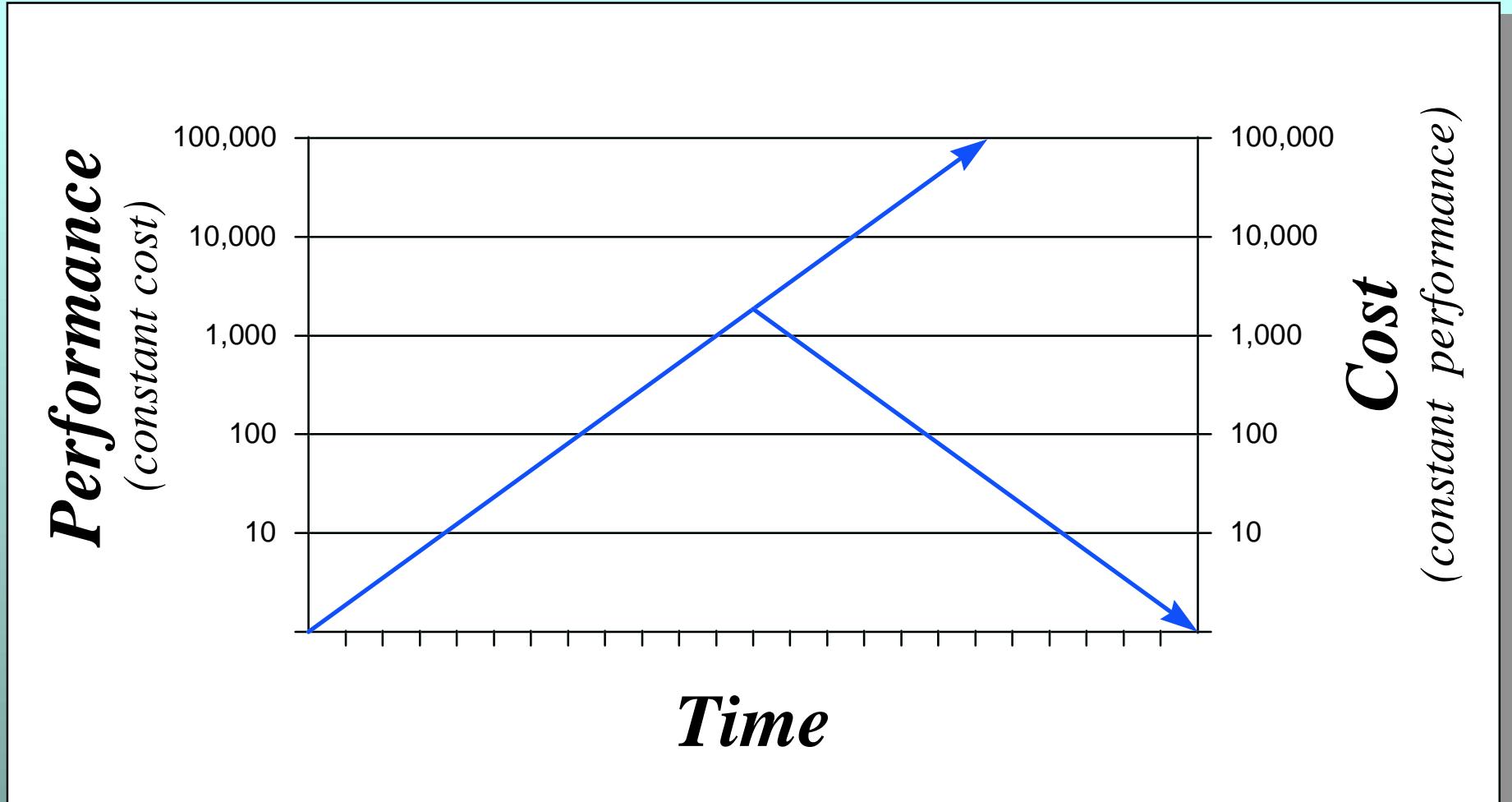
Every eighteen months, the number of transistors that can be placed on a chip doubles.

Gordon Moore, co-founder of Intel...

Moore's Law: *The Effect*



Moore's Law: *The Effect*



Moore's Law: *The Effect*

Three Phases of Novel IT Applications

- It's Impossible

Moore's Law: *The Effect*

Three Phases of Novel IT Applications

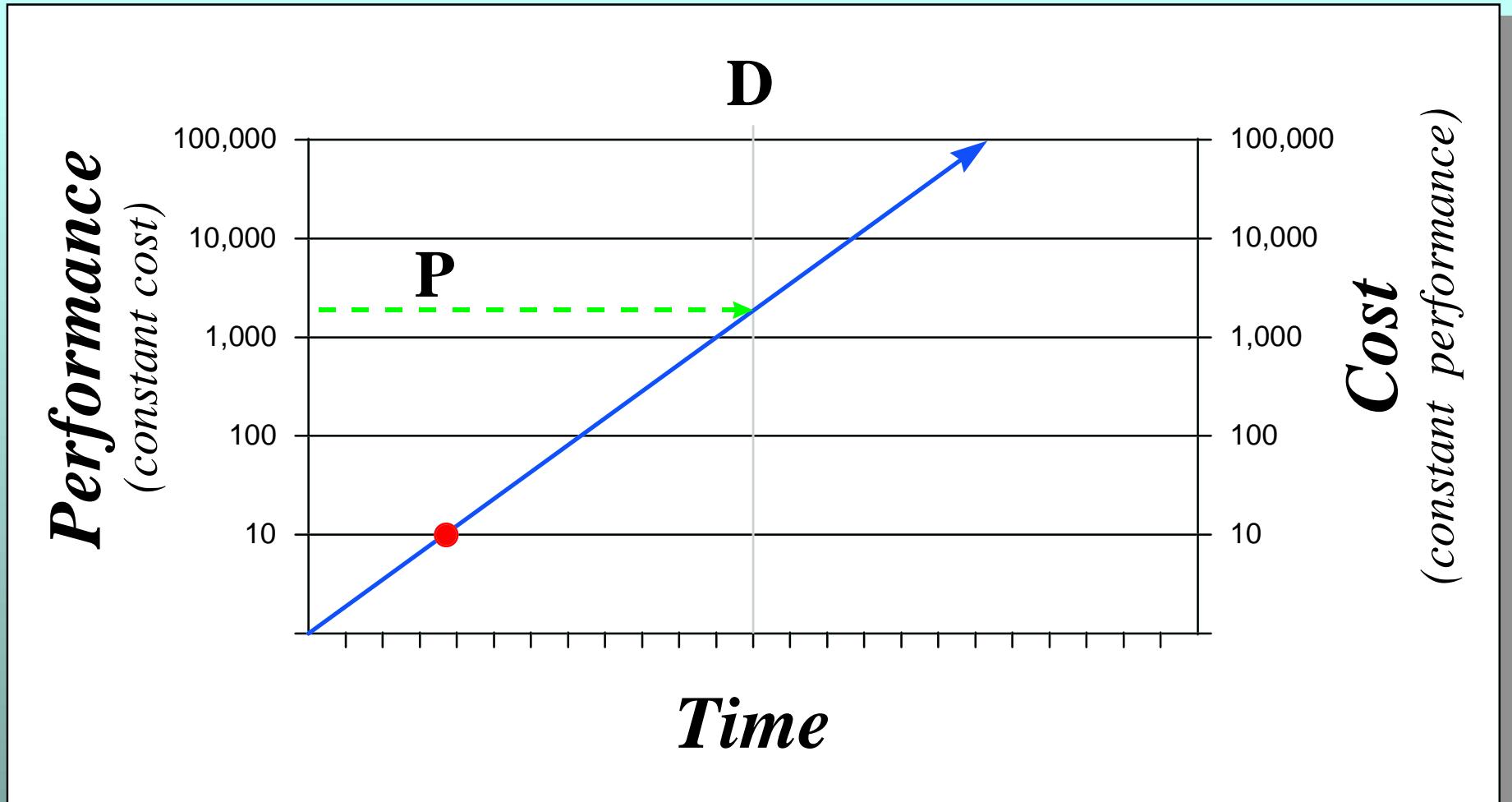
- It's Impossible
- It's Impractical

Moore's Law: *The Effect*

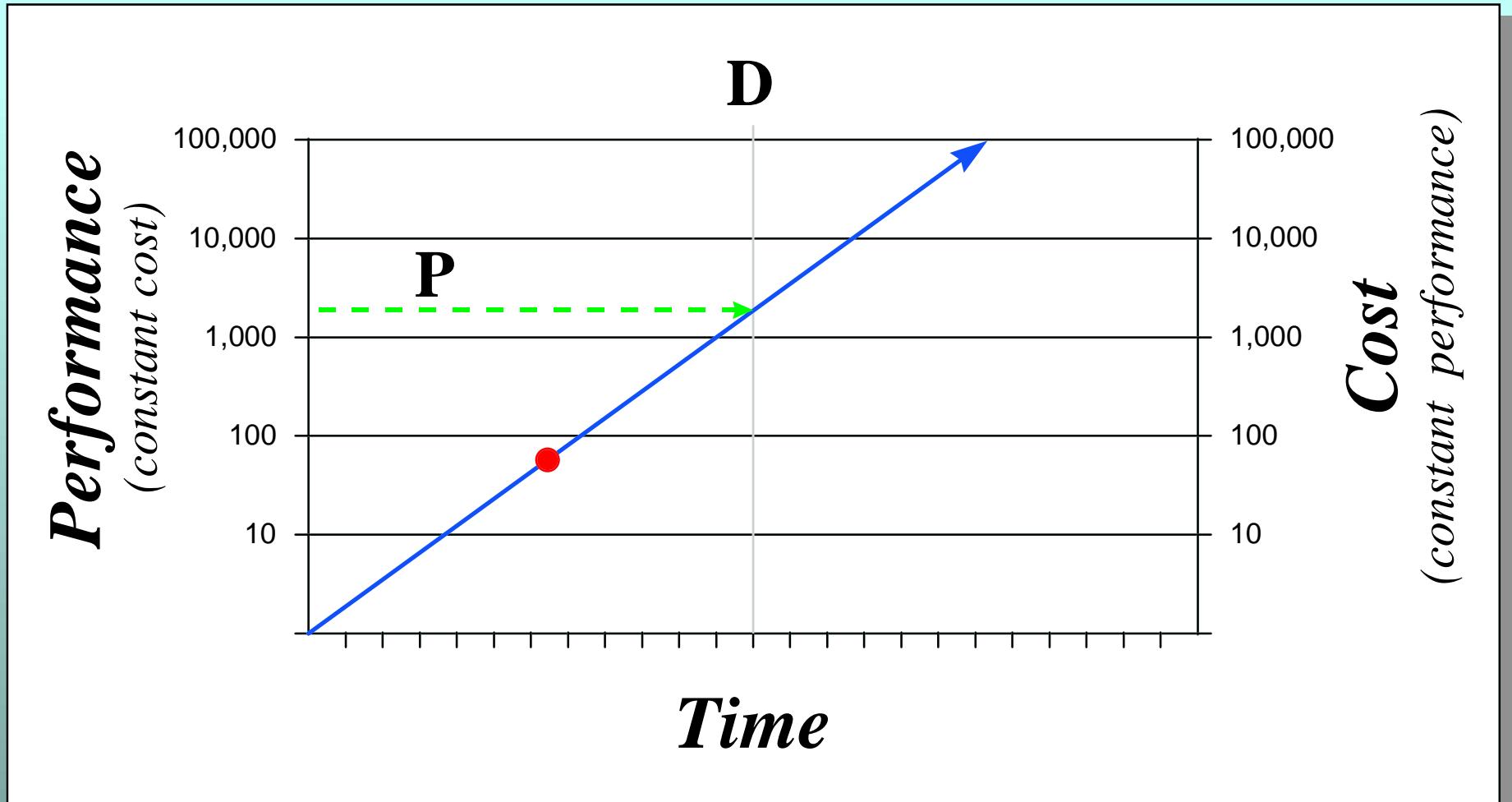
Three Phases of Novel IT Applications

- It's Impossible
- It's Impractical
- It's Overdue

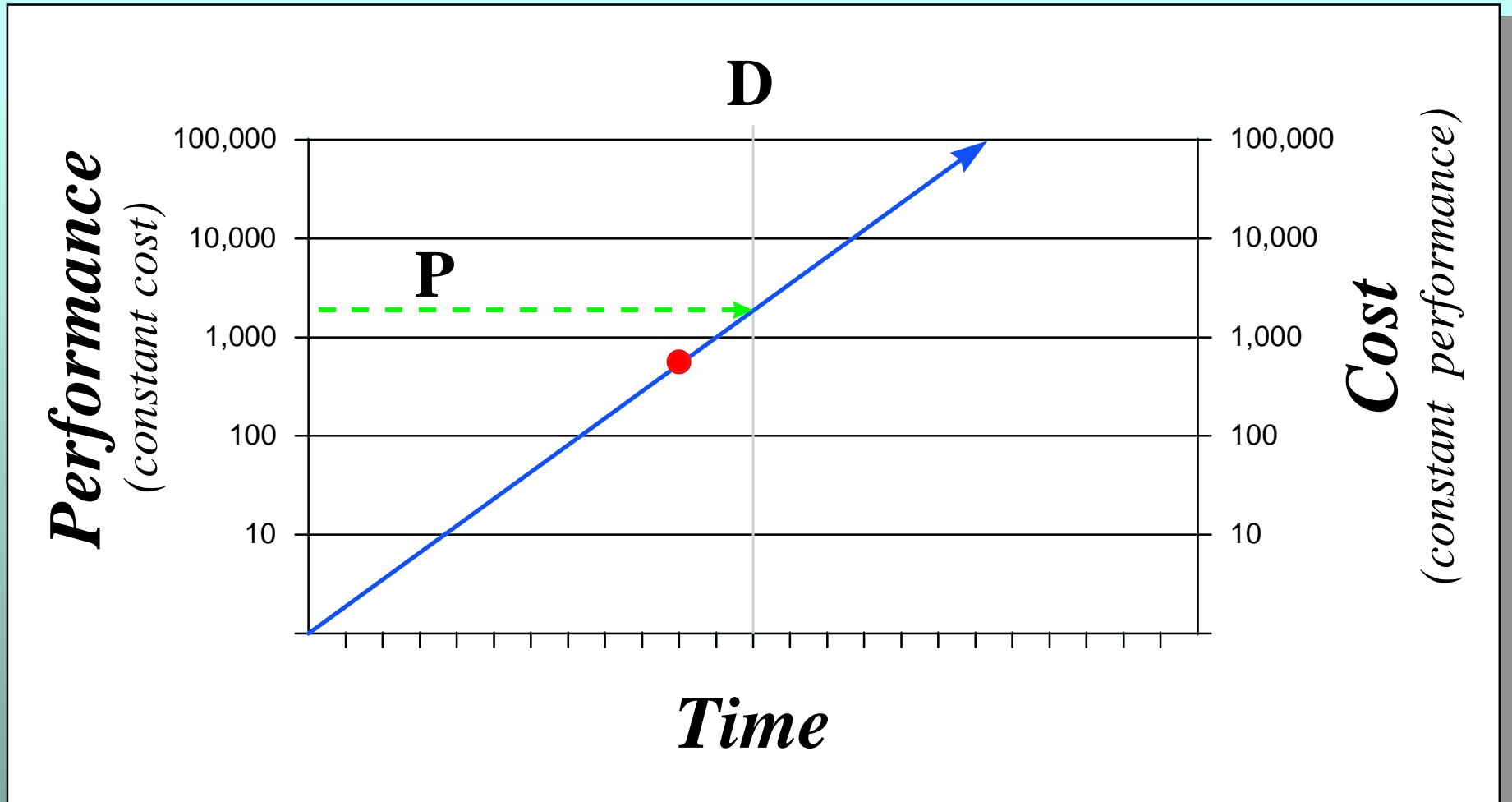
Moore's Law: *The Effect*



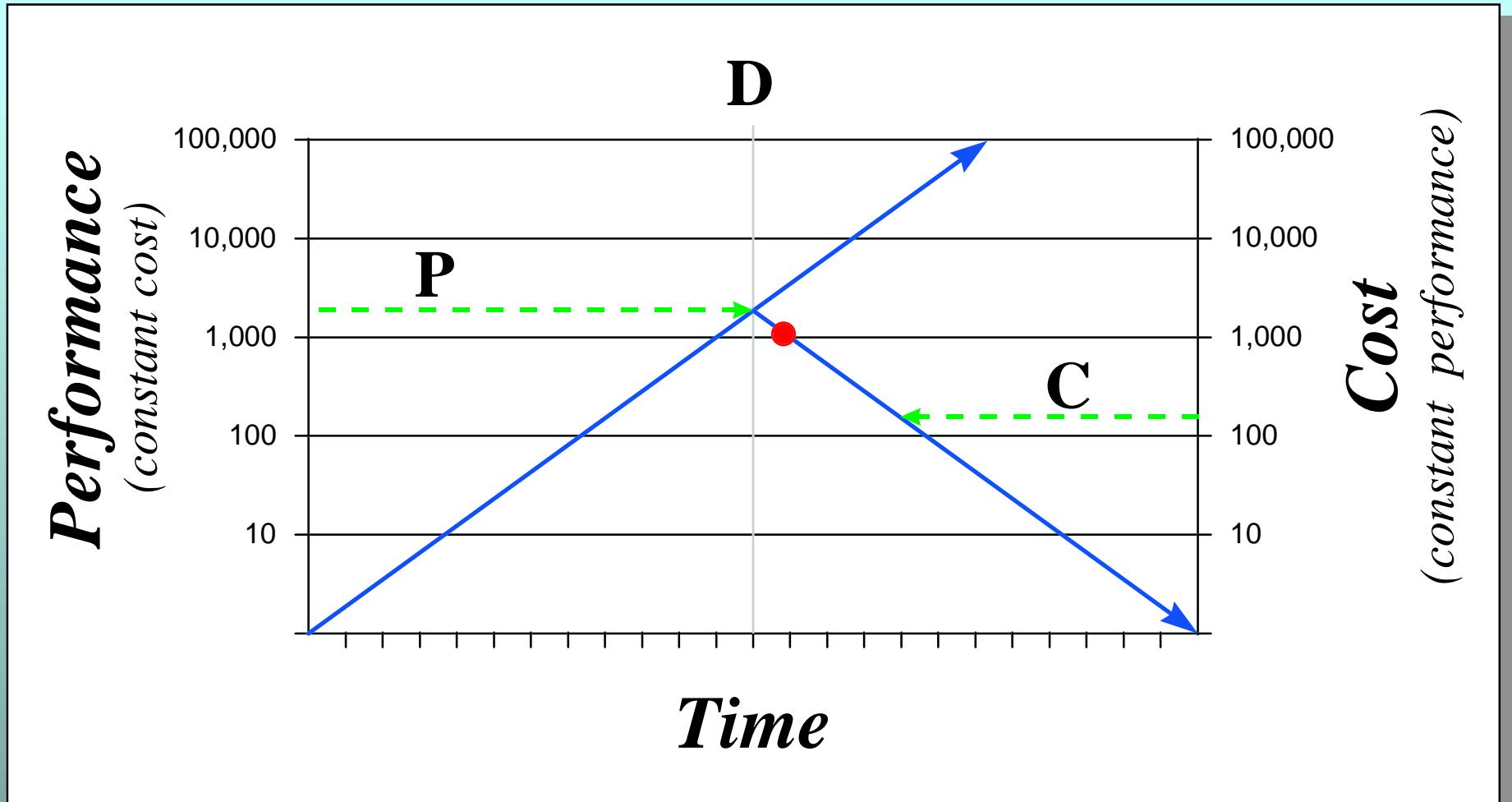
Moore's Law: *The Effect*



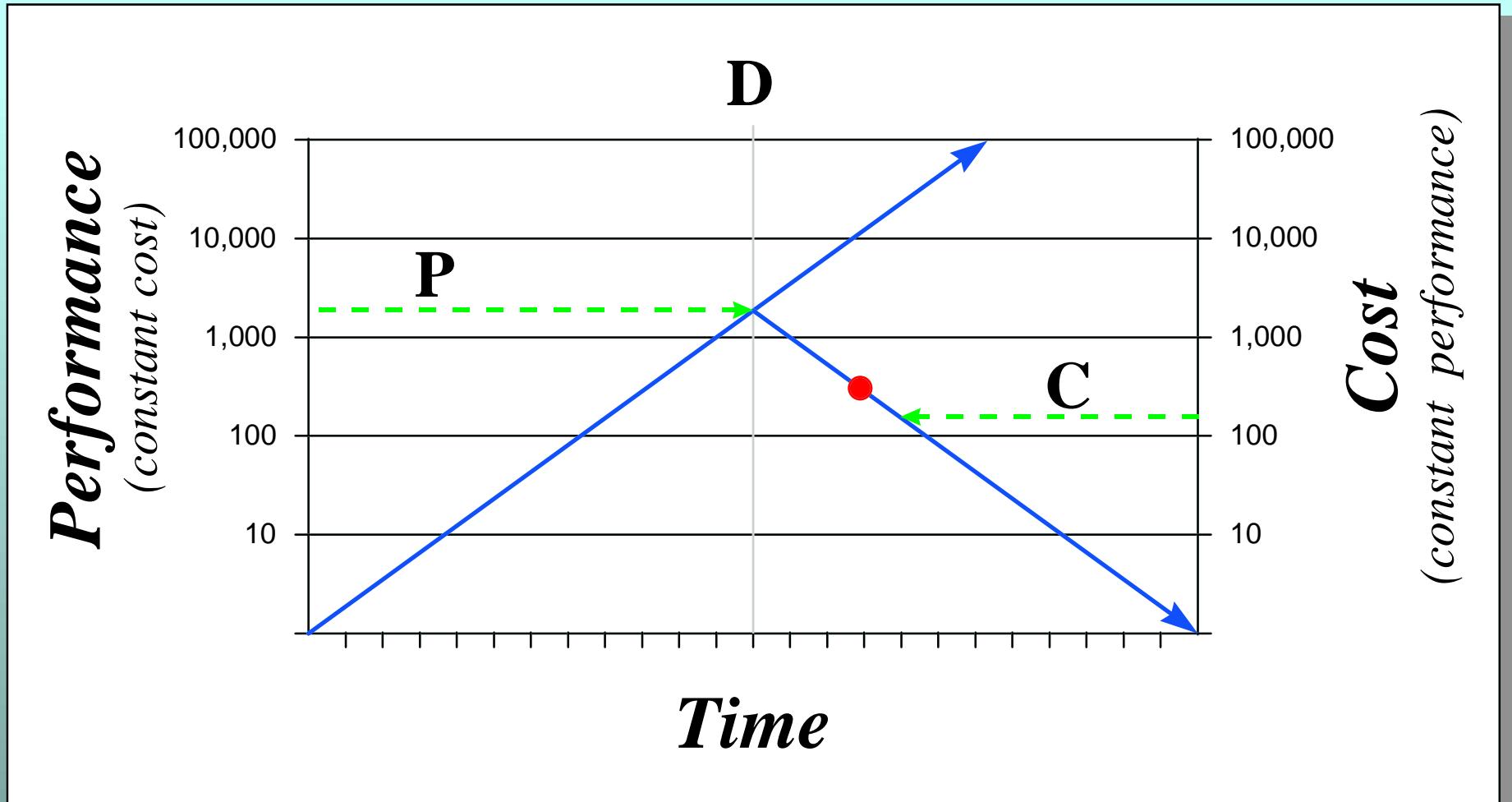
Moore's Law: *The Effect*



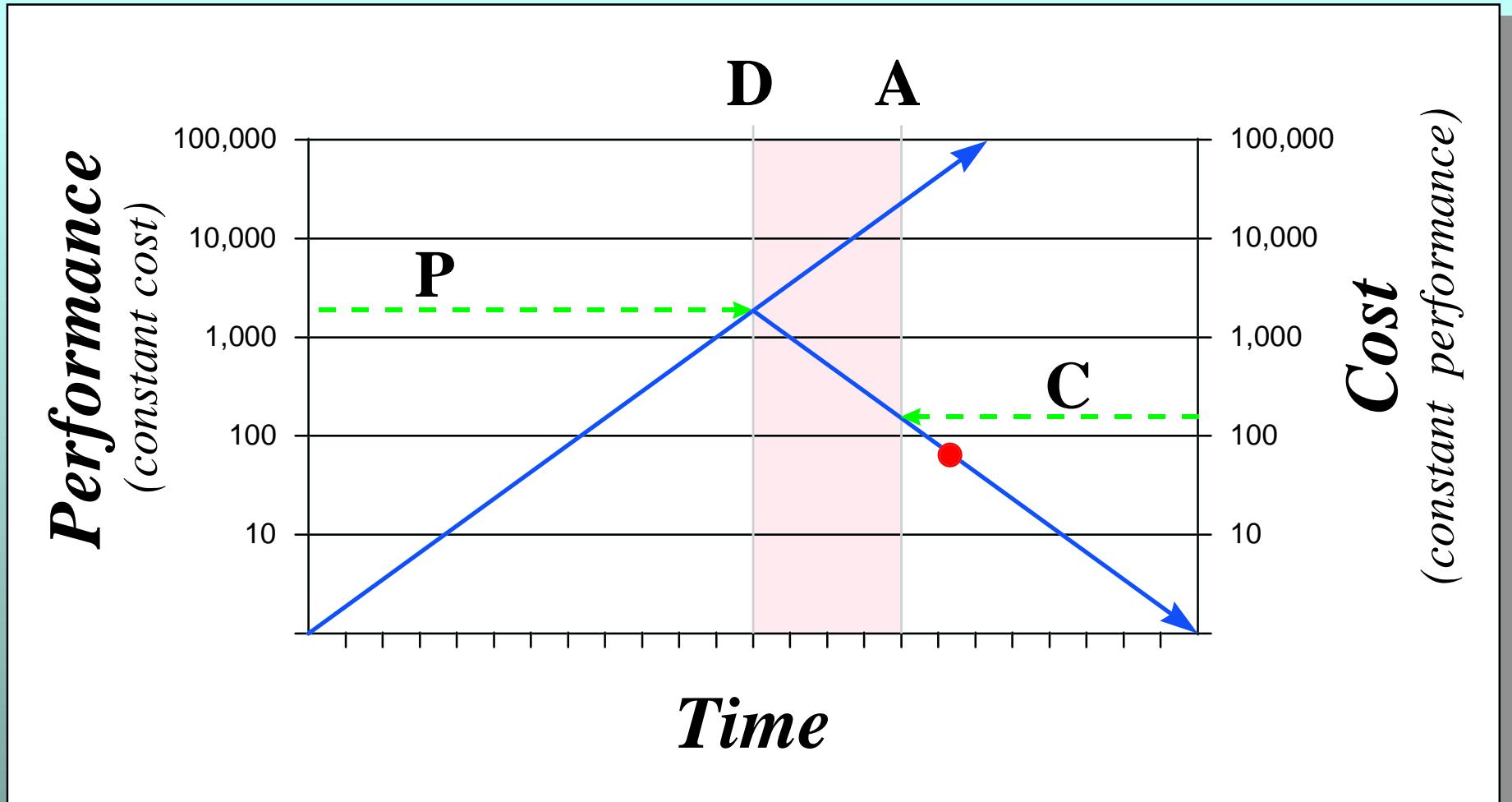
Moore's Law: *The Effect*



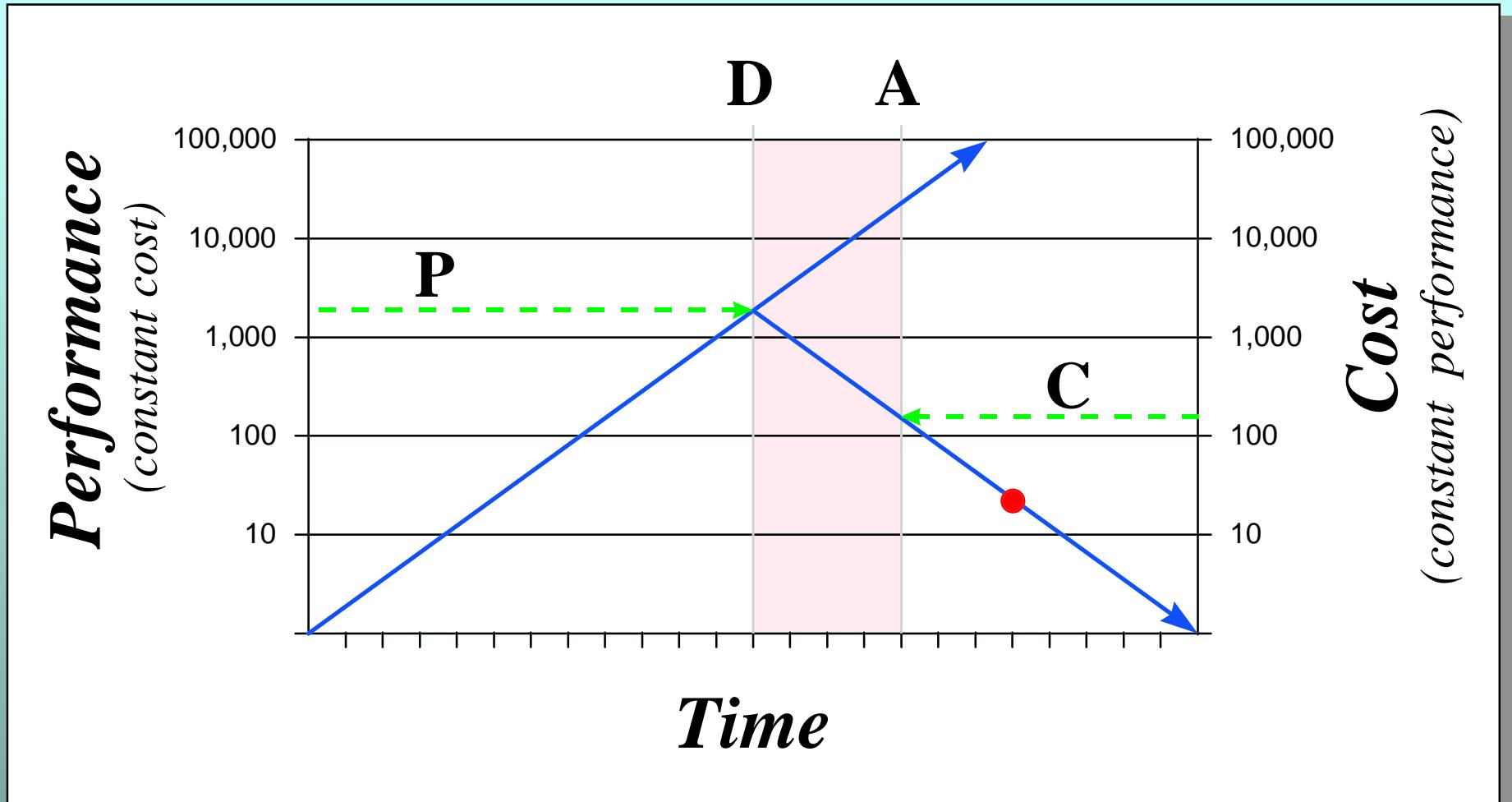
Moore's Law: *The Effect*



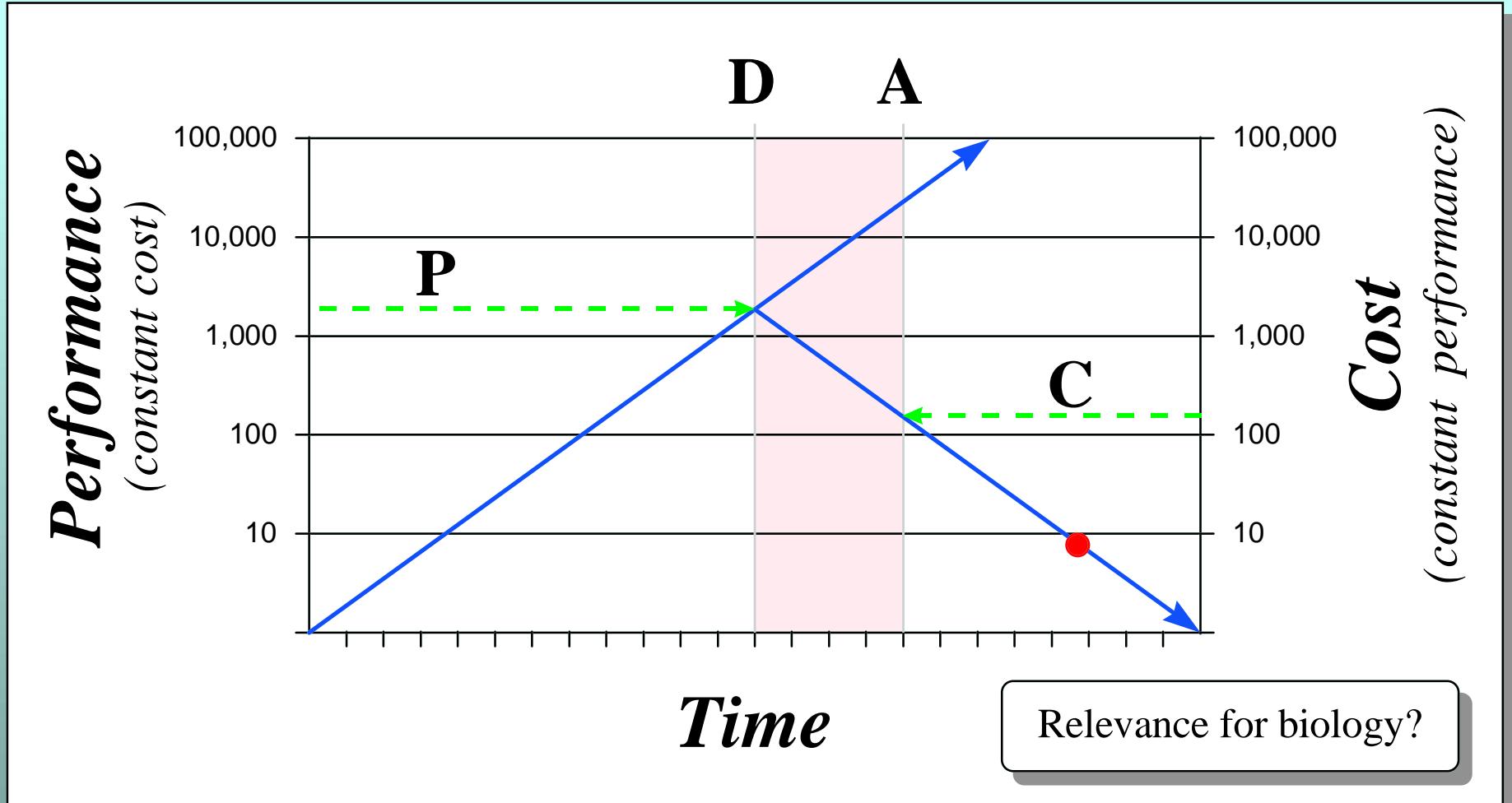
Moore's Law: *The Effect*



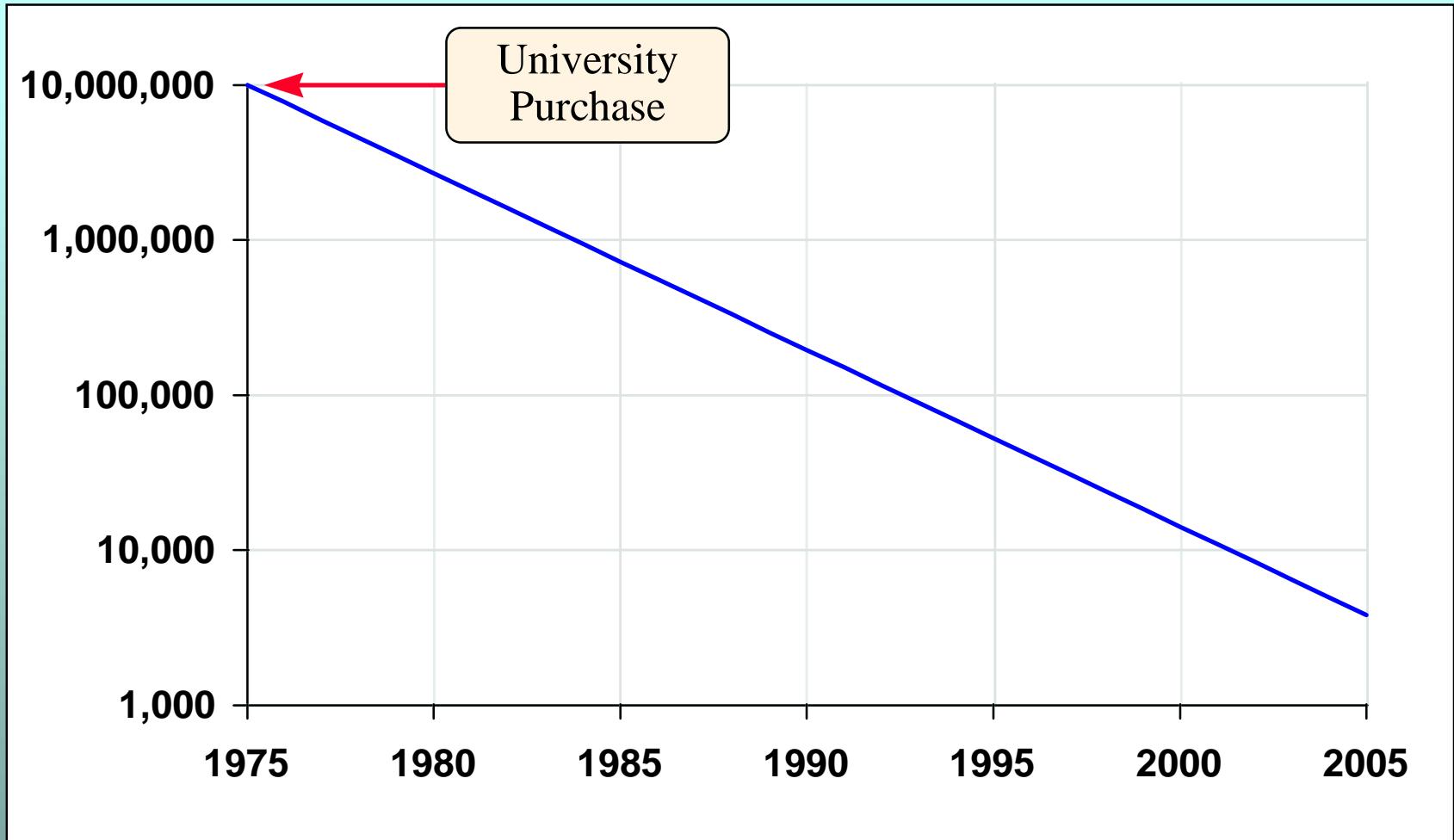
Moore's Law: *The Effect*



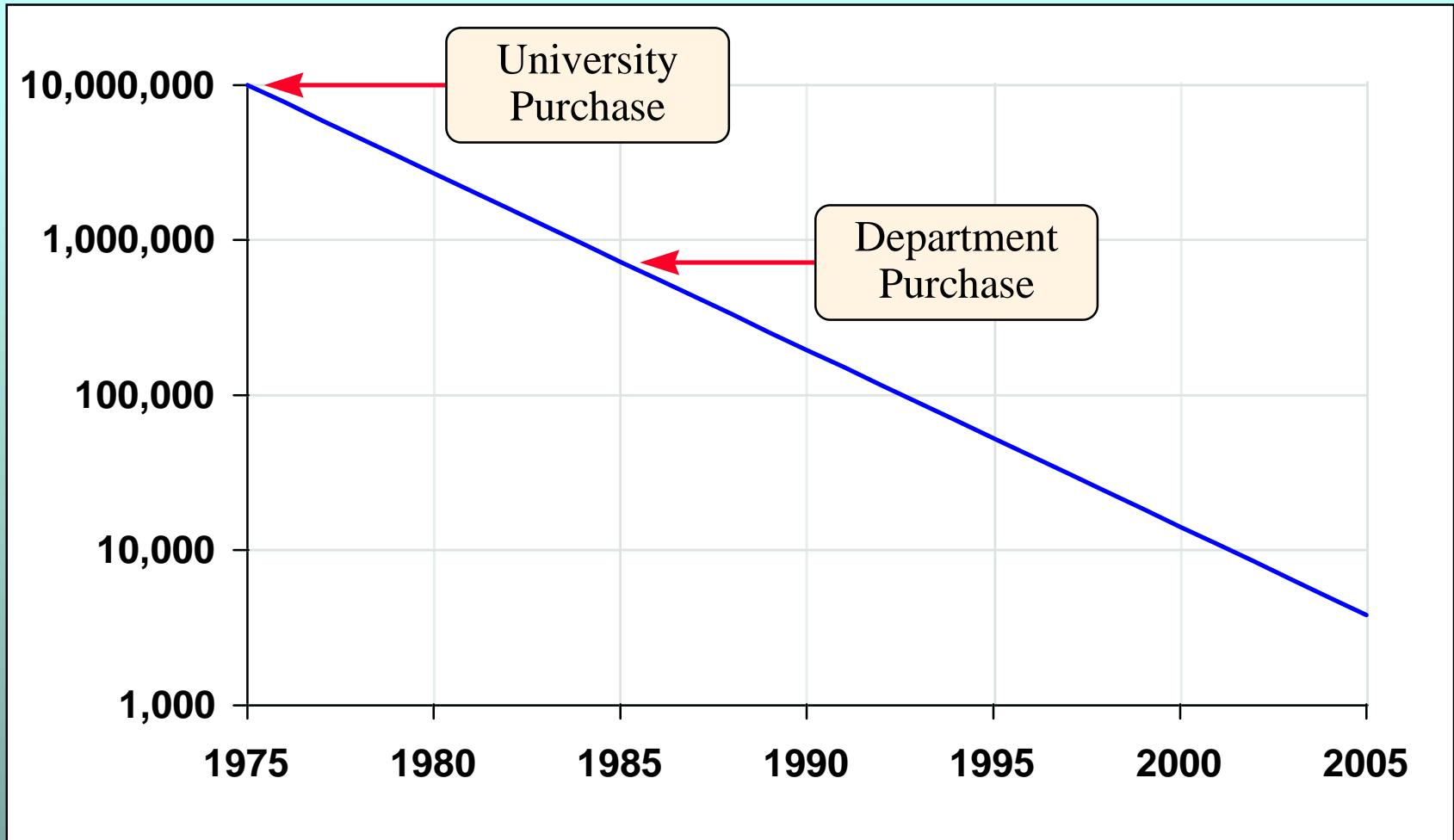
Moore's Law: *The Effect*



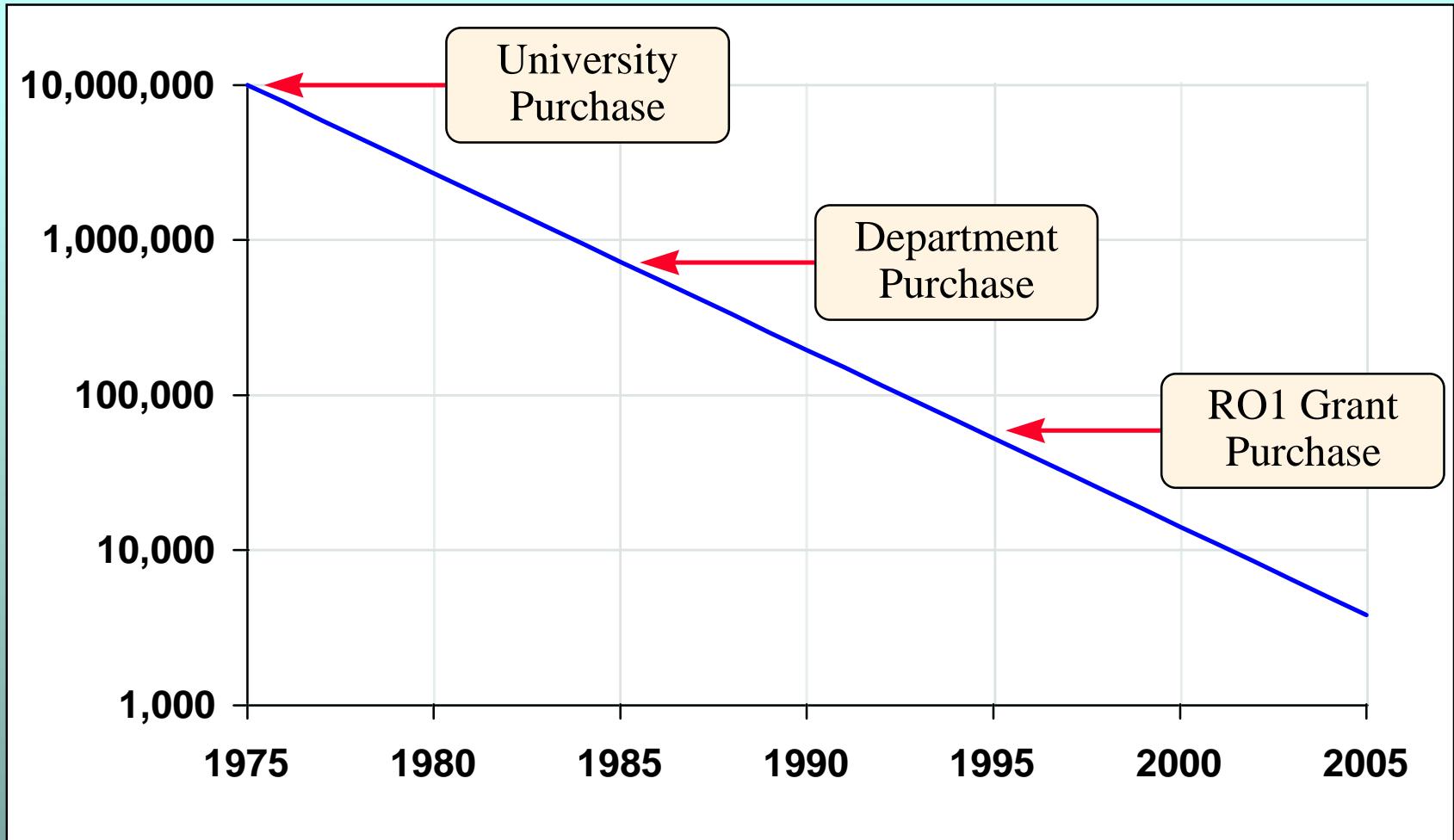
Cost (constant performance)



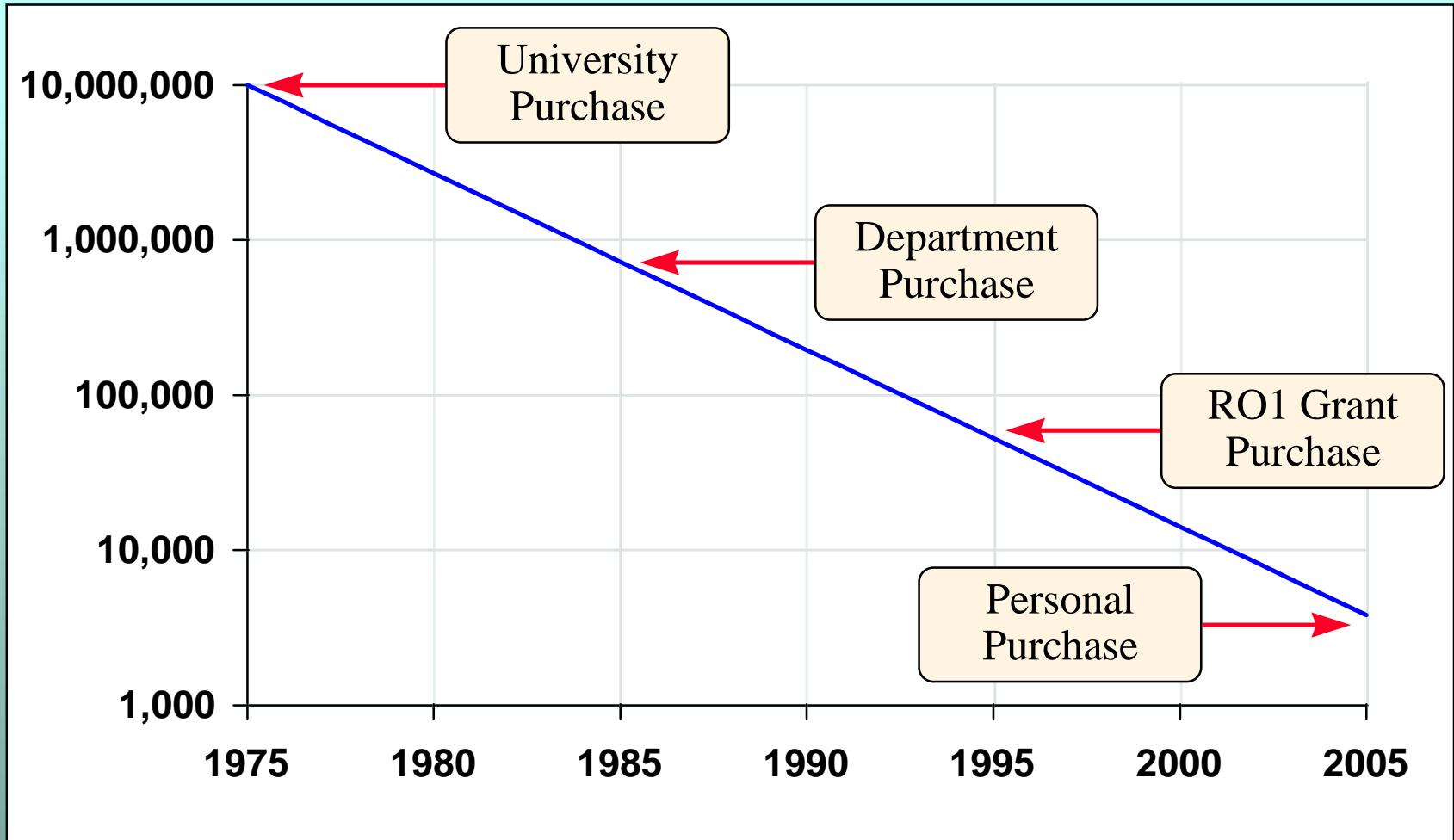
Cost (constant performance)



Cost (constant performance)



Cost (constant performance)



Funding for Information Infrastructure

The Problem

- IT moves at “Internet Speed” and responds rapidly to market forces.

The Problem

- IT moves at “Internet Speed” and responds rapidly to market forces.
- IT will play a central role in 21st Century biology.

The Problem

- IT moves at “Internet Speed” and responds rapidly to market forces.
- IT will play a central role in 21st Century biology.
- Current levels of support for public bio-information infrastructure are too low.

The Problem

- IT moves at “Internet Speed” and responds rapidly to market forces.
- IT will play a central role in 21st Century biology.
- Current levels of support for public bio-information infrastructure are too low.
- Reallocation of federal funding is difficult, and subject to political pressures.

The Problem

- IT moves at “Internet Speed” and responds rapidly to market forces.
- IT will play a central role in 21st Century biology.
- Current levels of support for public bio-information infrastructure are too low.
- Reallocation of federal funding is difficult, and subject to political pressures.
- Federal-funding decision processes are ponderously slow and inefficient.

Federal Funding of Bio-Databases

The challenges:

Federal Funding of Bio-Databases

The challenges:

- providing adequate funding levels

Federal Funding of Bio-Databases

The challenges:

- providing adequate funding levels
- making timely, efficient decisions

Federal Funding of Bio-Databases

Appropriate funding level:

Federal Funding of Bio-Databases

Appropriate funding level:

- approx. 10% of research funding

Federal Funding of Bio-Databases

Appropriate funding level:

- approx. 10% of research funding
- *i.e.*, 1 - 2 **billion** dollars per year

Federal Funding of Bio-Databases

Appropriate funding level:

- approx. 10% of research funding
- *i.e.*, 1 - 2 billion dollars per year

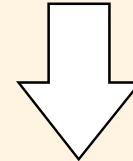
Source of estimate:

- Experience of IT-transformed industries.
- Current support for IT-rich biological research.

Market Forces

In a simple market economy, vendors try to anticipate the needs of buyers and offer products and services to meet those needs.

Vendors



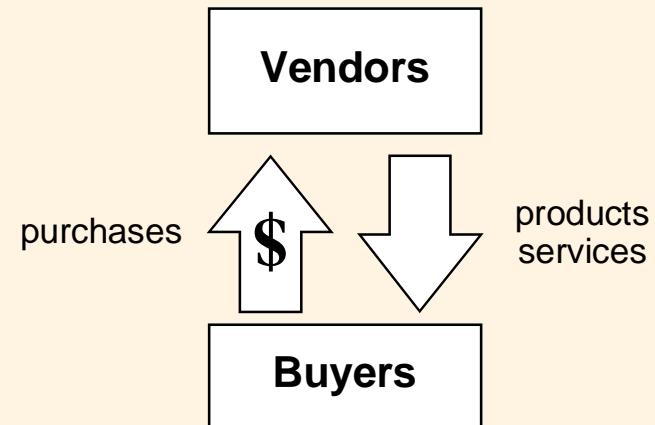
products
services

Buyers

Market Forces

In a simple market economy, vendors try to anticipate the needs of buyers and offer products and services to meet those needs.

Real users decide whether or not to buy a product or service, depending upon whether or not it meets a real need at a reasonable price.



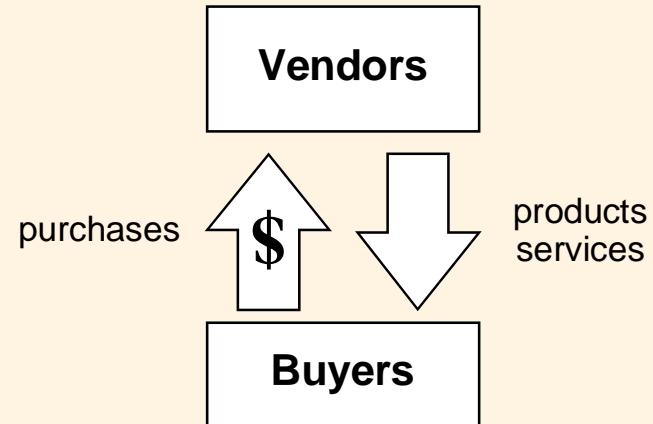
Market Forces

In a simple market economy, vendors try to anticipate the needs of buyers and offer products and services to meet those needs.

Real users decide whether or not to buy a product or service, depending upon whether or not it meets a real need at a reasonable price.

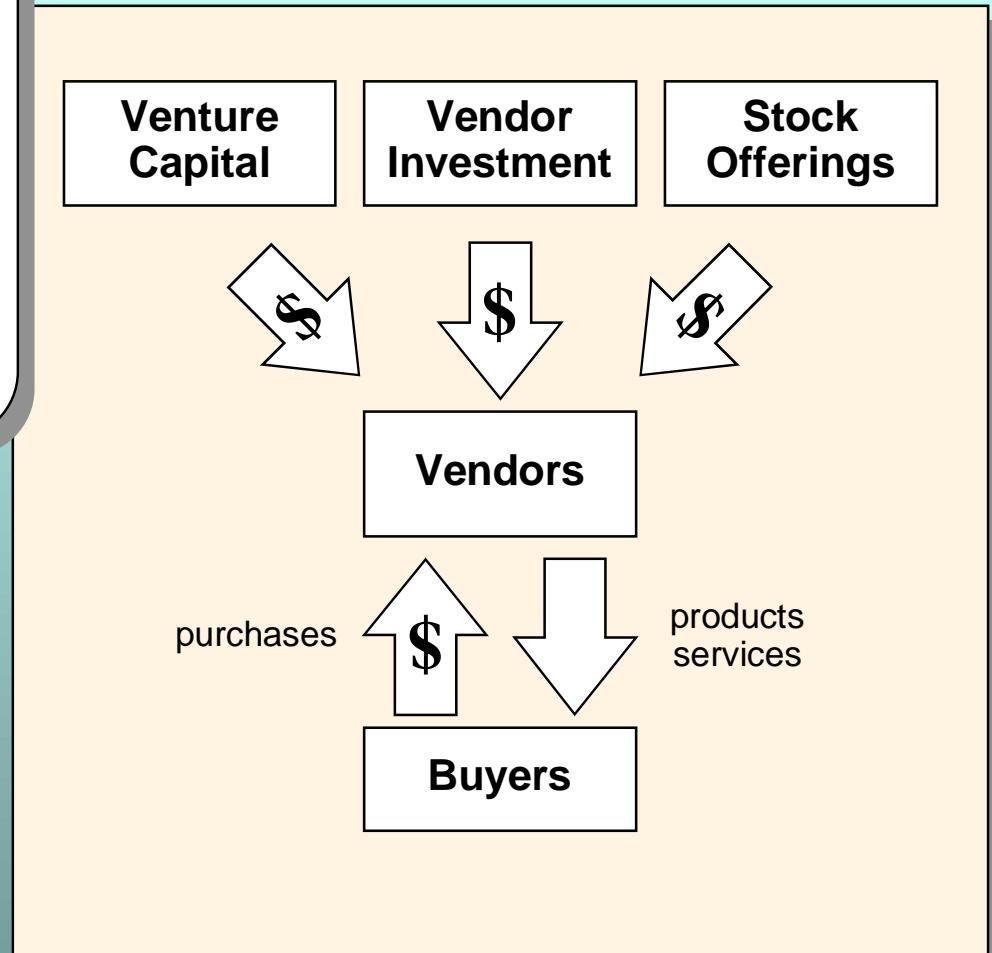
Business 101 Insight:

Successful vendors target a niche and excel at meeting the needs of that niche.



Market Forces

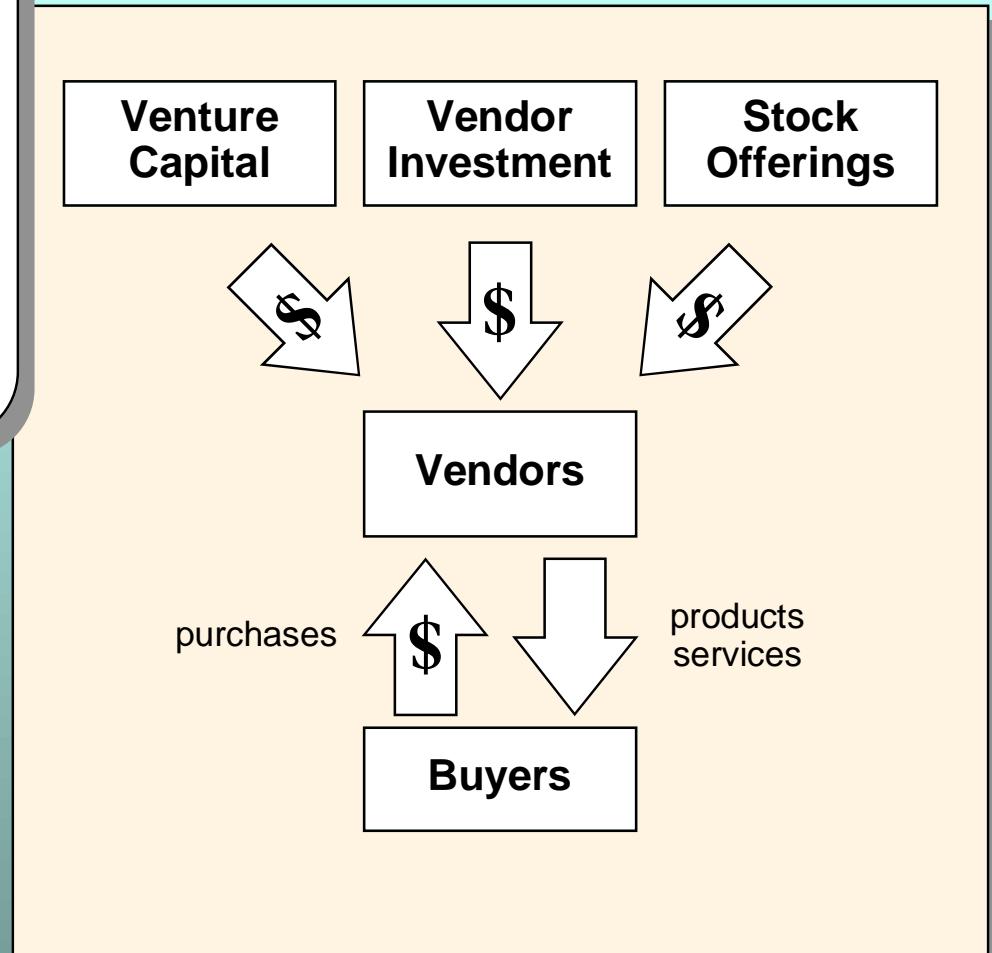
Funding to initiate the development of products and services come from investors, not from buyers.



Market Forces

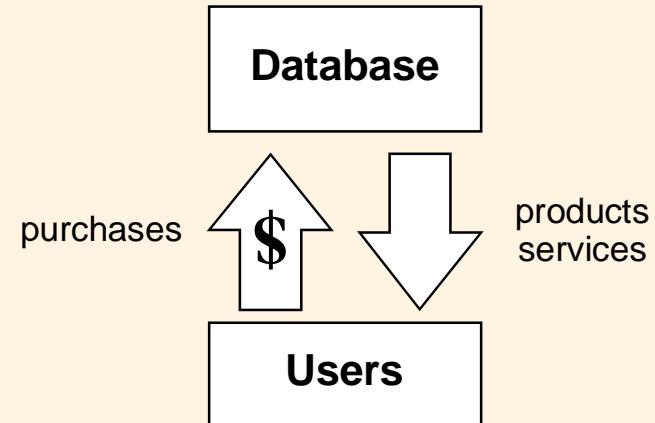
Funding to initiate the development of products and services come from investors, not from buyers.

Investors decide whether or not to provide start-up funding based upon the estimated ability of the vendor to create products and services that will meet real needs at competitive prices.



Federal Funding

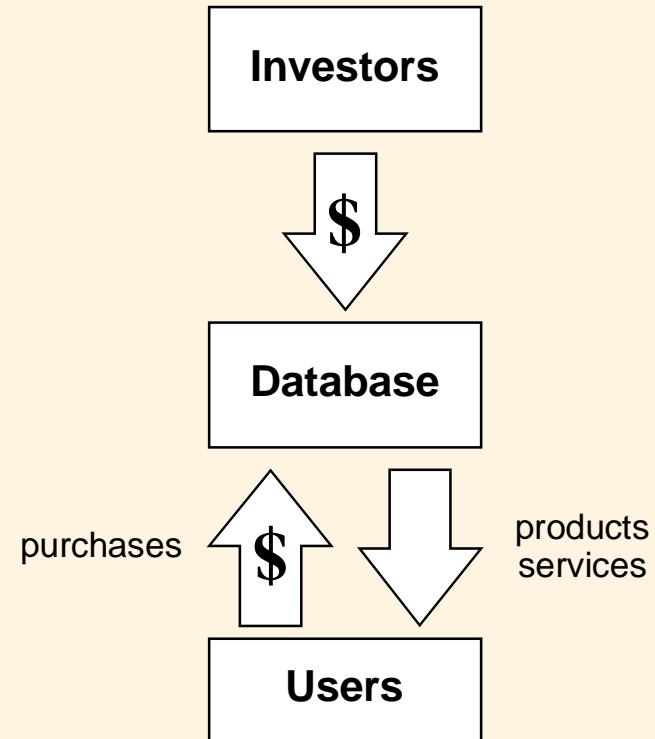
If biological databases were driven by market forces, individual users would choose what services they need and individual database providers would choose what services to make available.



Federal Funding

If biological databases were driven by market forces, individual users would choose what services they need and individual database providers would choose what services to make available.

Investors would provide start-up money on the likelihood of successful products and services being developed.

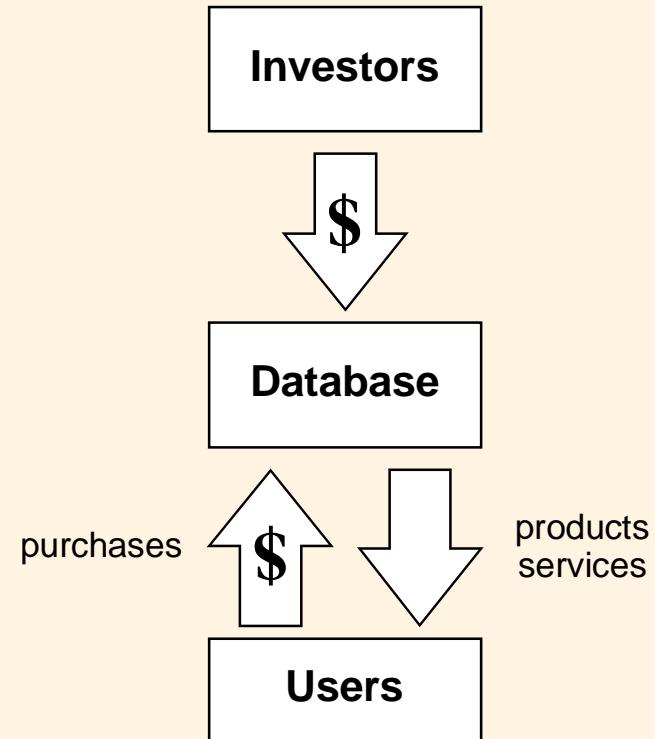


Federal Funding

If biological databases were driven by market forces, individual users would choose what services they need and individual database providers would choose what services to make available.

Investors would provide start-up money on the likelihood of successful products and services being developed.

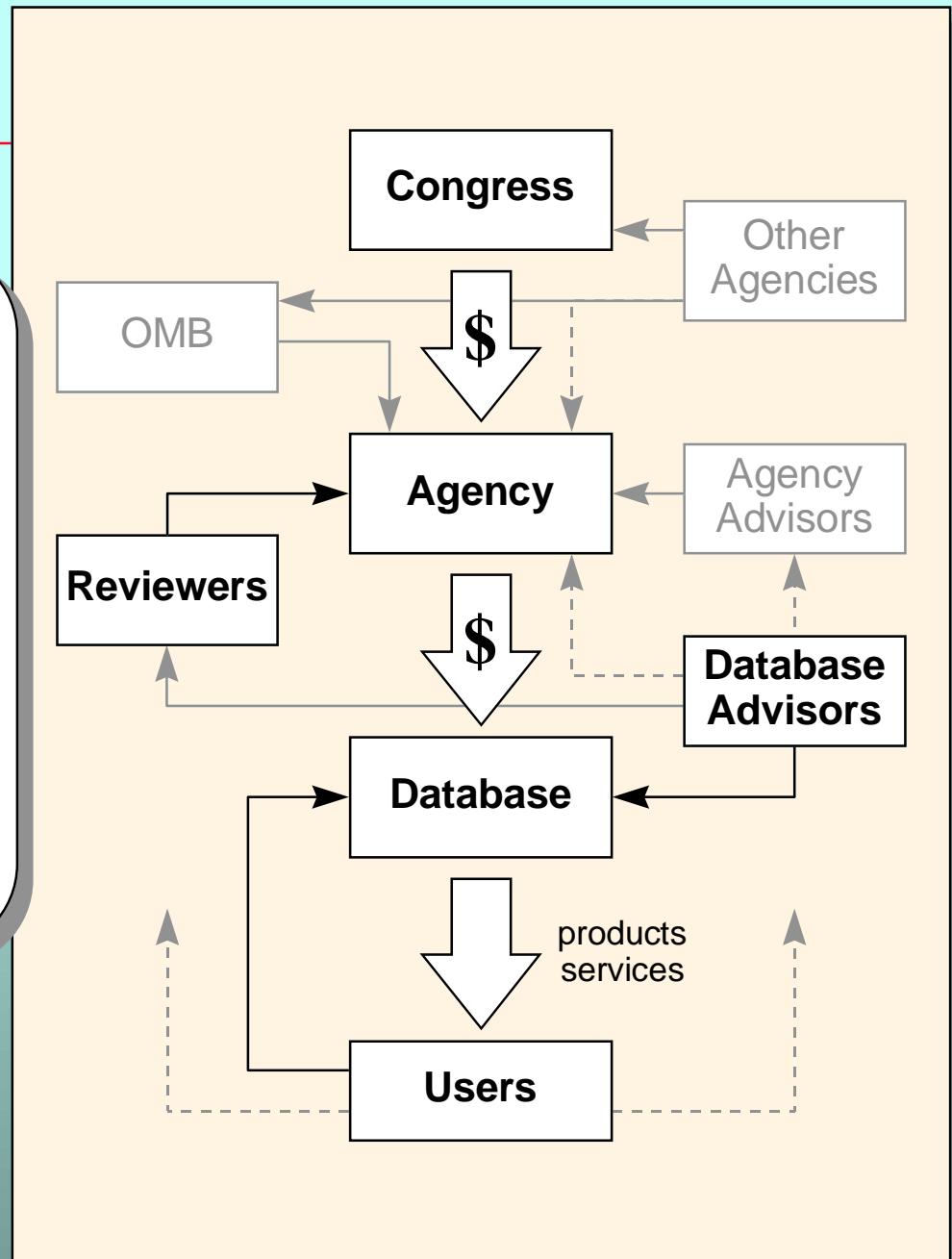
Ultimate success would depend on meeting the needs of real users. Decisions could be made rapidly, in response to changing needs and emerging opportunities.



Federal Funding

Instead, funding decisions for biological databases can follow a ponderously slow course, with almost no opportunity for input from real users.

Those most knowledgeable about a particular database are often excluded from participating in the review process because of a possible “conflict of interest” status with the database provider.



Federal Funding of Bio-Databases

Possible solution - create market forces:

Federal Funding of Bio-Databases

Possible solution - create market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.

Federal Funding of Bio-Databases

Possible solution - create market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.
- start supporting the demand side through fast, efficient processes.

Federal Funding of Bio-Databases

Possible solution - create market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.
- start supporting the demand side through fast, efficient processes.
- provide guaranteed supplementary funding, redeemable only for access to bio-databases.

Federal Funding of Bio-Databases

Possible solution - create market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.
- start supporting the demand side through fast, efficient processes.
- provide guaranteed supplementary funding, redeemable only for access to bio-databases.
- **data stamps**

Federal Funding of Bio-Databases

Possible solution - create market forces:

- stop supporting the supply side of biodatabases through slow, inefficient processes.
- start supporting the demand side through fast, efficient processes.
- provide guaranteed supplementary funding, redeemable only for access to bio-databases.
- **data stamps, AKA *food (for-thought) stamps* ?!**

Food (for thought) Stamps

Funding Agencies could:

- provide a 10% supplement to **every** research grant in the form of “stamps” redeemable only at database providers.

Food (for thought) Stamps

Funding Agencies could:

- provide a 10% supplement to **every** research grant in the form of “stamps” redeemable only at database providers.
- allow the “stamps” to be transferable among scientists, so that a market for them could emerge.

Food (for thought) Stamps

Funding Agencies could:

- provide a 10% supplement to **every** research grant in the form of “stamps” redeemable only at database providers.
- allow the “stamps” to be transferable among scientists, so that a market for them could emerge.
- provide funding only after the stamps have been redeemed at a database provider.

Food (for thought) Stamps

Problems:

- how to estimate the amount of FFT stamps that would actually be redeemed (and thus the required budget set-aside).

Food (for thought) Stamps

Problems:

- how to estimate the amount of FFT stamps that would actually be redeemed (and thus the required budget set-aside).
- how to identify “approved” database providers.

Food (for thought) Stamps

Problems:

- how to estimate the amount of FFT stamps that would actually be redeemed (and thus the required budget set-aside).
- how to identify “approved” database providers.
- how to initiate the FFT system.

Food (for thought) Stamps

Problems:

- how to estimate the amount of FFT stamps that would actually be redeemed (and thus the required budget set-aside).
- how to identify “approved” database providers.
- how to initiate the FFT system.
- etc etc

Food (for thought) Stamps

Alternatives (if no solution emerges):

- increasingly inefficient research activities (abject failure will occur when it becomes easier to redo research than to obtain the results of prior work).

Food (for thought) Stamps

Alternatives (if no solution emerges):

- increasingly inefficient research activities (abject failure will occur when it becomes easier to redo research than to obtain the results of prior work).
- loss of access to bio-databases for public-sector research.

Food (for thought) Stamps

Alternatives (if no solution emerges):

- increasingly inefficient research activities (abject failure will occur when it becomes easier to redo research than to obtain the results of prior work).
- loss of access to bio-databases for public-sector research.
- movement of majority of “important” biological research into the private sector.

Food (for thought) Stamps

Alternatives (if no solution emerges):

- increasingly inefficient research activities (abject failure will occur when it becomes easier to redo research than to obtain the results of prior work).
- loss of access to bio-databases for public-sector research.
- movement of majority of “important” biological research into the private sector.
- loss of American pre-eminence (if other countries solve the problems first).

Food (for thought) Stamps

Bottom Line:

- This might not be the answer, but

Food (for thought) Stamps

Bottom Line:

- This might not be the answer, but
- it's FOOD FOR THOUGHT

Slides:

<http://www.esp.org/rjr/beckman.pdf>