分子生物计算
(*Perl 语言编程*)

天津医科大学
生物医学工程与技术学院

2016-2017 学年上学期（秋）
2014 级生信班

第一章　绪论

伊现富（Yi Xianfu）

天津医科大学（TIJMU）
生物医学工程与技术学院

2016 年 11 月

教学提纲

教学提纲

课程安排 | 授课教材


Beginning Perl for Bioinformatics
An Introduction to Perl for Biologists
Perl语言在生物信息学中的应用
——基础篇
O'REILLY®
James Tisdall 著
伊现富 译

课程安排 | 理论课

后 9 周，每周三，下午四节（13:30-17:30），西楼 610

授课内容

- 教材内容：第一章……～第十章
- 补充知识：Markdown, Git, ...

课程安排 | 理论课 | 自主学习

- 上课期间以小组形式编写程序
- 任务主题不限（生信方面的最好）
- 理论课的最后一/两次课（2～4 学时）进行报告
- 小组成员人人都要参与
- 既参与编程又进行报告者加分
- 作为平时成绩的一大部分

目的

按照用户的需求生成随机密码。

需求

- 指定生成密码的元素（默认：混合使用数字、大小写字母、符号）
- 指定生成密码的长度（默认：12 个字符）
- 指定生成密码的个数（默认：1 个）
- 指定密码中元素的比例/字符数（默认：无比例/字符数要求）
- 指定需要排除的元素（比如：0 和 O，1 和 l；默认：无）
- 挖掘更多人性化需求……

代码仓库

- Yixf-Education/project_Perl

目的

按照用户的需求生成随机密码。

需求

- 指定生成密码的元素（默认：混合使用数字、大小写字母、符号）
- 指定生成密码的长度（默认：12 个字符）
- 指定生成密码的个数（默认：1 个）
- 指定密码中元素的比例/字符数（默认：无比例/字符数要求）
- 指定需要排除的元素（比如：0 和 O，1 和 l；默认：无）
- 挖掘更多人性化需求……

代码仓库

- Yixf-Education/project_Perl

目的

按照用户的需求生成随机密码。

需求

- 指定生成密码的元素（默认：混合使用数字、大小写字母、符号）
- 指定生成密码的长度（默认：12 个字符）
- 指定生成密码的个数（默认：1 个）
- 指定密码中元素的比例/字符数（默认：无比例/字符数要求）
- 指定需要排除的元素（比如：0 和 O，1 和 l；默认：无）
- 挖掘更多人性化需求……

代码仓库

- Yixf-Education/project_Perl

课程安排 | 实验课

后 9 周，每周四，上午后两节（10:10-12:00），教一楼 304

实验内容

- 紧跟理论课进度
- 教材上的程序/实例

课程安排 | 考核方式

1. 理论课：60%
   1. 平时表现：10%
   2. 闭卷考试：50%
2. 实验课：40%
   1. 平时表现：20%
   2. 实验报告：20%

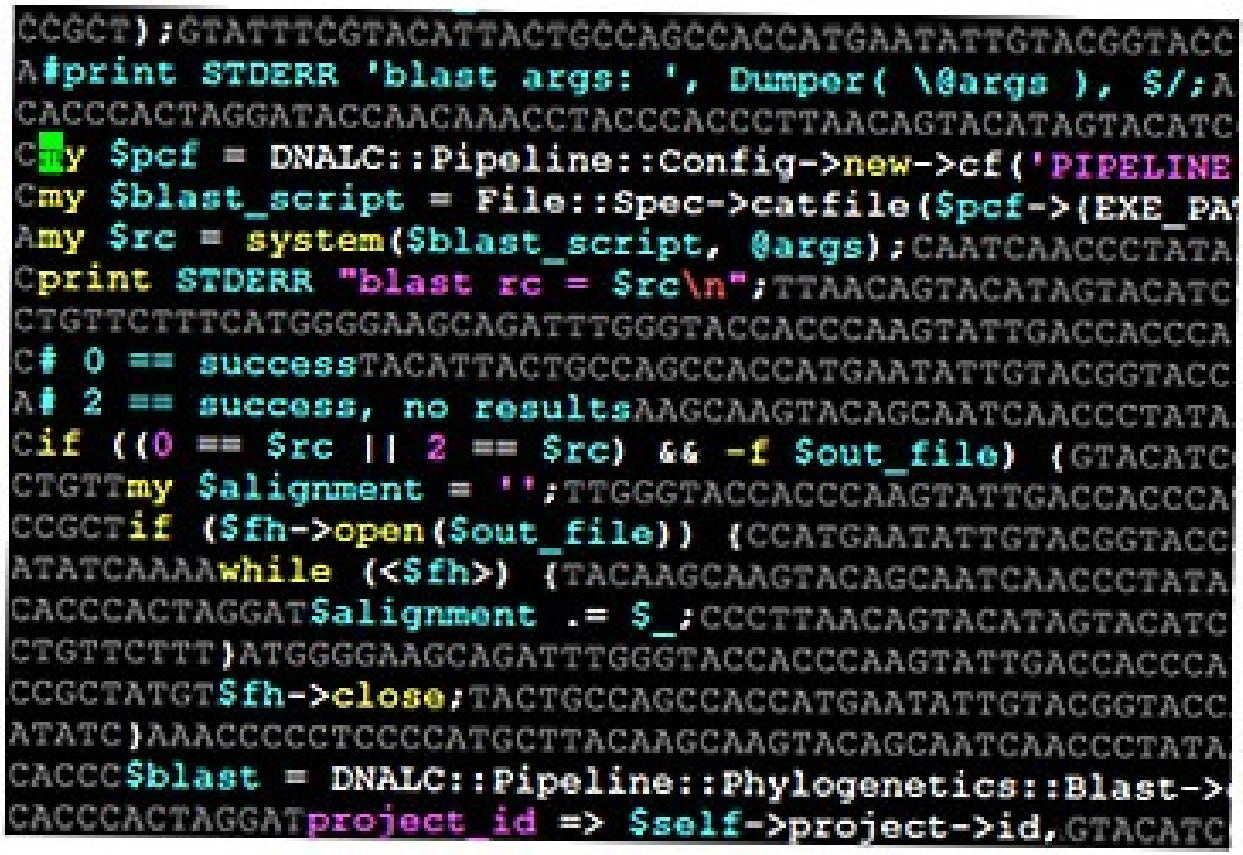
教学提纲

生物信息学（bioinformatics）

“If you can’t do bioinformatics, you can’t do biology.”

Lincoln Stein

(Biologist and former CSHL Professor)



生物信息学 | 维基百科

生物信息学（bioinformatics）利用应用数学、信息学、统计学和计算机科学的方法研究生物学的问题。目前的生物信息学基本上只是分子生物学与信息技术（尤其是互联网技术）的结合体。

生物信息学的研究材料和结果就是各种各样的生物学数据，其研究工具是计算机，研究方法包括对生物学数据的搜索（收集和筛选）、处理（编辑、整理、管理和显示）及利用（计算、模拟）。

目前主要的研究方向有：序列比对、基因识别、基因重组、蛋白质结构预测、基因表达、蛋白质反应的预测，以及建立进化模型。

生物信息学 | 维基百科（续）

生物学技术往往生成大量的嘈杂数据。与数据挖掘类似，生物信息学利用数学工具从大量数据中提取有用的生物学信息。生物信息学所要处理的典型问题包括：重新组装在霰弹枪测序法测序过程中被打散的 DNA 序列，从蛋白质的氨基酸序列预测蛋白质结构，利用 mRNA 微阵列或质谱仪的数据检验基因调控的假说。

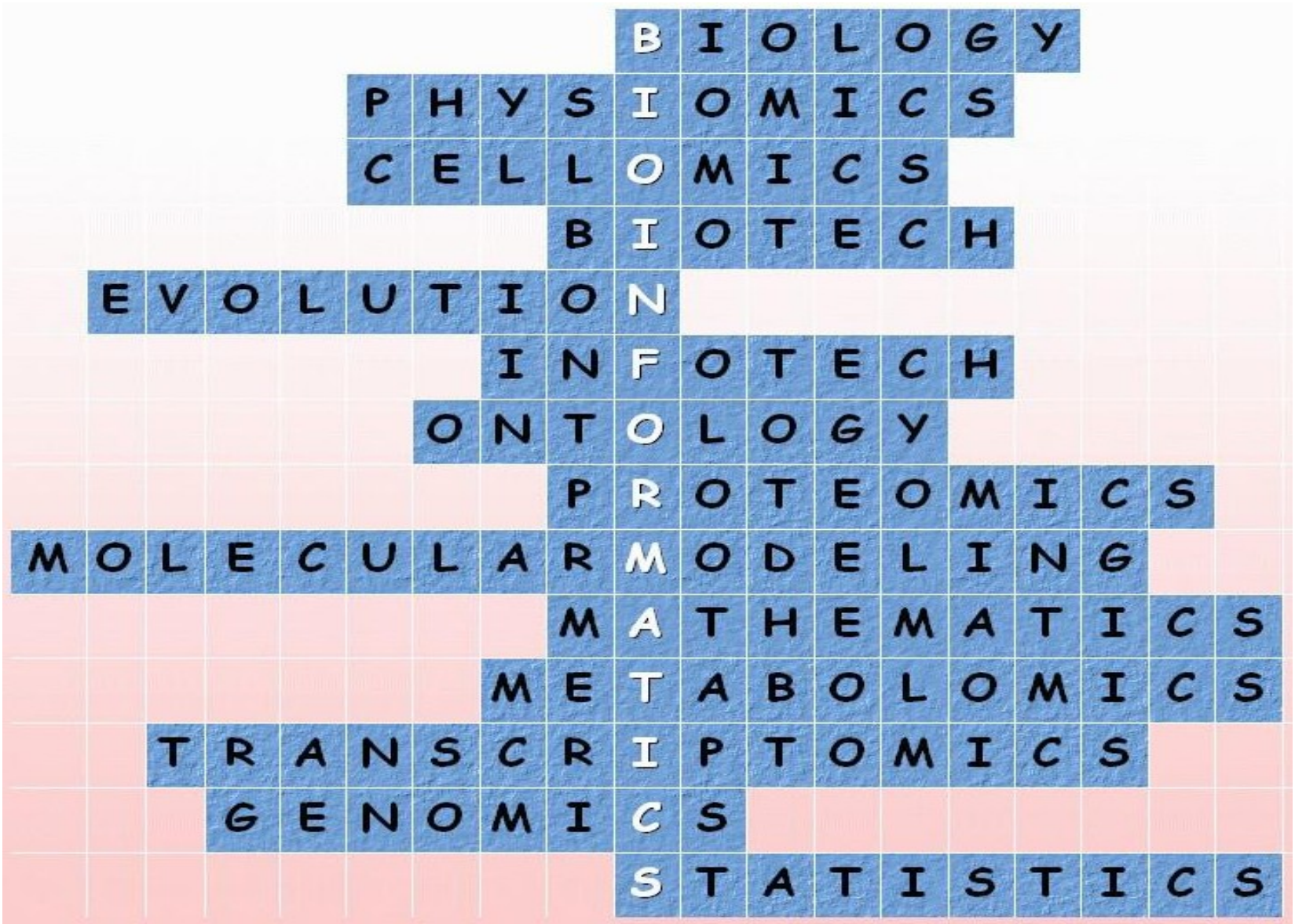
某些人将计算生物学作为生物信息学的同义词处理；但是另外一些人认为计算生物学和生物信息学应当被当作不同的条目处理，因为生物信息学更侧重于生物学领域中计算方法的使用和发展，而计算生物学强调应用信息学技术对生物学领域中的假说进行检验，并尝试发展新的理论。

生物信息学可以定义为对分子生物学中两类信息流的研究：

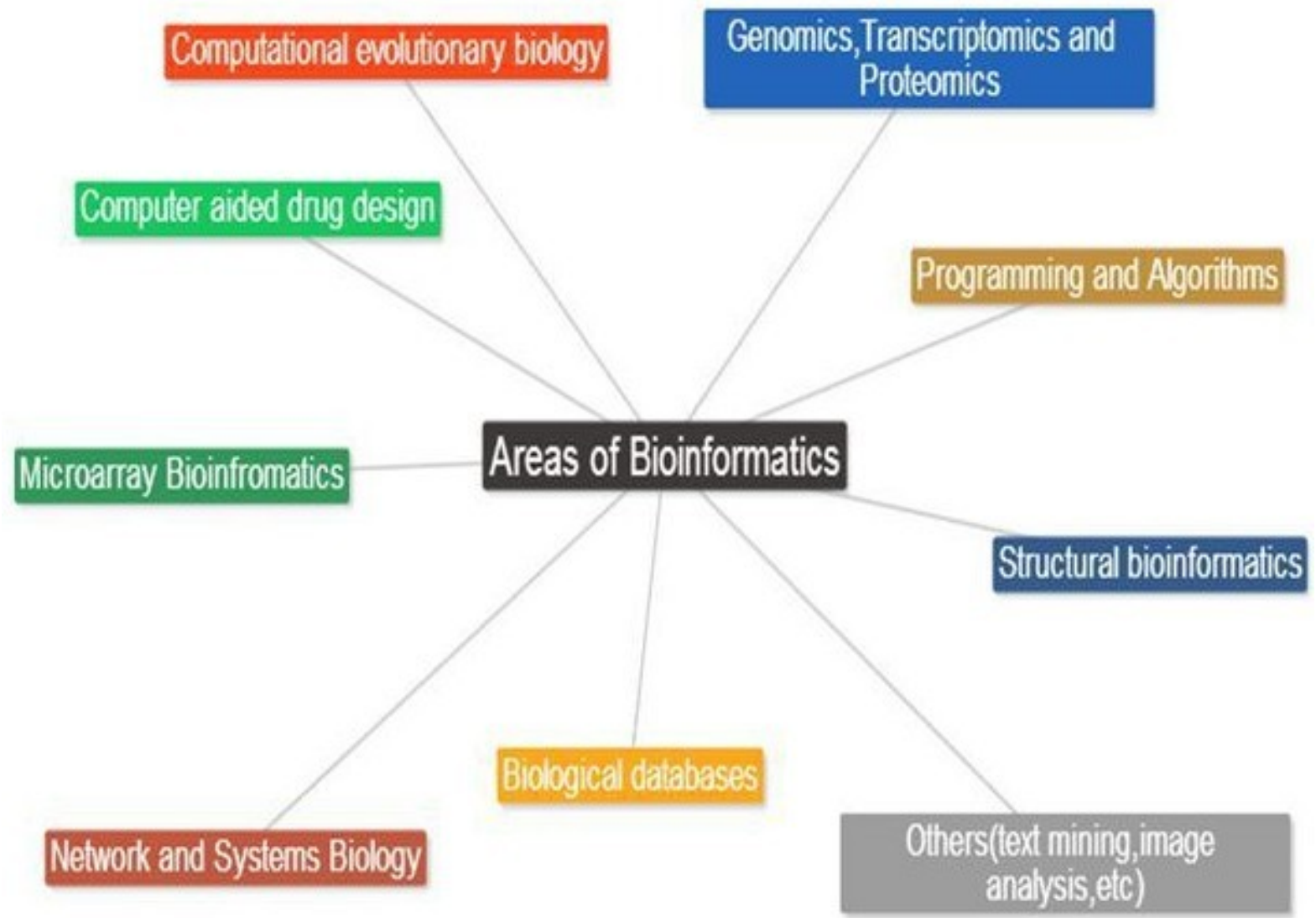
第一类信息流源于分子生物学的中心法则：DNA 序列被转录为 mRNA 序列，后者被翻译为蛋白质序列。蛋白质序列继而折叠为具有功能的三维结构。按照达尔文演化理论，这些功能被生物体的环境所选择，从而驱动群体中 DNA 序列的进化。因此，第一类的生物信息学应用关注于中心法则中任一阶段的信息传递，包括 DNA 序列中基因的组织与控制、确定 DNA 中的转录单位、从序列预测蛋白质结构以及分子功能分析。

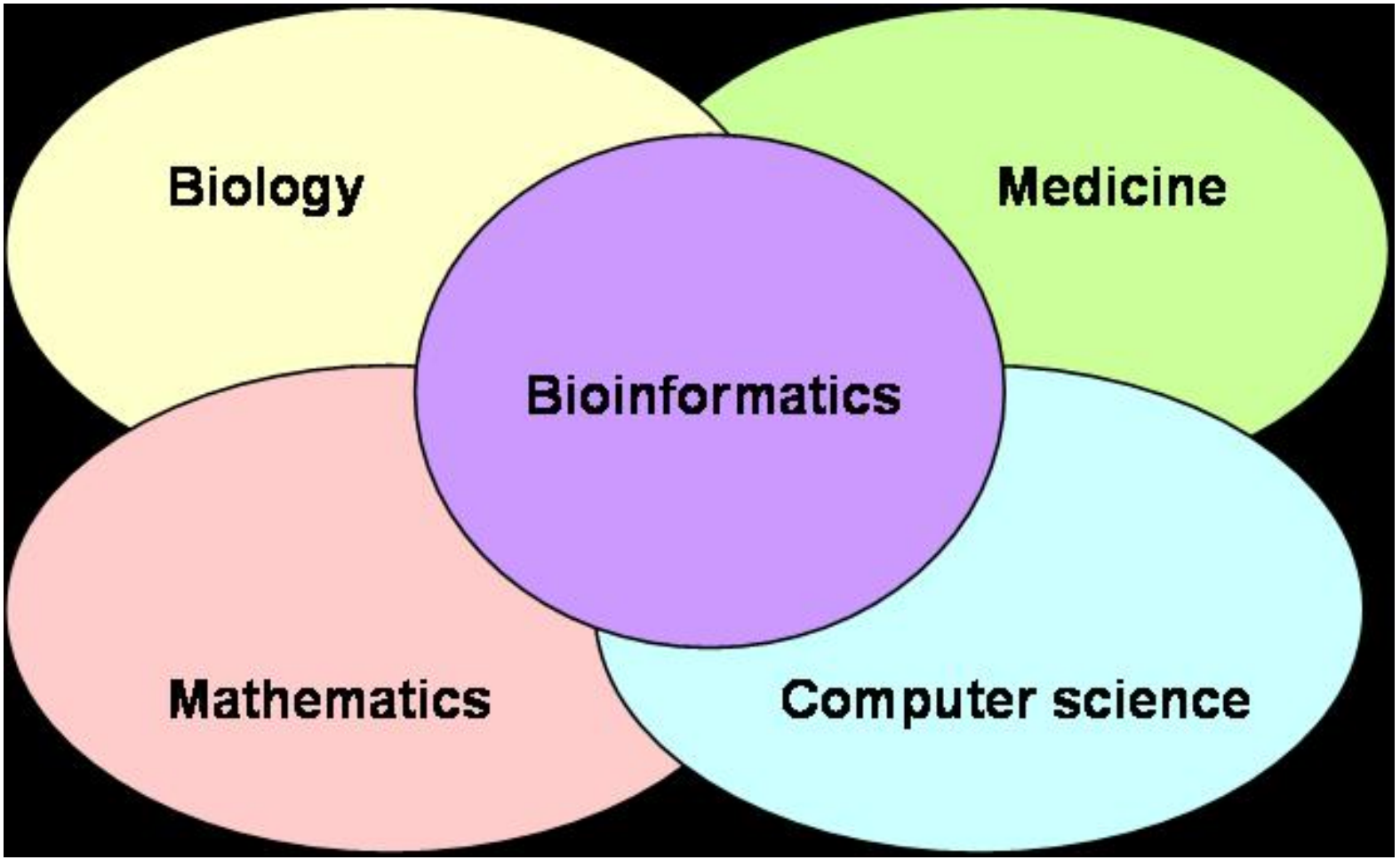
第二类信息流是基于科学方法：提出关于生物学活动的假设，设计实验以验证这些假设，评估结果与假设的相容性，然后根据实验数据对原假设作扩展或修正。第二类的生物信息学应用关注于这一流程中的信息传递，包括产生假设、设计实验、通过数据库将实验结果组织起来、检验数据与模型的相容性以及修正假设的各个系统。

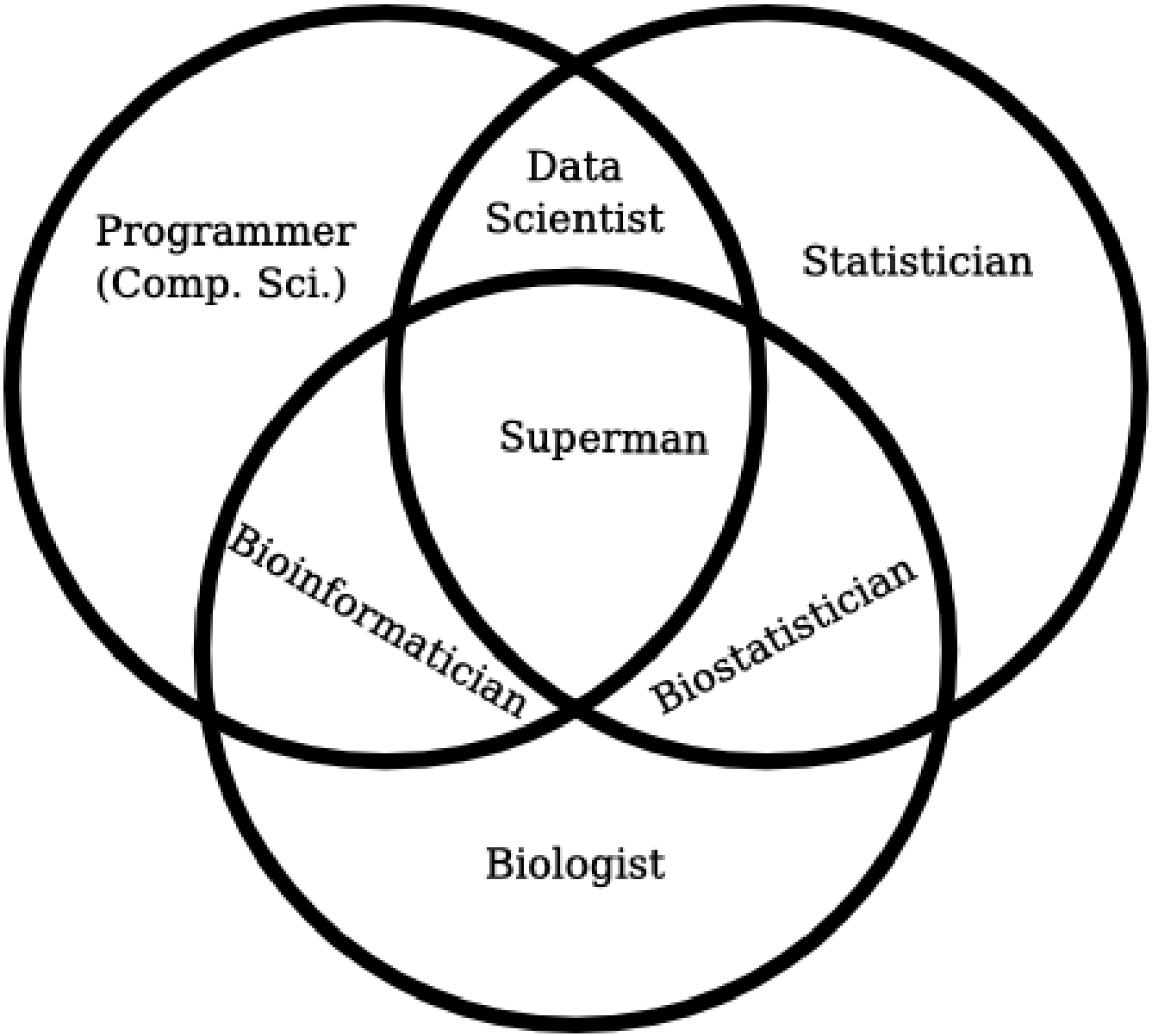
生物信息学 | 是什么？| 学科角度

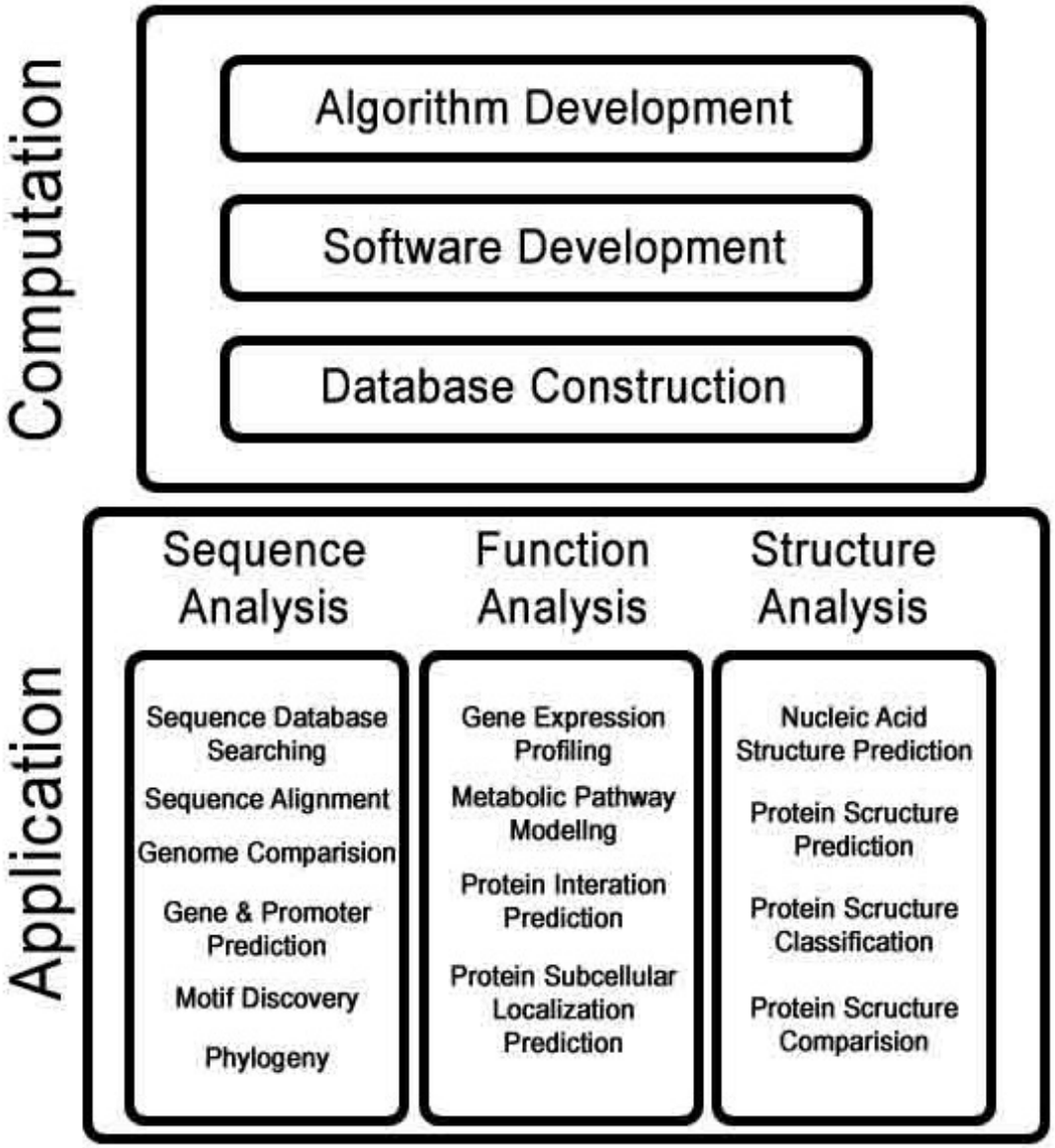


生物信息学 | 是什么？| 学科角度




Biology
Medicine
Bioinformatics
Mathematics
Computer science

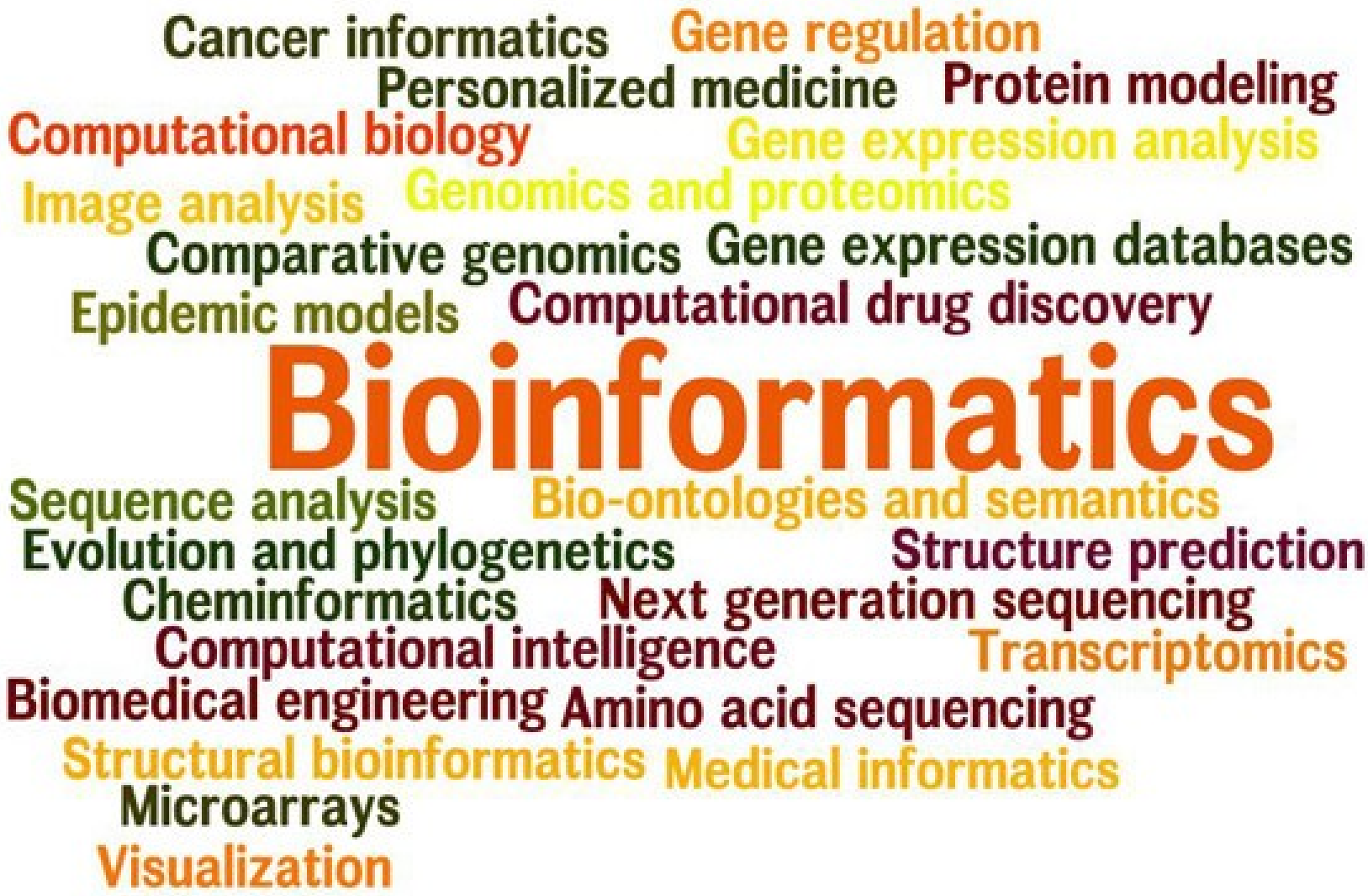

Programmer
(Comp. Sci.)
Data
Scientist
Statistician
Superman
Bioinformatician
Biostatistician
Biologist


Computation
Algorithm Development
Software Development
Database Construction
Application
Sequence Analysis
Function Analysis
Structure Analysis
Sequence Database Searching
Sequence Alignment
Genome Comparision
Gene & Promoter Prediction
Motif Discovery
Phylogeny
Gene Expression Profiling
Metabolic Pathway Modelling
Protein Interation Prediction
Protein Subcellular Localization Prediction
Nucleic Acid Structure Prediction
Protein Scructure Prediction
Protein Scructure Classification
Protein Scructure Comparision

生物信息学 | 资源（数据库与工具）

生物信息学 | 资源（技术与工具）

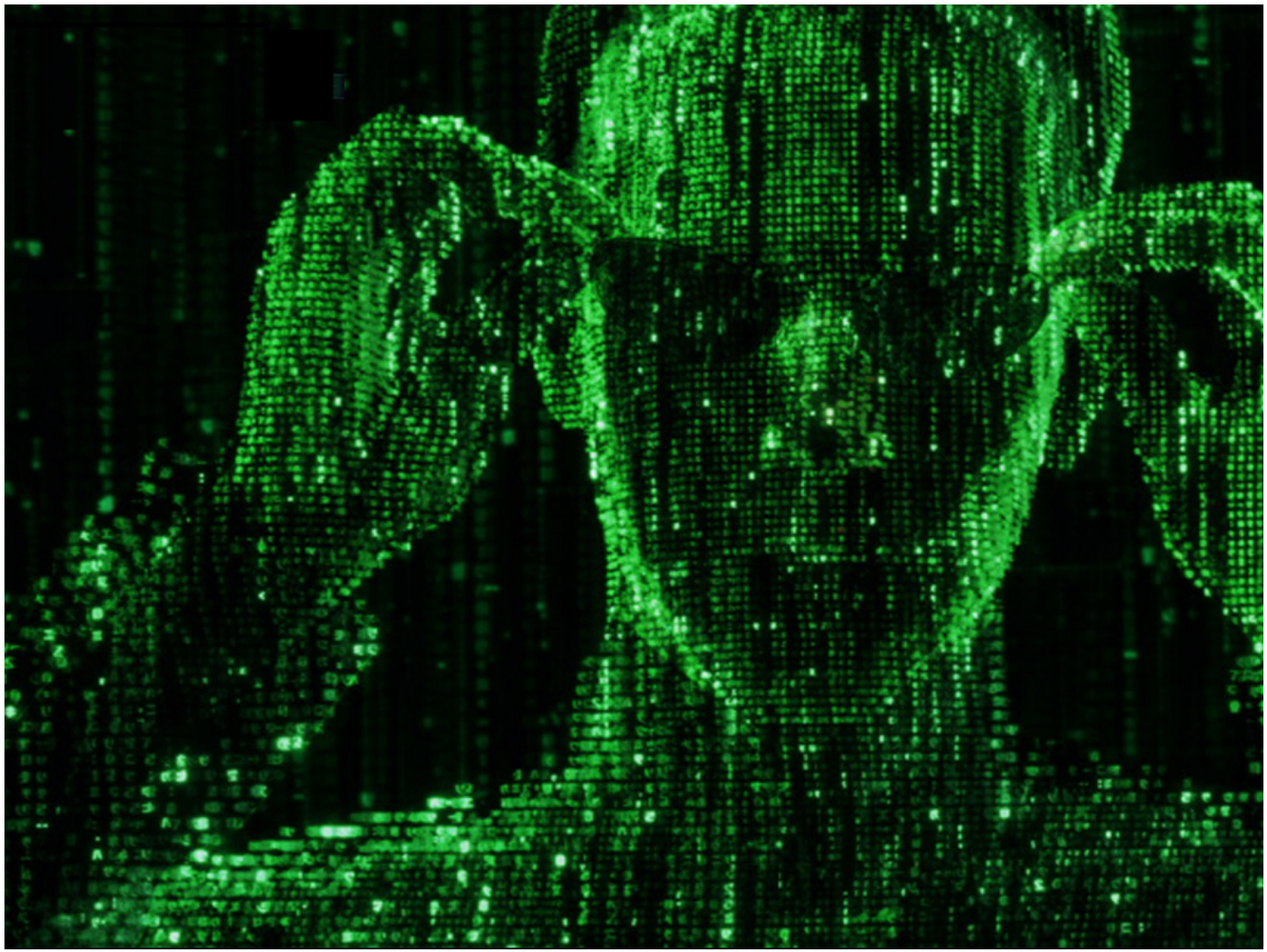
生物信息学 | 主题词


Cancer informatics
Gene regulation
Personalized medicine
Protein modeling
Computational biology
Gene expression analysis
Image analysis
Genomics and proteomics
Comparative genomics
Gene expression databases
Epidemic models
Computational drug discovery
Bioinformatics
Sequence analysis
Bio-ontologies and semantics
Evolution and phylogenetics
Structure prediction
Cheminformatics
Next generation sequencing
Computational intelligence
Transcriptomics
Biomedical engineering
Amino acid sequencing
Structural bioinformatics
Medical informatics
Microarrays
Visualization

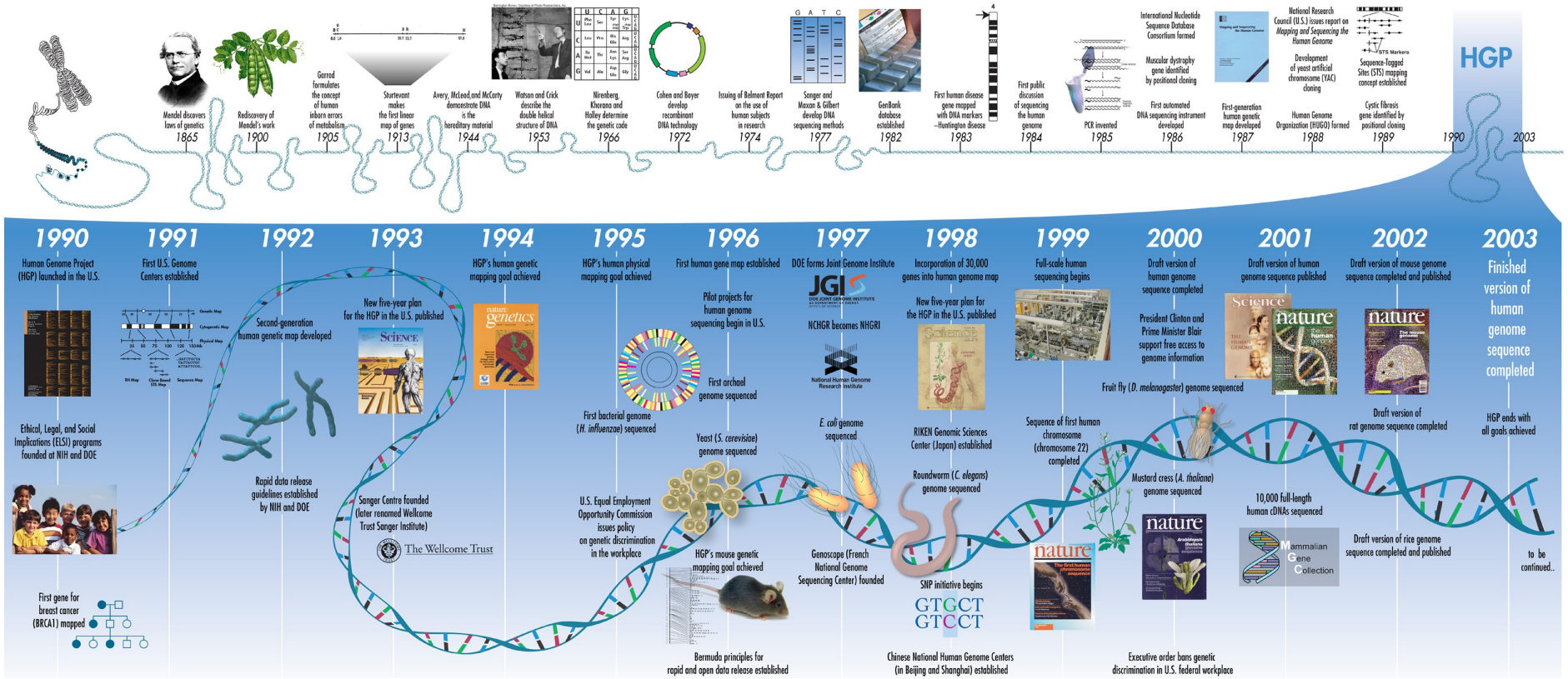
生物信息学 | 主题词

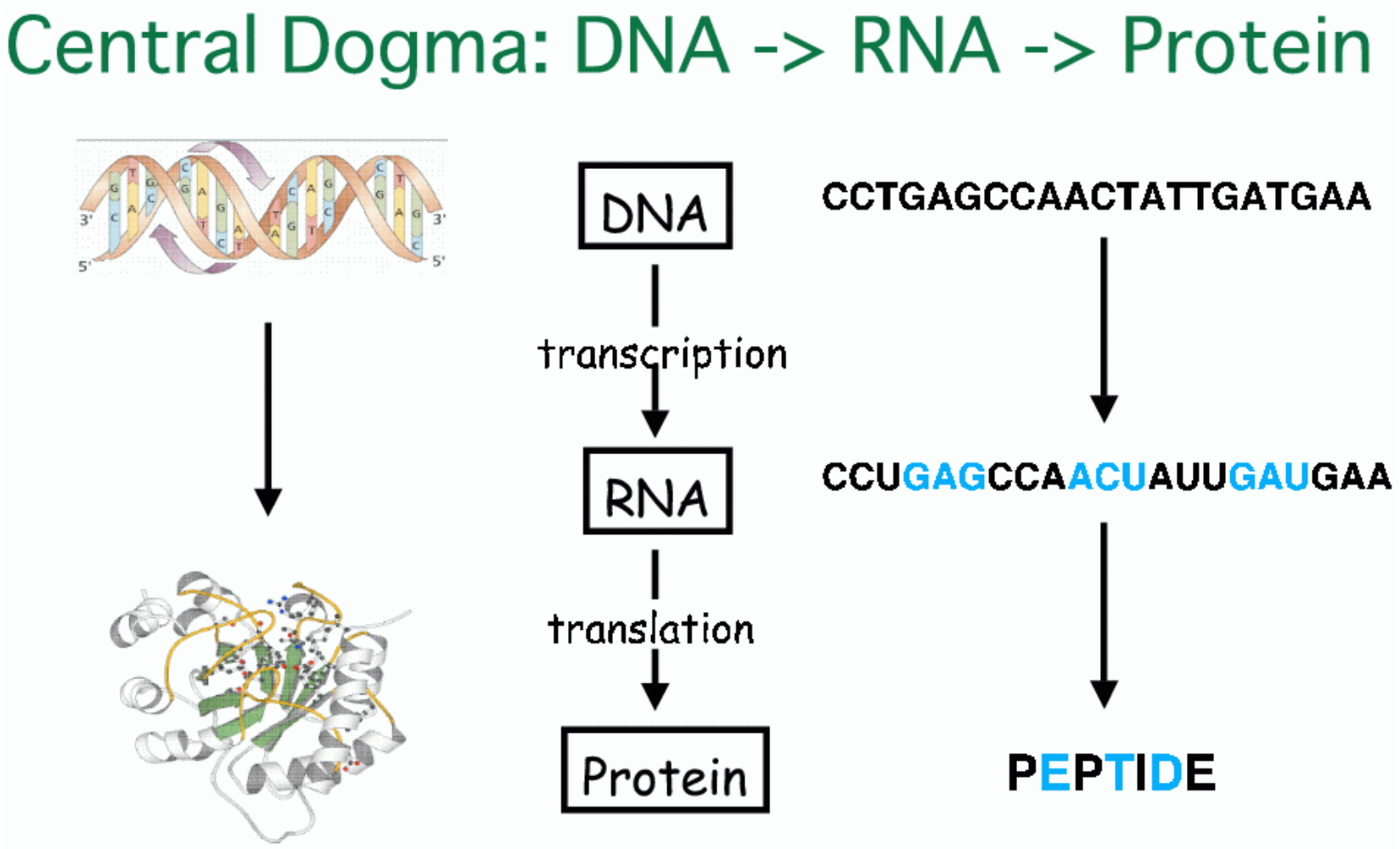

conducted
post-processing
GEO
Structural
field
provide
databases
evaluation
available
Results
GO-term
search
robust
matrix
structure
proteins
gene
developed
TRISUM-169
substitution
similar
linear
score
similarities
background
improvements
describe
network
predictive
GO
relationships
proteomic
expression
Gene
significant
method
highly
peptide
time
cross-validation
terms
true-path
model
computationally
Geneva
based
protein
biological
previous
times
protein-protein
form
regression
3D
substantial
structures
alphabet
used
using
positive
sets
1D
changes
network-based
design
EGF
wanted
i.e
prediction
retention
database
SA-FAST
algorithm
can
rules
sequence
sequences
approach
use
performance
task
shotgun
pairs
implementation
results
hierarchical
alphabets
analysis
local
standard
Analysis
Conclusions
structural
framework
several
tools
data
searches
found
biologically
interaction
predictions
positives
genes
among
identifications
functional
function
advantages
introduce
within
experiments
automated
clusters
successive
FASTA
false
also
information
applied
consistency
number
matches
annotation
alignment
methods
related
increased

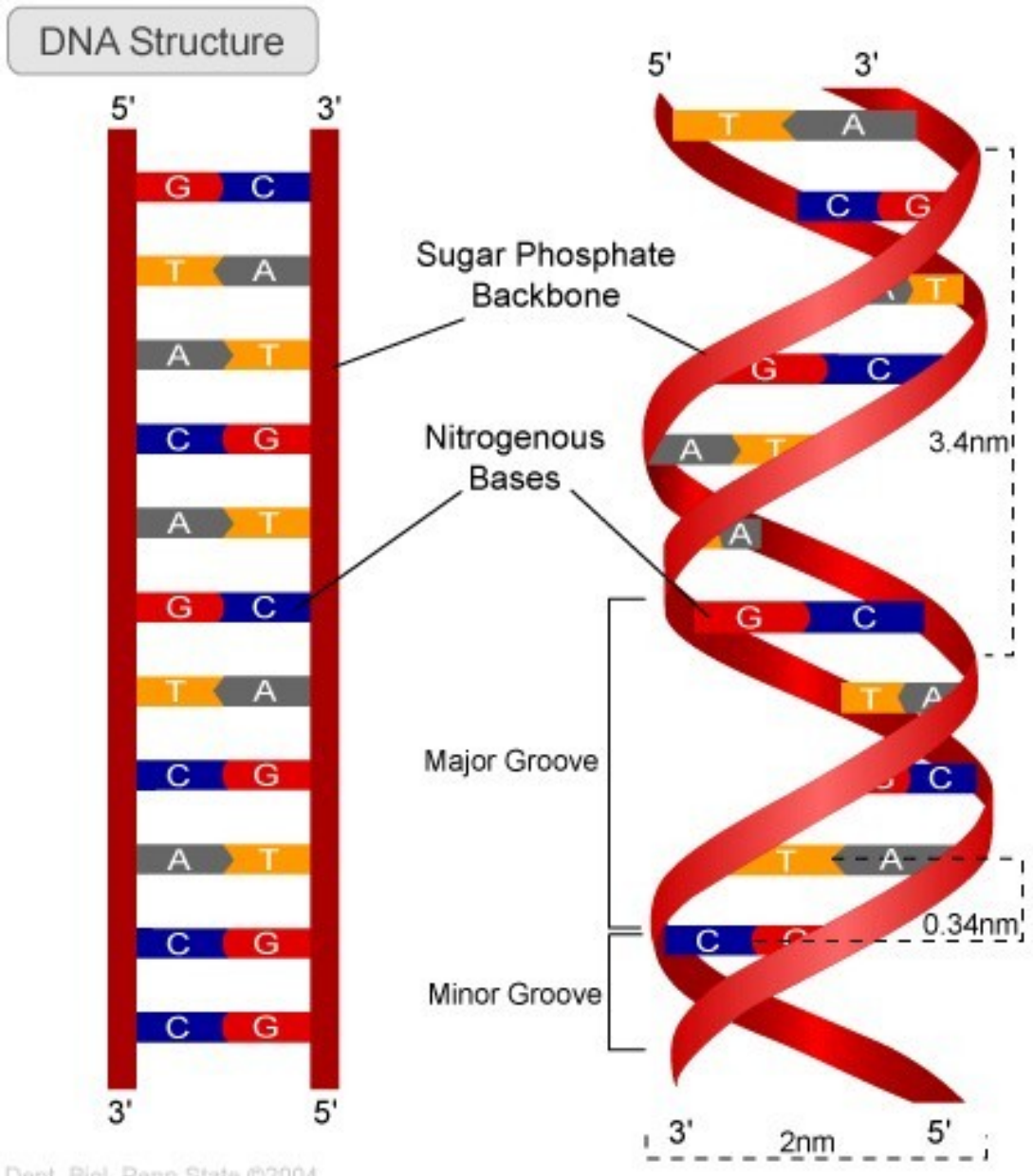
生物信息学家 = 生命科学中的黑客！

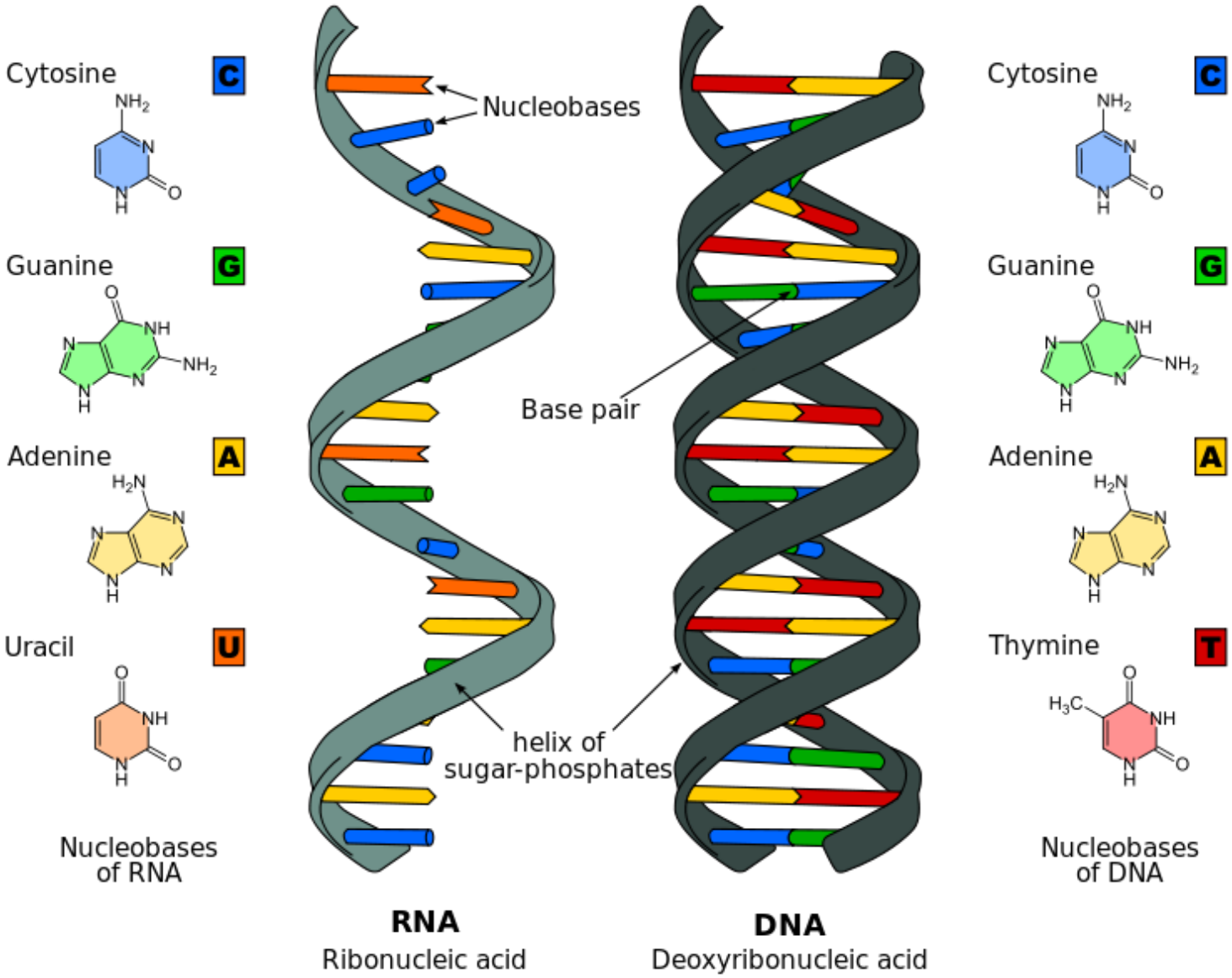


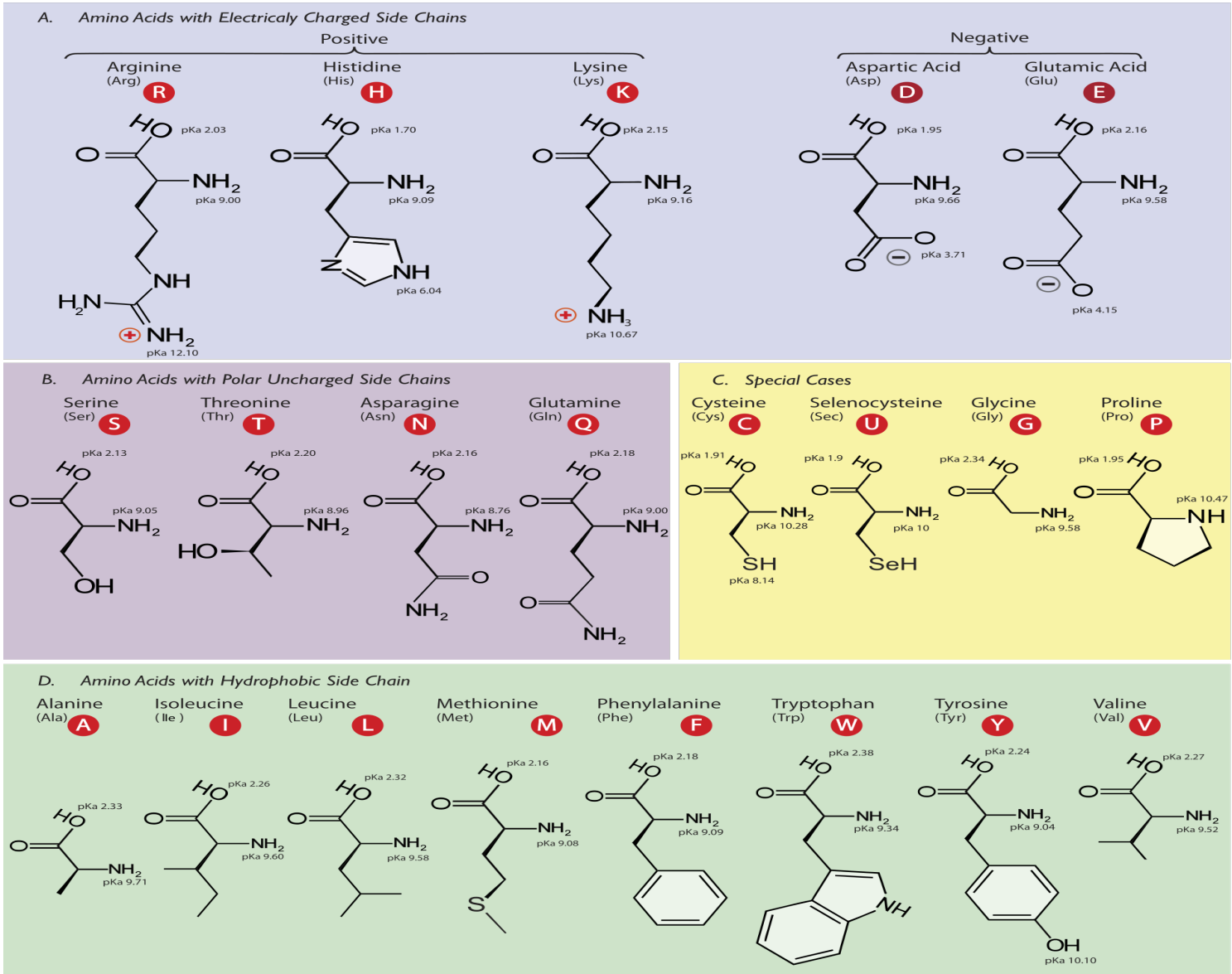
教学提纲


HGP
1865 Mendel discovers laws of genetics
1900 Rediscovery of Mendel's work
1905 Garrod formulates the concept of human inborn errors of metabolism
1913 Sturtevant makes the first linear map of genes
1944 Avery, McLeod, and McCarty demonstrate DNA is the hereditary material
1953 Watson and Crick describe the double helical structure of DNA
1966 Nirenberg, Khorana and Holley determine the genetic code
1972 Cohen and Boyer develop recombinant DNA technology
1974 Issuing of Belmont Report on the use of human subjects in research
1977 Sanger and Maxam & Gilbert develop DNA sequencing methods
1982 GenBank database established
1983 First human disease gene mapped with DNA markers --Huntington disease
1984 First public discussion of sequencing the human genome
1985 PCR invented
1986 International Nucleotide Sequence Database Consortium formed
Muscular dystrophy gene identified by positional cloning
First automated DNA sequencing instrument developed
1987 First-generation human genetic map developed
1988 National Research Council (U.S.) issues report on Mapping and Sequencing the Human Genome
Development of yeast artificial chromosome (YAC) cloning
Human Genome Organization (HUGO) formed
1989 STS Markers
Sequence-Tagged Sites (STS) mapping concept established
Cystic fibrosis gene identified by positional cloning
1990
2003
1990 Human Genome Project (HGP) launched in the U.S.
Ethical, Legal, and Social Implications (ELSI) programs founded at NIH and DOE
First gene for breast cancer (BRCA1) mapped
1991 First U.S. Genome Centers established
Genetic Map
Cytogenetic Map
Physical Map
Clone-Based STS Map
Sequence Map
1992 Second-generation human genetic map developed
Rapid data release guidelines established by NIH and DOE
1993 New five-year plan for the HGP in the U.S. published
SCIENCE
Sanger Centre founded (later renamed Wellcome Trust Sanger Institute)
The Wellcome Trust
1994 HGP's human genetic mapping goal achieved
nature genetics
1995 HGP's human physical mapping goal achieved
First bacterial genome (H. influenzae) sequenced
U.S. Equal Employment Opportunity Commission issues policy on genetic discrimination in the workplace
1996 First human gene map established
Pilot projects for human genome sequencing begin in U.S.
First archael genome sequenced
Yeast (S. cerevisiae) genome sequenced
HGP's mouse genetic mapping goal achieved
Bermuda principles for rapid and open data release established
1997 DOE forms Joint Genome Institute
JGI
DOE JOINT GENOME INSTITUTE
NCHGR becomes NHGRI
National Human Genome Research Institute
E. coli genome sequenced
Genoscope (French National Genome Sequencing Center) founded
1998 Incorporation of 30,000 genes into human genome map
Now five-year plan for the HGP in the U.S. published
Science
RIKEN Genomic Sciences Center (Japan) established
Roundworm (C. elegans) genome sequenced
SNP initiative begins
GTGCT
GTCCT
Chinese National Human Genome Centers (in Beijing and Shanghai) established
1999 Full-scale human sequencing begins
Sequence of first human chromosome (chromosome 22) completed
nature
2000 Draft version of human genome sequence completed
President Clinton and Prime Minister Blair support free access to genome information
Fruit fly (D. melanogaster) genome sequenced
Mustard cress (A. thaliana) genome sequenced
nature
Executive order bans genetic discrimination in U.S. federal workplace
2001 Draft version of human genome sequence published
Science
THE HUMAN GENOME
nature
the human genome
10,000 full-length human cDNAs sequenced
Mammalian Gene Collection
2002 Draft version of mouse genome sequence completed and published
nature
the mouse genome
Draft version of rat genome sequence completed
Draft version of rice genome sequence completed and published
2003 Finished version of human genome sequence completed
HGP ends with all goals achieved
to be continued..


Central Dogma: DNA -> RNA -> Protein
DNA
transcription
RNA
translation
Protein
CCTGAGCCAACTATTGATGAA
CCUGAGCCAACUAUUGAUGAA
PEPTIDE


DNA Structure
5'
3'
G C
T A
A T
C G
A T
G C
T A
C G
A T
C G
C G
3'
5'
Sugar Phosphate Backbone
Nitrogenous Bases
Major Groove
Minor Groove
3.4nm
0.34nm
2nm


Cytosine
C
Guanine
G
Adenine
A
Uracil
U
Nucleobases of RNA
Nucleobases
Base pair
helix of sugar-phosphates
RNA
Ribonucleic acid
DNA
Deoxyribonucleic acid
Cytosine
C
Guanine
G
Adenine
A
Thymine
T
Nucleobases of DNA


A. Amino Acids with Electricaly Charged Side Chains
Positive
Arginine (Arg) R
Histidine (His) H
Lysine (Lys) K
Negative
Aspartic Acid (Asp) D
Glutamic Acid (Glu) E
B. Amino Acids with Polar Uncharged Side Chains
Serine (Ser) S
Threonine (Thr) T
Asparagine (Asn) N
Glutamine (Gln) Q
C. Special Cases
Cysteine (Cys) C
Selenocysteine (Sec) U
Glycine (Gly) G
Proline (Pro) P
D. Amino Acids with Hydrophobic Side Chain
Alanine (Ala) A
Isoleucine (Ile) I
Leucine (Leu) L
Methionine (Met) M
Phenylalanine (Phe) F
Tryptophan (Trp) W
Tyrosine (Tyr) Y
Valine (Val) V

必需氨基酸

甲携来一本亮色书（甲硫氨酸、缬氨酸、赖氨酸、异亮氨酸、苯丙氨酸、亮氨酸、色氨酸、苏氨酸）

20 种氨基酸

苏缬亮异亮，苯丙属芳香。
还有色赖蛋，缺一人遭殃。
（以上是必需氨基酸）
丙组丝甘半，天谷建酸胺。
精酪加一脯，20 氨基酸。
（以上是非必需氨基酸）

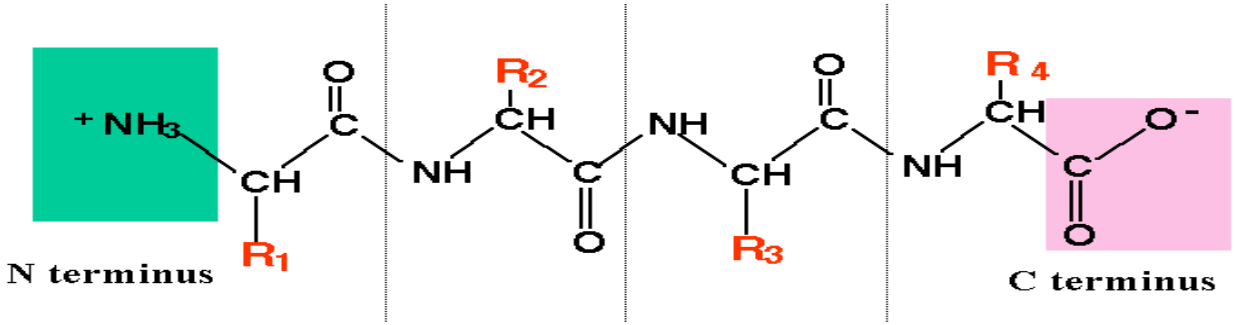
必需氨基酸

甲携来一本亮色书（甲硫氨酸、缬氨酸、赖氨酸、异亮氨酸、苯丙氨酸、亮氨酸、色氨酸、苏氨酸）

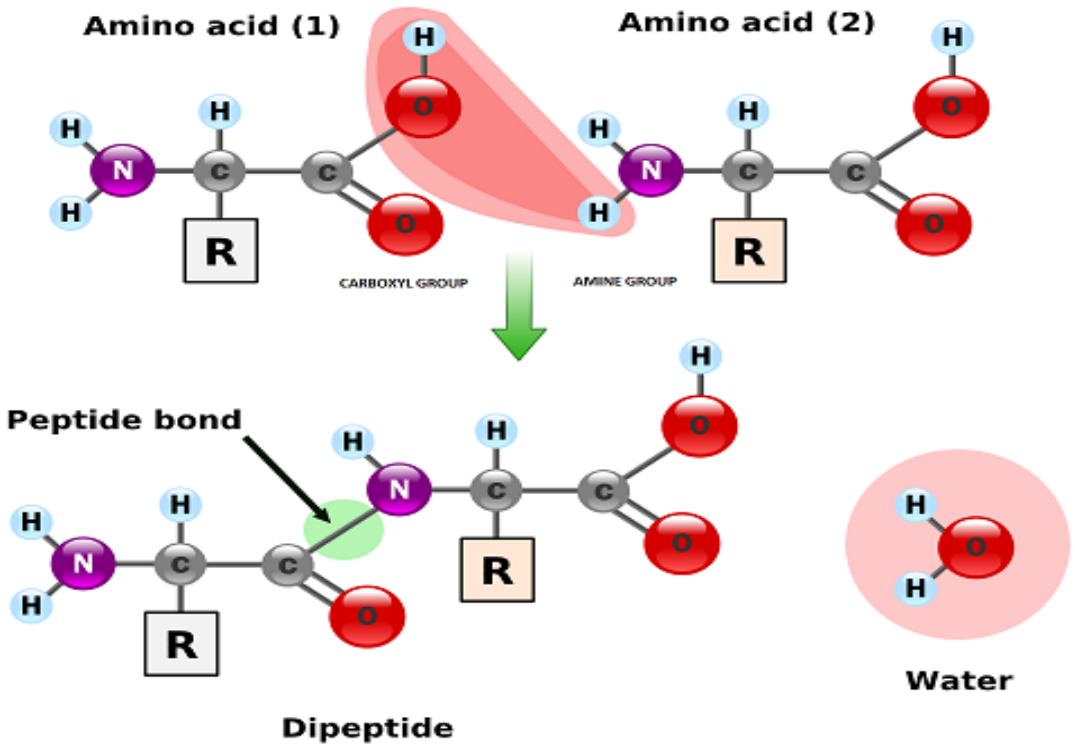
20 种氨基酸

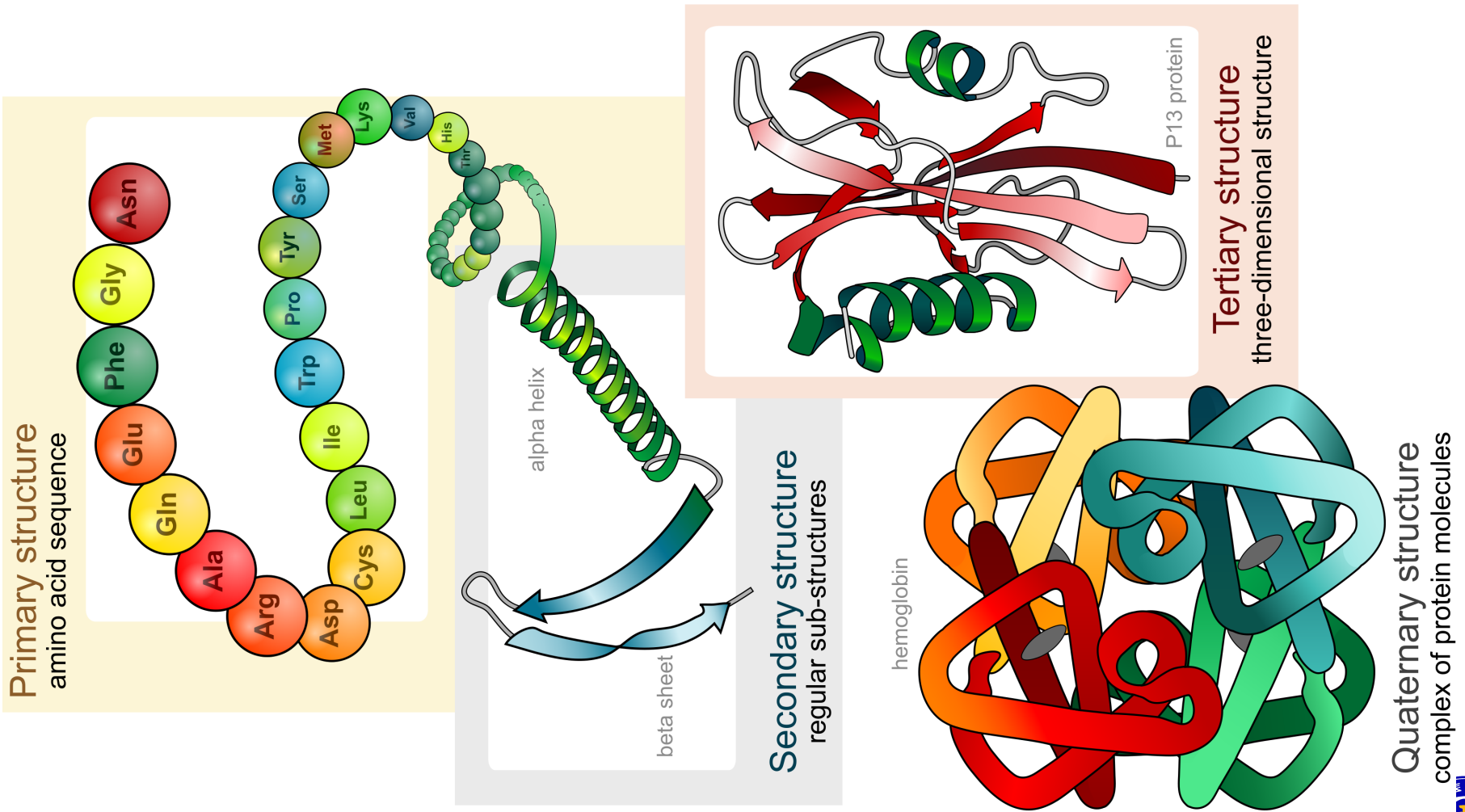
苏缬亮异亮，苯丙属芳香。
还有色赖蛋，缺一人遭殃。
(以上是必需氨基酸)
丙组丝甘半，天谷建酸胺。
精酪加一脯，20 氨基酸。
(以上是非必需氨基酸)

Peptide = chain of amino acids



polypeptide chain

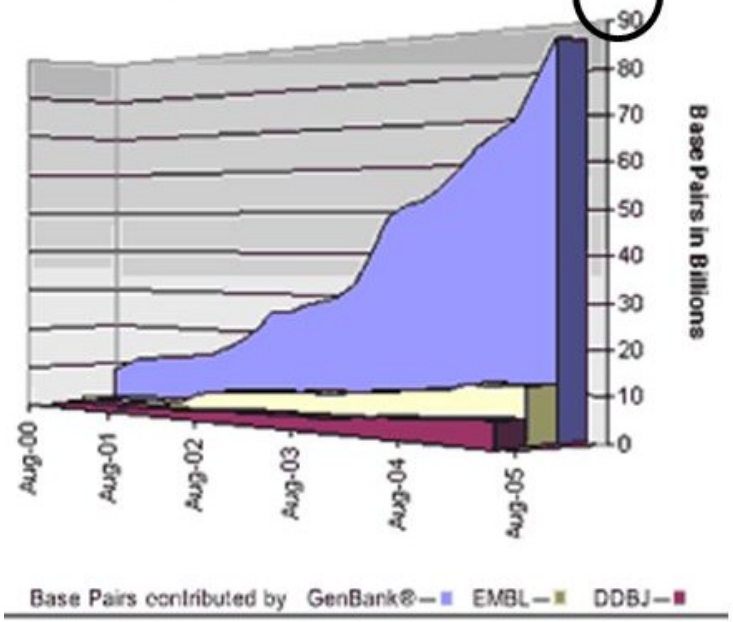



Primary structure
amino acid sequence
Asn
Gly
Phe
Glu
Gln
Ala
Arg
Asp
Cys
Leu
Ile
Trp
Pro
Tyr
Ser
Met
Lys
Val
His
Thr
Secondary structure
regular sub-structures
alpha helix
beta sheet
Tertiary structure
three-dimensional structure
P13 protein
Quaternary structure
complex of protein molecules
hemoglobin

100 Billion Bases!

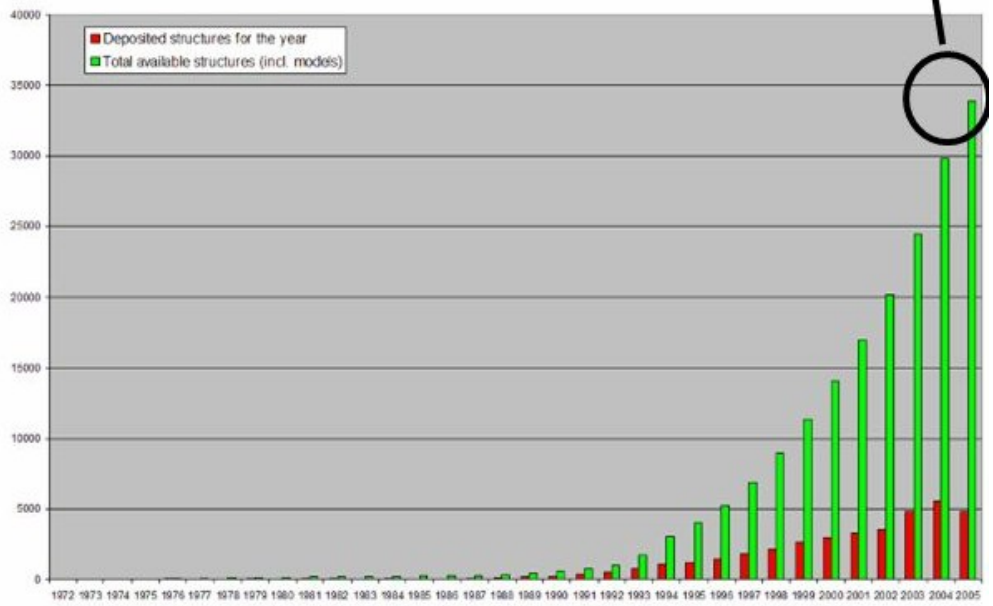
35,000 Structures

GenBank

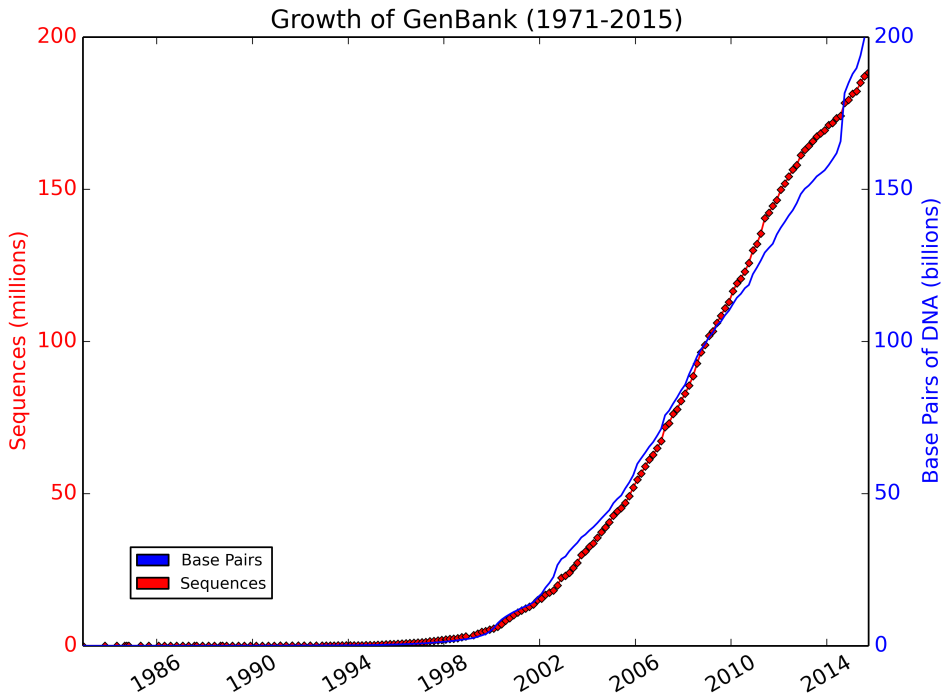


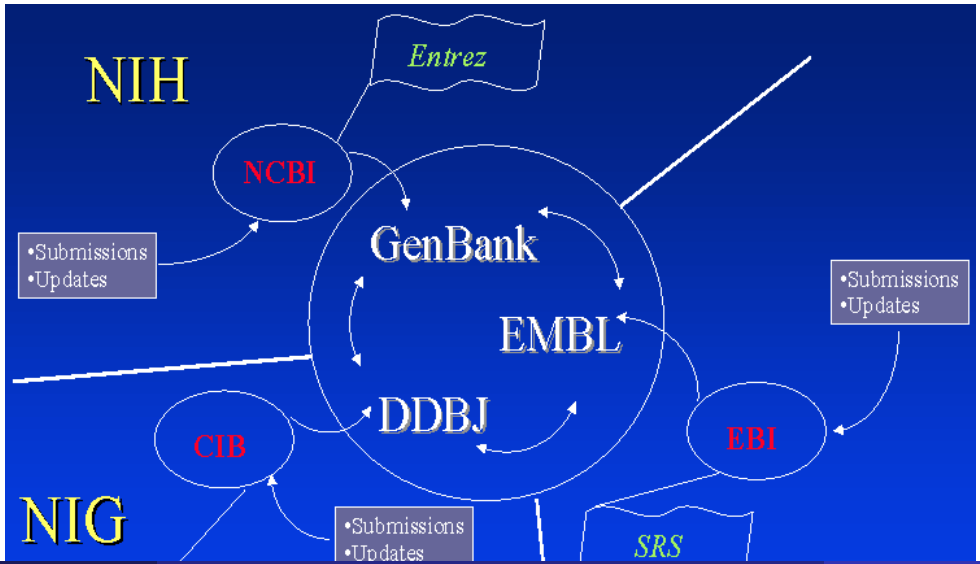
www.ncbi.nlm.nih.gov/Genbank

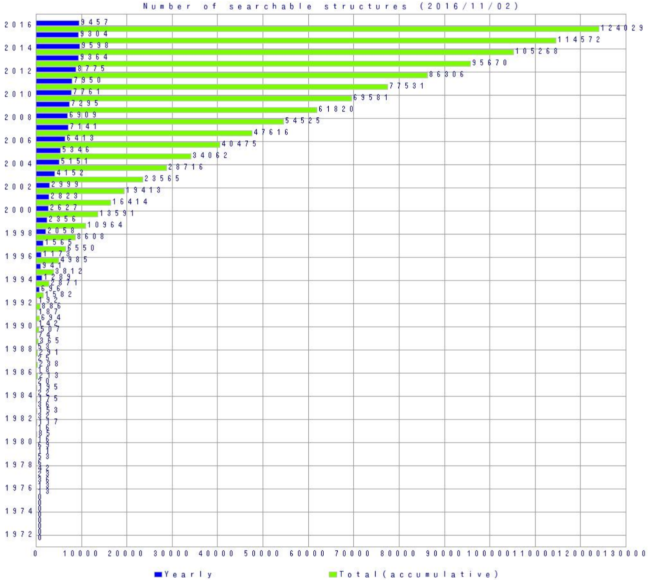
Protein Data Bank



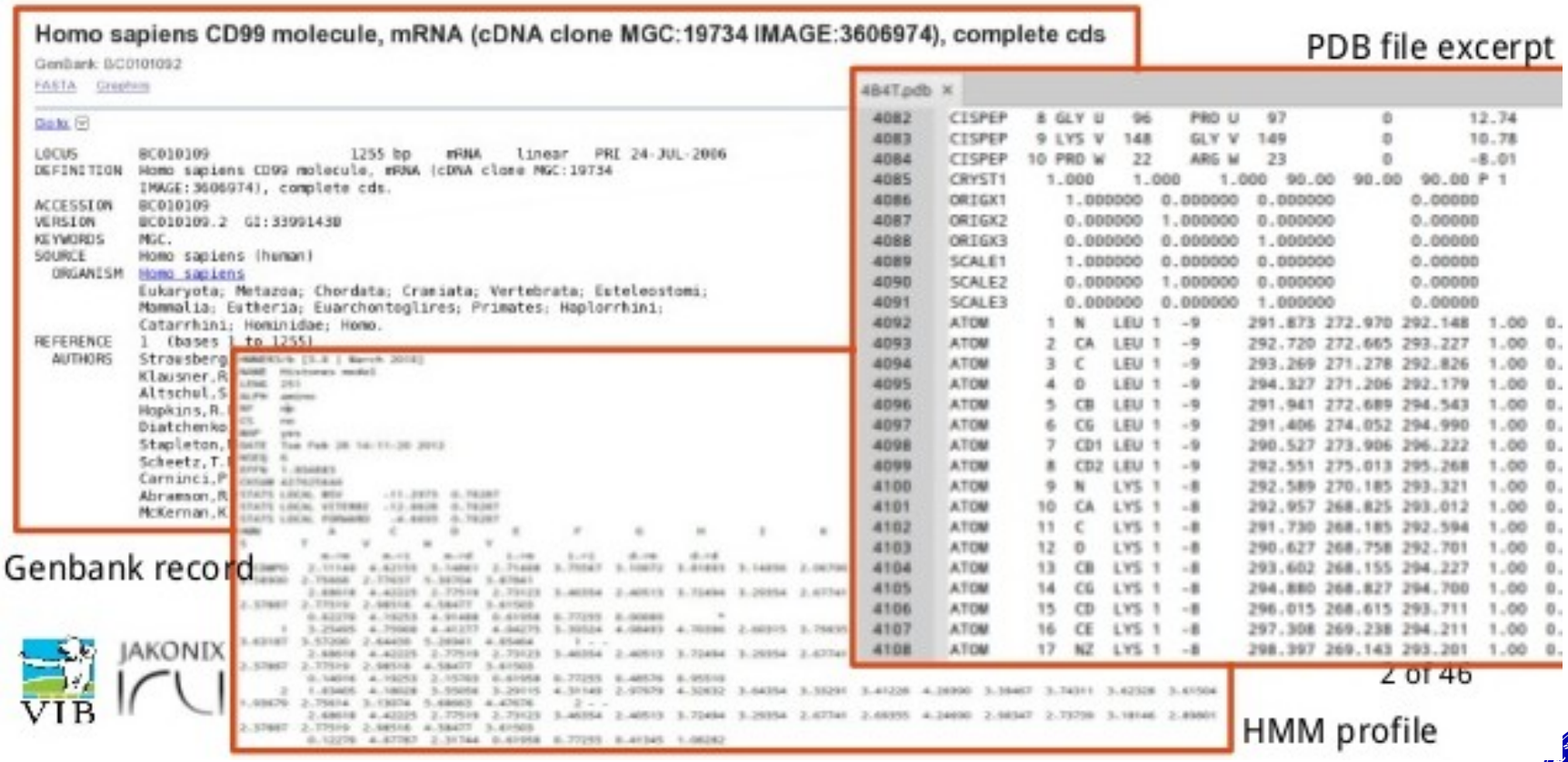
www.rcsb.org/pdb/holdings.html


Growth of GenBank (1971-2015)
Sequences (millions)
Base Pairs of DNA (billions)
Base Pairs
Sequences
0
50
100
150
200
1986
1990
1994
1998
2002
2006
2010
2014


NIH
Entrez
NCBI
•Submissions
•Updates
GenBank
EMBL
DDBJ
CIB
NIG
•Submissions
•Updates
EBI
•Submissions
•Updates
SRS

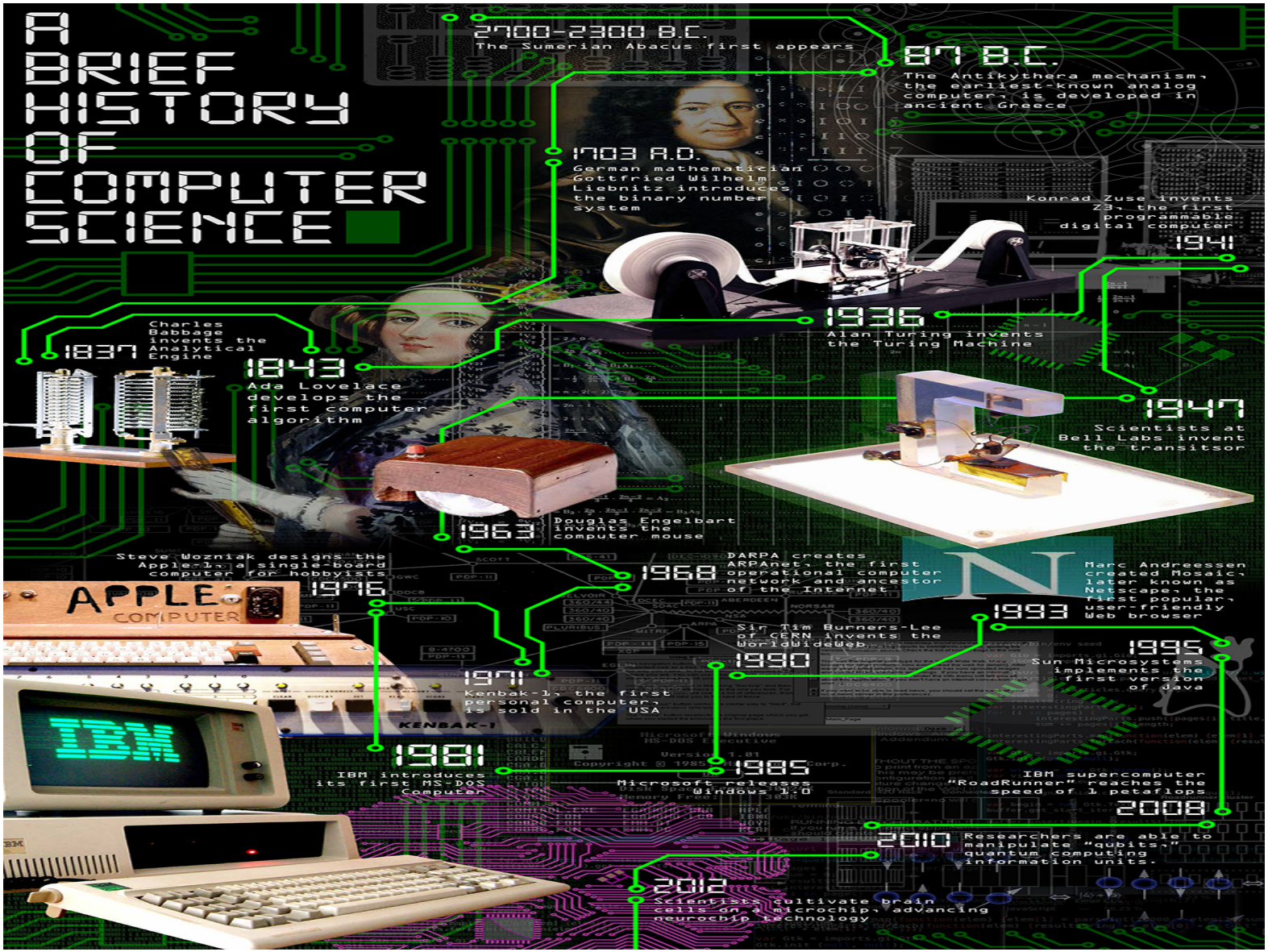

Number of searchable structures (2016/11/02)
2016
2014
2012
2010
2008
2006
2004
2002
2000
1998
1996
1994
1992
1990
1988
1986
1984
1982
1980
1978
1976
1974
1972
9457
124029
9304
114572
9598
105268
9364
95670
8775
86306
7950
77531
7761
69581
7295
61820
6909
54525
7141
47616
6413
40475
5346
34062
5151
28716
4152
23565
2999
19413
2823
16414
2627
13591
2356
10964
2058
8608
1565
6550
1173
4985
941
3812
1289
2871
696
1582
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
110000
120000
130000
Yearly
Total(accumulative)

Historically, bioinformatics has always used text files to store data.



教学提纲

计算机科学 | 历史


A BRIEF HISTORY OF COMPUTER SCIENCE
2700-2300 B.C.
The Sumerian Abacus first appears
87 B.C.
The Antikythera mechanism, the earliest-known analog computer, is developed in ancient Greece
1703 A.D.
German mathematician Gottfried Wilhelm Liebnitz introduces the binary number system
Konrad Zuse invents Z3, the first programmable digital computer
1941
1936
Alan Turing invents the Turing Machine
1837
Charles Babbage invents the Analytical Engine
1843
Ada Lovelace develops the first computer algorithm
1947
Scientists at Bell Labs invent the transistor
1963
Douglas Engelbart invents the computer mouse
Steve Wozniak designs the Apple-1, a single-board computer for hobbyists
1976
APPLE COMPUTER
1968
DARPA creates ARPAnet, the first operational computer network and ancestor of the Internet
Marc Andreessen created Mosaic, later known as Netscape, the first popular, user-friendly Web browser
1993
1990
Sir Tim Burners-Lee of CERN invents the WorldWideWeb
1995
Sun Microsystems implements the first version of Java
1971
Kenbak-1, the first personal computer, is sold in the USA
KENBAK-1
1981
IBM introduces its first MS-DOS computer
IBM
1985
Microsoft releases Windows 1.0
IBM supercomputer "RoadRunner" reaches the speed of 1 petaflops
2008
2010
Researchers are able to manipulate "qubits," quantum computing information units.
2012
Scientists cultivate brain cells on a microchip, advancing neurochip technology

in vivo

In vivo 为拉丁文“在活体内”之意。在科学文献中，*in vivo* 常指进行于完整且存活的个体内的组织的实验，以区别在生物体上移除下来的组织或死亡的组织上进行的实验（对应的拉丁文为 *in vitro* ）。

in vitro

In vitro 是拉丁语中“在玻璃里”的意思，意指进行或发生于试管内的实验与实验技术。更广义的意思，则指活生物体之外的环境中的操作。

in silico

In silico 是指“在硅之中”，也就是说“进行于电脑中，或是经由电脑模拟”之意，此用语是衍生自另外两个在生物学上常用的短语：*in vivo*（生物活体内）及 *in vitro*（生物活体外）。

in vivo

In vivo 为拉丁文“在活体内”之意。在科学文献中，*in vivo* 常指进行于完整且存活的个体内的组织的实验，以区别在生物体上移除下来的组织或死亡的组织上进行的实验（对应的拉丁文为 *in vitro* ）。

in vitro

In vitro 是拉丁语中“在玻璃里”的意思，意指进行或发生于试管内的实验与实验技术。更广义的意思，则指活生物体之外的环境中的操作。

in silico

In silico 是指“在硅之中”，也就是说“进行于电脑中，或是经由电脑模拟”之意，此用语是衍生自另外两个在生物学上常用的短语：*in vivo*（生物活体内）及 *in vitro*（生物活体外）。

in vivo

In vivo 为拉丁文“在活体内”之意。在科学文献中，*in vivo* 常指进行于完整且存活的个体内的组织的实验，以区别在生物体上移除下来的组织或死亡的组织上进行的实验（对应的拉丁文为 *in vitro* ）。

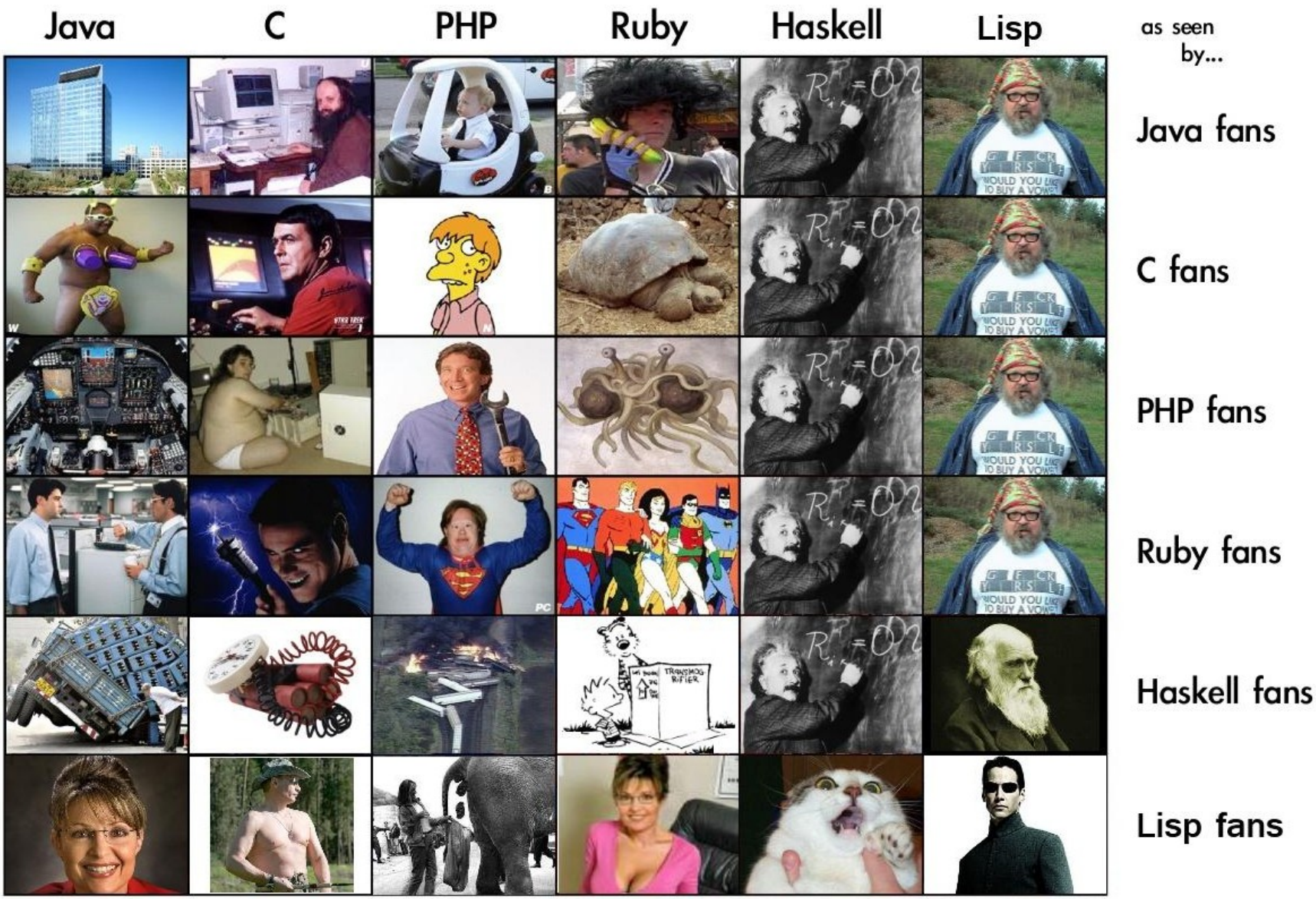
in vitro

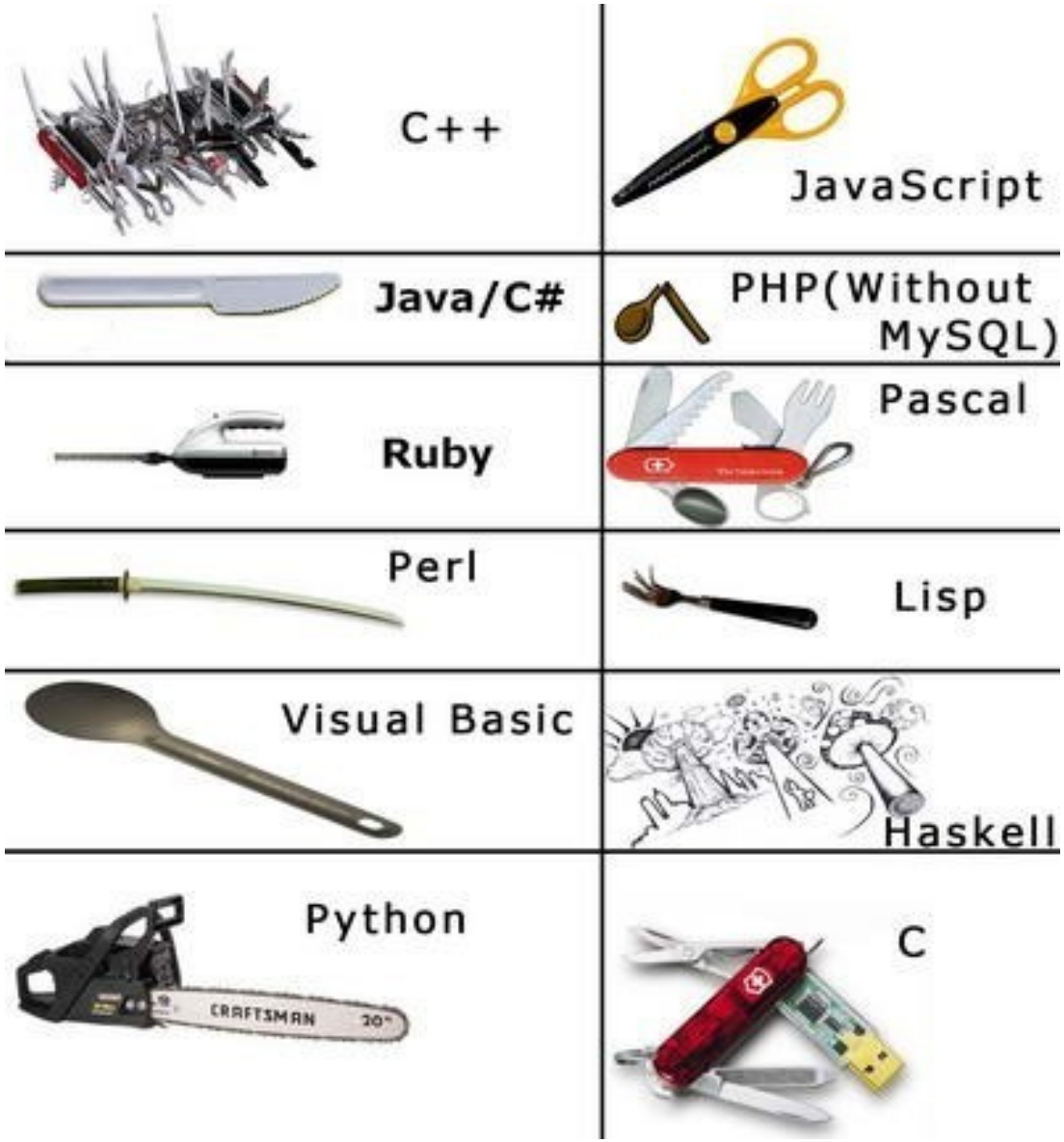
In vitro 是拉丁语中“在玻璃里”的意思，意指进行或发生于试管内的实验与实验技术。更广义的意思，则指活生物体之外的环境中的操作。

in silico

In silico 是指“在硅之中”，也就是说“进行于电脑中，或是经由电脑模拟”之意，此用语是衍生自另外两个在生物学上常用的短语：*in vivo*（生物活体内）及 *in vitro*（生物活体外）。


JavaScript
Perl
.NET
PHP
XML
Markup
SQL
C++
Pascal
Scala
Functional
VisualBasic
DataStructured
Erlang
Declarative
Ruby
Scripting
Ada
Squeak
Multiparadigm
Logic-based
LabVIEW
Python
MATLAB
Cobol
Prototype-based
Java
Clojure
Assembly
Object-Oriented
Scheme
Haskell
Logo
Distributed
Imperative
Interpreted
BASIC
HyperCard
Procedural
PostScript
Fortran
Prolog
Reflective
HTML
Obfuscated
Lisp
Visual
C-sharp
Smalltalk
AppleScript
Occam
Delphi
Yorick
Simula
Miranda
Euphoria
Cocoa
Eiffel
Groovy
Oberon
Sather
CobolScript
Befunge
DOSBatch
Tcl-Tk
Database
ObjectiveCaml
BETA
Maple
S-Lang
Mumps
Verilog
CurlyBracket
LaTeX
Clean
Caml
Turing
UML
GLSL
ICI
Awk
ElasticC
Pike
Lua
Rexx
Jal
Intercal
Lagoona
Dataflow
ADL
Simkin
Moto
CLU
Cecil
RPG
Pliant
Goedel
Wirth
Fortress
Clarion
REBOL
TADS
Limbo
YAFL
VBScript
APL
SAS
Leda
Lisp

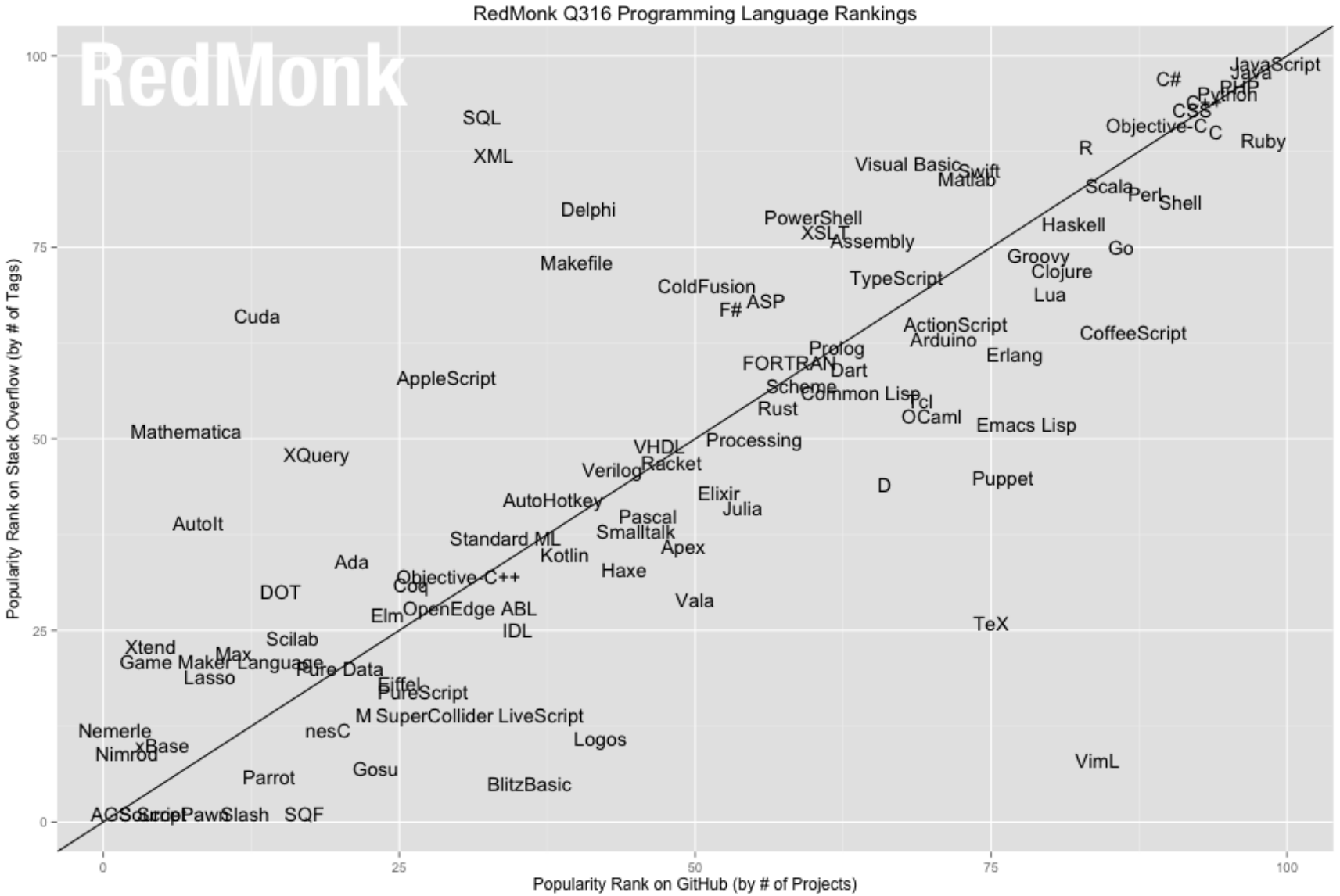

Java
C
PHP
Ruby
Haskell
Lisp
as seen by...
Java fans
C fans
PHP fans
Ruby fans
Haskell fans
Lisp fans


C++
JavaScript
Java/C#
PHP(Without MySQL)
Ruby
Pascal
Perl
Lisp
Visual Basic
Haskell
Python
C
CRAFTSMAN

语言	宗教	超级英雄	《哈利波特》
C	犹太教	—	伏地魔
C++	伊斯兰教	—	西弗勒斯·斯内普
Java	正统基督教	万磁王	洛雷斯·乌姆里奇
Lisp	佛教	Xavier 教授	—
Perl	巫毒教	—	罗恩·韦斯莱
PHP	Cafeteria 基督教	小丑王	德拉科·马尔福
Python	人文主义	蝙蝠侠	哈利·波特
Ruby	新异教主义	钢铁侠	—
shell	—	—	鲁伯·海格

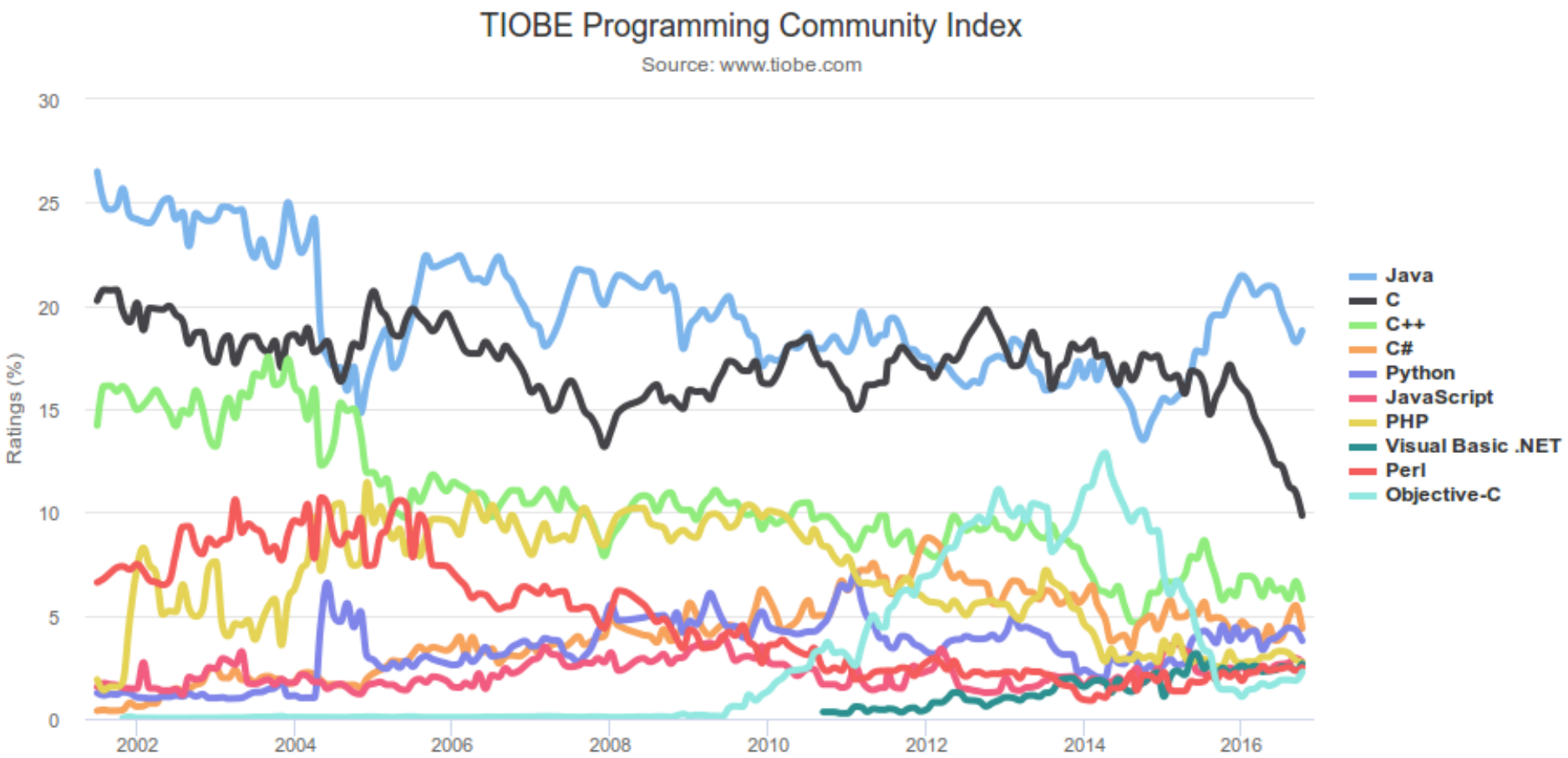
语言	女人	武器	船
C	霸道女总裁	M1 式加兰德步枪	核潜艇
C++	—	双截棍	—
Java	娇妻贤内助	M240 通用弹夹式自动机枪	大货船
Lisp	女博士	剃须刀	—
Perl	—	燃烧弹	拖船
PHP	—	水管子	竹筏
Python	万人迷	双管枪	—
Ruby	—	宝刀	摩托艇
shell	女公务员	锤子	—

计算机科学 | 编程语言 | 流行度



计算机科学 | 编程语言 | 排名 | 2016

Oct 2016	Oct 2015	Change	Programming Language	Ratings	Change
1	1		Java	18.799%	-0.74%
2	2		C	9.835%	-6.35%
3	3		C++	5.797%	+0.05%
4	4		C#	4.367%	-0.46%
5	5		Python	3.775%	-0.74%
6	8	⌃	JavaScript	2.751%	+0.46%
7	6	⌄	PHP	2.741%	+0.18%
8	7	⌄	Visual Basic .NET	2.660%	+0.20%
9	**9**		**Perl**	**2.495%**	**+0.25%**
10	14	⌃⌃	Objective-C	2.263%	+0.84%
11	12	⌃	Assembly language	2.232%	+0.66%
12	15	⌃	Swift	2.004%	+0.73%
13	10	⌄	Ruby	2.001%	+0.18%
14	13	⌄	Visual Basic	1.987%	+0.47%
15	11	⌄⌄	Delphi/Object Pascal	1.875%	+0.24%
16	65	⌃⌃	Go	1.809%	+1.67%
17	32	⌃⌃	Groovy	1.769%	+1.19%
18	20	⌃	R	1.741%	+0.75%
19	17	⌄	MATLAB	1.619%	+0.46%
20	18	⌄	PL/SQL	1.531%	+0.46%


TIOBE Programming Community Index
Source: www.tiobe.com
Ratings (%)
30
25
20
15
10
5
0
2002
2004
2006
2008
2010
2012
2014
2016
Java
C
C++
C#
Python
JavaScript
PHP
Visual Basic .NET
Perl
Objective-C

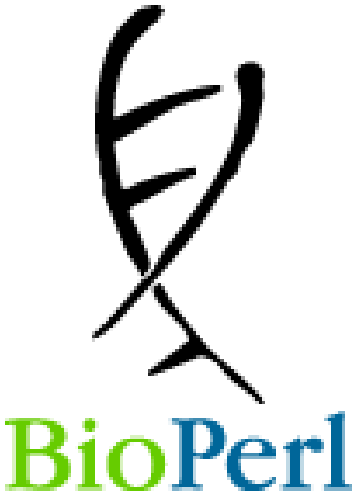
Programming Language	2016	2011	2006	2001	1996	1991	1986
Java	1	1	1	2	15	-	-
C	2	2	2	1	1	1	1
C++	3	3	3	3	2	2	5
C#	4	5	6	10	-	-	-
Python	5	6	7	23	24	-	-
PHP	6	4	4	8	-	-	-
JavaScript	7	9	8	7	19	-	-
Visual Basic .NET	8	28	-	-	-	-	-
Perl	9	8	5	4	3	-	-
Ruby	10	10	17	31	-	-	-
Lisp	27	12	12	15	7	4	3
Ada	28	16	15	16	6	5	2
Pascal	74	14	16	13	4	3	6

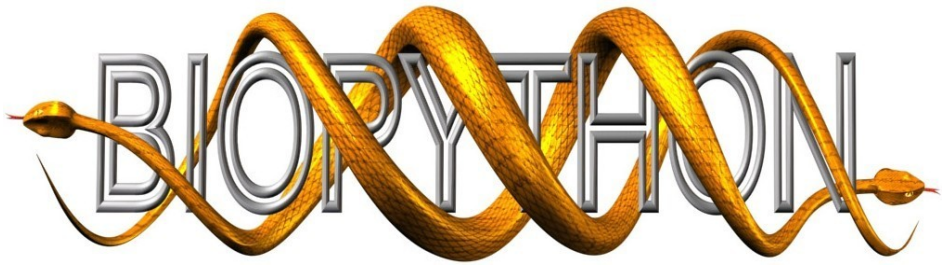
生物信息学常用编程语言

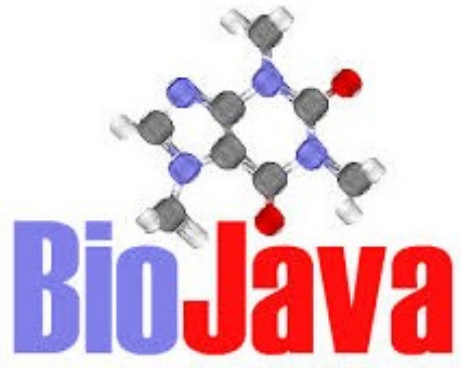
Perl（1987）、Python（1991）、Ruby（1995）、Java（1995）

生物信息学专用

BioPerl、Biopython、BioRuby、BioJava







BioSQL

BioSQL is a generic relational model covering sequences, features, sequence and feature annotation, a reference taxonomy, and ontologies (or controlled vocabularies).

Each Bio* project has a language binding (object-relational mapping, ORM) to BioSQL.

教学提纲

不需要！

- 有现成的工具可以使用
- 五湖四海皆兄弟
- 有钱能使磨推鬼
- ……

需要！

- 没有现成的工具
- 兄弟们都在忙着混江湖
- 工资涨如龟速，物价涨如赤兔
- ……

不需要！

- 有现成的工具可以使用
- 五湖四海皆兄弟
- 有钱能使磨推鬼
- ……

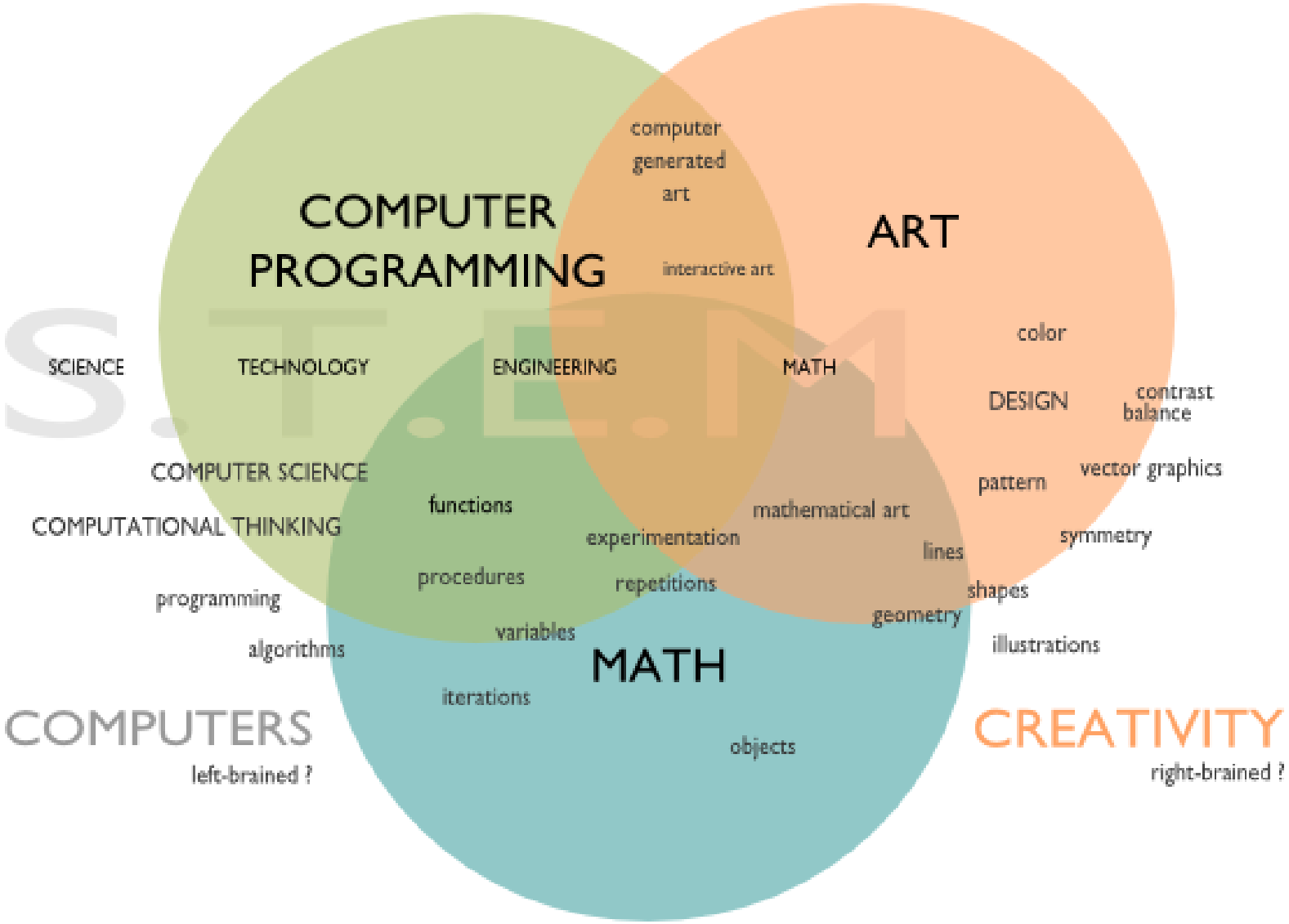
需要！

- 没有现成的工具
- 兄弟们都在忙着混江湖
- 工资涨如龟速，物价涨如赤兔
- ……

编程 | 值得学习

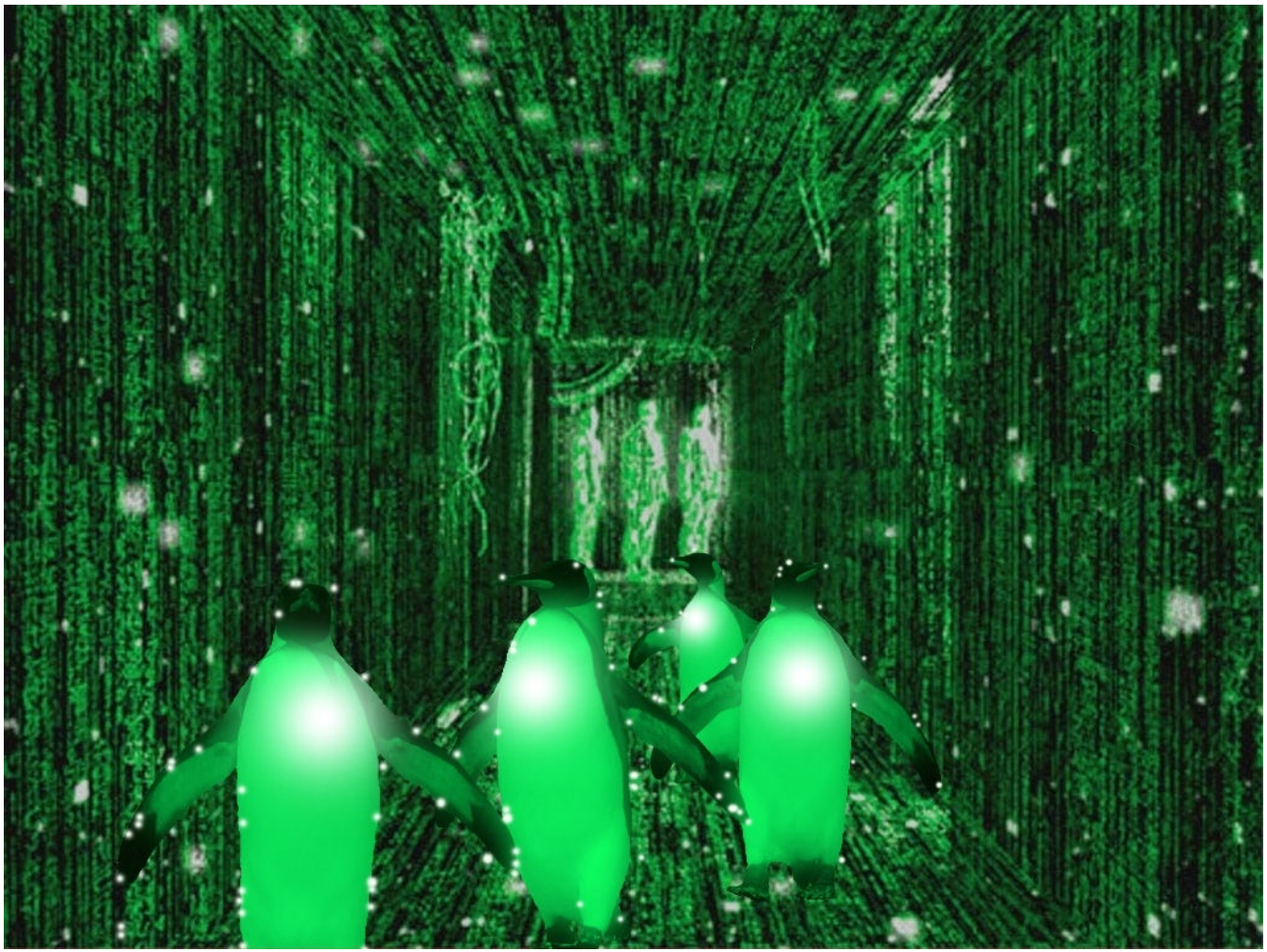
- 有助于理解现有工具的配置与个性化修改
- 编写程序来批量运行现有程序
- 数据分析很简单，前期数据处理很难难难……（二八定律，80/20法则，帕雷托法则）
- 增强研究工作的可重复性
- 人生苦短，学习编程
- ……

编程 | 与艺术


COMPUTER PROGRAMMING
ART
MATH
S.T.E.M
SCIENCE
TECHNOLOGY
ENGINEERING
COMPUTER SCIENCE
COMPUTATIONAL THINKING
programming
algorithms
functions
procedures
variables
iterations
computer generated art
interactive art
experimentation
repetitions
objects
mathematical art
lines
geometry
color
DESIGN
contrast balance
pattern
vector graphics
symmetry
shapes
illustrations
COMPUTERS
left-brained ?
CREATIVITY
right-brained ?

每一位程序员都是一位艺术家！
每一个程序都是一件艺术品！
编写程序 = 艺术创作！

If you program enough,
it can change the way you look at the world ···



任务

1. 同时精确测量粒子的位置和动量
2. 寻找第四种可以镶嵌平面的凸六边形
3. 寻找可以镶嵌平面的边数大于 6 的凸多边形
4. 分析敲除所有必需基因后的基因表达

原理

1. 海森堡不确定性原理
2. 有且只有三种凸六边形可以镶嵌平面
3. 边数大于 6 的凸多边形都无法镶嵌平面
4. 都说了是必需基因……

任务

1. 同时精确测量粒子的位置和动量
2. 寻找第四种可以镶嵌平面的凸六边形
3. 寻找可以镶嵌平面的边数大于 6 的凸多边形
4. 分析敲除所有必需基因后的基因表达

原理

1. 海森堡不确定性原理
2. 有且只有三种凸六边形可以镶嵌平面
3. 边数大于 6 的凸多边形都无法镶嵌平面
4. 都说了是必需基因……

任务

1. 同时精确测量粒子的位置和动量
2. 寻找第四种可以镶嵌平面的凸六边形
3. 寻找可以镶嵌平面的边数大于 6 的凸多边形
4. 分析敲除所有必需基因后的基因表达

原理

1. 海森堡不确定性原理
2. 有且只有三种凸六边形可以镶嵌平面
3. 边数大于 6 的凸多边形都无法镶嵌平面
4. 都说了是必需基因……

任务

1. 同时精确测量粒子的位置和动量
2. 寻找第四种可以镶嵌平面的凸六边形
3. 寻找可以镶嵌平面的边数大于 6 的凸多边形
4. 分析敲除所有必需基因后的基因表达

原理

1. 海森堡不确定性原理
2. 有且只有三种凸六边形可以镶嵌平面
3. 边数大于 6 的凸多边形都无法镶嵌平面
4. 都说了是必需基因……

编程 | 局限 | （现阶段）实际上不可解

任务

1. 破解 RSA 秘钥
2. 暴力破解由 94 个字符（26 小写、26 大写、10 数字、32 标点）随机组合成的长 12 个字符的密码
3. 蛋白质折叠中通过随机尝试找到总能量最低的构象状态

原因

1. 对极大整数做因数分解非常困难（破解 RSA-2048（2048-bit）的密钥可能需要耗费传统电脑 10 亿年的时间，而量子计算机只需要 100 秒就可以完成。）【量子计算机 + 秀尔算法】
2. 普通台式机，每秒运算 40 亿次，需要大约 30 万年以上才能破解
3. 利文索尔佯谬（Levinthal's paradox）：100 氨基酸，每个 2 种构象，每次尝试耗时 10^{-13}s，穷举需要 40 亿年

编程 | 局限 | （现阶段）实际上不可解

任务

1. 破解 RSA 秘钥
2. 暴力破解由 94 个字符（26 小写、26 大写、10 数字、32 标点）随机组合成的长 12 个字符的密码
3. 蛋白质折叠中通过随机尝试找到总能量最低的构象状态

原因

1. 对极大整数做因数分解非常困难（破解 RSA-2048（2048-bit）的密钥可能需要耗费传统电脑 10 亿年的时间，而量子计算机只需要 100 秒就可以完成。）【量子计算机 + 秀尔算法】
2. 普通台式机，每秒运算 40 亿次，需要大约 30 万年以上才能破解
3. 利文索尔佯谬（Levinthal's paradox）：100 氨基酸，每个 2 种构象，每次尝试耗时 10^{-13}s，穷举需要 40 亿年

编程 | 局限 | （现阶段）实际上不可解

任务

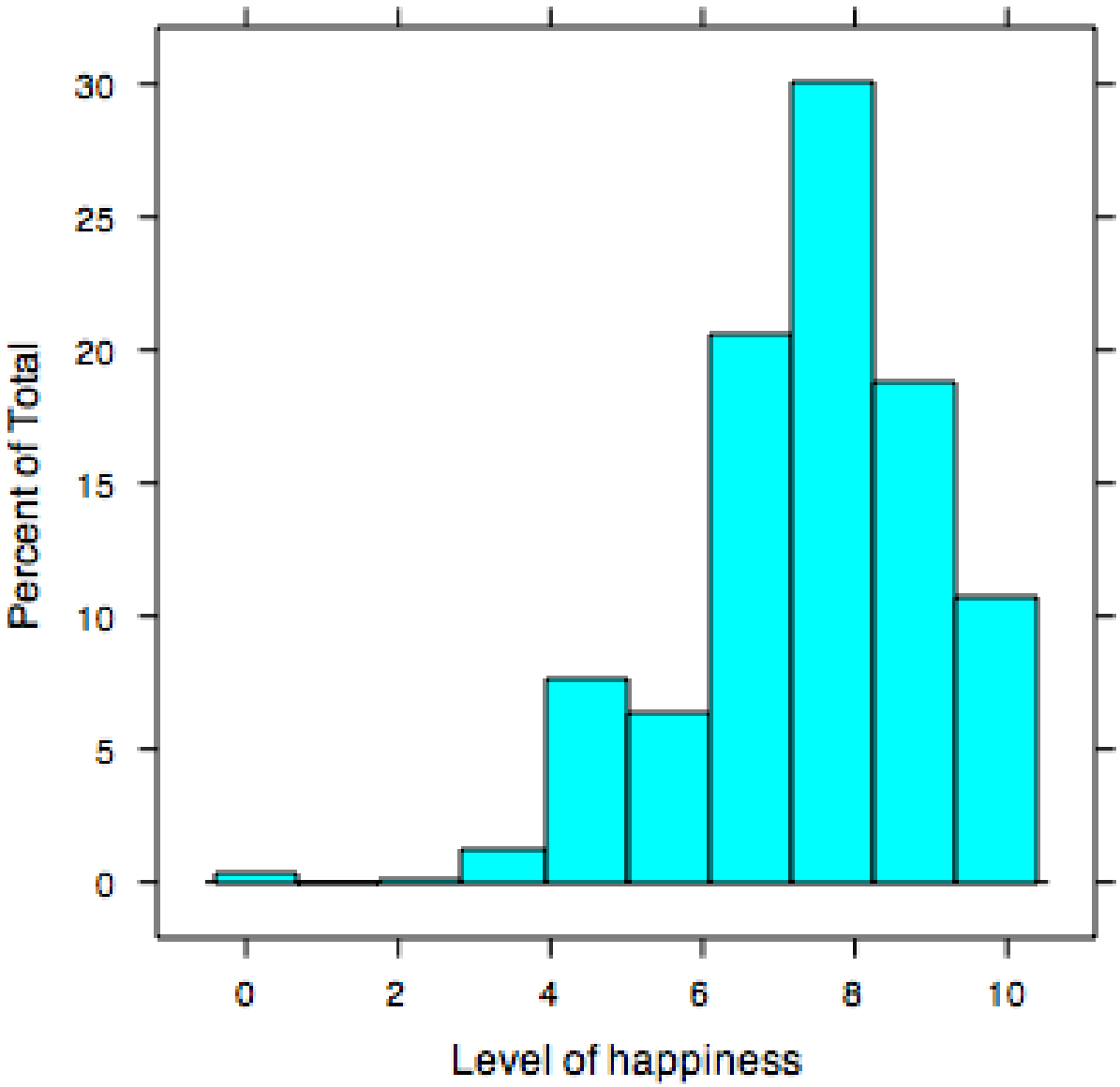
1. 破解 RSA 秘钥
2. 暴力破解由 94 个字符（26 小写、26 大写、10 数字、32 标点）随机组合成的长 12 个字符的密码
3. 蛋白质折叠中通过随机尝试找到总能量最低的构象状态

原因

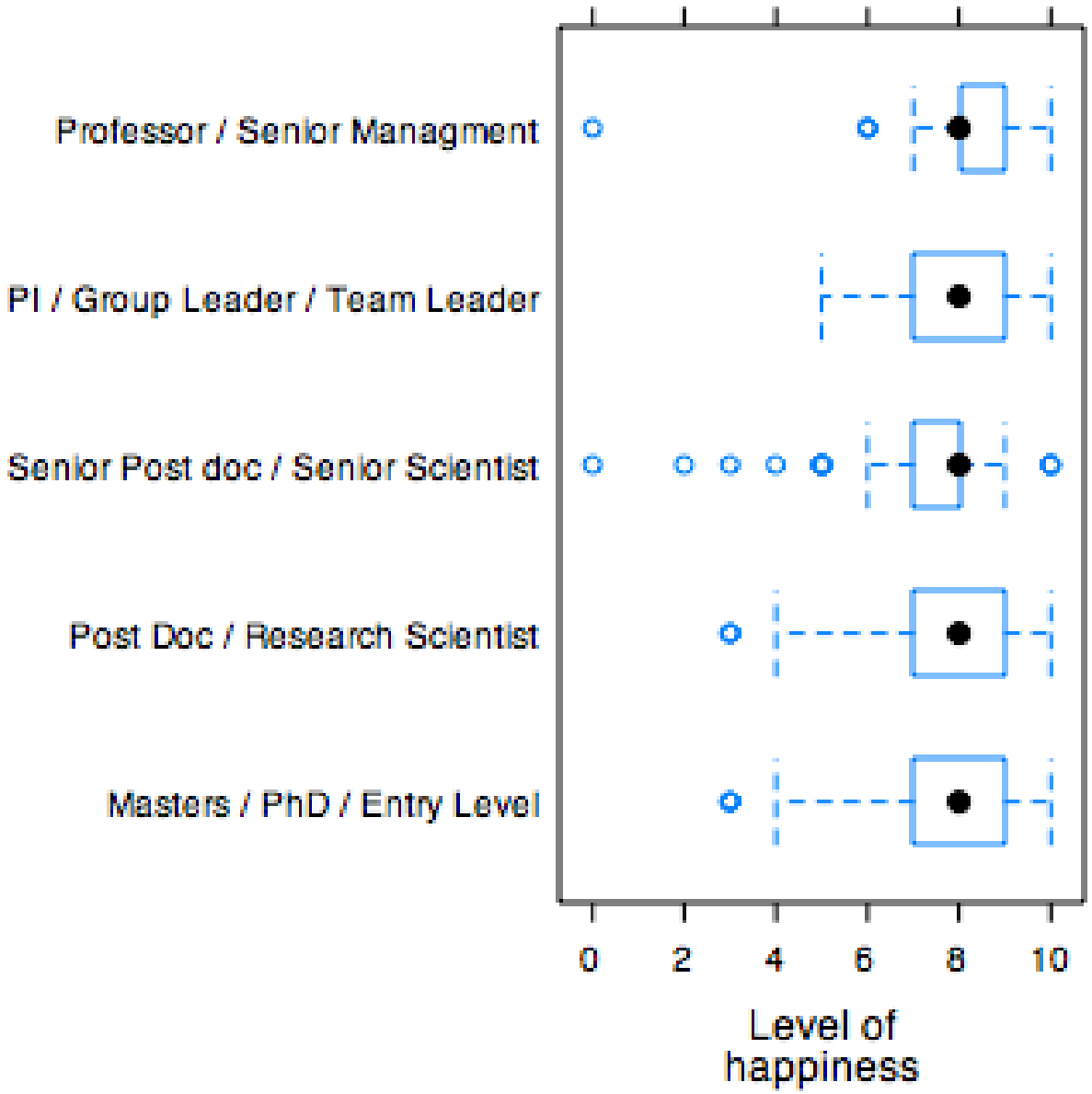
1. 对极大整数做因数分解非常困难（破解 RSA-2048（2048-bit）的密钥可能需要耗费传统电脑 10 亿年的时间，而量子计算机只需要 100 秒就可以完成。）【量子计算机 + 秀尔算法】
2. 普通台式机，每秒运算 40 亿次，需要大约 30 万年以上才能破解
3. 利文索尔佯谬（Levinthal's paradox）：100 氨基酸，每个 2 种构象，每次尝试耗时 10^{-13}s，穷举需要 40 亿年

教学提纲

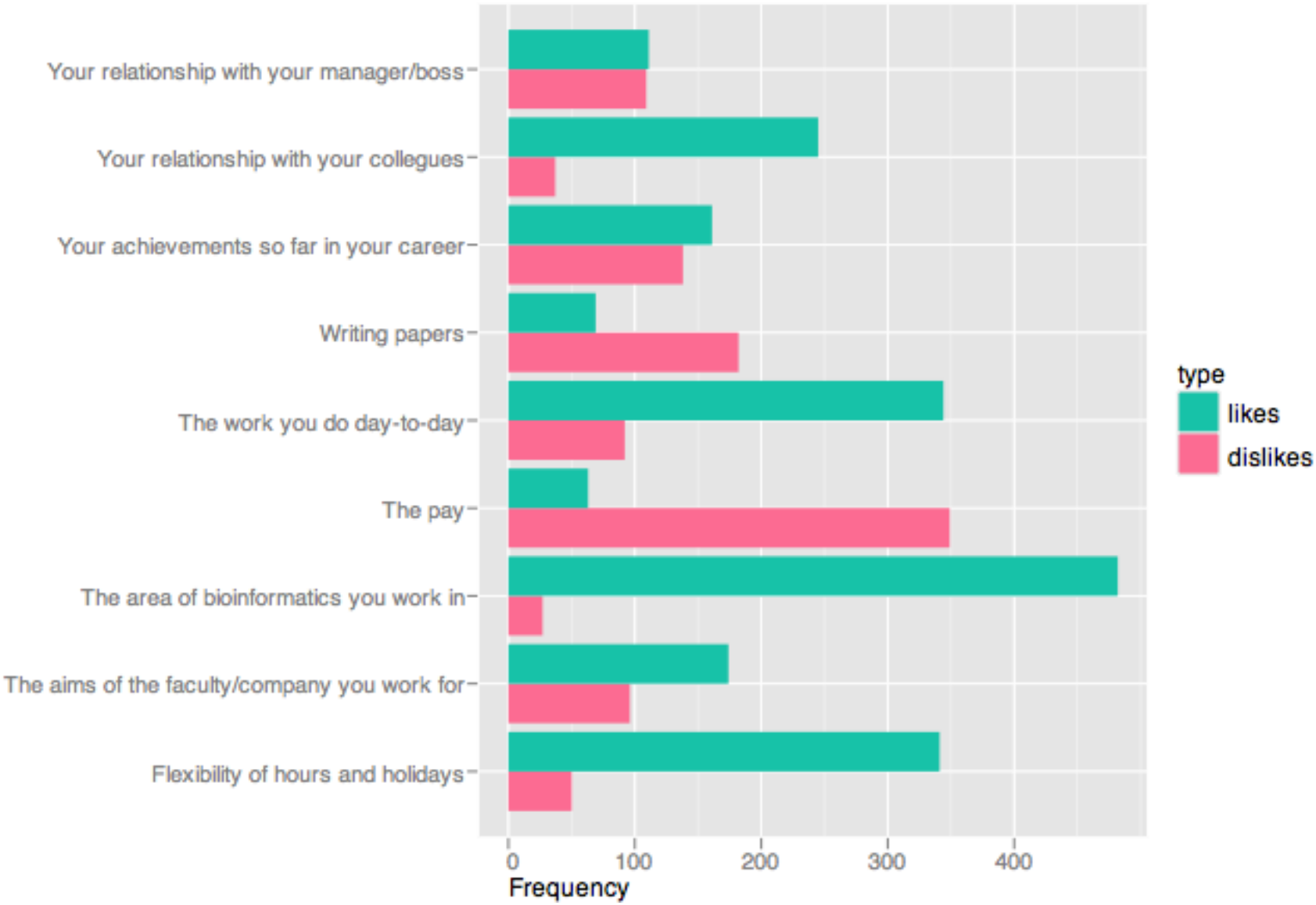
2008 | happiness



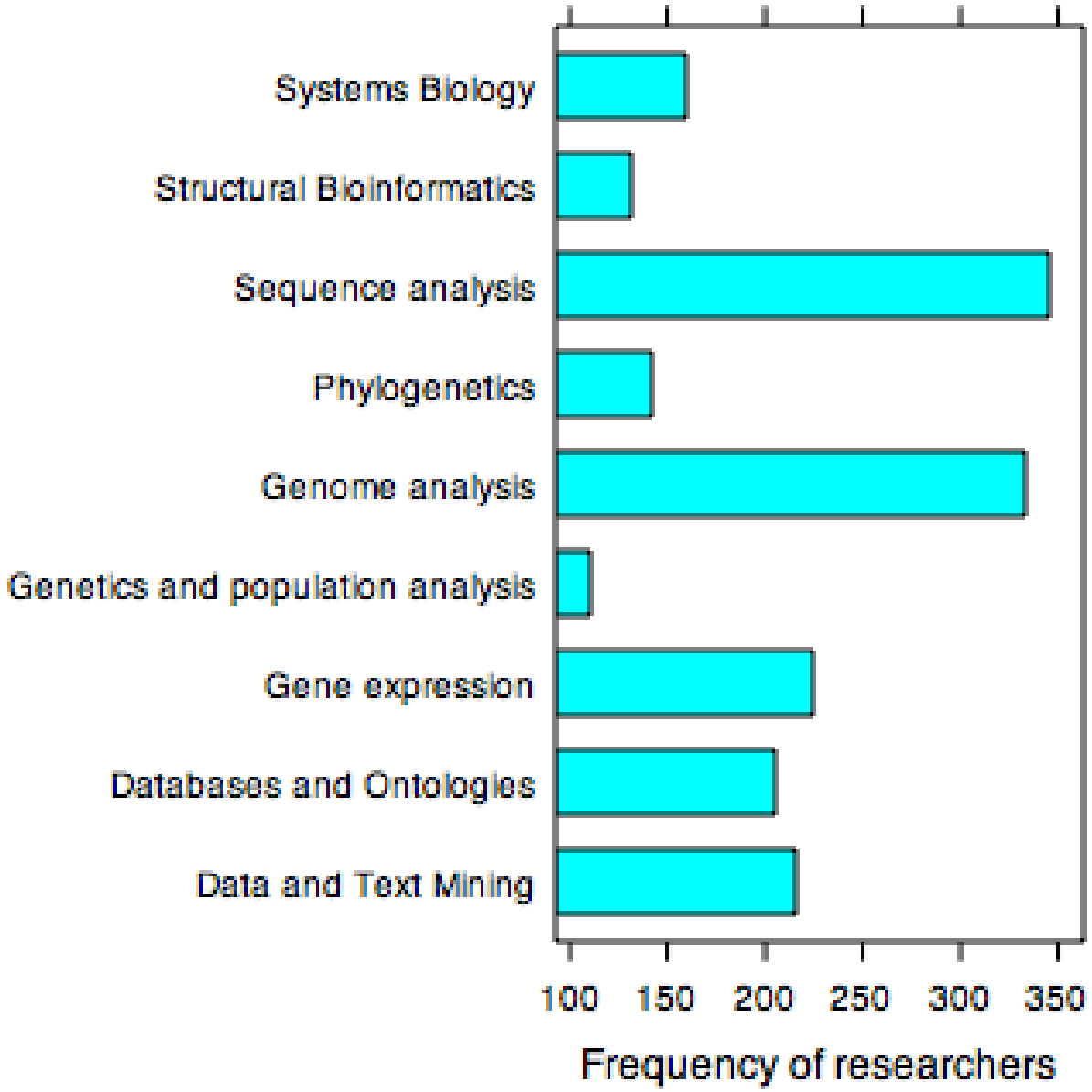
2008 | happiness vs. career position



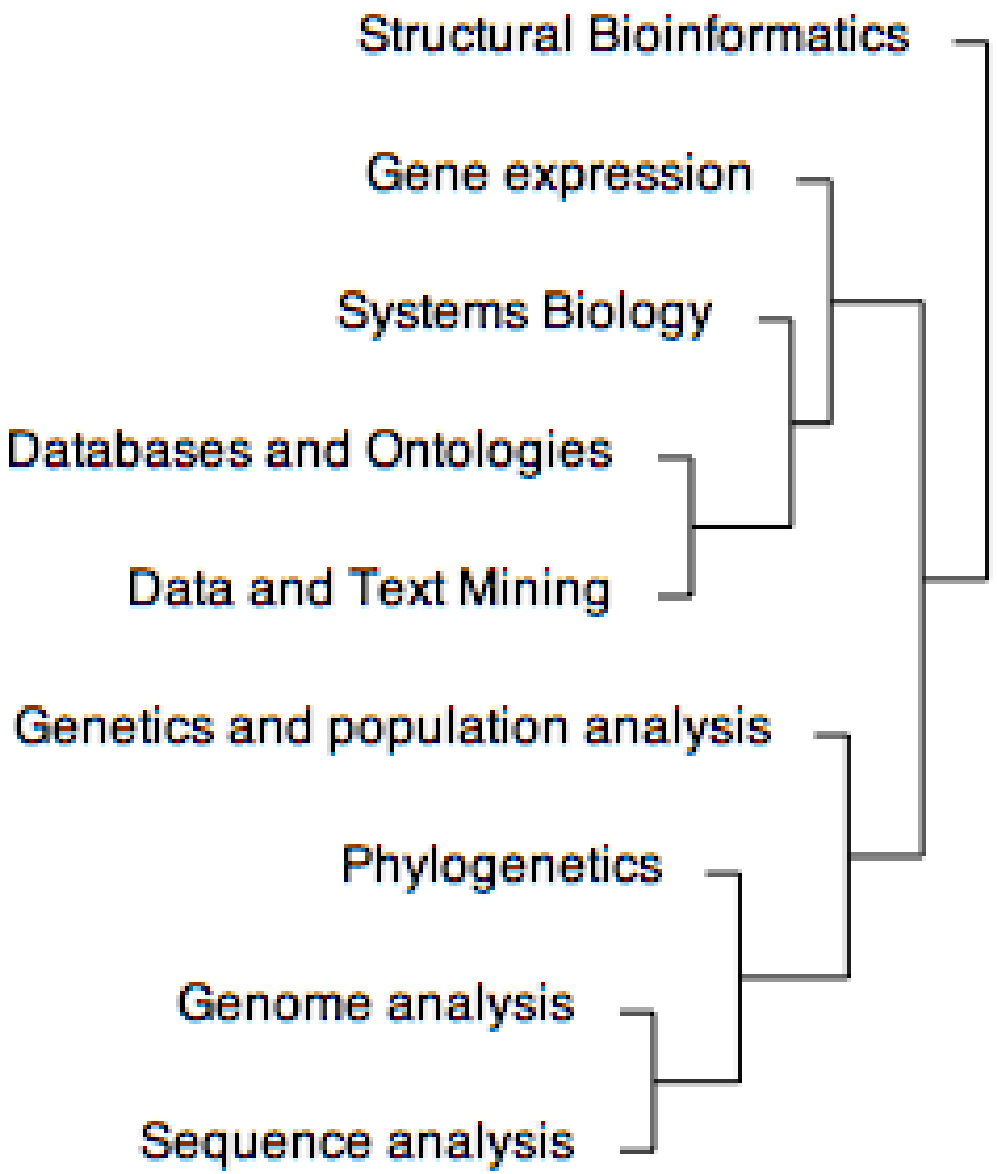
2008 | likes vs. dislikes



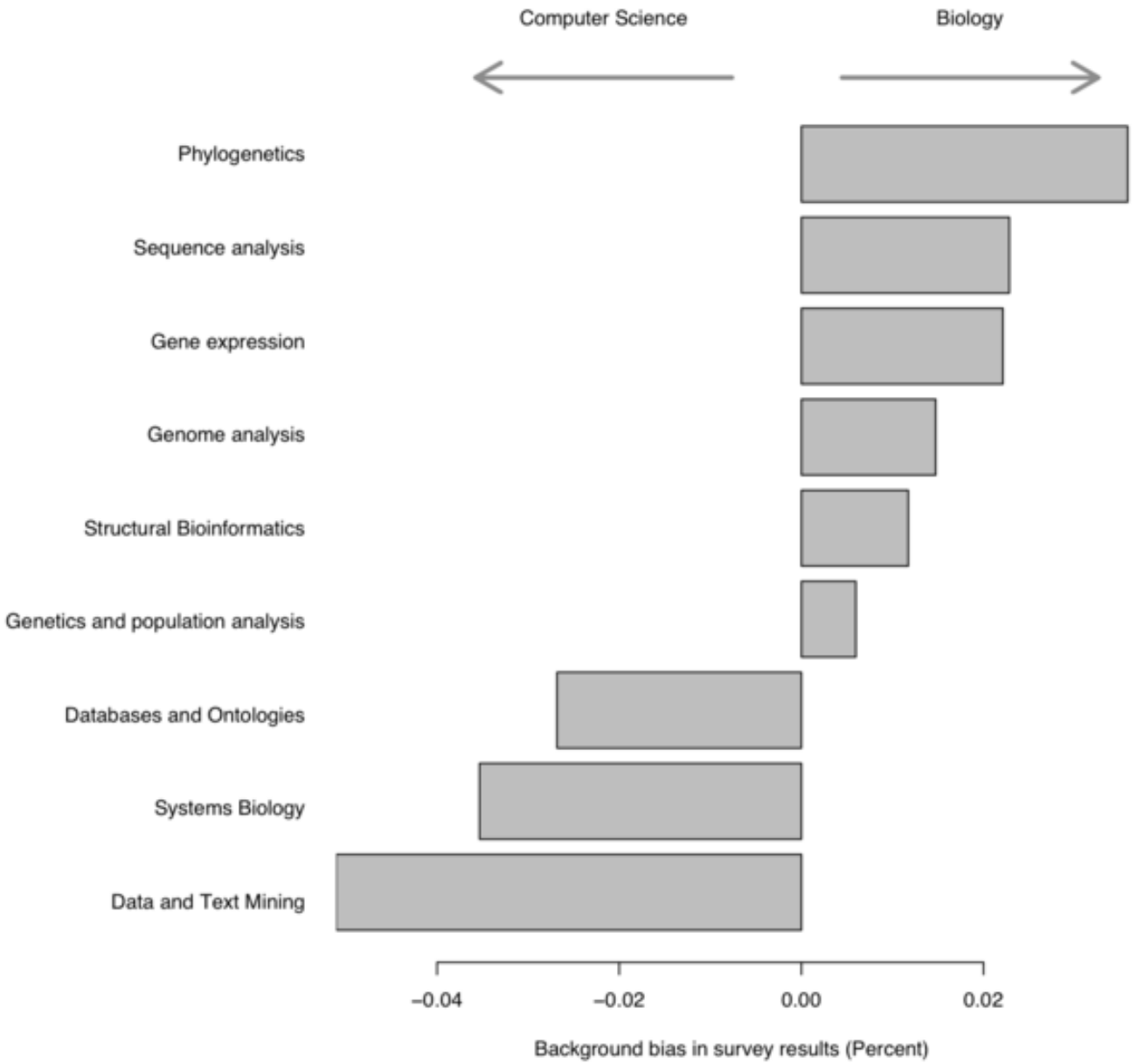
2008 | research area



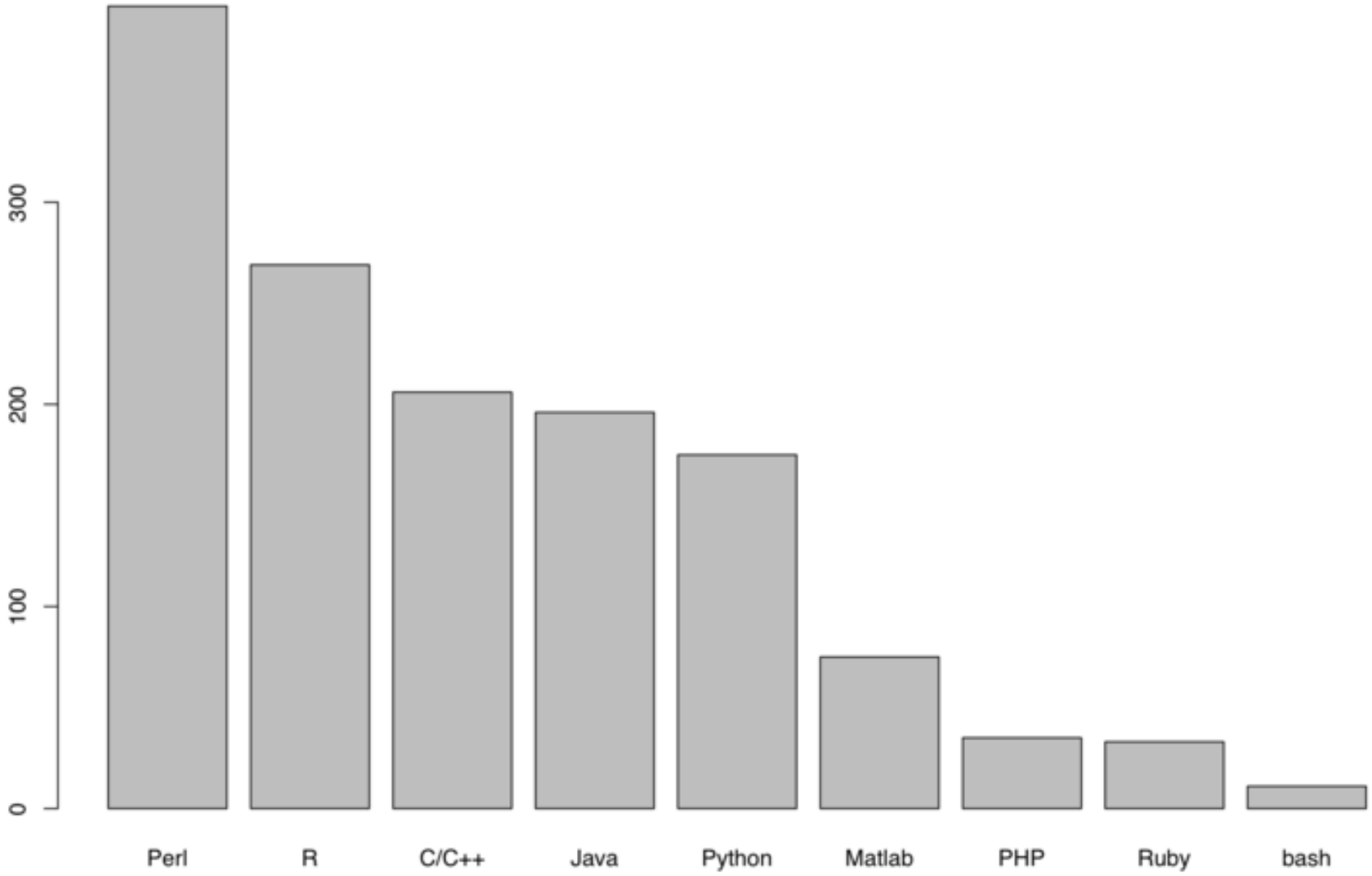
2008 | research area | clustering



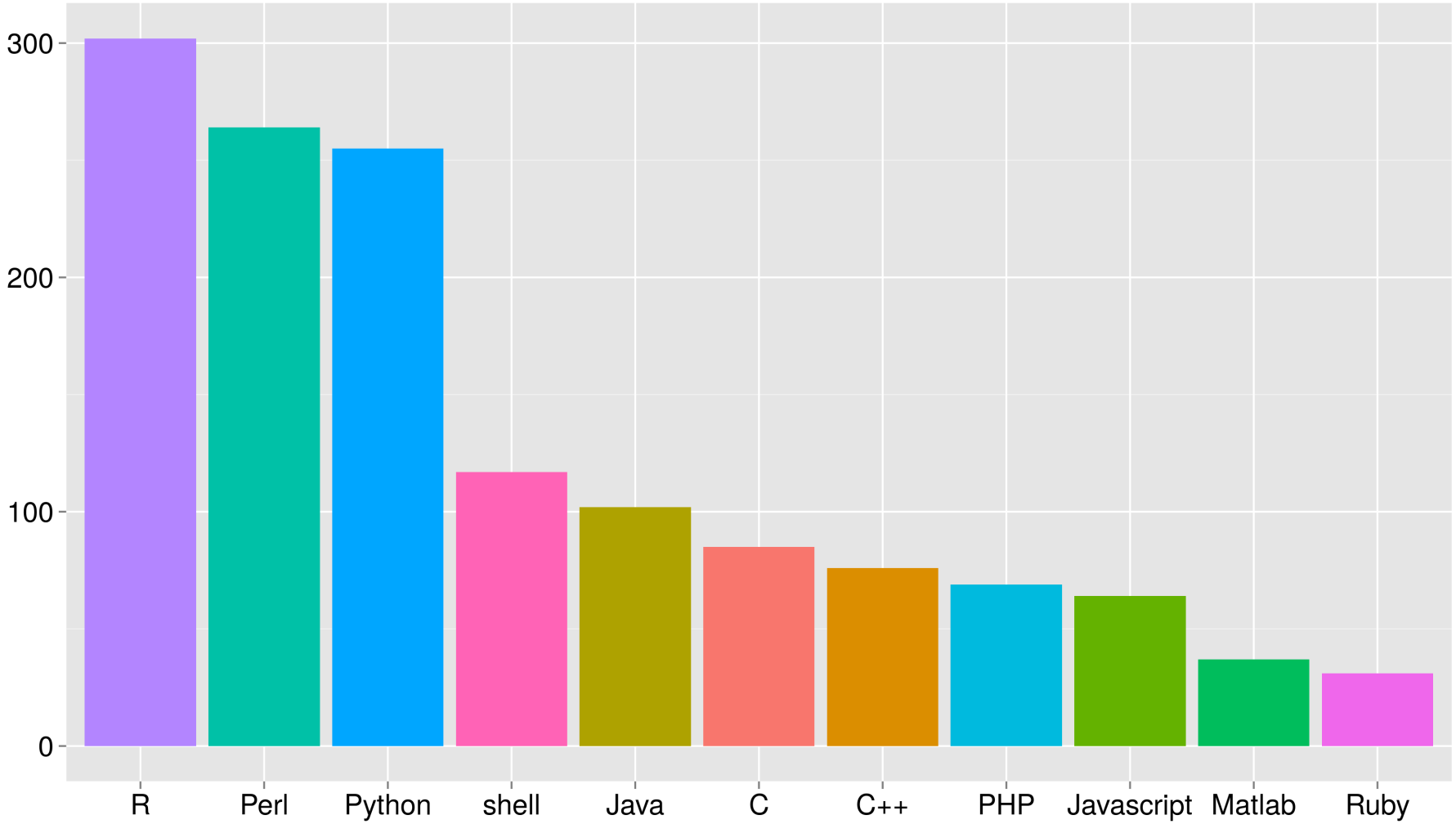
2008 | background



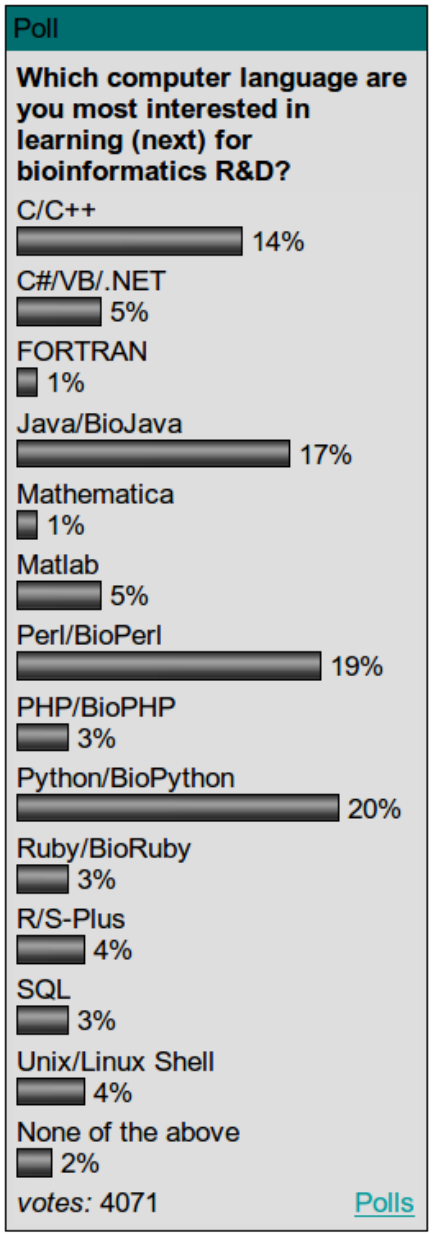
2008 | language



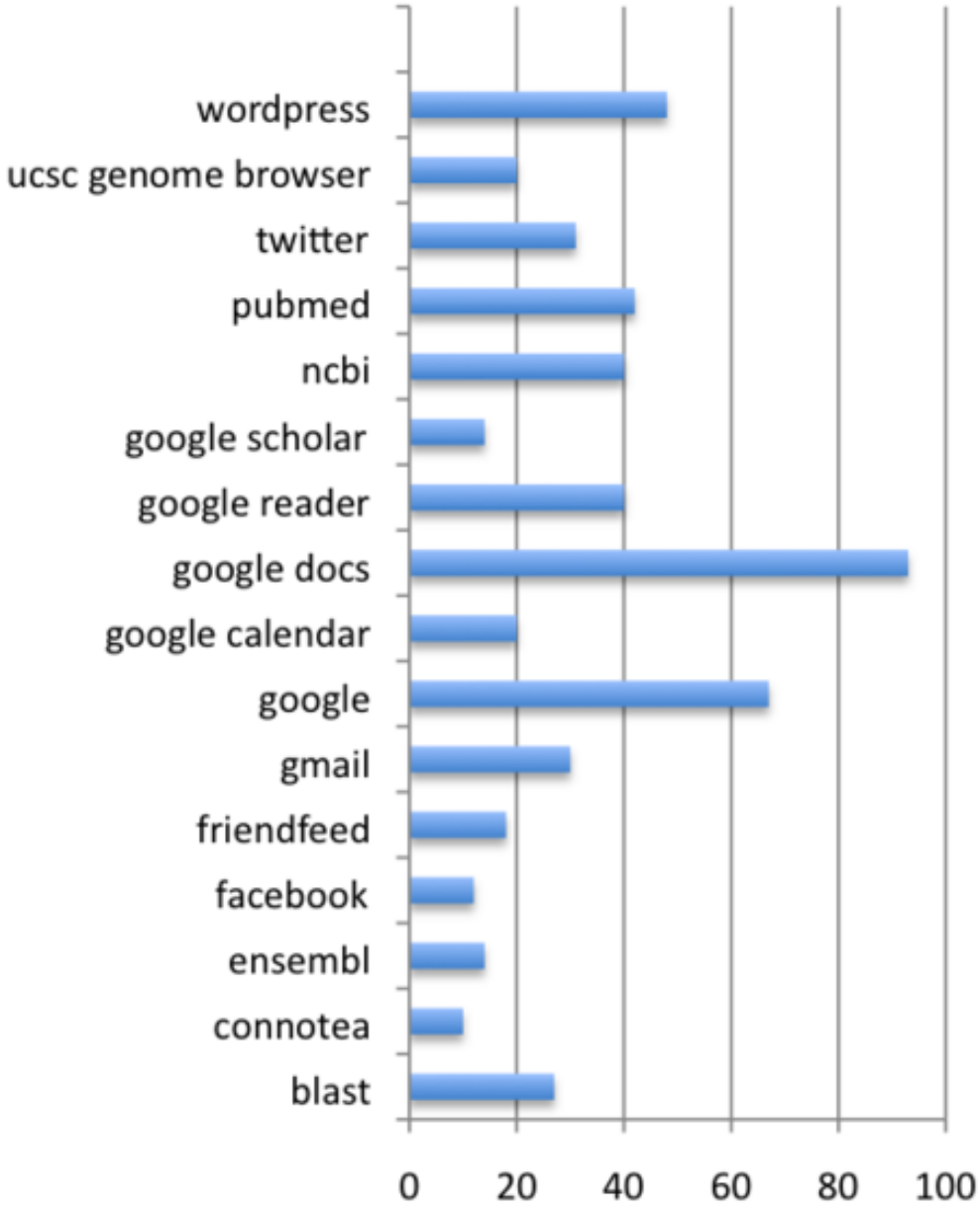
2012 | language



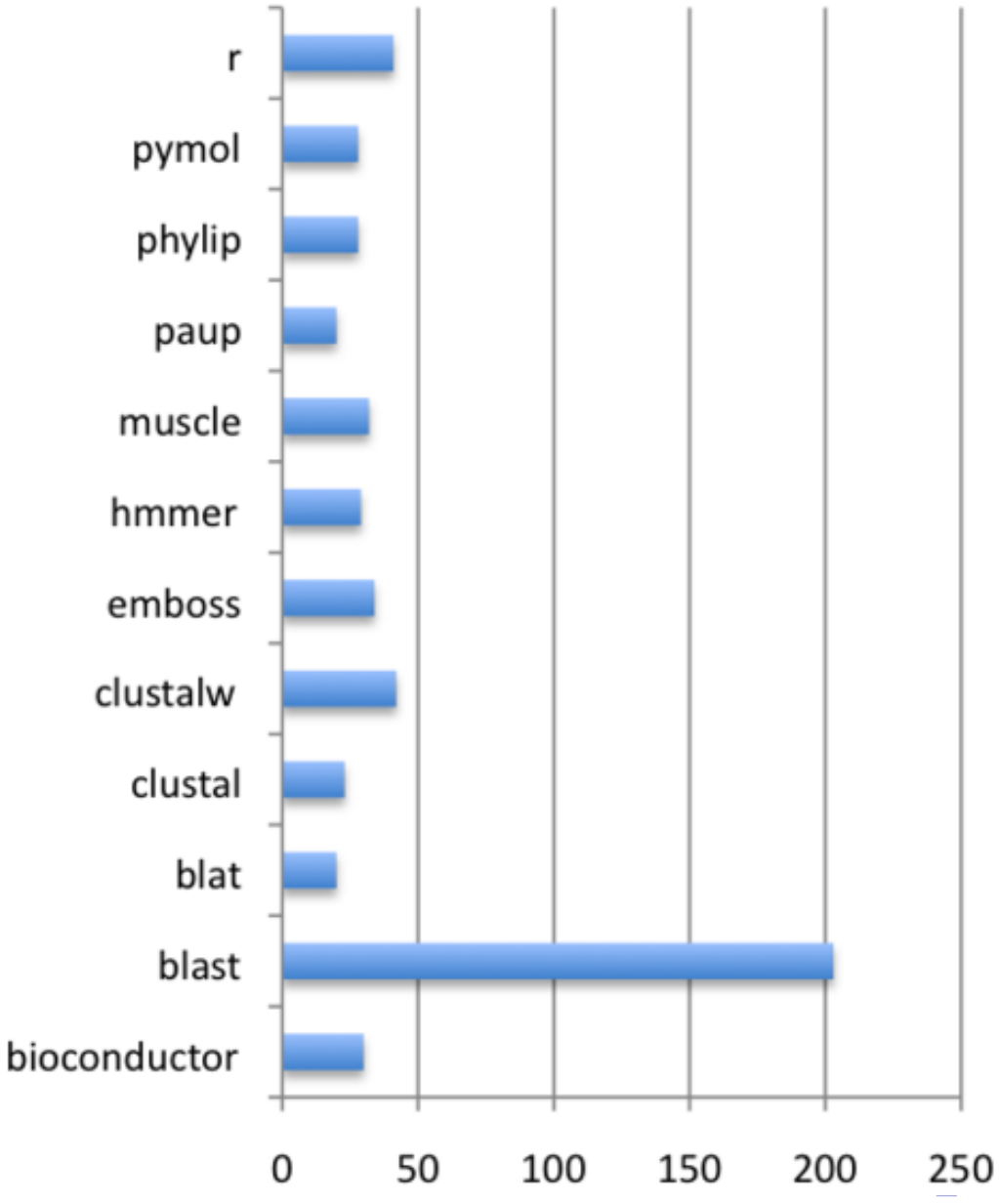
Bioinformatics.org (2007) | language



2008 | applications



2008 | tools


r
pymol
phylip
paup
muscle
hmmer
emboss
clustalw
clustal
blat
blast
bioconductor
0
50
100
150
200
250

2008 | salary

industry vs. academic

- Academic Salary Mean/Median: $36,520 / $33,712
- Industry Salary Mean/Median: $66,239 / $64,235

Salary/Years Experience

- Academic Mean/Median: $10,970 / $8,333
- Industry Mean/Median: $17,410 / $12,000

2008 | salary

industry vs. academic

- Academic Salary Mean/Median: $36,520 / $33,712
- Industry Salary Mean/Median: $66,239 / $64,235

Salary/Years Experience

- Academic Mean/Median: $10,970 / $8,333
- Industry Mean/Median: $17,410 / $12,000

- Bioinformatics Career Survey 2008
- Bioinformatics Career Survey 2008 Results
- bioinformatics-career-survey (GitHub)
- A comparison of common programming languages used in bioinformatics
- A comparison of bioinformatics programming languages
- Programming Languages of Bioinformatics

教学提纲

R

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

RStudio

RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

RStudio is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro (Debian/Ubuntu, RedHat/CentOS, and SUSE Linux).

R

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

RStudio

RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

RStudio is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro (Debian/Ubuntu, RedHat/CentOS, and SUSE Linux).

Useful and popular packages: Load data

readr	readr makes it easy to read many types of tabular data including: delimited files, fixed width files and web log files.
XLConnect	Help you read, write and format **Micorsoft Excel** files from R.
foreign	Functions for reading and writing data stored by some versions of Epi Info, Minitab, S, SAS, SPSS, Stata, Systat and Weka and for reading and writing some dBase files.
httr	A set of useful tools for working with **http** connections.
XML	Read and create **XML** documents with R.
jsonlite	Read and create **JSON** data tables with R.

Useful and popular packages: Manipulate data

tidyr Tools for changing the layout of your data sets. Use the gather and spread functions to convert your data into the tidy format, the layout R likes best.

magrittr magrittr provides a mechanism for chaining commands with a new forward-pipe operator, %>%.

dplyr Essential shortcuts for subsetting, summarizing, rearranging, and joining together data sets for fast data manipulation.

Useful and popular packages: Manipulate data

lubridate Tools that make working with dates and times easier.

stringr Easy to learn tools for regular expressions and character strings.

plyr plyr is a set of tools for a common set of problems: you need to **split** up a big data structure into homogeneous pieces, **apply** a function to each piece and then **combine** all the results back together.

DT The R package DT provides an R interface to the JavaScript library DataTables. R data objects (matrices or data frames) can be displayed as tables on HTML pages, and DataTables provides filtering, pagination, sorting, and many other features in the tables.

Useful and popular packages: Visualize data

ggplot2	R's famous package for making beautiful graphics. ggplot2 lets you use the grammar of graphics to build layered, customizable plots.
ggvis	**Interactive, web based graphics** built with the grammar of graphics.
DiagrammeR	Create **graph diagrams and flowcharts** using R.
htmlwidgets	The htmlwidgets package provides a framework for easily creating R bindings to JavaScript libraries.

Useful and popular packages: Report results

rmarkdown rmarkdown lets you insert R code into a markdown document. R then generates a final document, in a wide variety of formats, that replaces the R code with its results.

knitr knitr is an elegant, flexible and fast dynamic report generation that combines R with TeX, Markdown, or HTML. For open access publishing, and reproducible research in statistics.

xtable The xtable function takes an R object (like a data frame) and returns the latex or HTML code you need to paste a pretty version of the object into your documents. Copy and paste, or pair up with R Markdown.

Useful and popular packages: Report results

Shiny Easily make **interactive, web apps** with R. A perfect way to explore data and share findings with non-programmers.

shinydashboard shinydashboard makes it easy to use Shiny to create dashboards.

Useful and popular packages: High performance

data.table	An alternative way to organize data sets for very, very fast operations. Useful for big data.
parallel	Use **parallel processing** in R to speed up your code or to crunch large data sets.

Useful and popular packages: Development

devtools	An essential suite of tools for turning your code into an R package.
roxygen2	A quick way to **document** your R packages. roxygen2 turns inline code comments into documentation pages and builds a package namespace.
testthat	testthat provides an easy way to write **unit tests** for your code projects.

Easily install and load packages from the tidyverse

The tidyverse is a collection of R packages that share common philosophies and are designed to work together.
The core tidyverse packages that you are likely to use in almost every analysis:

- readr, for data import.
- tidyr, for data tidying.
- dplyr, for data manipulation.
- ggplot2, for data visualisation.
- tibble, for tibbles, a modern re-imagining of data frames.
- purrr, for functional programming.

Easily install and load packages from the tidyverse

It also installs a selection of other tidyverse packages that you're likely to use frequently, but probably not in every analysis. This includes packages for:

- Working with specific types of vectors: hms (times), stringr (strings), lubridate (date/times), forcats (factors).
- Importing other types of data: DBI (databases), haven (SPSS, SAS and Stata files), httr (web apis), jsonlite (JSON), readxl (.xls and .xlsx files), rvest (web scraping), xml2 (XML).
- Modelling: modelr (modelling within a pipeline), broom (turning models into tidy data).

Others

viridis	Use the color scales in this package to make plots that are pretty, better represent your data, easier to read by those with colorblindness, and print well in grey scale.
argparse	A command line parser to be used with Rscript to write "#!" shebang scripts that gracefully accept positional and optional arguments and automatically generate usage.
ggtree	Visualization and annotation of phylogenetic trees.
ComplexHeatmap	ComplexHeatmap package provides a highly flexible way to arrange multiple heatmaps and supports self-defined annotation graphics.
...	...

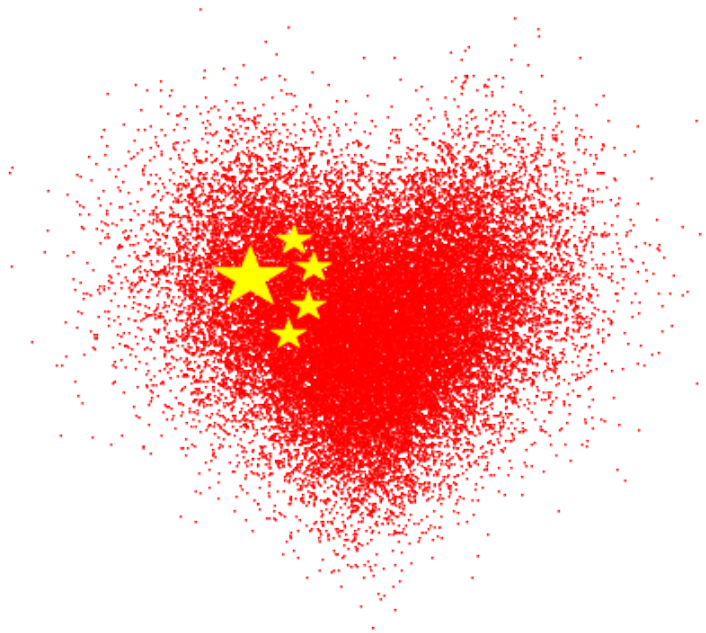
Bioconductor

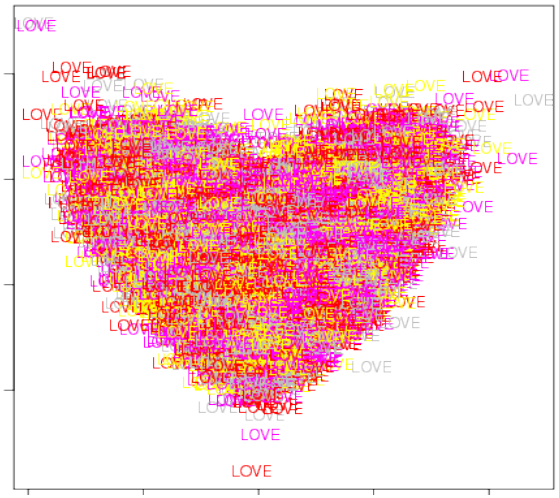
Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development.

R | Example

```
1 n <- 50000
2 r <- 0.7
3 r_e <- (1 - r * r) ^ 0.5
4 X <- rnorm(n)
5 Y <- X * r + r_e * rnorm(n)
6 Y <- ifelse(X>0, Y, -Y)
7 plot(X, Y, col="pink")
```

R | Example

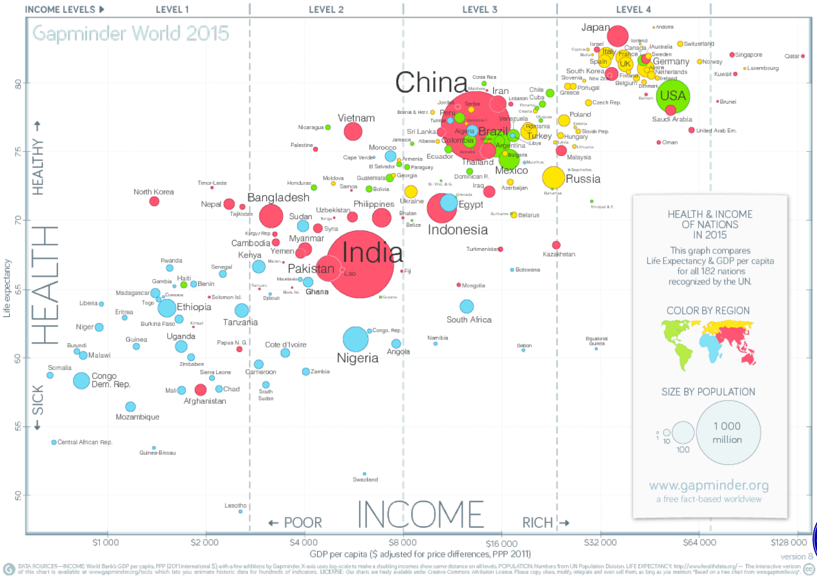




R | Example

```
1  xrange <- c(-15, 15); yrange <- c(0, 16)
2  plot(0, xlim=xrange, ylim=yrange, type="n")
3
4  yr <- seq(yrange[1], yrange[2], len=50)
5  offsetFn <- function(y) { 2 * sin(0 + y/3) }
6  offset <- offsetFn(yr)
7  leftE <- function(y) { -10 - offsetFn(y) }
8  rightE <- function(y) { 10 + offsetFn(y) }
9  xp <- c(leftE(yr), rev(rightE(yr)))
10 yp <- c(yr, rev(yr))
11 polygon(xp, yp, col="#ffeecc", border=NA)
12
13 h <- 9
14 xt <- seq(0, rightE(h), len=100)
15 yt <- log(1 + log(1 + log(xt + 1)))
16 yt <- yt - min(yt); yt <- h * yt/max(yt)
17 x <- c(leftE(h), rightE(h), rev(xt), -xt)
18 y <- c(h, h, rev(yt), yt)
19 polygon(x, y, col="red", border=NA)
```

R | Example

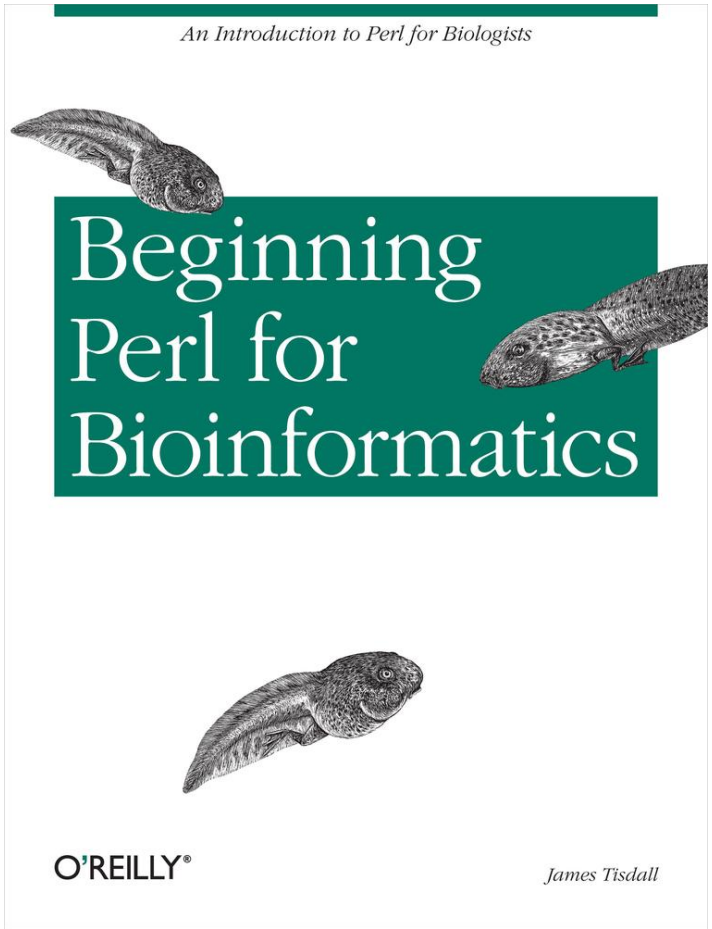

INCOME LEVELS ▶
LEVEL 1
LEVEL 2
LEVEL 3
LEVEL 4
Gapminder World 2015
HEALTH
HEALTHY →
← SICK
Life expectancy
INCOME
← POOR
RICH →
GDP per capita ($ adjusted for price differences, PPP 2011)
$1 000
$2 000
$4 000
$8 000
$16 000
$32 000
$64 000
$128 000
50
55
60
65
70
75
80
China
India
USA
Japan
Indonesia
Brazil
Pakistan
Nigeria
Bangladesh
Russia
Mexico
Germany
Vietnam
Philippines
Ethiopia
Egypt
Iran
Turkey
Thailand
South Africa
Congo Dem. Rep.
Tanzania
Kenya
Myanmar
Sudan
Uganda
Afghanistan
Nepal
North Korea
Mozambique
HEALTH & INCOME
OF NATIONS
IN 2015
This graph compares
Life Expectancy & GDP per capita
for all 182 nations
recognized by the UN.
COLOR BY REGION
SIZE BY POPULATION
1 000 million
100
10
1
www.gapminder.org
a free fact-based worldview
version 8

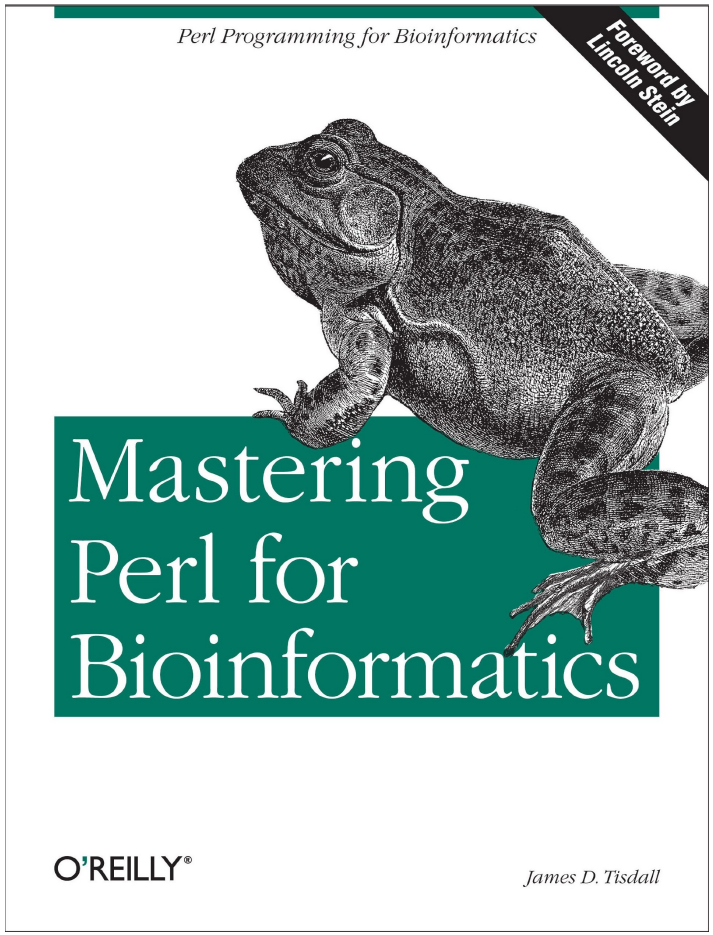
- The R Project for Statistical Computing
- RStudio
- Learn R
- RStudio Cheat Sheets
- R packages inspired by R and its community
- Quick list of useful R packages
- Bioconductor

教学提纲

生物信息学角度

Beginning $\Longrightarrow$ *Mastering Perl for Bioinformatics*





编程语言角度：代号

小骆驼 ⇒ 羊驼书 ⇒ 小羊驼母亲和她的孩子 ⇒ 大骆驼 ⇒ 黑豹书

英文名

Learning Perl ⇒ Intermediate Perl ⇒ Mastering Perl ⇒ Programming Perl ⇒ Advanced Perl Programming

中文名

《Perl 语言入门》⇒《Perl 进阶》⇒《精通 Perl》⇒《Perl 语言编程》⇒《高级 Perl 编程》

编程语言角度：代号

小骆驼 ⇒ 羊驼书 ⇒ 小羊驼母亲和她的孩子 ⇒ 大骆驼 ⇒ 黑豹书

英文名

Learning Perl ⇒ *Intermediate Perl* ⇒ *Mastering Perl* ⇒ *Programming Perl* ⇒ *Advanced Perl Programming*

中文名

《Perl 语言入门》⇒ 《Perl 进阶》⇒ 《精通 Perl》⇒ 《Perl 语言编程》⇒ 《高级 Perl 编程》

编程语言角度：代号

小骆驼 ⇒ 羊驼书 ⇒ 小羊驼母亲和她的孩子 ⇒ 大骆驼 ⇒ 黑豹书

英文名

Learning Perl ⇒ *Intermediate Perl* ⇒ *Mastering Perl* ⇒ *Programming Perl* ⇒ *Advanced Perl Programming*

中文名

《Perl 语言入门》⇒ 《Perl 进阶》⇒ 《精通 Perl》⇒ 《Perl 语言编程》⇒ 《高级 Perl 编程》

编程语言角度：代号

小骆驼 ⇒ 羊驼书 ⇒ 小羊驼母亲和她的孩子 ⇒ 大骆驼 ⇒ 黑豹书

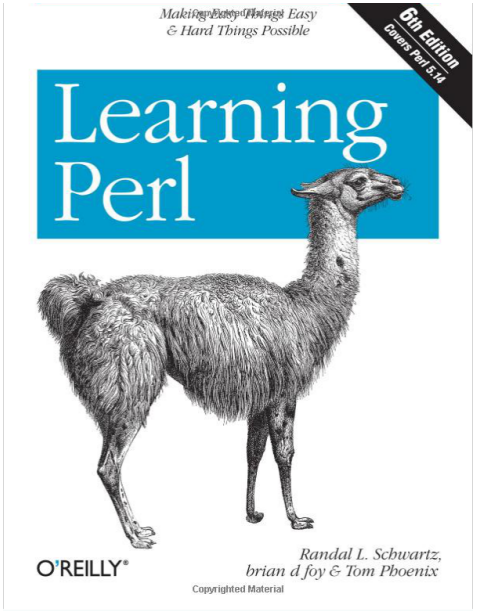
英文名

Learning Perl ⇒ *Intermediate Perl* ⇒ *Mastering Perl* ⇒ *Programming Perl* ⇒ *Advanced Perl Programming*

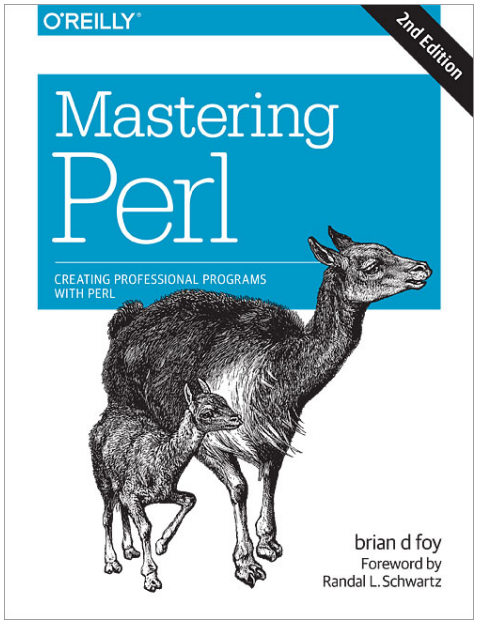
中文名

《Perl 语言入门》⇒ 《Perl 进阶》⇒ 《精通 Perl》⇒ 《Perl 语言编程》⇒ 《高级 Perl 编程》

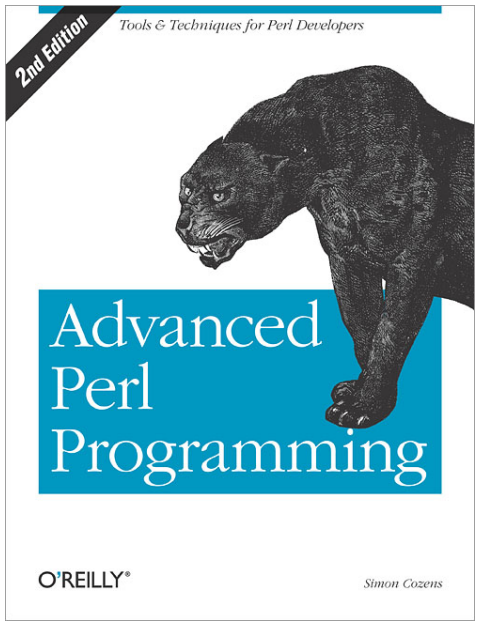
书籍 | Perl


Making Easy Things Easy & Hard Things Possible
6th Edition
Covers Perl 5.14
Learning Perl
O'REILLY
Randal L. Schwartz, brian d foy & Tom Phoenix
Copyrighted Material


Beyond the Basics of Learning Perl
2nd Edition
Covers Perl 5.14
Intermediate Perl
O'REILLY
Randal L. Schwartz, brian d foy & Tom Phoenix


O'REILLY
2nd Edition
Mastering Perl
CREATING PROFESSIONAL PROGRAMS WITH PERL
brian d foy
Foreword by Randal L. Schwartz


Unmatched Power for Text Processing and Scripting
4th Edition
Covers Version 5.14
Programming Perl
O'REILLY
Tom Christiansen, brian d foy & Larry Wall
with Jon Orwant


2nd Edition
Tools & Techniques for Perl Developers
Advanced Perl Programming
O'REILLY
Simon Cozens

书籍 | Perl

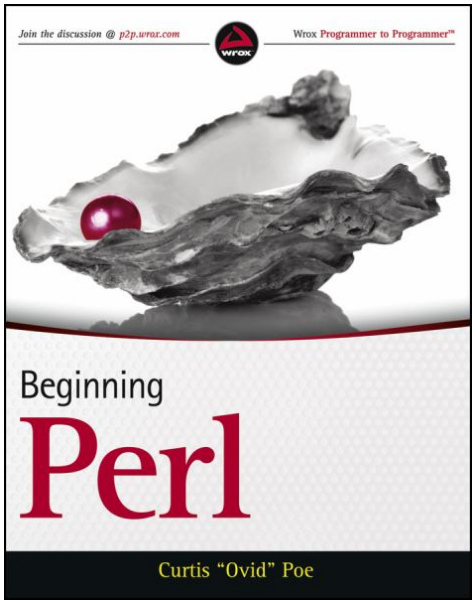
中文名

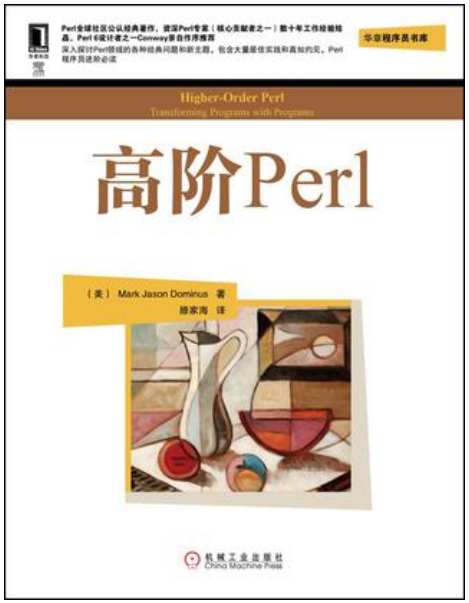
- 《Perl 入门经典》
- 《高阶 Perl》
- 《Perl 高效编程》
- 《Perl 最佳实践》
- *Perl Cookbook*

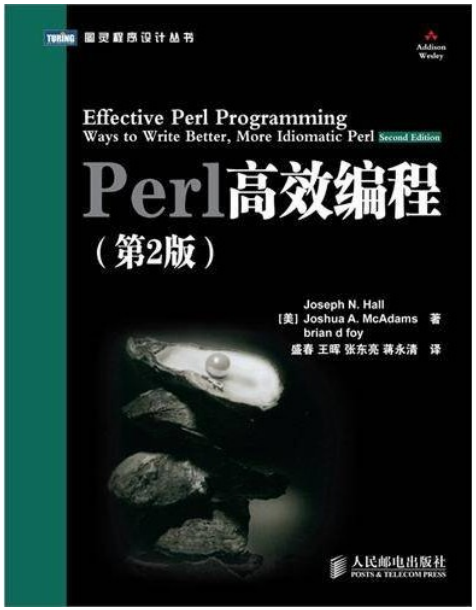
英文名

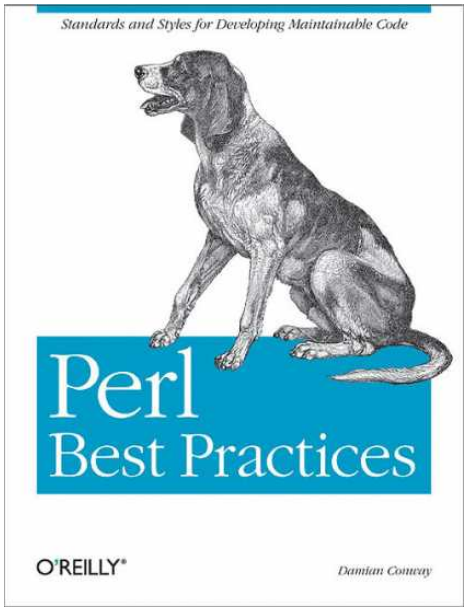
- *Beginning Perl*
- *Higher-Order Perl*
- *Effective Perl Programming*
- *Perl Best Practices*
- *Perl Cookbook*

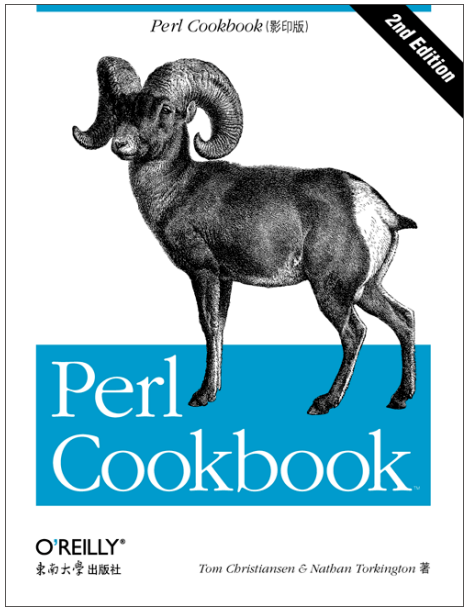
书籍 | Perl


Join the discussion @ p2p.wrox.com
Wrox Programmer to Programmer™
Beginning
Perl
Curtis "Ovid" Poe


Higher-Order Perl
高阶Perl
（美） Mark Jason Dominus 著
China Machine Press


Effective Perl Programming
Ways to Write Better, More Idiomatic Perl Second Edition
Perl高效编程
（第2版）
Joseph N. Hall
[美] Joshua A. McAdams 著
brian d foy
盛春 王晖 张东亮 蒋永清 译
人民邮电出版社


Standards and Styles for Developing Maintainable Code
Perl
Best Practices
O'REILLY
Damian Conway


Perl Cookbook（影印版）
2nd Edition
Perl
Cookbook
O'REILLY
东南大学出版社
Tom Christiansen & Nathan Torkington 著

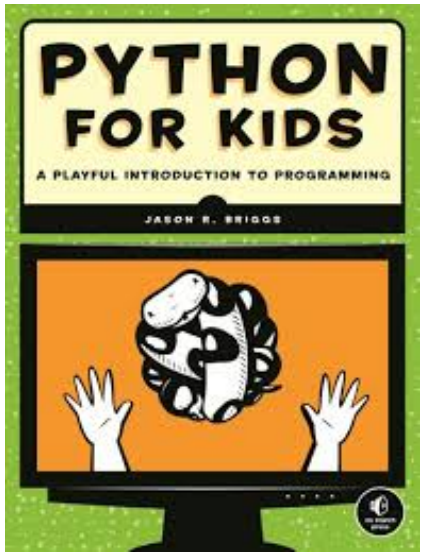
中文名

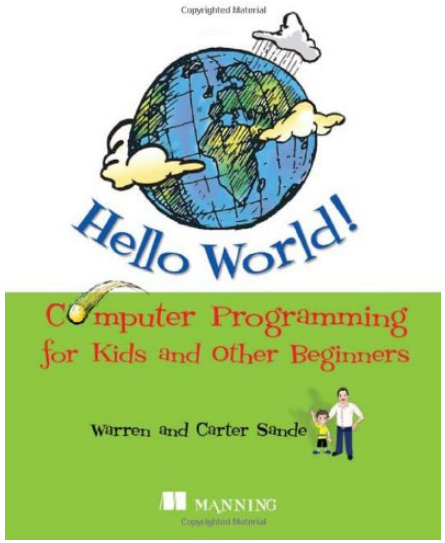
- 《趣学 Python 编程》
- 《父与子的编程之旅：与小卡特一起学 Python 》
- 《Python 学习手册》
- 《深入浅出 Python》
- 《“笨办法”学 Python》
- 《像计算机科学家一样思考 Python》
- *Python Cookbook*

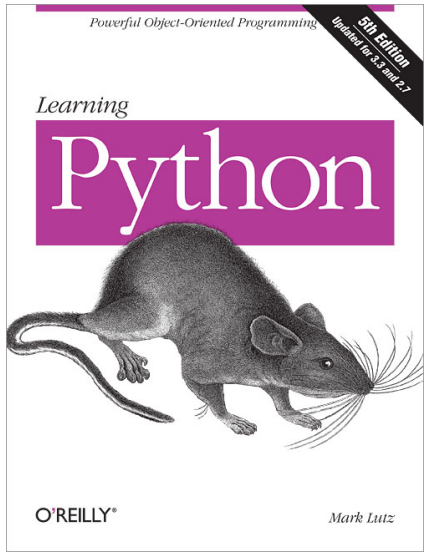
英文名

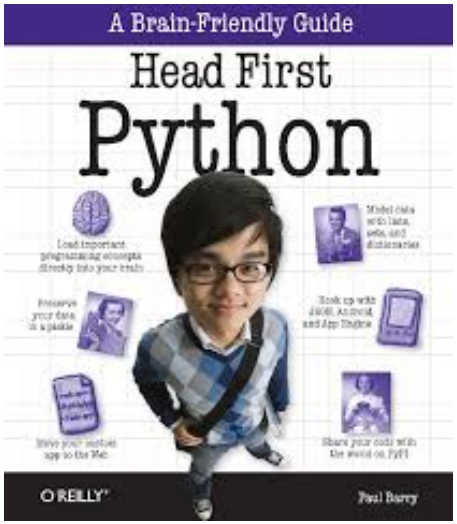
- *Python for Kids*
- *Computer Programming for Kids and Other Beginners*
- *Learning Python*
- *Head First Python*
- *Learn Python the Hard Way*
- *Think Python*
- *Python Cookbook*

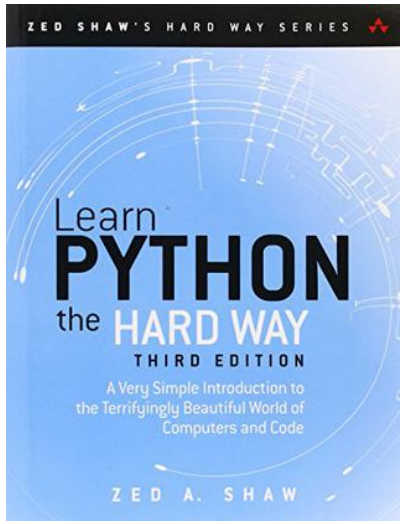
书籍 | Python


PYTHON
FOR KIDS
A PLAYFUL INTRODUCTION TO PROGRAMMING
JASON R. BRIGGS

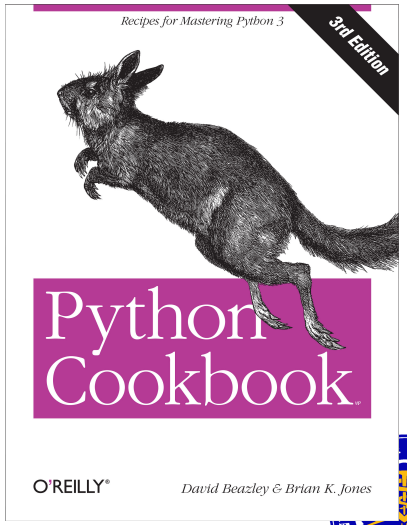

Hello World!
Computer Programming
for Kids and Other Beginners
Warren and Carter Sande
MANNING


Powerful Object-Oriented Programming
5th Edition
Learning
Python
O'REILLY
Mark Lutz


A Brain-Friendly Guide
Head First
Python
O'REILLY
Paul Barry


ZED SHAW'S HARD WAY SERIES
Learn
PYTHON
the HARD WAY
THIRD EDITION
A Very Simple Introduction to
the Terrifyingly Beautiful World of
Computers and Code
ZED A. SHAW


How to Think Like a Computer Scientist
Think
Python
O'REILLY
Allen B. Downey


Recipes for Mastering Python 3
3rd Edition
Python
Cookbook
O'REILLY
David Beazley & Brian K. Jones

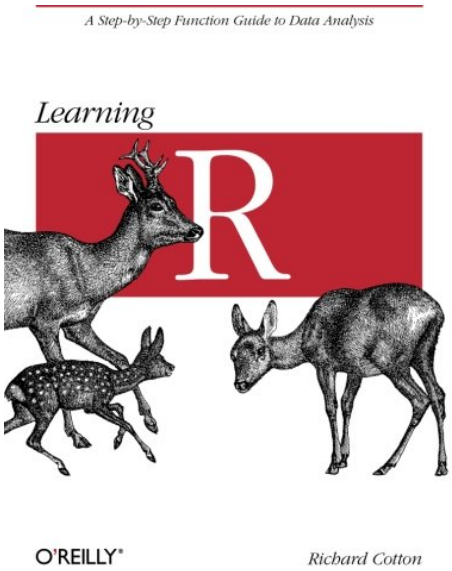
中文名

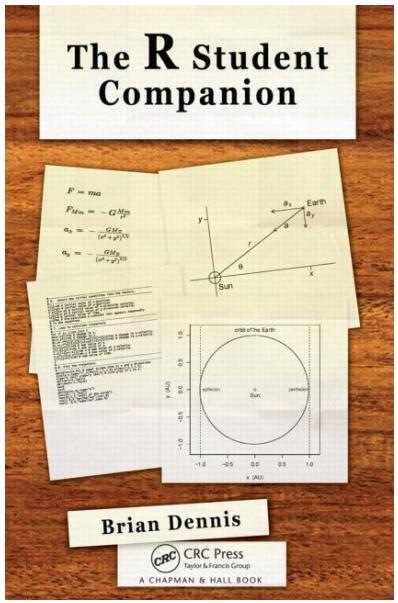
- 《学习 R》
- 《R 语言初学指南》
- 《R 语言实战》
- 《R 语言经典实例》
- 《R 数据可视化手册》

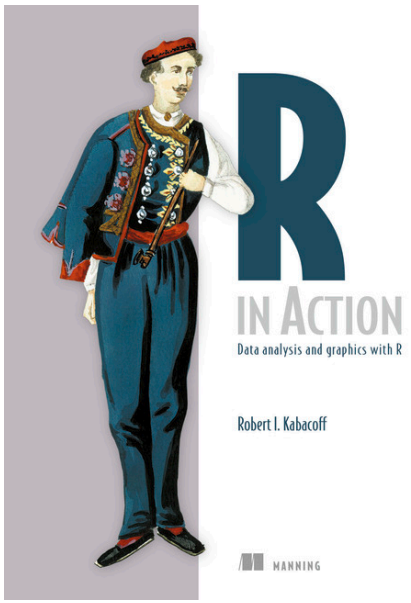
英文名

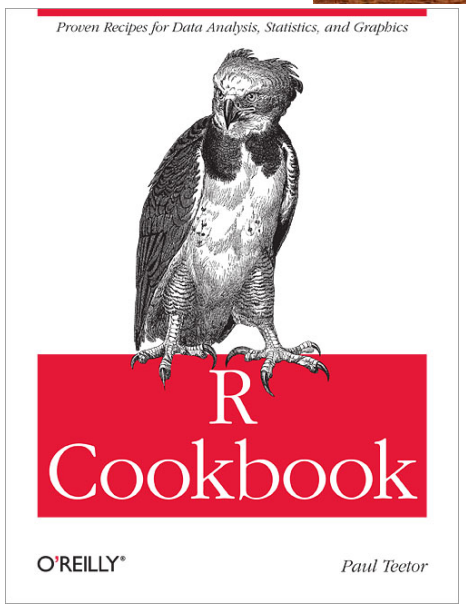
- *Learning R*
- *The R Student Companion*
- *R in Action*
- *R Cookbook*
- *R Graphics Cookbook*

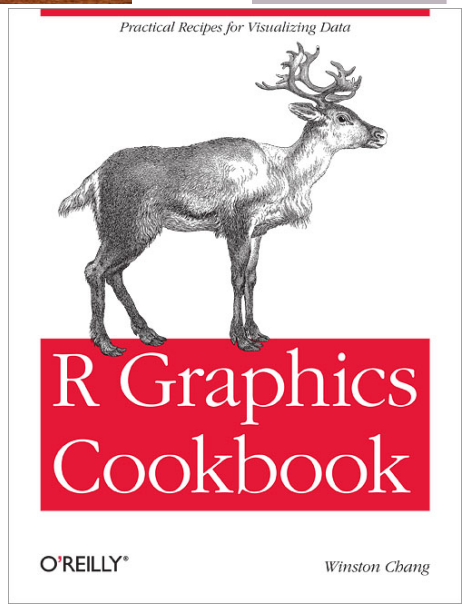
书籍 | R


A Step-by-Step Function Guide to Data Analysis
Learning
R
O'REILLY
Richard Cotton


The R Student Companion
Brian Dennis
CRC Press
A CHAPMAN & HALL BOOK


R
IN ACTION
Data analysis and graphics with R
Robert I. Kabacoff
MANNING


Proven Recipes for Data Analysis, Statistics, and Graphics
R
Cookbook
O'REILLY
Paul Teetor


Practical Recipes for Visualizing Data
R Graphics
Cookbook
O'REILLY
Winston Chang

计算机科学与程序设计

- 《我的第一本编程书》
- 《深入浅出程序设计》
- 《程序是怎样跑起来的》
- 《写给大家看的算法书》
- 《啊哈！算法》
- 《算法的乐趣》
- 《算法帝国》
- 《算法笔记》
- 《轻松学算法》
- 《大话数据结构》
- 《程序员的数学》
- 《程序员的数学 2：概率统计》
- 《程序员的数学 3：线性代数》
- 《统计思维：程序员数学之概率统计》
- 《程序员的数学思维修炼》

统计学与数据分析

- 《命令行中的数据科学》
- 《深入浅出数据分析》
- 《深入浅出统计学》
- 《白话统计学》
- 《爱上统计学》
- 《赤裸裸的统计学》
- 《介绍丛书: 统计学》
- 《从零开始读懂统计学》
- 《菜鸟侦探挑战数据分析》
- 《你一定爱读的极简统计学》
- 《漫画玩转统计学》
- 《漫画统计学》
- 《数字唬人》
- 《统计数字会撒谎》
- 《统计数据的真相》
- 《生活中的概率趣事》
- 《改变世界的 134 个概率统计故事》

教学提纲

教学提纲

知识点

- 生物信息学：交叉学科，多种技能，领域宽泛
- 生物学：DNA，RNA，蛋白质
- 计算机科学：三个术语，编程语言
- R：readr、dplyr、ggplot2 等常用包

教学提纲

1. 生物信息学主要是由哪些学科交叉而来的？
2. 生物信息学主要需要哪些方面的知识和技能？
3. DNA 是由哪四种碱基组成的？RNA 与之有何不同？
4. 列举常见的 20 种氨基酸，它们三字母和单字母的缩写分别是什么？
5. 列举常见的存储 DNA 序列和蛋白质结构的数据库。
6. *in vivo*、*in vitro* 和 *in silico* 分别代表什么含义？
7. 列举常见的编程语言，在生物信息学中常用的编程语言，专用于生物信息学的工具集。
8. 列举 R 的常用包并简介其主要用途。

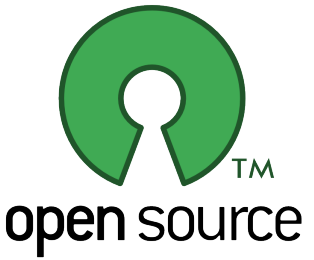
下节预告

Markdown

回顾、总结 Markdown 标记语言的基本语法：

- 标题
- 强调
- 列表
- 代码
- 引用
- 链接
- ……

Powered by



TEX LATEX XETEX Beamer