

# 生物信息学 (Bioinformatics)

## 第一章：概论



## 背景

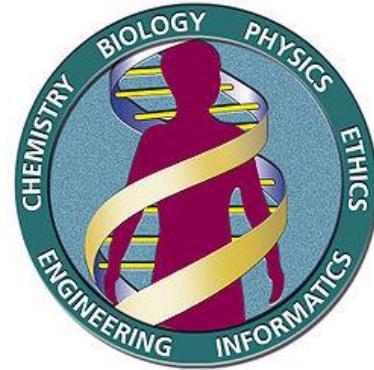
- 人类基因组计划(Human Genome Project, HGP): 1990年正式启动，旨在完成人类基因组约30亿个碱基的全序列测定。
- 海量生物数据的迅速膨胀: DNA、RNA和蛋白质序列，蛋白质二级结构和三维结构数据，蛋白质相互作用数据等。
- 对大量生物数据的管理、分析和信息化需求促进了生物信息学的迅速发展。

# 人类基因组计划

## HGP (Human Genome Project)



- 美国NIH和能源部提出和带头，美、英、德、法、日、中共同参与的国际合作项目。
- 重大国际研究项目：测定人类基因组全部DNA序列，构建人类基因组遗传图谱和物理图谱。
- 1990年：正式启动，30亿美元。
- 2001年：人类基因组草图公开发表。
- 2005年：测序工作已经基本完成。





# 生物信息学定义的历史演变

- 定义一：生物信息学是一门收集、分析遗传数据以及分发给研究机构的新学科（Bioinformatics is a new subject of genetic data collection, analysis and dissemination to the research community）。(Dr. Hwa A. Lim, 1987)
- 定义二：生物信息学特指数据库类的工作，包括持久稳固的在一个稳定的地方提供对数据的支持（Bioinformatics refers to database-like activities, involving persistent sets of data that are maintained in a consistent state over essentially indefinite periods of time）。(Dr. Hwa A. Lim, 1994)
- 定义三：生物信息学是在大分子方面的概念型的生物学，并且使用了信息学的技术，这包括了从应用数学、计算机科学以及统计学等学科衍生而来各种方法，并以此在大尺度上来理解和组织与生物大分子相关的信息。 (Luscombe,2001)

# 什么是生物信息学？

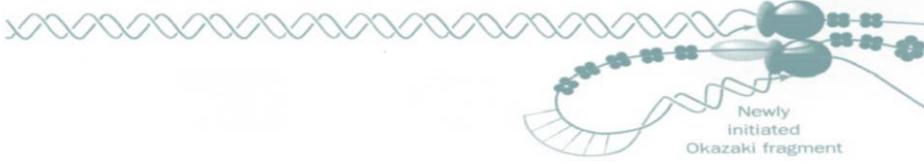


□ Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned.

Biology in the 21st century is being transformed from a purely lab-based science to an information science as well.

from NCBI's science primer

[www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html](http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html)



# 什么是生物信息学？

□ 生物信息学是一门交叉学科，它包含了生物信息的获取、处理、存储、分发、分析和解释等在内的所有方面，它综合运用数学、计算机科学和生物学的各种工具，来阐明和理解大量数据所包含的生物学意义。

(美国人类基因组计划第一个五年总结报告，1995)

□ 生物信息学最重要的任务：

**数据 → 信息 → 知识**

# 广义生物信息学观点

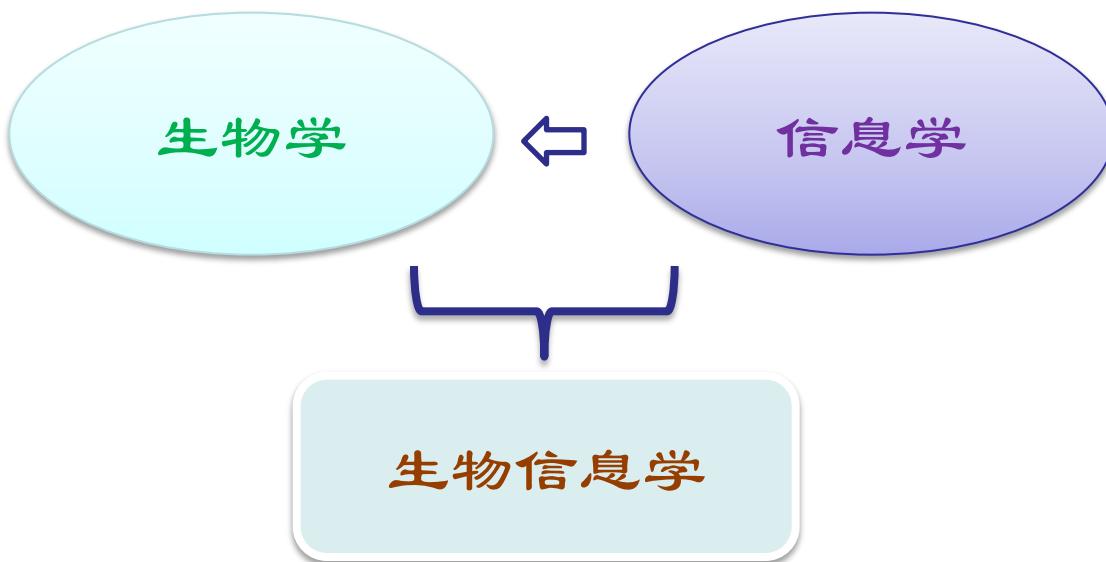


- Biology may be viewed as the study of transmission of information: from mother cell to daughter cell, from one cell or tissue type to another, from one generation to the next, and from one species to another. This informational viewpoint is termed bioinformatics.
- 生物学研究可以被看成是研究信息的传递：从DNA经转录翻译到蛋白质，从细胞质中到细胞核内，从母细胞到子细胞，从一个细胞或一个组织到另一个细胞或另一个组织，从一代到下一代，从一个物种到另一个物种的进化演变。这种信息论的观点即可称为生物信息学。

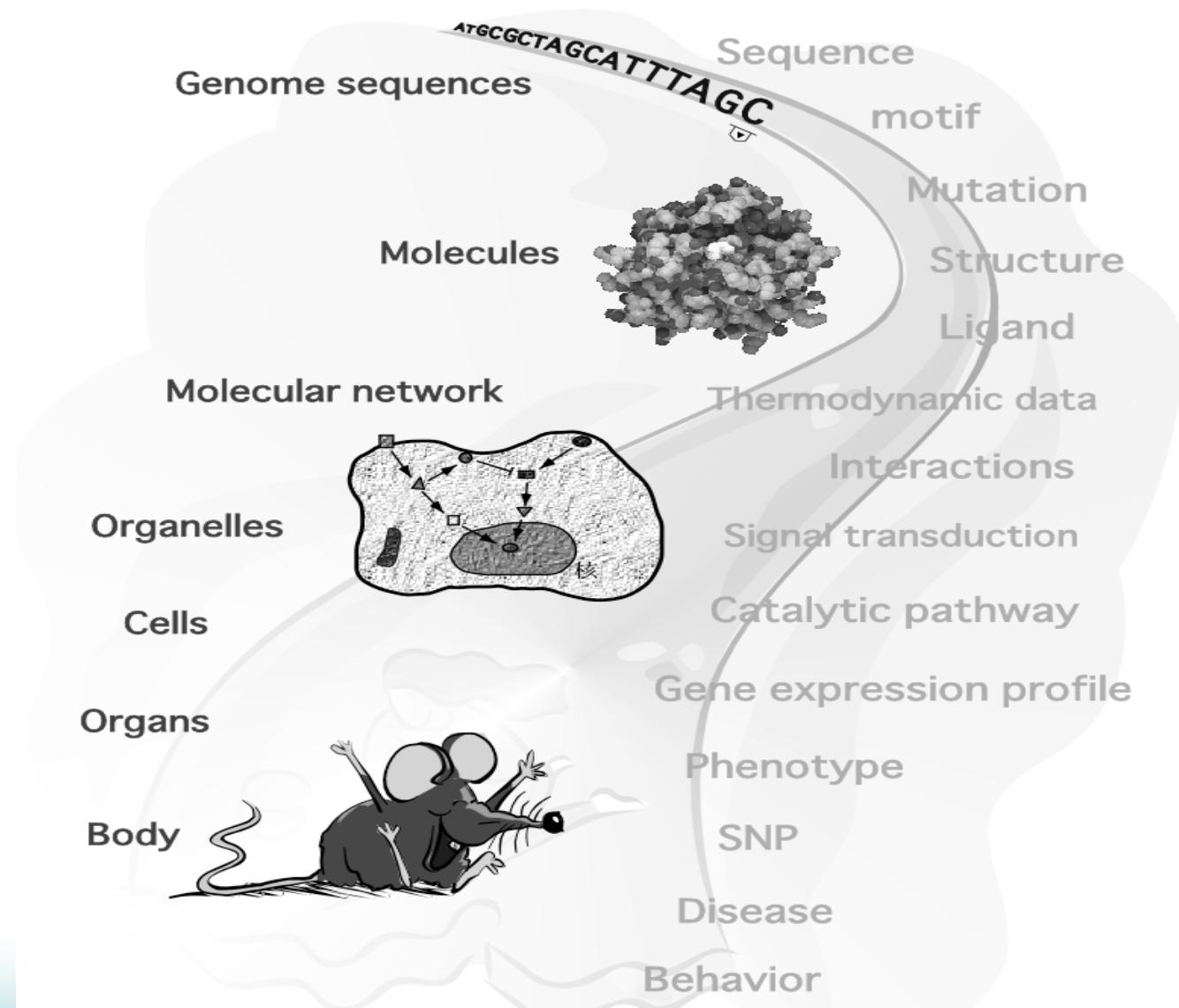
(Bioinformatic challenges for the next decade(s), David Eisenberg et al., 2006)



- 口 生物数据的获取、处理、存储、分发、分析和解释等在内的所有方面；
- 口 生物体内外的信息/信号的传递；



# Roadmap of Bioinformatics



# 生物信息学的主要研究内容



- 生物信息的存储与数据库检索；
  - 序列比对；
  - 序列模式识别；
  - 基因预测及基因组分析；
  - 分子进化与系统发育分析；
  - RNA结构预测；
  - 蛋白质结构分析与预测；
  - 生物芯片；
  - 生物网络；
- ...

# 生物信息学的发展历程



- 1952年，Sanger根据胰岛素蛋白质的测序结果，推断蛋白质是排列完美的分子。
- 1955年，Sanger与合作者分别对牛、猪和羊的胰岛素蛋白质进行了测序并做了序列上的比较。-最早的序列比对。
- 1962年，鲍林提出分子进化的理论，推测在人中可能存在50,000~100,000个不同的基因/蛋白质。-分子进化理论的奠定。
- 1965年，Margaret Dayhoff构建蛋白质序列图谱
- 1970年，Needleman-Wunsch算法：全局优化比对。
- 1981年，Smith-Waterman算法开发：局部优化比对。
- 1990年，快速序列相似性搜索工具BLAST的开发



# 生物信息学发展过程中的里程碑性事件

时间	重大事件
1962	鲍林的分子进化理论
1965	Margaret Dayhoff 构建蛋白质序列图谱
1970	Needleman-Wunsch 算法开发
1977	DNA 测序以及使用利用 Staden 软件进行分析
1981	Smith-Waterman 算法开发
1981	序列模体的概念 (motif)
1983	序列数据库搜索算法的开发 (Wilbur-Lipman)
1985	快速序列相似性搜索工具 FASTP/FASTN 的开发
1990	快速序列相似性搜索工具 BLAST 的开发
1996	酵母基因组的完全测序
1997	PSI-BLAST 工具的开发
1998	秀丽线虫基因组的完全测序
1999	果蝇基因组的完全测序

# 早期的发展阶段示意图



	large-scale data	computational tools	databases of biological knowledge
1990	ESTs	BLAST	
1995	complete genomes gene expressions SNPs	CLUSTAL W HMM PSI-BLAST	SCOP KEGG COG Pfam SMART
2000	protein interactions		GO



# 最早的序列分析：胰岛素蛋白质

Insulin Chain A: 8-10位存在着不同（牛， ASV；猪， TSI；羊， AGV）(Brown *et al.*, 1955)。

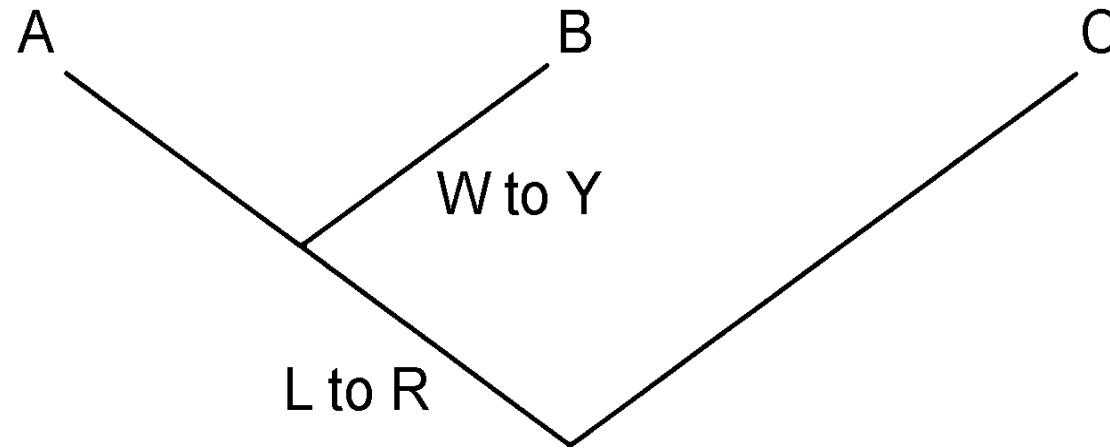
Bovine	GIVEQCC <b>A</b> S <b>V</b> CSLYQLENYCN
Pig	GIVEQCC <b>T</b> S <b>I</b> CSLYQLENYCN
Sheep	GIVEQCC <b>A</b> G <b>V</b> CSLYQLENYCN
Human	GIVEQCC <b>T</b> S <b>I</b> CSLYQLENYCN

Made by GeneDoc

# 不同物种的系统发育分析



ORGANISM A	A	W	T	V	A	S	A	V	R	T	S	I
ORGANISM B	A	Y	T	V	A	A	A	V	R	T	S	I
ORGANISM C	A	W	T	V	A	A	A	V	L	T	S	I



# 80年代：DNA序列数据库

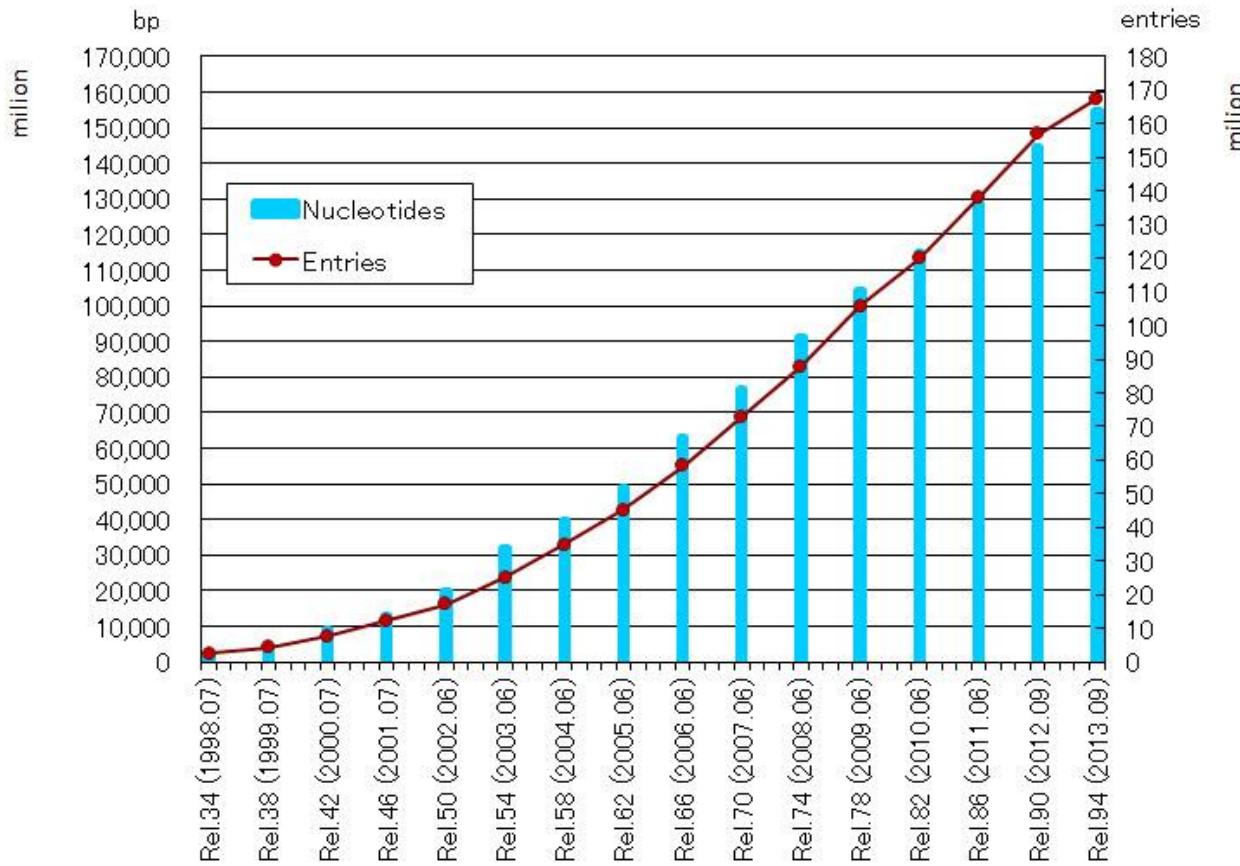


- 1974年，George I.Bell等人收集DNA序列，构建GenBank数据库。1982~1992开发第一个版本。
- 1980年，EMBL数据库成立。
- 1984年，日本DDBJ数据库成立。
- 核酸序列数据的去冗余：Refseq数据库，对于相同的序列只列一条目录。

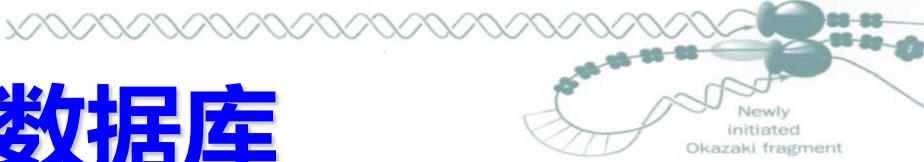
# 核酸数据库数据的增长



DDBJ/EMBL/GenBank database growth



Note: CON division is not counted in statistics of DDBJ periodical



# 获取序列及检索公共数据库

- NCBI：可通过Entrez系统进行检索。
- 提供关键字的搜索的方法。
- “硬搜索”：包含关键字的，完全匹配的结果。
- “软搜索”：与查询内容相关的信息。
- 查询内容：基因/蛋白质的名称、标识符，文献、蛋白质结构，等等。



*Entrez, The Life Sciences Search Engine.*

HOME | SEARCH | SITE MAP

PubMed

All Databases

Human Genome

GenBank

Map Viewer

BLAST

Search across databases

GO

Clear

Help

### Welcome to the Entrez cross-database search page



**PubMed:** biomedical literature citations and abstracts



**Books:** online books



**PubMed Central:** free, full text journal articles



**OMIM:** online Mendelian Inheritance in Man



**Site Search:** NCBI web and FTP sites



**OMIA:** online Mendelian Inheritance in Animals



**CoreNucleotide:** Core subset of nucleotide sequence records



**dbGaP:** genotype and phenotype



**EST:** Expressed Sequence Tag records



**UniGene:** gene-oriented clusters of transcript sequences



**GSS:** Genome Survey Sequence records



**CDD:** conserved protein domain database



**Protein:** sequence database



**3D Domains:** domains from Entrez Structure



**Genome:** whole genome sequences



**UniSTS:** markers and mapping data



**Structure:** three-dimensional macromolecular structures



**PopSet:** population study data sets



**Taxonomy:** organisms in GenBank



**GEO Profiles:** expression and molecular abundance profiles



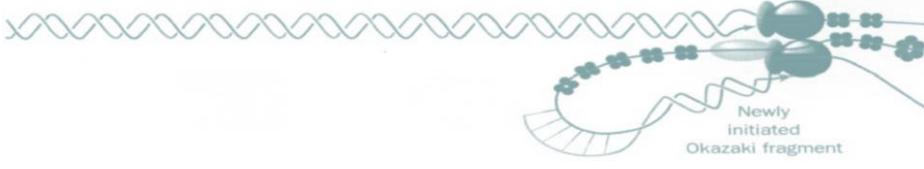
**SNP:** single nucleotide polymorphism



**GEO DataSets:** experimental sets of GEO data

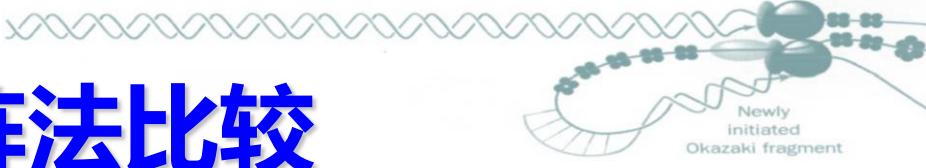


<http://www.ncbi.nlm.nih.gov/sites/gquery>

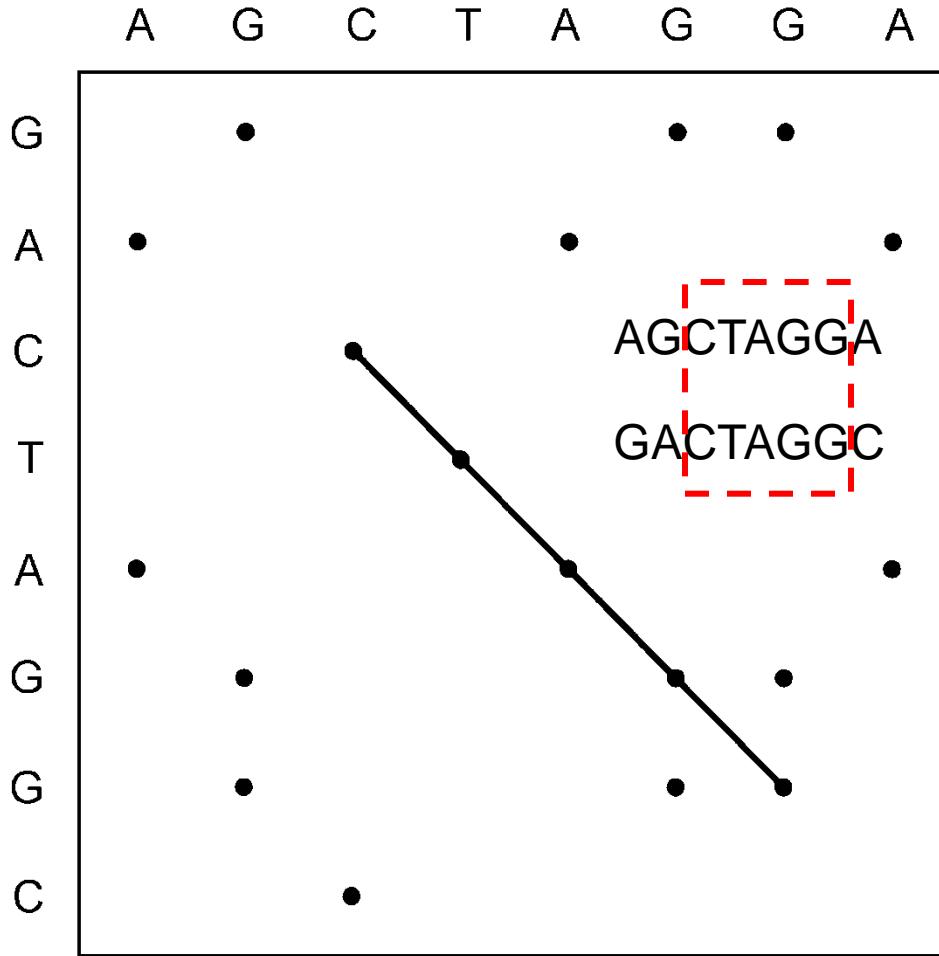


# 序列比对工具的开发

- 1970年，Gibbs AJ 和 McIntyre GA，点阵法进行氨基酸和核酸的序列比较：当相同的字母在两条序列中同时出现时，在交叉处置点。
- 1970年，Needleman-Wunsch，全局优化的序列比对算法：允许匹配、错配和缺失。动态规划的算法：任务可分割，分成更小的子问题进行解决。
- 1981年，Smith-Waterman，局部优化的序列比对算法。
- FASTA & BLAST的开发，启发式优化算法。
- 多序列比对：CLustalW/X, POA, MUSCLE.



# 两条DNA序列的点阵法比较



# 全局优化 vs. 局部优化



ACTGTTCCGAA... ...100kbp.....AGCCTGA... ...100kbp.....ACTACTG  
ACGCCTG

## 全局优化

ACTGTTCCGAA... ...100kbp.....AGCCTGA... ...100kbp.....ACTACTG

AC-----GCC-----TG

## 局部优化

ACTGTTCCGAA... ...100kbp.....A-GCCTGA.....100kbp.....ACTACTG

ACGCCTG

# Needleman-Wunsch 算法



	G	A	T	C	T	A	
G	1						
A		2				1	GATCTA
T			3		1		GATCA
C				4			
A		1			5	(minus gap penalty)	

Deduced alignment with gap  $\Delta$

G	A	T	C	T	A	
G	A	T	C	$\Delta$	A	



# 数据库中搜索相似序列

□通过搜索数据库中相似序列发现基因功能。

- 例如反转录病毒编码的致癌基因 **v-sis**和**v-src**
- 通过和模式生物已知遗传或生化信息的基因序列进行相似性搜索，预测新基因功能。

□FASTA和BLAST

- FASTA**:以几个残基长度的'word'为单元进行检索； **W. Pearson**和**D. Lipman**开发。
- BLAST**:应用最广泛的序列相似性搜索工具，相比**FASTA**有更多改进，速度更快。
  - ◆ **PSI-BLAST**: 位点特异性迭代**BLAST**
  - ◆ **PHI-BLAST**: 模式发现迭代**BLAST**



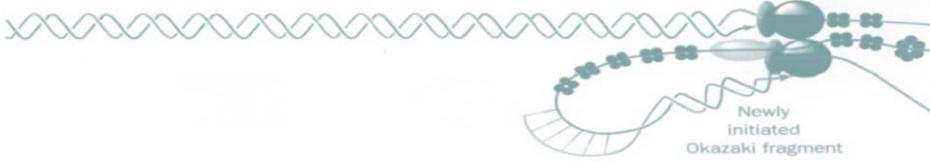
# 全局性的多序列比对

DmPka-C2	100	ARFPFLIYLVDSTKCFDYLYLIIPLVNGGELFSYHRRVRKFNEKHARFYAAQVALALEYMHKMHLMYRD	168
DmCG12069	102	MTFPNTVNLIASYKDEDSLYLVIPLIGGELFTYHRKVRKFTEKQARFYAAQVFLALEYLHHCSLLYRD	170
SctPK2	125	VEHPFLIRMWGTFQDARMIFMVMMDYIEGGELFSLLRKSQRFNPVPAKFYAAEVILALEYIHHMIIYRD	193
SctPK1	142	VTHPFIIRMWGTFQDAQQIFMIMDYIEGGELFSLLRKSQRFNPVPAKFYAAEVCLALEYIHSKEIIYRD	210
SctPK3	143	VSHPFIIRMWGTFQDSQQVFMVMDYIEGGELFSLLRKSQRFNPVPAKFYAAEVCLALEYIHSKEIIYRD	211
Cekin-1	136	IEFPFLVNMTFSFKDNNSLYMVLFISGGEMFSHLRRIGRFSEPHSRFYAAQIVLAFEYIHSLELIYRD	204
DmPka-C1	101	IQFPFLVSIHYHFKDMSLYMVLEYVPGGEMFSHLRKVGRFSEPHSRFYAAQIVLAFEYIHYLDLIYRD	169
HsPKACg	99	IEFPFLVKLQFSFKDNNSLYLVMEYVPGGEMFSRLQRVGRFSEPHACFYAAQVVIAVQYIHSLELIHRD	167
HsPKACa	99	VNFPFLVKLEFSFKDNNSLYMVMEMYVPGGEMFSHLRRIGRFSEPHARFYAAQIVLTFEYIHSLELIYRD	167
HsPKACb	99	VNFPFLVRLEYAFKDMSLYMVMEMYVPGGEMFSHLRRIGRFSEPHARFYAAQIVLTFEYIHSLELIYRD	167
DmPKA-C3	329	IRHPFVISLEWSTKDDSNLYMIFDYVCGGELEFTYLRNAGKFTSQTSNFYAAEIVSALEYIHSLGIVYRD	397
HsPRKX	104	VSHPFLIRLFWTWHDERFLYMLMEYVPGGELFSYLRNRGRFSSTTGLFYSAEIIICATEYIHSKEIVYRD	172
HsPRKY	104	VSHPFLIRLFWTWHEERFLYMLMEYVPGGELFSYLRNRGHFSSTTGLFYSAEIIICATEYIHSKEIVYRD	172

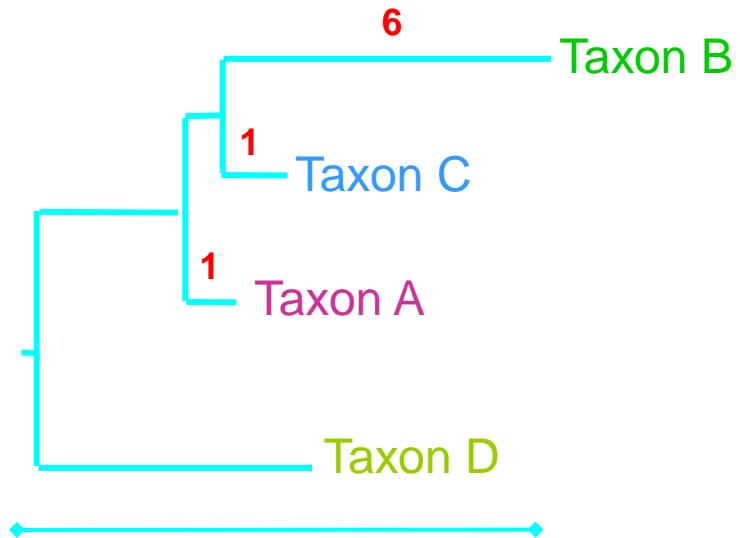


# 基于序列信息研究分子进化

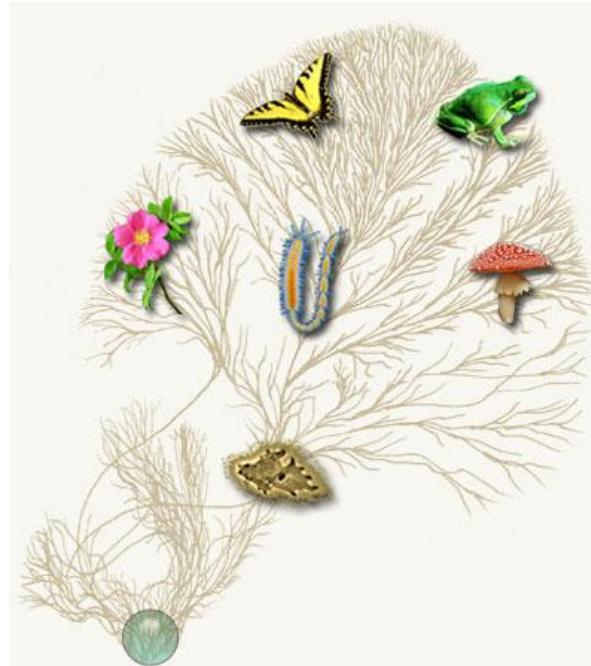
- 构建进化树，分析蛋白质的超家族及亚家族分类。
- 寻找Ortholog (直系同源物)或者Paralog (旁系同源物)。
- 分子进化树的构建方法：邻接法(Neighbor-Joining), 最大简约法(Maximum Parsimony), 最大似然性法(Maximum Likelihood), 以及贝叶斯类算法(MCMC)。
- 构建进化树的第一步：可靠的多序列比对。



# 构建进化树



遗传变化



Tree of Life



# 模式识别

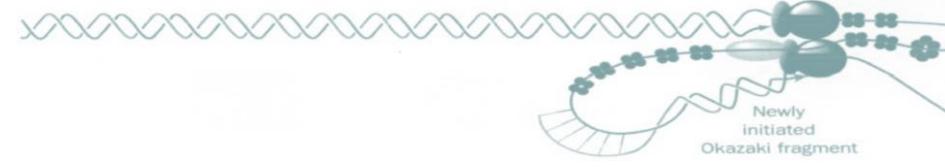
EcORF708	EEVI-----	AALEEGEFDLAIGLSVFFHIVHLHGIDEVKRLLSRLADVTQAVILELAVKEE
DmPka-C2	-----	TMLENYITRAVLGNNSFGTVMLVREKSGKNYYAAKMIISKEDLVRLK-----
DmCG12069	-----	TGLDDYEIKATLGSGSFGRVQLVRERESGWYYASKQLSKDQIWKTK-----
ScTPK2	-----	YTLDHFQIMRTLGTGSFGRVHLVRSVHNGRYYAIKVLLKKQQVVKMK-----
ScTPK1	-----	YSLQDFQILRTLGTGSFGRVHLIRSRHNGRYYAMKVLKKEIVVRLK-----
ScTPK3	-----	YSLSDFQILRTLGTGSFGRVHLIRSNHNGRFYALKTLKKHTIVKLK-----
Cekin-1	-----	ACQDDDFDRIKTLGTGSFGRVMLVVKHKQSGNYYAMKILDQKQVVKLK-----
DmPka-C1	-----	AAUDDFFERIKTLGTGSFGRVMLVVKHKQPTKDYYAMKILDQKQVVKLK-----
HsPKACg	-----	ASSDQFERLRTLGMGSFGRVMLVRHQETGGHYAMKILNKQKVVVKMK-----
HsPKACa	-----	AHLDQFERIKTLGTGSFGRVMLVVKHKETGNHYAMKILDQKQVVKLK-----
HsPKACb	-----	AGLEDFERKKTLGTGSFGRVMLVVKHKATEQYYAMKILDQKQVVKLK-----
DmPKA-C3	DGNETDDEDDDESEESSVQTAKGVRKYHLDYQIIKTVGTGTFGRVCLCRDRISEKYCAMKILAMTEVIRLK	-----
HsPRKX	-----	YSLQDFDTLATVGTGTFGRVHLVKEKTAKHFFALKVMSIPDVIRLK-----
HsPRKY	-----	YRLQDCDALVMTGTFGRVHLVKEKTAKHFFALKVMSIPDVIRRK-----

**模块/BLOCK**

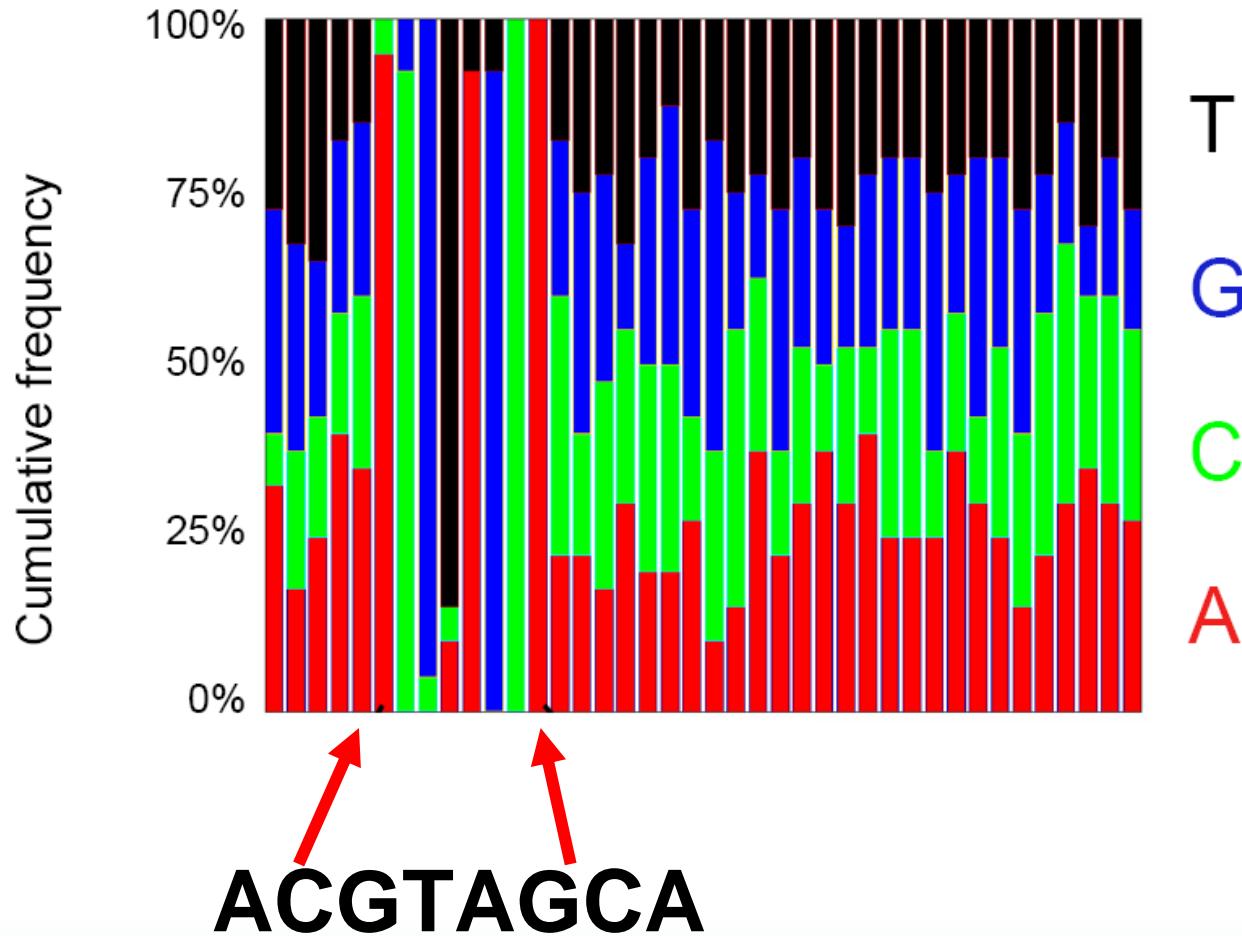
□SUMO化的序列模体:

**Ψ-K-X-E** (Ψ:A, I, L, V, M, F, P; X:任意氨基酸)

**模体/Motif**



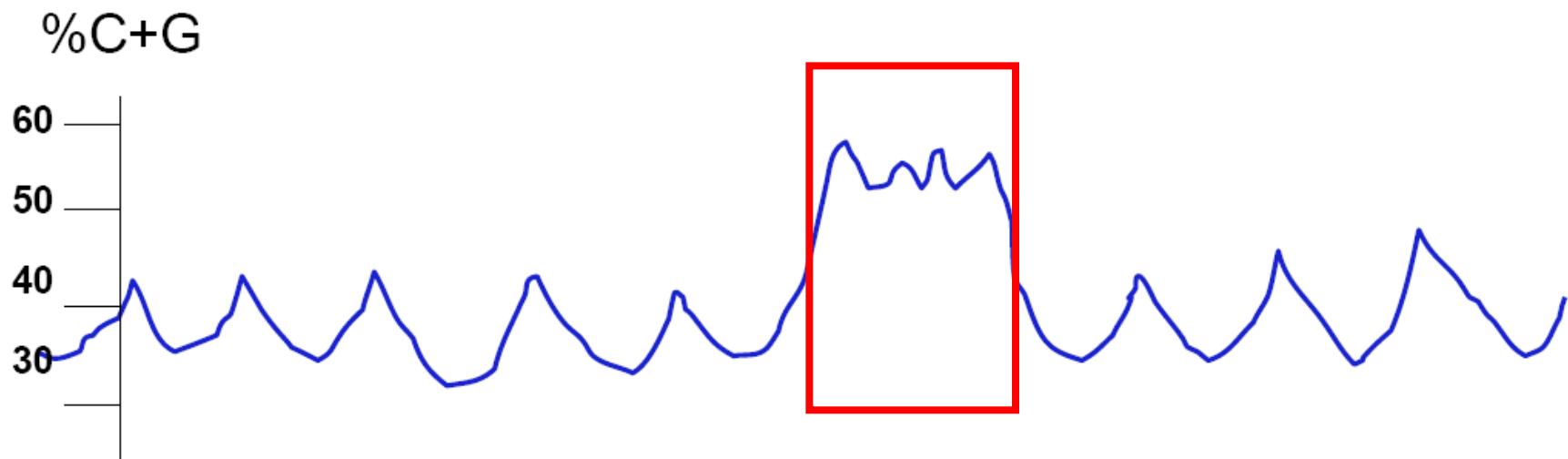
# Strong Motif





# CpG岛 : HMM

□ 存在两种状态：是CpG岛 (CpG Island, I),  
不是CpG岛 (Genome, G)





# 基因组大小 & 基因数

Sequenced genomes vary from 470-30,000 genes

Species	Genome (Mb)	Genes	Lethal loci
<i>Mycoplasma genitalium</i>	0.58	470	~300
<i>Rickettsia prowazekii</i>	1.11	834	
<i>Haemophilus influenzae</i>	1.83	1,743	
<i>Methanococcus jannaschi</i>	1.66	1,738	
<i>B. subtilis</i>	4.2	4,100	
<i>E. coli</i>	4.6	4,288	1,800
<i>S. cerevisiae</i>	13.5	6,034	1,090
<i>S. pombe</i>	12.5	4,929	
<i>A. thaliana</i>	119	25,498	
<i>O. sativa</i> (rice)	466	~30,000	
<i>D. melanogaster</i>	165	13,601	3,100
<i>C. elegans</i>	97	18,424	
<i>H. sapiens</i>	3,300	~30,000	

Minimum gene numbers range from 500 to 30,000

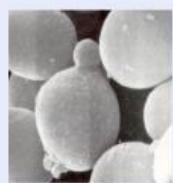
500 genes  
Extracellular (parasitic) bacterium



1,500 genes  
Free-living bacterium



5,000 genes  
Unicellular eukaryote



13,000 genes  
Multicellular eukaryote



25,000 genes  
Higher plants

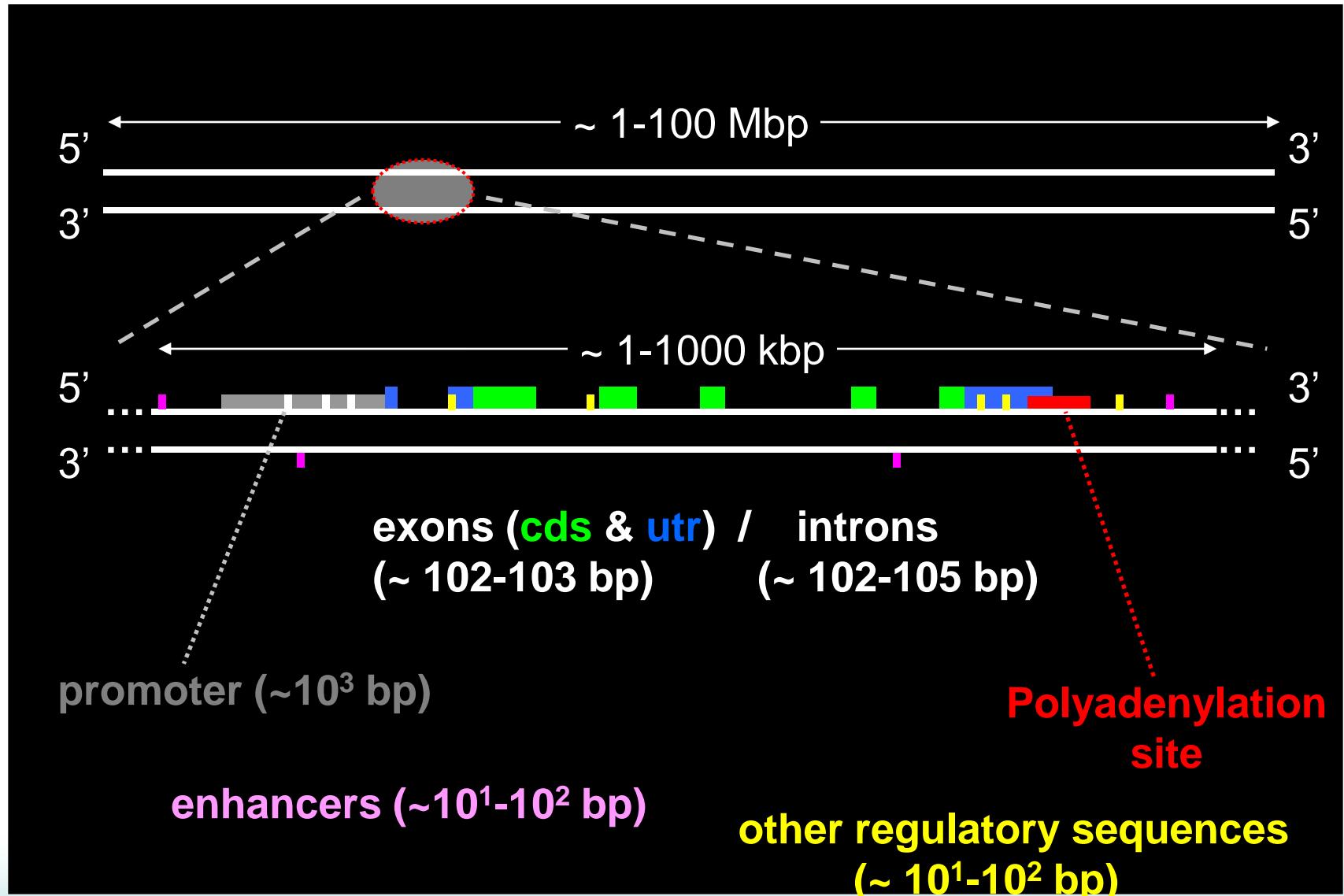


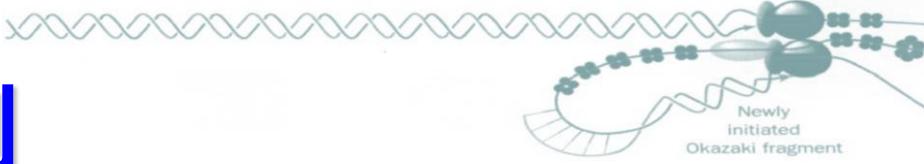
30,000 genes  
Mammals





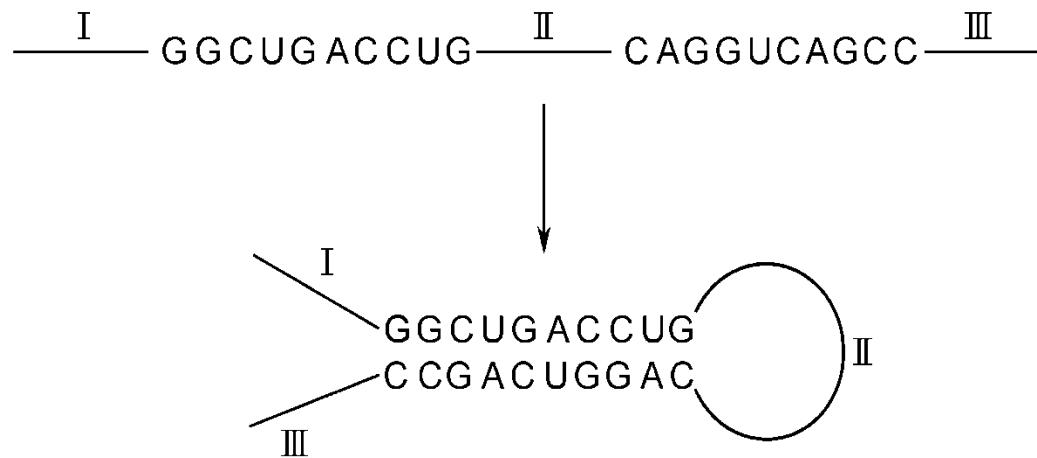
# 真核生物的基因结构





# RNA二级结构的预测

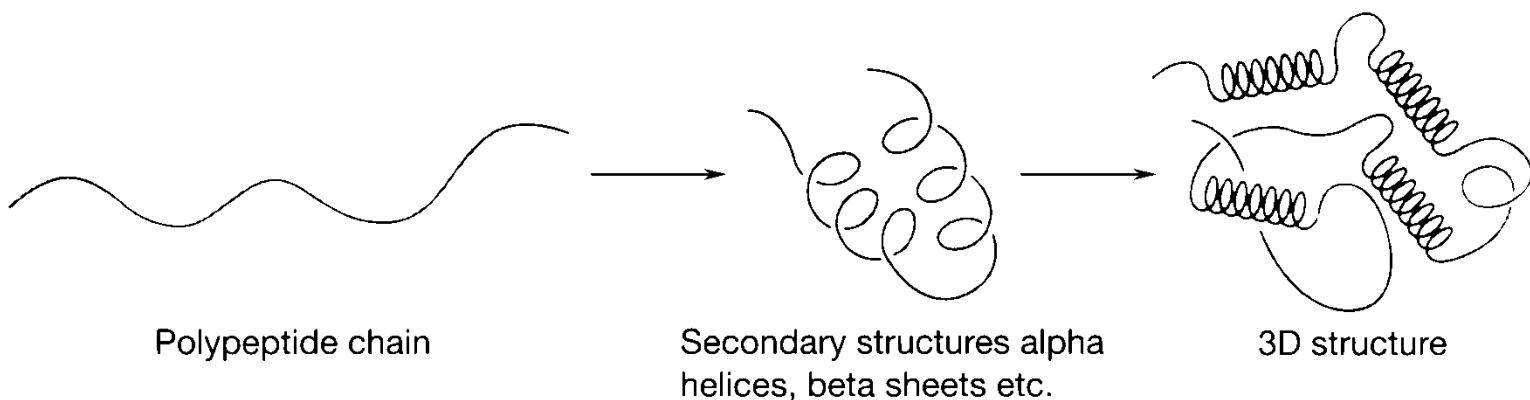
- RNA分子中，如果存在重复且反向互补，则可以形成发卡结构。
- 动态规划法：最多配对的碱基对。
- 应用：构建RNA分子数据库。



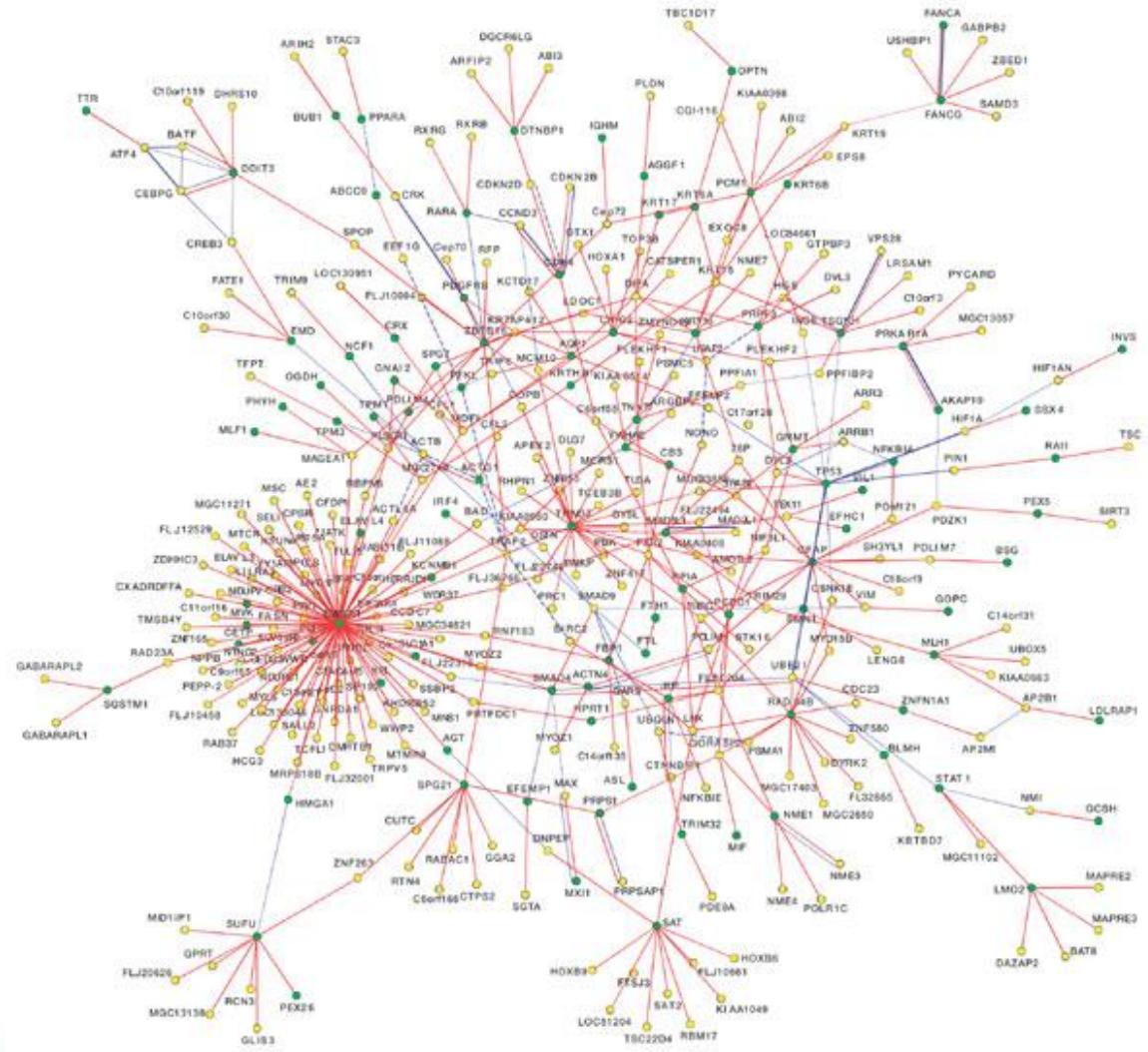
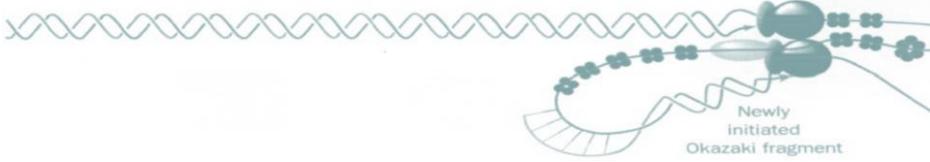


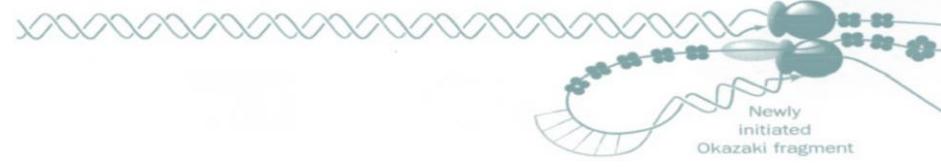
# 蛋白质结构预测

- 大量序列已知 vs . 少量结构解析
- 实验方法测定：X射线晶体学，核磁共振等
- 三级结构主要由一级序列决定。
- 蛋白质折叠的类型相对有限 (~1,000)。
- 蛋白质二级结构的预测 -> 三级结构预测



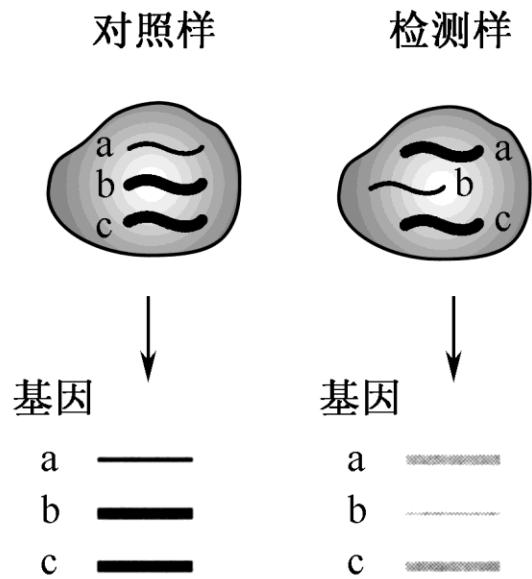
# 蛋白质相互作用网络



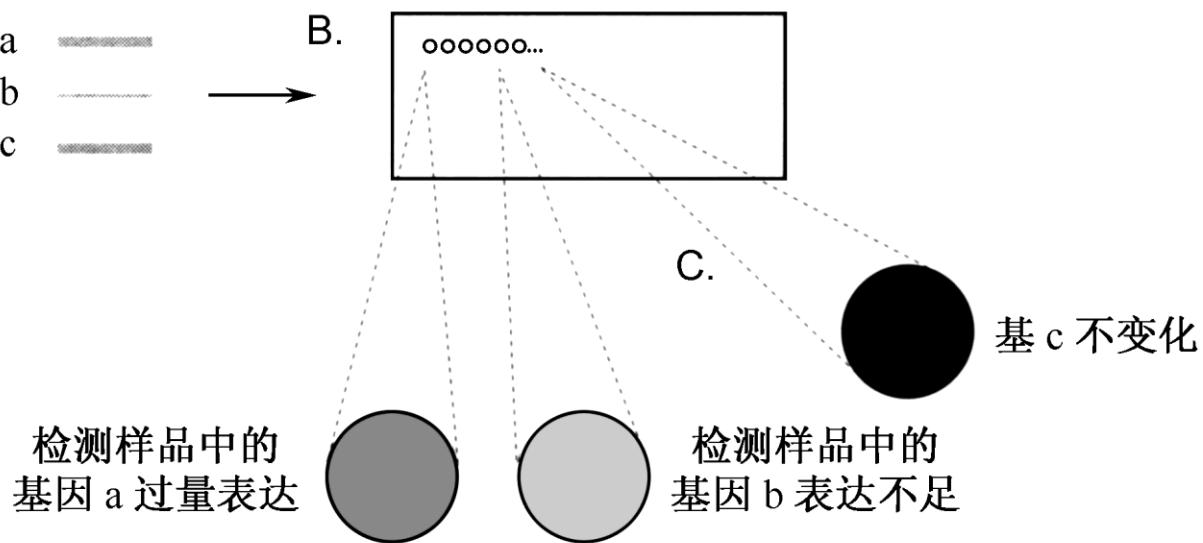


# 基因芯片

A.



B.



# 生物信息学的过去、现在和未来



## bioinformatics in the 90s

primary databases containing large-scale experimental data in genomics and proteomics



understanding functions and utilities of individual genes or proteins

## bioinformatics now

secondary databases (knowledge bases) containing accumulated biological knowledge



understanding functions and utilities at the molecular, cellular and organism levels

## bioinformatics in the future

a complete computer representation of the cell and the organism



understanding basic principles of the higher complexity of biological systems

# 生物信息学：学科交叉



bioinformatics now

mathematics

informatics

physics

biology

chemistry

medicine

bioinformatics in the future

mathematics

informatics

physics

biology

chemistry

medicine



# 生物信息学的相关知识储备

- 生物学背景：e.g., 分子生物学、细胞生物学、发育生物学、生物化学，...
- 数学知识：概率论与统计学，...
- 算法及编程能力：JAVA, Perl/Python, PHP+MySQL, ...



# 生物信息学的常用算法与方法

- 动态规划算法(Dynamic programming);
- 贝叶斯统计(bayesian statistic);
- 人工神经网络(ANNs);
- 马尔可夫模型和隐马尔科夫模型(HMM);
- 遗传算法(Genetic Algorithm);
- 蒙特卡洛方法(Monte Carlo);
- 模拟退火算法(Simulated Annealing);
- 支持向量机(SVM);

...



# 科研机构及网络资源中心

□ NCBI：美国国立卫生研究院NIH下属国立生物技术信息中心NCBI。

<http://www.ncbi.nlm.nih.gov>

□ EMBL-EBI：欧洲分子生物学实验室下属欧洲生物信息学研究所。 <http://www.ebi.ac.uk/>

□ EMBnet：欧洲分子生物学网络

<http://www.embnet.org/>

□ ExPASy: (Expert Protein Analysis System)瑞士生物信息研究所SIB下属的蛋白质分析专家系统； <http://www.expasy.org/>

# 科研机构及网络资源中心



## □ Bioinformatics Links Directory :

[http://bioinformatics.ca/links\\_directory/](http://bioinformatics.ca/links_directory/)

## □ 各种数据库等

如 GenBank, PDB (Protein Data Bank)

UniProt 数据库等

## □ 软件资源：

<http://www.expasy.org/tools/>

<http://www.ebi.ac.uk/Tools/>

<http://www.ncbi.nlm.nih.gov/Tools/>



# 国内生物信息中心举例

□CBIPKU：北京大学生物信息中心

[www.cbi.pku.edu.cn/chinese/](http://www.cbi.pku.edu.cn/chinese/)

□BioSino：中国生物信息

<http://www.biosino.org/>

中国科学院上海生命科学院生物信息中心

□上海生物信息技术研究中心

<http://www.scbit.org/>

# 生物信息学的相关杂志



生物信息学相关期刊名称	网址
Bioinformatics	<a href="http://bioinformatics.oxfordjournals.org/">http://bioinformatics.oxfordjournals.org/</a>
BMC Bioinformatics	<a href="http://www.biomedcentral.com/bmcbioinformatics/">http://www.biomedcentral.com/bmcbioinformatics/</a>
Genome Biology	<a href="http://genomebiology.com/">http://genomebiology.com/</a>
Genome Research	<a href="http://www.genome.org/">http://www.genome.org/</a>
Nucleic Acids Research	<a href="http://nar.oxfordjournals.org/">http://nar.oxfordjournals.org/</a>
Briefings in Bioinformatics	<a href="http://www.henrystewart.com/briefings_in_bioinformatics/">http://www.henrystewart.com/briefings_in_bioinformatics/</a>
FEBS letters	<a href="http://www.febsletters.org/">http://www.febsletters.org/</a>
Biochemical and Biophysical Research Communications	<a href="http://www.sciencedirect.com/science/journal/0006291X">http://www.sciencedirect.com/science/journal/0006291X</a>
Molecular Systems Biology	<a href="http://www.nature.com/msb/index.html">http://www.nature.com/msb/index.html</a>
Molecular Biology and Evolution	<a href="http://mbe.oxfordjournals.org/">http://mbe.oxfordjournals.org/</a>
PLoS Computational Biology	<a href="http://www.ploscompbiol.org/">http://www.ploscompbiol.org/</a>
PLoS ONE	<a href="http://www.plosone.org/">http://www.plosone.org/</a>
Protein Science	<a href="http://www.proteinscience.org/">http://www.proteinscience.org/</a>
Proteins	<a href="http://www3.interscience.wiley.com/cgi-bin/jhome/36176">http://www3.interscience.wiley.com/cgi-bin/jhome/36176</a>
Protein Engineering Design and Selection	<a href="http://peds.oxfordjournals.org/">http://peds.oxfordjournals.org/</a>

# 生物信息学研究的一般步骤



1. 确立研究的生物学体系。例如：生物芯片数据分析；蛋白质三级结构与功能；
2. 确定研究的问题。已有哪些计算方面的工作？是否需要实验的支持？
3. 构建生物学/数学模型，例如：ligand结合位点的预测，构建特异性识别位点的结构模式模型。
4. 计算方法的选择或开发：HMM, SVM, ANN或新方法。
5. 计算结果分析，与同类工具做比较。构建相应的数据库/软件/在线网站等。
6. 扩展及应用：有哪些用处？



# 计算生物学 vs. 生物信息学

- 计算生物学（Computational Biology）：  
根据美国国家卫生研究所（NIH）的定义，  
它是指开发和应用数据分析及理论的方法、  
数学建模和计算机仿真技术，用于生物学、  
行为学和社会群体系统的研究的一门学科。
- 生物信息学主要侧重于对生物学信息的存贮、  
分析处理、有效信息和知识的提取方面。
- 计算生物学侧重于使用计算技术研究生物学  
问题。



# “生物信息学和计算生物学”计划

□ NIH于2003年形成了一个通向生命科学未来的“中长期发展规划”--国立卫生研究院路线图（NIH Roadmap）。NIH路线图中启动了一个“生物信息学和计算生物学”计划，希望通过这个项目的实施而铺设一条通向生命科学未来的“信息高速公路”。该项目计划从2004年开始，建立数个“国立生物医学计算中心”，以便开发相关软件和数据管理工具。



# 本课程特点及注意事项

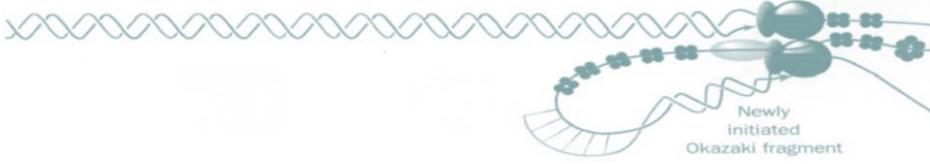
- 多种算法讲解及实例计算：  
要能保证平时练习，加强理解及时巩固！
- 课后作业：开放选题，计入平时成绩
- 实验环节：注重应用、实践，与实际结合
- 知识性、理解、灵活运用

到课听讲，随堂消化，重在平时，切勿压题！



# 参考教材

- 生物信息学：序列与基因组分析，(第二版)，  
David W.Mount，科学出版社
- 生物信息学， DRWesthead, JHParish,  
RMTwymann，科学出版社



课件资料下载：

<http://staff.ustc.edu.cn/~zhaowei/>

赵伟 [zhaowei@ustc.edu.cn](mailto:zhaowei@ustc.edu.cn)

张志勇 [zzyzhang@ustc.edu.cn](mailto:zzyzhang@ustc.edu.cn)

助教：黄斌 [hb080@mail.ustc.edu.cn](mailto:hb080@mail.ustc.edu.cn)

本课程部分PPT整理自薛宇教授和任间教授课件或其它资源，  
特致谢！