



中华人民共和国国家标准

GB/T 29859—2013

生物信息学术语

Bioinformatics terms

2013-11-12 发布

2014-04-15 实施

中华人民共和国国家质量监督检验检疫总局 发布
中国国家标准化管理委员会

目 次

前言	I
1 范围	1
2 术语和定义	1
索引	9
参考文献	15

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准由中国标准化研究院提出并归口。

本标准起草单位：中国标准化研究院、深圳华大基因研究院、天津华大基因科技有限公司。

本标准主要起草人：戚菲、孙广芝、张育润、江洲、任冠华、张蕊、王俊、李英睿、杨玲。

生物信息学术语

1 范围

本标准规定了生物信息学领域的基本术语及其定义。

本标准适用于生物信息学领域相关概念的统一、协调,以及学术交流和知识传播。

2 术语和定义

2.1 生物信息学

2.1.1

生物信息学 **bioinformatics**

应用信息科学及相关学科的方法和技术,研究和分析生物体系和生物过程中信息存储、处理和传递的一门交叉学科。

2.2 生物序列比对

2.2.1

序列比对 **sequence alignment**

比较两个或两个以上核苷酸或者氨基酸序列间的相似性的过程。

2.2.2

结构比对 **structural alignment**

比较两个或两个以上蛋白质或核酸分子空间结构的相似性的过程。

2.2.3

基本局部比对搜索工具 **basic local alignment search tool; BLAST**

基于局部比对的序列数据库搜索工具。

2.2.4

核酸序列 **nucleic acid sequence**

核酸分子中的(脱氧)核苷酸的排列顺序,常用 IUPAC 的核苷酸字符集进行描述。

2.2.5

结构域 **structural domain**

蛋白质或核酸分子中具有特定折叠结构和功能的单元。

2.2.6

非翻译区 **untranslated region; UTR**

在信使核糖核酸中位于编码序列两侧的未被翻译成蛋白质的序列。

2.2.7

保守序列 **conserved sequence**

在进化过程中基本上不变的核苷酸序列或氨基酸序列。

2.2.8

外显子 **exon**

真核生物基因的一部分,在剪接后会被保留在成熟核糖核酸分子中的序列。

2.2.9

表达序列标签 **expressed sequence tag; EST**

特定条件下基因表达序列的片段。

2.2.10

序列标记位点 **sequenced tagged site; STS**

在绘制基因组物理图谱时基因组中特定的基因座,可用作染色体定位的标记。

2.2.11

串联重复序列 **tandem repeat sequence**

染色体上首尾相连的多次重复的序列片段。

2.2.12

基因家族 **gene family**

存在于多个物种、来源于共同祖先的一组基因,或同一物种中结构与功能相似、进化起源上密切相关的一组基因。

2.2.13

基因作图 **gene mapping**

对脱氧核糖核酸染色体或染色质中基因的相对位置和距离进行确定的过程。

2.2.14

遗传连锁图谱 **genetic linkage map**

显示基因组内基因以及专一的多态性脱氧核糖核酸在基因组中的相对位置的图谱。

2.2.15

物理图谱 **physical map**

显示脱氧核糖核酸在基因组中精确位置的图谱。

2.2.16

限制酶图谱 **restriction map**

描述限制性内切酶的特异识别序列在脱氧核糖核酸链上的出现频率和它们之间的相对位置的图谱。

2.2.17

基因预测 **gene prediction**

利用算法探索脱氧核糖核酸序列中特异的区域,如基因的起始区域和终止区域,对潜在基因作出预测或根据与已知基因之间的相似性来预测新基因的过程。

2.2.18

转录 **transcription**

脱氧核糖核酸的遗传信息被拷贝成核糖核酸遗传信息的过程。

2.2.19

转录因子 **transcription factor**

参与识别脱氧核糖核酸序列中启动子、增强子或特定序列且调控基因表达的蛋白质。

2.2.20

内含子 **intron**

真核生物基因的一部分,在剪接后未被保留在成熟核糖核酸分子中的序列。

2.2.21

密码子 **codon**

由三个相邻核苷酸组成的信使核糖核酸(mRNA)基本编码单位,由它们决定在翻译中生成的氨基酸链的一级结构以及翻译的起始和终止。

2.2.22

能在基因组中移动位置的脱氧核糖核酸片段。

2.2.23

开放阅读框 open reading frame; ORF

一段从起始密码子到终止密码子的核苷酸序列,可编码一条完整的氨基酸序列链,其间不存在使翻译中断的终止密码子。

2.2.24

操纵子 operon

原核生物中由调节基因、启动子和结构基因组成的一个转录功能单位。

2.2.25

启动子 promoter

脱氧核糖核酸分子上能与核糖核酸聚合酶结合并形成转录起始复合体的序列

2.2.26

终止子 terminator

位于基因或操纵子末端,在转录过程中能够终止核糖核酸聚合酶转录,使核糖核酸合成终止的脱氧核糖核酸序列。

2.2.27

非编码 RNA noncoding RNA

能被转录但不能编码蛋白质的核糖核酸序列。

2.2.28

同源 homology

两种或多种生物起源于共同祖先的生物特征。

2.2.29

同源基因 homologous gene

具有同源特征的基因。

2.2.30

同源序列 homologs

具有同源特征的序列。

旁系同源基因 paralogous gene

在同一物种内由基因复制产生的具有共同起源的基因。

2.2.32

直系同源基因 orthologous gene

在不同物种中具有共同起源的基因。

2.2.33

单核苷酸多态性 single nucleotide polymorphism; SNP

在基因组水平上,由单个核苷酸位点的变异(替代、插入或缺失)所引起的脱氧核糖核酸序列多态性。

2.2.34

保守结构域 conserved domain

在进化过程中基本保持不变的结构域。

2.2.35

同线性 **synteny**

两个或多个基因组(片段)中基因座的排列相同的特征。

2.2.36

基因座 **gene locus**

基因在染色体上所处的特定位置。

2.2.37

Needleman-Wunsch 算法 **Needleman-Wunsch algorithm**

使用动态规划对序列进行全局比对的算法。

2.2.38

Smith-Waterman 算法 **Smith-Waterman algorithm**

使用动态规划对序列进行局部比对的算法。

2.2.39

序列相似性 **sequence similarity**

序列比对过程中,用来描述检测序列和目标序列之间相同脱氧核糖核酸碱基或氨基酸残基序列所占比例高低的性质。

2.2.40

全局比对 **global alignment**

比对结果中包含所比较序列全长范围内所有位点的比对。

2.2.41

局部比对 **local alignment**

对相似性水平较高的局部片段的比对。

2.2.42

沉默子 **silencer**

调控真核生物基因转录的负调控元件。

2.2.43

增强子 **enhancer**

调控真核生物基因转录的正调控元件。

2.2.44

供体位点 **donor site**

核糖核酸剪接加工过程内含子 5'-端接头序列。

2.2.45

受体位点 **acceptor site**

核糖核酸剪接加工过程内含子 3'-端接头序列。

2.3 分子系统发生

2.3.1

邻接法 **neighbour-joining method; NJ**

一种利用距离作分子系统分析的方法。

2.3.2

系统发育分类 **phylogenetic taxonomy**

按不同物种进化过程中的亲缘关系进行分类的方法。

2.3.3

系统发育图 phylogram

用枝长表示进化时间的系统树的示意图。

2.3.4

系统发生树 phylogenetic tree

用以表示系统发生过程的示意图。

2.3.5

进化支 clade

生物进化过程中进化树上的分支。

2.3.6

遗传漂移 genetic drift

与选择压力无关的基因的随机漂移。

2.3.7

进化分支图 cladogram

进化树中代表每个节点分支的最近共同祖先的树枝状图。

2.3.8

自举检验 bootstrap test

对置信程度进行的量化检验。

2.3.9

同义替代 synonymous substitution

在脱氧核糖核酸编码序列水平上不使所表达的氨基酸序列发生改变的核苷酸替换。

2.3.10

非同义替代 non synonymous substitution

在脱氧核糖核酸编码序列水平上使所表达的氨基酸序列发生改变的核苷酸替换。

2.3.11

正向选择 positive selection

物种或生物分子在环境特定的选择压力下,使得某种突变被固定下来的概率明显升高的现象。

2.3.12

负向选择 negative selection**纯化选择 purifying selection**

物种或生物分子在环境特定的选择压力下,使得突变被固定下来的概率明显降低的现象。

2.3.13

替代比率 K_a/K_s ; dN/dS

非同义替代率(K_a 或 dN)和同义替代率(K_s 或 dS)之间的比例,用于判断是否有选择压力作用于基因。

2.4 组学类相关

2.4.1

比较基因组学 comparative genomics

用计算机和实验室的高通量筛选方法,对各种各样的生物体的基因组信息进行比较研究,以达到对生物过程和现象在基因组水平上的理解的学科。

2.4.2

转录组学 transcriptomics

在整体水平上研究细胞中基因转录的情况及转录调控规律的学科。

2.4.3

转录组测序技术 RNA-Seq

用高通量测序技术对细胞内转录组进行测序的方法。

2.4.4

生物芯片 biochips

采用生物技术制备或应用于生物技术的微处理器。

2.4.5

微阵列 microarray

在一个固体基片上的已知位置固定了 DNA 探针或蛋白质探针的有序阵列。

2.4.6

功能基因组学 functional genomics

研究基因组中各基因的功能,包括基因的表达及其调控模式的学科。

2.4.7

蛋白质组学 proteomics

从整体的角度分析细胞内动态变化的蛋白质组成成分、表达水平与修饰状态,了解蛋白质之间的相互作用与联系,揭示蛋白质功能与细胞生命活动规律的学科。

2.4.8

基因表达系列分析 serial analysis of gene expression;SAGE

通过构建较短的表达序列标签规模化地检测基因表达种类及其丰度的实验技术。

2.4.9

基因组学 genomics

用全基因组序列信息和高通量基因技术,在基因组水平上研究生物系统的结构、功能和进化的分子机制的学科。

2.4.10

结构基因组学 structural genomics

以基因组图谱、测序、组成以及基因组水平的蛋白结构鉴定为方法,对基因和蛋白的基因组水平的结构进行研究的学科。

2.4.11

基因组注释 genome annotation

鉴别和标记基因组序列的功能单位和基因组上特殊信号的过程。

2.4.12

基因本体论 gene ontology;GO

对基因和蛋白质功能进行限定和描述的语义词汇集合。

2.4.13

测序 sequencing

测定氨基酸或者核苷酸序列的过程。

2.4.14

叠连群 contig

测序得到的 DNA 片段,根据相互间的重叠性,构成的一个长的无缺失的 DNA 片段。

2.4.15

鸟枪法测序 shotgun sequencing

将基因组打断为 DNA 片段并进行测序的方法。

2.4.16

表观遗传组学 epigenomics

研究表观遗传变异的基因组学分支学科。

2.4.17

组蛋白修饰 histone modification

发生在染色体组成成分组蛋白上的化学修饰。

2.4.18

染色质免疫共沉淀-芯片技术 Chromatin Immunoprecipitation-chip; ChIP-chip

结合染色质免疫共沉淀技术(ChIP)与芯片技术,能够快速在目标基因组的染色体中确定特异DNA结合蛋白的准确结合区段的技术方法。

2.4.19

染色质免疫共沉淀技术 Chromatin Immunoprecipitation; ChIP

研究体内蛋白质与DNA相互作用的技术方法。

2.4.20

染色质免疫沉淀-测序 Chromatin Immunoprecipitation-Sequencing; ChIP-Seq

结合染色质免疫共沉淀技术(ChIP)与高通量测序技术,能够快速在目标基因组的染色体中确定特异DNA结合蛋白的准确结合区段的技术方法。

2.4.21

系统生物学 systems biology

整合不同层次组学数据以理解生物系统如何行使功能的学科。

2.4.22

药物基因组学 pharmacogenomics

在基因组水平上研究不同个体及人群对药物反应的差异,并探讨用药个性化和以特殊人群为对象的新药开发的学科。

2.4.23

基因调控网络 gene regulation network

由细胞中参与基因表达调控的DNA、RNA、蛋白质以及代谢中间体所形成的相互作用的网络。

2.5 生物大分子的结构及其预测

2.5.1

结构生物学 structural biology

以生物物理学和生物化学手段研究生物大分子的三维结构,以及结构与对应功能的关系的学科。

2.5.2

模体 motif

在蛋白质或核酸序列中涉及相互作用的较小且高度保守的区域。

2.5.3

蛋白质二级结构预测 protein secondary structure prediction

预测氨基酸序列形成可能的二级结构的方法。

2.5.4

蛋白质三级结构预测 protein tertiary structure prediction

预测氨基酸序列形成可能的三级结构的方法。

2.5.5

同源模建法 homology modeling method

通过同源分析找到一个已知结构的同源蛋白质,然后以该蛋白质的结构为模板为未知结构的蛋白质建立结构模型的方法。

2.5.6

分子对接 molecular docking

模拟两个或多个分子之间的识别过程的方法。

2.6 生物数据库挖掘

2.6.1

生物数据挖掘 biological data mining

从生物数据库的大量数据中揭示出隐含的、先前未知的并有潜在价值的信息的过程。

2.6.2

隐马尔可夫模型 Hidden Markov Model; HMM

一种用参数表示的用于描述随机过程统计特性的概率模型,由马尔可夫链和一般随机过程组成。

注:在生物信息学中常用于模式搜索和识别。

支持向量机 Support Vector Machines; SVM

一种在数据点支撑起的高维空间中用超平面把点分隔开的办法。

注:在生物信息学中常用于模式识别和分类。

2.6.4

期望最大化算法 expectation maximization algorithm; EM

在不完全数据情况下,计算极大似然估计或者后验分布的迭代算法。

注:在生物信息学中常用于模型参数的估计。

2.6.5

主成分分析 principle component analysis; PCA

将分散在一组变量上的信息,集中到某几个综合指标(主成分)上的一种统计分析方法。

注:在生物信息学中常用于多维度聚类。

索引

中文索引

- B**
- 保守结构域 2.2.34
 保守序列 2.2.7
 比较基因组学 2.4.1
 表达序列标签 2.2.9
 表观遗传组学 2.4.16
- C**
- 操纵子 2.2.24
 测序 2.4.13
 沉默子 2.2.42
 串联重复序列 2.2.11
 纯化选择 2.3.12
- D**
- 单核苷酸多态性 2.2.33
 蛋白质二级结构预测 2.5.3
 蛋白质三级结构预测 2.5.4
 蛋白质组学 2.4.7
 叠连群 2.4.14
- F**
- 非编码 RNA 2.2.27
 非翻译区 2.2.6
 非同义替代 2.3.10
 分子对接 2.5.6
 负向选择 2.3.12
- G**
- 功能基因组学 2.4.6
 供体位点 2.2.44
- H**
- 核酸序列 2.2.4
- J**
- 基本局部比对搜索工具 2.2.3
- 基因本体论 2.4.12
 基因表达系列分析 2.4.8
 基因家族 2.2.12
 基因调控网络 2.4.23
 基因预测 2.2.17
 基因组学 2.4.9
 基因组注释 2.4.11
 基因作图 2.2.13
 基因座 2.2.36
 结构比对 2.2.2
 结构基因组学 2.4.10
 结构生物学 2.5.1
 结构域 2.2.5
 进化分支图 2.3.7
 进化支 2.3.5
 局部比对 2.2.41
- K**
- 开放阅读框 2.2.23
- L**
- 邻接法 2.3.1
- M**
- 密码子 2.2.21
 模体 2.5.2
- N**
- Needleman-Wunsch 算法 2.2.37
 内含子 2.2.20
 鸟枪法测序 2.4.15
- P**
- 旁系同源基因 2.2.31
- Q**
- 期望最大化算法 2.6.4
 启动子 2.2.25

全局比对 2.2.40

R

染色质免疫沉淀-测序 2.4.20

染色质免疫共沉淀技术 2.4.19

染色质免疫共沉淀-芯片技术 2.4.18

S

Smith-Waterman 算法 2.2.38

生物数据挖掘 2.6.1

生物芯片 2.4.4

生物信息学 2.1.1

受体位点 2.2.45

T

替代比率 2.3.13

同线性 2.2.35

同义替代 2.3.9

同源 2.2.28

同源基因 2.2.29

同源模建法 2.5.5

同源序列 2.2.30

W

外显子 2.2.8

微阵列 2.4.5

物理图谱 2.2.15

X

系统发生树 2.3.4

英文索引

系统发育分类 2.3.2

系统发育图 2.3.3

系统生物学 2.4.21

限制酶图谱 2.2.16

序列比对 2.2.1

序列标记位点 2.2.10

序列相似性 2.2.39

Y

药物基因组学 2.4.22

遗传连锁图谱 2.2.14

遗传漂移 2.3.6

隐马尔可夫模型 2.6.2

Z

增强子 2.2.43

正向选择 2.3.11

支持向量机 2.6.3

直系同源基因 2.2.32

终止子 2.2.26

主成分分析 2.6.5

转录 2.2.18

转录因子 2.2.19

转录组测序技术 2.4.3

转录组学 2.4.2

转座子 2.2.22

自举检验 2.3.8

组蛋白修饰 2.4.17

A

acceptor site 2.2.45

B

basic local alignment search tool 2.2.3

biochips 2.4.4

bioinformatics 2.1.1

biological data mining 2.6.1

BLAST 2.2.3

bootstrap test 2.3.8

C

ChIP 2.4.19

ChIP-chip 2.4.18

ChIP-Seq 2.4.20

Chromatin Immunoprecipitation 2.4.19

Chromatin Immunoprecipitation-chip 2.4.18

Chromatin Immunoprecipitation-Sequencing 2.4.20

clade 2.3.5

cladogram 2.3.7

codon 2.2.21

comparative genomics 2.4.1

conserved domain 2.2.34

conserved sequence 2.2.7

contig 2.4.14

D

dN/Ds 2.3.13

donor site 2.2.44

E

EM 2.6.4

enhancer 2.2.43

epigenomics 2.4.16

EST 2.2.9

exon 2.2.8

expectation maximization algorithm 2.6.4

expressed sequence tag 2.2.9

F

functional genomics 2.4.6

G

gene family 2.2.12

gene locus 2.2.36

gene mapping 2.2.13

gene ontology 2.4.12

gene prediction 2.2.17

gene regulation network 2.4.23

genetic drift 2.3.6

genetic linkage map 2.2.14

genome annotation 2.4.11

genomics	2. 4. 9
global alignment	2. 2. 40
GO	2. 4. 12

H

Hidden Markov Model	2. 6. 2
histone modification	2. 4. 17
HMM	2. 6. 2
homologous gene	2. 2. 29
homologs	2. 2. 30
homology	2. 2. 28
homology modeling method	2. 5. 5

I

intron	2. 2. 20
---------------------	----------

K

Ka/Ks	2. 3. 13
--------------------	----------

L

local alignment	2. 2. 41
------------------------------	----------

M

microarray	2. 4. 5
molecular docking	2. 5. 6
motif	2. 5. 2

N

Needleman-Wunsch algorithm	2. 2. 37
negative selection	2. 3. 12
neighbour-joining method	2. 3. 1
NJ	2. 3. 1
non synonymous substitution	2. 3. 10
noncoding RNA	2. 2. 27
nucleic acid sequence	2. 2. 4

O

open reading frame	2. 2. 23
operon	2. 2. 24
ORF	2. 2. 23
orthologous gene	2. 2. 32

P

paralogous gene	2. 2. 31
------------------------------	----------

PCA	2.6.5
pharmacogenomics	2.1.22
phylogenetic taxonomy	2.3.2
phylogenetic tree	2.3.4
phylogram	2.3.3
physical map	2.2.15
positive selection	2.3.11
principle component analysis	2.6.5
promoter	2.2.25
protein secondary structure prediction	2.5.3
protein tertiary structure prediction	2.5.4
proteomics	2.4.7
purifying selection	2.3.12

R

restriction map	2.2.16
RNA-Seq	2.4.3

S

SAGE	2.4.8
sequence alignment	2.2.1
sequence similarity	2.2.39
sequenced tagged site	2.2.10
sequencing	2.4.13
serial analysis of gene expression	2.4.8
shotgun sequencing	2.4.15
silencer	2.2.42
single nucleotide polymorphism	2.2.33
Smith-Waterman algorithm	2.2.38
SNP	2.2.33
structural alignment	2.2.2
structural biology	2.5.1
structural domain	2.2.5
structural genomics	2.4.10
STS	2.2.10
Support Vector Machines	2.6.3
SVM	2.6.3
synonymous substitution	2.3.9
synteny	2.2.35
systems biology	2.4.21

T

tandem repeat sequence	2.2.11
------------------------------	--------

terminator	2.2.26
transcription	2.2.18
transcription factor	2.2.19
transcriptomics	2.4.2
transposon	2.2.22

U

untranslated region	2.2.6
UTR	2.2.6

参 考 文 献

- [1] 郝柏林. 生物信息学手册(第二版). 上海:上海科学技术出版社,2002.
- [2] (美)坎贝尔,(美)海尔. 探索(基因组学、蛋白质组学和生物信息学). 孙之荣,译. 北京:科学出版社,2007.
- [3] 王禄山,高培基. 生物信息学应用技术. 北京:化学工业出版社,2008.
- [4] 许忠能. 生物信息学. 北京:清华大学出版社,2008.
- [5] 陶士珩. 生物信息学. 北京:科学出版社,2007.
-
- [7] (美)恩格中. 蛋白质功能基因组学. 北京:科学出版社,2007.
-
- [9] (美)P. E. 波恩,(美)H. 魏西希. 结构生物信息学. 刘振明,刘海燕,译. 北京:化学工业出版社,2009.
- [10] 薛庆中. DNA 和蛋白质序列数据分析工具. 北京:科学出版社,2009.
- [11] 胡松年. 基因表达序列标签(EST)数据分析手册. 杭州:浙江大学出版社,2005.

中 华 人 民 共 和 国
国 家 标 准
生 物 信 息 学 术 语
GB/T 29859—2013

*

中国标准出版社出版发行
北京市朝阳区和平里西街甲2号(100013)
北京市西城区三里河北街16号(100045)

网址 www.spc.net.cn

总编室:(010)64275323 发行中心:(010)51780235

读者服务部:(010)68523946

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

*

开本 880×1230 1/16 印张 1.25 字数 32 千字
2013年12月第一版 2013年12月第一次印刷

*

书号: 155066·1-47886

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68510107



GB/T 29859-2013