



中华人民共和国国家标准

GB/T 34798—2017

核酸数据库序列格式规范

Formats specifications of nucleotide sequence database

2017-11-01 发布

2018-05-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 核酸序列格式规范制定的总则	2
6 核酸序列描述规范	2
7 核酸序列特征描述规范	3
8 核酸序列格式规范	5
9 核酸序列文件整体格式规范	5
附录 A (资料性附录) 核苷酸含义表	6
附录 B (资料性附录) 与核酸相关的特征关键词表	7
附录 C (资料性附录) 密码子表	10
附录 D (资料性附录) 修饰碱基表	11
附录 E (资料性附录) 限定词中英文对照表	13
附录 F (资料性附录) 核酸序列文件样例	14
附录 G (资料性附录) 行首大写字母含义表	15
参考文献	16

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准由全国生化检测标准化技术委员会(SAC/TC 387)提出并归口。

本标准起草单位:深圳华大基因研究院、深圳华大基因科技有限公司、广东省标准化研究院、广东产品质量监督检验研究院。

本标准主要起草人:魏晓锋、陈凤珍、刘克、杜佳婷、李倩一、沈维燕、李启沅、谢强、王娟、谭嘉力、宋祚锜、黄江勇。

核酸数据库序列格式规范

1 范围

本标准规定了核酸数据库的序列格式,包括生物体基因组核酸序列特征规范制定的总则、序列描述格式规范、序列特征描述规范和序列格式规范等。

本标准适用于生物体基因组核酸数据库序列文件的编写。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 29859 生物信息学术语

ZC 0003 核苷酸和/或氨基酸序列列表和序列列表电子文件标准

3 术语和定义

GB/T 29859 界定的以及下列术语和定义适用于本文件。

3.1

核酸数据库 nucleic acid database

以核酸序列为基本内容,并附有核酸序列注释信息的数据库。

3.2

编码序列 coding sequence

编码一段蛋白产物的序列,始于起始密码子,终于终止密码子。

3.3

序列组装 sequence assembly

基因组长序列打断之后形成较短的序列,通过算法和计算机的帮助,把这些短的序列组装起来成为一条完整有序的序列的过程。

3.4

甲基化 methylation

蛋白质和核酸的一种重要的修饰,调节基因的表达和关闭。

3.5

识别码 identifier

某个体系中相对唯一的编码。

3.6

位置 location

一个或一段碱基在另一段较长碱基上的相对坐标位置。

3.7

特征限定词 feature qualifier

用来进一步描述序列的某一类特征的词。