

国家重大出版工程项目

Principes des techniques de biologie moléculaire

图解常用分子生物学原理

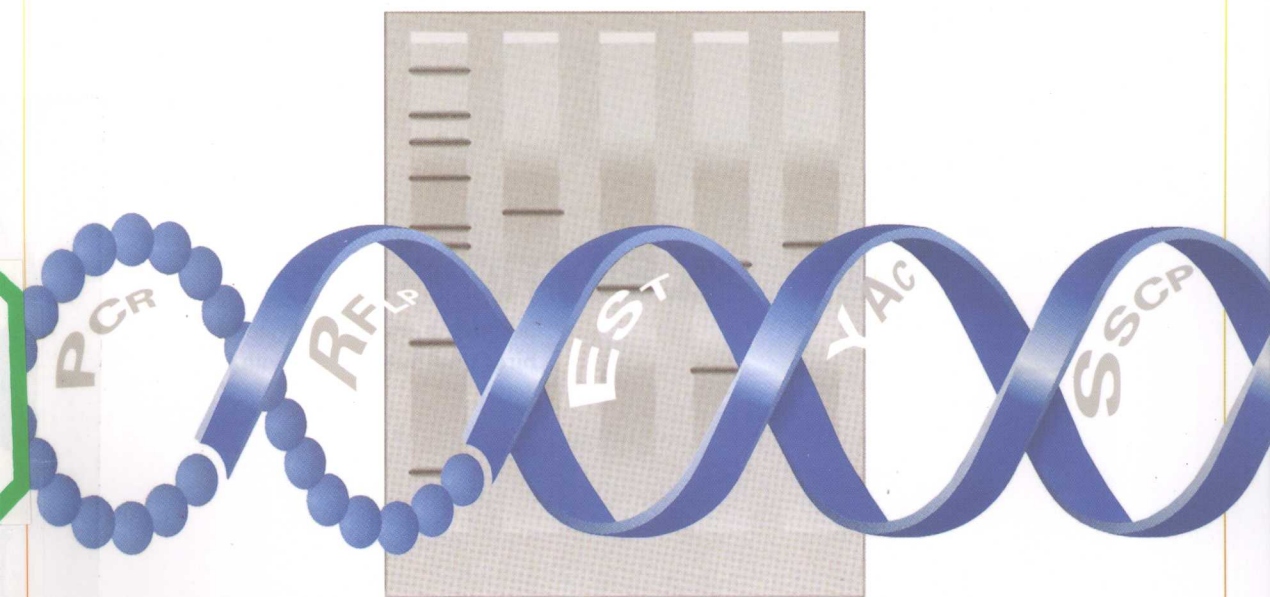
第2次修订补充版

D. Tagu, C. Moussard 编著

康定明 主译

陈章良 审校

(根据英文版翻译)



中国农业大学出版社
China Agricultural University Press

éditions
Quæ

国家重大出版工程项目

Principes des techniques de biologie moléculaire

图解常用分子生物学原理

第2次修订补充版

D. Tagu, C. Moussard 编著

中国农业大学出版社

图书在版编目(CIP)数据

图解常用分子生物学原理/(法)塔格(Tagu, D.), (法)穆萨德(Moussard, C.)编著;康定明主译,陈章良审校. —北京:中国农业大学出版社, 2008. 11

书名原文:Principes des techniques de biologie moléculaire

ISBN 978-7-81117-272-0

I. 分… II. ①塔…②穆…③康…④陈… III. 分子生物学 IV. Q7

中国版本图书馆 CIP 数据核字(2008)第 010008 号

Principes des techniques de biologie moléculaire

by Tagu D., Moussard C.

2ème édition revue et augmentée

Published by arrangement with INRA Paris and translated from the English version published in 2006 by Science Publishers

INRA Paris, 2003

ISBN 2-7380-1067-9

著作权合同登记图字:01-2007-0495 号

书 名 Principes des techniques de biologie moléculaire

图解常用分子生物学原理

作 者 D. Tagu, C. Moussard 编著

康定明 主译

陈章良 审校

策划编辑 宋俊果

封面设计 郑 川

出版发行 中国农业大学出版社

社 址 北京市海淀区圆明园西路 2 号

电 话 发行部 010-62731190, 2620

编辑部 010-62732617, 2618

网 址 <http://www.cau.edu.cn/caup>

经 销 新华书店

印 刷 涿州市星河印刷有限公司

版 次 2008 年 11 月第 1 版 2008 年 11 月第 1 次印刷

规 格 787×1 092 16 开本 11.5 印张 204 千字

印 数 1~2 500

定 价 50.00 元

责任编辑 冯雪梅

责任校对 王晓凤 陈 莹

邮政编码 100193

读者服务部 010-62732336

出 版 部 010-62733440

e-mail cbsszs@cau.edu.cn

图书如有质量问题本社发行部负责调换

作者信息

D. Tagu, INRA Rennes, UMR INRA / AGROCAMPUS BiO3P BP 35327,
35653 Le Rheu cedex, 法国

C. Moussard Service de Biochimie Médicale, Faculté de Médecine Place
Saint-Jacques, 25000 Besançon, 法国

中文版前言

在高校多年的《分子生物学》教学实践中,我一直习惯用不同类型的图例来讲解其中的概念和原理,也经常指导同学用直观的图例增强记忆和理解,帮助同学们把看似死板枯燥的分子生物学书本知识与形象有寓意的图例联系起来。这样,在锻炼同学空间想像思维能力的同时,也使得非常抽象、生活中难以触及的分子生物学知识在同学们的大脑中较容易地以图例的形式记忆下来,得到了多届同学的好评,帮助他们分子生物学的学习。因此,我早想能有一本以图例的形式把分子生物学知识更加形象地描述讲解出来的读本,以帮助对分子生物学有兴趣的同志在学习分子生物学及其技术时参考。在国外参加会议期间,我发现这本书正好符合了这方面的要求。因此,决定组织力量将其翻译出版,贡献给国内的读者。

本书是 2006 年英文版《Techniques for Molecular Biology》的翻译,英文版是 2003 年法文原版《Principes des techniques de biologie moléculaire, 2ème édition revue et augmentée》的译本。全书分为 8 个部分有 57 项。书中内容涵盖了分子生物学技术最基本的概念知识,以及常用技术和近些年热点研究领域如基因功能分析技术中的基因标签技术、酵母双杂交系统、基因敲出技术和 RNAi 技术,以及基因组多态性研究技术中的 SNP 标记、SSR 标记等常用技术。本书的特点是几乎对每项技术的原理讲解都配了简洁的图例,使得读者对常常感到看不见摸不着的深奥的分子生物学理论和原理有了比较直观的认识,并且更加容易理解。掌握了这些分子生物学技术的概念和原理后,在自己的学习和研究中,就能够方便地找到解决问题的方向和准备具体方案。

本书适合于对分子生物学知识和技术有兴趣的学生和相关研究人员,以及从事有关分子生物学技术学习和研究的大中专学生、研究生和专业技术人员。

参加本书翻译的研究生有王燕冰、张海燕和兰阿峰等同学。全书由康定明教授统稿整理。

陈章良

2008-7-8

目 录

I	定义	1
	图解 1 真核生物基因的结构与表达	3
	图解 2 基因组相关参数	5
	图解 3 基因组测序	8
II	载体及克隆	11
	图解 4 限制性内切酶	13
	图解 5 核酸电泳技术	15
	图解 6 质粒和噬菌体质粒	17
	图解 7 噬菌体和考斯粒	19
	图解 8 YAC(酵母人工染色体)和其他大载体	20
	图解 9 分子克隆	22
	图解 10 细菌和酵母的遗传转化	26
III	核的标记和杂交	27
	图解 11 DNA 标记	29
	图解 12 分子杂交	32
	图解 13 mRNA 的原位杂交	35
IV	DNA 文库和筛选	37
	图解 14 基因文库的构建与筛选	39
	图解 15 cDNA 文库的构建	41
	图解 16 文库的筛选	43
	图解 17 差异性筛选:差减文库,cDNA-AFLP	46
	图解 18 差异显示 RT-PCR	49
	图解 19 运用 SSH(抑制差减杂交)的差异性筛选	52
	图解 20 运用 RDA(代表性差异分析)的差异性筛选	55
	图解 21 EST:表达序列标签	58
	图解 22 DNA 微阵列:DNA 芯片,cDNA 吸附膜	61
V	基因鉴定	71
	图解 23 DNA 测序	73

图解 24	PCR:多聚酶链式反应	77
图解 25	RACE:cDNA 末端的快速扩增	79
图解 26	PCR 方法的基因组步查	82
图解 27	RT-PCR:逆转录 PCR	84
图解 28	体外转录	86
图解 29	转录起始位点的确定	87
图解 30	启动子功能分析	89
图解 31	凝胶阻滞	91
图解 32	DNA 酶 I 酶切足迹	93
VI	真核生物的遗传转化	95
图解 33	用土壤农杆菌进行植物的遗传转化	97
图解 34	将外源基因转入植物原生质	102
图解 35	运用基因枪法导入外源基因	103
图解 36	动物细胞的遗传转化	104
图解 37	动物克隆	107
图解 38	瞬时表达	111
VII	基因功能分析	113
图解 39	重组蛋白质	115
图解 40	昆虫的杆状病毒,外源基因表达载体	117
图解 41	酵母双杂交系统	121
图解 42	定点诱变	125
图解 43	酵母的突变互补	129
图解 44	酵母基因的敲除	131
图解 45	基因标签	133
图解 46	RNA 干扰	137
VIII	基因组多态性	141
图解 47	分子标记	143
图解 48	遗传和物理图谱	149
图解 49	PFGE:脉冲电泳	152
图解 50	RLFP:限制性片段长度多态性	154
图解 51	RAPD:随机扩增多态性 DNA	156
图解 52	AFLP:扩增片段长度多态性	158
图解 53	逆转录标记	160

图解 54	SSCP:单链构象多态性	164
图解 55	DGGE:变性凝胶梯度电泳	166
图解 56	SNP:单核苷酸多态性	168
图解 57	单序列重复(SSR):微卫星	170
作者列表		174

I

定义

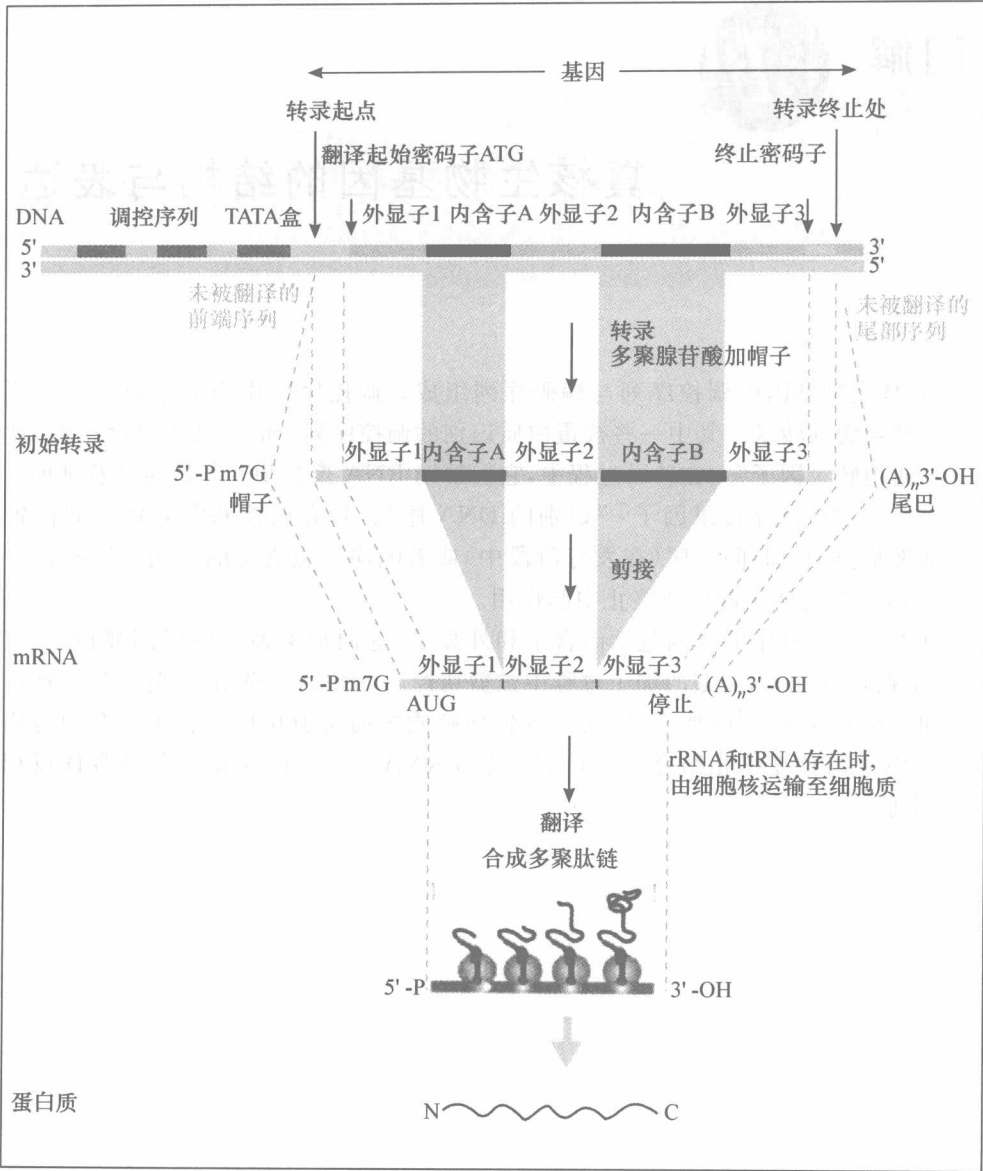
真核生物基因的结构与表达

真核生物基因由调控序列与编码序列组成。调控序列作为信号起始或终止 RNA 聚合酶的转录。其中一些被蛋白质识别的调控序列(如 TATA 启动子盒)被称为“通用转录因子”。在转录过程中,它们帮助 RNA 聚合酶 II 起始转录及延伸。

另一些被“特异转录因子^①”识别的 DNA 序列,具有调控基因表达空间位置(依据细胞类型)、时间顺序(在发育阶段中)或者响应逆境表达的作用。这些蛋白质在 DNA 链的转录起始或终止中起作用。

DNA 编码链中的基因包括内含子和外显子,这两种类型的序列都被转录,但内含子在随后 mRNA 链的加工过程中被剪切掉。mRNA 首先在 5' 端和 3' 端被修饰,即 5' 端加帽子,3' 端加尾巴,在转运到细胞质之前又剪切掉内含子。在细胞质中,mRNA 与核糖体相结合,在 tRNA(转移 RNA)的参与下,mRNA 被翻译成相应的多肽。

^①在编码序列的上下游发现它们。



DENIS TAGU

基因组相关参数

参数

- 每 2C 核中 DNA 的量: 大约 1 pg ($1 \text{ pg} = 10^{-12} \text{ g}$)
- DNA 总长度: 表达的单位为: 碱基对(bp), 千碱基对(kbp), 兆碱基对(1 Mb = 10^6 bp)。
- G+C 的百分含量: 胞嘧啶和鸟嘌呤在 DNA 序列中的含量。
- DNA 的直径: $20 \text{ \AA}$ 。一条 DNA 链的长度含有的物质为 $193/6 \times 10^{23} \text{ g}$ 。
- 1 个碱基对含有的物质为 $660/N$ 或 $660/6 \times 10^{23} \text{ g}$ 。

不同物种的基因组大小

- 病毒(DNA 或 RNA 病毒): 含 $10^3 \sim 10^5 \text{ bp}$, 或长 $100 \mu\text{m}$ 的展开链。
- 细菌: 含 $10^6 \sim 10^7 \text{ bp}$; 长大约 1 mm。
- 酵母: $1.2 \times 10^7 \text{ bp}$ 。
- 高等植物:
 - 细胞核: $10^8 \sim 10^{10} \text{ bp}$, 约 1 m 长, 倍性一般 $2n$, 遵循孟德尔遗传规律。
 - 线粒体: $1.8 \sim 6 \times 10^5 \text{ bp}$, 长度变化很大, 倍性不定, 属细胞质母性遗传。
 - 叶绿体: $1.5 \times 10^5 \text{ bp}$, 长 $50 \mu\text{m}$ 。细胞中叶绿体的倍性变化是 10 的倍数, 每个细胞大约 10~100 个叶绿体, 属细胞质母性遗传。
- 人类:
 - 细胞核: $3 \times 10^9 \text{ bp}$, 长约 1 m。
 - 线粒体: $16 \sim 20 \times 10^3 \text{ bp}$ 。

核基因组序列^①的三种类型

- 简单序列: 指单倍体的基因组中一些单一的或拷贝很少的基因序列(指作为一个小的多基因家族, 如: 编码禾本科作物的保守蛋白基因)这些序列包括编码蛋白的基因。

^①根据 Francis Quétier。

● 中度重复序列:每个基因拷贝约 1 000~10 000 倍的重复。它们包括一些编码核糖体 rRNA,tRNA 及部分蛋白质的基因。

高度重复序列:10 万~100 万倍的重复,这些序列不是基因,可以或也可以不被转录,其功能不清楚(如卫星 DNA)。

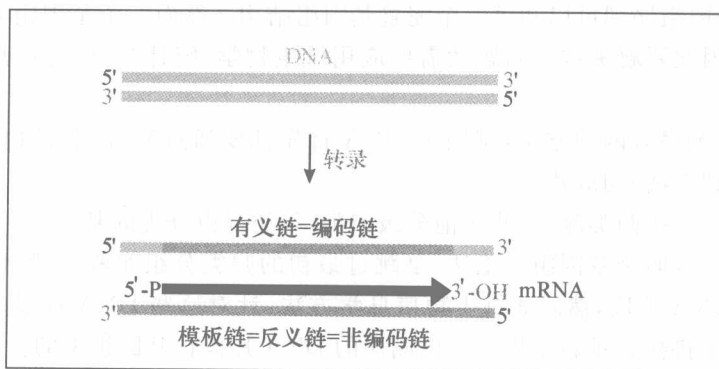
目前已公开的一些基因组数据

许多单核和真核生物的基因组已测序或正在进行测序表。下表给出一些基因的参数:

		大小(bp)	基因数	
真细菌	根癌土壤杆菌	$4.8 \cdot 10^6$ g	4 554	
	枯草芽孢杆菌	$4.2 \cdot 10^6$ g	4 100	
	布氏杆菌	$0.6 \cdot 10^6$ g	564	
	粪肠球菌	$3.2 \cdot 10^6$ g	3 334	
	大肠杆菌	$4.6 \cdot 10^6$ g	4 288	
	流感嗜血杆菌	$1.8 \cdot 10^6$ g	1 709	
	幽门螺杆菌	$1.6 \cdot 10^6$ g	1 566	
	乳酸乳球菌	$2.3 \cdot 10^6$ g	2 566	
	单胞李斯特氏单孢菌	$2.9 \cdot 10^6$ g	2 855	
	生殖道支原体	$0.5 \cdot 10^6$ g	480	
	铜绿假单孢菌	$6.2 \cdot 10^6$ g	5 565	
	伤寒沙门氏菌	$4.8 \cdot 10^6$ g	4 600	
	苜蓿中华根瘤菌	$6.7 \cdot 10^6$ g	6 204	
	天蓝色链霉菌	$8.6 \cdot 10^6$ g	7 848	
	苛养木杆菌	$2.6 \cdot 10^6$ g	2 766	
古细菌		$4.6 \cdot 10^6$ g	4 008	
	嗜热古菌	$4.6 \cdot 10^6$ g	2 694	
	甲烷杆菌属	$1.7 \cdot 10^6$ g	2 407	
	嗜盐古菌	$2.0 \cdot 10^6$ g	2 058	
	甲烷杆菌	$1.7 \cdot 10^6$ g	1 869	
	詹式甲烷球菌	$1.6 \cdot 10^6$ g	1 715	
	火棒菌属	$2.2 \cdot 10^6$ g	2 605	
	嗜热球菌	$1.7 \cdot 10^6$ g	1 765	
	嗜超高温古菌	$2.9 \cdot 10^6$ g	2 977	
	嗜酸热原体	$1.5 \cdot 10^6$ g	1 478	
		大小	染色体数	基因数
真核生物	拟南芥	$125 \cdot 10^6$ g	5	25 000
	秀丽广杆线虫	$97 \cdot 10^6$ g	6	19 000
	黑腹果蝇	$140 \cdot 10^6$ g	4	13 600
	兔脑炎原虫	$2.9 \cdot 10^6$ g	11	1 997
	人类	$3\,100 \cdot 10^6$ g	21	35 000
	水稻	$420 \cdot 10^6$ g	12	40 000
	酿酒酵母	$12 \cdot 10^6$ g	16	6 100

编码链和非编码链的定义

DNA 由两条互补且相反的平行链组成。在转录成 RNA 的过程中,只有一条链在 RNA 聚合酶的作用下发挥模板作用。RNA 链与模板 DNA 链互补,这个与 RNA 链互补的模板 DNA 链被称为反义链;另一条链与 RNA 有相同的序列,称有义链。与有义链序列相同的 mRNA 链被阅读并被解码成蛋白质,这一过程叫作翻译。相应地,与 RNA 相同的 DNA 链称为编码链;书写时,从起始的 5' 端到 3' 端书写与编码链相一致的基因序列。



DENIS TAGU

基因组测序

通过基因组序列可以知道一个完整基因组结构。然而一个基因组有数百万的碱基序列,因此要解决这一问题就需要应用微生物学,统计学,生物信息学等多种方法。

基因组测序有两种方法,基因组 DNA 首先都要通过酶(限制性内切酶)或物理方法(如超声波)切成片段。

方法一:分级归类测序,测序前分级归类,每类都由分类的基因组片段组成。

方法二:也叫全基因组鸟枪法,是跳过最初的归类分组实验阶段,而直接测序任意组的 DNA 片段,然后通过生物信息学方法,针对这些 DNA 片段之间共同的重叠序列,来拼接并重新排出各个已测序的 DNA 片段在基因组中的位序。

方法一:分级归类测序

提取 DNA 后,将其切成片段,一般通过超声波法,切成可操作的单元(如:用大的载体可以被克隆的片段)50~200 kb,在载体中将其克隆,如细菌人工染色体(BACs,见图解 8);克隆的数量一定要大,即达到总基因量的 5~10 倍。这些克隆的基因组片段然后通过特定探针(分子标记)杂交或限制性内切图谱分析或 BAC 末端(500~600 bp)杂交和测序后排序,在这些克隆片段(基因组片段)被排序后,选取单个克隆再打断成短的,可测序片段,然后由统计学方法来拼接和排列成基因组序列(比对排列)。

分级归类测序方法有两个优点:(1)这个方法很容易根据物理图谱和遗传距离图谱获取数据,或者由 BACs 的重叠来拼接基因组序列。(2)方便多个实验室合作共同测序,每个实验室可以作染色体的一部分。主要缺点是重复序列的拼接很困难。

方法二:全基因组鸟枪法

该法是将基因组 DNA 打成小片段采用“磨盘滚动式”测序,是私人公司 Celera

Genomic 在细菌基因组测序中发展起来的,随后应用在果蝇基因组测序中,最终在人和老鼠的基因组测序中得到应用。主要步骤如下:编辑 2~3 个大小不同的随机 DNA 片段文库。测出大量的克隆片段序列然后拼接。全基因组序列是通过拼接和重新组装得到的。

该方法以快速、低消耗而优于第一种方法。缺点是拼接时问题重重,尤其是遇到高度重复序列时更难办,哺乳动物基因组就是典型的例子。

测序和拼接

在两种方法之前,DNA 片段用同样的方法拼接和测序。测序原理是基于 Sanger 法(见图解 23)。全基因组测序要求特别高效的 DNA 测序工作者,而且要借助机器人助手才能自动完成成千上万的如同自动化的移液工作和酶反应工作。

在最原始的序列数据得到后,序列必须被拼接成最初序列。因此得到的所有序列必须是有秩序的,按照重叠端所指示的位序一个一个地排列好,即片段端部含有一部分一样的核苷酸链(除了序列错误外),两个序列的重叠是毗连的,并形成一个“毗连群”(contig),由此确认两个或多个重叠片段,这个工作的目标就是把要分析的一个染色体的每个毗连群按 DNA 分子的真实顺序拼接起来。错误率可达 1%,应该说这是一定的(比如在重复区段),这部分是很难测序和拼接的。100%地重新组装和拼接一个完整基因组序列是非常耗时耗力,而且花费也很大。为此,一些基因组的测序过程所得的结果,只能称之为“工作草图”。

注释

基因注释是最后一步,基于序列比对和统计学校正研究,这个工作包括(1)定位编码区域(基因);(2)确定转录方向;(3)利用基因结构和功能决定编码蛋白质;(4)鉴定基因家族(相似型^①)。

FRÉDÉRIQUE BARLOY-HUBLER

^① 对于一个确定的基因组中的基因,表现出一个很高的相似性,而且很可能来自一个单个先祖的基因,被称为相似体(paralogue)。