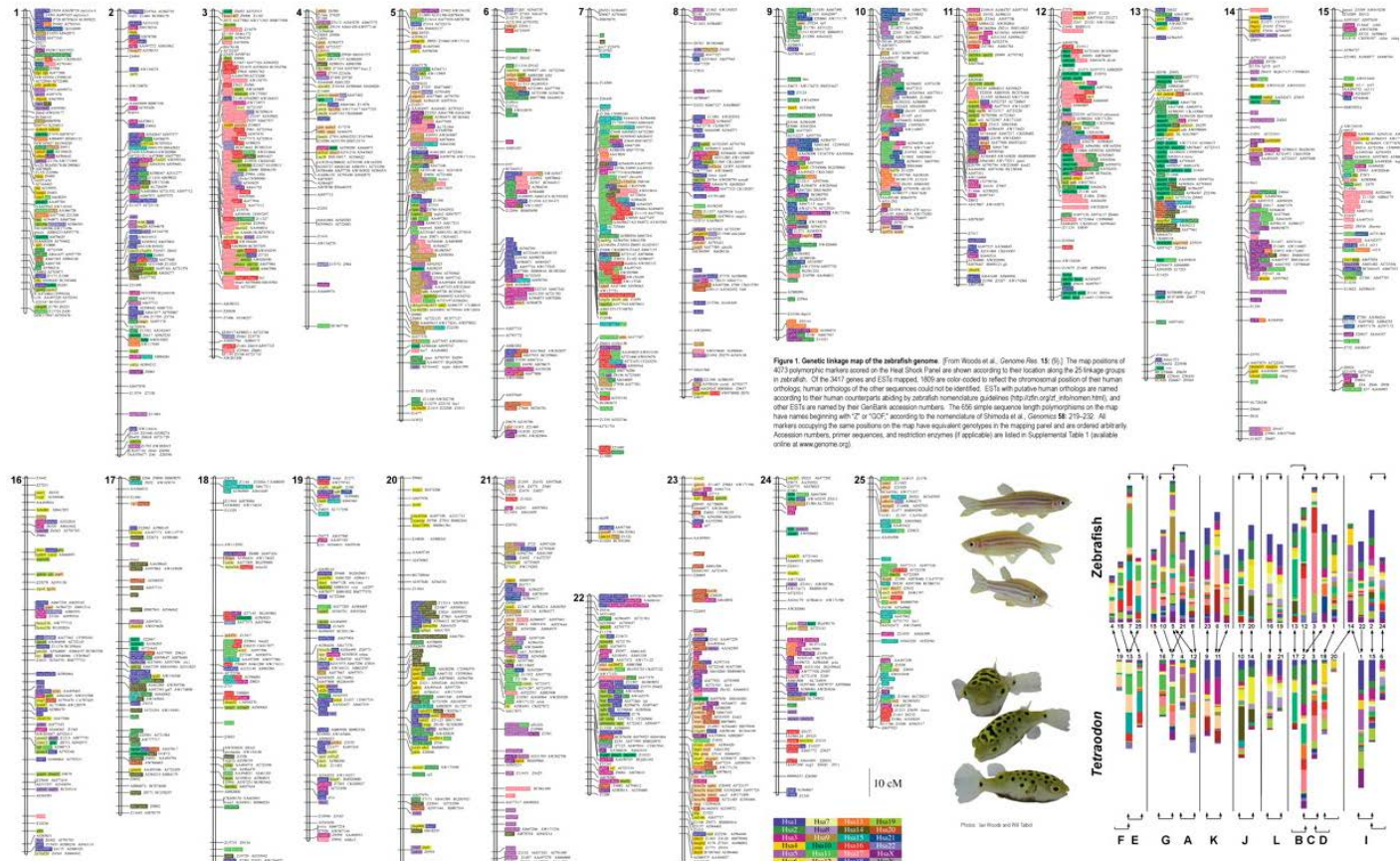


第八章 遗传图的制作和基因定位

GENE MAPPING

GENOME RESEARCH

The Zebrafish Gene Map Defines Ancestral Vertebrate Chromosomes



Acknowledgments: Based on work supported by grants from the NH&HR&NIH under the Centers of Excellence

主要内容

(1) 基本概念、原理与方法

(2) 遗传标记、物理图谱与遗传图谱

(3) 人类基因定位的基本方法

(4) 真菌类生物的遗传分析

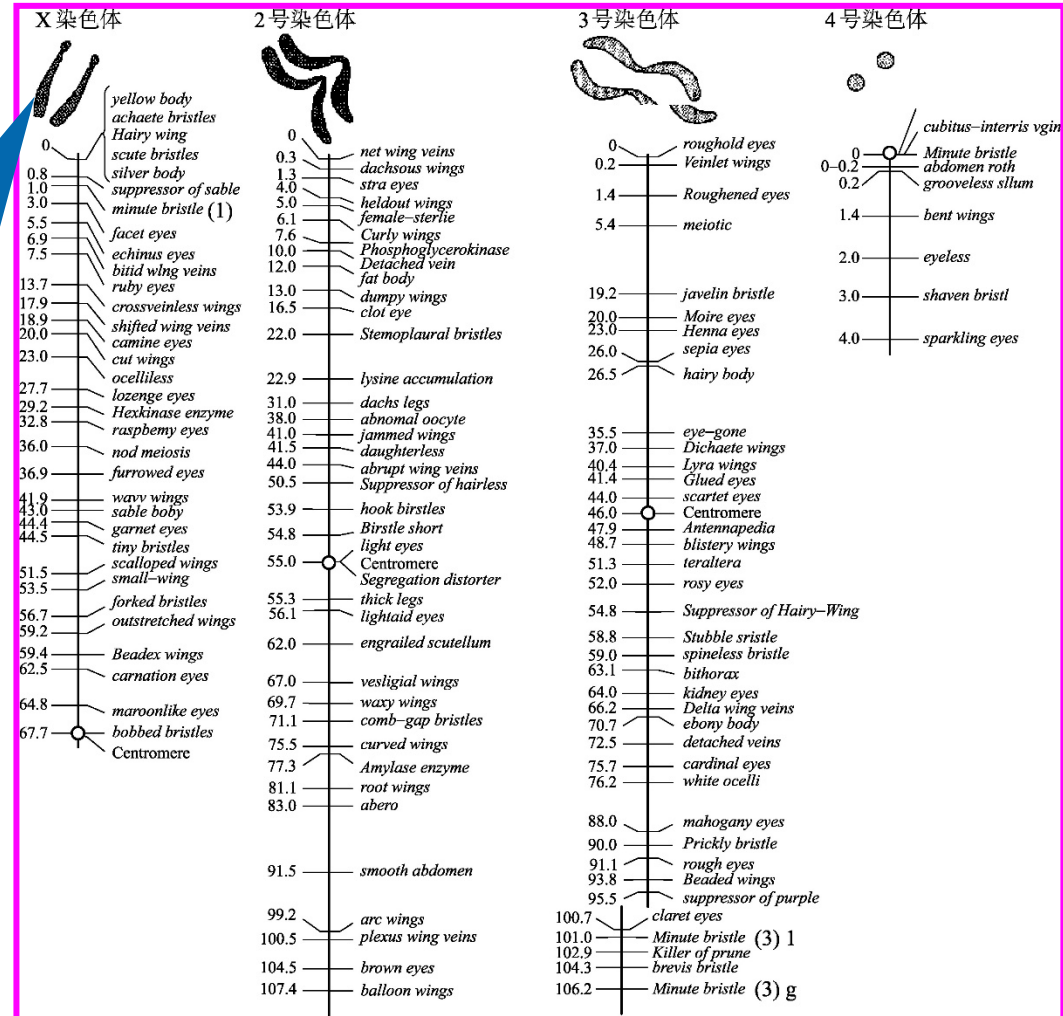
(5) 有丝分裂交换与基因定位

(6) 细菌的基因定位

(7) 噬菌体的遗传分析与作图

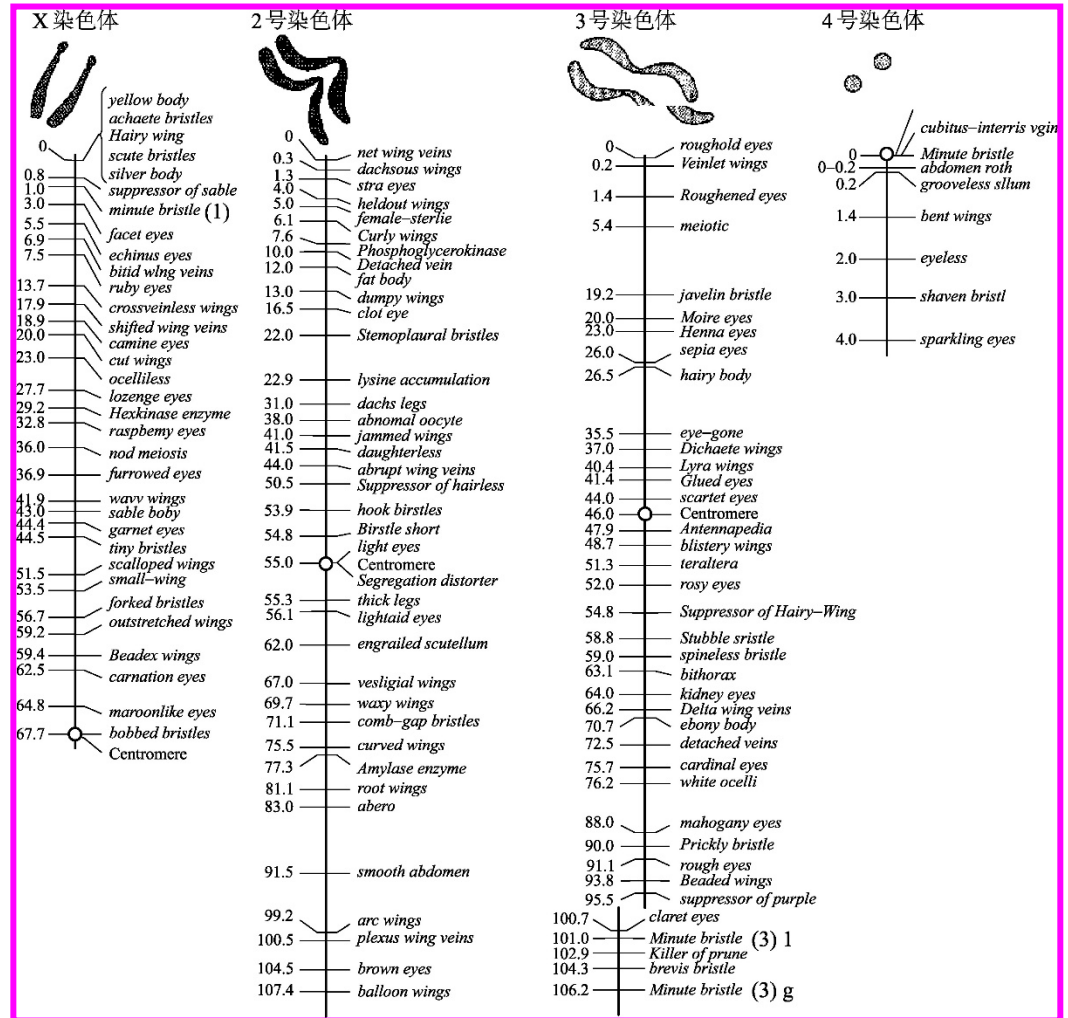
基本概念

• Gene mapping: 将基因定位于某一特定染色体上, 以及测定基因在染色体上线性排列顺序与距离的过程。



- 遗传图（Genetic map）：连锁图

（linkage map），
以遗传距离为单位，
对基因或遗传标记
在染色体上的相对
位置进行排列所绘
制的线性图。



- 科学史话：Sturtevant于1913年首次构建遗传图（P154）

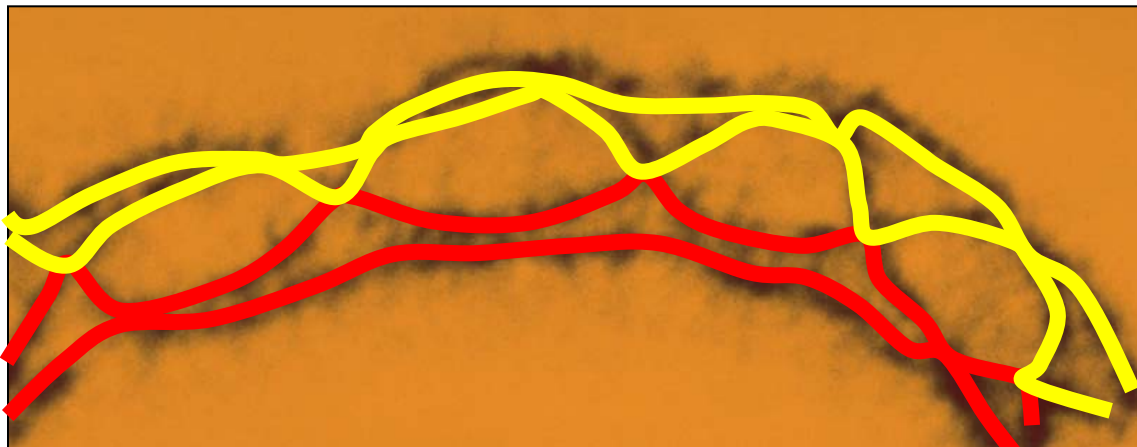
- 遗传距离即图距（map distance）：两个基因在染色体图上距离的数量单位，基因在染色体上的一个距离单位（map unit, mu）。
- 将图距单位称“厘摩”（centimorgan, cM）。

基因定位的基本原理

回顾：什么叫不完全连锁？

Morgan的三条假设——交换（crossing over）

- 连锁的基因在染色体上呈线性排列
- 在任何两个基因之间都能发生交换
- 距离越远的基因之间发生交换的次数也越多

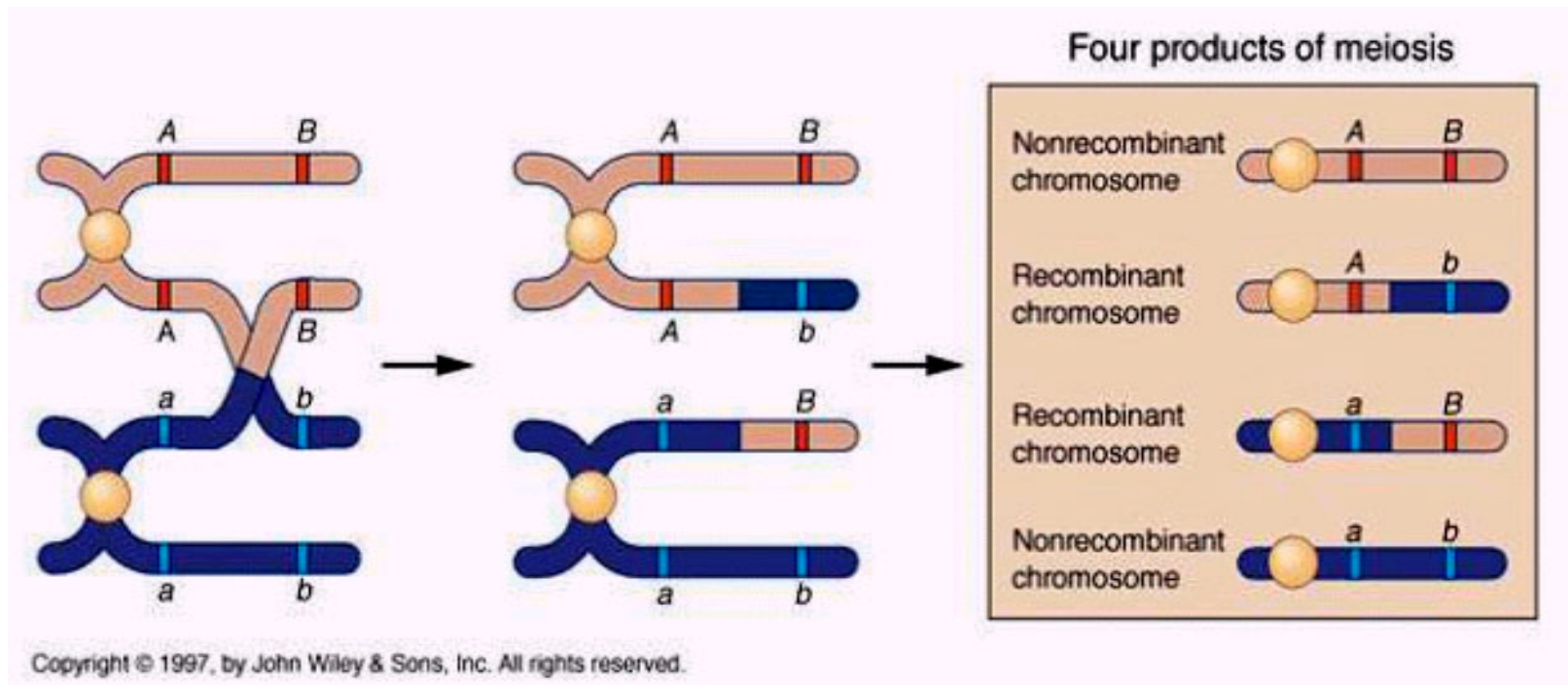


摩尔根于1911年提出：

重组值大小反映基因座在染色体上距离的远近。

- 将交换的百分率确定为染色体上基因之间的相对距离单位，就可能根据交换率确定基因在染色体上的相对位置。
- 以重组值去掉%号表示基因在染色体上的一个距离单位（map unit, mu），即“厘摩”（cM）。
- 两基因 $a-b$ 间重组率为17%，两个基因相距17个图距单位（17 cM）。

- 基因定位的基础：遗传重组
- 遗传重组大多发生在减数分裂时同源染色体间的交换，但有丝分裂中也可发生遗传物质的交换——有丝分裂交换。



基因定位的意义

以囊性纤维化（cystic fibrosis, CF）病为例

- 欧洲中世纪（公元5世纪后期到公元15世纪中期）有一个传说。



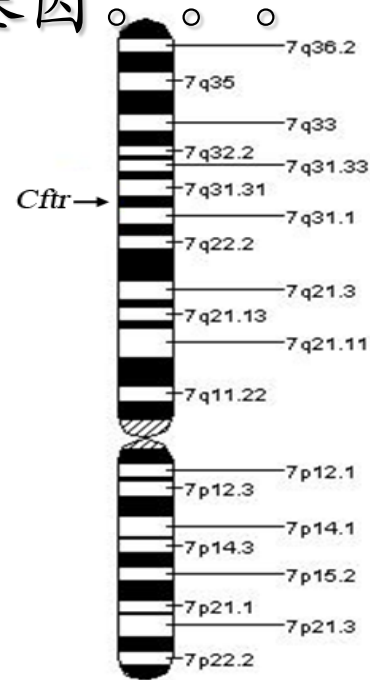
- 直到1938年，美国医生安德森（Andersen）在解剖一些死于营养不良的儿童尸体时，发现有些儿童的胰腺存在很多胰管扩张呈囊状和广泛的纤维化，称其为“胰腺囊性纤维化”，“囊性纤维化”病名由此而来。

囊性纤维化病变导致**反复胰腺炎**和**慢性胰腺功能不全**。还有肺部病变，导致反复的呼吸道感染；消化道病变，导致消化液分泌不足，吸收不良，进一步加重营养不良。



- 1949年，纽约市急救中心的圣依溺斯医生偶然发现了CF患儿的汗液比别的儿童盐分更高，相应的更容易发生脱水，死亡率也更高，证实了欧洲中世纪的那个传说。其所使用的汗液测试也成了CF诊断的最常用检查，沿用至今。 **(CF研究的第一大突破)**
- 此后的二三十年里对解开囊性纤维化发病机制之谜进展缓慢。该病为常染色体隐性遗传病。
- 对CF研究的突破发生在1980年代。其中，加州大学洛杉矶分校生理学博士后保罗·昆顿发现自己长期的呼吸道疾病很像CF。被确诊后，他致力于CF研究。昆顿发现了CF患者汗腺上皮细胞的**氯离子通道异常**。
(CF研究的第二大突破)

- 离子通道就是细胞表面的特殊蛋白质，找基因。
- 1985年，通过家系连锁分析将CF相关基因定位于7q31~7q32约500kb区域。
- 1988年，在第7染色体上发现CF的第一个突变 $\Delta F508$ （三个核苷酸的缺失导致CFTR第508位的苯丙氨酸缺失），该突变占全球CF的2/3，在美国占到CF的90%。
- 1989年，J. M. Rommens等鉴定了CF基因*Cftr*，全长250 kb，编码一由1 480个氨基酸组成的蛋白质-囊性纤维化跨膜转运调节因子（CFTR）。基因突变使CFTR功能失活，导致具有外分泌功能的上皮组织出现结构缺陷或功能障碍。（**第三大突破**）



- 同年（1989），其基因突变谱立刻受到关注，基因突变分析立即被用于家系患者的诊断、产前诊断和携带者筛查。建立了*Cftr*基因突变数据库及其在不同种族中的分布频率。目前已报道1 200多种*Cftr*突变等位基因形式与疾病关联和700多种良性变异形式，包括外显子、内含子和启动子在内的整个基因。发现了CF临床异质性如影响胰腺功能和导致先天性无输精管（至少97%的CF男性患者不育）的遗传因素。
- 1992年，CF的胚胎植入前遗传学诊断成功；首次构建了携带*Cftr*突变基因的小鼠模型。

- 1994年，首次尝试将正常*Cftr*基因转入肺上皮细胞治疗CF病（基因治疗）。
- 1997年，在美国开展了对几十种*Cftr*杂合子的筛查。
- 2003年，根据*Cftr*突变图谱研制开发出治疗CF的药物取得初步进展。这些药物作用于CFTR蛋白表达和功能及由*Cftr*基因突变所引起的离子和水分子转运异常。

- 2019年10月，美国FDA批准Vertex制药公司的新药Trikafta用于年龄 ≥ 12 岁人群治疗某些类型的CF，适用于*Cftr*基因至少携带1个拷贝F508del突变的CF患者。
- 2020年9月，Vertex制药公司宣布，美国FDA已批准Kalydeco (ivacaftor)，用于治疗4-6个月大的CF婴儿患者，具体治疗对象为：在*Cftr*基因中存在至少一个突变对Kalydeco治疗有反应的CF婴儿患者。

科学家们在囊性纤维化疾病研究领域的新成果

- **【1】 Nat Commun:** 科学家有望利用CRISPR-Cas技术剔除囊性纤维化致病基因突变。
- **【2】 Science:** 揭示铰链状蛋白结构，为治疗囊性纤维化带来新思路
- **Science子刊:** 揭示囊性纤维化患者为何容易遭受铜绿假单胞菌感染
- **【4】 Nat Commun:** 新的检测方法可能导致囊性纤维化的个性化治疗
- **【5】 Nature:** 常用的抗真菌药物—两性霉素有望治疗囊性纤维化

仍存在的问题：

- 尽管引发囊性纤维化的原因很简单，但其治疗方法却不简单。大部分成功只限于特定种类基因突变的CF患者，许多其他患者仍未得到有效治疗。
- 导致大多数CF基因治疗失败的主要挑战是将治疗方法传送到肺部。对于肺，需要大量的病毒载体，特别是如果要将其雾化，可能有一半的载体会因为雾化而丢失。
- CRISPR/Cas9技术有可能直接恢复*Cftr*的突变。

- 从上面的CF研究历史可以看出，通过对CF相关基因的定位，推动了*Cftr*的克隆和功能研究，进而促进了对CF的基础病理生理学、致病途径、动物模型、药物开发、治疗方案的研究和遗传咨询的分子诊断、产前诊断、携带者筛查及种族间CF 发病率差异和不同种族基因组的特点的临床应用等。

遗传图或基因定位的意义

- 通过遗传连锁将一种性状的遗传与另一种性状或标记相联系，因此基因定位有助于理解生物性状的遗传性质。
- 将表型差异与染色体改变相联系，可用于改良动植物性状、定位和克隆人类及动植物的相关基因等。
- 遗传连锁图也是结构基因组学研究不可缺少的技术平台。从生物性状的遗传改良到功能基因组学研究，基因定位和遗传连锁图的制作都发挥着不可替代的作用。

基因定位的基本方法

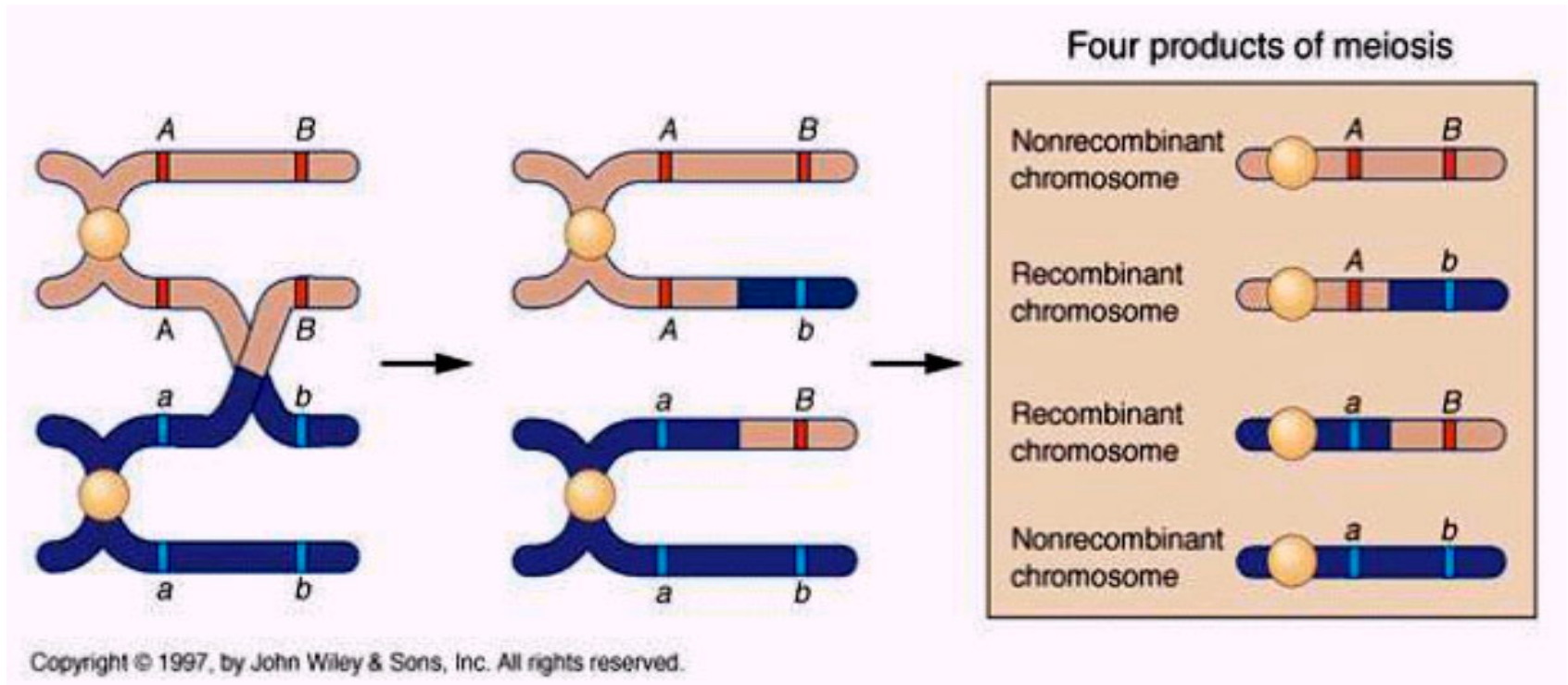
基因定位的基础或原理是什么？

遗传重组



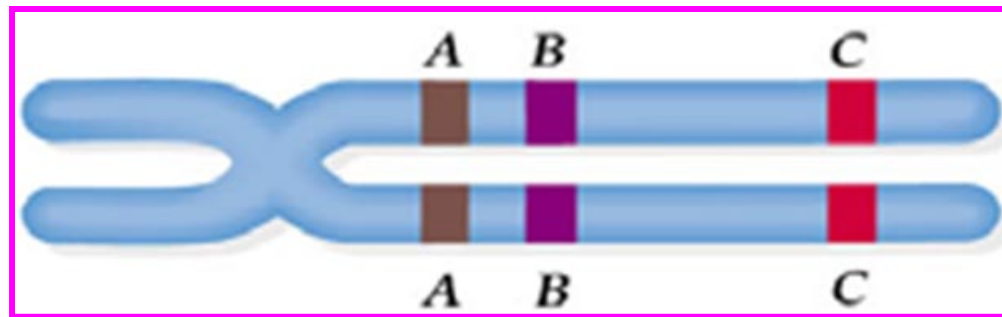
要获得可观察到的重组后代，需要注意什么？

杂合体




基因定位有两种基本方法：**两点测交**和**三点测交**。

- **两点测交**（two-point testcross）：每次测定两个基因间遗传距离而进行基因定位的方法。
- **三点测交**（three-point testcross）：通过一次测交，同时确定三对等位基因的排列顺序和它们之间遗传距离的方法。是基因定位的常用方法。



- 两点测交

- 玉米粒  糊粉层有色 C (无色 c)、胚乳饱满 Sh (凹陷 sh)，胚乳非糯性 Wx (糯性 wx)，连锁遗传。

CSh/CSh	×	csh/csh		
	↓			
	CSh/csh	×	csh/csh	
		↓		
CSh/csh	csh/csh	Csh/csh	cSh/csh	
4 032	4 035	149	152	

- 重组率=

- $[(152+149) / (4032+4035+149+152)] \times 100\% = 3.6\%$

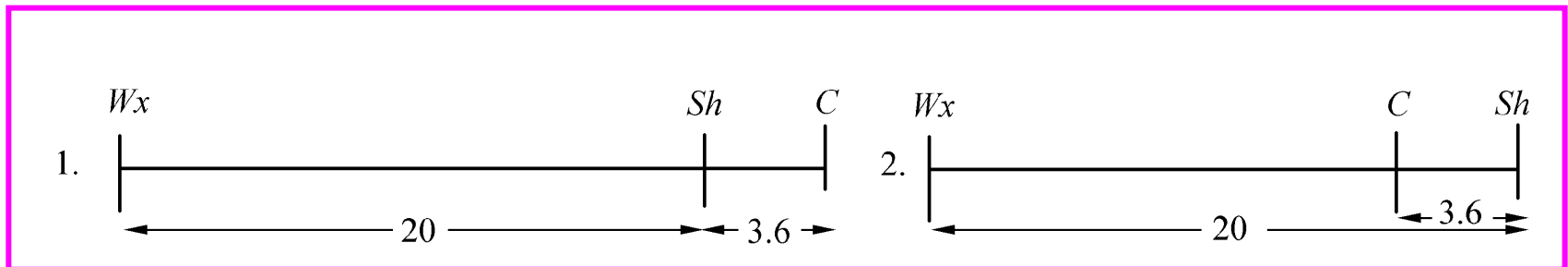
- C 与 Sh 之间的图距为3.6 cM。

• 同理： W_x 和 Sh

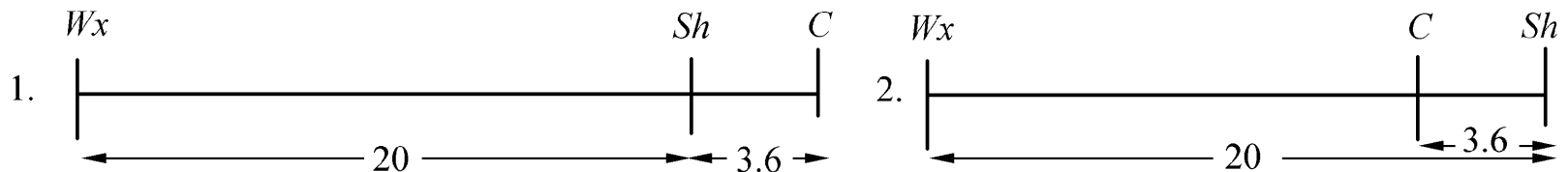
$wxSh/wxSh$		×	$Wxsh/Wxsh$	
		↓		
$wxSh/Wxsh$		×	$wxsh/wxsh$	
		↓		
$wxSh/wxsh$	$Wxsh/wxsh$		$wxsh/wxsh$	$WxSh/wxsh$
5 885	5 991		1 531	1 488

• 交换率为20%， W_x 与 Sh 间的距离为20 cM。

• 根据上述两个实验，三个基因可有两种排列方式：

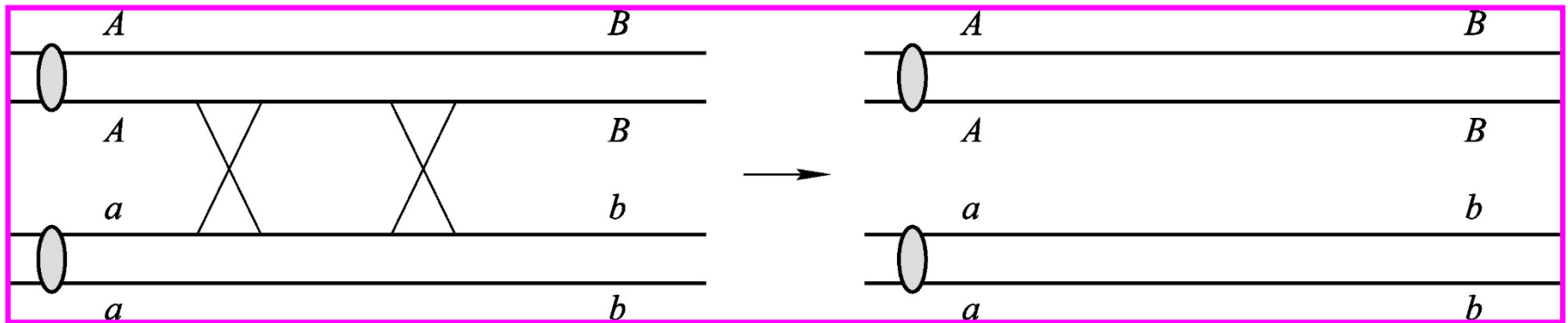


- 还需第三次杂交和测交。
- $WxC/wxc \times wxc/wxc$
- WxC/wxc (2 542) , wxc/wxc (2 716) ,
- Wxc/wxc (739) , wxC/wxc (717) 。
- 交换率为22%。
- W_x 和 C 之间的距离22“接近” 23.6，故认为第一种排列顺序更符合实际。



两点测交法的弊端：

1. 繁琐复杂。如测定3对或3对以上基因位点间的距离和顺序，需分别进行3次或多次杂交和3次或多次测交。
2. 结果不稳定。由于环境和基因间的影响，结果会发生偏差。
3. 双交换或偶数次交换无法检出。在两个基因之间双交换的结果等于没有发生交换。



三点测交

+++ / +++ × *c sh wx / c sh wx*,

F₁: +++ / *c sh wx*,

F₁ × *c sh wx / c sh wx*:

玉米三对基因的三点测交结果

F ₁ 配子的基因型			实得籽粒数
+	+	+	2 238
<i>c</i>	<i>sh</i>	<i>wx</i>	2 198
<i>c</i>	+	+	98
+	<i>sh</i>	<i>wx</i>	107
+	+	<i>wx</i>	672
<i>c</i>	<i>sh</i>	+	662
<i>c</i>	+	<i>wx</i>	39
+	<i>sh</i>	+	19
总数			6 033

亲组合

单交换

双交换

RF *c-sh*:

亲组合: 2238 + 2198 + 672 + 662 = 5770,

重组组合:

98 + 107 + 39 + 19 = 263,

263 / 6033 = 4.36%,

距离 = 4.36

计算 *sh-wx* 间距离

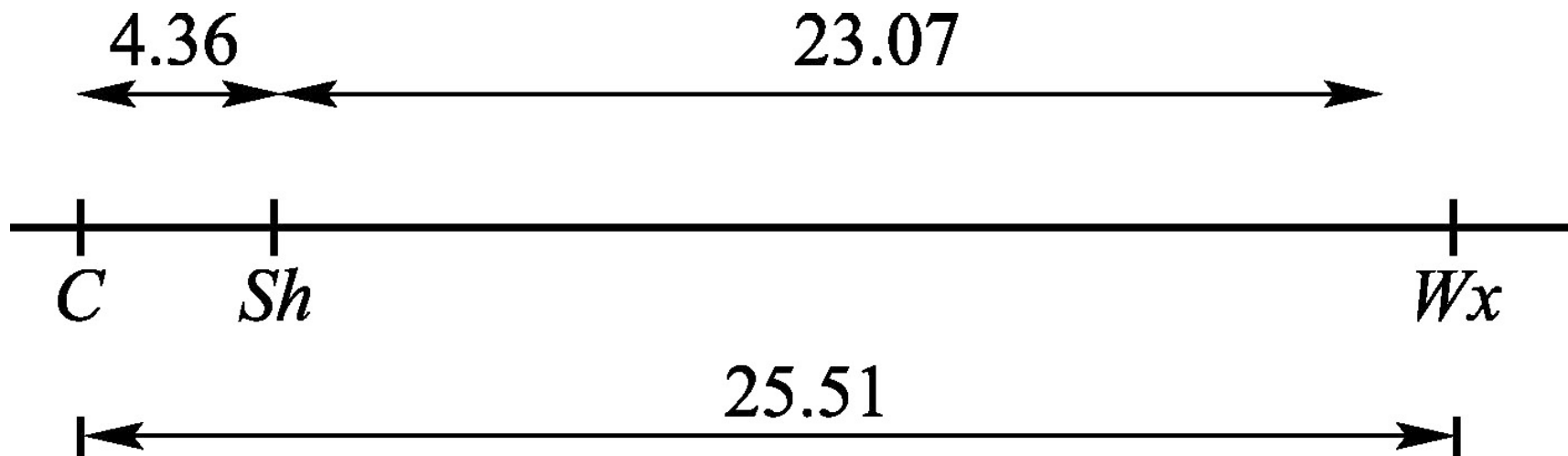
1392 / 6033 = 23.07%,

距离 = 23.07

RF *c-wx*:

1539 / 6033 = 25.51%

距离 = 25.51



分析：为什么 $25.51 \neq 4.36 + 23.07$?

因为 *C* 和 *Wx* 之间发生了双交换

F ₁ 配子的基因型			实得籽粒数
+	+	+	2 238
<i>c</i>	<i>sh</i>	<i>wx</i>	2 198
<i>c</i>	+	<i>wx</i>	39
+	<i>sh</i>	+	19

双交换

$$= (19 + 39) / 6033$$

$$= 0.96\%$$

亲组合

双交换

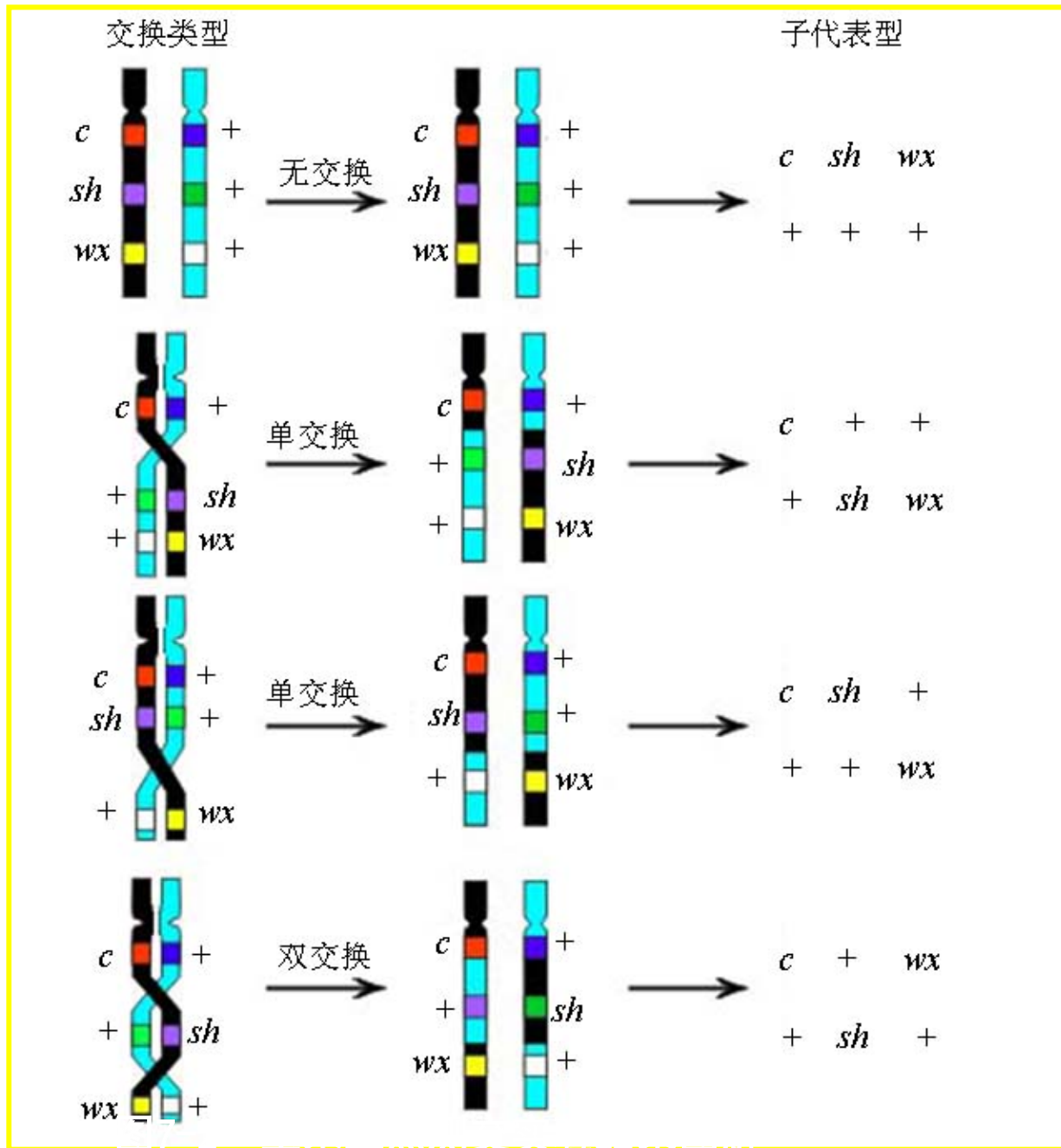
注意：由于一个双交换实际上包含了二次单交换，因此，应加上2倍双交换值。

RF_{c—wx}:

$$1539 / 6033 = 25.51\%$$

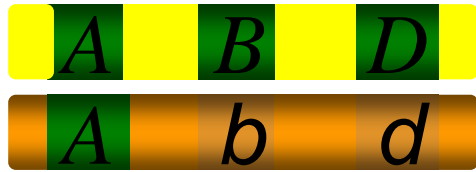
$$\text{距离} = 25.51 + 2 \times 0.96 = 27.43$$

- 三个基因的顺序是 $c-sh-wx$ ，
即遗传作图

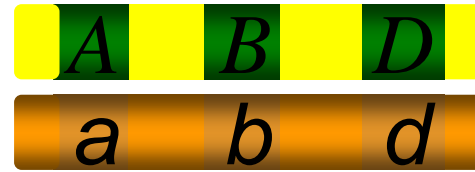


遗传作图三条原则 (Three criteria)

① 基因型都是杂合体 (heterozygous)



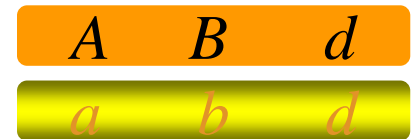
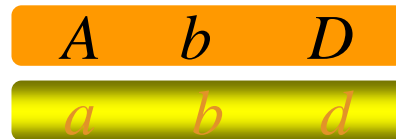
纯合体不能检测交换



杂合体才能检测交换

② 表型 (phenotypes) 对应基因型 (genotypes)

表型→



③ 子代数量足够多

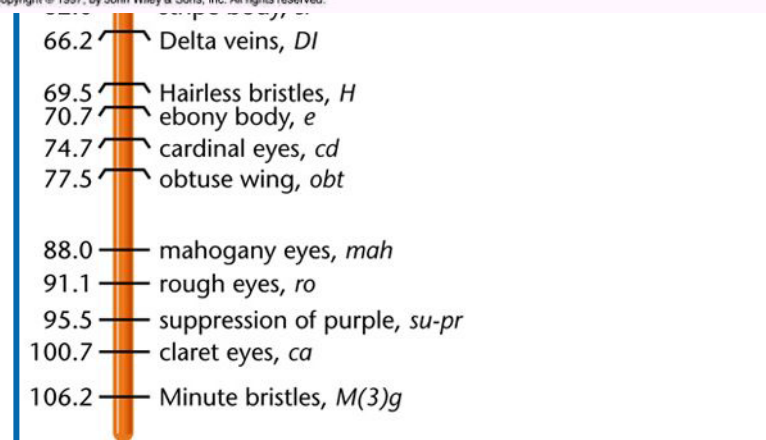
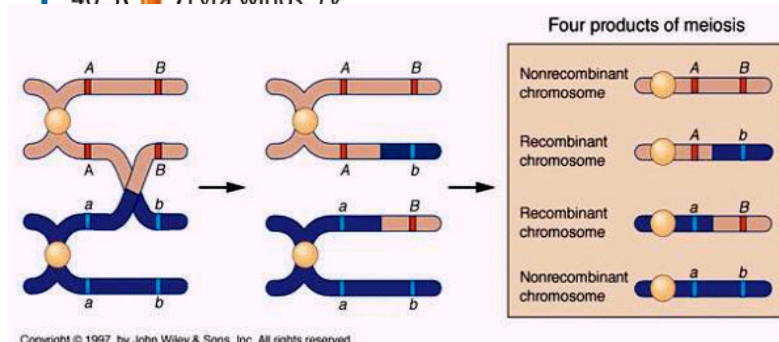
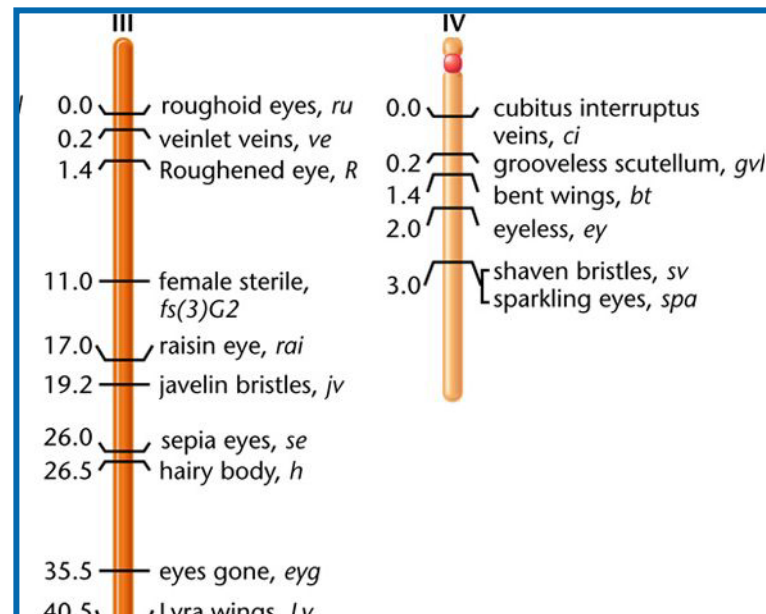
- **基因座 (locus)** : 基因在遗传图上的位置。

- **座位怎么排?** 绘制遗传图时, 以最先端基因为0, 发现新基因在更先端时, 让出0点, 其余基因位置相应移动。

- 交换值介于0到50%。

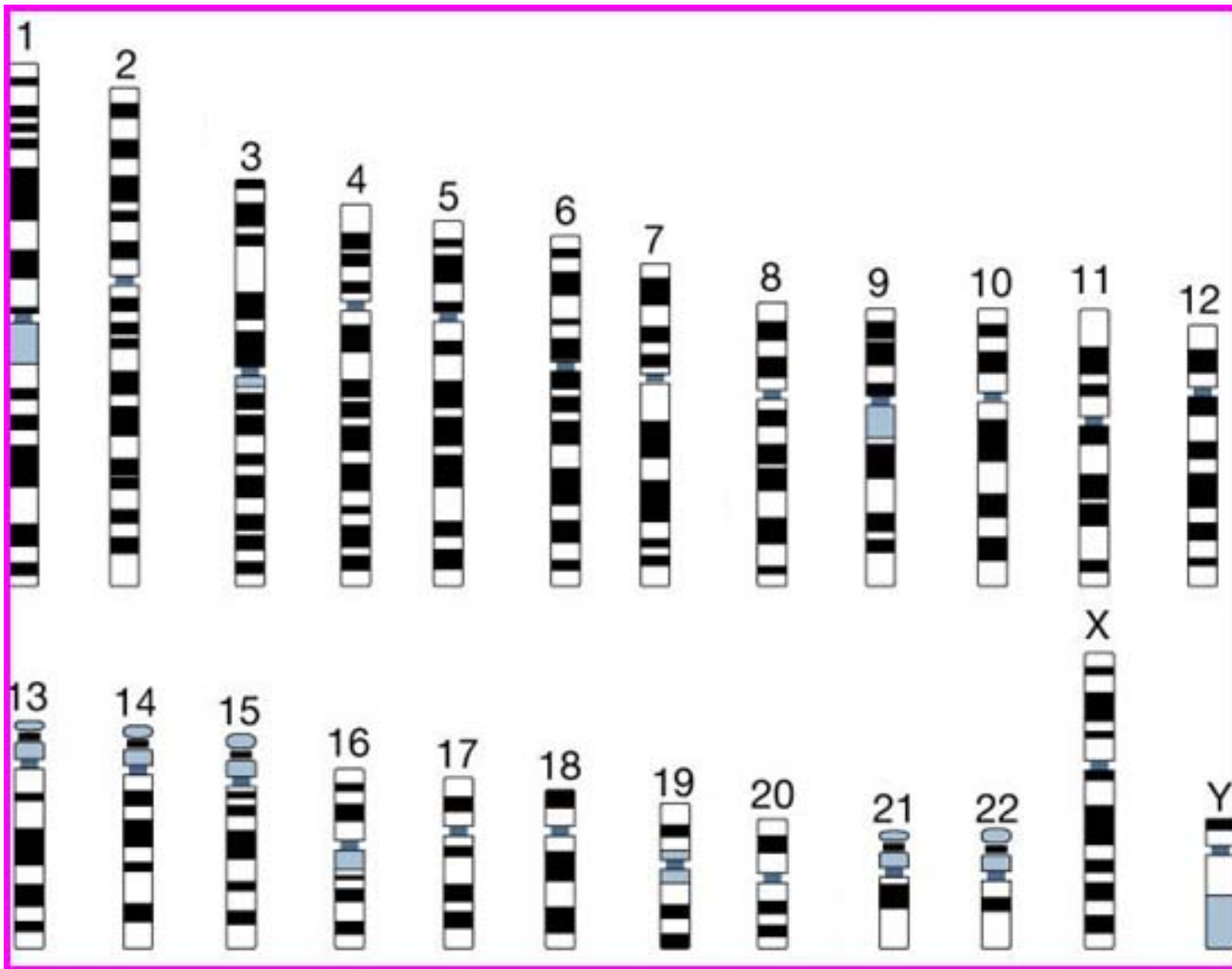
- 遗传图上可看到两基因间超过50图距单位, 是因为两基因间多次交换累加。实际上两基因间重组值不会超过50%。

- 要从图上知道基因间的重组值只限于邻近基因座位间。

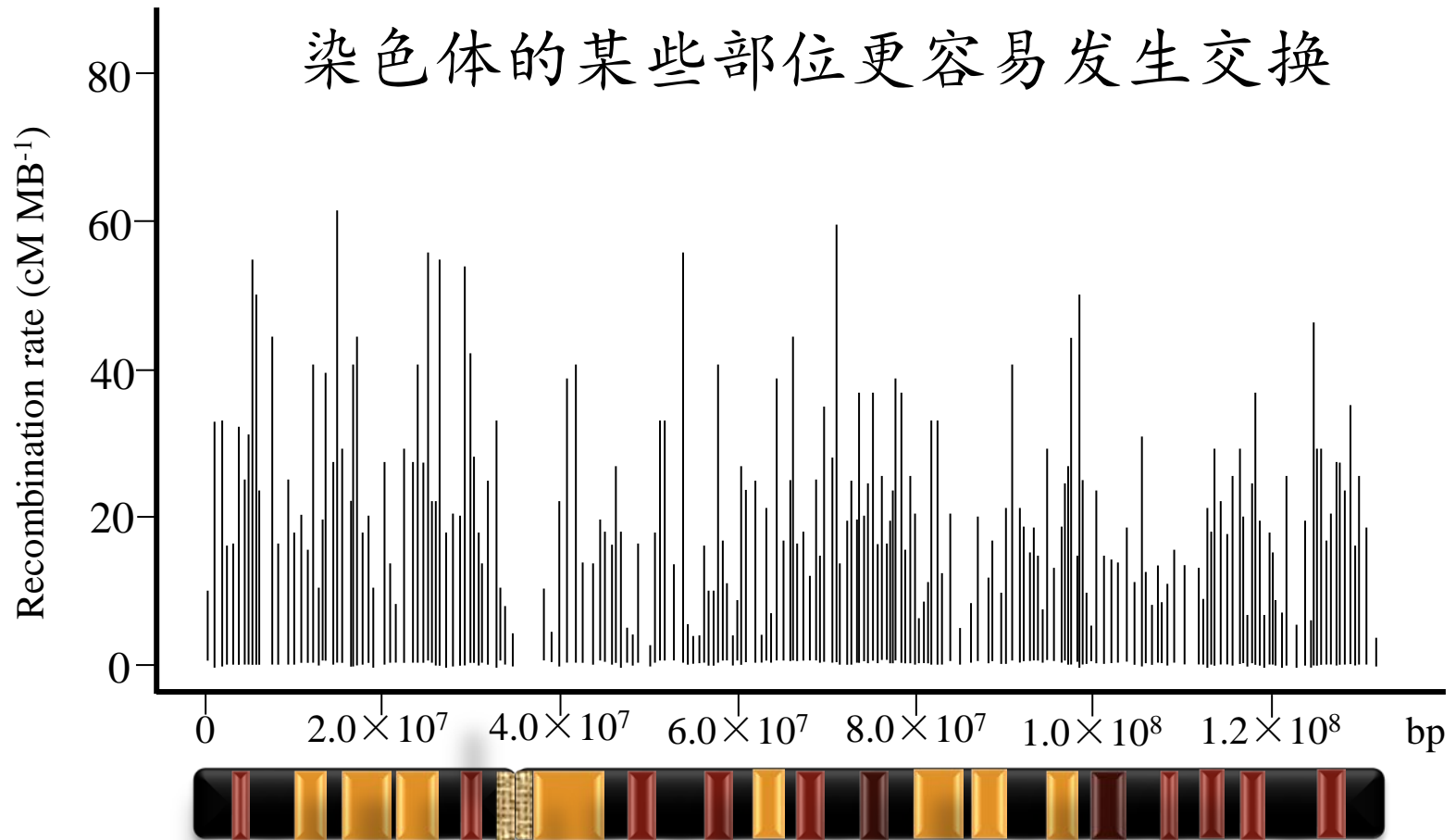


什么叫连锁群 (linkage group) ?

位于一对同源染色体上的所有基因的组合。

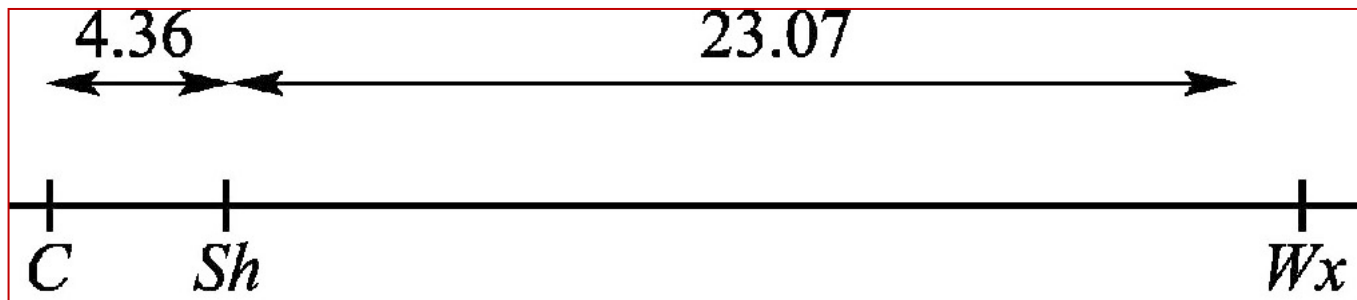


所有基因间的交换是不是一样频率？



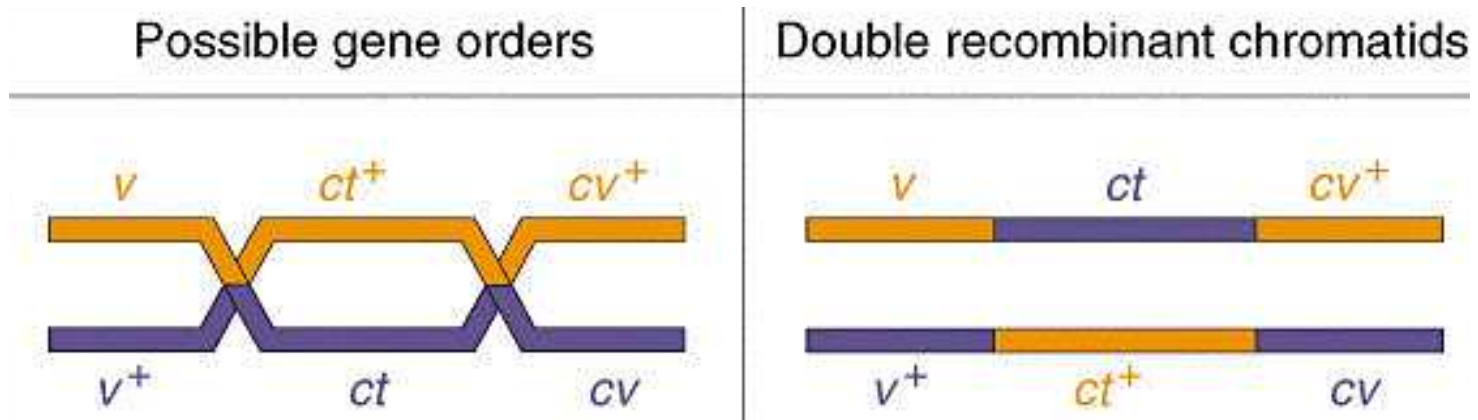
干涉和并发系数

- 染色体上三个基因之间发生双交换的**理论双交换率**是两个单交换概率的乘积，上例中为：



- 理论双交换率： $23.07\% \times 4.36\% = 1.0\%$ ，
- 可实际双交换值只有： 0.96% ，

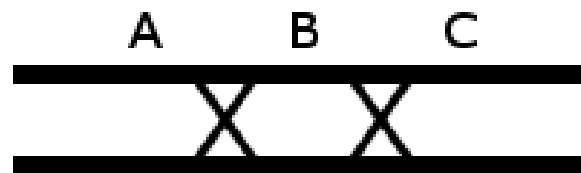
- **干涉 (interference)**：每发生一次单交换时，其邻近基因间也发生一次交换的机会减少的现象。



干涉由美国著名遗传学家穆勒于1916年最先在果蝇中报道并命名 (Muller, American Naturalist, 1916)



- 干涉程度常用**并发系数**表示



并发系数 (coefficient of coincidence, c) : 实际双交换率与理论双交换率的比值。

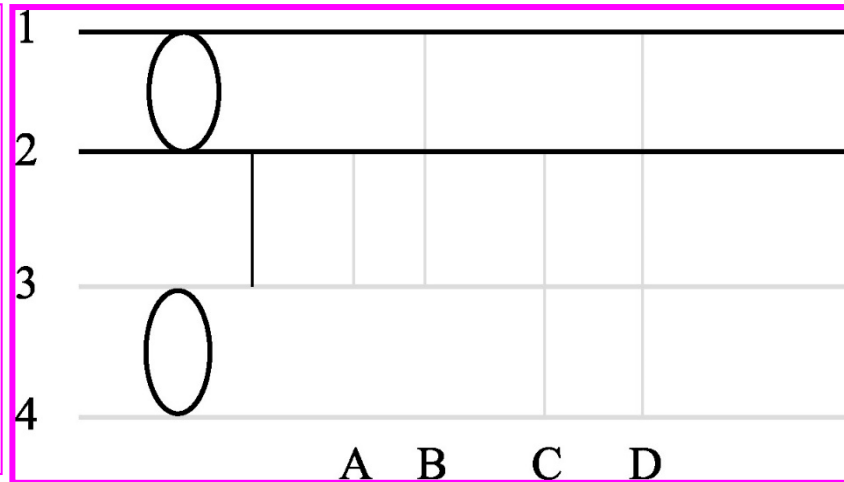
$$\text{并发系数}(c) = (\text{实际双交换值}) / (\text{理论双交换值})$$

负干涉 (negative interference) : 并发系数大于1, 表示一次交换的发生使第二次发生交换的频率增加了。

染色体干涉 (chromosomal interference) : 染色体内基因区段之间的干涉。

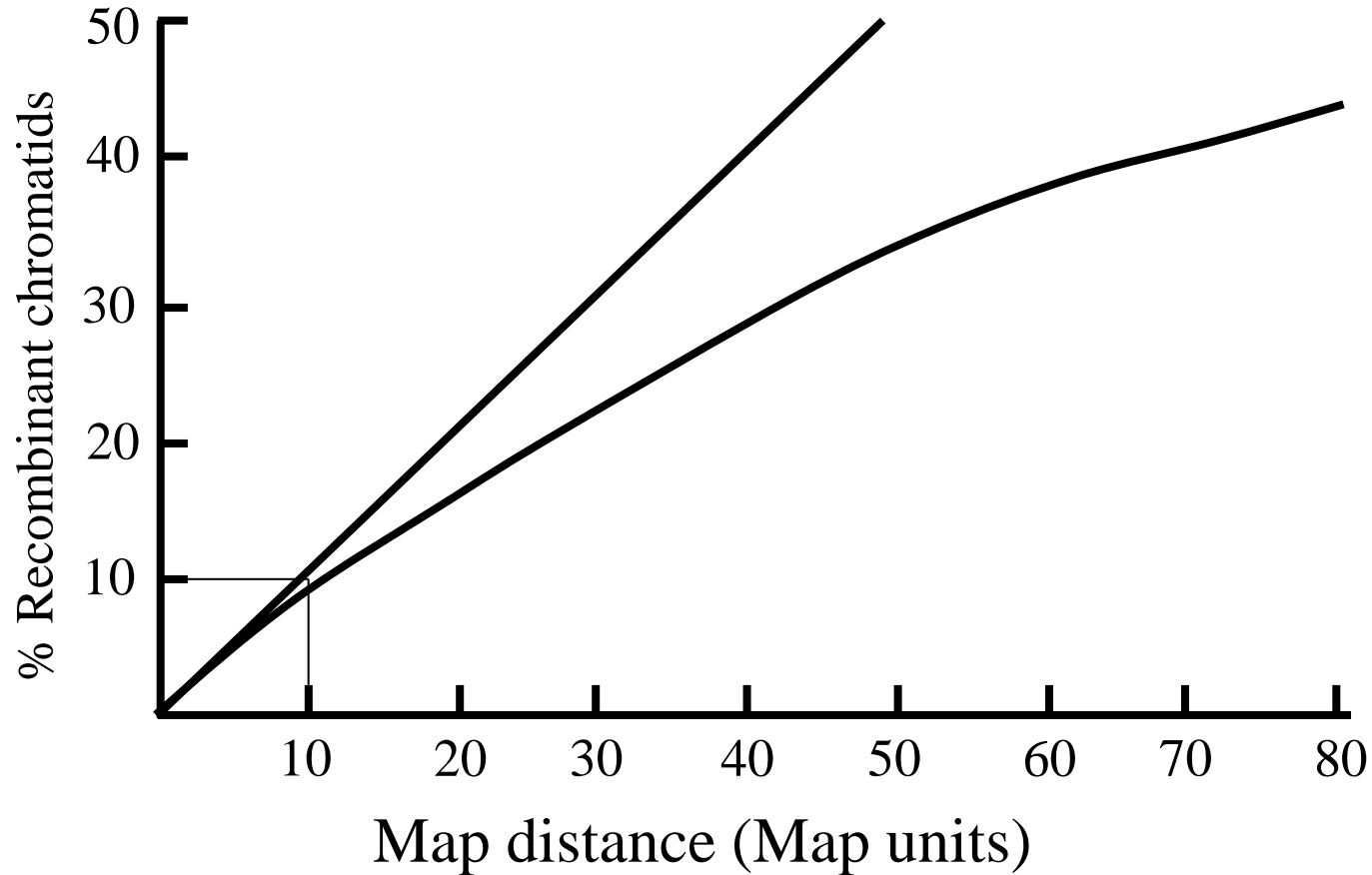
染色单体干涉 (chromatid interference) : 减数分裂前期, 交换并不限于某二条非姊妹染色单体之间。非姊妹染色单体间的一次单交换会干涉其它非姊妹染色单体间的交换。

通常, 第一次交换除对同线双交换有干涉外, 对其他双交换没有干涉



- A. 2-3二线双交换
- B. 1-3三线双交换
- C. 2-4三线双交换
- D. 1-4四线双交换

基因越近，干涉越强，双交换越少，作图越精准



利用基因编辑技术实现水稻遗传干涉现象的完全消除

2021年10月1日，Frontiers in Plant Science基因编辑专刊在线发表了王克剑团队和程祝宽团队合作的题为“Concurrent Disruption of Genetic Interference and Increase of Genetic Recombination Frequency in Hybrid Rice Using CRISPR/Cas9”的研究论文。该研究利用基因编辑技术对杂交稻减数分裂联会复合体中央元件基因进行编辑，同步实现了遗传干涉现象的完全消除与遗传重组频率的增加，为水稻种质资源创新和遗传多样性创造提供了一个高效途径。

进一步对突变体进行授粉并对后代进行遗传分析发现：突变体中遗传重组频率提高了1.5倍左右，特别是在着丝粒重组冷点区域甚至提高到3.3倍之多。令人意外的是，在同一染色体的相邻区间相互干涉的强度几乎降为0，表明突变体中交叉干涉现象基本完全消失。该结果证明水稻联会复合体可能参与介导交叉干涉的形成。此外，该研究也为利用基因组编辑技术打破遗传累赘、实现优异基因聚合、提高育种效率提供了一种有效策略。

遗传标记、物理图谱与遗传图谱

什么是遗传标记 (Genetic marker) ?

指可追踪染色体、染色体某一节段、某个基因座、某性状在家系中传递的任何一种遗传特性。

可识别性

可遗传性

表 8-2 用于遗传作图的常用标记

标记	定义和应用
形态学表型	形态学表型变异（包括疾病等）。用于符合孟德尔遗传的形态学变异的作图。受显性关系复杂性、非等位基因相互作用和环境效应的影响，人类疾病基因一般是二态表型变异
细胞学标记	指染色体数目和形态的变异，主要包括：染色体核型和带型及缺失、重复、易位、倒位等
蛋白多态性	在电泳和等电聚焦时蛋白迁移率的差异。共显性和中等的多态性在数量上有限。很多蛋白的多态性不能通过电泳检测，因为有些氨基酸替代不改变蛋白的物理化学性质。同时，基因编码的多态性蛋白本身可能不被作图
限制性片段长度多态性 (RFLP)	等位 DNA 片段中限制性位点的差异。共显性且丰富，但常为二态性。如果变异位于限制性位点之外，很多顺序多态性不能被检测
VNTR, STR	由于不同个体串联重复序列的数目和位置不同所引起的片段长度差异。呈共显性和高度多态性。小卫星 DNA 组成的染色体座位具有丰富的多态性，即 VNTR 序列。微卫星 DNA (STR) 在整个基因组中平均分布，可以用 PCR 快速测定基因型
随机扩增的多态性 DNA (RAPD)	随机引物扩增的 DNA 片段。丰富，但并非总可以区分纯合和杂合基因型，有时重复性较差。主要用于植物基因组作图
单核苷酸多态性 (SNP)	基因组内某一特定核苷酸位置上的核苷酸差异。在人基因组中平均每 1 000 个核苷酸中有一个差别，数目多，覆盖密度大

微卫星标记

分子标记为什么具有遗传标记特性？

亲本间存在差异即多态性，即具可识别性；

亲本间存在的差异在后代中可以重演，即具有可遗传性。

A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes

The Huntington's Disease Collaborative Research Group*

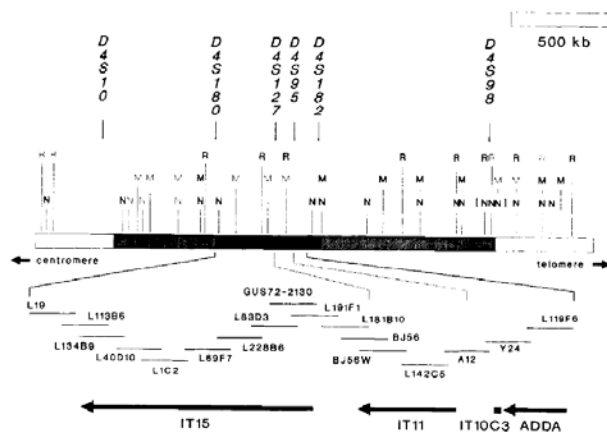


Figure 1. Long-Range Restriction Map of the HD Candidate Region

*The Huntington's Disease Collaborative Research Group comprises:

Group 1:

Marcy E. MacDonald,¹ Christine M. Ambrose,¹ Mabel P. Duyao,¹ Richard H. Myers,² Carol Lin,¹ Lakshmi Srinidhi,¹ Glenn Barnes,¹ Sherryl A. Taylor,¹ Marianne James,¹ Nicolet Groot,¹ Heather MacFarlane,¹ Barbara Jenkins,¹ Mary Anne Anderson,¹ Nancy S. Wexler,³ and James F. Gusella^{1†}

¹Molecular Neurogenetics Unit Massachusetts General Hospital and Department of Genetics Harvard Medical School Boston, Massachusetts 02114

²Department of Neurology Boston University Medical School Boston, Massachusetts 02118

³Hereditary Disease Foundation 1427 7th Street, Suite 2 Santa Monica, California 90401

Group 2:

Gillian P. Bates, Sarah Baxendale, Holger Hummerich, Susan Kirby, Mike North, Sandra Youngman, Richard Mott, Gunther Zehetner, Zdenek Sedlacek, Annemarie Poustka, Anna-Maria Frischauf, and Hans Lehrach

Genome Analysis Laboratory Imperial Cancer Research Fund Lincoln's Inn Fields London, WC2A 3PX, England

Group 3:

Alan J. Buckler,¹ Deanna Church,¹ Lynn Doucette-Stamm,¹ Michael C. O'Donovan,¹

Laura Riba-Ramirez,¹ Manish Shah,¹ Vincent P. Stanton,¹ Scott A. Strobel,² Karen M. Draths,² Jennifer L. Wales,² Peter Dervan,² and David E. Housman¹

¹Center for Cancer Research Massachusetts Institute of Technology Cambridge, Massachusetts 02139

²Division of Chemistry and Chemical Engineering California Institute of Technology Pasadena, California 91125

Group 4:

Michael Altherr, Rita Shiang, Leslie Thompson, Thomas Fielder, and John J. Wasmuth
Department of Biological Chemistry University of California Irvine, California 92717

Group 5:

Danilo Tagle, John Valdes, Lawrence Elmer, Marc Allard, Lucio Castilla, Manju Swaroop, Kris Blanchard, and Francis S. Collins
Department of Internal Medicine and Human Genetics and The Howard Hughes Medical Institute University of Michigan Ann Arbor, Michigan 48109

Group 6:

Russell Snell, Tracey Holloway, Kathleen Gillespie, Nicole Datson, Duncan Shaw, and Peter S. Harper
Institute of Medical Genetics University of Wales College of Medicine Cardiff, CF4 4XN, Wales

[†]Correspondence should be addressed to James F. Gusella.

基因是非常有用的遗传标记，但为什么还需要其他标记？

- 可用作标记的基因十分有限。许多性状都涉及多基因（数量性状遗传，通过实验无法区分）；只有部分基因其等位基因成员可以通过常规实验予以区分，导致遗传图不完整。
- 高等生物基因组中存在大量间隔区和重复序列，使得大片段无标记。

人类遗传图用6000多个遗传标记作路标，把基因组分成6000多个区域，连锁分析时，找到某表型（疾病）基因与其中某一标记紧密连锁的证据，即可定位基因。

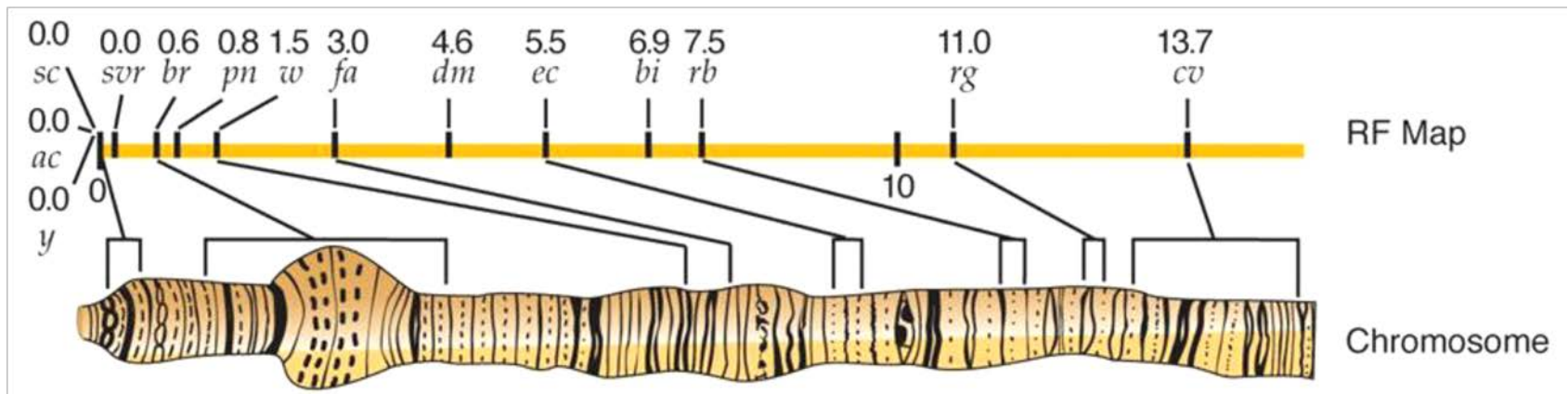
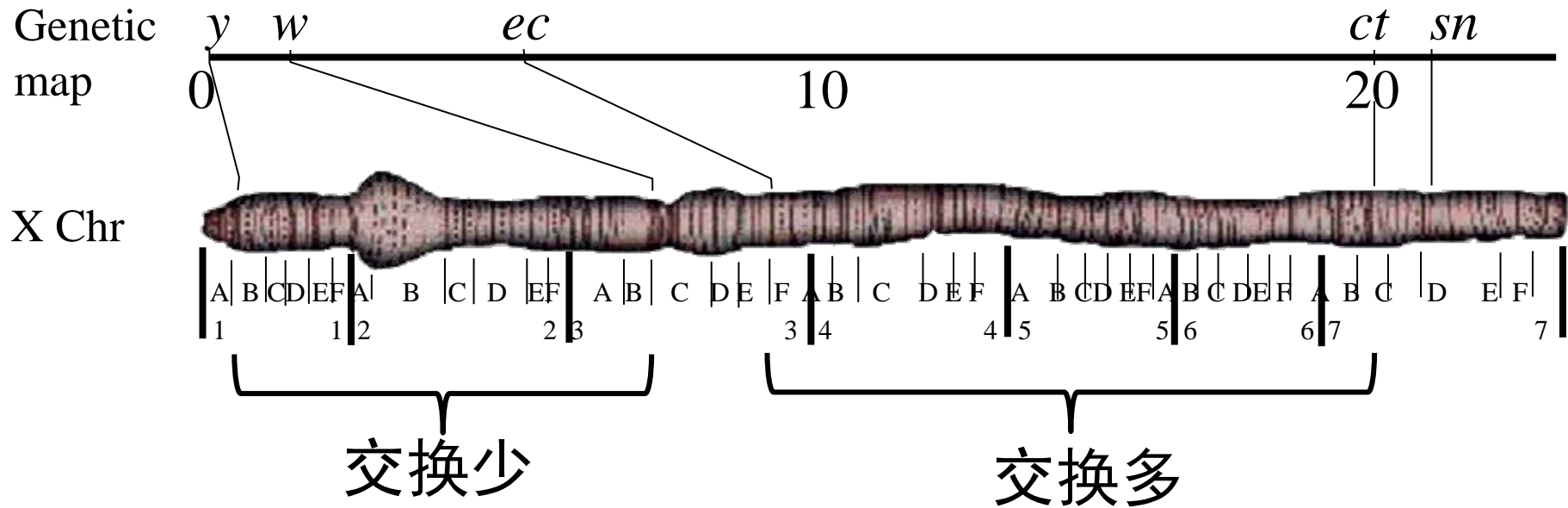
什么叫物理图？

应用分子生物学技术直接将DNA分子标记、基因等标定在染色体（基因组）实际位置所制作的图谱。

为什么要绘制物理图？

- ◆遗传图谱的分辨有限：测序要求标记间小于100kb，人类遗传图标记的平均密度实际是599kb。在人类中由于子代群体大小、婚配方式受限，无法获得高分辨率遗传图。
- ◆遗传图谱的精确性有限：重组热点、着丝粒区、不同性别具有不同的重组热点。
- ◆遗传图谱的准确率有限：环境因素和取样误差。可以相互校正。
- ◆基因组太大，必须分散测序再组装，组装时需要图谱指导；
基因组中存在大量重复序列，干扰排序。

遗传图距 (Genetic Distance) 与物理图距 (Physical Distance)



制作物理图谱的主要方法：

限制性作图 (restriction mapping) 。

荧光原位杂交 (fluorescent hybridization, FISH)

序列标签位点作图 (sequence tagged sites mapping, STS)

遗传图谱与物理图谱的异同：

- ①遗传图谱基于重组频率，物理图谱基于染色体结构或直接测量的DNA结构。
- ②减数分裂重组的频率沿染色体并不一致，有热点和冷点位点。
- ③遗传图谱是表示基因或标记间的相对距离，以重组值表示；物理图谱表示的是基因或标记间的物理距离，距离的单位为长度单位，如 μm 或者碱基对数（bp或kb）等。

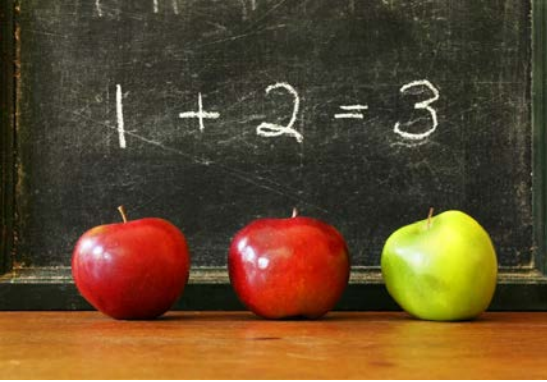
遗传图的意义：通过遗传图谱，我们可以大致了解各个基因或DNA片段之间的相对距离与方向，如哪个基因更靠近着丝粒，哪个更靠近端粒等。遗传图谱不仅是现阶段定位基因的重要手段，即使在很多物种基因组全物理图谱建立起来之后，它依然是研究基因组遗传与变异以及克隆基因的重要手段。



不同生物的基因定位方法是否一样？

- 细菌：根据细菌之间遗传物质转移的不同方式（转化、接合、转导）而异，有中断杂交、高频重组与性导、转化、重组和转导作图等；
- 噬菌体：遗传重组进行作图；
- 真菌：顺序四分子分析和非顺序四分子分析；
- 高等动植物：杂交法，利用基因间的交换值定位；
- 人类：家系分析法、体细胞杂交、全基因组扫描及核酸杂交技术等；
- 利用有丝分裂交换进行基因定位。

但基因定位的原理及基本方法不变。



Summary

基本概念：基因定位，干涉，并发系数，基因座，连锁群，遗传标记，遗传图，物理图，。。

基因定位的基本原理：遗传重组。

基因定位的基本方法：两点测交和三点测交。

不同的生物类群，基因定位的方法有所不同。