Chapter 13

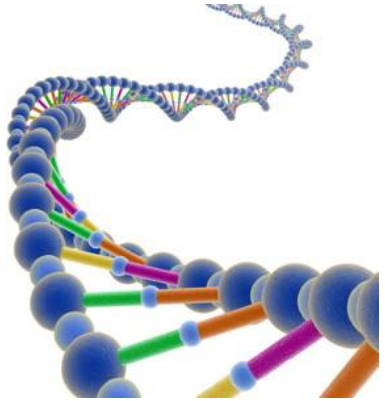
Microbial genetics

part 1

Deep in the cavern of the infant's breast The father's nature lurks, and lives anew

---

Horace ,Odes



遗传学的基本概念

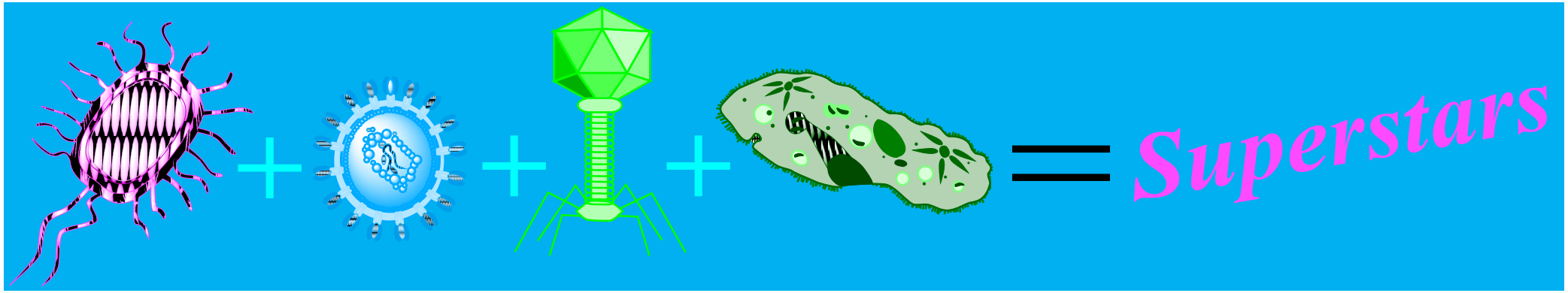
遗传学（genetics）：研究基因的结构、传递、功能和表达规律的学科。
微生物遗传学（microbial genetics）:研究微生物基因的结构、传递、功能和表达规律的学科。

微生物细胞结构简单，营养体一般为单倍体，方便建立纯系

很多常见微生物都易于人工培养，快速、大量生长繁殖

对环境因素的作用敏感，易于获得各类突变株，操作性强

微生物是遗传学研究中的明星！

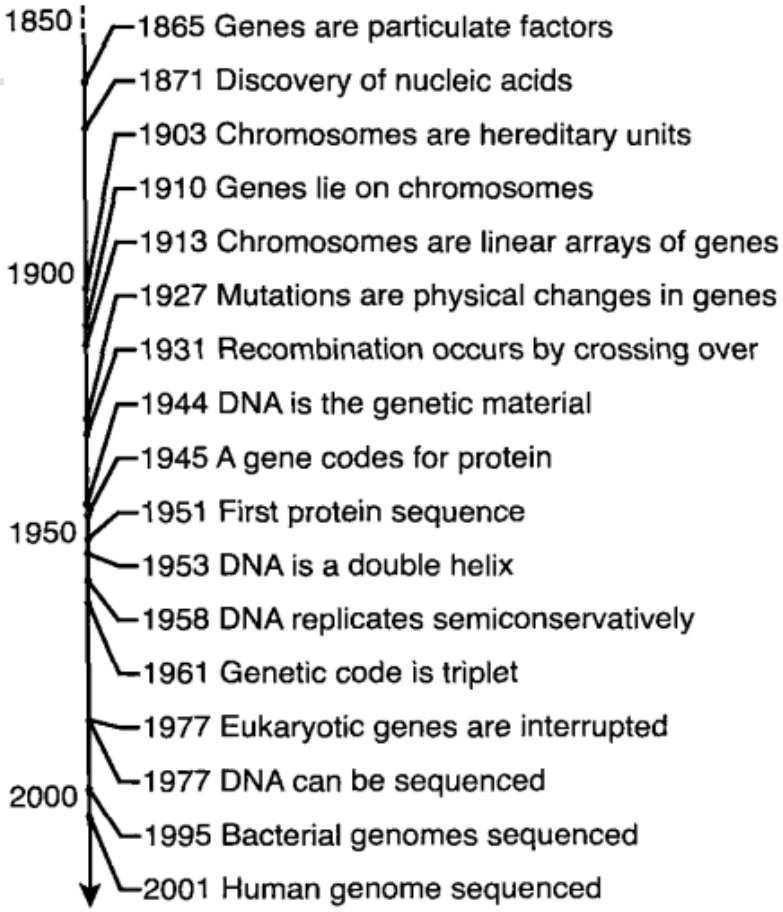


But the most important qualification of bacteria for genetic studies is their extremely rapid rate of growth. . . .

—R. F. Weaver and P. W. Hedric

微生物遗传学的研究内容

1.微生物的遗传物质、基因、基因组及染色体结构；
2.质粒：概念、结构、形态、类型
3.微生物基因组学
4.DNA的复制、转录和表达和调控（略）
5.基因突变：概念、机制和类型；
6.DNA修复；
7.转座因子：概念、类型、机制；
8.基因转移及其定位：结合、转化和转导；
9.微生物育种。

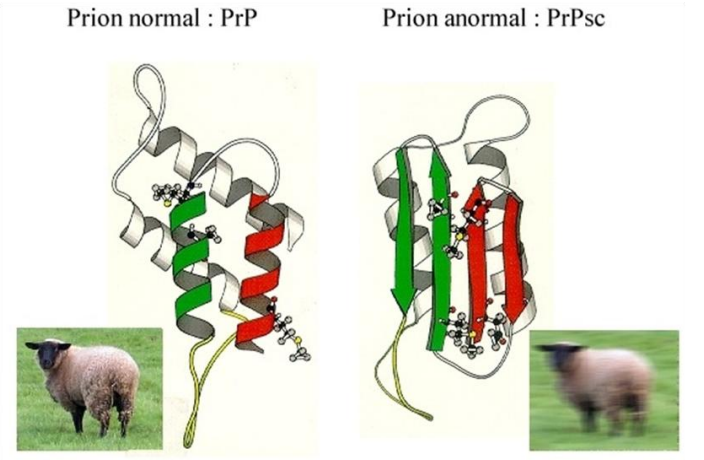


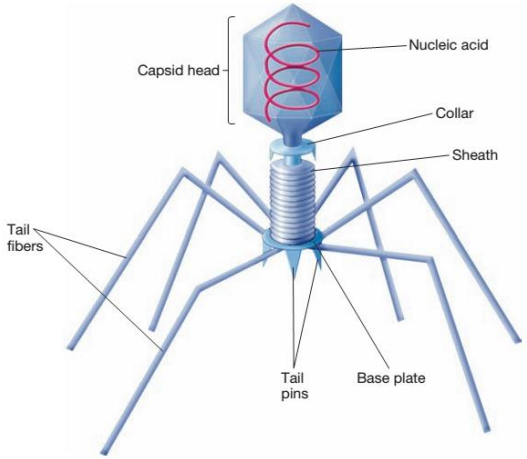
遗传研究大事记

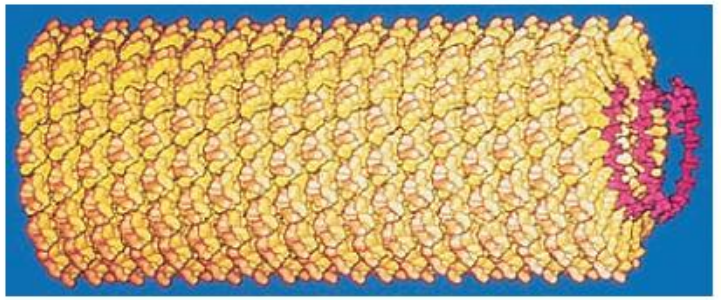
遗传的物质基础

一、DNA作为遗传物质
二、RNA作为遗传物质
三、朊病毒的发现和思考

Classic image of a cow with BSE. A feature of such disease is the inability of the infected animal to stand.







TMV

1).DNA 作为遗传物质.1928

英国微生物学家Fred. Griffith发现了转化现象。

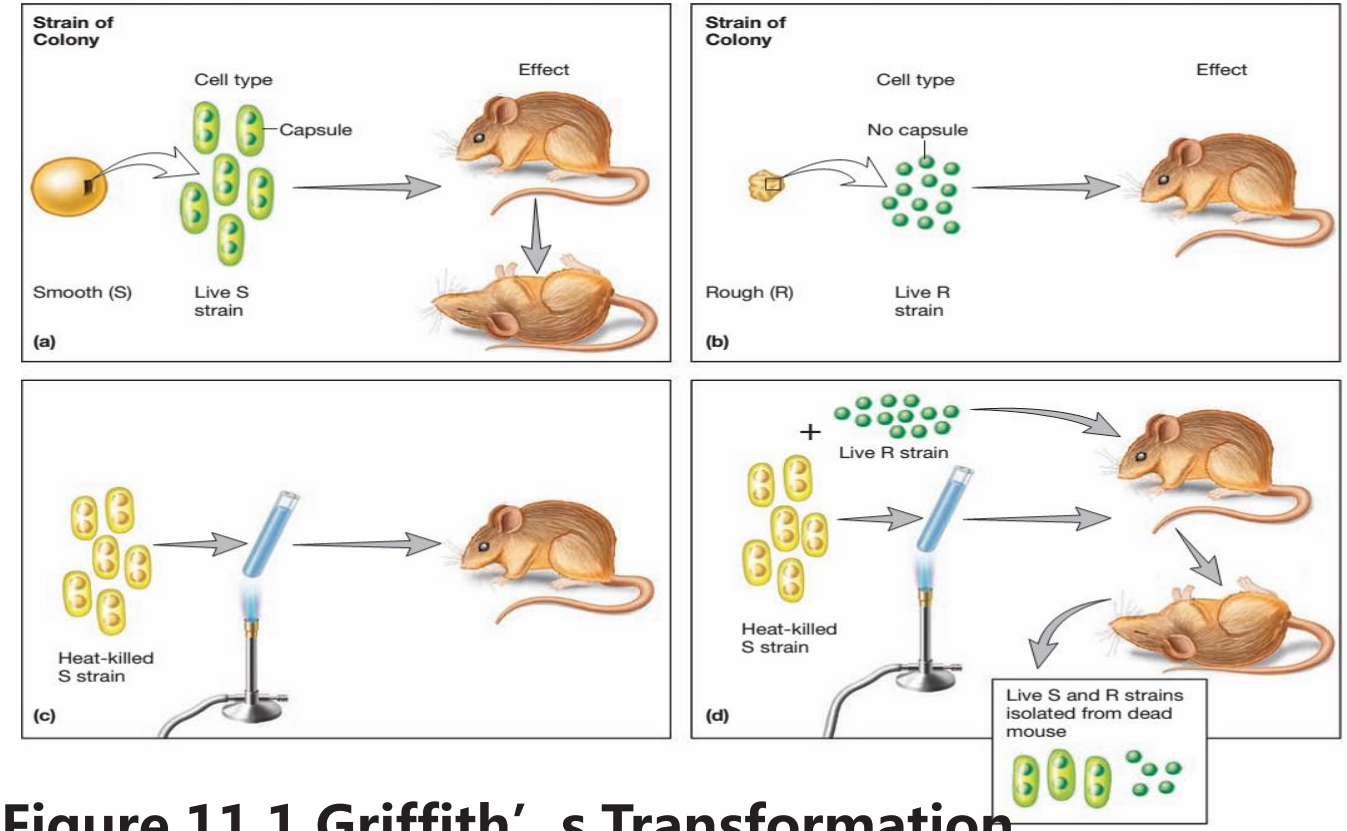


Figure 11.1 Griffith’s Transformation Experiments.

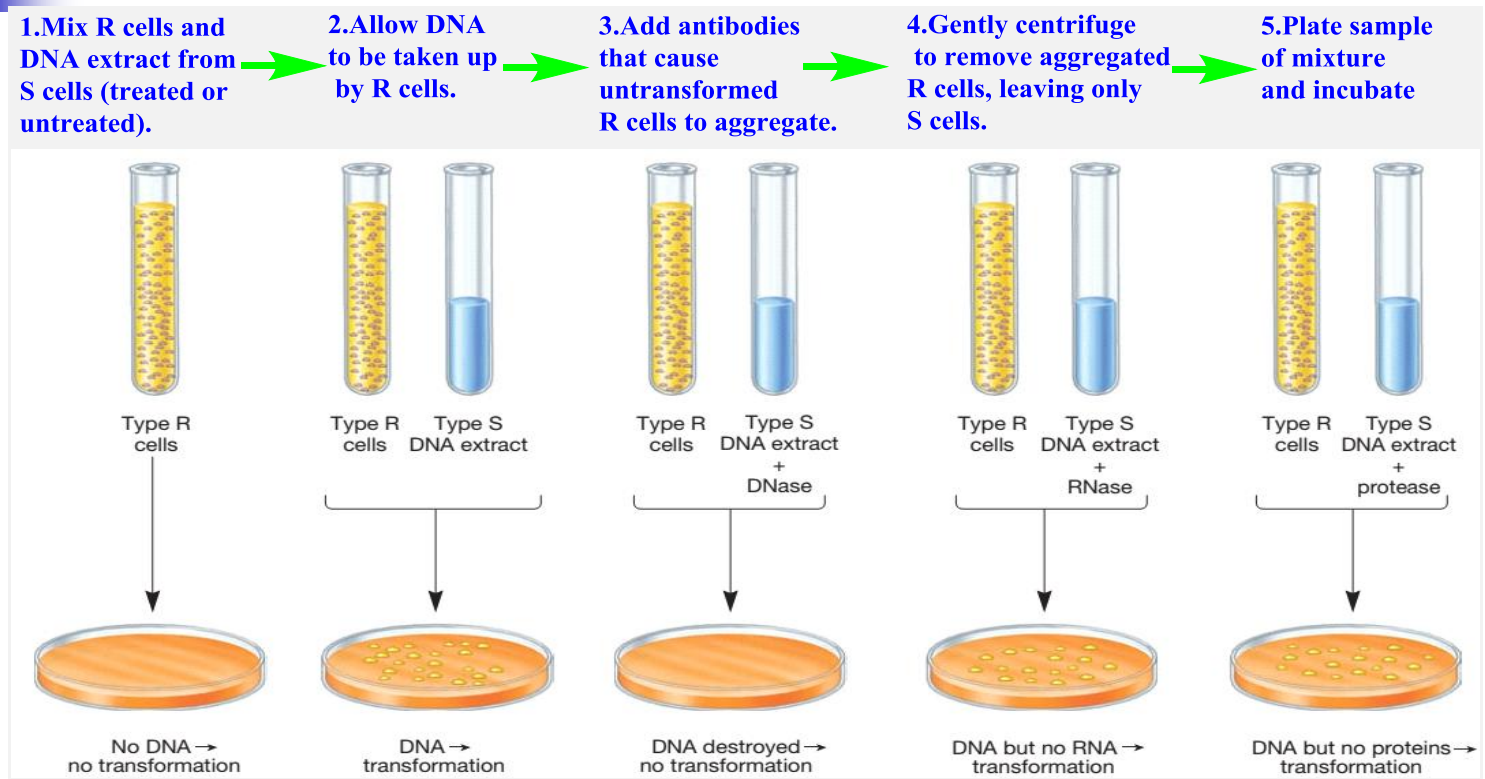
实验结果：

活的非致病性的R型肺炎型球菌能够从死的致病性的S型肺炎球菌中得到某种物质或成分，并转化为致病性的S型肺炎球菌。

问题：

这种物质是什么？通过什么途径进入cell体内？这种物质如何起作用？

Oswald Avery 设计实验回答了转化实验留下的问题.1944

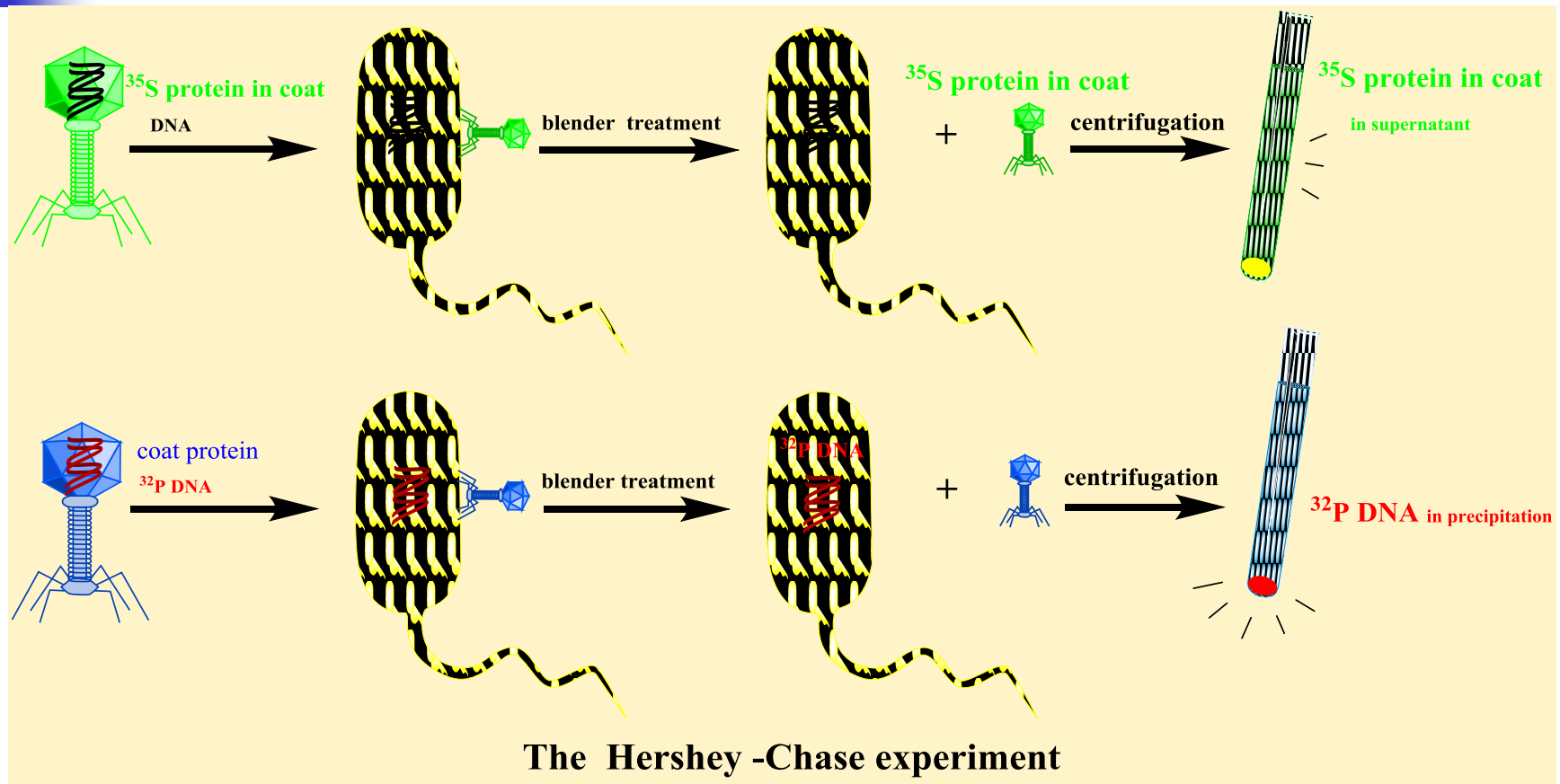


Some Experiments on the Transforming Principle.

Oswald Avery
(1877～1955)

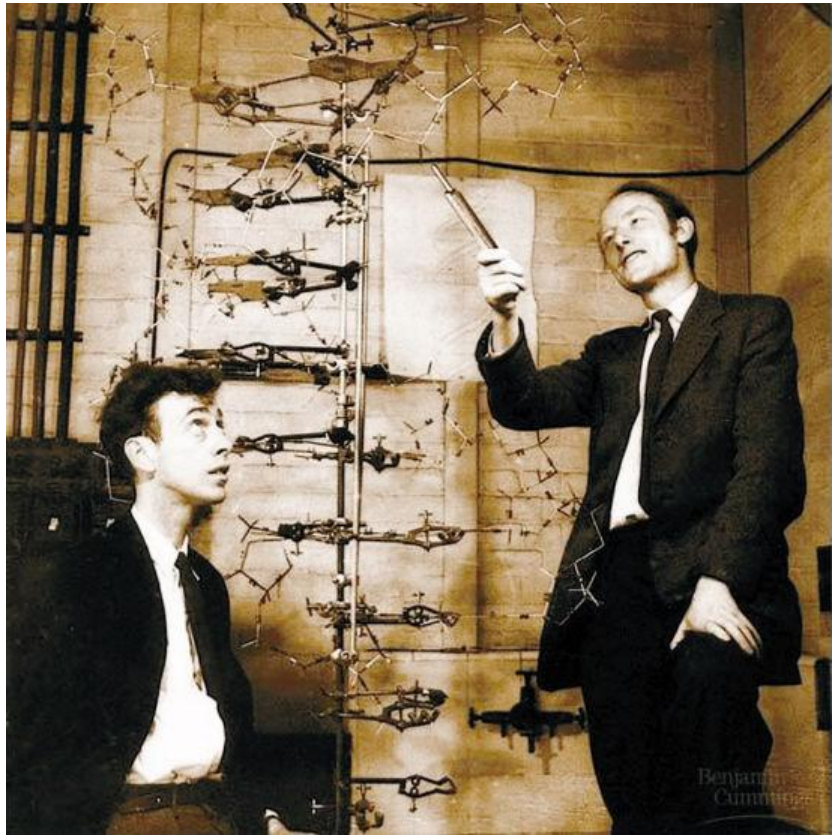
结论：DNA作为遗传物质，携带遗传信息，是导致S型转化为R型的关键因子。

2).DNA 作为遗传物质. 1952



结论：
噬菌体T2感染宿主菌时，只有其核酸进入细胞内，而蛋白质外壳留在细胞外。经此进入宿主细胞的病毒核酸能够复制出子代T2病毒，并且，与感染的病毒一样。证明DNA是遗传物质

Watson and Crick published the article(below) in NATURE in 1953



equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson
F. H. C. Crick

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

While the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury[1]) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1–3], semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxypentose nucleic acid ('structure *B*' in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the internucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[5] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the nth layer line being proportional to the square of J_n, the nth order Bessel function. A straight line may be drawn approximately through

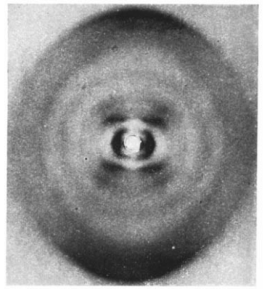


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. coli*. Fibre axis vertical

the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats n times along the helix there will be a meridional reflexion (J_n^2) on the nth layer line. The helical configuration produces side-bands on this fundamental frequency, the effect[5] being to reproduce the intensity distribution about the origin around the new origin, on the nth layer line, corresponding to C in Fig. 2.

We will now briefly analyse in physical terms some of the effects of the shape and size of the repeat unit or nucleotide on the diffraction pattern. First, if the nucleotide consists of a unit having circular symmetry about an axis parallel to the helix axis, the whole diffraction pattern is modified by the form factor of the nucleotide. Second, if the nucleotide consists of a series of points on a radius at right-angles to the helix axis, the phases of radiation scattered by the helices of different diameter passing through each point are the same. Summation of the corresponding Bessel functions gives reinforcement for the inner-

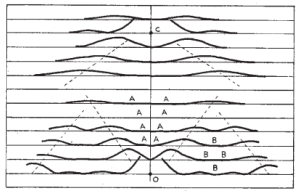
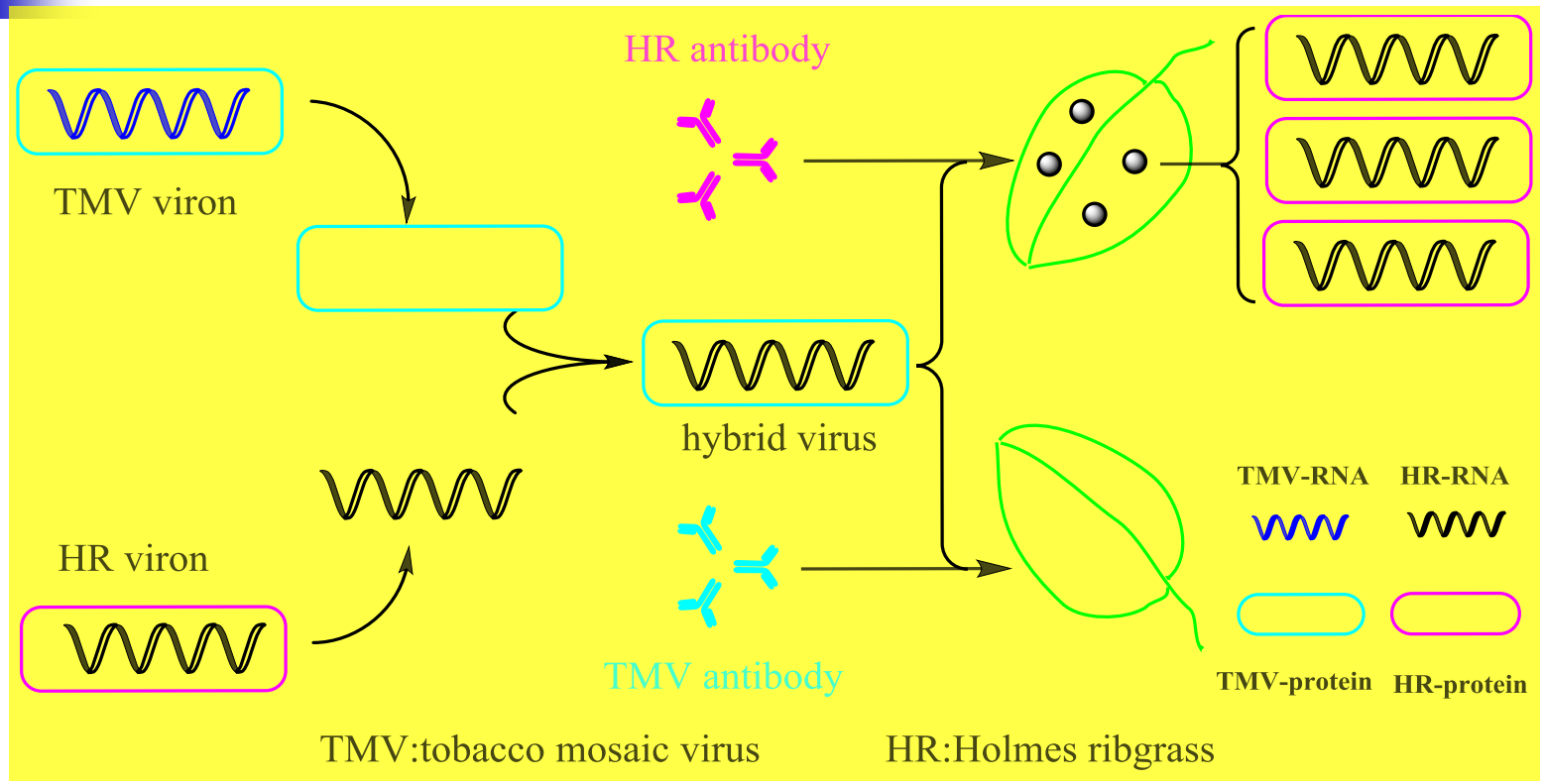


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide mass at 29 A. diameter and remainder distributed along a radius, the mass at a given radius being proportional to the radius. About C on the tenth layer line similar functions are plotted for an outer diameter of 12 A.

3).RNA作为遗传物质.1956



H.Franenkel-Conrat

H.Franenkel-Conrat设计的烟草花叶病毒（TMV）拆分重建实验

结论：RNA是遗传物质

Summary

三个经典实验的结果证明：
DNA或者RNA是遗传物质，是遗传信息的携带者。

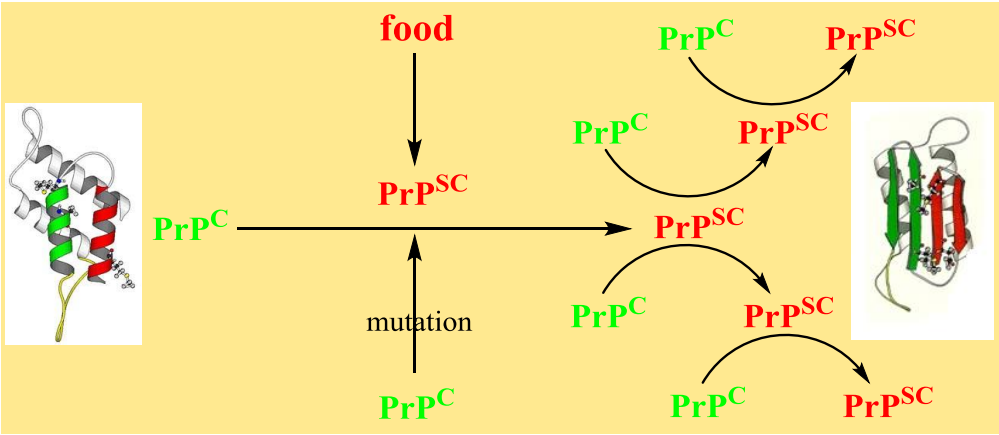
Questions:

1. Define genome, genotype, and phenotype.
2. Briefly summarize the experiments of Griffith; Avery, MacLeod, and McCarty; and Hershey and Chase. What did each show, and why were these experiments important to the development of microbial genetics?

朊病毒引发的思考

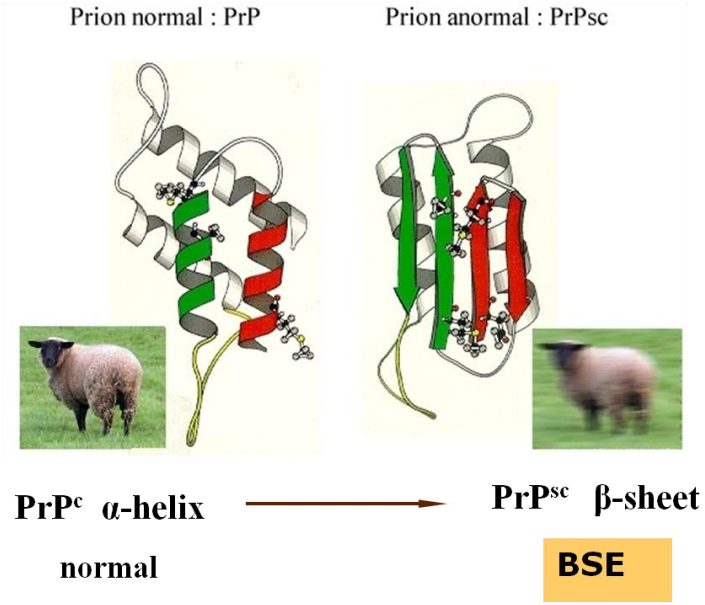
蛋白感染因子：Prion
(proteinaceous infectious particle)

PrP^{sc} (scrapie-associated prion protein)



正常PrP^c转化为PrP^{sc}的机制

Classic image of a cow with BSE. A feature of such disease is the inability of the infected animal to stand.

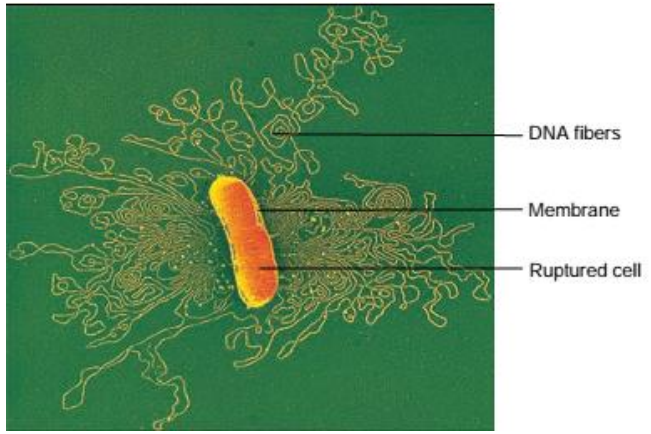


大量研究表明：朊病毒没有遗传物质。朊病毒是不是生命？如果不是，那么它属于什么？生命与非生命之间的界限存在吗？是“非此即彼”还是“亦此亦彼”？

1.微生物的基因、基因组及染色体结构

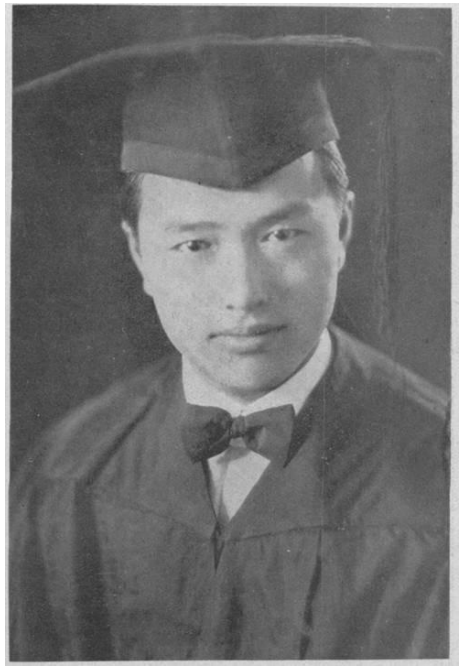
基因（gene）概念的演变

1.Mendel的遗传因子
2.染色体上的遗传功能单位
3.有遗传功能的DNA片段
4.完整的功能单位
5.不同功能的核苷酸序列



基因（gene）:是一个化学实体，是含有特定遗传信息的核苷酸序列（DNA or RNA）,是遗传物质的最小功能单位，彼此可有重叠，基因是可分的，也是可以移动的遗传因子，基因本身在结构和功能上存在着差别。

对基因本质的认识还在前进的路上…



著名遗传学家谈家桢
（1909--2008）

1.微生物的染色体结构

微生物基因的结构

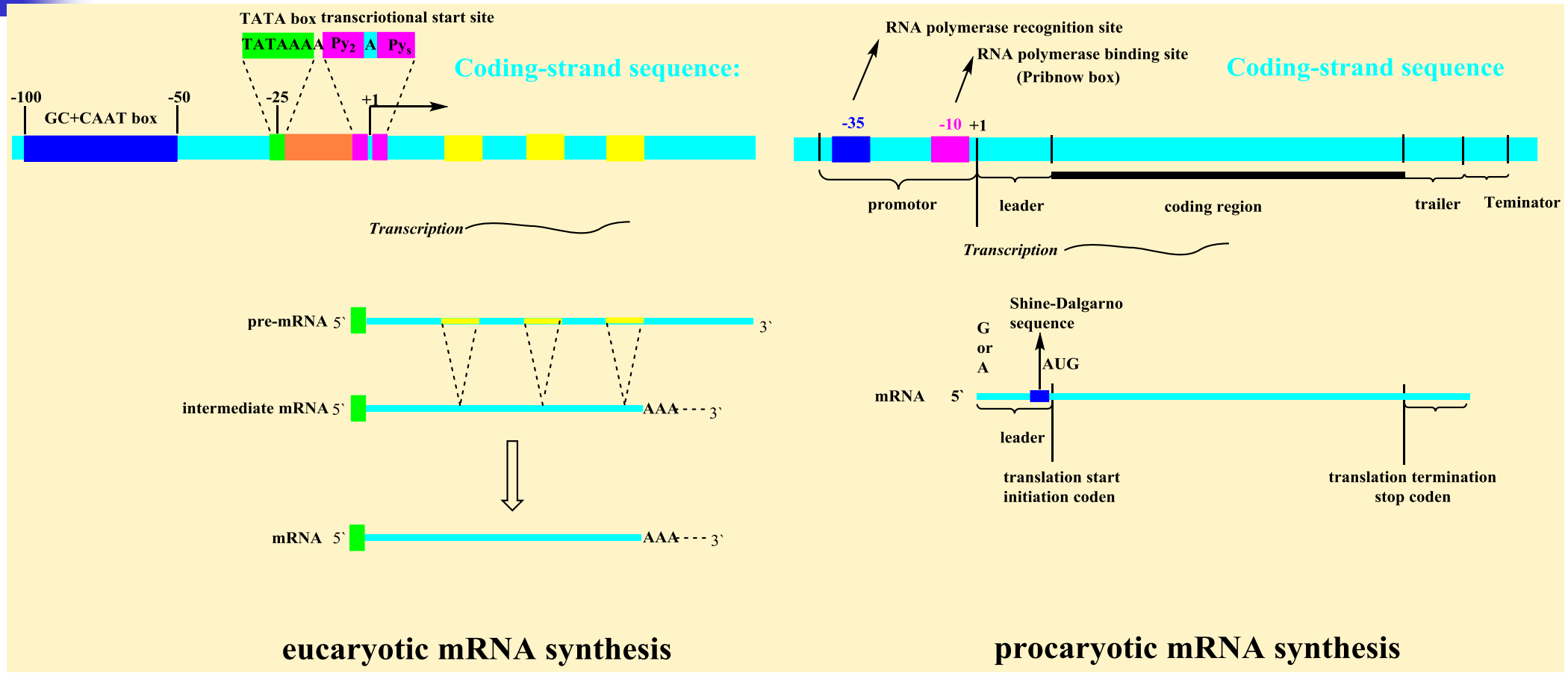
原核生物的基因：

（bacteria）、（archaea）的基因包括：蛋白质编码基因，非编码基因。后者主要包括tDNA, rDNA基因。基因内部不含内含子（intron）；DNA序列中存在着操纵子（operon）结构，即几个功能相关的基因串联排列，公用一个启动子（promotor）,同时转录（transcription）同时翻译（translation），同时关闭的遗传功能单位。

真核微生物的基因

酵母菌（yeast）,真菌（fungi）和原生动物（protist）基因有内含子(intron),少有操纵子结构，一个基因一个启动子。

真核mRNA和原核mRNA比较



微生物的基因组结构

基因组（genome）:一个微生物细胞或病毒中全部基因的总和。

MICROORGANISM	PROCARYOTES	EUCARYOYTES	VIRUS
GENOME	Chromatin，plasmids, haploid	Chromosome, haploid, rare diploid plasmid, chloroplast. mitochondria	Haploid
GENETIC MATEIAL	DNA	DNA	DNA or RNA
SIZE	<10M	>10M	<0.5M
COMFORMATION	circle or linear	linear including linear	linear or circle
HISTONE	histone-like protein	yes	no
GENE CLUSTER	yes	yes	no
SECONDARY METABOLISM GENE	yes	yes	no
TELOMERE	yes or no	yes	no

微生物的基因组结构

Table 15.2 Estimated Number of Genes Involved in Various Cell Functions[a]

Gene Function	*Escherichia coli K12*	*Bacillus subtilis*	*Mycoplasma genitalium*	*Treponema pallidum*	*Rickettsia prowazekii*	*Chlamydia trachomatis*	*Mycobacterium tuberculosis*	*Methanocaldococcus jannaschii*	*Pyrococcus abyssi*
Approximate total number of genes[b]	4,289	4,100	484	1,040	834	894	4,425	1,728	1,765
Cellular processes[c]	190	374	6	77	40	46	132	26	64
Cell envelope components	172	185	29	53	74	45	152	25	106
Transport and binding proteins	315	400	33	59	49	58	168	56	140
DNA metabolism	102	122	30	51	63	48	68	53	63
Transcription	41	114	13	25	26	18	40	21	37
Protein synthesis	122	161	90	99	104	133	110	118	108
Regulatory functions	176	293	5	22	26	12	165	19	66
Energy metabolism[d]	368	439	33	54	89	56	234	158	180
Central intermediary metabolism[e]	73	96	7	6	19	12	293	19	79
Amino acid biosynthesis	114	143	0	7	13	18	91	64	76
Fatty acid and phospholipid metabolism	67	84	8	11	22	27	158	9	18
Purines, pyrimidines, nucleosides, and nucleotides	77	81	19	21	19	14	57	37	51
Biosynthesis of cofactors and prosthetic groups	100	113	5	15	24	27	109	50	53

[a]Data adapted from TIGR (The Institute for Genomic Research) databases.
[b]The number of genes with known or hypothetical functions.
[c]Genes involved in cell division, chemotaxis and motility, detoxification, transformation, toxin production and resistance, pathogenesis, adaptations to atypical conditions, etc.
[d]Genes involved in amino acid and sugar catabolism, polysaccharide degradation and biosynthesis, electron transport and oxidative phosphorylation, fermentation, glycolysis/gluconeogenesis, pentose phosphate pathway, Entner-Doudoroff, pyruvate dehydrogenase, TCA cycle, photosynthesis, chemoautotrophy, etc.
[e]Amino sugars, phosphorus compounds, polyamine biosynthesis, sulfur metabolism, nitrogen fixation, nitrogen metabolism, etc.

原核、真核和古菌的基因组结构比较

Genome of *E.coli*	Genome of *Saccharomyces cerevisiae*	Gene of *Methanococcus jannaschii*
无细胞核，环形染色质，4-5.45M； DNA结合组蛋白样的蛋白质压缩成脚手架的致密结构； 没有端粒和组蛋白； 遗传信息的连续性； 大量的操纵子结构； 结构基因的单拷贝； 重复序列少而短； 有次级代谢产物。	**16**条线状染色体,**13.5M**； 有着丝粒和端粒； 线粒体中有环状染色质； 有质粒； 没有明显的操纵子结构； 基因内有内含子； 约**200**个拷贝的**rDNA**和**tDNA**基因； 大量的重复序列。	无核，环形染色体，**1.66M**； 有组蛋白； 有质粒； 有操纵子结构； 无内含子； 复制、转录和翻译类似真核生物； 有**40%**的基因与真核和细菌有同源性。

微生物与人类基因组计划

人类基因组计划 （Human Genome Project）

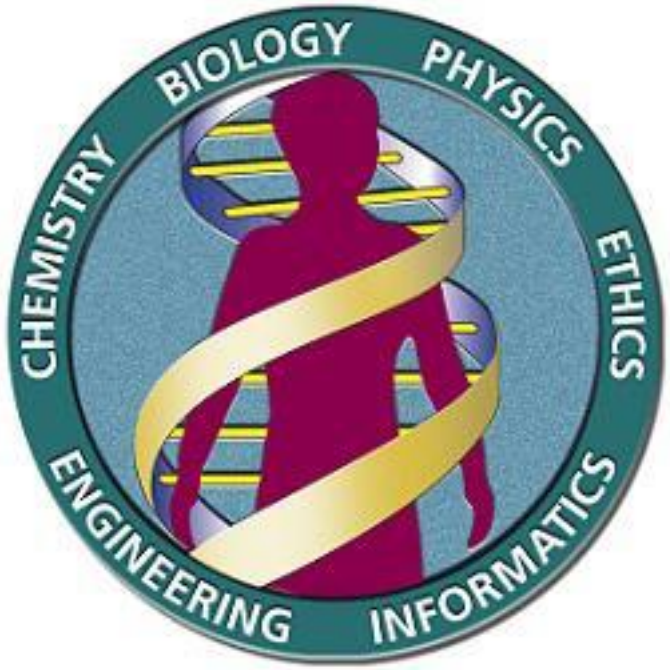
测定组成人类染色体（指单倍体）中所包含的30亿个碱基对组成的核苷酸序列，从而绘制人类基因组图谱，并且辨识其载有的基因及其序列，达到破译人类遗传信息的最终目的

发起国：美国
参与国：美、英、法、德、日、中

遗传图 序列图 物理图 转录图

1985年提出；
1990年正式开始实施；
2001年2月，测序工作完成；
2003年4月，草图的精细化图谱完成
2006年6月，最后染色体测序完成

大肠杆菌、酵母、线虫、果蝇 和 小鼠的基因组测序



后基因组时代（Postgenomics Era）

(functional genomics)

拓展阅读

Nature: 2015 Oct 1;526(7571):29-31

Twenty-five years of big biology

The Human Genome Project, which launched a quarter of a century ago this week, still holds lessons for the consortium-based science it ushered in, say **Eric D. Green**, **James D. Watson** and **Francis S. Collins**.

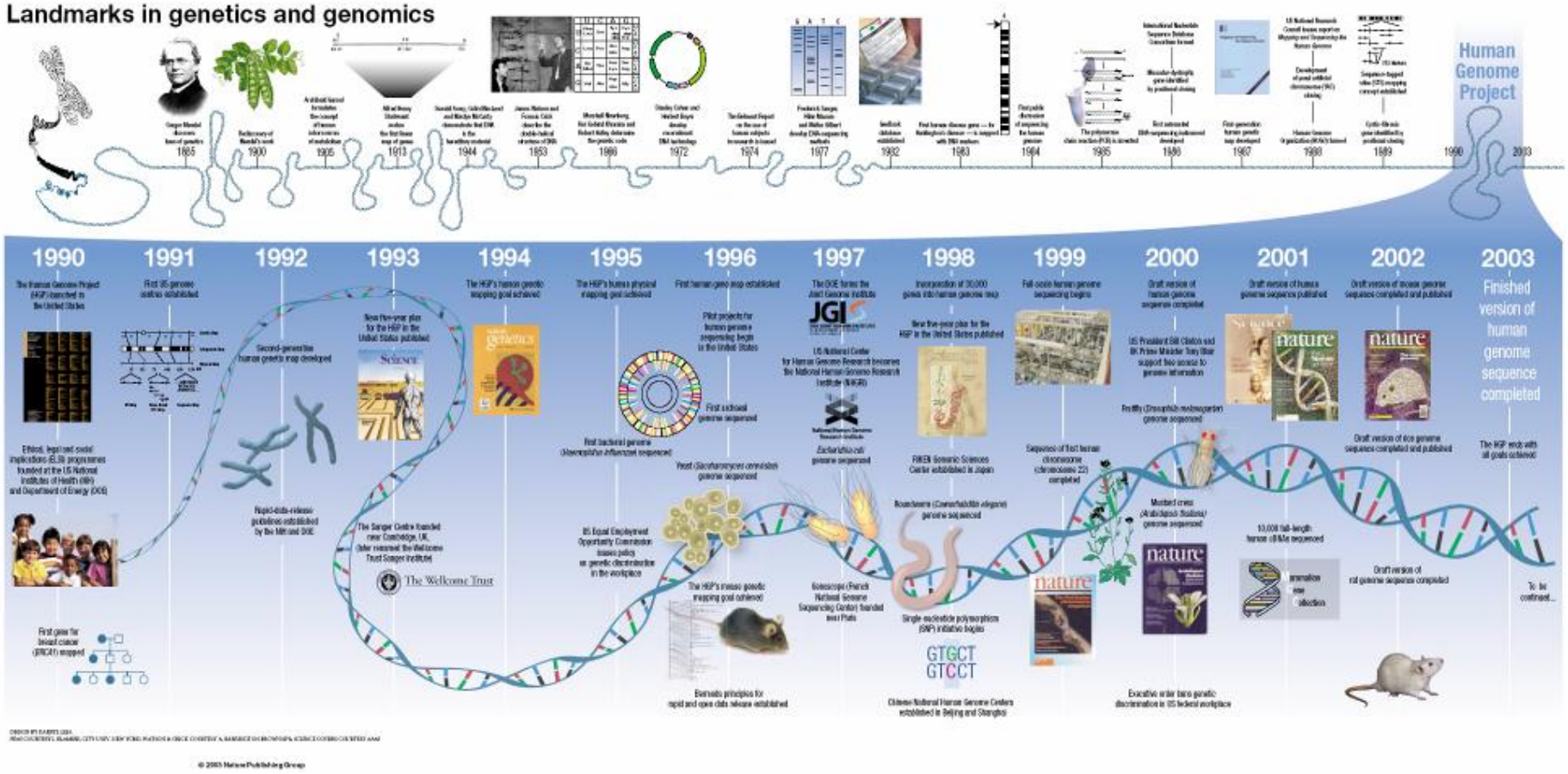
TOP: BOTTOM RIGHT: COLD SPRING HARBOR LAB. LIBRARY & ARCHIVES; BOTTOM LEFT: ERIC GREEN

1989, Pre-HGP Banbury meeting. Francis Collins & James Watson in top row.

1990, Washington University School of medicine. Maynard Olsen & Eric Green.

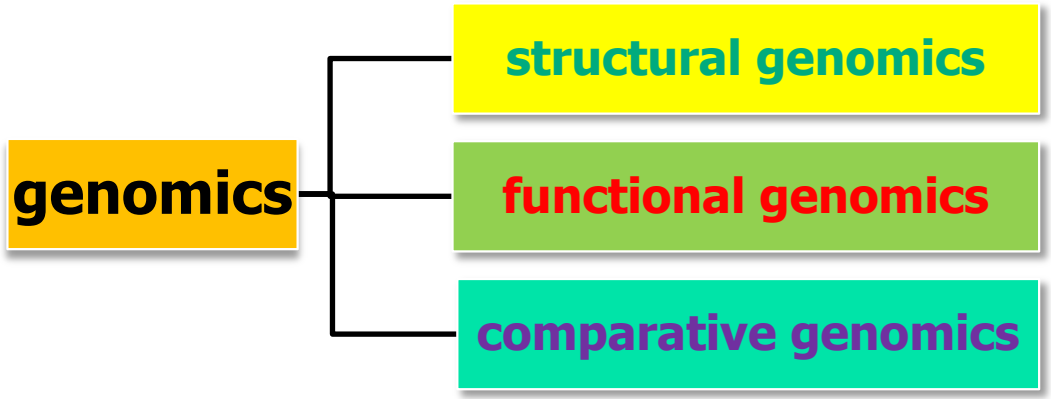
1997, HGP meeting at Cold Spring Harbor. EGreen, R Myers, J Witkowski & R Gibbs.

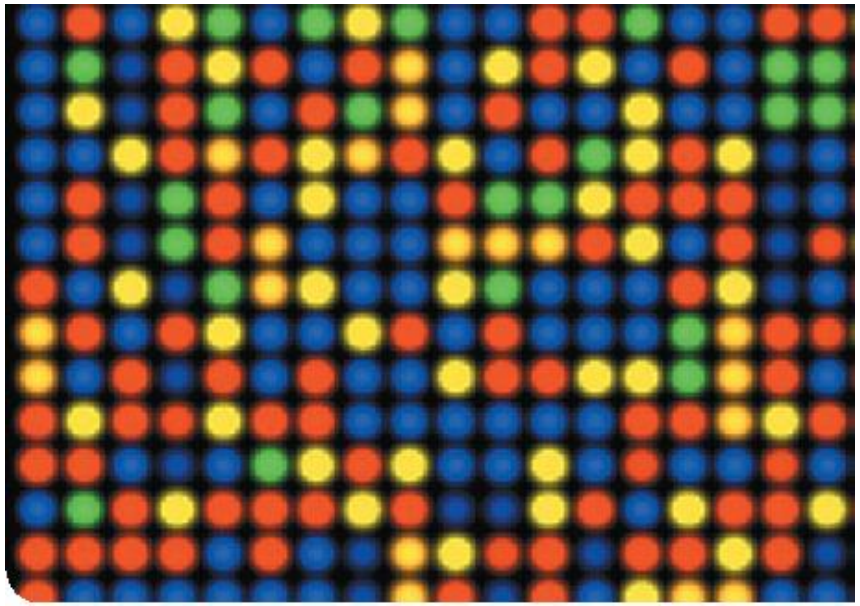
微生物的基因组学


Landmarks in genetics and genomics
Human Genome Project
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
Finished version of human genome sequence completed
JGI
nature
genetics
Science
GTGCT
GTCCT

微生物基因组学（microbial genomics）

研究基因组的分子组织，遗传信息和编码产物。





Microarray picture

A prerequisite to understanding the complete biology of an organism is the determination of its entire genome sequence.

—J. Craig Venter, et al.

基因组测序技术

1995, J. Craig Venter, Hamilton Smith和同事开发了全基因鸟枪法测序法和所需的计算机软件，开启了生命科学的基因组学时代。95-05年十年间，发表了249个全基因组，目前达到5635个基因组。

了解：中国的华大基因公司

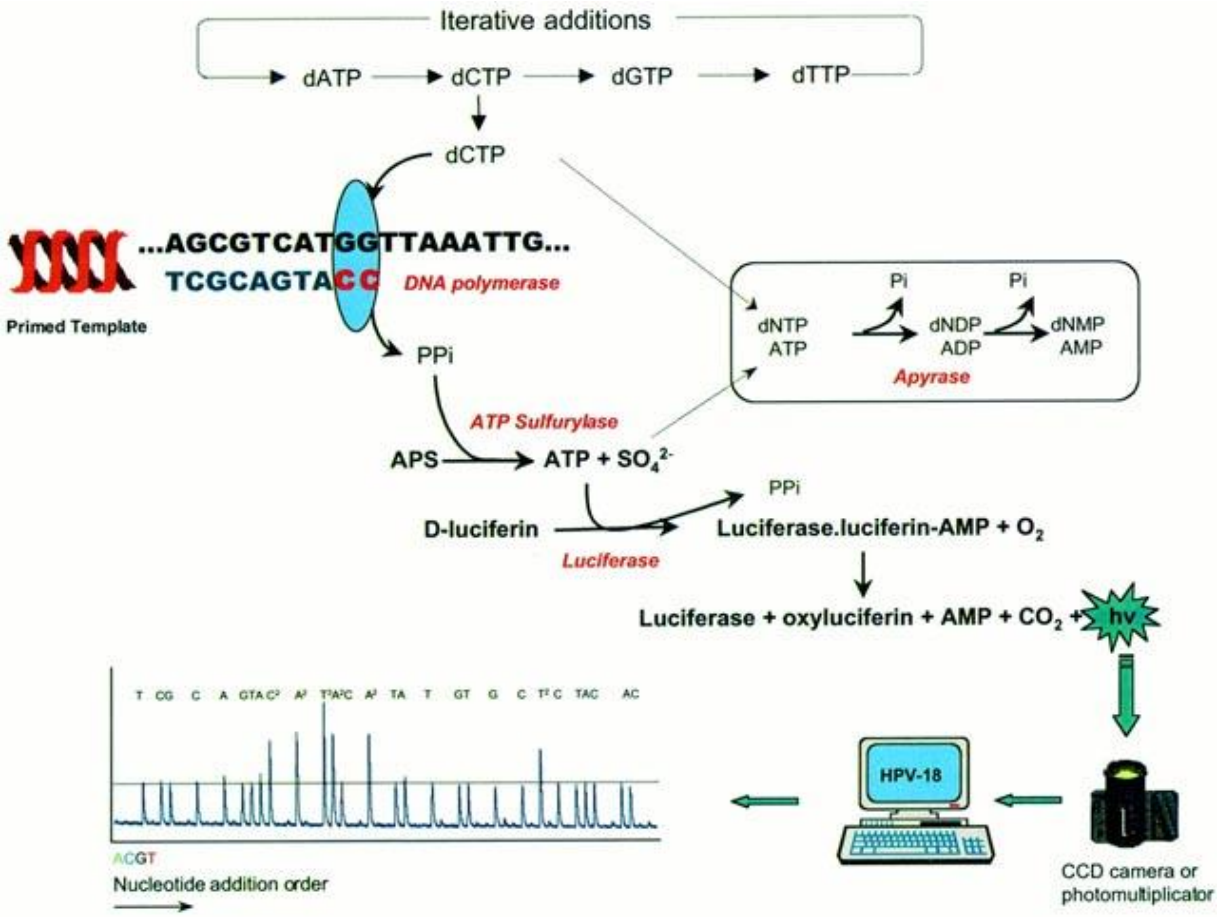
Table 15.1 Examples of Complete Published Microbial Genomes

Genome	Domain[a]	Number of Strains Sequenced	Size (Mb)	% G + C
Agrobacterium tumefaciens	B	2	4.92	60
Aquifex aeolicus	B	1	1.55	43
Archaeoglobus fulgidus	A	1	2.18	48
Bacillus anthracis	B	4	5.09–5.23	36
Bacillus subtilis	B	1	4.21	43
Borrelia burgdorferi	B	1	1.44	28
Campylobacter jejuni	B	1	1.64	31
Caulobacter crescentus	B	1	4.02	62–67
Chlamydia pneumoniae	B	2	1.23	40
Chlamydia trachomatis	B	2	1.05–1.07	41
Chlorobium tepidum	B	1	2.15	57
Clostridium perfringens	B	1	3.03	29
Corynebacterium glutamicum	B	1	3.31	55–58
Deinococcus radiodurans	B	1	3.06	67
Escherichia coli	B	6	4–5.45	50
Geobacter sulfurreducens PCA	B	1	3.81	61
Haemophilus influenzae Rd	B	1	1.83	39
Halobacterium sp. NRC-1	A	1	2.01	68
Helicobacter pylori	B	2	1.64–1.67	39
Listeria monocytogenes	B	2	2.9	37–39
Methanobacterium thermoautotrophicum	A	1	1.75	49
Methanocaldococcus jannaschii	A	1	1.66	31
Mycobacterium leprae	B	1	3.27	58
Mycobacterium tuberculosis	B	2	4.40	65
Mycoplasma genitalium	B	1	0.58	31
Mycoplasma pneumoniae	B	1	0.82	40
Nanobacterium equitans	A	1	0.49	32
Neisseria meningitidis	B	3	2.18–2.27	51
Prochlorococcus marinus	B	3	1.66–2.41	31–51
Pseudomonas aeruginosa	B	1	6.26	67
Pyrococcus abyssi	A	1	1.77	44
Pyrococcus horiksohii	A	1	1.74	42
Rhodopseudomonas palustris	B	1	5.46	65
Rickettsia prowazekii	B	1	1.11	29
Saccharomyces cerevisiae	E	1	12.14	38
Salmonella enterica serovar Typhimurium	B	1	4.86	50–53
Staphylococcus aureus	B	7	2.80–2.90	33
Streptococcus mutans	B	1	2.03	37
Streptococcus pneumoniae	B	2	2.16	40
Streptococcus pyogenes	B	6	1.84–1.90	39
Streptomyces coelicolor	B	1	8.67	72
Sulfolobus tokodaii	A	1	2.69	33
Synechocystis sp.	B	1	3.57	47
Thermoplasma acidophilum	A	1	1.56	46
Thermotoga maritima	B	1	1.86	46
Treponema pallidum	B	1	1.14	52
Vibrio cholerae	B	1	4.03	48
Yersinia pestis	B	3	4.60–4.65	48
Yersinia pseudotuberculosis	B	1	4.74	48

[a]The following abbreviations are used: A, *Archaea*; B, *Bacteria*; E, *Eucarya*.

454焦磷酸测序(pyrosequencing)技术-2005年

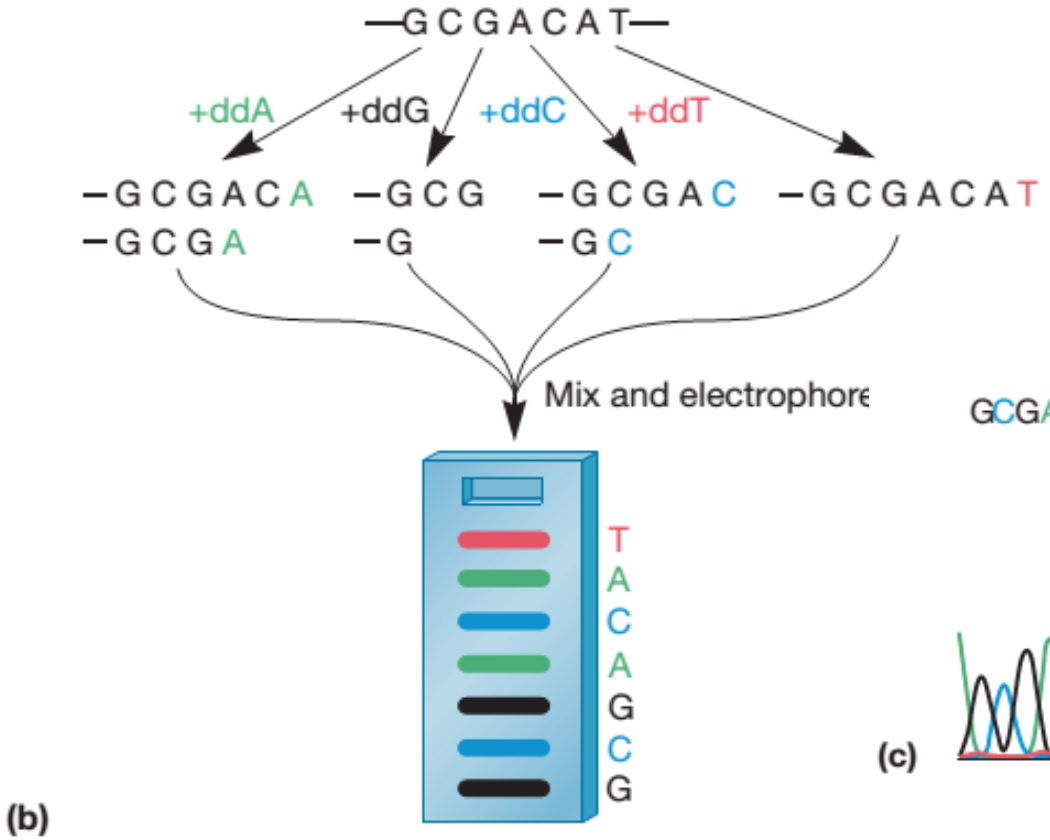
焦磷酸测序技术的原理是：引物与模板DNA退火后，在DNA聚合酶(DNA polymerase)、ATP硫酸化酶(ATP sulfurytase).荧光素酶(luciferase)和三磷酸腺苷双磷酸酶(Apyrase)4种酶的协同作用下，将引物上每一个dNTP的聚合与一次荧光信号的释放偶联起来，通过检测荧光的释放和强度，达到实时测定DNA序列的目的。焦磷酸测序技术的反应体系由反应底物、待测单链、测序引物和4种酶构成。反应底物为5’-磷酰硫酸(adenosine- 5’-phosphosulfat，APS)、荧光素(luciferin)。



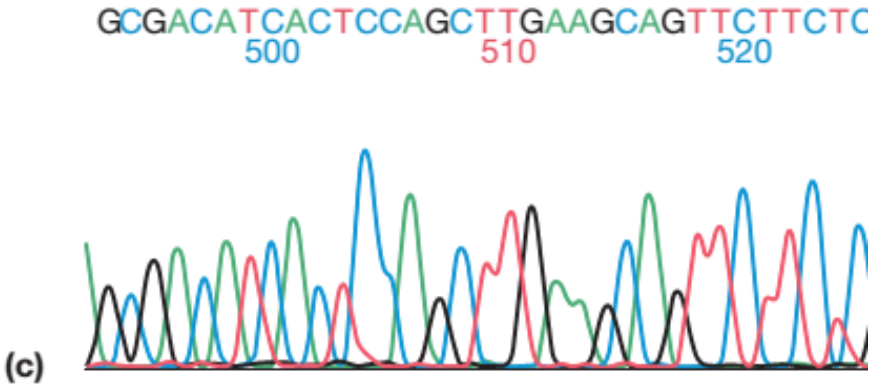
Early days: a DNA-sequencing lab in 1994.

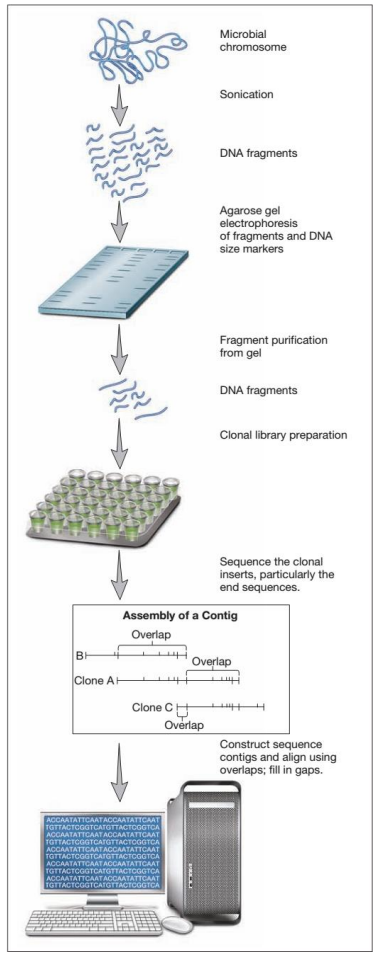
By 2006, DNA sequencing required much less manpower.

Structural genomics is the study of the physical nature of genomes. Its primary goal is to determine and analyze the DNA sequence of the genome



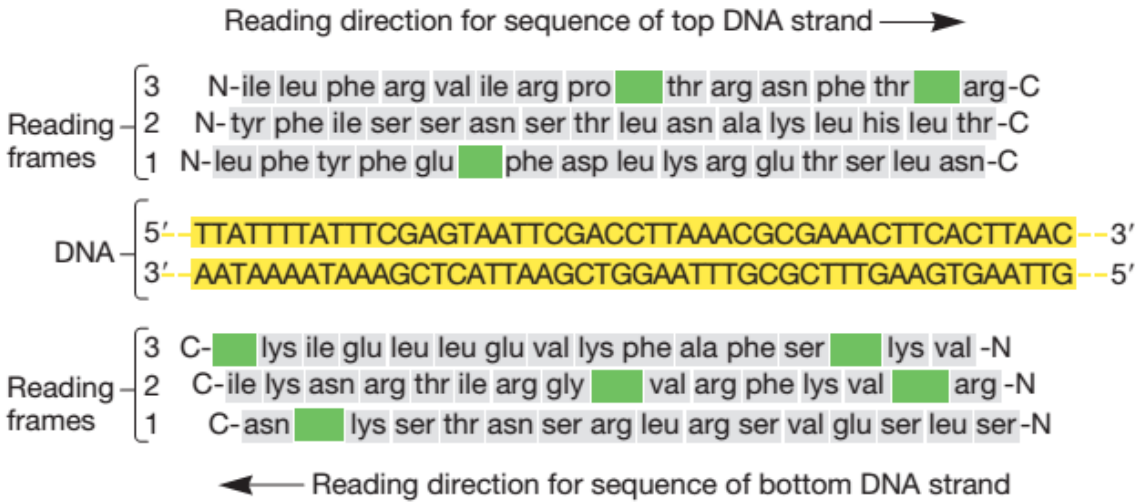
1. *Library construction.*
2. *Random sequencing.*
3. *Fragment alignment and gap closure.*
4. *Editing.*





Functional genomics is concerned with the way in which the genome functions. That is, it examines the transcripts produced by the genome and the array of proteins they encode.

The base-by-base comparison of two or more gene sequences is called alignment.
the nucleotide sequences are so alike that they most probably arose through gene duplication; such genes are called paralogs.
Alignments of genes found in two or more different organisms may reveal that they are so strikingly similar that they are predicted to have the same function; these genes are called orthologs.



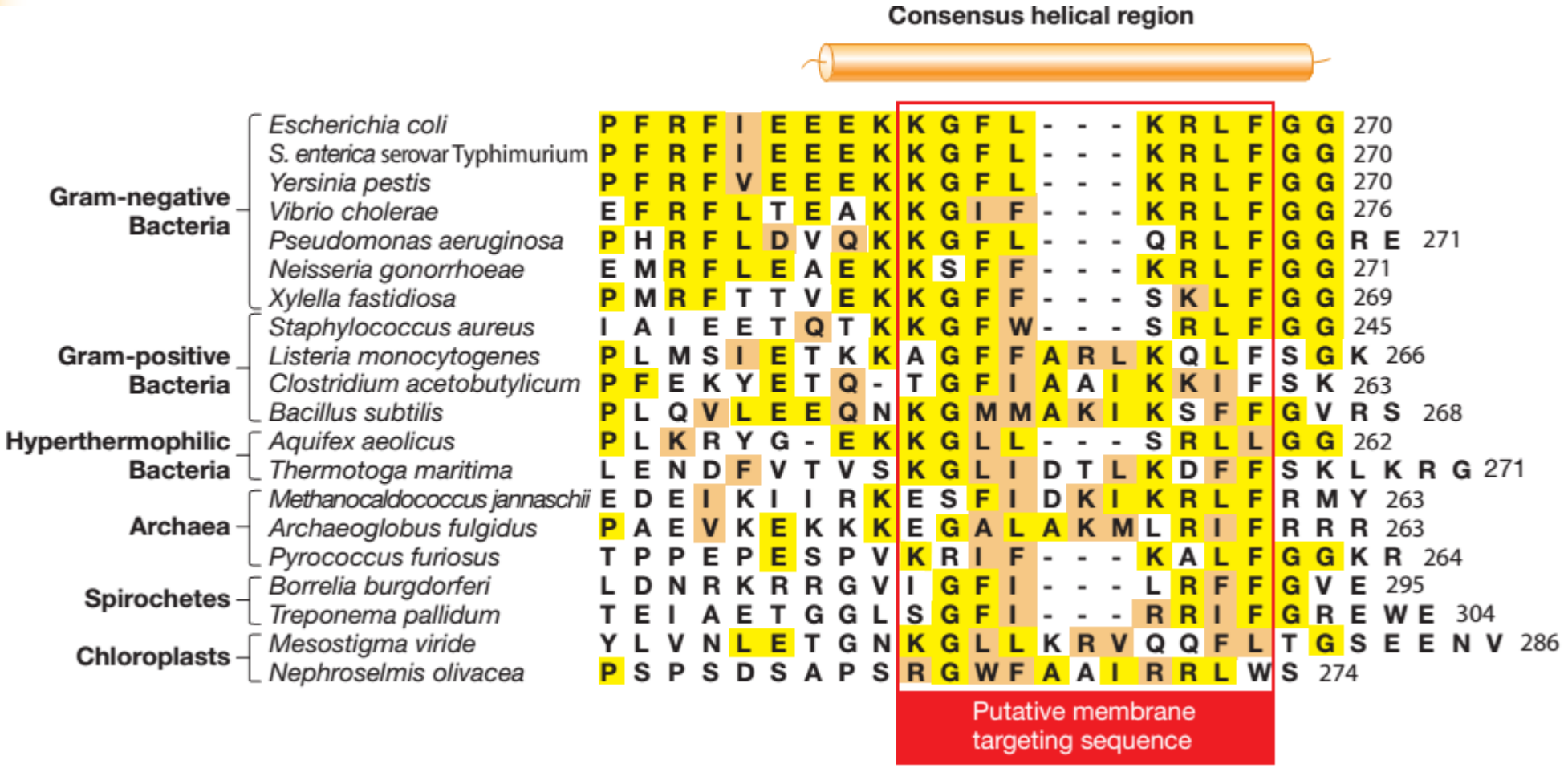


Figure 15.5 Analysis of Conserved Regions of Phylogenetically Well-Conserved Proteins.

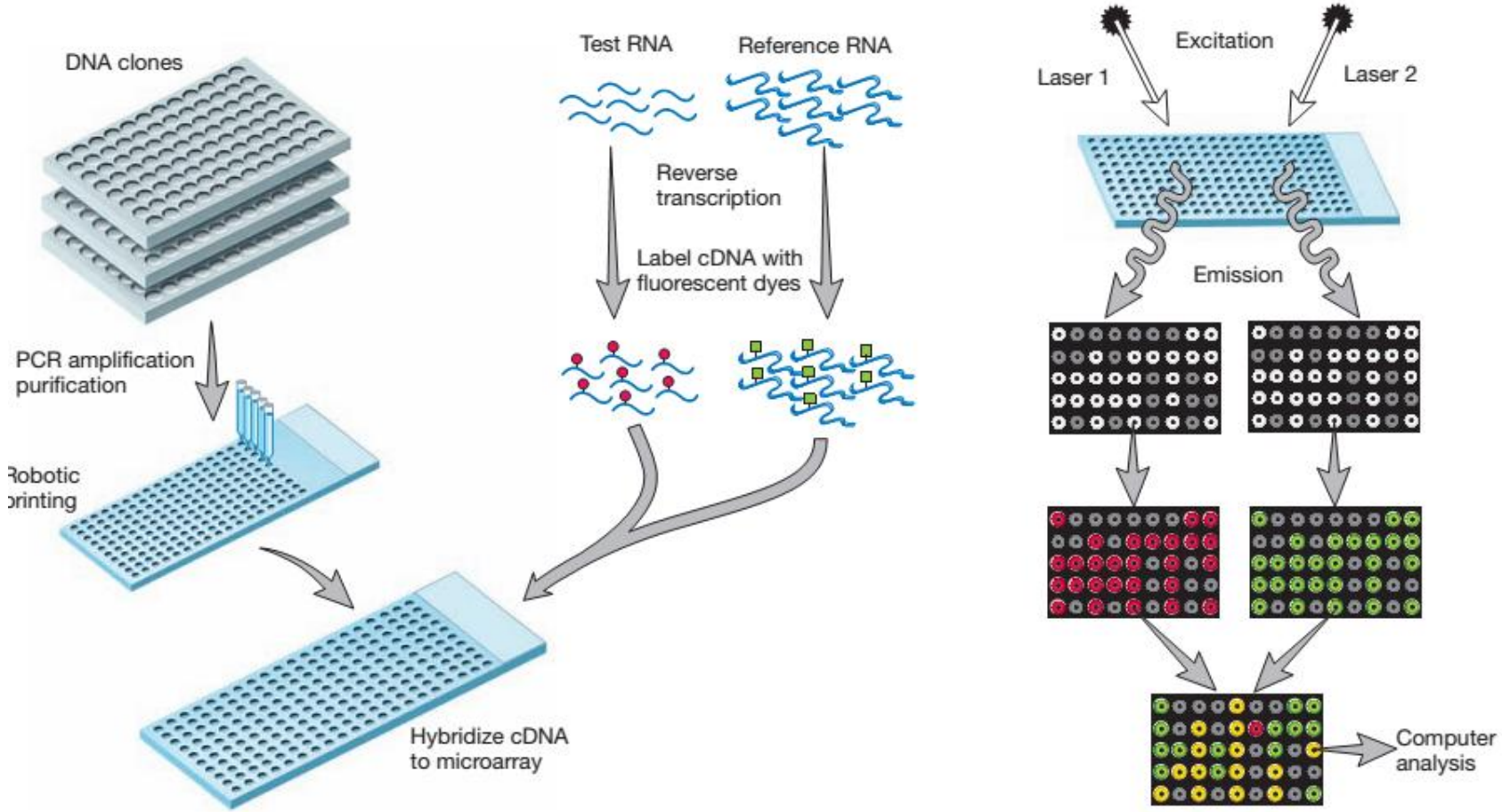


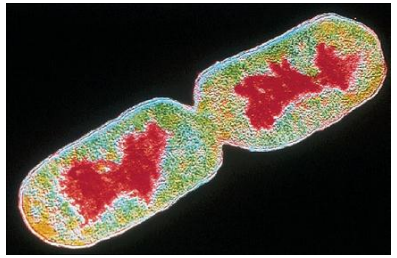
Figure 15.9 A Microarray System for Monitoring Gene Expression.

comparative genomics, in which genomes from different organisms are compared to look for significant differences and similarities.

microbial genomes are not as static as once thought.

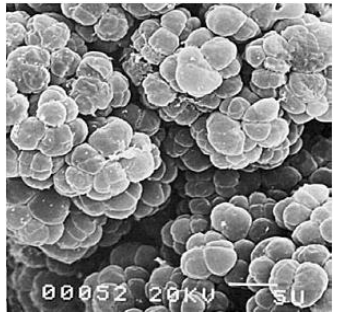
HGT is a major evolutionary force in short-term microbial evolution and long-term.

Genome analysis has revealed that HGT is frequently mediated by phages, and that lysogeny may be the rule, rather than the exception.



Methanosarcina mazei **appears to have acquired about one-third of its genes from other procaryotes.**

E. coli **may have acquired the lactose (lac) operon from another microbe ,and became capable of colonizing the mammalian colon, where milk sugar is a common carbon source**



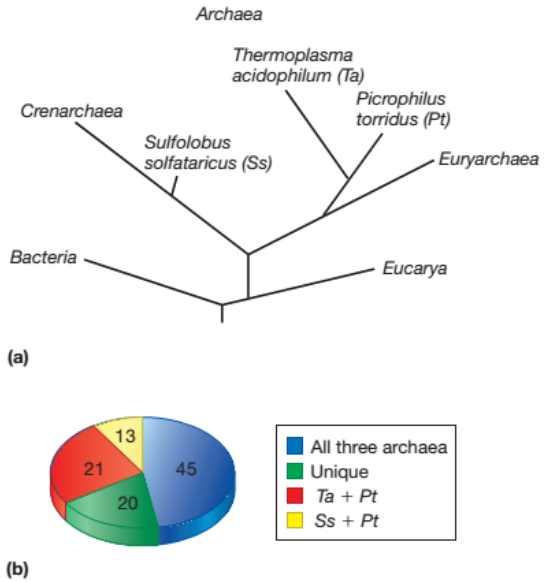
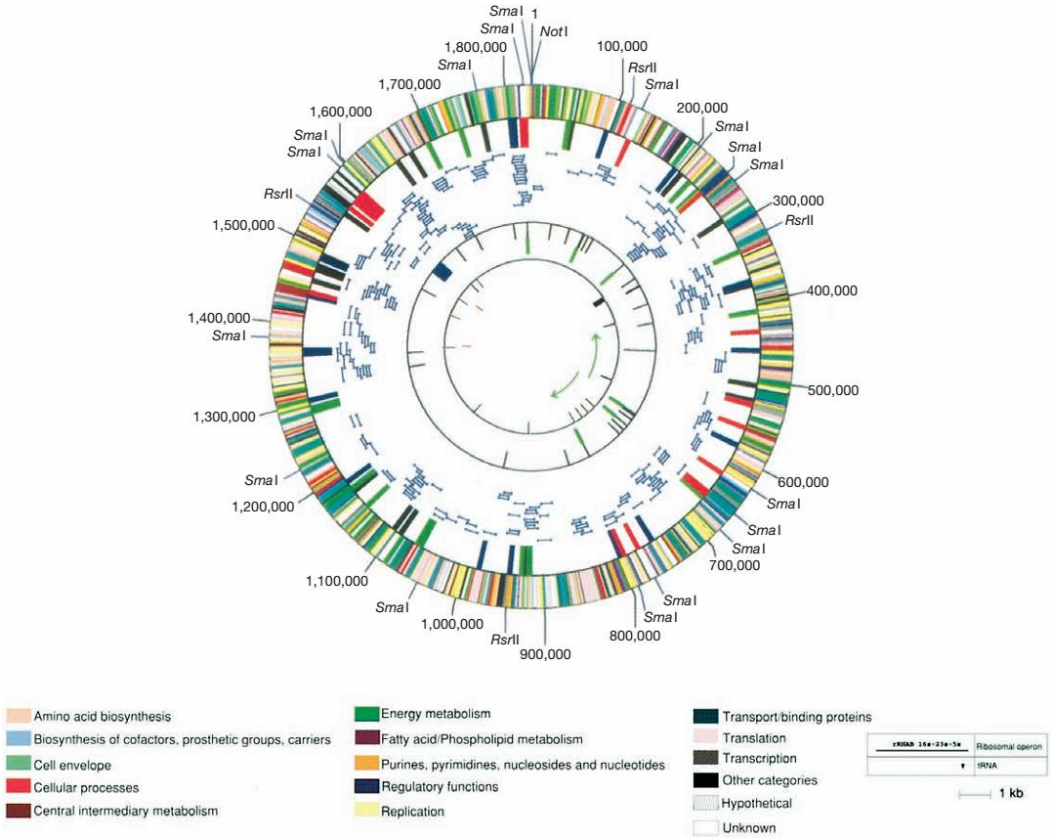


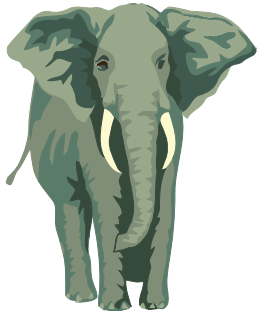
Figure 15.10 Comparative Genomics of *Picrophilus*

基因组学的发展，为我们提供了微生物基因组结构、生理学和系统发育及病原微生物致病、微生物进化和生态等方面的大量信息，为基础研究，应用研究如开发疫苗、药物、生态治理等提供理论依据！

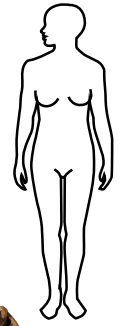


微生物宏基因组学

宏基因组（Megagenomic）:指特定微生态区域或某个动、植物生物个体所含有的全部微生物的基因组的总和。



人体微生物组(microbiome)则构成一个宏基因组。人体中微生物细胞的数量约是人体细胞的10倍，含有非冗余的ORF为4.1百万，是人基因2.2万个的186倍，被誉为人体的第二套基因组。正成为迅速发展的研究领域。



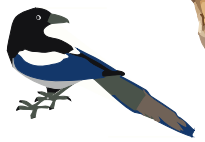
宏基因组学（Megagenomics）：对特定环境的微生物组(microbiome)的宏基因组进行克隆，并通过构建宏基因组文库和筛选等手段获得新的生理活性物质；或者根据rDNA数据库设计引物，通过系统学分析获得该环境中微生物的遗传多样性和分子生态学信息。











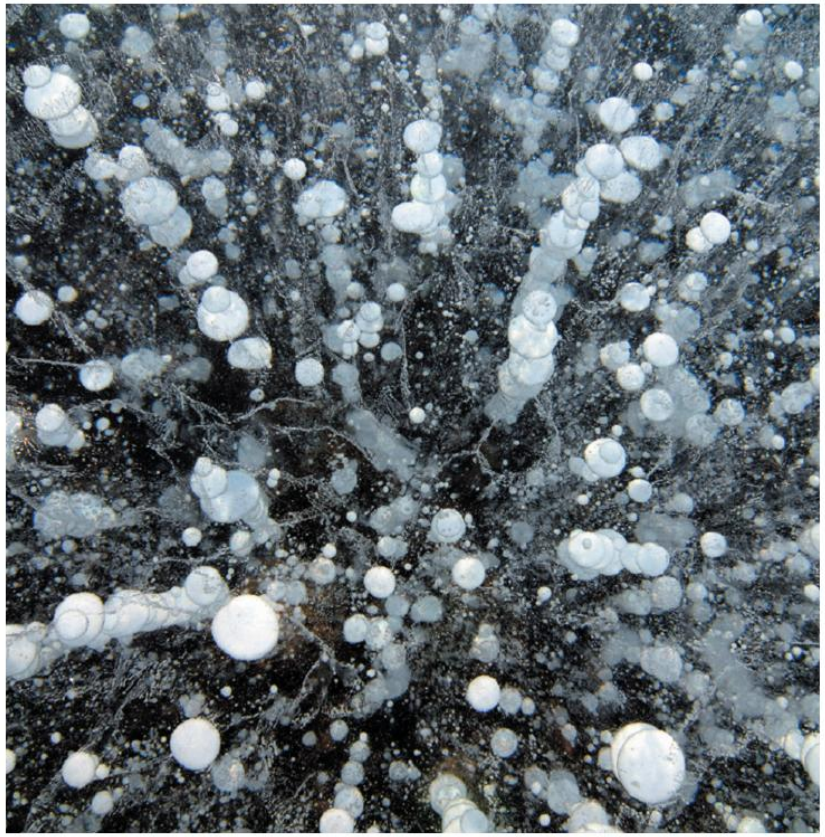

IMI: International Microbiome Institution

“完整地认识地球微生物群落（微生物组）在生物圈和人类健康中起到的作用，是解决21世纪人类社会从能源、传染病到农业等领域面临的许多难题的关键。”

Create a global microbiome effort

Understanding how microbes affect health and the biosphere requires an international initiative, argue **Nicole Dubilier**, **Margaret McFall-Ngai** and **Liping Zhao**.

CLIFF LEIGHT/GETTY



Even extreme environments such as Antarctic ice lakes host microbes.

ARTICLE

Received 21 Feb 2013 | Accepted 13 Jun 2013 | Published 23 Jul 2013

DOI: 10.1038/ncomms3151

Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota

Yongfei Hu[1,*], Xi Yang[1,*], Junjie Qin[2], Na Lu[1], Gong Cheng[1], Na Wu[1], Yuanlong Pan[1], Jing Li[1], Liying Zhu[3], Xin Wang[3], Zhiqi Meng[3], Fangqing Zhao[4], Di Liu[1], Juncai Ma[1], Nan Qin[5], Chunsheng Xiang[5], Yonghong Xiao[5], Lanjuan Li[5], Huanming Yang[2], Jian Wang[2], Ruifu Yang[6], George F. Gao[1,7], Jun Wang[2] & Baoli Zhu[1]

BRIEFINGS IN BIOINFORMATICS. VOL 13. NO 6. 728–742
Advance Access published on 9 September 2012

doi:10.1093/bib/bbs039

Current opportunities and challenges in microbial metagenome analysis—a bioinformatic perspective

Hanno Teeling and Frank Oliver Glöckner

Submitted: 30th March 2012; Received (in revised form): 9th June 2012

Review

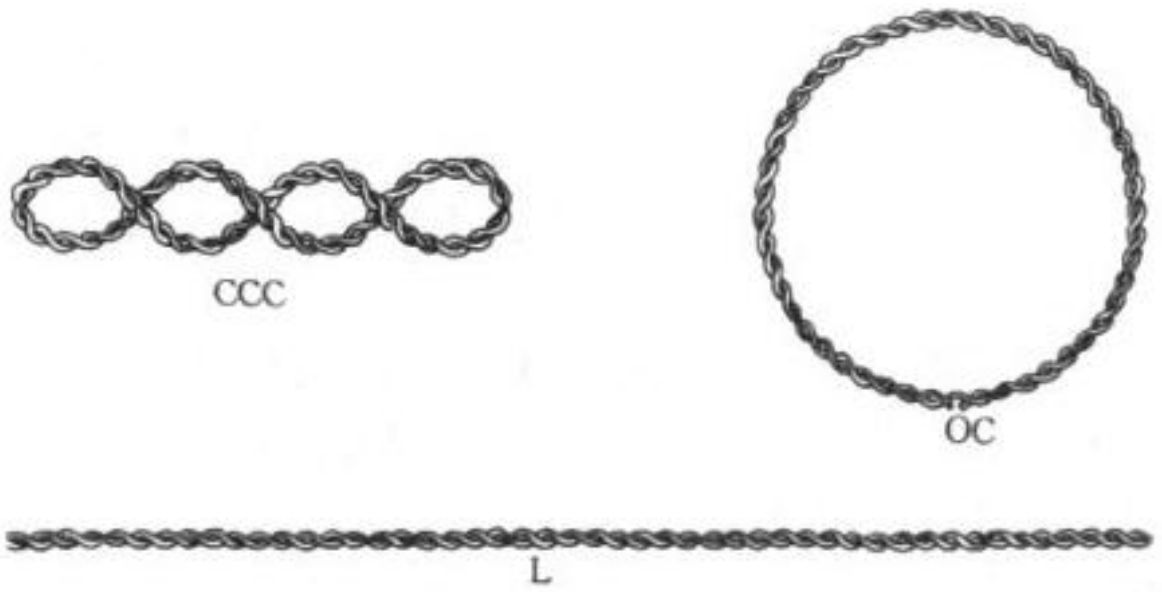
The gut microbiota, obesity and insulin resistance

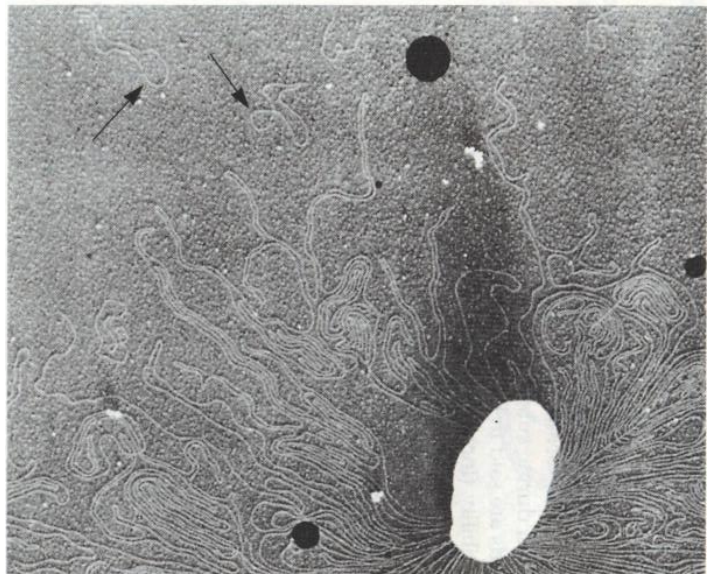
Jian Shen [a], Martin S. Obin [c], Liping Zhao [a,b,*]

[a] *Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Dongchuan Road 800, Shanghai 200240, China*
[b] *School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Dongchuan Road 800, Shanghai 200240, China*
[c] *Obesity and Metabolism Laboratory, JMUSDA-Human Nutrition Research Center on Aging at Tufts University, Boston, MA 02111, USA*

2.微生物的质粒

微生物细胞（主要指原核细胞，也包括一些酵母和真菌）中独立存在于染色体外的、小的双链的**DNA**分子，称为质粒**(plasmid)**。





质粒的性质和遗传学意义

项　　目	性　　质
大小(size)	1kb-1000kb，一般少于30基因
类型(type)	绝大多数为双链DNA, RNA质粒较为少见
构型(configuration)	共价闭合环CCC型,开环型OC，线型L
复制(replication)	独立于染色体进行自主复制
拷贝数(copy number)	单拷贝质粒，低拷贝质粒和高拷贝质粒
遗传学功能(genetics function)	能赋予宿主一些新性状，或增强其适应性，或提升竞争优势，丢失亦不影响宿主的正常生命活动。

遗传学意义

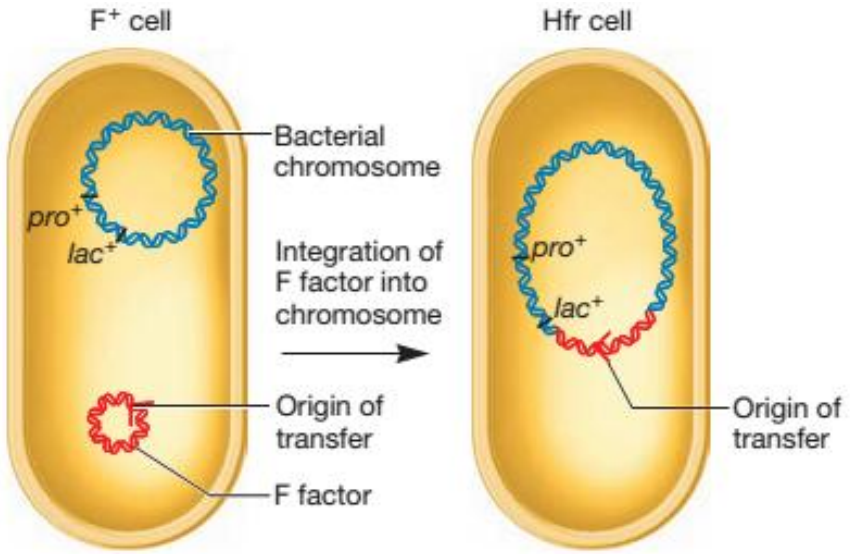
Plasmids play many important roles in the lives of the organisms that have them.

They also have proved invaluable to microbiologists and molecular geneticists in constructing and transferring new genetic combinations and in cloning genes.

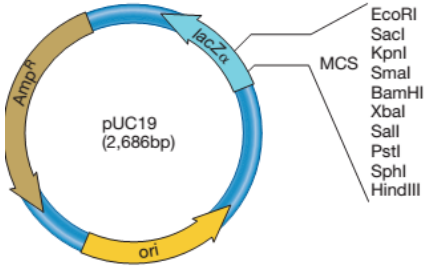
几个概念

附加体(episomes): 能整合在宿主的染色体上，并随着基因组的复制而进行复制的质粒。例如介导结合作用的F因子。

质粒的消除(curing): 质粒的丢失。可以人为，亦可自然丢失。通常可用丫啶橙、UV、离子辐射、胸腺嘧啶饥饿等方法达到目的



质粒的不亲和性(incompatibility):指不能共存于一个宿主菌细胞的不同的质粒。根据此便能将质粒分为许多不亲和群。能在同一个细胞共存的质粒属于不同的不亲和群。



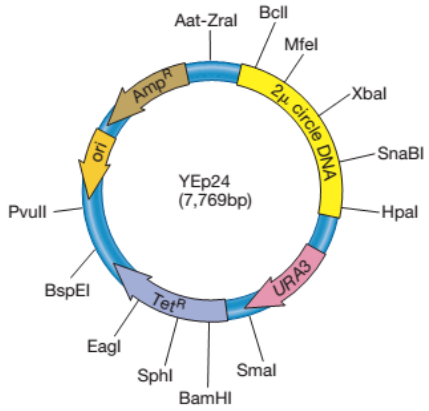
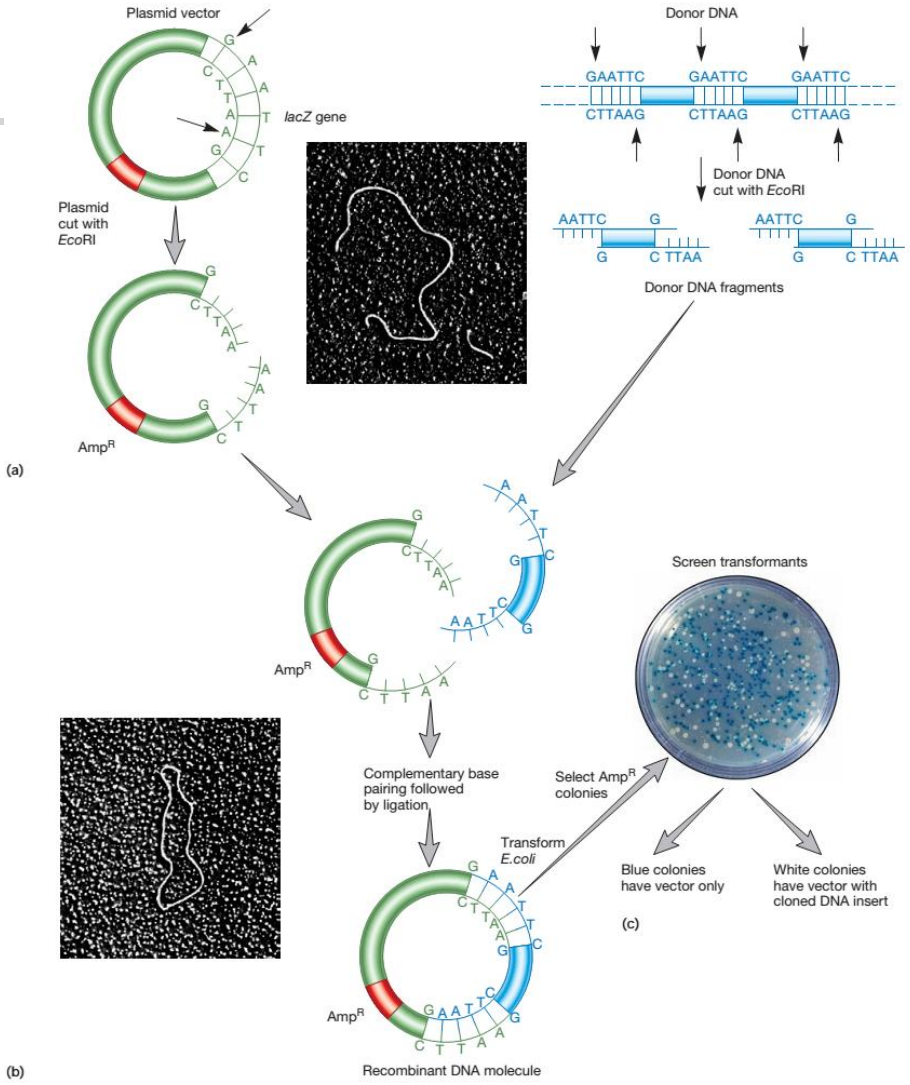


Table 14.3 Recombinant DNA Cloning Vectors

Vector	Insert Size (kb, 1 kb = 1,000 bp)	Example	Features
Plasmid	<20 kb	pBR322, pUC19	Replicates independently of microbial chromosome so many copies may be maintained in a single cell
Bacteriophage	9–25 kb	λ1059, λ gt11, M13mp18, EMBL3	Packaged into lambda phage particles; single-stranded DNA viruses like M13 have been modified (e.g., M13mp18) to generate either double- or single-stranded DNA in the host
Cosmids	30–47 kb	pJC720, pSupercos	Can be packaged into lambda phage particles for efficient introduction into bacteria, then replicates as a plasmid
PACs (P1 artificial chromosomes)	75–100 kb	pPAC	Based on the bacteriophage P1 packaging mechanism
BACs (bacterial artificial chromosomes)	75–300 kb	pBAC108L	Modified F plasmid that can carry large DNA inserts; very stable within the cell
YACs (yeast artificial chromosomes)	100–1,000 kb	pYAC	Can carry largest DNA inserts, replicates in *Saccharomyces cerevisiae*



根据表型效应分以下几类质粒

- 致育因子（Fertility factor，F因子）:致育性
- 抗性因子（Resistance factor，R因子）：对药物和金属产生抗性
- Col质粒（Bacteriocin production plasmid）：产生抗菌物质
- 毒性质粒（virulence plasmid）：产毒素
- 代谢质粒（Metabolic plasmid）：利用异常营养物
- 隐秘质粒（cryptic plasmid）：破解谜底待后生！

1、致育因子(Fertility factor，F因子）

1946年Lederbeger 首次发现*E.coli*的“有性生殖”：结合(conjugation)现象。是由F因子介导的。

E.Coli F因子100kb,单拷贝。
携带F质粒的菌株称为F^+菌株（相当于雄性），菌体表面有性菌毛；
没有F质粒的菌株称为F^-菌株（相当于雌性），菌体表面无性菌毛。

F因子介导结合作用，F^+菌株能够将自己的一份拷贝导入雌性菌株（F^-）而使后者变成F^+菌株。

F因子可游离于菌体细胞质中，也可整合在染色体上。前者为F+菌株，后者为Hfr菌株，故称F因子为附加体(episome)。

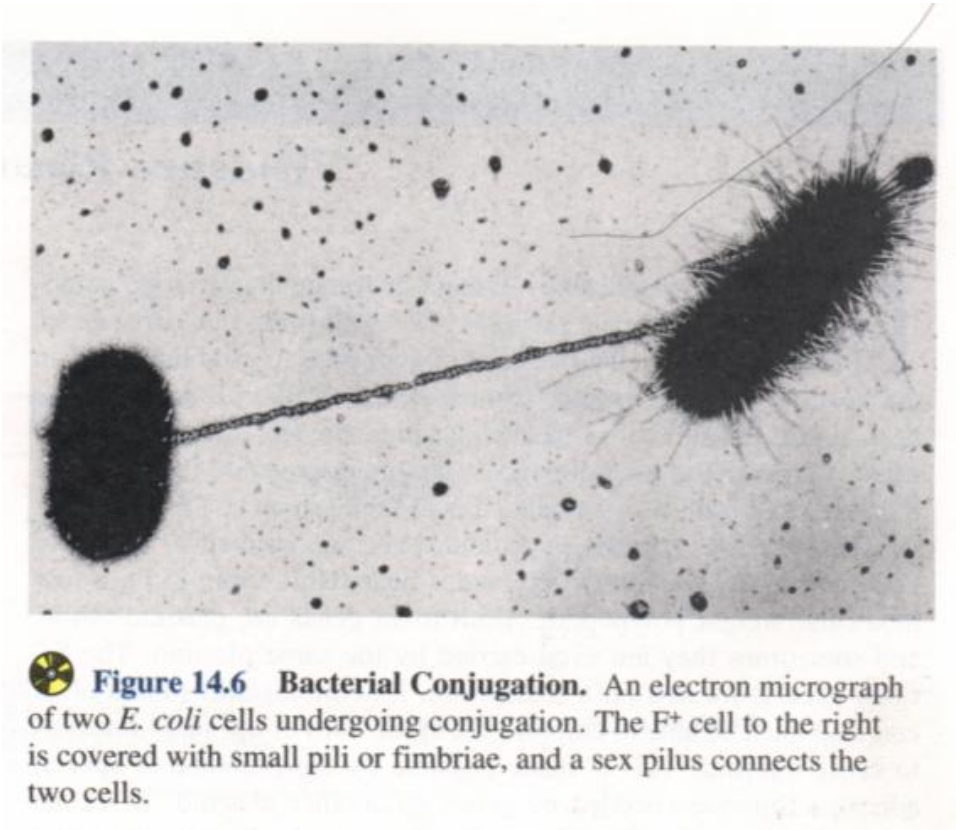


Figure 14.6 Bacterial Conjugation. An electron micrograph of two *E. coli* cells undergoing conjugation. The F^+ cell to the right is covered with small pili or fimbriae, and a sex pilus connects the two cells.

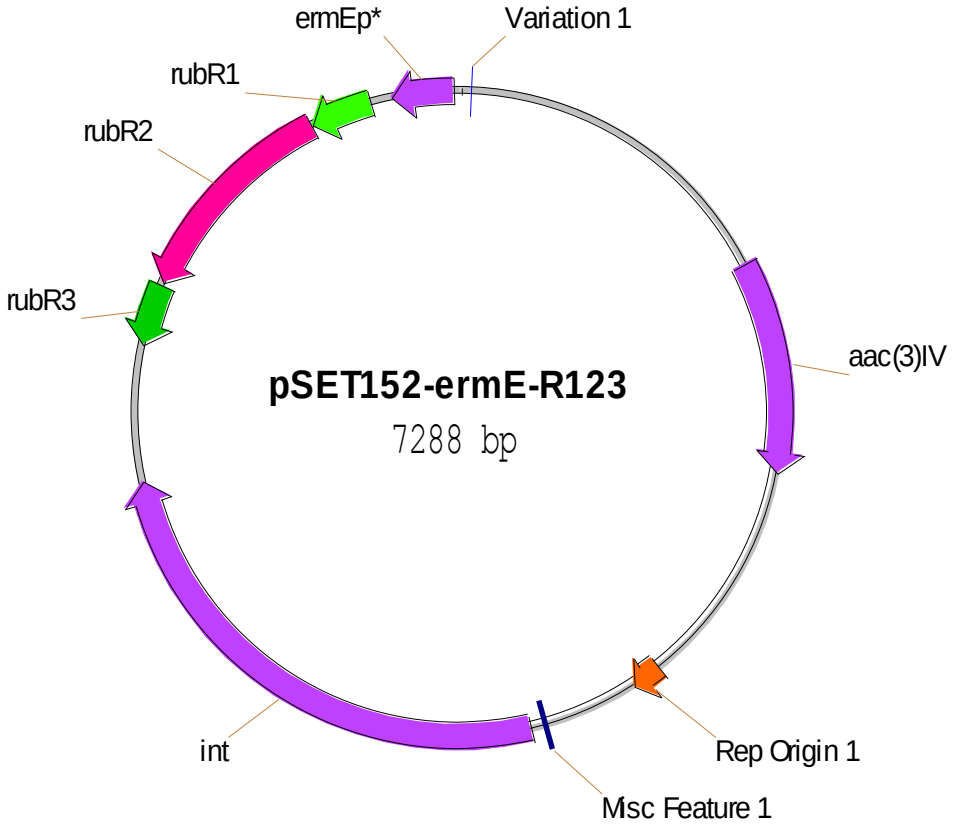
有关F因子的内容在讲细菌的接合作用（conjugation）时具体介绍

F⁺种类：

大肠杆菌(*E.coli*)： **F⁺**

绿脓杆菌(*Bacillus pyocyaneus*)： **FP2, FP5, FP39**

放线菌中(*Actinobacteria*)： **Scp1**



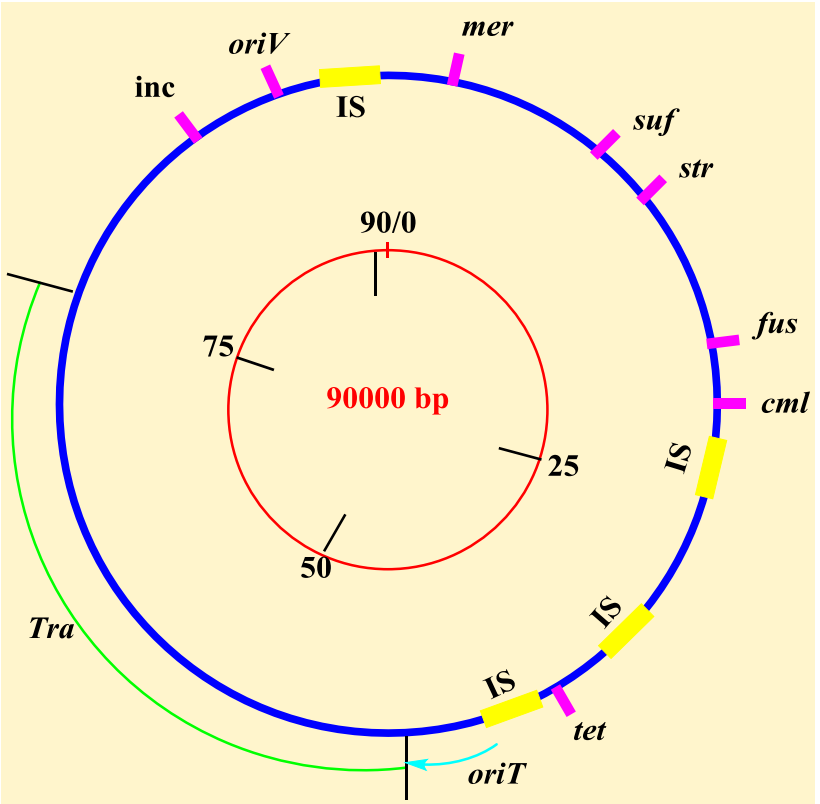
2、抗性因子（Resistance factor，R因子）

包括抗药性和抗重金属二大类，简称R质粒。

质粒类型	名称	大小（kb）	拷贝数	宿主菌	遗传表型
R Plasmids	RP4	54	1–3	*Pseudomonas* and many other gram-negative bacteria	Sex pilus, conjugation, resistance to Amp, Km, Nm, Tet
	R1	80	1–3	Gram-negative bacteria	Resistance to Amp, Km, Su, Cm, Sm
	R6	98	1–3	*E. coli, Proteus mirabilis*	Su, Sm, Cm, Tet, Km, Nm
	R100	90	1–3	*E. coli, Shigella, Salmonella, Proteus*	Cm, Sm, Su, Tet, Hg
	pSH6	21		*Staphylococcus aureus*	Gm, Tet, Km
	pSJ23a	36		*S. aureus*	Pn, Asa, Hg, Gm, Km, Nm, Em, etc.
	pAD2	25		*Enterococcus faecalis*	Em, Km, Sm

R100质粒(89kb)可使宿主对下列药物及重金属具有抗性：汞（mercuric ion ， mer）、四环素（tetracycline，tet ）链霉素(Streptomycin, Str)、磺胺(Sulfonamide, Su)、氯霉素(Chlorampenicol, Cml)、夫西地酸（fusidic acid，fus）并且负责这些抗性的基因是成簇(clustering)存在于抗性质粒上。

某些抗性质粒也是F因子，如右图R100质粒就是F因子，能介导抗性基因在细菌间的传递。某些抗性质粒能够存在于含有F因子的菌体中，也能随着F因子的转移而“搭便车”，也是细菌产生抗药性的重要原因之一。



plasmid R100 map

3、 Col质粒(colicin-producing plasmid)

Col Plasmids	ColE1	9	10–30	*E. coli*	Colicin E1 production
	ColE2		10–15	*Shigella*	Colicin E2

细菌的种类进行命名一般根据产生菌细菌素

比较项目	细菌素(bacteriocin)	(抗生素antibiotics)
来源	常见于G+细菌	绝大多数源于微生物
合成基因位置	质粒和染色体上	多见染色体抗生素基因簇
基因结构	一个基因	多个基因成簇排列
生物合成途径	核糖体合成的短肽	次级代谢合成途径
抗菌谱	窄	寬
作用机制	在菌体表面打孔，不易产生抗性	抗拒机制多样，易产生抗性
蛋白酶降解	容易	适度到不降解
生物活性强度	纳-微摩尔级	微-毫摩尔级
PH范围	寬	窄
热稳定性	高	低
色/味/臭	无	有
对真核细胞毒性	无	有

产细菌素的质粒（Bacteriocin production plasmid）

细菌素(bacteriocin):由某些细菌在代谢过程中由基因编码核糖体合成的一类具有生物活性的多肽或蛋白类物质

由G^+细菌产生的细菌素或与细菌素类似的因子与colicins有所不同，多数由质粒基因编码。细菌素基因并不都在Col等质粒上，铜绿假单胞菌(*Pseudomonas aeruginosa*)细菌素pyocins的基因在染色体上。人肠道内的菌群也能产生细菌素，能发挥免疫保护作用。有些细菌素已经获准应用在临床和食品工业、动物养殖和日常用品。如含有ninisA的牙膏、洗手液等。

Perez *et al. Microbial Cell Factories* 2014, **13**(Suppl 1):S3
http://www.microbialcellfactories.com/content/13/S1/S3

MICROBIAL CELL FACTORIES

PROCEEDINGS Open Access

Novel bacteriocins from lactic acid bacteria (LAB): various structures and applications

4、毒性质粒（virulence plasmid）

许多致病菌的致病性是由其所携带的质粒引起的，这些质粒具有编码毒素的基因，其产物对宿主（动物、植物）造成伤害。

产毒素大肠杆菌是引起人类和动物腹泻的主要病原菌之一，其中许多菌株含有为一种或多种肠毒素编码的质粒。

质粒类型	名称	大小（kb）	拷贝数	宿主菌	遗传表型
Virulence Plasmids	Ent (P307)	83		*E. coli*	Enterotoxin production
	K88 plasmid			*E. coli*	Adherence antigens
	ColV-K30	2		*E. coli*	Siderophore for iron uptake; resistance to immune mechanisms
	pZA10	56		*S. aureus*	Enterotoxin B
	Ti	200		*Agrobacterium tumefaciens*	Tumor induction

苏云金杆菌含有编码δ内毒素(伴孢晶体中)的质粒

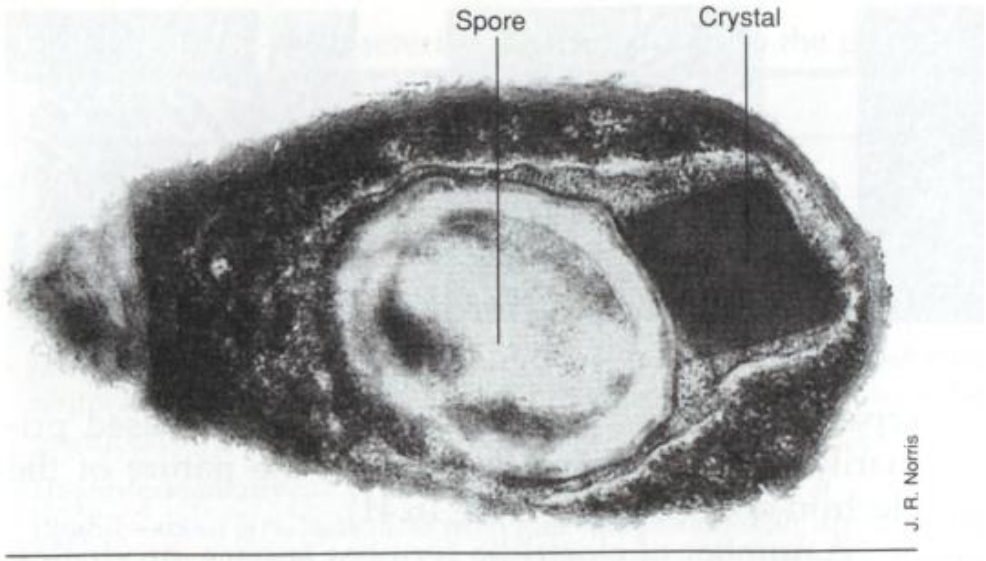


Figure 16.88 Formation of the toxic parasporal crystal in the insect pathogen *Bacillus thuringiensis*. Electron micrograph of a thin section.

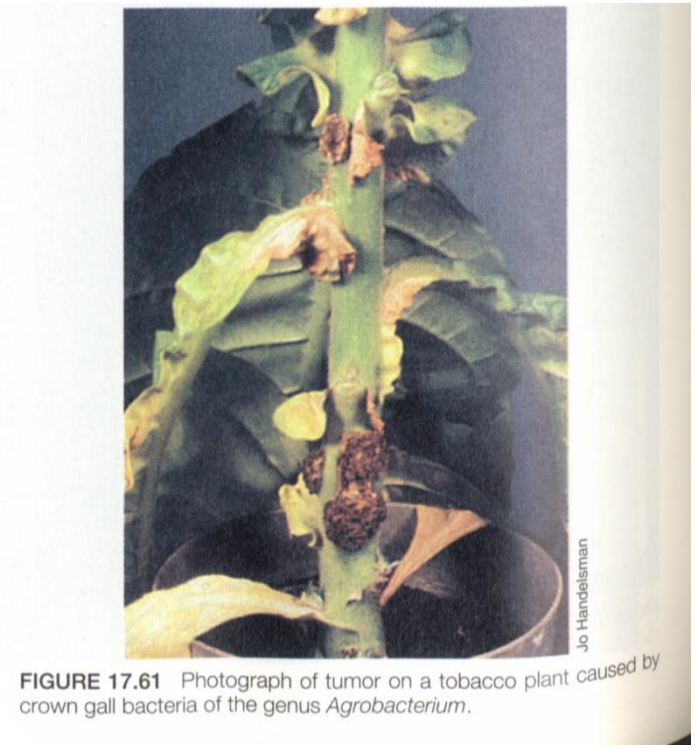


FIGURE 17.61 Photograph of tumor on a tobacco plant caused by crown gall bacteria of the genus *Agrobacterium*.

根癌土壤杆菌所含Ti质粒是引起双子叶植物冠瘿瘤的致病因子

5、代谢质粒（Metabolic plasmid）

质粒上携带有有利于微生物生存的基因，如能降解某些基质的酶，进行共生固氮，或产生抗生素（某些放线菌）等。

降解质粒： 将复杂的有机化合物降解成能被其作为碳源和能源利用的简单形式，环境保护方面具有重要的意义!

质粒类型	名称	大小（kb）	拷贝数	宿主菌	遗传表型
Metabolic Plasmids	CAM	230		*Pseudomonas*	Camphor degradation
	SAL	56		*Pseudomonas*	Salicylate degradation
	TOL	75		*Pseudomonas putida*	Toluene degradation
	pJP4			*Pseudomonas*	2,4-dichlorophenoxyacetic acid degradation
				E. coli, Klebsiella, Salmonella	Lactose degradation
				Providencia	Urease
	sym			*Rhizobium*	Nitrogen fixation and symbiosis

6、隐秘质粒（cryptic plasmid）

很多原核微生物细胞天生的大部分的质粒，都是隐秘性质粒，对宿主细胞而言可有可无，但却能长期稳定存在，某些*E. coli*菌株中，在没有抗性压力的条件下，可以存在上百代而不丢失。在幽门螺旋杆菌（*Helicobacter pylori*）中，一半以上的小质粒属于隐秘质粒。.

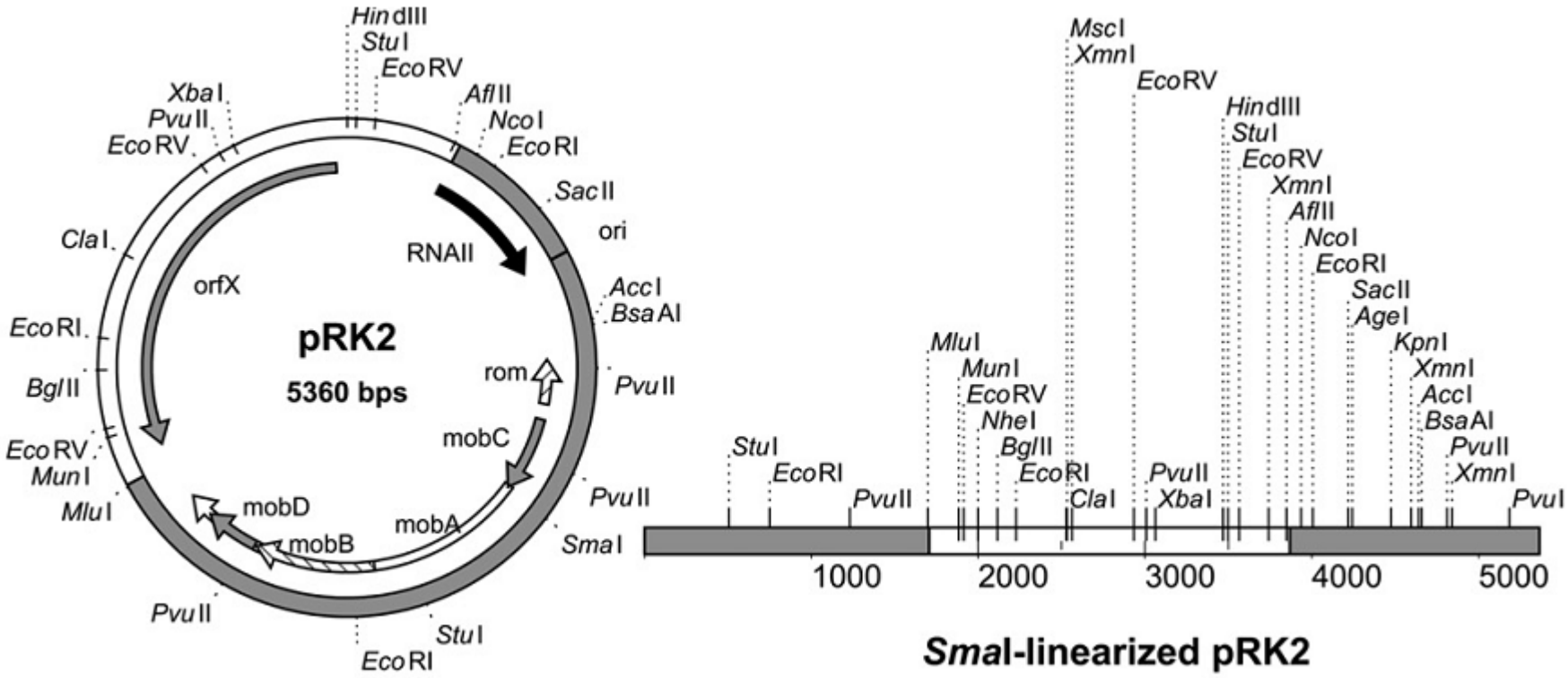
阅读：幽门螺旋杆菌方面（**Cryptic plasmids in *Helicobacter pylori***）的文献

在应用上，很多隐秘质粒被加以改造作为基因工程的载体（一般加上抗性基因）

隐秘质粒不显示任何表型效应，它们的存在只有通过物理的方法，例如用凝胶电泳检测细胞抽提液等方法才能发现。它们存在的生物学意义，目前了解还不够深入。

撩开神秘的面纱，等待在座的每一位…

6、隐秘质粒（cryptic plasmid）



Maps of the cryptic plasmid pRK2 *E.Coli* W

质粒的起源？

借鸡生蛋，经营存身之道；利人利己，奉行双赢策略；
道流天下，不改自由本色；四海为家，一贯逍遥快活。

面对一个质粒，你想到什么？你该如何思考？

生命乎？非命乎？生命物质乎？
抑或是最原始的形式，最简单的形态，还是最简形式？
是遗迹，还是选择？是自然的演化，还是最优的结果？

思接千载，视通万里，穿越时空，追求真相的衣袂…

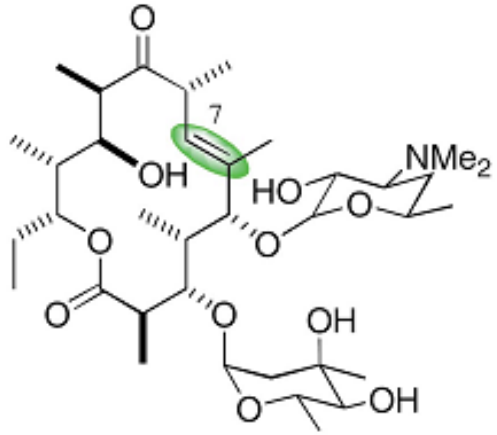
作为遗传操作工具的质粒

Table 14.1	Some Milestones in Biotechnology and Recombinant DNA Technology
1958	DNA polymerase purified
1970	A complete gene synthesized in vitro Discovery of the first sequence-specific restriction endonuclease and the enzyme reverse transcriptase
1972	First recombinant DNA molecules generated
1973	Use of plasmid vectors for gene cloning
1975	Southern blot technique for detecting specific DNA sequences
1976	First prenatal diagnosis using a gene-specific probe
1977	Methods for rapid DNA sequencing Discovery of "split genes" and somatostatin synthesized using recombinant DNA
1978	Human genomic library constructed
1979	Insulin synthesized using recombinant DNA First human viral antigen (hepatitis B) cloned
1981	Foot-and-mouth disease viral antigen cloned First monoclonal antibody-based diagnostic kit approved for use
1982	Commercial production by *E. coli* of genetically engineered human insulin Isolation, cloning, and characterization of a human cancer gene Transfer of gene for rat growth hormone into fertilized mouse eggs
1983	Engineered Ti plasmids used to transform plants

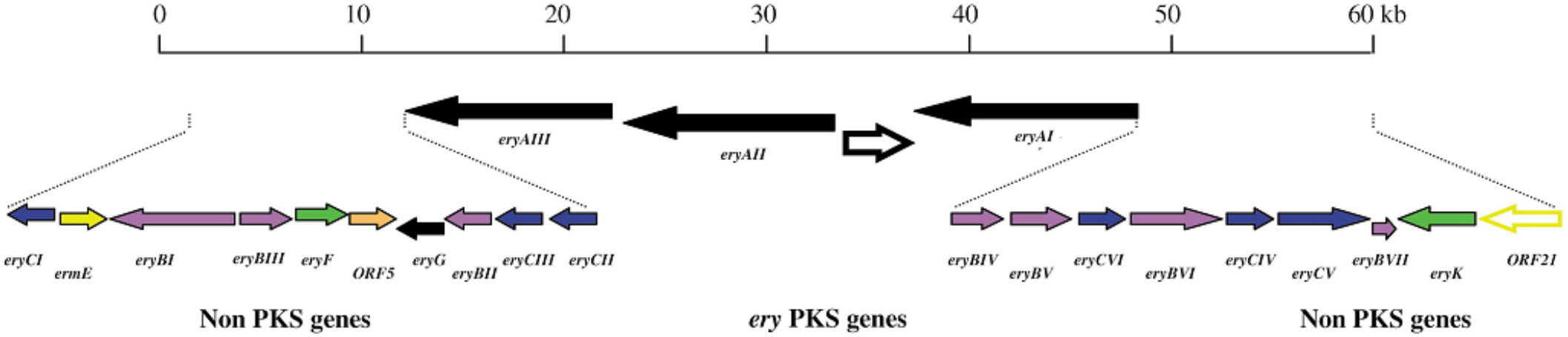
1985	Tobacco plants made resistant to the herbicide glyphosate through insertion of a cloned gene from *Salmonella* Development of the polymerase chain reaction technique
1987	Insertion of a functional gene into a fertilized mouse egg cures the shiverer mutation disease of mice, a normally fatal genetic disease
1988	The first successful production of a genetically engineered staple crop (soybeans) Development of the gene gun
1989	First field test of a genetically engineered virus (a baculovirus that kills cabbage looper caterpillars)
1990	Production of the first fertile corn transformed with a foreign gene (a gene for resistance to the herbicide bialaphos)
1991	Development of transgenic pigs and goats capable of manufacturing proteins such as human hemoglobin First test of gene therapy on human cancer patients
1994	The Flavr Savr tomato introduced, the first genetically engineered whole food approved for sale Fully human monoclonal antibodies produced in genetically engineered mice
1995	*Haemophilus influenzae* genome sequenced
1996	*Methanocaldococcus jannaschii* and *Saccharomyces cerevisiae* genomes sequenced
1997	Human clinical trials of antisense drugs and DNA vaccines begun; *E. coli* genome sequenced
1998	First cloned mammal (the sheep Dolly)
2002	*Plasmodium falciparum* genome sequenced
2003	Completion of the draft of the human genome
2005	Reconstruction of 1918 influenza virus

7.基因簇(gene cluster)结构

微生物世界是天然产物的重要来源和抗生素的资源宝藏。在微生物的基因组上，负责抗生素生物合成或同一代谢途径的多个基因集中排列，成簇地组织在一起，成簇状分布，我们称之为基因簇(gene cluster)。基因簇广泛地存在于原核生物和真菌的染色体及质粒。



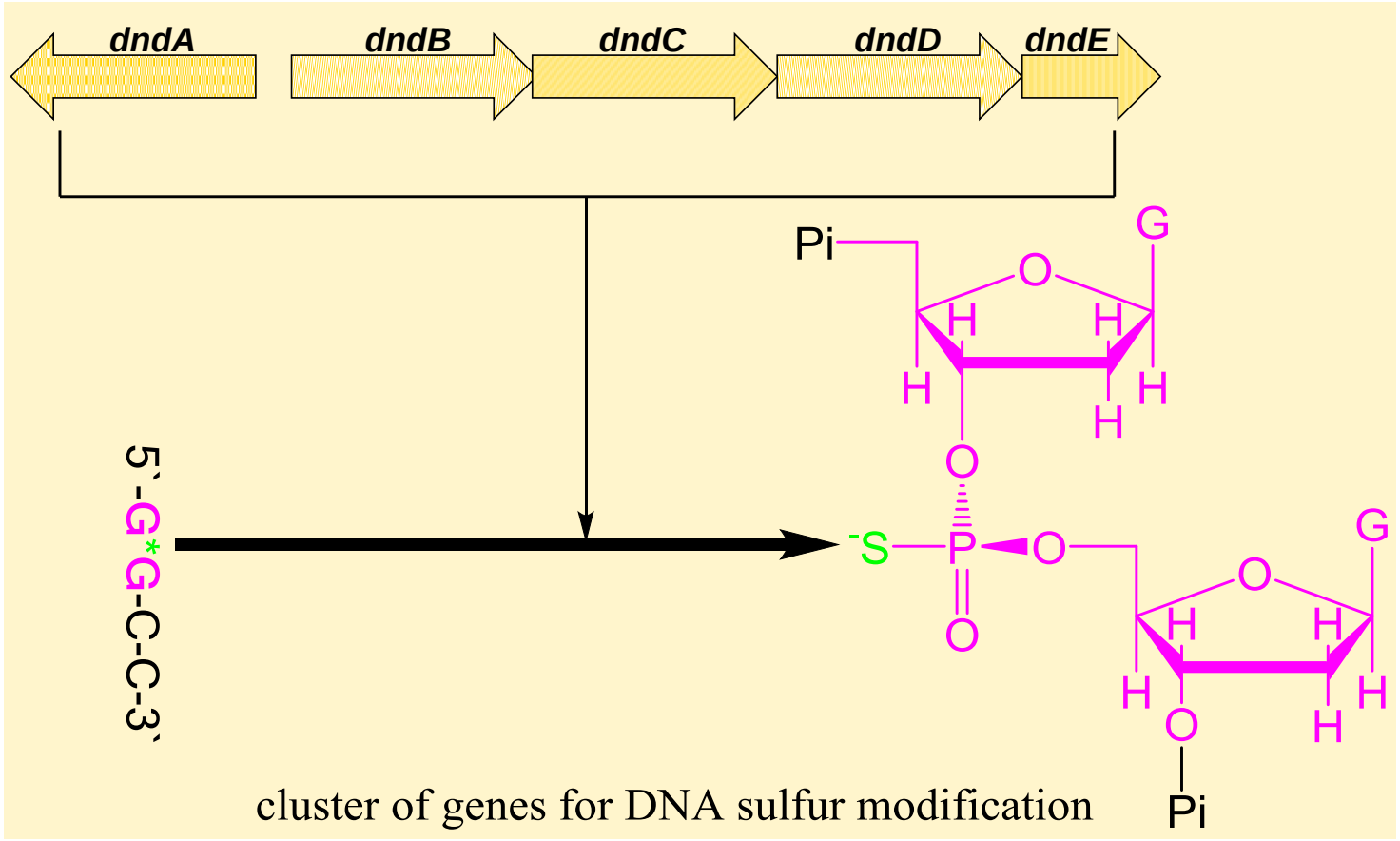
红霉素(erythromycin)



红霉素生物合成的基因簇

负责链霉菌DNA硫修饰的基因簇和作用原理

邓子新院士的成名作：DNA的硫修饰

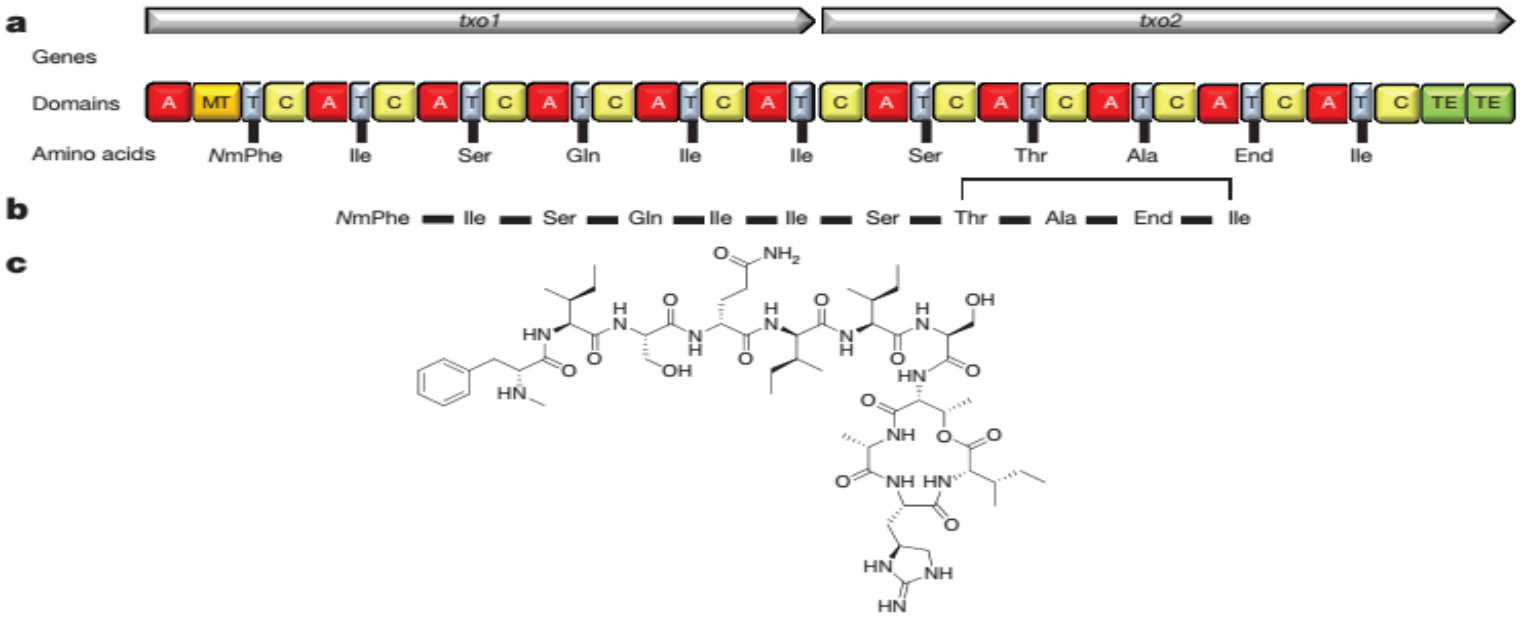


cluster of genes for DNA sulfur modification

ARTICLE

doi:10.1038/nature14098

A new antibiotic kills pathogens without detectable resistance



Thanks for your attention!