Annu. Rev. Genet. 1998. 32:339–77

THE DIVERSE AND DYNAMIC STRUCTURE OF BACTERIAL GENOMES

Sherwood Casjens
Division of Molecular Biology and Genetics, Department of Oncological Sciences, University of Utah, Salt Lake City, Utah, 84132;
e-mail: sherwood.casjens@hci.utah.edu

KEY WORDS: bacterial genome, bacterial chromosome, chromosome structure, genome stability, bacterial diversity

ABSTRACT

Bacterial genome sizes, which range from 500 to 10,000 kbp, are within the current scope of operation of large-scale nucleotide sequence determination facilities. To date, 8 complete bacterial genomes have been sequenced, and at least 40 more will be completed in the near future. Such projects give wonderfully detailed information concerning the structure of the organism's genes and the overall organization of the sequenced genomes. It will be very important to put this incredible wealth of detail into a larger biological picture: How does this information apply to the genomes of related genera, related species, or even other individuals from the same species? Recent advances in pulsed-field gel electrophoretic technology have facilitated the construction of complete and accurate physical maps of bacterial chromosomes, and the many maps constructed in the past decade have revealed unexpected and substantial differences in genome size and organization even among closely related bacteria. This review focuses on this recently appreciated plasticity in structure of bacterial genomes, and diversity in genome size, replicon geometry, and chromosome number are discussed at inter- and intraspecies levels.

CONTENTS

0066-4197/98/1215-0339$08.00

INTRODUCTION

One might consider that a full understanding of the genetic structure of a species' genome is achieved when the complete nucleotide sequence of the genome of a member of that species is known. Such an assumption would likely be wrong for most bacteria. Recent findings have shown an unexpected level of structural plasticity in many bacterial genomes. Even among the genomes of different individuals belonging to the same species there may be very substantial differences. This review explores the structural diversity and "fluidity" of the (eu)bacterial genome, and attempts to delineate the known ways in which individuals differ within species and ways in which related species differ from one another. Archaeal genomes are not covered. This discussion is aimed at a broad readership, which necessitates omitting interesting details, but the reader should garner a flavor of our current knowledge [additional details can be found in other reviews of this topic (25, 52, 128, 130) and the various informative chapters in (60)].

The structure of bacterial genomic DNAs can be analyzed at many levels, including nucleotide nearest neighbor and oligonucleotide frequencies, G+C content, GC skew, nucleotide sequence, gene organization, overall size, and replicon geometry. This discussion focuses on the overall size and geometry and the diversity in these parameters for bacterial genomes. The recent rapid expansion of knowledge of bacterial genome structure is largely due to advances in pulsed-field gel electrophoretic technology, which allows separation of large DNA molecules. This technology, which has made the measurement of bacterial genome size and construction of physical (macro-restriction enzyme cleavage site) maps of bacterial chromosomes relatively straightforward and much more accurate than previous methods, has been amply reviewed elsewhere (52, 60, 77, 226) and is not discussed here.

The new field of bacterial genomics, the study and comparison of whole bacterial genomes, is blossoming and will continue to expand during the coming

decade, as many complete bacterial genome sequences are determined. The complete sequences of eight bacterial genomes are published at this writing. The individual bacteria whose genomes have been sequenced are members of the following species: *Haemophilus influenzae* (75), *Mycoplasma genitalium* (81), *Mycoplasma pneumoniae* (101), *Synechocystis* sp. PCC6803 (121), *Helicobacter pylori* (246), *Escherichia coli* (17), *Bacillus subtilis* (133), and *Borrelia burgdorferi* (80). Over 40 additional complete bacterial genome sequences are anticipated within a few years (the World Wide Web site at The Institute for Genomic Research maintains a current listing and status of bacterial genome sequencing projects; http://www.tigr.org/tdb/mdb/mdb.html).

To understand themes and variations on those themes in bacterial genome structure, this new knowledge must be placed in the context of the bacterial phylogenetic tree. Modern bacterial phylogenetic classification is based mainly on nucleotide sequence comparisons, with rRNA sequence comparisons the most useful for relating distant phyla (110, 261). This tree probably does not accurately describe the phylogenetic relationships of many bacterial genes [e.g. horizontal transfer and perhaps even genome fusion events may have occurred (22, 91, 138, 170)], and, although still incomplete and subject to controversy, the rRNA tree provides an initial framework for discussion and shows that there are currently 23 named major bacterial phylogenetic divisions [and probably at least as many unstudied ones (110)]. Figure 1 presents a diagrammatic version of the bacterial rRNA phylogenetic tree. Because bacteria have limited variation in physical shape and size, we need to be reminded that the branches of their phylogenetic tree are deep and separated by immense time spans, as great or greater than those separating the deep eukaryotic branches (e.g. fungi and vertebrates). Thus, although the bacteria form a self-consistent clade of organisms, substantial differences may well be found among them.

Large DNA replicons are here referred to as chromosomes and smaller ones as either extrachromosomal elements, plasmids, or small chromosomes, as appropriate in each circumstance. However, the definitions of these terms have become fuzzy as new paradigms have emerged. Currently, the de facto definition of a chromosome is as a carrier of housekeeping genes, but even if a replicon is dispensable under some laboratory conditions, its universal presence in natural isolates might suggest that it is essential in the real habitat of the organism. Should we call smaller DNA elements of the latter type plasmids or small chromosomes? And should elements that are essential in particular situations be called dispensable chromosomes? New terminology, and certainly additional information about the nature of particular replicons, may be required for accurate discussion and complete understanding of all these elements.

Archaea
Eucarya
Bacteria
Verrucomicrobiales
Chlamydia 1050
Acidobacterium group
Coprothermobacter group
Planctomycete group 5200
Bacteriodaceae 4800-5300
Chlorobium 2150-2650
Cytophagales
Green sulfur group
Fibrobacter group 3600
Spirochetes
Leptospira 4400 to 4600
Treponema 1000 to 3000
Borrelia 910 to 930
Thermus 1750
Deinococcus 3600
Green non-sulfur bacteria
Dictoglomus group
Thermotogales 1800
Thermodesulfobacterium group
Aquificales 1600
Campylobacter 1200 to 2100
Helicobacter 1700
Nitrospira group
Desulfovibrio 3100 to 3700
Myxococcus 9200
Flexistipes group
Haemophilus 1900 to 2400
Escherichia 4600-5300
Proteobacteria
Synergistes group
Fusobacteria 2400
Burkholdaria 3600+3200+1100
Neisseria 2200
Rickettsia 1200 to 1700
Bradyrhizobium 8700
Rhizobium 3400+1700+1400
Firmicutes (Gram positive)
Low G+C group
Clostridia 3500 to 6700
Actinomycetes
Mycoplasma 580 to 1325
Mycobacterium 2800 to 44000
Streptomycetes 6400 to 8000
Holophaga group
Synechococcus 2700
Anabena 6400
Cyanobacteria
α
β
γ
δ
ε

BACTERIAL CHROMOSOME SIZE, GEOMETRY, NUMBER, AND PLOIDY

Chromosome Size

Bacterial genome sizes can differ over a greater than tenfold range. The smallest known genome is that of *Mycoplasma genitalium* at 580 kbp (81) and the largest known genome is that of *Myxococcus xanthus* at 9200 kbp (99). The median size is near 2000 kbp for those studied to date (225), although this value is likely skewed toward the genome sizes of the more frequently studied pathogenic bacteria. Figure 2 compares this range in genome size to those of members of the other kingdoms of life. It overlaps the largest viruses [bacteriophage G—670 kbp (222)] and the smallest eukaryotes (the microsporidia protozoan *Spraguea lophii*—haploid 6200 kbp (13, but see also 89)]. The average bacterial gene size in the completely sequenced bacterial genomes is uniform so far, at 900 to 1000 bp, and genes appear to be similarly closely packed in these sequences (∼90% of the DNA encodes protein and stable RNA). Thus, larger bacterial genomes have commensurately more genes than smaller ones have. Gene number appears to reflect lifestyle; Bacteria with smaller genomes (down to ∼470 genes in *M. genitalium*) are specialists, such as obligate parasites that grow only within living hosts or under other very specialized conditions, and those with larger genomes (up to nearly 10,000 genes in *M. xanthus*) are metabolic generalists and/or undergo some form of development such as sporulation, mycelium formation, etc [see (225) for a more thorough discussion].

Even within a genus, bacterial chromosome size can be surprisingly variable (Figure 1, Table 1 and references therein). For example, different spirochete

←

Figure 1 Bacterial chromosome size and geometry. An unrooted rRNA phylogenetic tree containing the 23 named major bacterial phyla and some of their relevant subgroups [from Hugenholtz et al (110) and the NCBI Taxonomy web site (http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy)]; at least 12 additional, as yet unstudied major phyla exist (110). The branch lengths are not meant to indicate actual phylogenetic distances, and the *fig leaf* in the center indicates a part of the phylogenetic tree where branching order is less firmly established. Chromosome geometry (circular or linear) is indicated by symbols near the ends of the branch lines; *multiple circles* on the α and β proteobacteria and a spirochete branch indicate that some members of these groups have multiple circular chromosomes. At the end of each branch, representative genera (or higher group names) are given with their known range of chromosome sizes in kbp; if no value is given, none has been determined from that group. Above each name the *black* and *gray circles* indicate the number of genome sequencing projects completed and under way, respectively, for that group in early 1998. Small circular extrachromosomal elements (plasmids) have been found in nearly all bacterial phyla, but linear extrachromosomal elements (indicated by *white squares* below the groups in which they have been found) are rare except in the *Borrelias* and *Actinomycetes*, where they are common. For a color version of this figure, see the color section at the back of the volume.

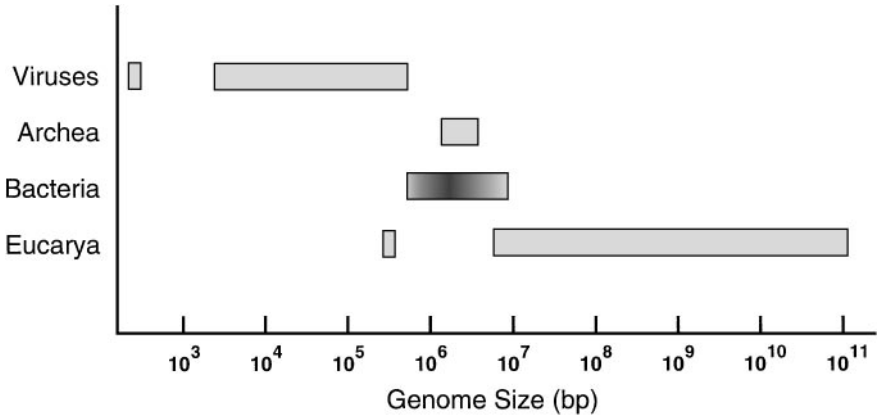


Figure 2 Known genome size ranges for extant life forms on Earth. The range of genome sizes is shown for viruses and the three kingdoms of cellular life forms. The smaller outliers in the virus and eucarya groups are viroids and green algal endosymbionts (89), respectively. The *shading* in the bacterial range reflects the fact that the largest fraction of known bacterial genomes are in the 1–3 mbp range. Figure was modified from Figure 1 of Shimkets (225).

treponemes can vary nearly threefold in chromosome size [the *Treponema pallidum* chromosome is 1040 kbp in length (252), whereas that of *T. denticola* is 3000 kbp (162)] and the firmicute mycoplasmas vary at least 2.3-fold (*M. genitalium*, 580 kbp, to *M. mycoides*, 1350 kbp). Perhaps more typical are the genera *Streptomyces* and *Rickettsia*, which vary from 6400 to 8200 and 1200 to 1700 kbp, respectively. Such variation, although apparently typical, is not universal, since in the ten *Borrelia* species studied in detail, the chromosomes vary less than 15 kbp in size (38, 189); these bacteria may restrict this type of variation to the many plasmids they harbor (see below). Too few species have been studied in most genera and even in many higher phylogenetic groups to give us even a minimal understanding of natural variation in genome size in most groups.

Very large variations in chromosome size within higher phylogenetic divisions appears to be the rule rather than the exception (e.g. proteobacteria, 1200 to 9400 kbp; firmicutes, 580 kbp to 8200 kbp; spirochetes, 910 to 4600 kbp). There is also significant variation in size within many species, as is discussed below. Inspection of the chromosome sizes in Figure 1 and Table 1 shows that genome size is diverse in bacteria, and it seems simplest to imagine that much of this variation in genome size has arisen by rapid gene loss when a species finds happiness in a very specific niche. However, recent discoveries of horizontal genetic transfer among bacteria phyla (see below) make significant increases in genome size also seem plausible. At present, less than half of the

Table 1 Physical analysis of bacterial chromosomes: size, geometry, and number

Major division[1] Subphyla Species	Size (kbp)[2]	Geometry[3]	References
Aquificales			
Aquifex pyrophilus	1600	C	224
Chlamydiae			
Chlamydia trachomatis	1045	C	16
Cyanobacteria			
Anabena sp. PCC7120	6400	C	3
Synechococcus sp. PCC6301	2700	C	120
Synechococcus sp. PCC7002	2700	C	46
Synechocystis sp. PCC6803	3573	C	121
Fibrobacter			
Fibrobacter succinogenes	3600	C	188
Firmicutes			
Low G+C group			
Acholeplasma hippikon	1540	C[4]	180
Acholeplasma laidlawii [2]	1580–1650	C[4]	180, 204
Acholeplasma oculi	1630	C	244
Bacillus cereus [10]	2400–6270	C [5]	28, 30, 32
Bacillus megaterium	4670	—	249
Bacillus subtilis [2]	4200	C [2]	112, 113
Bacillus thuringiensis [2]	5400–5700	C [2]	34–36
Carnobacterium divergens	3200	—	56
Clostridium beijerinckii [2]	4150–6700	C	258
Clostridium perfringens [8]	3650	C [8]	27, 28, 123
Clostridium [5 additional species]	2500–4000	—	148, 267
Enterococcus faecalis	2825	C	177
Lactobacillus [4 species]	1800–3400	—	57, 58
Lactococcus acidophilus	1900	C	212
Lactococcus helveticus [3]	1850–2000	—	158
Lactococcus lactis [20]	2100–3100	C [5]	57, 58, 139, 140, 247
Listeria monocytogenes [37]	2300–3150	C [2]	37, 100, 175
Leuconostoc [4 species]	1750–2170	—	242
Mycoplasma sp. PG50	1040	C	199
Mycoplasma [12 species]	735–1300	C[4]	180
Mycoplasma capricolum	800	C	176, 257
Mycoplasma flocculare	890	—	204
Mycoplasma gallisepticum [3]	1000–1050	C [3]	92, 243
Mycoplasma genitalium	580	C	194
Mycoplasma hominis [5]	700–770 [5]	C	135
Mycoplasma hyopneumoniae	1070	—	204
Mycoplasma mobile	780	C	9
Mycoplasma mycoides [6]	1200–1350	C [5]	180, 198, 199

(*Continued*)

Table 1 (*Continued*)

Major division[1] Subphyla Species	Size (kbp)[2]	Geometry[3]	References
Mycoplasma pneumoniae	800–840	C	255, 256
Mycoplasma-like organisms [11]	575–1185[5]	—	179
Western X-disease agent	670	C	73
Oeoncoccus oeni [41]	1780–2100	—	136, 242
Pediococcus acidilactici [2]	1860–2130	—	159
Spiroplasma citri	1780	C	266
Sprioplasma [20 species]	940–2200	—	29
Staphylococcus aureus [2]	2750	C	4
Streptococcus caronsus	2590	C	251
Streptococcus thermophilis [3]	1824–1865	C [3]	211
Streptococcus mutans	2100	C	97, 269
Streptococcus pneunomiae	2270	C	87
Streptococcus pyogenes	1920	C	233
Ureaplasma diversum [3]	1100–1160	—	119
Ureaplasma urealyticum [16]	760–1140	C	51, 204
Ureaplasma [7 species]	760–1170	—	119
Actinomycetes			
Bifidobacterium [8 species]	1600–2000	—	185
Brevibacterium lactofermentum	3050	—	53
Corynebacteria glutamicum	2990	—	53
Micrococcus sp. Y-1	4010	C	192
Mycobacterium leprae	2800	—	70
Mycobacterium bovis	4350	C	195
Mycobacterium tuberculosis	4400	—	196
Rhodococcus facians	4000	L[4]	54
Rhodococcus sp. R312	6400	—	14
Streptomyces ambofaciens [4]	6420–8150	L	11, 141, 142
Streptomyces coelicolor	7900	L	125, 201, 248
Streptomyces griseus	7800	L	147
Streptomyces lividans	8000	L	143, 149
Fusobacteria			
Fusobacterium nucleatum [6]	2400	—	18
Green sulfur bacteria and Cytophagales			
Chlorobium limnicola [2]	2460–2650	—	171
Chlorobium tepidum	2150	C	178
Bacteriodes [7 species]	4800–5300	—	223
Planctomycetes and relatives			
Planctomyces limnophilus	5200	C	254
Proteobacteria			
α group			
Agrobacterium tumefaciens	3000 & 2100	C & L	1, 116
Bartonella bacilliformis	1600	C	131

(*Continued*)

Table 1 (*Continued*)

Major division[1] Subphyla Species	Size (kbp)[2]	Geometry[3]	References
Bradyrhizobium japonicum [2]	6200–8700	C	132, 227
Brucella abortus	2050 & 1100	C	174
Brucella suis [3]	3200, 2000 & 1150, 1850 & 1350	C [3]	117, 174
Brucella canis	2150 & 1170	C	174
Brucella ovis	2030 & 1150	C	174
Brucella neotomae	2050 & 1170	C	174
Brucella melitensis	2050 & 1130	C	116, 173
Caulobacter crescentus	4000	C	71
Rhizobium meliloti	3450, 1700 & 1340	C	109
Rhodobacter capsulatus	3800	C	76, 78, 79
Rhodobacter sphaeroides	3050 & 914	C	48, 49, 234
Rickettsia bellii	1660	—	213
Rickettsia melolanthae	1720	—	82
Rickettsia rickettsii	1270	—	213
Rickettsias [14 species]	1220–1400	—	213
Zymomonas mobilis	2080	C	71
β group			
Bordetella pertussis [5]	3750–4020	C [5]	229, 230
Bordetella parapertussis	4780	C	230
Burkholderia cepacia 17616	3400, 2500 & 900	C	47
Burkholderia cepacia 25416	3650, 3170 & 1070	C	205
Burkholderia cepacia [10]	4600–8100[6]	—	145
Burkholderia glumae	3800, 3000	C[4]	205
Neisseria gonorrhoeae [2]	2200–2330	C [2]	15, 62, 63
Neisseria meningitidis [2]	2200–2300	C [2]	10, 64, 85
Thiobacillus ferrooxidans	2900	C	111
Thiobacillus cuprinus	3800	C	169
γ group			
Acinetobacter sp. ADP1	3780	C	93
Alteromonas nigrifaciens	2040	—	235
Alteromonas sp. M-1	2240	—	235
Azotobacter vinelandii	4700	—	165
Dichelobacter nodosus	1540	C	134
Escherichia coli [15]	4600–5300	C	12, 193
Haemophilus ducreyi	1760	C	106
Haemophilus influenzae [2]	1980–2100	C [2]	24, 144
Haemophilus parainfluenza	2340	C	124
Proteus mirabilis	4200	—	2
Salmonella enteritidis	4600	C	152
Salmonella paratyphi	4600–4660	C	155
Salmonella typhi [2]	4780	C	124, 156, 157
Salmonella typhimurium	4700	C	153

(*Continued*)

Table 1 (*Continued*)

Major division[1] Subphyla Species	Size (kbp)[2]	Geometry[3]	References
Shewanella putrefaciens	2380	—	235
Shigella flexnerii [2]	4600	C	190
Vibrio cholerae	3200	C	164
Yersinia pestis	4400	—	160
Yersinia ruckeri	4600	—	206
γ^* group			
Chromatium vinosum	3670	C	86
Moraxella catarrhalis	1940	C	84
Pseudomonas aeruginosa [23]	5900–6600	C [23]	108, 207–210, 219
Pseudomonas fluorescens	6630	C	200
Pseudomonas putida	3610–5960	C	50, 107
Pseudomonas solanacearum	5550	C	107
Pseudomonas stutzeri [20]	3060–4640	C	50, 90
Pseudomonas syringae	5640	C	61
δ group			
Desulfovibrio desulfuricans	3100	—	65
Desulfovibrio propionicus	3600	—	65
Desulfovibrio vulgaris	3700	—	65
Myxococcus xanthus	9200	C	99
Stigmatella aurantiaca [11]	9200–9900	C	181, 182
Stigmatella erecta [2]	9800	—	181
ε group			
Campylobacter coli	1690	C	241, 265
Campylobacter fetus [3]	1300–2120	C [3]	43, 83, 216, 217
Campylobacter jejuni [4]	1720	C [4]	20, 43, 122, 126, 183, 184, 241
Camplyobacter laridis	1470	—	43
Campylobacter upsaliensis	2000	C	20
Helicobacter mustelae [15 isolates]	1700	—	239
Helicobacter pylori [5]	1670–1740	C [5]	23, 115, 240
Radioresistant micrococci and relatives			
Deinococcus radiodurans	3580	—	94
Thermus thermophilus [2]	1740–1820	C [2]	19, 237
Spirochetes			
Borrelia afzelii [3]	910	L [3]	38, 189
Borrelia andersonii [2]	910	L [2]	38
Borrelia burgdorferi [7]	910–925	L [7]	38, 41, 59
Borrelia garinii [3]	910	L [3]	38, 189
Borrelia hermsii	950	L[4]	127
Borrelia japonica [2]	910–930	L [2]	38
Borrelias [5 additional species]	910–920	L [5]	38
Borrelias [4 additional species]	900–950	L[4]	41

(*Continued*)

Table 1 (*Continued*)

Major division[1] Subphyla Species	Size (kbp)[2]	Geometry[3]	References
Leptospira interrogans [2]	4500 & 350	C [2]	7, 270, 271
Serpulina hyodysenteria	3200	C	272
Spirochaeta aurantia	~3000	C[4]	72
Treponema denticola	3000	C	162
Treponema pallidum	1080	C	252

[1]Phylogenetic grouping is as described in the legend to Figure 1. Numbers in square brackets after name are the number of isolates whose genome size has been measured if it is >1.

[2]Only those sizes are included that have been determined directly by pulsed-field gel electrophoretic size measurement of whole DNA or a small number of fragments.

[3]L, linear map; C, circular map; —, not determined. Square brackets in "Geometry" column indicate the number of isolates for which physical maps have been constructed if that number is greater than 1. In those cases where no map has been constructed the number of replicons that gave rise to the DNA fragments that were summed to calculate genome size is generally not known.

[4]Chromosomes thought to be linear or circular from behavior in pulsed-field electrophoresis gels (linear molecules enter gel, while similarly sized circles do not). No map has been constructed.

[5]Some findings suggest multiple chromosomes may be present in some of these non-culturable mycoplasma-like plant pathogens.

[6]All *Burkholderia cepacia* appear to have 2 to 4 >1000-kbp replicons.

"characterized" major bacterial phylogenetic divisions have been analyzed for genome size and geometry (Figure 1).

Replicon Geometry

Until recently, all bacterial replicons were assumed to be circular. Although this rule is true for most bacteria, increasing numbers of exceptions are being identified. The known geometries of bacterial chromosomes are summarized in Figure 1 and Table 1. In the early 1990s, two bacterial genera, the spirochete *Borrelias* and the actinomycete *Streptomyces*, were proven to have linear chromosomes (41, 45, 59, 149), and to date all species studied in these two genera have this chromosome geometry, although the *Streptomyces* may naturally interchange between linear and circular (250). Both genera also commonly carry linear plasmids. Other studied members of the spirochete group (*Treponemas*, *Leptospiras*, and a *Serpulina*) have circular chromosomes, and genera related to *Streptomyces* also have circular chromosomes [*Micrococcus* and *Mycobacterium*; however, the actinomycete *Rhodococcus facians* has been suggested to have a linear chromosome (54)]. Furthermore, the proteobacterial species *Agrobacterium tumefaciens* is reported to have a 2100-kbp linear replicon in addition to a 3000-kbp circular replicon (1, 116). Finally, rare linear plasmids have been described in the proteobacteria *Thiobacillus* (260), *Klebsiella* (231),

and *Escherichia* (236). That two different types of replicon linearity exist locally on the bacterial phylogenetic tree strongly suggests that linearity arose at least twice from circular progenitors; however, we have no experimental indication as yet of what advantage may have been gained by becoming linear in these cases.

The structures of the telomeres of linear replicons have been studied only in *Borrelia*, *Escherichia*, and *Streptomyces*, and they take two very different forms. The *Borrelia* replicon ends are covalently closed hairpins, where one DNA strand loops around and becomes the other strand (42, 103, 104). This type of telomere is rare in cellular organisms. The *Borrelias* carry many linear extrachromosomal elements, which also have this type of terminal structure. Only one other bacterial replicon is known to have this type of telomere, the linear N15 prophage plasmid of *E. coli* (166), and some eukaryotic organelle DNAs have at least one hairpin end (105). The *Borrelia* linear DNAs and the N15 plasmid have 20- to 30-bp inverted sequence repeats at the two ends, but in neither case is it understood how the telomeres are replicated (26, 42, 167). The *Streptomyces* telomere DNAs, on the other hand, are open ended and have specific proteins covalently attached to the 5′ ends; these proteins are thought to have primed terminal replication (45, 215). Many linear plasmids with this type of telomere have been described in the *Streptomyces* and other actinomycete genera (see 45, 215). Lin et al (149) used an insertion vector to circularize the *S. lividans* chromosome, and Ferdows et al (72) found a naturally occurring, circularized version of a *Borrelia* linear plasmid, indicating that linearity is not required for their replication in culture.

Chromosome Number

Most bacteria have a single large chromosome (Figure 1; Table 1). In addition, extrachromosomal DNA elements (plasmids) can be found in many if not all species. Plasmids are not universally present (isolates exist that carry no extrachromosomal elements), but they can be very common. *Borrelias*, for example, appear to always carry multiple small replicons (see below). Extrachromosomal elements have been documented in virtually all genera examined to date.

However, members of several bacterial genera have recently been found to contain two or three large replicons (chromosomes) greater than several hundred kilobase pairs: the α-proteobacterial genera *Agrobacterium* (1, 116), *Brucella* (116, 117, 173), *Rhizobium* (109), and *Rhodobacter* (49, 234), the β-proteobacterial genus *Burkholderia* (47, 145, 205), some isolates of the firmicute *Bacillus thuringiensis* (35), and the spirochete genus *Leptospira* (271). In at least two, *Brucella* and *Burkholderia*, current analyses are sufficient to tentatively conclude that multiple chromosomes are a stable property of the genus (Table 1). In particular, six species of *Brucella* each have two chromosomes of

about 2100 and 1200 kbp (174) (but see below), each carrying hybridization targets of important housekeeping genes. The *B. cereus*/*thuringiensis* complex represents a different paradigm; in the isolates analyzed, four have a single chromosome in the 5500- to 6300-kbp size range, whereas another has a 2400-kbp chromosome (30–32, 34). In the latter isolate, a number of probes that hybridized with the chromosomes of the other isolates hybridized with large extrachromosomal DNA. Ribosomal RNA genes were found only on the 2400-kbp replicon, but some housekeeping gene hybridization targets were found on the extrachromosomal DNAs. In *Rhodobacter sphaeroides*, both the 3000- and 900-kbp replicons appear to carry important housekeeping genes; in *Agrobacterium tumefaciens*, both large replicons carry rRNA genes; in two isolates of *Burkholderia cepacia*, all three large replicons carried rRNA genes, and in two isolates of *Leptospira interrogans*, the 4500-kbp replicon carries most essential genes, but the hybridization target for a gene thought to be essential in cell wall synthesis lies on the 350-kbp replicon. Thus, in all these cases, important genes are found on all large replicons, justifying the moniker chromosome.

By contrast, in *Borrelia* and *Rhizobium* extrachromosomal elements are found that are always present, and are essential for their lifestyles in the wild, but which carry no genes essential for growth in culture. In *Rhizobium meliloti*, all three rRNA operons lie on the 3400-kbp replicon, whereas the 1400- and 1700-kbp replicons appear to carry genes required for plant symbiosis. *Borrelia* harbors the largest number of extrachromosomal elements yet found in bacteria, and all natural isolates carry multiple linear DNAs in the 5- to 180-kbp size range and several circular plasmids (8 to 60 kbp) (see 5, 263). In only one isolate, the type strain B31, has the complete complement of extrachromosomal elements been delineated; it carries 12 linear and 9 circular replicons between 5 and 54 kbp in size (80; S Casjens, WM Huang, G Sutton, N Palmer, R van Vugt, B Stevenson, P Rosa, R Lathigra & C Fraser, unpublished data). Its complete genome sequence is known, and only a small number of plasmid-encoded genes, *guaA* and *guaB*, whose products convert IMP to GMP, and several transporter genes, appear to be potentially metabolically or structurally critical for life as a cell (80, 168). Nearly all of these DNA elements can be lost without affecting growth in culture (214). Some, and possibly most of these plasmids are required for the complex, but very specialized, life in which it obligatorily exists alternately inside arthropods and vertebrates (220, 264). Several of these extrachromosomal elements are present in all natural isolates examined, and where they have been analyzed, they have the same gene order in all isolates (167, 245). Their required presence for successful parasitization of their obligate hosts indicates their importance in real life (220, 264), and it has been persuasively argued that they should be considered mini-chromosomes (6).

Once again, the localized presence of multiple chromosomes in the phylogenetic tree suggests they have independently arisen from single chromosomes a number of times, but their advantage remains a mystery. Interestingly, Itaya & Tanaka (114) have recently divided the single, circular *Bacillus subtilis* chromosome into two circular large parts that appear to function well independently.

Chromosome Copy Number

The characterization of bacteria as haploid is an oversimplification. In exponential growth phase, bacteria, especially fast-growing bacteria, contain on average four or more times as many copies of sequences near the origin as near the terminus of replication, and in a few cases, particular species have been observed to carry more than one complete copy of their chromosome per cell. Three such cases are *Azotobacter vinelandii*, *Deinococcus radiodurans*, and *Borrelia hermsii*. Studies of *A. vinelandii* indicate that in rich medium at late exponential phase, it exhibits a large increase in chromosomes per cell (165), but the function of this increase is unknown. *D. radiodurans*, an extremely radiation-resistant bacterium, appears to have about four copies of its chromosome in stationary phase, which can homologously recombine to regenerate an intact chromosome after severe radiation damage (55). Chromosome copy number measurements for *B. hermsii* indicated that during growth phase in mice it contains 13–18 copies of its genome, and it has been hypothesized that this could allow nonreciprocal recombination among plasmids during the cassette replacement mechanism that generates diversity in expression of its major outer-surface protein (127) (see below). Finally, the streptomycetes are partially diploid in that they carry large 24- to 550-kbp inverted terminal duplications on their linear chromosomes. Housekeeping genes have not been found in the duplicated regions (250). Plasmids found in natural bacterial isolates typically have low copy number, close to that of the chromosome, but this is not reviewed here.

GENE CLUSTERING, ORIENTATION, AND LACK OF OVERALL GENOME SYNTENY

The construction of detailed genetic maps of *E. coli* and *B. subtilis* first indicated that overall gene order in bacteria is fluid over long evolutionary time spans. There is little similarity in overall gene order (synteny) among the different major phyla. For example, Figure 3 indicates the lack of overall colinearity in gene order even between two proteobacteria, *H. influenzae* and *H. pylori* (see also Figure 1 in Reference 128 and Figure 3 in Reference 238). No compelling rationales for overall bacterial gene orders have been devised, although genes near the origin of replication will be present in a higher copy

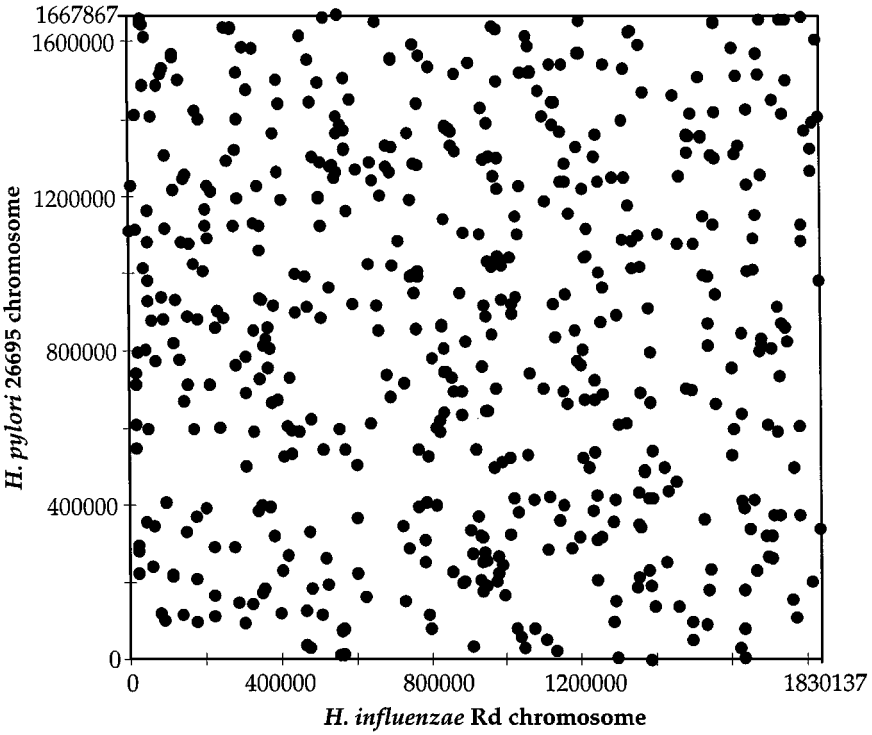


Figure 3 Lack of conservation of gene order between *Haemophilus influenzae* and *Helicobacter pylori*. Linearized chromosomes of *H. influenzae* and *H. pylori* are plotted on the horizontal and vertical axes, respectively, beginning with position one as defined in Fleischmann et al (75) and Tomb et al (246). Each *dot* indicates the intersection of perpendicular lines from the positions of orthologous genes in the two genomes. Genes in similar operons, which do exist, are too close together to give separated points on the scale used. Data for this figure were compiled by O White (personal communication).

number than those near the terminus during times when chromosomes are being duplicated, and superhelical densities could vary systematically across chromosomes. Analysis of additional complete genome sequences should help in answering this question. On the other hand, gene orientation is often more regular. The chromosomes of *M. genitalium*, *B. subtilis*, and *B. burgdorferi*, for example, have 85%, 75%, and 66% of their genes, respectively, oriented so that the (putative) origins of replication program replication forks to pass over them in the same direction as transcription. However, the chromosome of *E. coli* has a lower fraction of genes oriented in this manner (55%) (17). The preferred orientation of the first three chromosomes above is speculated to

minimize head-on collisions between replication and transcription complexes (151).

In contrast to the lack of conservation of overall gene order, cognate operons often, but not always, have similar arrangements in distantly related species, i.e. genes that work together appear to stay together, for extremely long time spans. Examples of parallel operon structure in different phyla are too numerous to list here, but macromolecular synthesis gene clusters are highly conserved across all bacteria; two examples suffice. The universally present *dnaA* gene encodes the protein that recruits the DNA replication apparatus to the origin of replication. It has convincing orthologs in all bacteria where adequate searches have been performed. There are notable exceptions, but a striking observation is that the genes *dnaA*, *dnaN* (DNA polymerase subunit), *gyrB* (DNA gyrase subunit), and *rpmH* (ribosomal protein) are usually found in immediate proximity to one another (187, 191). Likewise, similar ribosomal protein gene clusters are present in such distantly related bacteria as the proteobacteria, spirochetes, and firmicutes (232). Why should this be the case, given the apparent evolutionary mobility of DNA regions relative to one another? Ease of advantageous co-regulation could be a factor, but even where genes have stayed together, regulatory mechanisms have sometimes changed; for example the very similar *E. coli* and *B. subtilis* tryptophan operons have different regulatory mechanisms (172). And genes in an operon whose structure is conserved in some species can be properly regulated as dispersed genes in other species. Two nonmutually exclusive hypotheses to explain the fact that gene clusters often remain intact are that (*a*) horizontal transfer between lineages drives the clustering of genes whose products work together, since the transferred DNA is more likely to survive to fixation in the recipient species if it contains all the genes necessary to perform an advantageous biochemical task (138), and (*b*) clustering minimizes the formation of inactive hybrids that might form as a result of horizontal transfer, e.g. in which two different genetic specificities that must work together cannot do so, such as a regulatory protein and its operator or two interacting proteins (40, 74).

SIMILARITIES AND DIFFERENCES IN GENOME STRUCTURE BETWEEN CLOSELY RELATED SPECIES

The detailed genetic maps of the *E. coli* and *Salmonella typhimurium* chromosomes revealed that these two species have largely similar gene orders (see 154, 218), and physical mapping of these species and their close relatives *S. enteriditis* and *S. paratyphi* showed that they also have similar gene contents and orders (17, 152, 154). *Salmonella* and *Escherichia* are closely related genera in the enterobacteria group, which are thought to have separated about

140 MYA (186). This apparent genome stability may have led to a narrow (incorrect?) view that considers bacterial genome structure in general to be stable over evolutionary times. More recently, significant differences in genome structure have been found in a number of closely related species. Table 2 summarizes some of the currently available information in this area (see references in Table 1). Many, but not all, intermediate level bacterial phyla (genera and sibling genera) display substantial internal differences in overall gene content

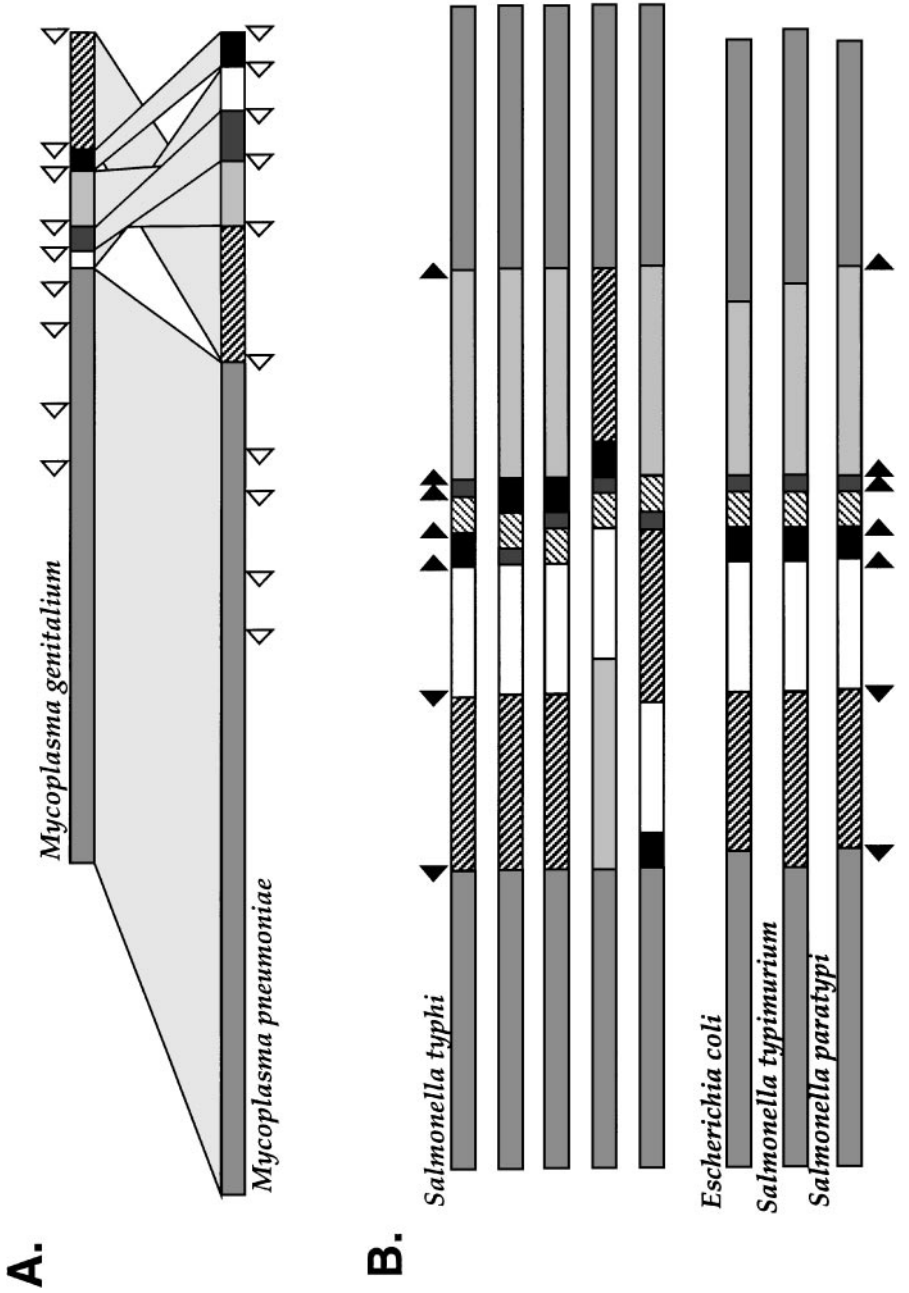
Table 2 Genome structure relationships among some closely related species

Group Genus	Physical map comparisons
α Proteobacteria	
Brucella	*B. abortus* has a 640-kbp inversion relative to five other species. In addition, the two chromosomes appear to be fused in at least one isolate of *B. suis*
Rhodobacter	Substantial differences in gene order exist between *R. capsulatus*, which has one chromosome, and *R. sphaeroides*, which has two
β Proteobacteria	
Bordetella	*B. pertussis* and *B. parapertussis* have little conservation of gene order at present resolution
Neisseria	*N. menigitidis* and *N. gonorrhoeae* have some similarity in gene order but also have complex rearrangements relative to one another
γ Proteobacteria	
Salmonella & *Escherichia*	*E. coli*, *S. typhimurium*, and *S. paratyphi* have similar overall gene orders and *S. typhi* is variable (see text)
ε Proteobacteria	
Campylobacter	*C. jejuni* and *C. upsalensis* appear to have rather similar gene orders
Firmicutes	
Bacillus	Gene orders in *B. subtilis* and *B. cereus* are rather different
Clostridium	Maps of *C. perfringens* and *C. beijerinckii* chromosomes are of very different size, and gene order is not identical
Mycoplasma	*M. genitalium* and *M. pneumoniae* have large rearranged segments relative to one another (see text)
Streptomyces	*S. lividans* and *S. coelicolor* have similar gene orders despite little apparent conservation of restriction sites
Cyanobacteria	
Synechococcus	*Synechococcus* sps. PCC6301 and PCC7002 have similar chromosome sizes, but do not appear to have highly conserved gene orders
Spirochetes	
Borrelia	Little detectable variation in chromosome size or gene order in ten closely related species (multiple isolates from each species)

and arrangement, and numerous examples of such differences between closely related species, or even within species, are now known. Given this complexity, when only one individual from a particular species is shown to be different from other related species (e.g. *Brucella* in Table 2), it is not clear whether this represents an isolated event or a systematic difference in that species. Given the surprisingly high level of large-scale differences in genome structure among closely related bacteria, detailed analysis of multiple independent isolates is essential to draw firm conclusions about genome structure relationships. Such overall genome structure differences have not yet been incorporated into taxonomic classification schemes, and in some cases, they may be pointing out areas in need of taxomonic re-evaluation.

The very low resolution of most bacterial chromosome physical and genetic maps makes it premature to draw detailed conclusions about the exact nature of most of the observed structural genome differences. However, a very detailed comparison can be made between the 580-kbp chromosome of *M. genitalium* and the 816-kbp chromosome of *M. pneumoniae*, the only pair of close relatives with completely sequenced genomes. Himmelreich et al (102) compared these two genomes in detail (summarized in Figure 4*A*). There has been considerable change in genome structure since the two species diverged—both deletions/insertions and other rearrangements are required to convert the gene order of one into that of the other; the chromosome can be viewed as six segments whose order, but not orientation, has been shuffled between the two species. Interestingly, directly repeated sequences, called MgPa sequences, between the six moveable sections may have mediated the rearrangements. An ortholog of every gene in *M. genitalium* is found in *M. pneumoniae*, suggesting that the former may have arisen though a set of deletions and segment re-ordering events from a *M. pneumoniae*–like precursor; such presumed deletions have occurred in each of the six segments shown in Figure 4*A*.

Figure 4 Large differences in genome structure among species within a genus and within a species. *A.* Relationship between *M. pneumoniae* and *M. genitalium*. The *different shades* represent different genomic segments, and the *open triangles* indicate the locations of the repeated MgPa sequence (102). Each of the *M. genitalium* segments also contains deletions relative to *M. pneumoniae* (after Reference 102). *B.* Rearrangements among individuals within the species *Salmonella typhi*. The first five lines represent 5 of the 17 observed *S. typhi* genome arrangements. The *different shades* represent different genomic segments, and the *closed triangles* indicate the locations of the seven rRNA operons. The genomes of three related species of bacteria are shown below. On the right of the *S. typhi* lines, the number of individuals (among 127 analyzed) with that arrangement is given. Strain names are given to the right of the bottom three lines. All the chromosomes are opened for linearization approximately at the replication terminus; the origin of replication is in the *white segment* (after Reference 157).
For a color version of this figure, see the color section at the back of the volume.


A.
Mycoplasma genitalium
Mycoplasma pneumoniae
B.
Salmonella typhi
2
22
57
4
1
Escherichia coli
K12
Salmonella typimurium
LT2
Salmonella paratypi
B

INTRA-SPECIES VARIATION IN GENOME STRUCTURE

Differences in Overall Genome Arrangement

Even more surprising than macrogenomic differences between closely related species are recent findings of substantial genome differences among independent natural isolates of the same species. The best-studied species with this property is *Salmonella typhi*, in which there appear to be substantial ongoing chromosomal rearrangements mediated through homologous recombination among the seven rRNA operons (Figure 4*B*). Liu & Sanderson (157) found 17 of the 21 possible inter-rRNA genome segment orders among 127 independent isolates examined, suggesting that many segment shuffling recombination events have occurred in nature that are not strongly selected against. Curiously, the analyzed members of the several closely related species with very similar genomic content, *S. paratyphi* B (1 isolate), *S. typhimurium* (50 isolates), *S. enteriditis* (1 isolate), and *E. coli*, all have the segment arrangement corresponding to one found only twice among the 127 *S. typhi* isolates studied (157). *Brucella* isolates typically have two chromosomes of about 1200 and 2100 kbp (174), but individual *Brucella suis* isolates have been found that contain one 3250-kbp chromosome or two rearranged chromosomes that could have arisen through recombination events involving rRNA genes (117). Numerous other cases of intra-species, large-scale, genomic variation have also been found. For example, one of five *Mycoplasma hominis* isolates studied has a 300-kbp inversion relative to the others (135), and less well-understood genome rearrangements appear to be frequent in many species including *Bacillus cereus* (32, 33), *Bacillus subtilis* (112), *Helicobacter pylori* (115), *Bordetella pertussis* (230), *Neisseria gonorrhoeae* (88), *Campylobacter jejuni* (184, 240), *Lactococcus lactis* (57, 139, 140), and *Leptospira interrogans* (271). In none of these are the true dynamics of the situation understood. In no case is the rate of rearrangement known or whether particular rearrangements are under selection in the wild.

Accessory Elements

In most cases where sufficient information is available, the genomes of different isolates of the same bacterial species contain multiple insertion/deletion differences, each in the few kbp to 200-kbp range. These are thought to be largely due to integrated accessory elements (e.g. 146), the transposons, integrons, conjugative transposons, retrons, invertrons, prophages, defective prophages, pathogenicity islands, and plasmids. These are not reviewed in detail here, except to note that these and related elements are present in many if not all of the major bacterial phylogenetic branches and thus contribute to the observed intraspecies variability in genome structure. The characteristic of each of these

types of elements relevant to this discussion is that various members of each family of elements may be found integrated into the chromosomes of different individuals, and that any given element may or may not be present in a particular natural isolate's chromosome. The evolutionary reasons for keeping genes on accessory elements is thought to involve the advantages of genetic mobility, and these have been previously discussed (25, 44, 146).

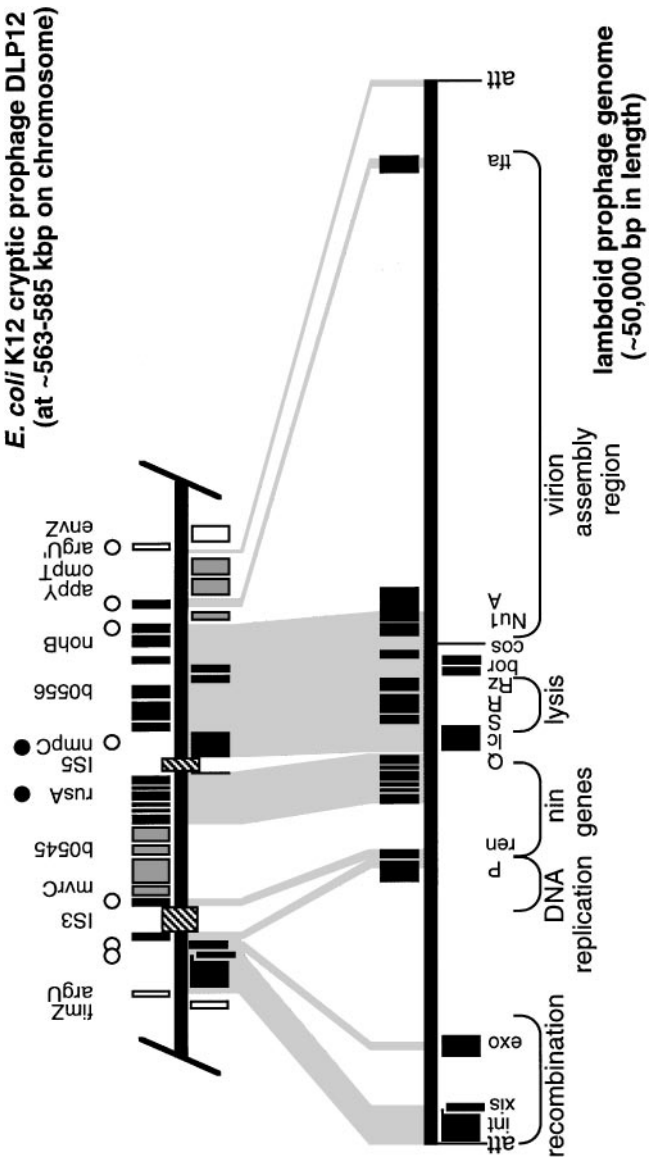
Thus, the genome of each bacterial species can be thought of as being composed of a universally present core of genes that carries with it in each individual a smattering of accessory elements that can be free replicons or be integrated into the chromosome at various places. This universally present core of genetic material is referred to here as the *endo*genome or *endo*chromosome; the accessory elements, both free and integrated, are called the *exo*genome (from the prefixes *endo-*, inside or at the core, and *exo-*, outside or extra). The individual accessory elements can be functional or nonfunctional (in a state of evolutionary decay) and apparently may be selfish (insertion sequence) or be of great value to the bacterium under particular circumstances (pathogenicity island, integron with drug-resistance genes), or be a combination of the two (prophage that carries a bacterial virulence gene). Many of these elements can be recognized in nucleotide sequence by the types of genes they carry, e.g. transposase in transposons, known virulence genes in pathogenicity islands, virus structural genes in prophages, etc. However, some accessory element gene homologs can exist outside of accessory elements, and the variability in these types of genes is large enough that outliers, nonorthologous analogs, or previously uncharacterized accessory element genes would be missed. Thus, at present, when sequence information often substantially precedes functional analysis, it is not always possible to recognize integrated accessory elements from pure sequence information from one individual organism.

Recent genome comparisons and sequencing efforts have shed new light particularly on the nature of defective prophages and pathogenicity islands. Next to transposons with their unique and recognizable transposase genes, prophages may be the easiest to recognize. The bacterial genomes whose sequences have been completed contain numerous recognizable prophages, both apparently intact and defective, as well as other accessory elements: *H. influenzae* Rd contains at least one possible functional prophage (75) and a probable defective prophage (R Hendrix & G Hatfull, personal communication); *E. coli* K12 contains one intact prophage (λ), at least eight apparently defective prophages, and 42 insertion sequences or parts thereof (17); *B. subtilis* contains three known defective prophages and seven additional A+T rich regions that could be integrated accessory elements (133). The *Helicobacter*, *Synechococcus*, *Mycoplasma*, and *Borrelia* completed genomes do not have currently recognizable prophages, but *Helicobacter* has at least one putative pathogenicity island.

In these four cases, the temperate bacteriophages that might naturally infect them have not been characterized, and bacteriophage diversity is extremely great (e.g. 39), so they could easily have been missed; these four genomes harbor 9, 99, 0, and 1 transposase genes, respectively, that indicate the probable presence of that number of transposable elements. (The *Borrelia* genome carries about a dozen transposase genes or fragments thereof, none of which appears to be intact. The least damaged appears to have one frameshift mutation in its coding region (80).) So far, the ratio of inserted accessory elements to endogenome seems to be smaller in small-genome bacteria than in larger-genome bacteria. Perhaps the process of minimizing their genome size has resulted in removal of integrated accessory elements?

Integrated accessory elements usually carry with them genes that encode the enzymatic machinery required for mobility, but if random mutation inactivates that machinery, the element may enter a state of immobile stability or decay in which genes that are not useful to the host are slowly inactivated and lost by stochastic nucleotide changes and deletion. Genes useful to the host could be kept and even modified to better suit the host's needs. Defective prophages are thought to be integrated bacteriophage DNAs undergoing this process. The complete sequence of *E. coli* combined with the large body of knowledge concerning the lambda-like bacteriophages provide a relatively good understanding of its lambdoid defective prophages. As an example, defective prophage DLP12 (150) is compared to the bacteriophage genome in Figure 5. A typical lambdoid phage genome contains about 60 genes, and many of them, especially the virion assembly and control genes, have been deleted from DLP12, leaving about 30

Figure 5 *E. coli* cryptic prophage DLP12. *E. coli* strain K12 prophage DPL12 is shown above, and a lambdoid bacteriophage genome is shown below; connecting *gray areas* indicate homologies between the two DNAs. In each map genes are indicated by boxes; those above each map are transcribed rightward and below leftward; selected gene names are shown above and below the two maps. *Black and gray boxes* indicate all the open reading frames in DLP12 and selected genes on the lambdoid phage genome; *white boxes* indicate endochromosomal genes outside of DLP12. Only genes with homologs in DLP12 are shown on the lambdoid chromosome. *Cross-hatched boxes* on the DLP12 map indicate insertion sequences. The lambdoid phage shown is a composite of several closely related phages, but shows all the genes in their correct positions. The DLP12 genes with known phage homologs have different closest known relatives as follows: *int* and *xis*-P22; *P* and *ren*-λ; nin region genes −82; *Q*-21; *lc*-PA-2; lysis and virion assembly genes −λ. *Black circles* indicate genes that are naturally poorly expressed and require a mutation or IS excision to be expressed. *White circles* indicate genes that are obviously truncated or disrupted compared to their phage counterparts. Since analyses of lambdoid phage genes have not yet reached saturation (39), the gray genes in DLP12 might in fact have been a part of the prophage genome that originally occupied this site. Curiously, one, orf b0545, is very similar to *E. coli* orf yblB, which is immediately adjacent to the phage λ attachment site.


E. coli K12 cryptic prophage DLP12
(at ~563-585 kbp on chromosome)
lambdoid prophage genome
(~50,000 bp in length)
fimZ
argU
IS3
mvrC
b0545
rusA
IS5
nmpC
b0556
nohB
appY
ompT
argU'
envZ
att
int
xis
exo
recombination
P
ren
DNA
replication
nin
genes
cI
cro
R
Z
S
lysis
bor
cos
Nu1
A
virion
assembly
region
tfa
att

open reading frames. Of these genes, two have been apparently damaged by IS insertion and five by deletion; five genes are known to be potentially functional (if one selects for excision of the IS5 from the *nmpC* gene) by virtue of various genetic analyses. The status of the remaining genes is unknown, but one is very similar to the lambda *bor* gene, a possible *E. coli* virulence gene (8). Little is known about the rates of decay of such "dead" elements. The presence of DLP12-like elements at the same location in other *E. coli* isolates has not been investigated, but insertions at the attachment site of lambda-like phage 21 have. Of 78 isolates examined, 3 carry the defective prophage e14 or a relative, and 30 have phage 21-like prophages integrated at that site (at least some appear to be defective) (253).

Pathogenicity islands have been defined as regions of the chromosome in pathogenic bacteria that contain clusters of virulence factor encoding (host interaction) genes such as toxins, pili, and host cell adsorption and invasion factors (95, 96). They also have one or more of the following properties: association with mobile DNA factors (e.g. IS sequences, integrase genes, temperate phage attachment sites), less than universal distribution in natural isolates, unusual codon usage or G+C content, and the ability to occasionally undergo precise excision. They can be of any size; the largest known to date is 190 kbp (96). Because of their sporadic presence and apparent mobility, they contribute significantly to the variability in genome content in many pathogens. The distributions of four different chromosomal virulence determinants were studied by Boyd & Hartl (21), and they found that the four sequences tested were present in 10, 11, 28, and 28 of 72 *E. coli* isolates examined. Although much less is known in most other bacterial phyla, pathogenicity islands likely represent one aspect of a more universal phenomenon that we might call specialization islands, which confer particular metabolic capabilities, defenses, etc, and so allow a fraction of the members of a bacterial species to occupy very specialized niches.

The accessory elements comprise a group of genetic elements that may or may not be integrated and whose definitions can overlap and are not yet always precise. These definitions are evolving, and as new paradigms emerge, they will be no doubt modified, but in the final analysis, specific functional information will be required in all cases to understand their particular properties. Nonetheless, given the significant ranges in chromosome size within species in many bacterial phylogenetic branches, most species will likely have within them at least some integrated genetic elements that fall into this category. Figure 6 diagrammatically shows these elements in a generic bacterial genome.

Systematic Structural Variation at the Gene to Nucleotide Level

Different bacterial genomes have characteristic codon usage, G+C content (ranging from about 75% to 25%), GC strand bias, nearest neighbor frequencies,

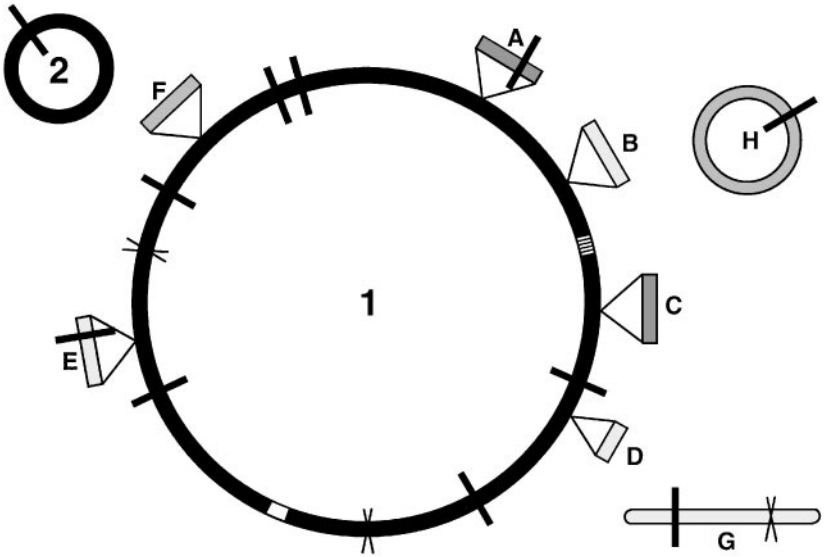


Figure 6 A generic dynamic bacterial genome. According to our current view, a generic bacterial genome is primarily made up of an endogenome consisting of one or more endochromosomes, shown in black (1 & 2). Components of the endogenome can be linear or circular and may or may not be undergoing segment shuffling or inversion. Other systematic physical changes (e.g. white invertable section and slipped strand mispairing changes indicated by Xs) may be occurring at a noticeable frequency, but the sequences in the endogenome are present in all healthy individuals (except for nonorthologous replacements, in *crosshatch*). In addition to this core of genetic material common to all members of a species, a number of accessory elements may or may not be present (*A–H*). These include multiple types of plasmids that may exist as free replicons (*H*) or integrated (*F*), prophages or fragments thereof that may be integrated or in free plasmid form (*B, D, E, G*), and pathogenicity islands (*A, C*). Finally, smaller mobile elements of various kinds (transposons, integrons, etc) can be inserted into any of the above genome components (*thick black lines*).

and oligonucleotide frequencies, and although their mechanistic origins are not always entirely clear, these characteristics likely evolve slowly enough to be of use in attempting to decipher evolutionary histories of horizontally transferred DNA regions (see 137, 138). Neither these nor nucleotide modifications or random sequence polymorphisms are discussed further here, since they are not systematic changes.

Nonetheless, there are programmed smaller-scale structural genome differences in bacterial genomes that (*a*) can occur randomly in time or (*b*) are built into particular responses. Both will contribute to individual genome differences within the affected species. The former cause semistable, reversible phase variations in phenotype, in which members of a population spontaneously change properties (due to switches between gene expression states), with a

small probability at each cell division. Although these can have a number of alternate states, they are reversible with low probability. The known response-programmed genome changes occur irreversibly in particular cells that are destined not to give rise to progeny (the latter are analogous in a sense to somatic cells of multicellular organisms). In addition, transient amplifications of particular or random regions of the genome occur in bacteria and can be stabilized by selection (161, 250), but it is unclear whether they should be considered systematic changes (i.e. programmed to occur in a particular way).

There are two characterized examples of somatic cell analogs in which dead-end genome rearrangements are known to occur: (*a*) the mother cell chromosome in *B. subtilis* spore formation and (*b*) the chromosome of the heterocyst in some cyanobacteria. In both cases the rearranged chromosome is not destined to be replicated, and a specific rearrangement causes important changes in gene structure and expression (reviewed in 98).

The first examples of reversible systematic structural changes in (nondead-end) bacterial DNAs were the invertase-catalyzed inversion of a promoter region in *Salmonella* that switches the orientation of a promoter so that two alternate flagellin genes are expressed and inversions that exchange two tail-fiber protein domains in phage Mu and P1 lysogens (summarized in 197). Similar systems have subsequently been identified in other bacteria (e.g. 21, 67–69), suggesting that it is a common mechanism for allowing a slow alternating between two gene expression states. Some of these may be catalyzed by homologous recombination rather than an invertase (66), and overlapping inversions can cause more complex rearrangements (67, 68).

In addition, there are cassette mechanisms in which DNA sequence information from unexpressed pseudogenes is physically placed in an expression site. These can be either distributive, where the pseudogenes do not appear to be clustered in one location, or organized, in which the pseudogenes are tandemly arrayed in one or a few locations. The only organized cassette mechanisms known in bacteria are found in *Borrelia burgdorferi* (268) and *B. hermsii* (202), where outer-surface proteins are changed using information brought into the expression site from the pseudogenes. In *B. burgdorferi*, the causative agent of Lyme disease, the system has a single plasmid-borne expression site and 15 different, silent, tandem pseudogenes near the expression site. *Neisseria gonorrhoeae* pilin expression site information appears to be changed through a distributive cassette mechanism. Information is imported from a number of silent (sometimes partial) pilin genes scattered about the chromosome (129).

Yet another mechanism by which two individuals from the same bacterial species can differ at the gene level is by nonorthologous replacement of genes by nonhomologous genes of similar function. The best-studied examples of this type of gene replacement are in the temperate bacteriophage genomes (see

39), but it can also happen in the bacterial endochromosome. For example, different individual *Salmonella enterica* isolates can have alternate, nonhomologous endochromosomal genes (at a particular location) that encode different enzymes involved in the synthesis of its surface polysaccharide (262). In this case, there may be selective pressure to diversify the structure of the polysaccharide, and so horizontal transfer of endochromosal information can be seen; normally it may be infrequent enough to make it unlikely, in the absence of such diversifying selection, to catch such an event before either loss or fixation occurs within the species.

Phase variations in phenotype can also be caused by mechanisms that vary the number of repeats in a repeated sequence (or length of a run of the same nucleotide) in which the genes or their control regions are changed so that the reading frame or the transcription of a gene is changed (68). Such structural changes cause switches between on and off states for the expression of the gene. Examples of this type of phase variation are changes in the number of tandem CTTCTs in the leader peptide of the *opa* surface protein genes in *Neisseria gonorrhoeae* (228) and variation in the length of a run of Cs in the promoter of the *Bordetella pertussis fim* gene (259). These changes are thought to occur randomly with a small probability each generation through slipped-strand mispairing during replication or unequal homologous recombination.

There are also unexpressed cryptic genes in bacteria that can be activated by mutation. Such changes could be considered systematic, but are these genes that are in the early stages of decay (resulting from disuse) that can still be reactivated under selection, or are they genes that are being somehow kept in an inactive state in case they are needed? It is difficult to imagine how the latter state could be maintained, since there should be no selection maintaining the functionality of dormant genes. Genes of this type are now known to be located on defective prophages [e.g. *recE* and *rusA* (118, 163)] as well as in apparently endochromosomal regions (203) in *E. coli*.

SUMMARY

All these phenomena, and no doubt additional ones that we are not aware of yet, contribute to bacterial interspecies genome plasticity and to the individuality of members of a given bacterial species. However, beyond this, it is difficult if not foolish to draw universal conclusions about genome structure in a kingdom as diverse as the bacteria. The genomes of bacteria are often (usually?) very fluid on an evolutionary time scale, both in terms of gene content and gene order, such as the rapidly re-ordering chromosomal segments of *Salmonella typhi* and the genus *Mycoplasma*. Some bacterial chromosomes seem stable in overall gene content and order (e.g. some of the enterobacteria and members

of the genus *Borrelia*), but even these may carry a wide variety of plasmids, prophages, and pathogenicity islands, etc.

Bacterial genome structure is much more dynamic and diverse than was originally expected, and so, to understand this aspect of bacterial biology, many independently isolated individuals within a species must compared in some detail before reliable conclusions regarding genome structure can be drawn. Simply making a physical map (or even determining the complete nucleotide sequence) of the genome of one individual from a species tells us little or nothing of possible genomic plasticity that may be present, and conclusions drawn from lone maps regarding overall genome structure in related species or even all members of the same species will often be misleading.

At present, only a narrow view exists of the overall picture of bacterial genome structural diversity and fluidity, but the coming complete sequences of more bacterial genomes should allow more informative comparisons to be made, but perhaps more important in the present context, they will provide the detailed standards necessary for comparison with other natural isolates from the phylogenetic groups in which they reside. In the coming decade, the anticipated rapid expansion of knowledge of bacterial genome structure should open whole fields of inquiry as yet only dimly perceived as questions.

ACKNOWLEDGMENTS

I thank Wai Mun Huang, Roger Hendrix, Patti Rosa, Costa Georgopoulos, John Roth, and Claire Fraser for stimulating my interest in this area of biological science, and Owen White for supplying the data for Figure 3.

Literature Cited

1. Allardet-Servent A, Michaux-Charachon S, Jumas-Bilak E, Karayan L, Ramuz M. 1993. Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. *J. Bacteriol.* 175:7869–74
2. Allison C, Hughes C. 1991. Closely linked genetic loci required for swarm cell differentiation and multicellular migration by *Proteus mirabilis. Mol. Microbiol.* 5:1975–82
3. Bancroft I, Wolk CP, Oren EV. 1989. Physical and genetic maps of the genome of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* 171:5940–48
4. Bannantine JP, Pattee PA. 1996. Construction of a chromosome map for the phage group II *Staphylococcus aureus* Ps55. *J. Bacteriol.* 178:6842–48
5. Barbour AG. 1988. Plasmid analysis of *Borrelia burgdorferi*, the Lyme disease agent. *J. Clin. Microbiol.* 26:475–78
6. Barbour AG. 1993. Linear DNA of *Borrelia* species and antigenic variation. *Trends Microbiol.* 1:236–39
7. Baril C, Herrmann JL, Richaud C, Margarita D, Girons IS. 1992. Scattering of the rRNA genes on the physical map of the circular chromosome of *Leptospira interrogans* serovar *icterohaemorrhagiae. J. Bacteriol.* 174:7566–71

8. Barondess JJ, Beckwith J. 1995. bor gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. *J. Bacteriol.* 177:1247–53
9. Bautsch W. 1988. Rapid physical mapping of the *Mycoplasma mobile* genome by two-dimensional field inversion gel electrophoresis techniques. *Nucleic Acids Res.* 16:11461–67
10. Bautsch W. 1993. A NheI macrorestriction map of the *Neisseria meningitidis* B1940 genome. *FEMS Microbiol. Lett.* 107:191–97
11. Berger F, Fischer G, Kyriacou A, Decaris B, Leblond P. 1996. Mapping of the ribosomal operons on the linear chromosomal DNA of *Streptomyces ambofaciens* DSM40697. *FEMS Microbiol. Lett.* 143:167–73
12. Bergthorsson U, Ochman H. 1995. Heterogeneity of genome sizes among natural isolates of *Escherichia coli*. *J. Bacteriol.* 177:5784–89
13. Biderre C, Pages M, Metenier G, David D, Bata J, et al. 1994. On small genomes in eukaryotic organisms: molecular karyotypes of two microsporidian species. *C. R. Acad. Sci.* 317:399–404
14. Bigey F, Janbon G, Arnaud A, Galzy P. 1995. Sizing of the *Rhodococcus* sp. R312 genome by pulsed-field gel electrophoresis. Localization of genes involved in nitrile degradation. *Antonie Van Leeuwenhoek* 68:173–79
15. Bihimaier A, Romling U, Meyer TF, Tummler B, Gibbs CP. 1991. Physical and genetic map of the *Neisseria gonorrhoeae* strain MS11–N198 chromosome. *Mol. Microbiol.* 5:2529–39
16. Birkelund S, Stephens RS. 1992. Construction of physical and genetic maps of *Chlamydia trachomatis* serovar L2 by pulsed-field gel electrophoresis. *J. Bacteriol.* 174:2742–47
17. Blattner FR, Plunkett GR, Bloch CA, Perna NT, Burland V, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–74
18. Bolstad AI. 1994. Sizing the *Fusobacterium nucleatum* genome by pulsed-field gel electrophoresis. *FEMS Microbiol. Lett.* 123:145–51
19. Borges KM, Bergquist PL. 1993. Genomic restriction map of the extremely thermophilic bacterium *Thermus thermophilus* HB8. *J. Bacteriol.* 175:103–10
20. Bourke B, Sherman P, Louie H, Hani E, Islur P, Chan VL. 1995. Physical and genetic map of the genome of *Campylobacter upsaliensis*. *Microbiology* 141:2417–24
21. Boyd EF, Hartl DL. 1998. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J. Bacteriol.* 180:1159–65
22. Brown JR, Doolittle WF. 1997. Archea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* 61: 1092–2172
23. Bukanov NO, Berg DE. 1994. Ordered cosmid library and high-resolution physical-genetic map of *Helicobacter pylori* strain NCTC11638. *Mol. Microbiol.* 11:509–23
24. Butler PD, Moxon ER. 1990. A physical map of the genome of *Haemophilus influenzae* type b. *J. Gen. Microbiol.* 136: 2333–42
25. Campbell A. 1981. Evolutionary significance of accessory DNA elements in bacteria. *Annu. Rev. Microbiol.* 35:55–83
26. Campbell AM. 1993. Genome organization in prokaryotes. *Curr. Opin. Genet. Dev.* 3:837–44
27. Canard B, Cole ST. 1989. Genome organization of the anaerobic pathogen *Clostridium perfringens*. *Proc. Natl. Acad. Sci. USA* 86:6676–80
28. Canard B, Saint-Joanis B, Cole ST. 1992. Genomic diversity and organization of virulence genes in the pathogenic anaerobe *Clostridium perfringens*. *Mol. Microbiol.* 6:1421–29
29. Carle P, Laigret F, Tully JG, Bove JM. 1995. Heterogeneity of genome sizes within the genus *Spiroplasma*. *Int. J. Syst. Bacteriol.* 45:178–81
30. Carlson C, Gronstad A, Kolsto A. 1992. Physical maps of the genomes of three *Bacillus cereus* strains. *J. Bacteriol.* 174:3750–56
31. Carlson C, Kolsto A. 1993. A complete physical map of a *Bacillus thuringiensis* chromosome. *J. Bacteriol.* 175:1053–60
32. Carlson C, Kolsto A. 1994. A small (2.4 Mb) *Bacillus cereus* chromosome corresponds to a conserved region of a larger (5.3 Mb) *Bacillus cereus* chromosome. *Mol. Microbiol.* 13:161–69
33. Carlson CR, Gronstad A, Kolsto AB. 1992. Physical maps of the genomes of three *Bacillus cereus* strains. *J. Bacteriol.* 174:3750–56
34. Carlson CR, Johansen T, Kolsto AB. 1996. The chromosome map of *Bacillus thuringiensis* subsp. *canadensis* HD224 is highly similar to that of the *Bacillus*

cereus type strain ATCC 14579. *FEMS Microbiol. Lett.* 141:163–77
35. Carlson CR, Johansen T, Lecadet M, Kosto A. 1996. Genomic organization of the entomopathogenic bacterium *Bacillus thuringiensis* subsp. *berliner* 1715. *Microbiology* 142:1625–34
36. Carlson CR, Kolsto AB. 1993. A complete physical map of a *Bacillus thuringiensis* chromosome. *J. Bacteriol.* 175:1053–60
37. Carriere C, Allardet-Servent A, Bourg G, Audurier A, Ramuz M. 1991. DNA polymorphism in strains of *Listeria monocytogenes*. *J. Clin. Microbiol.* 29: 1351–55
38. Casjens S, Delange M, Ley HL, Rosa P, Huang WM. 1995. Linear chromosomes of Lyme disease agent spirochetes: genetic diversity and conservation of gene order. *J. Bacteriol.* 177:2769–80
39. Casjens S, Hatfull G, Hendrix R. 1992. Evolution of dsDNA tailed-bacteriophage genomes. *Semin. Virol.* 3:383–97
40. Casjens S, Hendrix R. 1974. Comments on the arrangement of the morphogenetic genes of bacteriophage lambda. *J. Mol. Biol.* 90:20–25
41. Casjens S, Huang W. 1993. Linear chromosomal physical and genetic map of *Borrelia burgdorferi*, the Lyme disease agent. *Mol. Microbiol.* 8:967–80
42. Casjens S, Murphy M, DeLange M, Sampson L, van Vugt R, Huang WM. 1997. Telomeres of the linear chromosomes of Lyme disease spirochaetes: nucleotide sequence and possible exchange with linear plasmid telomeres. *Mol. Microbiol.* 26:581–96
43. Chang N, Taylor DE. 1990. Use of pulsed-field agarose gel electrophoresis to size genomes of *Campylobacter* species and to construct a SalI map of *Campylobacter jejuni* UA580. *J. Bacteriol.* 172:5211–17
44. Cheetham F, Katz M. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol. Microbiol.* 18:201–8
45. Chen CW. 1996. Complications and implications of linear bacterial chromosomes. *Trends Genet.* 12:192–96
46. Chen X, Widger WR. 1993. Physical genome map of the unicellular cyanobacterium *Synechococcus* sp. strain PCC 7002. *J. Bacteriol.* 175:5106– 16
47. Cheng HP, Lessie TG. 1994. Multiple replicons constituting the genome of *Pseudomonas cepacia* 17616. *J. Bacteriol.* 176:4034–42
48. Choudhary M, Mackenzie C, Nereng K, Sodergren E, Weinstock G, Kaplan G. 1994. Multiple chromosomes in bacteria: structure and function of chromosome II of *Rhodobacter sphaeroides* $2.4.1^T$. *J. Bacteriol.* 176:7694–702
49. Choudhary M, Mackenzie C, Nereng K, Sodergren E, Weinstock G, Kaplan S. 1997. Low-resolution sequencing of *Rhodobacter sphaeroides* $2.4.1^T$: Chromosome II is a true chromosome. *Microbiology* 143:3085–99
50. Claus H, Rotlish H, Filip A. 1992. DNA fingerprints of *Pseudomonas* spp. using rotating field electrophoresis. *Microb. Releases* 1:11–16
51. Cocks BG, Pyle LE, Finch LR. 1989. A physical map of the genome of *Ureaplasma urealyticum* 960T with ribosomal RNA loci. *Nucleic Acids Res.* 17:6713–19
52. Cole ST, Saint Girons I. 1994. Bacterial genomics. *FEMS Microbiol. Rev.* 14:139–60
53. Correia A, Martin JF, Castro JM. 1994. Pulsed-field gel electrophoresis analysis of the genome of amino acid producing corynebacteria: chromosome sizes and diversity of restriction patterns. *Microbiology* 140:2841–47
54. Crespi M, Messens E, Caplan AB, van Montagu M, Desomer J. 1992. Fasciation induction by the phytopathogen *Rhodococcus fascians* depends upon a linear plasmid encoding a cytokinin synthase gene. *EMBO J.* 11:795–804
55. Daly MJ, Minton KW. 1995. Interchromosomal recombination in the extremely radioresistant bacterium *Deinococcus radiodurans*. *J. Bacteriol.* 177:5495–505
56. Daniel P. 1995. Sizing the *Lactobacillus plantarum* genome and other lactic acid bacterial species by transverse alternating field electrophoresis. *Curr. Microbiol.* 30:243–46
57. Davidson BE, Kordias N, Baseggio N, Lim A, Dobos M, Hillier AJ. 1995. Genomic organization of Lactococci. *Dev. Biol. Stand.* 85:411–22
58. Davidson BE, Kordias N, Dobos M, Hillier AJ. 1996. Genomic organization of lactic acid bacteria. *Antonie Van Leeuwenhoek* 70:161–83
59. Davidson BE, MacDougall J, Saint Girons I. 1992. Physical map of the linear chromosome of the bacterium *Borrelia burgdorferi* 212, a causative agent of Lyme disease, and localization of rRNA genes. *J. Bacteriol.* 174:3766–74

60. de Bruijn F, Lupski J, Weinstock G, eds. 1998. *Bacterial Genomes: Physical Structure and Analysis*. New York: Chapman & Hall. 793 pp.
61. De Ita M, Marsch-Moreno R, Guzman P, Alvarez-Morales A. 1998. Physical map of the chromosome of the phytophathogenic bacterium *Pseudomonas syringae* pv. *phaseolicola*. *Microbiology* 144:493–501
62. Dempsey JA, Cannon JG. 1994. Locations of genetic markers on the physical map of the chromosome of *Neisseria gonorrhoeae* FA1090. *J. Bacteriol.* 176:2055–60
63. Dempsey JA, Litaker W, Madhure A, Snodgrass TL, Cannon JG. 1991. Physical map of the chromosome of *Neisseria gonorrhoeae* FA1090 with locations of genetic markers, including *opa* and *pil* genes. *J. Bacteriol.* 173:5476–86
64. Dempsey JA, Wallace AB, Cannon JG. 1995. The physical map of the chromosome of a serogroup A strain of *Neisseria meningitidis* shows complex rearrangements relative to the chromosomes of the two mapped strains of the closely related species *N. gonorrhoeae*. *J. Bacteriol.* 177:6390–400
65. Devereux R, Willis SG, Hines ME. 1997. Genome sizes of *Desulfovibrio desulfuricans*, *Desulfovibrio vulgaris*, and *Desulfobulbus propionicus* estimated by pulsed-field gel electrophoresis of linearized chromosomal DNA. *Curr. Microbiol.* 34:337–39
66. Dworkin J, Blaser MJ. 1997. Molecular mechanisms of *Campylobacter fetus* surface layer protein expression. *Mol. Microbiol.* 26:433–40
67. Dworkin J, Blaser MJ. 1997. Nested DNA inversion as a paradigm of programmed gene rearrangement. *Proc. Natl. Acad. Sci. USA* 94:985–90
68. Dybvig K. 1993. DNA rearrangements and phenotypic switching in prokaryotes. *Mol. Microbiol.* 10:465–71
69. Dybvig K, Yu H. 1994. Regulation of a restriction and modification system via DNA inversion in *Mycoplasma pulmonis*. *Mol. Microbiol.* 12:547–60
70. Eiglmeier K, Honore N, Woods SA, Caudron B, Cole ST. 1993. Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. *Mol. Microbiol.* 7:197–206
71. Ely B, Gerardot CJ. 1988. Use of pulsed-field-gradient gel electrophoresis to construct a physical map of the *Caulobacter crescentus* genome. *Gene* 68:323–33
72. Ferdows M, Serwer P, Griess G, Norris S, Barbour A. 1996. Conversion of a linear to a circular plasmid in the relapsing fever agent *Borrelia hermsii*. *J. Bacteriol.* 178:793–800
73. Firrao G, Smart CD, Kirkpatrick BC. 1996. Physical map of the Western X-disease phytoplasma chromosome. *J. Bacteriol.* 178:3985–88
74. Fisher R. 1930. *The Genetical Theory of Natural Selection.* Oxford: Clarendon
75. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512
76. Fonstein M, Haselkorn R. 1993. Chromosomal structure of *Rhodobacter capsulatus* strain SB1003: cosmid encyclopedia and high-resolution physical and genetic map. *Proc. Natl. Acad. Sci. USA* 90:2522–26
77. Fonstein M, Haselkorn R. 1995. Physical mapping of bacterial genomes. *J. Bacteriol.* 177:3361–69
78. Fonstein M, Koshy EG, Nikolskaya T, Mourachov P, Haselkorn R. 1995. Refinement of the high-resolution physical and genetic map of *Rhodobacter capsulatus* and genome surveys using blots of the cosmid encyclopedia. *EMBO J.* 14:1827–41
79. Fonstein M, Zheng S, Haselkorn R. 1992. Physical map of the genome of *Rhodobacter capsulatus* SB 1003. *J. Bacteriol.* 174:4070–77
80. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, et al. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390:580–86
81. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, et al. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270:397–403
82. Frutos R, Pages M, Bellis M, Roizes G, Bergoin M. 1989. Pulsed-field gel electrophoresis determination of the genome size of obligate intracellular bacteria belonging to the genera *Chlamydia*, *Rickettsiella*, and *Porochlamydia*. *J. Bacteriol.* 171:4511–13
83. Fujita M, Amako K. 1994. Localization of the sapA gene on a physical map of *Campylobacter fetus* chromosomal DNA. *Arch. Microbiol.* 162:375–80
84. Furihata K, Sato K, Matsumoto H. 1995. Construction of a combined NotI/SmaI physical and genetic map of *Moraxella (Branhamella) catarrhalis*

strain ATCC25238. *Microbiol. Immunol.* 39:745–51
85. Gaher M, Einsiedler K, Crass T, Bautsch W. 1996. A physical and genetic map of *Neisseria meningitidis* B1940. *Mol. Microbiol.* 19:249–59
86. Gaju N, Pavon V, Marin I, Esteve I, Guerrero R, Amils R. 1996. Chromosome map of the phototrophic bacterium *Chromatium vinosum*. *FEMS Microbiol. Lett.* 126:241–48
87. Gasc AM, Kauc L, Barraille P, Sicard M, Goodgal S. 1991. Gene localization, size, and physical map of the chromosome of *Streptococcus pneumoniae*. *J. Bacteriol.* 173:7361–67
88. Gibbs C, Meyer T. 1996. Genome plasticity in *Neisseria gonorrhoeae*. *FEMS Microbiol. Lett.* 145:173–79
89. Gilson PR, McFadden GI. 1997. Good things in small packages: the tiny genomes of chlorarachniophyte endosymbionts. *BioEssays* 19:167–73
90. Ginard M, Lalucat J, Tummler B, Romling U. 1997. Genome organization of *Pseudomonas stutzeri* and resulting taxonomic and evolutionary considerations. *Int. J. Syst. Bacteriol.* 47:132–43
91. Golding GB, Gupta RS. 1995. Protein-based phylogenies support a chimeric origin for the eukaryotic genome. *Mol. Biol. Evol.* 12:1–6
92. Gorton TS, Goh MS, Geary SJ. 1995. Physical mapping of the *Mycoplasma gallisepticum* S6 genome with localization of selected genes. *J. Bacteriol.* 177:259–63
93. Gralton EM, Campbell AL, Neidle EL. 1997. Directed introduction of DNA cleavage sites to produce a high-resolution genetic and physical map of the *Acinetobacter* sp. strain ADP1 (BD413UE) chromosome. *Microbiology* 143:1345–57
94. Grimsley JK, Masters CI, Clark EP, Minton KW. 1991. Analysis by pulsed-field gel electrophoresis of DNA double-strand breakage and repair in *Deinococcus radiodurans* and a radiosensitive mutant. *Int. J. Radiat. Biol.* 60:613–26
95. Groisman EA, Ochman H. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* 87:791–94
96. Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* 23:1089–97
97. Hantman MJ, Sun S, Piggot PJ, Daneo-Moore L. 1993. Chromosome organization of *Streptococcus mutans* GS-5. *J. Gen. Microbiol.* 139:67–77
98. Haselkorn R. 1992. Developmentally regulated gene rearrangements in prokaryotes. *Annu. Rev. Genet.* 26:113– 30
99. He Q, Chen H, Kuspa A, Cheng Y, Kaiser D, Shimkets LJ. 1994. A physical map of the *Myxococcus xanthus* chromosome. *Proc. Natl. Acad. Sci. USA* 91:9584–87
100. He W, Luchansky JB. 1997. Construction of the temperature-sensitive vectors pLUCH80 and pLUCH88 for delivery of Tn917::NotI/SmaI and use of these vectors to derive a circular map of *Listeria monocytogenes* Scott A, a serotype 4b isolate. *Appl. Environ. Microbiol.* 63:3480–87
101. Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 24:4420–49
102. Himmelreich R, Plagens H, Hilbert H, Reiner B, Herrmann R. 1997. Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res.* 25:701–12
103. Hinnebusch J, Barbour A. 1991. Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. *J. Bacteriol.* 173:7233–39
104. Hinnebusch J, Bergstrom S, Barbour AG. 1990. Cloning and sequence analysis of linear plasmid telomeres of the bacterium *Borrelia burgdorferi*. *Mol. Microbiol.* 4:811–20
105. Hinnebusch J, Tilly K. 1993. Linear plasmids and chromosomes in bacteria. *Mol. Microbiol.* 10:917–22
106. Hobbs MM, Leonardi MJ, Zaretzky FR, Wang TH, Kawula TH. 1996. Organization of the *Haemophilus ducreyi* 35000 chromosome. *Microbiology* 142:2587–94
107. Holloway BW, Escuadra MD, Morgan AF, Saffery R, Krishnapillai V. 1992. The new approaches to whole genome analysis of bacteria. *FEMS Microbiol. Lett.* 79:101–5
108. Holloway BW, Romling U, Tummler B. 1994. Genomic mapping of *Pseudomonas aeruginosa* PAO. *Microbiology* 140:2907–29
109. Honeycutt RJ, McClelland M, Sobral BW. 1993. Physical map of the genome of *Rhizobium meliloti* 1021. *J. Bacteriol.* 175:6945–52
110. Hugenholtz P, Pitulle C, Hershberger KL, Pace NR. 1998. Novel division level

bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* 180:366–76
111. Irazabal N, Marin I, Amils R. 1997. Genomic organization of the acidophilic chemolithoautotrophic bacterium *Thiobacillus ferrooxidans* ATCC 21834. *J. Bacteriol.* 179:1946–50
112. Itaya M. 1997. Physical map of the *Bacillus subtilis* 166 genome: evidence for the inversion of an approximately 1900 kb continuous DNA segment, the translocation of an approximately 100 kb segment and the duplication of a 5 kb segment. *Microbiology* 143:3723–32
113. Itaya M, Tanaka T. 1991. Complete physical map of the *Bacillus subtilis* 168 chromosome constructed by a gene-directed mutagenesis method. *J. Mol. Biol.* 220:631–48
114. Itaya M, Tanaka T. 1997. Experimental surgery to create subgenomes of *Bacillus subtilis* 168. *Proc. Natl. Acad. Sci. USA* 94:5378–82
115. Jiang Q, Hiratsuka K, Taylor DE. 1996. Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. *Mol. Microbiol.* 20:833–42
116. Jumas-Bilak E, Maugard C, Michaux-Charachon S, Allardet-Servent A, Perrin A, et al. 1995. Study of the organization of the genomes of *Escherichia coli*, *Brucella melitensis* and *Agrobacterium tumefaciens* by insertion of a unique restriction site. *Microbiology* 141:2425–32
117. Jumas-Bilak E, Michaux-Charachon S, Bourg G, O'Callaghan D, Ramuz M. 1998. Differences in chromosome number and genome rearrangements in the genus *Brucella*. *Mol. Microbiol.* 27:99–106
118. Kaiser K, Murray N. 1980. On the nature of *SbcA* mutations in *E. coli* K-12. *Mol. Gen. Genet.* 179:555–63
119. Kakulphimp J, Finch LR, Robertson JA. 1991. Genome sizes of mammalian and avian *Ureaplasmas*. *Int. J. Syst. Bacteriol.* 41:326–27
120. Kaneko T, Matsubayashi T, Sugita M, Sugiura M. 1996. Physical and gene maps of the unicellular cyanobacterium *Synechococcus* sp. strain PCC6301 genome. *Plant Mol. Biol.* 31:193–201
121. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, et al. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* 3:109–36
122. Karlyshev AV, Henderson J, Ketley JM, Wren BW. 1998. An improved physical and genetic map of *Campylobacter jejuni* NCTC 11168 (UA580). *Microbiology* 144:503–8
123. Katayama S, Dupuy B, Garnier T, Cole ST. 1995. Rapid expansion of the physical and genetic map of the chromosome of *Clostridium perfringens* CPN50. *J. Bacteriol.* 177:5680–85
124. Kauc L, Goodgal SH. 1989. The size and a physical map of the chromosome of *Haemophilus parainfluenzae*. *Gene* 83:377–80
125. Kieser HM, Kieser T, Hopwood DA. 1992. A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome. *J. Bacteriol.* 174:5496–507
126. Kim NW, Lombardi R, Bingham H, Hani E, Louie H, et al. 1993. Fine mapping of the three rRNA operons on the updated genomic map of *Campylobacter jejuni* TGH9011 (ATCC 43431). *J. Bacteriol.* 175:7468–70
127. Kitten T, Barbour AG. 1992. The relapsing fever agent *Borrelia hermsii* has multiple copies of its chromosome and linear plasmids. *Genetics* 132:311–24
128. Kolsto AB. 1997. Dynamic bacterial genome organization. *Mol. Microbiol.* 24:241–48
129. Koomey M. 1994. Mechanisms of pilus antigentic variation in *Neisseria gonorrheae*. In *Molecular Genetics of Bacterial Pathogenesis*, ed. V Miller, J Kaper, D Portnoy, R Isberg. Washington, DC: Am. Soc. Microbiol.
130. Krawiec S, Riley M. 1990. Organization of the bacterial chromosome. *Microbiol. Rev.* 54:502–39
131. Krueger CM, Marks KL, Ihler GM. 1995. Physical map of the *Bartonella bacilliformis* genome. *J. Bacteriol.* 177: 7271–74
132. Kundig C, Hennecke H, Gottfert M. 1993. Correlated physical and genetic map of the *Bradyrhizobium japonicum* 110 genome. *J. Bacteriol.* 175:613–22
133. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390:249–56
134. La Fontaine S, Rood JI. 1997. Physical and genetic map of the chromosome of *Dichelobacter nodosus* strain A198. *Gene* 184:291–98

135. Ladefoged SA, Christiansen G. 1992. Physical and genetic mapping of the genomes of five *Mycoplasma hominis* strains by pulsed-field gel electrophoresis. *J. Bacteriol.* 174:2199–207
136. Lamoureux M, Prevost H, Cavin JF, Divies C. 1993. Recognition of *Leuconostoc oenos* strains by the use of DNA restriction profiles. *Appl. Microbiol. Biotechnol.* 39:547–52
137. Lawrence JG, Ochman H. 1997. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44:383–97
138. Lawrence JG, Roth JR. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143:1843–60
139. Le Bourgeois P, Lautier M, Mata M, Ritzenthaler P. 1992. Physical and genetic map of the chromosome of *Lactococcus lactis* subsp. lactis IL1403. *J. Bacteriol.* 174:6752–62
140. Le Bourgeois P, Lautier M, van den Berghe L, Gasson MJ, Ritzenthaler P. 1995. Physical and genetic map of the *Lactococcus lactis* subsp. *cremoris* MG1363 chromosome: comparison with that of *Lactococcus lactis* subsp. *lactis* IL 1403 reveals a large genome inversion. *J. Bacteriol.* 177:2840–50
141. Leblond P, Fischer G, Francou FX, Berger F, Guerineau M, Decaris B. 1996. The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. *Mol. Microbiol.* 19:261–71
142. Leblond P, Francou FX, Simonet JM, Decaris B. 1990. Pulsed-field gel electrophoresis analysis of the genome of *Streptomyces ambofaciens* strains. *FEMS Microbiol. Lett.* 60:79–88
143. Leblond P, Redenbach M, Cullum J. 1993. Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 175:3422–29
144. Lee JJ, Smith HO, Redfield RJ. 1989. Organization of the *Haemophilus influenzae* Rd genome. *J. Bacteriol.* 171:3016–24
145. Lessie TG, Hendrickson W, Manning BD, Devereux R. 1996. Genomic complexity and plasticity of *Burkholderia cepacia*. *FEMS Microbiol. Lett.* 144: 117–28
146. Levin BR. 1993. The accessory genetic elements of bacteria: existence conditions and (co)evolution. *Curr. Opin. Genet. Dev.* 3:849–54
147. Lezhava A, Mizukami T, Kajitani T, Kameoka D, Redenbach M, et al. 1995. Physical map of the linear chromosome of *Streptomyces griseus*. *J. Bacteriol.* 177:6492–98
148. Lin WJ, Johnson EA. 1995. Genome analysis of *Clostridium botulinum* type A by pulsed-field gel electrophoresis. *Appl. Environ. Microbiol.* 61:4441–47
149. Lin YS, Kieser HM, Hopwood DA, Chen CW. 1993. The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Mol. Microbiol.* 10:923–33
150. Lindsey DF, Mullin DA, Walker JR. 1989. Characterization of the cryptic lambdoid prophage DLP12 of *Escherichia coli* and overlap of the DLP12 integrase gene with the tRNA gene argU. *J. Bacteriol.* 171:6197–205
151. Liu B, Alberts BM. 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* 267:1131–37
152. Liu SL, Hessel A, Sanderson KE. 1993. The XbaI-BlnI-CeuI genomic cleavage map of *Salmonella enteritidis* shows an inversion relative to *Salmonella typhimurium* LT2. *Mol. Microbiol.* 10:655–64
153. Liu SL, Hessel A, Sanderson KE. 1993. The XbaI-BlnI-CeuI genomic cleavage map of *Salmonella typhimurium* LT2 determined by double digestion, end labelling, and pulsed-field gel electrophoresis. *J. Bacteriol.* 175:4104–20
154. Liu SL, Sanderson KE. 1992. A physical map of the *Salmonella typhimurium* LT2 genome made by using XbaI analysis. *J. Bacteriol.* 174:1662–72
155. Liu SL, Sanderson KE. 1995. The chromosome of *Salmonella paratyphi* A is inverted by recombination between rrnH and rrnG. *J. Bacteriol.* 177:6585–92
156. Liu SL, Sanderson KE. 1995. Genomic cleavage map of *Salmonella typhi* Ty2. *J. Bacteriol.* 177:5099–107
157. Liu SL, Sanderson KE. 1995. Rearrangements in the genome of the bacterium *Salmonella typhi*. *Proc. Natl. Acad. Sci. USA* 92:1018–22
158. Lortal S, Rouault A, Guezenec S, Gautier M. 1997. *Lactobacillus helveticus*: strain typing and genome size estimation by pulsed field gel electrophoresis. *Curr. Microbiol.* 34:180–85
159. Luchansky JB, Glass KA, Harsono KD, Degnan AJ, Faith NG, et al. 1992. Genomic analysis of *Pediococcus* starter cultures used to control *Listeria monocy-*

togenes in turkey summer sausage. *Appl. Environ. Microbiol.* 58:3053–59
160. Lucier TS, Brubaker RR. 1992. Determination of genome size, macrorestriction pattern polymorphism, and nonpigmentation-specific deletion in *Yersinia pestis* by pulsed-field gel electrophoresis. *J. Bacteriol.* 174:2078–86
161. Lupski J, Roth J, Weinstock G. 1996. Chromosomal duplications in bacteria, fruit flies and humans. *Am. J. Hum. Genet.* 58:21–27
162. MacDougall J, Saint Girons I. 1995. Physical map of the *Treponema denticola* circular chromosome. *J. Bacteriol.* 177:1805–11
163. Mahdi AA, Sharples GJ, Mandal TN, Lloyd RG. 1996. Holliday junction resolvases encoded by homologous *rusA* genes in *Escherichia coli* K-12 and phage 82. *J. Mol. Biol.* 257:561–73
164. Majumder R, Sengupta S, Khetawat G, Bhadra RK, Roychoudhury S, Das J. 1996. Physical map of the genome of *Vibrio cholerae* 569B and localization of genetic markers. *J. Bacteriol.* 178:1105–12
165. Maldonado R, Jimenez J, Casadesus J. 1994. Changes of ploidy during the *Azotobacter vinelandii* growth cycle. *J. Bacteriol.* 176:3911–19
166. Malinin A, Vostrov A, Rybchin V, Sverchevsky A. 1992. Structure of the linear plasmid N15 ends. *Genetika* 29: 257–65 (In Russian)
167. Marconi RT, Casjens S, Munderloh UG, Samuels DS. 1996. Analysis of linear plasmid dimers in *Borrelia burgdorferi* sensu lato isolates: implications concerning the potential mechanism of linear plasmid replication. *J. Bacteriol.* 178:3357–61
168. Margolis N, Hogan D, Tilly K, Rosa PA. 1994. Plasmid location of *Borrelia* purine biosynthesis gene homologs. *J. Bacteriol.* 176:6427–32
169. Marin I, Amils R, Abad JP. 1997. Genomic organization of the metal-mobilizing bacterium *Thiobacillus cuprinus*. *Gene* 187:99–105
170. Martin W, Muller M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature* 392:37–41
171. Mendez-Alvarez S, Pavon V, Esteve I, Guerrero R, Gaju N. 1995. Genomic heterogeneity in *Chlorobium limicola*: chromosomic and plasmidic differences among strains. *FEMS Microbiol. Lett.* 134:279–85
172. Merino E, Babitzke P, Yanofsky C. 1995. trp RNA-binding attenuation protein (TRAP)-trp leader RNA interactions mediate translational as well as transcriptional regulation of the *Bacillus subtilis* trp operon. *J. Bacteriol.* 177:6362–70
173. Michaux S, Paillisson J, Carles-Nurit MJ, Bourg G, Allardet-Servent A, Ramuz M. 1993. Presence of two independent chromosomes in the *Brucella melitensis* 16M genome. *J. Bacteriol.* 175:701–5
174. Michaux-Charachon S, Bourg G, Jumas-Bilak E, Guigue-Talet P, Allardet-Servent A, et al. 1997. Genome structure and phylogeny in the genus *Brucella*. *J. Bacteriol.* 179:3244–49
175. Michel E, Cossart P. 1992. Physical map of the *Listeria monocytogenes* chromosome. *J. Bacteriol.* 174:7098–103
176. Miyata M, Wang L, Fukumura T. 1991. Physical mapping of the *Mycoplasma capricolum* genome. *FEMS Microbiol. Lett.* 63:329–33
177. Murray BF, Singh KV, Ross RP, Heath JD, Dunny GM, Weinstock GM. 1993. Generation of restriction map of *Enterococcus faecalis* OG1 and investigation of growth requirements and regions encoding biosynthetic function. *J. Bacteriol.* 175:5216–23
178. Naterstad K, Kolsto AB, Sirevag R. 1995. Physical map of the genome of the green phototrophic bacterium *Chlorobium tepidum*. *J. Bacteriol.* 177:5480–84
179. Neimark H, Kirkpatrick BC. 1993. Isolation and characterization of full-length chromosomes from non-culturable plant-pathogenic Mycoplasma-like organisms. *Mol. Microbiol.* 7:21–28
180. Neimark HC, Lange CS. 1990. Pulse-field electrophoresis indicates full-length *Mycoplasma* chromosomes range widely in size. *Nucleic Acids Res.* 18: 5443–48
181. Neumann B, Pospiech A, Schairer HU. 1992. Size and stability of the genomes of the myxobacteria *Stigmatella aurantiaca* and *Stigmatella erecta*. *J. Bacteriol.* 174:6307–10
182. Neumann B, Pospiech A, Schairer HU. 1993. A physical and genetic map of the *Stigmatella aurantiaca* DW4/3.1 chromosome. *Mol. Microbiol.* 10:1087–99
183. Newnham E, Chang N, Taylor DE. 1996. Expanded genomic map of *Campylobacter jejuni* UA580 and localization of 23S ribosomal rRNA genes by I-CeuI restriction endonuclease digestion. *FEMS Microbiol. Lett.* 142:223–29

184. Nuijten PJ, Bartels C, Bleumink-Pluym NM, Gaastra W, van der Zeijst BA. 1990. Size and physical map of the *Campylobacter jejuni* chromosome. *Nucleic Acids Res.* 18:6211–14
185. O'Riordan K, Fitzgerald G. 1997. Determination of genetic diversity within the genus *Bifidobacterium* and an estimation of chromosomal size. *FEMS Microbiol. Lett.* 156:259–64
186. Ochman H, Wilson AC. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* 26:74–86
187. Ogasawara N, Yoshikawa H. 1992. Genes and their organization in the replication origin region of the bacterial chromosome. *Mol. Microbiol.* 6:629–34
188. Ogata K, Aminov RI, Nagamine T, Sugiura M, Tajima K, et al. 1997. Construction of a *Fibrobacter succinogenes* genomic map and demonstration of diversity at the genomic level. *Curr. Microbiol.* 35:22–27
189. Ojaimi C, Davidson BE, Saint Girons I, Old IG. 1994. Conservation of gene arrangement and an unusual organization of rRNA genes in the linear chromosomes of the Lyme disease spirochaetes *Borrelia burgdorferi*, *B. garinii* and *B. afzelii*. *Microbiology* 140:2931–40
190. Okada N, Sasakawa C, Tobe T, Talukder KA, Komatsu K, Yoshikawa M. 1991. Construction of a physical map of the chromosome of *Shigella flexneri* 2a and the direct assignment of nine virulence-associated loci identified by Tn5 insertions. *Mol. Microbiol.* 5:2171–80
191. Old IG, Margarita D, Saint Girons I. 1993. Unique genetic arrangement in the *dnaA* region of the *Borrelia burgdorferi* linear chromosome: nucleotide sequence of the *dnaA* gene. *FEMS Microbiol. Lett.* 111:109–14
192. Park JH, Song JC, Kim MH, Lee DS, Kim CH. 1994. Determination of genome size and a preliminary physical map of an extreme alkaliphile, *Micrococcus* sp. Y-1, by pulsed-field gel electrophoresis. *Microbiology* 140:2247–50
193. Perkins JD, Heath JD, Sharma BR, Weinstock GM. 1993. XbaI and BlnI genomic cleavage maps of *Escherichia coli* K-12 strain MG1655 and comparative analysis of other strains. *J. Mol. Biol.* 232:419–45
194. Peterson SN, Lucier T, Heitzman K, Smith EA, Bott KF, et al. 1995. Genetic map of the *Mycoplasma genitalium* chromosome. *J. Bacteriol.* 177:3199–204
195. Philipp WJ, Nair S, Guglielmi G, Lagranderie M, Gicquel B, Cole ST. 1996. Physical mapping of *Mycobacterium bovis* BCG pasteur reveals differences from the genome map of *Mycobacterium tuberculosis* H37Rv and from *M. bovis*. *Microbiology* 142:3135–45
196. Philipp WJ, Poulet S, Eiglmeier K, Pascopella L, Balasubramanian V, et al. 1996. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proc. Natl. Acad. Sci. USA* 93:3132–37
197. Plasterk RH, Brinkman A, van de Putte P. 1983. DNA inversions in the chromosome of *Escherichia coli* and in bacteriophage Mu: relationship to other site-specific recombination systems. *Proc. Natl. Acad. Sci. USA* 80:5355–58
198. Pyle LE, Finch LR. 1988. A physical map of the genome of *Mycoplasma mycoides* subspecies *mycoides* Y with some functional loci. *Nucleic Acids Res.* 16:6027–39
199. Pyle LE, Taylor T, Finch LR. 1990. Genomic maps of some strains within the *Mycoplasma mycoides* cluster. *J. Bacteriol.* 172:7265–68
200. Rainey PB, Bailey MJ. 1996. Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Mol. Microbiol.* 19:521–33
201. Redenbach M, Kieser HM, Denapaite D, Eichner A, Cullum J, et al. 1996. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol. Microbiol.* 21:77–96
202. Restrepo BI, Barbour AG. 1994. Antigen diversity in the bacterium *B. hermsii* through "somatic" mutations in rearranged vmp genes. *Cell* 78:867–76
203. Reynolds AE, Felton J, Wright A. 1981. Insertion of DNA activates the cryptic *bgl* operon in *E. coli* K12. *Nature* 293:625–29
204. Robertson JA, Pyle LE, Stemke GW, Finch LR. 1990. Human ureaplasmas show diverse genome sizes by pulsed-field electrophoresis. *Nucleic Acids Res.* 18:1451–55
205. Rodley PD, Romling U, Tummler B. 1995. A physical genome map of the *Burkholderia cepacia* type strain. *Mol. Microbiol.* 17:57–67
206. Romalde JL, Iteman I, Carniel E. 1991. Use of pulsed field gel electrophoresis to size the chromosome of the bacterial fish

pathogen *Yersinia ruckeri*. *FEMS Microbiol. Lett.* 68:217–25
207. Romling U, Grothues D, Bautsch W, Tummler B. 1989. A physical genome map of *Pseudomonas aeruginosa* PAO. *EMBO J.* 8:4081–89
208. Romling U, Schmidt KD, Tummler B. 1997. Large genome rearrangements discovered by the detailed analysis of 21 *Pseudomonas aeruginosa* clone C isolates found in environment and disease habitats. *J. Mol. Biol.* 271:386–404
209. Romling U, Tummler B. 1991. The impact of two-dimensional pulsed-field gel electrophoresis techniques for the consistent and complete mapping of bacterial genomes: refined physical map of *Pseudomonas aeruginosa* PAO. *Nucleic Acids Res.* 19:3199–206
210. Romling U, Tummler B. 1993. Comparative mapping of the *Pseudomonas aeruginosa* PAO genome with rare- cutter linking clones or two-dimensional pulsed-field gel electrophoresis protocols. *Electrophoresis* 14:283–89
211. Roussel Y, Bourgoin F, Guedon G, Pebay M, Decaris B. 1997. Analysis of the genetic polymorphism between three *Streptococcus thermophilus* strains by comparing their physical and genetic organization. *Microbiology* 143:1335–43
212. Roussel Y, Colmin C, Simonet JM, Decaris B. 1993. Strain characterization, genome size and plasmid content in the *Lactobacillus acidophilus* group. *J. Appl. Bacteriol.* 74:549–56
213. Roux V, Raoult D. 1993. Genotypic identification and phylogenetic analysis of the spotted fever group *rickettsiae* by pulsed-field gel electrophoresis. *J. Bacteriol.* 175:4895–904
214. Sadziene A, Rosa P, Thompson P, Hogan D, Barbour A. 1992. Antibody-resistant mutants of *Borrelia burgdorferi*: in vitro selection and characterization. *J. Exp. Med.* 176:799–809
215. Sakaguchi K. 1990. Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. *Microbiol. Rev.* 54:66–74
216. Salama SM, Garcia MM, Taylor DE. 1992. Differentiation of the subspecies of *Campylobacter fetus* by genomic sizing. *Int. J. Syst. Bacteriol.* 42:446–50
217. Salama SM, Newnham E, Chang N, Taylor DE. 1995. Genome map of *Campylobacter fetus* subsp. *fetus* ATCC 27374. *FEMS Microbiol. Lett.* 132:239–45
218. Sanderson K, Roth J. 1988. Linkage map of *Salmonella typhimurium*, edition VII. *Microbiol. Rev.* 52:485–532
219. Schmidt KD, Tummler B, Romling U. 1996. Comparative genome mapping of *Pseudomonas aeruginosa* PAO with *P. aeruginosa* C, which belongs to a major clone in cystic fibrosis patients and aquatic habitats. *J. Bacteriol.* 178:85–93
220. Schwan T, Burgdorfer W, Garon C. 1988. Changes in infectivity and plasmid profile of the Lyme disease spirochete, *Borrelia burgdorferi*, as a result of in vitro cultivation. *Infect. Immun.* 56:1831–36
221. Deleted in proof
222. Serwer P, Estrada A, Harris RA. 1995. Video light microscopy of 670–kb DNA in a hanging drop: shape of the envelope of DNA. *Biophys. J.* 69:2649–60
223. Shaheduzzaman SM, Akimoto S, Kuwahara T, Kinouchi T, Ohnishi Y. 1997. Genome analysis of *Bacteroides* by pulsed-field gel electrophoresis: chromosome sizes and restriction patterns. *DNA Res.* 4:19–25
224. Shao Z, Mages W, Schmitt R. 1994. A physical map of the hyperthermophilic bacterium *Aquifex pyrophilus* chromosome. *J. Bacteriol.* 176:6776–80
225. Shimkets L. 1998. Structure and sizes of the genomes of the Archaea and Bacteria. See Ref. 60, pp. 5–11
226. Smith CL, Condemine G. 1990. New approaches for physical mapping of small genomes. *J. Bacteriol.* 172:1167–72
227. Sobral BW, Honeycutt RJ, Atherly AG. 1991. The genomes of the family *Rhizobiaceae*: size, stability, and rarely cutting restriction endonucleases. *J. Bacteriol.* 173:704–9
227a. Sonenshein A, Hoch J, Losick R, eds. 1993. *Bacillus subtilis and Other Gram-Positive Bacteria*. Washington, DC: Am. Soc. Microbiol.
228. Stern A, Brown M, Nickel P, Meyer TF. 1986. Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* 47:61–71
229. Stibitz S, Garletts TL. 1992. Derivation of a physical map of the chromosome of *Bordetella pertussis* Tohama I. *J. Bacteriol.* 174:7770–77
230. Stibitz S, Yang MS. 1997. Genomic fluidity of *Bordetella pertussis* assessed by a new method for chromosomal mapping. *J. Bacteriol.* 179:5820–26
231. Stoppel RD, Meyer M, Schlegel HG. 1995. The nickel resistance determinant cloned from the enterobacterium *Klebsiella oxytoca*: conjugational trans-

fer, expression, regulation and DNA homologies to various nickel-resistant bacteria. *Biometals* 8:70–79
232. Sugita M, Sugishita H, Fujishiro T, Tsuboi M, Sugita C, et al. 1997. Organization of a large gene cluster encoding ribosomal proteins in the cyanobacterium *Synechococcus* sp. strain PCC6301: comparison of gene clusters among cyanobacteria, eubacteria and chloroplast genomes. *Gene* 195:73–79
233. Suvorov AN, Ferretti JJ. 1996. Physical and genetic chromosomal map of an M type 1 strain of *Streptococcus pyogenes*. *J. Bacteriol.* 178:5546–49
234. Suwanto A, Kaplan S. 1989. Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: presence of two unique circular chromosomes. *J. Bacteriol.* 171:5850–59
235. Suzuki S, Kita-Tsukamoto K, Fukagawa T. 1994. The 16S rRNA sequence and genome sizing of tributyltin resistant marine bacterium, strain M-1. *Microbios* 77:101–9
236. Svarchevsky AN, Rybchin VN. 1984. Physical mapping of plasmid N15 DNA. *Mol. Genet. Mikrobiol. Virusol.* 10:16–22 (In Russian)
237. Tabata K, Hoshino T. 1996. Mapping of 61 genes on the refined physical map of the chromosome of *Thermus thermophilus* HB27 and comparison of genome organization with that of *T. thermophilus* HB8. *Microbiology* 142:401–10
238. Tatusov RL, Mushegian AR, Bork P, Brown NP, Hayes WS, et al. 1996. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr. Biol.* 6:279–91
239. Taylor DE, Chang N, Taylor NS, Fox JG. 1994. Genome conservation in *Helicobacter mustelae* as determined by pulsed- field gel electrophoresis. *FEMS Microbiol. Lett.* 118:31–36
240. Taylor DE, Eaton M, Chang N, Salama SM. 1992. Construction of a *Helicobacter pylori* genome map and demonstration of diversity at the genome level. *J. Bacteriol.* 174:6800–6
241. Taylor DE, Eaton M, Yan W, Chang N. 1992. Genome maps of *Campylobacter jejuni* and *Campylobacter coli*. *J. Bacteriol.* 174:2332–37
242. Tenreiro R, Santos MA, Paveia H, Vieira G. 1994. Inter-strain relationships among wine leuconostocs and their divergence from other *Leuconostoc* species, as revealed by low frequency restriction fragment analysis of genomic DNA. *J. Appl. Bacteriol.* 77:271–80
243. Tigges E, Minion FC. 1994. Physical map of *Mycoplasma gallisepticum*. *J. Bacteriol.* 176:4157–59
244. Tigges E, Minion FC. 1994. Physical map of the genome of *Acholeplasma oculi* ISM1499 and construction of a Tn4001 derivative for macrorestriction chromosomal mapping. *J. Bacteriol.* 176:1180–83
245. Tilly K, Casjens S, Stevenson B, Bono JL, Samuels DS, et al. 1997. The *Borrelia burgdorferi* circular plasmid cp26: conservation of plasmid structure and targeted inactivation of the ospC gene. *Mol. Microbiol.* 25:361–73
246. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388:539–47
247. Tulloch DL, Finch LR, Hillier AJ, Davidson BE. 1991. Physical map of the chromosome of *Lactococcus lactis* subsp. *lactis* DL11 and localization of six putative rRNA operons. *J. Bacteriol.* 173:2768–75
248. van Wezel GP, Buttner MJ, Vijgenboom E, Bosch L, Hopwood DA, Kieser HM. 1995. Mapping of genes involved in macromolecular synthesis on the chromosome of *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 177:473–76
249. Vary P. 1993. The genetic map of *Bacillus megaterium*. See Ref. 227a, pp. 475–81
250. Volff JN, Altenbuchner J. 1998. Genetic instability of the *Streptomyces* chromosome. *Mol. Microbiol.* 27:239–46
251. Wagner E, Doskar J, Gotz F. 1998. Physical and genetic map of the genome of *Staphylococcus carnosus* TM300. *Microbiology* 144:509–17
252. Walker EM, Howell JK, You Y, Hoffmaster AR, Heath JD, et al. 1995. Physical map of the genome of *Treponema pallidum* subsp. *pallidum* (Nichols). *J. Bacteriol.* 177:1797–804
253. Wang FS, Whittam TS, Selander RK. 1997. Evolutionary genetics of the isocitrate dehydrogenase gene (icd) in *Escherichia coli* and *Salmonella enterica*. *J. Bacteriol.* 179:6551–59
254. Ward-Rainey N, Rainey FA, Wellington EM, Stackebrandt E. 1996. Physical map of the genome of *Planctomyces limnophilus*, a representative of the phylogenetically distinct planctomycete lineage. *J. Bacteriol.* 178:1908–13
255. Wenzel R, Herrmann R. 1988. Physi-

cal mapping of the *Mycoplasma pneumoniae* genome. *Nucleic Acids Res.* 16: 8323–36
256. Wenzel R, Pirkl E, Herrmann R. 1992. Construction of an EcoRI restriction map of *Mycoplasma pneumoniae* and localization of selected genes. *J. Bacteriol.* 174:7289–96
257. Whitley JC, Muto A, Finch LR. 1991. A physical map for *Mycoplasma capricolum* Cal. kid with loci for all known tRNA species. *Nucleic Acids Res.* 19: 399–400
258. Wilkinson SR, Young M. 1995. Physical map of the *Clostridium beijerinckii* (formerly *Clostridium acetobutylicum*) NCIMB 8052 chromosome. *J. Bacteriol.* 177:439–48
259. Willems R, Paul A, van der Heide HG, ter Avest AR, Mooi FR. 1990. Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *EMBO J.* 9:2803–9
260. Wlodarczyk M, Nowicka B. 1988. Preliminary evidence for the linear nature of *Thiobacillus versutus* pTAV2 plasmid. *FEMS Microbiol. Lett.* 55:125–28
261. Woese CR. 1987. Bacterial evolution. *Microbiol. Rev.* 51:221–71
262. Xiang SH, Hobbs M, Reeves PR. 1994. Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. *J. Bacteriol.* 176:4357–65
263. Xu Y, Johnson RC. 1995. Analysis and comparison of plasmid profiles of *Borrelia burgdorferi* sensu lato strains. *J. Clin. Microbiol.* 33:2679–85
264. Xu Y, Kodner C, Coleman L, Johnson RC. 1996. Correlation of plasmids with infectivity of *Borrelia burgdorferi* sensu stricto type strain B31. *Infect. Immun.* 64:3870–76
265. Yan W, Taylor DE. 1991. Sizing and mapping of the genome of *Campylobacter coli* strain UA417R using pulsed-field gel electrophoresis. *Gene* 101:117–20
266. Ye F, Laigret F, Whitley JC, Citti C, Finch LR, et al. 1992. A physical and genetic map of the *Spiroplasma citri* genome. *Nucleic Acids Res.* 20:1559–65
267. Young M, Cole S. 1993. *Clostridium*. See Ref. 227a, pp. 35–52
268. Zhang JR, Hardham JM, Barbour AG, Norris SJ. 1997. Antigenic variation in Lyme disease *borreliae* by promiscuous recombination of VMP-like sequence cassettes. *Cell* 89:275–85
269. Zuccon FM, Hantman MJ, Sechi LA, Sun S, Piggot PJ, Daneo-Moore L. 1995. Physical map of *Streptococcus mutans* GS-5 chromosome. *Dev. Biol. Stand.* 85:403–7
270. Zuerner RL. 1991. Physical map of chromosomal and plasmid DNA comprising the genome of *Leptospira interrogans*. *Nucleic Acids Res.* 19:4857–60
271. Zuerner RL, Herrmann JL, Saint Girons I. 1993. Comparison of genetic maps for two *Leptospira interrogans* serovars provides evidence for two chromosomes and intraspecies heterogeneity. *J. Bacteriol.* 175:5445–51
272. Zuerner RL, Stanton TB. 1994. Physical and genetic map of the *Serpulina hyodysenteriae* B78T chromosome. *J. Bacteriol.* 176:1087–92

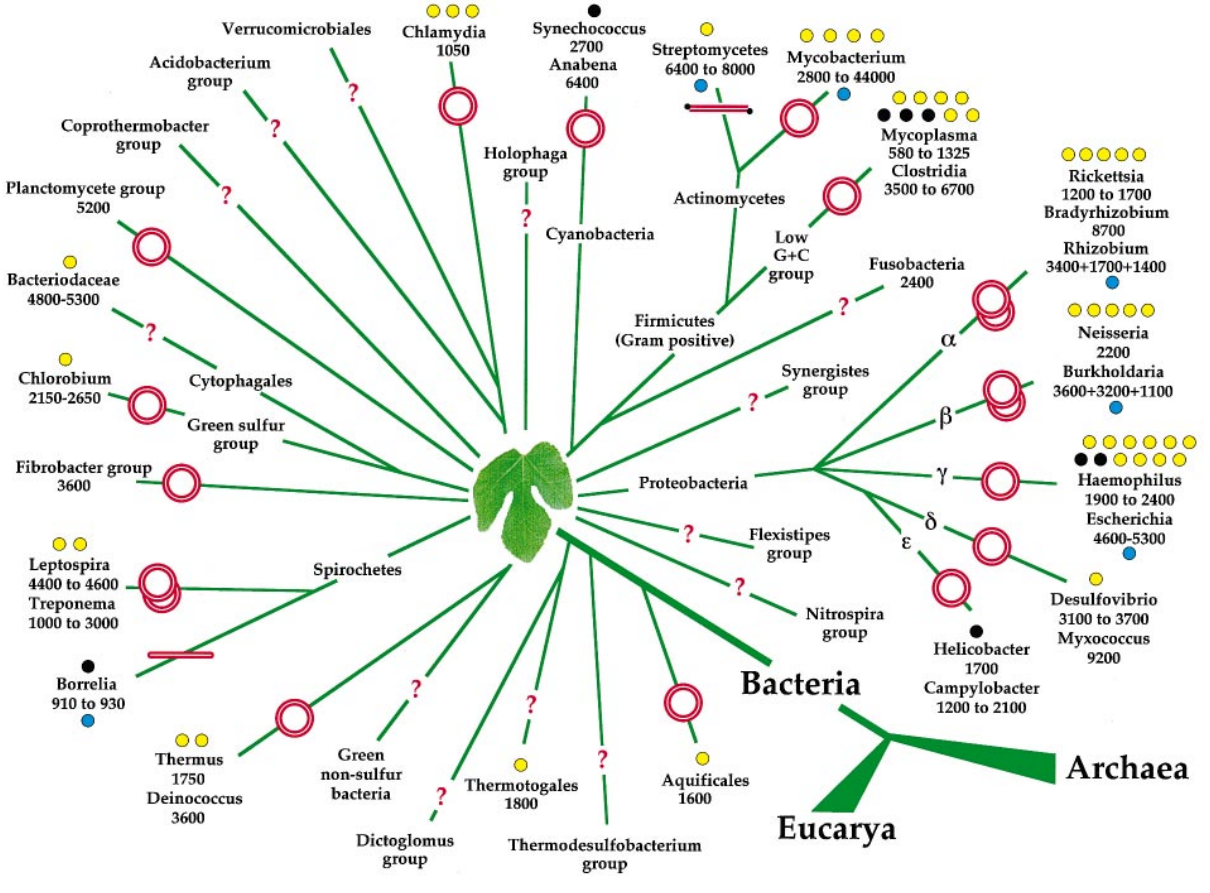

Verrucomicrobiales
Chlamydia
1050
Synechococcus
2700
Anabena
6400
Streptomycetes
6400 to 8000
Mycobacterium
2800 to 44000
Acidobacterium
group
Coprothermobacter
group
Holophaga
group
Mycoplasma
580 to 1325
Clostridia
3500 to 6700
Rickettsia
1200 to 1700
Bradyrhizobium
8700
Rhizobium
3400+1700+1400
Planctomycete group
5200
Actinomycetes
Cyanobacteria
Low
G+C
group
Bacteriodaceae
4800-5300
Fusobacteria
2400
Neisseria
2200
Burkholdaria
3600+3200+1100
Firmicutes
(Gram positive)
α
Synergistes
group
Chlorobium
2150-2650
Cytophagales
β
Green sulfur
group
Fibrobacter group
3600
Proteobacteria
γ
Haemophilus
1900 to 2400
Escherichia
4600-5300
δ
Flexistipes
group
ε
Leptospira
4400 to 4600
Treponema
1000 to 3000
Spirochetes
Nitrospira
group
Desulfovibrio
3100 to 3700
Myxococcus
9200
Helicobacter
1700
Campylobacter
1200 to 2100
Bacteria
Borrelia
910 to 930
Thermus
1750
Deinococcus
3600
Green
non-sulfur
bacteria
Thermotogales
1800
Aquificales
1600
Archaea
Dictoglomus
group
Thermodesulfobacterium
group
Eucarya

Figure 1 Bacterial chromosome size and geometry. An unrooted rRNA phylogenetic tree containing the 23 named major bacterial phyla and some of their relevant subgroups [from Hugenholtz et al (110) and the NCBI Taxonomy web site (http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy)]; at least 12 additional, as yet unstudied major phyla exist (110). The branch lengths are not meant to indicate actual phylogenetic distances, and the *fig leaf* in the center indicates a part of the phylogenetic tree where branching order is less firmly established. Chromosome geometry (circular or linear) is indicated by symbols near the ends of the branch lines; *multiple circles* on the α and β proteobacteria and a spirochete branch indicate that some members of these groups have multiple circular chromosomes. At the end of each branch, representative genera (or higher group names) are given with their known range of chromosome sizes in kbp; if no value is given, none has been determined from that group. Above each name the *black* and *gray circles* indicate the number of genome sequencing projects completed and under way, respectively, for that group in early 1998. Small circular extrachromosomal elements (plasmids) have been found in nearly all bacterial phyla, but linear extrachromosomal elements (indicated by *white squares* below the groups in which they have been found) are rare except in the *Borrelias* and *Actinomycetes*, where they are common. For a color version of this figure, see the color section at the back of the volume.

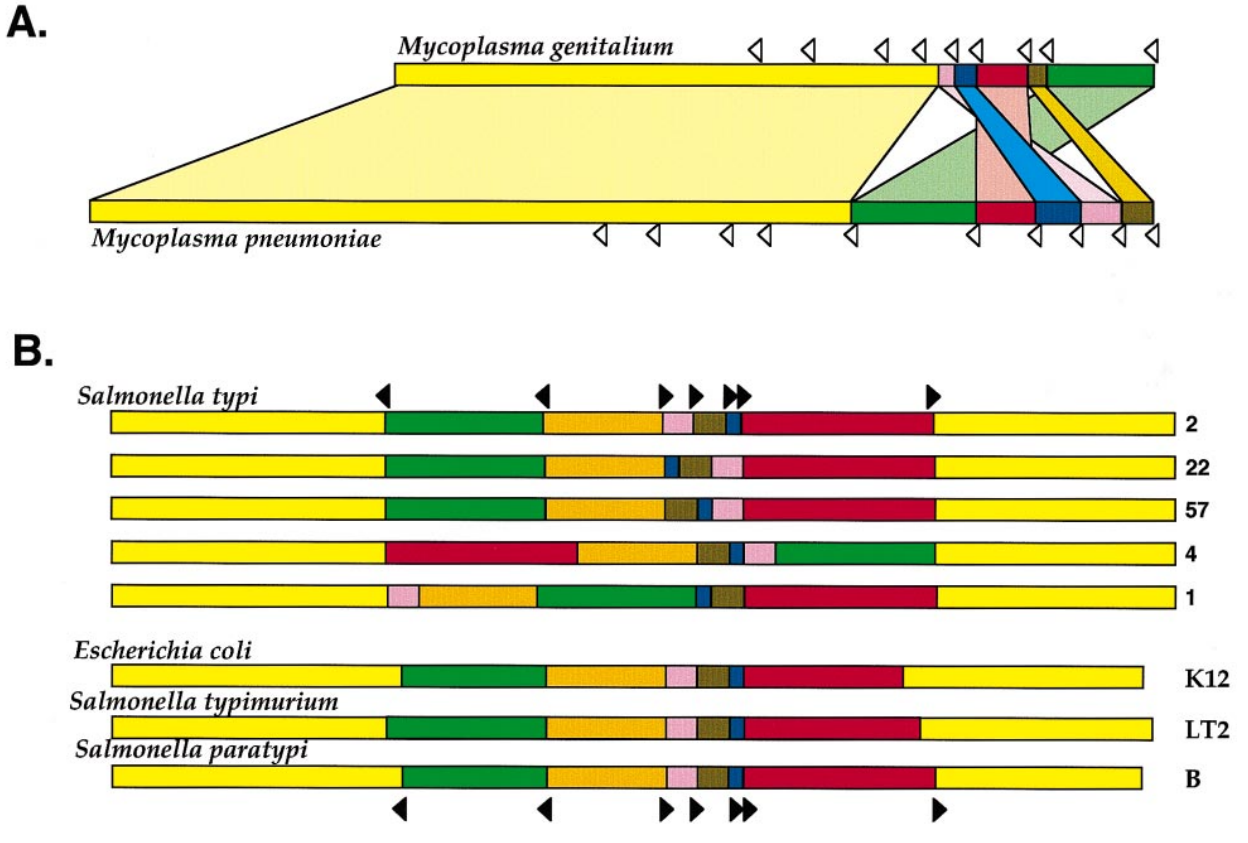

A.
Mycoplasma genitalium
Mycoplasma pneumoniae
B.
Salmonella typi
2
22
57
4
1
Escherichia coli
K12
Salmonella typimurium
LT2
Salmonella paratypi
B

Figure 4 Large differences in genome structure among species within a genus and within a species. *A*. Relationship between *M. pneumoniae* and *M. genitalium*. The *different shades* represent different genomic segments, and the *open triangles* indicate the locations of the repeated MgPa sequence (102). Each of the *M. genitalium* segments also contains deletions relative to *M. pneumoniae* (after Reference 102). *B*. Rearrangements among individuals within the species *Salmonella typhi*. The first five lines represent 5 of the 17 observed *S. typhi* genome arrangements. The *different shades* represent different genomic segments, and the *closed triangles* indicate the locations of the seven rRNA operons. The genomes of three related species of bacteria are shown below. On the right of the *S. typhi* lines, the number of individuals (among 127 analyzed) with that arrangement is given. Strain names are given to the right of the bottom three lines. All the chromosomes are opened for linearization approximately at the replication terminus; the origin of replication is in the *white segment* (after Reference 157).
For a color version of this figure, see the color section at the back of the volume.

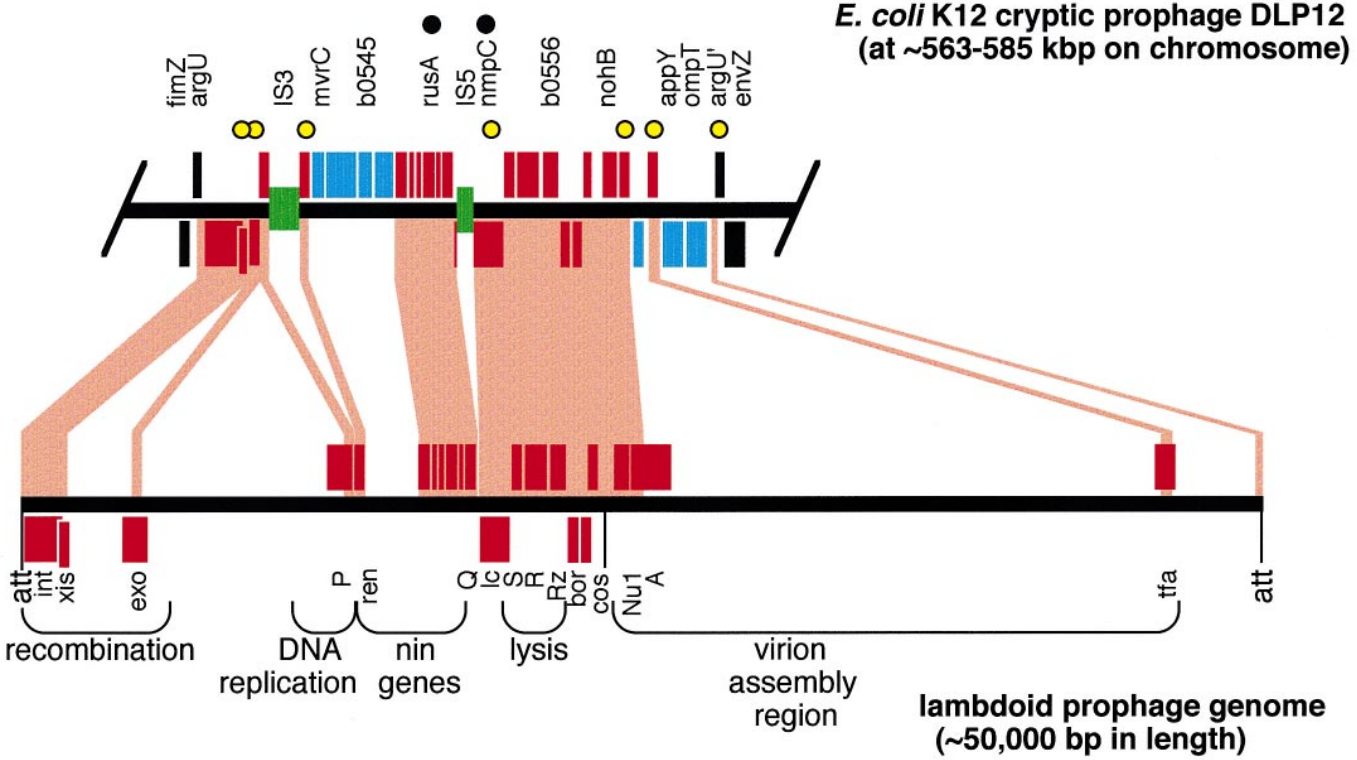

E. coli K12 cryptic prophage DLP12
(at ~563-585 kbp on chromosome)
fimZ
argU
IS3
mvrC
b0545
rusA
IS5
nmpC
b0556
nohB
appY
ompT
argU'
envZ
att
int
xis
exo
P
ren
Q
lc
S
R
Rz
bor
cos
Nu1
A
tfa
att
recombination
DNA replication
nin genes
lysis
virion assembly region
lambdoid prophage genome
(~50,000 bp in length)

Figure 5 *E. coli* cryptic prophage DLP12. *E. coli* strain K12 prophage DPL12 is shown above, and a lambdoid bacteriophage genome is shown below; connecting *gray areas* indicate homologies between the two DNAs. In each map genes are indicated by boxes; those above each map are transcribed rightward and below leftward; selected gene names are shown above and below the two maps. *Black and gray boxes* indicate all the open reading frames in DLP12 and selected genes on the lambdoid phage genome; *white boxes* indicate endochromosomal genes outside of DLP12. Only genes with homologs in DLP12 are shown on the lambdoid chromosome. *Cross-hatched boxes* on the DLP12 map indicate insertion sequences. The lambdoid phage shown is a composite of several closely related phages, but shows all the genes in their correct positions. The DLP12 genes with known phage homologs have different closest known relatives as follows: *int* and *xis*-P22; *P* and *ren*-λ; nin region genes −82; *Q*-21; *lc*-PA-2; lysis and virion assembly genes −λ. *Black circles* indicate genes that are naturally poorly expressed and require a mutation or IS excision to be expressed. *White circles* indicate genes that are obviously truncated or disrupted compared to their phage counterparts. Since analyses of lambdoid phage genes have not yet reached saturation (39), the gray genes in DLP12 might in fact have been a part of the prophage genome that originally occupied this site. Curiously, one, orf b0545, is very similar to *E. coli* orf yblB, which is immediately adjacent to the phage λ attachment site.

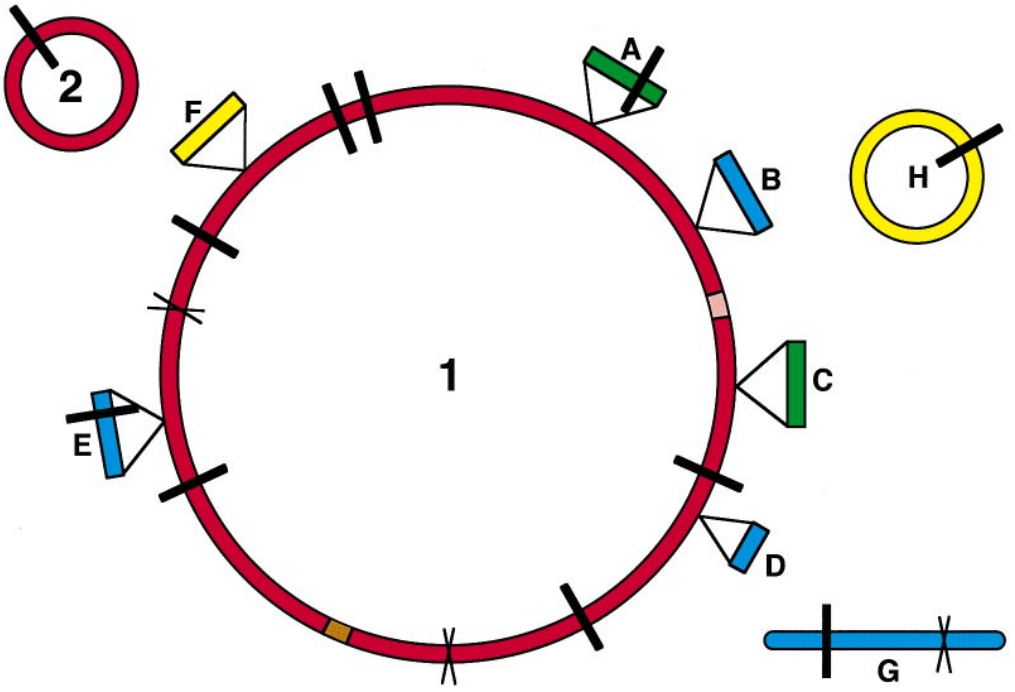

2
F
A
B
H
1
C
E
D
G

Figure 6 A generic dynamic bacterial genome. According to our current view, a generic bacterial genome is primarily made up of an endogenome consisting of one or more endochromosomes, shown in black (1 & 2). Components of the endogenome can be linear or circular and may or may not be undergoing segment shuffling or inversion. Other systematic physical changes (e.g. white invertable section and slipped strand mispairing changes indicated by Xs) may be occurring at a noticeable frequency, but the sequences in the endogenome are present in all healthy individuals (except for nonorthologous replacements, in *crosshatch*). In addition to this core of genetic material common to all members of a species, a number of accessory elements may or may not be present (*A–H*). These include multiple types of plasmids that may exist as free replicons (*H*) or integrated (*F*), prophages or fragments thereof that may be integrated or in free plasmid form (*B, D, E, G*), and pathogenicity islands (*A, C*). Finally, smaller mobile elements of various kinds (transposons, integrons, etc) can be inserted into any of the above genome components (*thick black lines*).