Organization of the Prokaryotic Genome
Organization of the Prokaryotic Genome

Edited by Robert L. Charlebois

Department of Biology, University of Ottawa, Ottawa, Ontario, Canada
CONTENTS

Contributors vii
Preface xi

1. Sequencing Microbial Genomes
   Christoph W. Sensen
   1

2. Statistical Predictions of Coding Regions in Prokaryotic Genomes by Using Inhomogeneous Markov Models
   Mark Borodovsky, William S. Hayes, and Alexander V. Lukashin
   11

3. Bacterial Genomes—All Shapes and Sizes
   Stewart T. Cole and Isabelle Saint-Girons
   35

4. Archaea: Whose Sister Lineage?
   Robert L. Charlebois
   63

5. The Eukaryotic Perspective: Similarities and Distinctions between Pro- and Eukaryotes
   Conrad L. Woldringh and Roel Van Driel
   77

6. Comparing Microbial Genomes: How the Gene Set Determines the Lifestyle
   Michael Y. Galperin, Roman L. Tatusov, and Eugene V. Koonin
   91

7. Impact of Homologous Recombination on Genome Organization and Stability
   Diarmuid Hughes
   109
8. Illegitimate Recombination in Bacteria
   Bénédicte Michel
   129

9. Insertion Sequences and Transposons
   Ronald Chalmers and Michel Blot
   151

10. Structure of DNA within the Bacterial Cell: Physics and Physiology
    Conrad L. Woldringh and Theo Odijk
    171

11. DNA Supercoiling and Its Consequences for Chromosome Structure and Function
    N. Patrick Higgins
    189

12. Local Genetic Context, Supercoiling, and Gene Expression
    Andrew St. Jean
    203

13. “Stable” Genomes
    Kenneth E. Sanderson, Michael McClelland, and Shu-Lin Liu
    217

14. Unstable Linear Chromosomes: the Case of Streptomyces
    Pierre Leblond and Bernard Decaris
    235

15. Genomic Flux: Genome Evolution by Gene Loss and Acquisition
    Jeffrey G. Lawrence and John R. Roth
    263

16. Gene Transfer in Escherichia coli
    Roger Milkman
    291

17. Genetic Inventory: Escherichia coli as a Window on Ancestral Proteins
    Bernard Labedan and Monica Riley
    311

18. Proteome Approach to the Identification of Cellular Escherichia coli Proteins
    Amanda S. Nouwens, Femia G. Hopwood, Mathew Traini, Keith L. Williams, and Bradley J. Walsh
    331

    Reginald K. Storms
    347

Index 367
CONTRIBUTORS

Michel Blot
Genomique Bactérienne et Evolution, Université Joseph Fourier-CNRS EP 2029-CEA
LRC12, Grenoble 38041 Cedex, France

Mark Borodovsky
Schools of Biology and Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0230

Ronald Chalmers
Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom

Robert L. Charlebois
Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada, and Canadian Institute for Advanced Research

Stewart T. Cole
Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cédex 15, France

Bernard Decaris
Laboratoire de Génétique et Microbiologie UA INRA 952, Université Henri Poincaré,
Nancy 1, Faculté des Sciences BP 239, 54506 Vandoeuvre-lès-Nancy, France

Michael Y. Galperin
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

William S. Hayes
School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230

N. Patrick Higgins
Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, AL 35294-2170
Femia G. Hopwood
Australian Proteome Analysis Facility (APAF), Macquarie University, Sydney, NSW 2109, Australia

Diarmaid Hughes
Department of Molecular Biology, Box 590, The Biomedical Center, Uppsala University, S-751 24 Uppsala, Sweden

Eugene V. Koonin
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

Bernard Labedan
Institut de Génétique et Microbiologie, CNRS UMR 8621, Bâtiment 409, Université de Paris-Sud, 91405 Orsay Cedex, France

Jeffrey G. Lawrence
Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

Pierre Leblond
Laboratoire de Génétique et Microbiologie UA INRA 952, Université Henri Poincaré, Nancy 1, Faculté des Sciences BP 239, 54506 Vandoeuvre-lès-Nancy, France

Shu-Lin Liu
Department of Medical Biochemistry, University of Calgary, Calgary, Alberta, Canada T2N 1N4

Alexander V. Lukashin
School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230

Michael McClelland
Sydney Kimmel Cancer Center, Room 300, 3099 Science Park Road, San Diego, CA 92121

Bénédicte Michel
Génétique Microbienne, Institut National de la Recherche Agronomique, Domaine de Vilvert, 78352 Jouy en Josas Cedex France

Roger Milkman
Department of Biological Sciences, The University of Iowa, Iowa City, IA 52242-1324

Amanda S. Nouwens
Australian Proteome Analysis Facility (APAF), Macquarie University, Sydney, NSW 2109, Australia

Theo Odijk
Faculty of Chemical Engineering and Materials Science, Delft University of Technology, P.O. Box 5045, 2600 GA Delft, The Netherlands

Monica Riley
Marine Biological Laboratory, Woods Hole, MA 02543

John R. Roth
Department of Biology, University of Utah, Salt Lake City, UT 84112

Isabelle Saint-Girons
Unité de Bactériologie Moléculaire et Médicale, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cédex 15, France
Kenneth E. Sanderson
Salmonella Genetic Stock Centre, Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada T2N 1N4

Christoph W. Sensen
National Research Council of Canada, Institute for Marine Biosciences, 1411 Oxford Street, Halifax, Nova Scotia, Canada B3H 3Z1

Andrew St. Jean
Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

Reginald K. Storms
Centre for Structural and Functional Genomics, Department of Biology, Concordia University, Montreal, Quebec, Canada H3G 1M8

Roman L. Tatusov
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

Mathew Traini
Australian Proteome Analysis Facility (APAF), Macquarie University, Sydney, NSW 2109, Australia

Roel Van Driel
E. C. Slater Instituut, BioCentrum Amsterdam, University of Amsterdam, Plantage Muidergracht 12, 1018 TV Amsterdam, The Netherlands

Bradley J. Walsh
Australian Proteome Analysis Facility (APAF), Macquarie University, Sydney, NSW 2109, Australia

Keith L. Williams
Australian Proteome Analysis Facility (APAF), Macquarie University, Sydney, NSW 2109, Australia

Conrad L. Woldringh
Institute for Molecular Cell Biology, BioCentrum Amsterdam, University of Amsterdam, Kruislaan 316, 1098 SM Amsterdam, The Netherlands
Until recently, genetic linkage analysis was restricted to a few model organisms for which genetic tools had been developed. Physical maps soon obviated the need for genetics in determining genome organization, but like their predecessors, these maps were necessarily coarse and incomplete. Genetic maps require genetic markers, and physical maps require discrete probes. Neither could suitably serve to describe a genome in its entirety. Due to this constraint, detailed structural, functional, and evolutionary analyses could only focus on the gene or on relatively tiny regions of the genome. The extensive work on genes as the objects of study has led to certain paradigms in molecular and evolutionary biology which still color our understanding of genomes.

To appreciate the distinction between genes and genomes, it is first necessary to identify which features are in common and which are different between a sequence of a gene and a sequence of genes. Both may suffer point mutations, including insertions, deletions, and replacements. Both contribute a phenotype that is subject to environmental controls; the gene produces a gene product, and the genome produces a cell. Both evolve with time. The gene, however, is far less modular than is the genome. Although functional domains in genes exist, functional domains in the genome—its genes—are by far its major organizational feature. In addition, structural elements within a gene, often designed to permit the correct placement of functional elements, are fundamentally different in purpose from the structural and functional elements in genomes. The question to be answered then is whether genomic organization contributes significantly to phenotype and to the evolution of phenotype. Linkage facilitates DNA replication as well as partitioning at cell division, and it has important effects in population genetics. But is linkage an important feature in cell and molecular biology? Is the genome an ordered set of genes or is it merely a set of genes?

To better appreciate what resources we have available for answering such fundamental questions, chapters 1 and 2 review the technology of collecting
sequence data and finding genes within that data, respectively. Next, a survey of high-level genomic characteristics is presented (chapter 3), followed by a discussion of what distinguishes Archaea from Bacteria (chapter 4) and eukaryotes from prokaryotes (chapter 5). Chapter 6 addresses the importance of genomic content on phenotype.

The next three chapters describe the major mechanisms by which genomic organization can change. These mechanisms include homologous recombination (chapter 7), illegitimate recombination (chapter 8), and transposition (chapter 9). Then we begin to look at DNA not only as a linear sequence of genes, but as three-dimensional, physical material (chapter 10) whose packaging (chapter 11) ties structure to function (chapter 12).

Next, the mutability of genomes is explored, first by operationally defining stability (chapter 13) and instability (chapter 14) and then by analyzing the causes and effects of horizontal genetic transfer (chapters 15 and 16) and the invention of new functions (chapter 17).

Finally, a more direct link between genomic content and cellular expression is covered, looking at proteomics (chapter 18) and then functional genomics (chapter 19).

Genome sequencing might be regarded as being similar to mountain-climbing expeditions, with surveyors (bioinformaticians) and geologists (biochemists) accompanying the explorers. Once having reached a summit, however, the team leader and resource manager can lose interest. Since the mountain has been climbed and there is nowhere else to go but down, the team sets its sights instead on another (usually higher) peak, dragging the surveyors and geologists away before their respective tasks on the present mountain have been completed.

It might be better to regard genomics metaphorically as akin to civics rather than to mountaineering and cartography. Genome sequencing as a discipline arose from the technology of genome mapping, but genomics should steer away from this physical mindset and return to its biological foundations. Biology is not so much about structure as it is about function; and the cell's function is complex indeed. The analogy to civics may be appropriate in that a city's efficiency is not only a function of the people and services within, but is also a function of their relative arrangement. I hope that the reader of this book learns to appreciate the importance of gene arrangement in the genome and its role in the shaping of life.

ROBERT L. CHARLEBOIS
INDEX

<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abridgement, incoming DNA</td>
<td>297-300</td>
</tr>
<tr>
<td>Acholeplasma oculi</td>
<td>40</td>
</tr>
<tr>
<td>Acinetobacter</td>
<td>36</td>
</tr>
<tr>
<td>addAB mutant</td>
<td>137</td>
</tr>
<tr>
<td>Aeropyrum pernix</td>
<td>65</td>
</tr>
<tr>
<td>Age structure, E. coli genome</td>
<td>276-277</td>
</tr>
<tr>
<td>Agrobacterium</td>
<td>48, 235</td>
</tr>
<tr>
<td>Agrobacterium tumefaciens</td>
<td>36, 47, 49-50, 236-237</td>
</tr>
<tr>
<td>Ameliorating genes</td>
<td>275-277, 279</td>
</tr>
<tr>
<td>Amino acid analysis, proteins</td>
<td>337-338</td>
</tr>
<tr>
<td>Amplifiable unit of DNA (AUD)</td>
<td>244-248, 253</td>
</tr>
<tr>
<td>type I</td>
<td>244-245</td>
</tr>
<tr>
<td>type II</td>
<td>244-245, 248</td>
</tr>
<tr>
<td>Amplification, see DNA amplification</td>
<td></td>
</tr>
<tr>
<td>Anabaena</td>
<td>42</td>
</tr>
<tr>
<td>Ancestral proteins</td>
<td>311-329</td>
</tr>
<tr>
<td>Anchoring of DNA</td>
<td>175</td>
</tr>
<tr>
<td>Ancient proteins</td>
<td>312</td>
</tr>
<tr>
<td>Antigenic novelty</td>
<td>300-303</td>
</tr>
<tr>
<td>Aquifex aelius</td>
<td>96, 101, 212</td>
</tr>
<tr>
<td>Aquifex pyrophilus</td>
<td>44</td>
</tr>
<tr>
<td>Archaea</td>
<td>63-76</td>
</tr>
<tr>
<td>chimerism</td>
<td>66-70</td>
</tr>
<tr>
<td>evolution</td>
<td>66-70</td>
</tr>
<tr>
<td>physical and genetic maps</td>
<td>44, 64-65</td>
</tr>
<tr>
<td>physical organization of nucleoid</td>
<td>65-66</td>
</tr>
<tr>
<td>Archaeal genomes</td>
<td>63-76</td>
</tr>
<tr>
<td>comparative genomic organization</td>
<td>70-71</td>
</tr>
<tr>
<td>genetic plasticity</td>
<td>64-65</td>
</tr>
<tr>
<td>genome statistics</td>
<td>46-47</td>
</tr>
<tr>
<td>mapping and sequencing</td>
<td>64-65</td>
</tr>
<tr>
<td>Archaeoglobus</td>
<td>272</td>
</tr>
<tr>
<td>Archaeoglobus fulgidus</td>
<td>23, 65, 93-95, 97, 99, 101, 103, 212</td>
</tr>
<tr>
<td>ATP synthase</td>
<td></td>
</tr>
<tr>
<td>F$_3$F$_1$-type</td>
<td>95-97</td>
</tr>
<tr>
<td>V-type</td>
<td>95-97</td>
</tr>
<tr>
<td>Atypical gene model</td>
<td>30</td>
</tr>
<tr>
<td>AUD, see Amplifiable unit of DNA</td>
<td></td>
</tr>
<tr>
<td>Automated sequencing machines</td>
<td>5</td>
</tr>
<tr>
<td>Automation, proteomics</td>
<td>338</td>
</tr>
<tr>
<td>Azorhizobium</td>
<td>50</td>
</tr>
<tr>
<td>BAC cloning system, see Bacterial artificial chromosome cloning system</td>
<td></td>
</tr>
<tr>
<td>Bacillus cereus</td>
<td>40, 50-51, 274</td>
</tr>
<tr>
<td>genome-wide strategies for studying gene function</td>
<td>347-365</td>
</tr>
<tr>
<td>Bacillus thuringiensis</td>
<td>40</td>
</tr>
<tr>
<td>Bacteria</td>
<td></td>
</tr>
<tr>
<td>comparison to eukaryotic nucleus</td>
<td>80-83</td>
</tr>
<tr>
<td>enteric, see Enteric bacteria</td>
<td></td>
</tr>
<tr>
<td>genome statistics</td>
<td>46-47</td>
</tr>
<tr>
<td>speciation</td>
<td>271-272, 277, 281-282</td>
</tr>
<tr>
<td>Bacterial artificial chromosome (BAC) cloning system</td>
<td>46</td>
</tr>
<tr>
<td>Bacterial interspersed mosaic elements (BIMEs)</td>
<td>196</td>
</tr>
<tr>
<td>Bacteriophage, gene acquisition from</td>
<td>281</td>
</tr>
<tr>
<td>Bar coded mutants</td>
<td>354</td>
</tr>
<tr>
<td>Bartonella bacilliformis</td>
<td>36</td>
</tr>
<tr>
<td>Basic genetic instability</td>
<td>242</td>
</tr>
<tr>
<td>bgly gene</td>
<td>138</td>
</tr>
<tr>
<td>BIMEs, see Bacterial interspersed mosaic elements</td>
<td></td>
</tr>
<tr>
<td>Bimolecular synopsis</td>
<td>159</td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>338</td>
</tr>
<tr>
<td>Bordetella pertussis</td>
<td>36</td>
</tr>
<tr>
<td>Borrelia</td>
<td>48, 235, 240</td>
</tr>
<tr>
<td>Borrelia afzelii</td>
<td>42</td>
</tr>
<tr>
<td>Borrelia andersonii</td>
<td>42</td>
</tr>
</tbody>
</table>
Borrelia burgdorferi, 42–43, 47, 93–96, 99–102, 212, 236–237, 239
Borrelia garinii, 43
Borrelia japonica, 43
Bradyrhizobium, 50
Bradyrhizobium japonicum, 36, 50
Breakage-fusion-bridge cycle, 252
Brucella, 45, 249
Brucella abortus, 36
Brucella canis, 36
Brucella melitensis, 36, 49
Brucella neotomae, 36
Brucella ovis, 36
Brucella suis, 36, 249
Burkholderia cepacia, 36, 49–50
CAI, see Codon adaptation index
Campylobacter coli, 36, 129, 140
Campylobacter fetus, 36
Campylobacter jejuni, 36
Campylobacter upsaliensis, 36
Cenancestor, 68–71
Chi site, 112, 116, 122, 152, 221
Chimerism
archaeal, 66–70
eukaryotic evolution, 70
Chlamydia, 47
physical and genetic maps, 42–43
Chlamydia trachomatis, 41, 47, 95–96, 100–101, 212
Chlorobium tepidum, 42
Chromatium vinosum, 36
Chromosomal splitting, 51
Chromosome(s)
circularized, 251–252
DNA supercoiling in, 189–202, 209–210
fused, 252
linear, see Linear chromosomes
local effects of supercoiling, 209–210
multiplicity, 45, 48–52
Chromosome arm, replacement, 249
Chromosome map, stabilizing on evolutionary time
scale, 109–110
Chromosome mechanics, sequence features required, 222
Chromosome segregation, 77–80
separation of daughter strands in bacteria, 83–84
separation of sister chromatids in eukaryotes, 77, 84–86
Chromosome territory model, 82–83
Cin recombinase, 139
dixQ site, 139
Clade, 303–304
Clonal frame, 295–296, 300, 303
Clonal segment, 295–296, 299, 304
Clonal sweep, 295, 304
Clone, species-wide, formation, 294–295
Clostridium acetobutylicum, 40, 45
Clostridium beijerinckii, 40, 45
Clostridium peptiﬁngens, 40, 45
Clumped genes, see Gene cluster
Clusters of orthologous groups (COGs), 98
Coding regions
length, 92
statistical predictions using Markov models, 11–33
Codon adaptation index (CAI), 304–305
Codon usage, 14, 16–17, 92, 273, 304–305
COGs, see Clusters of orthologous groups
Colinear genomes, 222
Colinearity, maintaining site for recombination to
insert foreign DNA, 222
Comparative genomics, archaeal genome
organization, 70–71
Composite transposons, 151
Confocal scanning light microscopy (CSLM),
nucleoid, 172–174
Conjugation, 292
Convergent evolution, 349
Copy break duplication, 249
Copy choice recombination, 133–134
Coregulation model, evolution of gene clusters, 284
Corynebacterium glutamicum, 40
Cotranscription, see also Gene cluster
facilitation of horizontal gene transfer, 285–286
prokaryotic cells and eukaryotic nucleus, 81
Coupled genes, see Gene cluster
Cross-hybridization analysis, 45
Cruciforms, 191
Cryofixation and freeze substitution, nucleoid
ﬁxation for electron microscopy, 176–177
CSLM, see Confocal scanning light microscopy
Cyanobacteria, physical and genetic maps, 42–43
Cytotrophic lineage, 68
D loop, 121
DARWIN program, 315–321
Daughter strands, separation in bacteria, 83–84
DAZ (Dif acting zone), 119–120
DDE catalytic triad, 152
Debye length, 179–180
Deletion(s), 110–114, 129
large-scale, leading to linear chromosomes, 248–249
M13 replication origin, 142–143
nested, 3
promoted by transposons, 153, 155–157
Streptomyces, 248–249, 254
at UV lesions, 143
Deletion hot spot, 143
Desulﬁurococcus, 97
Dichelobacter nodosus, 37
Dif acting zone, see DAZ
dif site, 118–119, 192, 220, 224
Directed-sequencing strategies, 2–3
Directional mutation pressure, 273–275
DNA
anchoring, 175
charge density, 179
charge interactions with itself, 179–181
concentration in bacterial cell, 80–81
concentration in eukaryotic nucleus, 80–81
conformation in prokaryotes and eukaryotes, 81–82
diameter, 179–180
electrostatic screening, 179–181
H-form, 191
hierarchical screening in bacteria, 78–80
lesions
illegitimate recombination and, 133
nonhomologous recombination induced by, 143
physics, 178–180
protein-nucleic acid interactions, 357–358
replication, see Replication
segregation in prokaryotes and eukaryotes, 78–80
structure within bacterial cell, 171–178
supercoiled, see Supercoiling of DNA
superhelical, in confined suspension of proteins, 180–183
unique blocks from lateral transfer, 226–228
wormlike chain model, 178–180
Z-DNA, 191
DNA amplification, 111, 129, 224, 252, see also Amplifiable unit of DNA
beyond duplication, 113
stability of amplified DNA, 248
Streptomyces, 242, 244–248, 255
DNA chip, 46, 354–356
DNA combing, 45
DNA gyrase, see Gyrase
DNA-binding proteins, 65, 174, 206, 210–211, 253
DNA–DNA hybridization, E. coli and S. enterica, 278–279
Domain
genome, 211
protein, 314
supercoiled DNA, 208–209
Dosage effect, 112–113, see also Gene dosage
Down mutants, illegitimate recombination, 136
Duplications, 313–314, 349–350
tandem, 111–113, 224

Ecological differentiation, genomic flux and, 270–271
Ecological niche, 281–282
ECOR strains, 293, 295–297, 300–303
Electron microscopy, nucleoid, 175–178
cryofixation and freeze substitution, 176–177
fixation conditions, 176
Electrostatic screening, in DNA, 179–181
Encyclopedias, 46
Endosymbionts, 271–272
Enteric bacteria
conservation of gene order, 217–233
gene rearrangements in stable bacteria, 222–228
Enterococcus faecalis, 40
Enterococcus hirae, 95–96
Escherichia coli, 4, 26, 46, 51, 85, 212, 227–228, 268
age structure of genome, 276–277
ATP synthase operons, 96
classes of genes, 16–17
clusters of orthologous groups, 99–100
compared to S. enterica, 273
conservation of gene order, 217–233
conserved gene strings, 92–95
divergence of Salmonella from, 278–280
essential genes, 264–265
eukaryotic repeats in microsatellite sequences, 131–132
evolution, 109–128
gene acquisition, 280–281
gene transfer, 291–309
GeneMark performance, 23
genome-wide strategies for studying gene function, 347–365
genomic organization, 293
genomics studies on, 311–312
horizontally transferred genes, 275
identifying foreign genes, 273–275
nondivisible zones, 116–117
paralogous proteins, numbering, 315–318
physical and genetic maps, 37
proteomics, 331–346
sequence similarities among proteins, 314–324
studying genes and genome evolution, 311–312
studying protein evolution, 312
studying relationships between sequence and functional similarities, 312
topoisomerases, 190
window on ancestral proteins, 311–329
Essential genes, 264–265, 348–349
strategies for analysis, 351–352
Eukaryotic cells
chimeric origins, 67
chromosomes, stability of repeated sequences, 131–132
comparison of nucleus to bacterial cell, 80–83
DNA segregation, 78
evolution, 70
prokaryotic cells vs., 63, 77–90
syntrophic origin, 67–68
Evolution
additivity results in many evolutionary trajectories, 304–305
applications of genome sequencing, 70
archaea, 66–70
Evolution (continued)
convergent, 349
gene clusters, 283–284
gene set and lifestyle, 91–107
genetic instability and, 254–256
genome, 263–289
studies in E. coli, 311–312
lack of conservation of gene order, 92–95
metabolic novelty, 282–283
metabolic pathways, 100–103
proteins, 312, 326–327
rate, 69
stabilizing chromosome maps, 109–110
universally conserved protein domains, 95–98
Exonuclease V, 137

Fisher theory, evolution of gene clusters, 283–284
Fitness
contributions of individual genes, 264–265
defined, 293
gene acquisition and, 277
genome maintenance and, 293–295
gene maintenance and fitness contribution, 265–267
gene rearrangements and, 122
transposons and, 163
Flagellar antigens, 223–224
Fluoroquinolones, 190
Forbidden inversions
not lethal, 118
S. typhimurium, 118
Foreign genes, see also Gene acquisition
identification, 272
F-prime factor, 51
Fused chromosomes, 252
Fusion proteins, 312

GAL1-GAL10 promoter region, 351–352
Gamma radiation, induction of nonhomologous
rearrangements, 143
γβ recombination system, 206, 211
γβ resolvase system, 194
γβ transposon, 204
gapA gene, 304–305
Gene(s)
ameliorating, 275–277, 279
as component of nucleoid, 203–204
essential, 264–265, 348–349
strategies for analysis, 351–352
existence implies a function, 264
fitness contribution, 264–265
functionally related, 349
intervals between, 226–228
nonequilibrium, 275
orphan, 347
Gene acquisition, 279
abridgement of incoming DNA, 296–300
assessment, 272–273
dynamics, 267–269
entry time, 272, 275–276
evidence for, 268
fitness and, 277
genomic flux and ecological differentiation, 270–271
identifying foreign genes in E. coli, 273–275
mechanisms, 280–281, 292
recombination and, 281
retention of given replacement, see Gene retention
Gene amplification, 224, see also DNA amplification
Gene boundaries, 23–26
Gene cluster
acquisition, 269
evolution, 283–284
facilitation of clustering, 284–285
facilitation of horizontal gene transfer, 285–286
Gene conversion, 114–115, 254, 263
Gene dosage, 220–221
tandem duplication and, 112–113
Gene expression, 203–215
assessing with transcript-specific DNA arrays,
354–356
local context and, 207–209
evidence from chromosome, 209–210
protein fluxes and networks, 341–342
proteomics, 331–346
rearrangements and, 253–254
supercoiling and, 221–222
transposon tagging for analysis, 352–353
Gene families, missing, 100–103
Gene finding
accuracy, 26–30
gene-finding algorithms, 13
in genomic sequences, 21–22
higher-order models, 30
location of gene boundaries, 23–26
models of typical and atypical genes, 30
similarity search, 13
using inhomogeneous Markov models, 11–33
Gene function
studying, 347–365
transposon tagging for analysis, 352–353
Gene loss
assessment, 272–273
dynamics, 264–267
evidence for, 267
Gene maintenance, see Gene retention
Gene order
conservation, 217–233
lack of conservation, 92–95
universally conserved, 94–95
Gene overlaps, 11, 26
Gene rearrangements
from cascade of structural instability events, 249–253
INDEX • 371

fitness and, 122
gene expression and, 253–254
genes which control variable surface properties, 223–224
genomic evolution and, 263
linear chromosomes of Streptomyces, 235–261
recombination and, 249, 252
rm operons, 224–227
stable enteric bacteria, 222–228
Streptomyces, 244–249, 254
Ter region, 224–225
transposable elements as cause, 152–159
Gene redundancy, 349
Gene replacement, 350–351
Gene retention, 265–267, 269, 291, 301–302
competition between genes for maintenance, 269–270
fitness and, 293–295
fitness contribution and, 265–267
selective, 265–267
Gene sets
lifestyle and, 91–107, 279
screening for functionally important genetic entities, 353–354
Gene shadow, 14
Gene shuffling, 304
Gene synteny, 349–350
Gene transfer
E. coli, 291–309
horizontal, 66–71, 103, 109, 152, 263–289, 291, see also Genomic flux
facilitation by gene clustering and cotranscription, 285–286
useful phenotypic information, 267–268
Gene tree, 295–296, 299
GeneMark core algorithm, 14–15
parameter variations, 15–16
predicting gene-containing ORFs in whole genomes, 22–23
prediction accuracy, 16–17
Genetic context, local, 203–215
Genetic drift, 265, 293–294, 302
Genetic instability
basic, 242
evolution and, 254–256
Streptomyces, 241–244
mutator states, 243–245
treatments that stimulate, 243
Genetic inventory
basic concepts, 312–314
E. coli as window on ancestral proteins, 311–329
Genetic maps, bacterial genomes, 35–62
Genome(s)
colinear, 222
domain structure, 211
evolution, 263–289
studies in E. coli, 311–312
genetic organization, 92
physical mapping, 35–46
screening for functionally important genetic entities, 353–354
shapes and sizes, 35–62
size, 46–47, 236, 263, 271–272
life cycle and, 236
size limit, 263, 266, 268–270
stability, effect of homologous recombination, 109–138
“stable,” 217–233
supercoiling and, 211–213
unstable, 235–261
Genome balance, 220
Genome fusion, 103
Genome sequence
facilitation of genetic and biochemical analysis, 350–352
facilitation of methods for characterizing ORF function, 352–356
facilitation of protein–protein and protein–nucleic acid analysis, 356–358
gene finding, see Gene finding
identification of new genes and, 348–350
prediction of gene-containing ORFs, 22–23
use in targeted gene replacement, 350–351
Genome sequencing, 1–9
archaeal genomes, 64–65
costs, 7–8
directed-sequencing strategies, 2–3
evolutionary applications, 70
gene set and lifestyle, 91–107
mixed strategy, 3–4
optimal strategies, 7–8
phases
finished sequence, 4
linking phase, 4
polishing phase, 4
primary sequencing phase, 4
primer-walking strategy, 2–4
sequencing factories, 7
tools
automated sequencing machines, 5
primer synthesis, 6
robotic workstations, 6
software, 6–7
total-genome shotgun sequencing, 1–2
“Genome signature,” 92
Genome studies, proteomics and, 332
Genomic comparisons, 91–107
Genomic flux, 263–289
dynamics, 269–272
ecological differentiation and, 270–271
evolution of gene clusters, 283–284
evolution of metabolic novelty, 282–283
facilitation of gene clustering, 284–285
impact, 280–286
Genomic flux (continued)
measuring, 272–280
point-mutational change vs., 277–278
speciation and, 271–272, 281–282
Genomic organization
E. coli, 293
effect of homologous recombination, 109–138
Genomic plasticity, 50, 64–65, 161, 228–229, 316
Genomics, 1
Giant linear plasmid, 237–240
GLIMMER program, 23
Gram-positive bacteria, physical and genetic maps, 40–41
Green sulfur bacteria, 47
physical and genetic maps, 42–43
gyrA mutant, 141
Gyrase, 141, 175, 190–192, 196
H antigen, 223
Haemophilus ducreyi, 37
Haemophilus influenzae, 4, 23, 37, 46, 92–96, 99–102, 228
Haemophilus parainfluenzae, 37
Halobacterium, 47
Halobacterium salinarium, 44, 65
Halofex mediterranei, 44, 65
Halofex volcanii, 44, 47, 65
“Hamiltonian” energy, 178
has locus, 249
Helicobacter pylori, 21, 23, 37, 46, 93–94, 96, 98–102, 153, 212
Heterogeneous nuclear particles (hnRNP), 81, 85
H-form DNA, 191
Hidden Markov model, 14, 24–26
with duration, 24
Hierarchical DNA segregation, 78–80
Hin recombinase system, 224
HisA protein, 325–326
HisF protein, 325–326
Histone(s), 82
Histone homologs, 65
hnRNP, see Heterogeneous nuclear particles
Homeostatic supercoil regulation model, 191–192
Homologous proteins, in different microorganisms, 98–99
Homologous recombination, 109–128, 291, see also Recombination
size of minimal sequence, 129
Homology, similarity vs., 313
Horizontal gene transfer, 66–71, 103, 109, 152, 263–289, 291, see also Genomic flux
facilitation by gene clustering and cotranscription, 285–286
rate, 277
useful phenotypic information, 267–268
hsd-mcr-mrr restriction region, 300–303
Hybridization technique, physical mapping, 35, 45
Hydratase family, 321–323
Hyperrecombination phenotype, 137–138
Hypervariability, 300–303
Streptomyces, 241–243
I–CeuI, 45, 218–219
Illegitimate recombination, 129–150, 153
between nonhomologous sequences, 140–143
between short homologous sequences, 129–139
between short repeats, 130–132
DNA lesions and, 133
functional parameters and direct repeat stability, 132
generic studies, 136–139
length and GC content of repeated sequences, 130
length of intervening sequence and, 130
mode of recombination and, 132
molecular models, 133–135
replication pauses and, 132–133
secondary structure and, 130–131
stability of repeats in bacterial chromosomes, 131
down mutants, 136
by replication slippage, 135
by single-strand annealing, 134–135
with site-specific elements, 139–140
Streptomyces, 248
transcription and, 133
up mutants, 136–137
ilvIH promoters, 207–208
Inhomogeneous Markov models, 11–33
Insertion(s), genomic evolution and, 263
Insertion sequences, see IS elements
Inside-out transposition, 157–159
Intramolecular transposition, 139
Intron, mobile, 151, 162–163
Inversions, 115–118, 222, 129
forbidden, not lethal, 118
forbidden inversions in S. typhimurium, 117–118
natural, 115
promoted by transposons, 153, 155–157
selected, 115–116
site-specific inversion systems, 224
Invertron, 254, 256
replication, 240–241
structure, 236–239
IS elements, 51–52, 151–169
archaea, 64
associated with acquired genes, 281
defined, 151
experimental evolution, 163–164
physiological regulation, 163–164
Isoelectric focusing, 333–334
with immobilized pH gradients, 334
INDEX  ■  373

Kinases of sugars, 323–324
Klebsiella aerogenes, 115
Klebsiella oxytoca, 226
Klebsiella pneumoniae, 51
Kullback-Liebler (K-L) distance, 17
LI3 operon, 94
Lactococcus, 45
Lactococcus lactis, 40
Lateral gene transfer, see Horizontal gene transfer
Leptospira interrogans, 43, 49, 51
leu-500 promoter, 191, 207–211
leuO promoter, 208
lexA mutant, 138–139
Life cycle, genome size and, 236
Lifestyle, gene set and, 91–107, 279
lig mutants, 136
Light microscopy, nucleoid, 172–174
Linear chromosomes, 47–48
  circularized, 251–252
  large-scale deletions leading to, 248–249
  plasmid interactions, 237–240
  Streptomyces, 235–261
Linear plasmid, 47–48, 237–240
  chromosome interactions, 237–240
Linking number (ΔLk), 204, 210–211
Linking-clone analysis, 45
Lipopolysaccharide, 223–224
Listeria monocytogenes, 45
Local context, gene expression and, 207–210
Local-context model, 212–213
Long terminal repeats, 244
Longest-ORF rule, 26
Loops, 226–228
“Low-protein chromatin,” 175

M13 replication origin, deletions at, 142–143
MAGPIE, 6–7
Markov model(s)
  hidden, 14, 24–26
  inhomogeneous, 11–33
  order of Markov model used, 17–21
Mass spectrometry (MS)
  electrospray ionization-TOF MS, 337
  identification of proteins and protein complexes, 357
  liquid chromatography-MS, 337
  MALDI-MS, 357
  MALDI-TOF MS, 337
Megaplasmids, 48
MEPS, see Minimum efficient processing segment
Meroclonal model, 304
Meroclone, 296
Metabolic novelty, evolution, 282–283
Metabolic pathways, evolution, 100–103
MetB protein, 325
MetC protein, 325

Methanobacterium thermoautotrophicum, 23, 44, 65, 93–95, 97, 99, 101, 103
Methanobacterium wolfei, 65
Methanococcus jannaschii, 23, 44, 65, 92–95, 97, 99, 101, 103, 212, 272
Methanococcus voltae, 44, 65
Methanococcus wolfei, 44
Methanosarcina mazei, 65, 97
Microcococcus, 37
Microsatellite sequences, E. coli, stability of introduced eukaryotic repeats in, 131–132
Minigels, 334
Minimum efficient processing segment (MEPS), 129, 136
Mitotic segregation, 77–78
Mixed strategy, genome sequencing, 3–4
Mobile intron, 151, 162–163
Module, protein, see Protein module
Moraxella catarrhalis, 37
MS, see Mass spectrometry
Multichromosomal bacteria, 48–52
Multifork replication, 78–79
Multimodular protein, 314, 318–319, 321
Multiplicity of chromosomes, 45, 48–52
Mutation
  directional mutation pressure, 273–275
  genomic evolution and, 263
  genomic flux vs., 277–278
Mutation rate, 265–266, 273
  distribution of transposable elements and, 159–161
  Streptomyces, 236
Mutator states, Streptomyces, 243–245
mutHLS gene, 137
Mycobacterium bovis, 40, 46
Mycobacterium branderi, 47
Mycobacterium celatum, 47
Mycobacterium leprae, 40
Mycobacterium leprae, 40, 46, 101, 140, 212
Mycobacterium xenopi, 40
Mycoplasma capricolum, 40
Mycoplasma gallisepticum, 41
Mycoplasma hominis, 41
Mycoplasma hyorhinis, 85
Mycoplasma mobile, 41
Mycoplasma mycoides, 41
Mycoplasma pneumoniae, 23, 41, 92–94, 96, 99–102, 272
Mycoplasma xanthus, 37

Natal theory, evolution of gene clusters, 283
Natural inversions, 115
Natural transformation, 292–293
Negatively frequency dependent selection (NFDS), 295, 301
Neisseria gonorrhoeae, 38
Neisseria meningitidis, 38
Nested deletions, 3
NFDS, see Negatively frequency dependent selection
Niche alteration, 271
Noncoding region, 14
Nondivisible zones, E. coli, 116–117
Nonequilibrium genes, 275
Nonhomologous sequences, illegitimate recombination between, 140–143
Nonrepetitive transposition, 153–154
Novel join points, 113
Nucleoid
archaeal, 65–66
arising from phase separation, 180–183
dNA structure within bacterial cell, 171–187
electron microscopy, 175–178
cryofixation and freeze substitution, 176–177
fixation conditions, 176
genes as components of, 203–204
light microscopy, 172–174
during segregation, 83–84
transcription activity, 176
transcription-mediated segregation model, 172–174, 176
Nucleosome, 82
Nucleotide composition
ameliorating genes, 275–276
position-specific, 273–275
Nucleus, eukaryotic, 77
comparison to bacterial cell, 80–83
O antigen, 223, 301
Oligonucleotide microarrays, 355
Open reading frame (ORF), 11–12
characterization of function, 352–356
defined, 11
density, 92
detecting protein-coding potential, 12–14
distances between, 212
length, 12, 26
longest-ORF rule, 24
prediction of gene-containing ORFs in whole genomes, 22–23
Operon, 95–97, 152, 221, 283–284
Operon fusions, 113
Optical mapping, 45
Ordered-library strategy, 46
ORF, see Open reading frame
oriC site, 220–221, 241
Origin of replication, 78–79, 116, 241, 252
M13, 142–143
multichromosomal bacteria, 51
Orphan genes, 347
Orthologous genes, 98, 100, 282, 313–314
PAGE, see Two-dimensional polyacrylamide gel electrophoresis
Palindromic tetranucleotides, 92
PAM units, 314–317
Paralogous genes, 98, 100, 282, 313–314, 316
Paralogous proteins, 349
E. coli, numbering, 315–318
Paranemic helix, 189–190
PAS, see Probabilistic ancestral sequence
Pathogenicity island, 226–228, 279
pdu-cob block, 279
Peptide mapping, 337
Peptide mass fingerprint (PMF), 337
Periplasmic binding proteins, family of repressors and binding proteins, 323, 325
PFGE, see Pulsed-field gel electrophoresis
Phenotypic capabilities
 genomic flux and ecological differentiation, 270–271
horizontal transfer of useful information, 267–268
spontaneous instability in Streptomyces, 241–243
phoA gene, 267
Phosphate transport system, 94
Phylogeny of organisms, vs. genealogy of proteins, 312–313
Physical maps, bacterial genomes, 35–62
Planctomyces limnophilus, 42
Planctomycetes, 47
physical and genetic maps, 42–43
Plasmids, 45
gene acquisition from, 281
illegitimate integration, 140
interactions with linear chromosome, 237–240
linear, 47–48, 237–240
Streptomyces, 236, 239–240
transfer, 292
Plectonemic helix, 189–190
Plectonemic supercoils, 204–205
Pleiotropic mutants, Streptomyces, 241–244
PMF, see Peptide mass fingerprint
Point-mutational change, genomic flux vs., 277–278
Porins, 323, 325
Posttranslational modification, proteins, 332, 338–341
PriA protein, 121
Primer synthesis, 6
Primer-walking sequencing strategy, 2–4
Primosome, 121
Probabilistic ancestral sequence (PAS), 321
Probability of ultimate survival, 294
Prokaryotic cells, see also Archaea; Bacteria
DNA segregation, 78
eukaryotic cells vs., 63, 77–90
Protein(s), see also Proteomics
ancestral, 311–329
ancient, 312
E. coli, proteomics, 331–346
evolution, 326–327
studies in E. coli, 312
fluxes and networks, 340–342
functional constraints on divergence among, 316
functional similarities, 312, 324–327
fusion, 312
genealogy, vs. phylogeny of organisms, 312–313
identification, 314–315
identification in proteome displays, 335–338
interactions with chromosomal DNA, 180–183
isoelectric focusing, 333–334
localization, transposon tagging for analysis, 352–353
motifs, 314
posttranslational modification, 332, 338–341
protein–nucleic acid interactions, 357–358
protein–protein interactions, 356
sequence similarities, 312, 324–326
E. coli, 314–324
sequence-related, 312
terminal, 236, 241
turnover, 322
Protein domains, 314
universally conserved, 95–98
Protein family
defined, 314
grouping modules into, 319–324
mixed, 323, 325
Protein module, 314
defined, 318–319
grouping into families, 319–324
length, 319–320
multimodal protein, 314, 318–319, 321
numbering, 319
Protein–coding genes, 11–12
detecting protein–coding potential, 12–14
gene finding using inhomogeneous Markov models, 11–33
Proteomics, 331–346
automation for high throughput, 338
bioinformatics, 338
challenges, 343
characterization of protein modifications, 338–341
defined, 331–332
protein fluxes and networks, 341–342
protein identification, 335–338
amino acid analysis, 337–338
N- and C-terminal tagging, 337
peptide mapping, 337
proteome and genome studies, 332
proteome displays, 332–335
Proteus mirabilis, 51
proU promoter, 210
Pseudomonas aeruginosa, 38, 51
Pseudomonas fluorescens, 38
Pulsed-field gel electrophoresis (PFGE), 35, 45, 218–220
Purine biosynthesis pathway, 100–102
PurK protein, 326
Purple bacteria, 46
physical and genetic maps, 36–39
PurT protein, 326
Pyrimidine biosynthesis pathway, 100–103
Pyrobaculum aerophilum, 65
Pyrococcus abyssi, 65
Pyrococcus furiosus, 65
Pyrococcus horikoshii, 65, 95, 97, 212
PurK protein, 326
PurT protein, 326
Pyrococcus abyssi, 65
Pyrococcus furiosus, 65
Pyrococcus horikoshii, 65, 95, 97, 212
Racket frame model, 239, 244
Random-sequencing phase, 3–4
Rearrangements, see Gene rearrangements
RecA, 111–112, 121, 136, 246, 251
RecBCD, 111–112, 121, 136
recD gene, 137
RecF, 111–112
Reciprocal recombination, 114
ref mutants, 136
Recombinase, site-specific, 139, 248
Recombination
between directly oriented repeat sequences, 110–114
between m loci, 249
between transposons, 155–156
chromosome rearrangement, 249
effect in terminus region, 118–120
effect on genome organization and stability, 109–128
gene acquisition and, 281
gene conversion, 114–115
homologous, 109–128, 291
size of minimal sequence, 129
illegitimate, see Illegitimate recombination
inversions generated, 115–118
nonhomologous, induced by DNA lesions, 143
rearrangement and, 252
reciprocal, 114
replication and, 120–122
Streptomyces, 237, 254–256
transcription and, 133
transspecies, 222
Recombination rate, 265
recQ mutant, 137
REP element, 113
rep sequence, 196
Repeated sequences
eukaryotic, stability, 131–132
short, illegitimate recombination between, 130–132
stability in bacterial chromosomes, 131
Replacement clock, 304
Replication, 78–79
damage-inducible stable, 121
direction, 241
effect on illegitimate recombination, 132
invertrons, 240–241
linear chromosome, 47–48
multichromosomal bacteria, 51
multifork, 78–79
pauses in, 132–133
problem of supercoiled DNA, 192–193
recombination and, 120–122
termination, 240–241, 251
Replication fork, collapsed, 121
Replication slippage, 133–134
illegitimate recombination by, 135
Replicative transposition, 153–155
intramolecular, 152
Replicon fusion, 153, 155–157
Repressor, family of repressors and periplasmic binding proteins, 323, 325
Restriction endonucleases, abridgement of incoming DNA, 297
Retrons, 151
rfb gene complex, 301–303
Rhizobium, 48, 50
Rhizobium mелилот, 38, 50
Rhodobacter capsulatus, 38, 211
Rhodobacter sphaeroides, 38, 49–50
Rhodococcus fascians, 47, 237
Rhodococcus formans, 140
Ribosomal proteins, 94
Rickettsia prowazekii, 101, 212
Rickettsia sibirica, 47
Rickettsiella melolonthae, 41
RNA, protein-nucleic acid interactions, 357–358
RNA polymerase, 94
problem with supercoiled DNA, 205–206
Robotic workstations, 6
rm operons, 45–46, 49–50, 111–112, 115, 218, 221, 249
gene rearrangements, 224–227
s value, 265–269
S10 operon, 94
Saccharomyces cerevisiae, 92, 97, 99, 249
analysis of essential genes, 351–352
assessing gene expression with transcript-specific DNA arrays, 354–356
gene-wide strategies for studying gene function, 347–365
identification of proteins and protein complexes, 356–357
protein–protein interactions, 356
screening gene sets for functionally important genetic entities, 353–354
targeted gene replacement, 350–351
Salmonella, 45, 51, 267
divergence from E. coli, 278–280
Salmonella hongori, 228
Salmonella enterica, 85, 268
compared to E. coli, 273
leu-500 promoters, 207–209
Salmonella enteritidis, 39, 115, 224, 227
Salmonella gallinarum, 226
Salmonella paratyphi, 39, 45, 224, 226–227
Salmonella pullorum, 226
Salmonella typhimurium, 39, 226, 249
conservation of gene order, 217–233
evolution, 109–128
forbidden inversions, 117–118
sbcCD mutant, 137
SDS-PAGE gel, 333
Selected inversions, 115–116
Selection coefficient, 294, 302
SELEX technology, 357–358
Selfish operon model, 284–285
SeqA protein, 195
Sequence-related proteins, 312
Serpulina hydysenteriae, 43
Shigella flexneri, 39, 220
Short repeated sequences, illegitimate recombination between, 130–132
Shotgun sequencing, total-genome, 1–2
Sigma factors, 254
σ value (specific linking difference), 204, 210–211
homeostatic control model, 191–192
total vs. effective, 204
Similarity, homology vs., 313
Similarity search, gene-finding, 13
Single-strand annealing (SSA), 133–136
Sister chromatid(s)
disentanglement, 84–85
separation in eukaryotes, 77, 84–86
Sister chromatid exchange, 249
Site-specific DNA inversion systems, 224
Site-specific element, illegitimate recombination associated with, 139–140
Site-specific enzymes, 139–140
Site-specific recombinase, 139, 248
Slab gel technologies, genome sequencing, 5
Slipped-mispairing model, 5
SMC proteins, 85
Software, genome sequencing, 6–7
Solenoidal supercoils, 204
SOS response, 121
spc operon, 94
Speciation
bacterial, 271, 277
genomic flux and, 281–282
Species tree, 295–296
INDEX • 377

Spirochetes, 47
  physical and genetic maps, 42–43
_Spiroplasma citri_, 41
SpoOJ protein, 195–196
SSA, see Single-strand annealing
ssb mutant, 138
“Stable” genomes, 217–233
Staden software package, 6
Start codon, location of gene boundaries, 23–26
_Stigmatella aurantiaca_, 39
str operon, 94
_Streptococcus mutans_, 41
_Streptococcus pneumoniae_, 41
_Streptococcus pyogenes_, 41
_Streptococcus thermophilus_, 41
_Streptomyces_ deletions, 248, 254
  DNA amplification, 242, 244–248, 255
  genetic instability, 241–244
  mutator states, 243–245
  treatments that stimulate, 243
  genome rearrangements, 236–237
  genome size, 236
  hypervariability, 241–243
  illegitimate recombination, 248
  interactions between plasmid and chromosome DNA, 237–240
  mutation rate, 236
  plasmids, 236–240
  rearrangements, 244–249, 254
  recombination, 237, 254–256
  spontaneous phenotypic instability, 241–243
  unstable linear chromosomes, 47–48, 235–261
_Streptomyces achromogenes_, 242, 244, 253
_Streptomyces ambrosfaciens_, 41, 236–239, 241–244, 246, 249, 251–256
_Streptomyces antibioticus_, 242
_Streptomyces bambergiensis_, 240
_Streptomyces catteya_, 242
_Streptomyces coelicolor_, 41, 236–239, 242, 253, 255
_Streptomyces fradiae_, 244
_Streptomyces griseus_, 41, 236, 239, 242
_Streptomyces lividans_, 41, 48, 236, 238–240, 242–246, 248, 253–256
_Streptomyces parvulus_, 238
_Streptomyces reticuli_, 242
_Streptomyces rimosus_, 236, 239–240, 242, 248, 253, 255–256
_Streptomyces rochei_, 239, 241
_Streptomyces tendae_, 242
_Streptomyces violaceus-niger_, 242
Sugar kinase family, 323–324
_Sulfolobus sulfataricus_, 6–7, 65
_Sulfolobus tokodaii_, 44, 65
Supercoiling of DNA, 174–175, 203–215
  association with DNA-binding proteins, 174
barrier structure, 194–195
conservation of gene order and, 221–222
domains, 208–209
effect on chromosome structure and function, 189–202
gene expression and, 204–206, 221–222
genome and, 211–213
homeostatic supercoil regulation model, 191–192
plectonemic supercoils, 204–205
replication and transcription problems, 192–193
solenoidal supercoils, 204
structural effects of supercoil distribution, 191
supercoil density, 191
supercoil diffusion
  plectonemic distribution and, 193–194
torsional distribution and, 193
transcription and, 204–206
Superhelical DNA, 180–183
Surface properties
  hypervariable regions, 301
  rearrangements in genes which control, 223–224
Synapse model, 254
Synapsis, bimolecular, 159
_Synechococcus_, 42
Syntenic conservation, 222
Syntrophic origin, eukaryotic cells, 67–68
Tandem duplications, 111–113, 224
TE, see Transposable elements
Telomere, 47–48
Ter sites, 117, 119–121, 137, 220, 251
gene rearrangements, 224–225
Ter-like sites, 251
Terminal inverted repeats (TIR), 236–241, 246–249, 251, 253–256
Terminal proteins, 236, 241
Terminus hyperrecombination zone (TRZ), 119–120
tetO promoter, 352
_Thermococcus celer_, 44, 65
_Thermoplasma acidophilum_, 65
_Thermus aquaticus_, 95
_Thermus thermophilus_, 39, 96
_Thiobacillus cuprinus_, 39
_Thiobacillus ferrooxidans_, 39
Three-domain view, 63
TIGR Assembler, 6
TIR, see Terminal inverted repeats
TMP, see 4,5',8-Trimethylpsoralen
Tn5 transposase, 139
Tn10 insertions, 220
Tn10 transposase, 139
topA gene, 141, 191–192, 207–209
topB gene, 138
Topoisomerase, 85, 138, 189–190, 192, 211
archaeal, 65–66
errors, 141
Topoisomerase I, 175, 190, 207
errors, 141–142
Topoisomerase II, 190
Topoisomerase III, 190
Topoisomerase IV, 175, 190, 196
Total-genome shotgun sequencing, 1–2
Transcription
activity at nucleoid, 176
cotranscriptional processes in prokaryotic cells and
eukaryotic nucleus, 81
orientation, 221
problem of supercoiled DNA, 192–193, 204–206
recombination and, 133
Transcription factors, 206
Transcription-mediated nucleoid segregation model,
172–174, 176
Transcript-specific DNA array, assessing global gene
expression, 354–356
Transduction, 292
Transformation, 292
natural, 292–293
Translation, location of gene boundaries, 23–26
Translocations, 113–114, 122, 129
Transposable elements (TE), 151–169, see also IS
elements; Transposon(s)
beneficial, 161
as cause of rearrangements, 152–159
classification, 151–152
common ancestry, 152
experimental evolution, 163–164
interactions with host, 159–164
mutation rate and distribution, 159–161
negative impact on host abilities, 161
physiological regulation, 163–164
positive impact on host, 163
transmission, 163
Transposase, 162–163
Transposition
control, 162–163
erroneous processing of intermediates, 140
inside-out, 157–159
intramolecular, 139, 152
mechanisms, 153
nonreplicative, 153–154
replicative, 152–155
Transposon(s), 151–169
bimolecular synapsis, 159
composite, 151
defined, 151
gene acquisition and, 281
homologous recombination between, 155–156
regions of portable homology, 155–156
Transposon excision, 137–138, 155–156
Transposon tagging, 352–353
Transposon-mediated sequencing, 3
Transspecies recombination, 222
Treponema denticola, 43
Treponema pallidum, 43, 95, 100–101
4,5',8-Trimethylpsoralen (TMP), 208, 210–211
tRNA genes, foreign genes adjacent to, 281
TRZ, see Terminus hyperrecombination zone
tuf genes, 115
Two-dimensional polyacrylamide gel electrophoresis
(2-D PAGE), 331–346
prefractionation techniques, 334–335
problems, 334
sample preparation, 334
Typical gene model, 30
Ultraviolet radiation, induction of nonhomologous
rearrangements, 143
Unstable genomes, 235–261
Up mutants, illegitimate recombination, 137–139
Ureaplasma urealyticum, 41
uup gene, 138
Vibrio cholerae, 39
Vitamin B₁₂ metabolism, 279
Xenologous genes, 313
Xenologous proteins, 316
Xer site, 118–119
xonA mutant, 137–138
xth mutants, 136
Yersinia cholerae, 228
Yersinia pestis, 228
Z-DNA, 191