Index

3-point formula, 118
4-point condition, 119

ab initio methods, 29, 40
accession number, 18, 19
additive distance, 117
additive tree, 117
Affymetrix, 147
AIDS, 96
Alanine, 24
AlignACE, 172
alignment, 39, 48
  approximate, 51
  global, 42, 43, 48
  local, 47, 51
  multiple, 53, 62, 71
  optimal global, 45
  optimal local, 46
  pairwise, 42
alignment score, 48
alignment scoring function, 43
allele, 80
alternative hypothesis, 32
Altschul, Stephen, 42, 51
amino acid, xiii, 44
amino acid alphabet, 24
amino acid sequence, 7
amino acid substitution matrix, 82
amino acids, 23
aniridia, 38, 60
Anopheles gambiae, *5*
approximate alignment, 51
Arabidopsis thaliana, *2*, *159*, *160*, *167*
Arber, Werner, 16
Archaea, 4
Arginine, 24
Asparagine, 24
Aspartic acid, 24
Avery, Oswald, xiv
Axel, Richard, 61
AZT, 97

baker's yeast, 1
base, 3, 4
best reciprocal hit, 132
binding site, 16, 160, 161
bioinformatics, xii
biological validation, 171
Birney, Ewan, 22
BLAST, xv, 40, 51, 60
BLOSUM matrix, 45, 83
blue whale, 95
bonobo, 78, 85, 93
bootstrap, 35
branch length, 115
branches, 113
BRH, *see* best reciprocal hit
Brown, Pat, 3, 143
buchnera, *136*
Buck Linda, 61

Caenorhabditis elegans, 5, *21*, *22*
cancer, 80
Cantor, Charles, 88, 95
Cavalli-Sforza Luigi Luca, 95
CCA1, 160
CCR5, 107
cDNA, 144
cDNA microarray, 145
central dogma, 25
change point analysis, 13, 62
chaos game representation, 15
chimpanzee, 78, 85, 93, 108
Chlamydia, xvi, *4*
Chlamydia pneumoniae, *129*, *131*
Chlamydia trachomatis, *2*, *129*, *131*
chloroplast, 5, 129
chromosome, 3, 30
Churchill, Gary, 63
circadian clock, xvii, 159
CLUSTALW, 54, 60
clustering, 151
  k-means, 151
  hierarchical, 151
codon, 26, 81
  start, 29, 33
  stop, 26, 28, 33
codon bias, 34, 100
codon distribution
  non-uniform, 34
  uniform, 34
codon sequence, 7
codon structure, 69
comparative genomics, 130
comparison-based methods, 29
complementarity, 3, 30
consensus sequence, 163
coronavirus, 111
correlation coefficient, 151
cow, 95
Crick, Francis, xiv, 21, 27, 37
CRS, 84
Cysteine, 24

D-loop, 84
Danio rerio, *5*
database searching, 40
Dayhoff, Margaret, 82
DDBJ database, 18
degradase, 23
delta method, 90
dendrogram, 155
diauxic shift, 141, 145, 147, 153
dimer, 14
dinucleotide, 14
diploid, 79
directed tree, 115
distance matrix, 117
distance measures, 151
distance-based methods, 117
DNA, xiii, xiv, 3, 26
  regulatory, 161
DNA alignment, 44
DNA fingerprinting, 81
DNA microarray, 142
DNA sequence, xiv, 3, 6, 7, 79
Drosophila melanogaster, *5*, *38*, *53*
drug resistance, xvi
duplications, 81
dynamic programming, 46, 56, 62, 75

E-value, 52
E. coli K12, *2*, *13*, *128*
elephant, 94
EM algorithm, *see* expectation-maximization algorithm
EMBL database, 18
emission matrix, 65, 66
emission probability, 63
ENV, 107, 121, 127
enzyme, xiii
epitopes, 98
EST, 144
Eubacteria, 4, 161
Eukaryote, xiii, 69
evening element, 167
evolution, 97, 98
exceptions, xii, 26
exon, 29, 70

CAMBRIDGE

Cambridge University Press
978-0-521-85603-4 - Introduction to Computational Genomics: A Case Studies Approach
Nello Cristianini and Matthew W. Hahn
Index
More information

expectation-maximization algorithm, 66, 76, 166
exploratory data analysis, 148
external node, 113
eyeless gene, 38, 41, 46, 49, 60

false negative, *see* Type II error
false positive, *see* Type I error
fast heuristic methods, 42
FASTA, 18, 51
FASTA format, 19
footprinting, 136
forward algorithm, 66, 76
frame shift mutation, 29
function
  prediction, 40, 62

GAG, 108
GC content, 12
Gehring, Walter, 38
GenBank, 18
gene, xiv, 3
gene duplication, 41, 130
gene expression, xv
gene family, 41, 132, 133
gene finding, 29, *see* gene prediction, 48, 62, 69
  prokaryotic, 22
gene function, 154
gene-finding, 28, 40
  prokaryotic, 29
GeneChip, 147
genetic algorithms, 98
genetic code, 26
  mitochondrial, 37
genetic distance, 86, 87, 88, 89, 91
genome, 3, 6
  annotation, 1
  eukaryotic, 4
  organelle, 5
  prokaryotic, 4
  viral, 4
genome signature, 15
germline mutation, 80
gibbon, 93
Gibbs sampling, 165, 166
Gibson, Toby, 54
Gish, Warren, 42
global alignment, 42, 43, 49
  optimal, 45
Glutamic acid, 24
Glutamine, 24
Glycine, 24
Gojobori, Takashi, 102
gorilla, 85, 93
gp120, 107
gp160, 107
gp41, 107
graph
  cyclic, 67
GTR model, 92

Haemophilus influenzae, *xv*, *1*, *2*, *4*, *10*, *12*, *14*, *16*, *18*, *21*, *25*, *31*, *35*, *36*
haplotype, 85
Harmer, Stacey, 167
Haussler, David, 63
heat maps, 155
helix, double, 3
hemoglobin, 23
hidden Markov model, xv, 13, 62, 63, 65, 67, 69
hierarchical clustering, 133
Higgins, Desmond, 54
Histidine, 24
HIV, 96, 127
HIV genome, 105
Holley, Robert, 26
homeobox domain, 38, 41
Homeobox, 52
Homo erectus, 78, 79, 93
Homo habilis, 78
Homo neanderthalis, 78, 85
Homo sapiens, *xv*, 2, 5, 79, 85, 92
Homo sapiens mitochondrion, *18*
homolog, 38, 40
homologous sequence, 67
homology, 40, 48, 51, 53
  multiple levels of, 39
homology-based methods, 29
horizontal gene transfer, 12
Hotel Metropole, 111, 121, 122
hotspot, 52
HOX, 38
HOX domain, 49
HOX gene, 49
human genome sequence, 2
human mtDNA, *2*
HVR, *see* hypervariable region
hybridization, 144
hydrophilic, 72
hydrophobic, 72
hydrophobicity, 82
hypervariable region, 18, 84, 92, 100
Hypothesis Testing, 31

i.i.d, 7, 12
image processing, 146
immune system, 97, 98
indel, 41, 81
insulin, 23
intergenic region, 136
internal node, 113
intron, 29, 63, 69, 70
inversions, 81
Isoleucine, 24

Java Man, 78
Jukes, Thomas, 88, 95
Jukes–Cantor formula, 88, 90, 91

k-mers, 14
Ka/Ks Ratio, 100
Khorana, Gobind, 26, 37
Kimura Model, 91
Kimura, Motoo, 91, 95, 100
kinking, 16
KL divergence, 165

Leakey, Louis, 78
leaves, 113
length of a tree, 118
Leucine, 24
LHY, 160
ligase, 23
likelihood, 8
Lipman, David, 42
loaded dice, 64
local alignment, 42, 47, 51
  optimal, 46
locus, 19
Lysine, 24

mad cow disease, 25
malaria, 96
mammoth, 94
manganese-dependent superoxide dismutase, 25
Markov chain, 8, 9, 88
Markov DNA sequence, 9
Markov model, 7, 9, 48
master regularity gene, 38
matrix
  substitution, 82
maximal segment pair, 51
MEGA, 109
MEME, 172
Mendel, Gregor, xiv
metabolism, xiii
methionine, 24, 28

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-85603-4 - Introduction to Computational Genomics: A Case Studies Approach
Nello Cristianini and Matthew W. Hahn
Index
More information

microarray, xv, 144
  one-dye, 147
  two-dye, 146
  cDNA, 145
microsatellites, 81
midpoint rooting, 120
Miller, Webb, 42
missing data, 150
mitochondria, 5, 26, 129
motif, 15, 16, 54, 161
  ungapped, 163
motif bias, 14
motif finding, 160, 164
mouse, 108, 160
mRNA, 26, 28, 70
mtDNA, 5, 82, 83, 85, 95
multidimensional scaling, 92, 124
multinomial model, 7, 9, 16, 48
multiple alignment, 53, 62, 71
multiple testing, 36
MUMmer, 140
Mus musculus, *5*, *53*
mutation, 79, 80, 81, 82, 86, 97
  fixed, 86
  germline, 80
  neutral, 86
  point, 80
Mycoplasma genitalium, *1*, *2*, *4*, *18*, *21*, *22*, *31*, *35*, *36*, *129*, *138*
Mycoplasma pneumoniae, *138*
Myers, Gene, 42

Nathans, Dan, 16
natural selection, 97
Neanderthal, xv, 19, 78, 85, 92, 93, 95, 127
Needleman, Saul, 42
Needleman–Wunsch algorithm, 45, 47, 51, 54, 55, 58, 60
Nei, Masatoshi, 102, 118
Nei–Gojobori Algorithm, 102, 105
neighbor-joining algorithm, 117, 118, 119, 120
neighborliness, 119
network
  phylogenetic, 112
neurogenesis, 49
neutral mutation, 86
Neutral Theory of Evolution, 100
Newick format, 115, 126
Nirenberg, Marshall, 26, 37
NJ algorithm, *see* neighbor-joining algorithm
non-synonymous substitution, 99, 101
nucleotides, 3, 4, 6, 8
null hypothesis, 32

odds ratio, 16
odorant receptors, 61, 69, 71, 72, 127
olfactory system, 61
open reading frame, 29, 30
optimal global alignment, 45
optimal local alignment, 46
orangutan, 85, 93
ORF, *see* open reading frame, *see* Open Reading Frame, 132
ORF finding, 69, 106
ortholog, 41
orthologous genes, 41
orthology, 41, 131
Oryza sativa, *5*
out of africa, 93
outgroup, 113
oxidative phosphorylation, 83

p-value, 32
paired-box domain, 38, 42
pairwise alignment, 42
palindrome, 17
palm civet, 111, 121
PAM matrix, 45, 83
paralogous genes, 41
paralogy, 41, 131
parasite, 129
pathway, xiv
pattern detection, 62
pattern discovery, 18, 166
pattern matching, 18, 166
PAX, 38
PAX domain, 49
PAX gene, 49, 50
Peking Man, 78
permutaton test, 35
phage, 1, 2, 13, 18, 20, 66, 77
Phenylalanine, 24
pheromone, 61
pHMM, *see* profile HMM, 68, 71
Phylip, 127
phylogenetic analysis, 114
phylogenetic footprinting, 136
phylogenetic network, 112
phylogenetic tree, xvi, 93, 112, 116
phylogenetics, 112
polymerase, 108, 161
polymorphism, 80,
polypeptides, 24
polyploid, 130
prediction of function, 40, 62
primate evolution, 93
prion, 25
probabilistic model, 6
probability, 8
  conditional, 9
  joint, 9
  transition, 8
probes, 144
profile, 68
profile HMM, 68
Prokaryote, xiii, 4
Proline, 24
promoter, 30, 163
protein, xiii, 23, 26
  domains, 41
protein coding, 60
protein fold prediction, 24
protein folding, 24
protozoa, 128
Prusiner, Stanley, 25
pseudocount, 165
pseudogenes, 61
PSSM, 68, 163, 166
PSWM, *see* PSSM
purines, 81
pyrimidines, 81

randomization test, 34
reading frame, 28
recombination, 116
reference design, 147
regulatory DNA, 161
repeats, 169
replicase, 121
restriction enzyme, 16
restriction site, 16
reversal distance, 137
reversal sort algorithm, 138
rhodopsin, 71
ribosome, 26, 28
RNA, 26
RNA sequence, 7
RNA-coding genes, 69
rooted tree, 113

sabre-tooth tiger, 95
Saccharomyces cerevisiae, *1*, *2*, *5*, *21*, *141*, 142
Saitou, Naruya, 118
Sanger, Fred, 1, 21
SARS, xvi, 2, 108, 110, 115
saturation, 87

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-85603-4 - Introduction to Computational Genomics: A Case Studies Approach
Nello Cristianini and Matthew W. Hahn
Index
More information

scoring function, 44, 164
  alignment, 43
segment pair, 51
  maximal, 51
segmentation, 62, 64
selection
  negative, 97
  positive, 97
  purifying, 97
sequence alignment, xv, 3, 39
sequence assembly, 40
sequence divergence, 40
sequence evolution, 88
sequence polymorphism, 80
sequence similarity, xv, 40
Serine, 24
Sharp, Paul, 54
significance level, 32
significant, 32
similarity, 51
Single Nucleotide Polymorphism, *see* SNP
single nucleotide polymorphisms, xvi
SIV, 127
sliding window, 107
Smith, Hamilton, 16
Smith, Temple, 42
Smith–Waterman algorithm, 47, 49, 50, 54, 58, 59, 60
SNP, 80, 81
sorting by reversals, 136
speciation, 41, 131
Spike gene, 121
splice sites, 70
standard nucleotide alphabet, 20
start codon, 29, 33
stop codon, 26, 28, 33, 34, 103
STR, 81
string, 7
subsequence, 7
substitution, 82, 86
  matrix, 44
  multiple, 87
  non-synonymous, 99
  synonymous, 99
substitution matrix, 82
substitution rate, 86
substring, 7
symbiont, 128
synonymous substitution, 99, 101
synteny, 135, 136

T Cells, 98
tabular computation, 56
tandem repeats, 81
TATA-box, 161
taxa, 112
test statistic, 32
theorem of total probability, 66
Thompson, Julie, 54
Threonine, 24, 27
tobacco, 160
TOC1, 160
topology, 113
trace-back, 57
transcription, 26
transcription factors, 160
transition matrix, 9, 65, 66
transition probability, 63
transitions, 81
translation, 26
transmembrane domain, 61
tranversions, 81
tree, 116
  additive, 117
  bifurcating, 113
  directed, 115
  oriented, 114
  phylogenetic, 112
  rooted, 113
  unrooted, 113
Tryptophan, 24
Type I error, 32, 33
Type II error, 33
Tyrosine, 24

ultrametricity, 120
ungapped motif, 163
unrooted tree, 113
untranslated regions, 28
UPGMA, 120
upstream region, 163
Urbani strain, 122
Urbani, Carlo, 110

Valine, 24
variation between species, 86
Venter, Craig, 1
visualization, 155
Viterbi algorithm, 66, 75

Waterman, Michael, 42
Watson, James, xiv, 21, 37
whole-genome analysis, xvi, 5
wine, 141
Wunsch, Christian, 42

yeast, 141, 142, 143, 172

zebrafish, 5

© Cambridge University Press www.cambridge.org