

Index

Locators for headings with subheadings refer to general or introductory aspects of that topic.
Locators in *italics* refer to pages/page ranges containing relevant figures and diagrams.

- a posteriori trees 28
- ABA multiple alignment program *81*
- abbreviations, amino acid 4
- acceptance regions 382
- accession numbers (AC) 36
- adaptation of molecular sequences *see* molecular adaptation
- adaptive evolution 13
- adaptive landscape models 13–14
- ADAPTSITE software package 453
- additive distance trees 148, 654, 659
- adenine 4
- affine gap penalties 73, 654
- AIC criteria (Akaike Information Criteria) 234, 305, 431
 - definitions 654, 663
 - molecular clock analysis 375, 381
 - selecting models of evolution 349, 352, 353
- Align-m multiple alignment program *81*, 91
- alignment 18, 45–47
- AllAll multiple alignment program *81*
- alleles 8, 9, 11
- allopatric speciation 15–16, 654, 665
- AltAVisT web alignment comparison tool 103–104
- alternative hypotheses 382
- Amap multiple alignment program *81*
- amino acids 4, 579
- ancestral recombination graph (ARG) 508, 654, 659; *see also* geneology
- Anemonia sulcata* (sea anemone) 154
- aneuploidy 8
- annotation 33
- approximately unbiased (AU) test 392–393, 403, 404
- ARG (ancestral recombination graph) 508, 654, 659; *see also* geneology
- assumptions, implicit/explicit 345–346
- asymptomatic normality 379, 435–436, 505, 654
- AU (approximately unbiased) test 392–393, 403, 404
- autapomorphy 654
- autocorrelated relaxed clocks 371, 654
- automated bootscanning 513
- average-over-window-based scanning approaches 523, 525
- background selection 562, 654
- backtracking algorithms 46
- bacteria 8, 494
- balancing selection 10, 411, 596, 654, 659
- BaliBase test case collection 89, 90, 92, 92–93
- BaliPhy multiple alignment program *81*
- Base-by-Base multiple alignment editor 97
- base trees 355
- Bayes factor (BF) 233–235, 351, 443–444, 574
 - definition 655
- Bayes, Thomas 211
- Bayesian analysis 27, 210–216; *see also* BEAST software; Markov Chain Monte Carlo sampling
 - burn-in 224
 - convergence 224–227
 - LAMARC 595
 - Markov Chain Monte Carlo sampling 220–224, 381, 564
 - Metropolis coupling 220, 227–229
 - mixing behavior 224–226
 - model choice/model averaging 232–235
 - phylogenetic inference 216–220, 640
 - phylogenetic models 230–232, 233–234
 - practice exercise *see* MrBayes software
 - prior probability distributions 236
 - protein sequences 330
 - recombination detection 512
 - selection pressures on alignment 438
 - summarizing results 229–230
 - trace plot 224
 - Trypanosomatid phylogeny 338–342
- Bayesian Information Criteria *see* BIC

- BEAST (Bayesian evolutionary analysis by sampling trees) software 564–566, 576
 amino acid sequences, translating data 579
 analysis of output 583–586
 Bayesian MCMC for geneology-base population genetics 566–569
 burn-in 583, 588–589
 computational performance 568–569
 conclusions 575, 590–591
 definitions/units 572
 discussion 569–575
 divergence time estimation 570–571
 evolutionary models 564, 579
 FigTree software program 590
 input/output format 568
 Markov Chain Monte Carlo 564, 581–582
 maximum likelihood methodology 564
 model comparison 572
 molecular clock models 570
 multiple data partitions 572
 NEXUS file loading 577
 parameters 572, 573–574
 posterior probability limit 589
 rate models among branches 570–571
 relaxed molecular clock models 565
 running analysis 582–583
 setting dates of taxa 577–579
 setting evolutionary model 579–580
 setting MCMC options 582
 setting operators 580–581
 substitution models 570
 trace plot 584, 585, 586, 587
 tree priors 571
 tree summarizing 586–589
 TreeAnnotator tool 588
 viewing annotated tree 589–590
 web-based aids 591
 BEAUti software package 572, 573–574, 576–582
 BF *see* Bayes factor
 bias 317, 431, 615, 655
 BIC (Bayesian Information Criteria) 234, 305, 351–353, 381, 655
 definitions 655, 663
 BioEdit multiple alignment editor 97
 BIONJ 154
 BLAST (Basic Local Alignment Search Tool) 33, 47–50, 52–54
 database searching 62–66
 definition 655
 neighborhood word list 48
 protein sequences 320, 321, 327–328, 335
 Trypanosomatid phylogeny 339
 BLAT database search tool 52
 BLOSUM substitution matrices 45, 327, 335
 definitions 655, 662
 multiple sequence alignment 72, 73, 76, 88
 bootscanning 198, 511, 513, 527, 528
 bootstrap analysis 28
 definitions 655, 660, 665, 667
 detecting recombination 527, 528
 maximum likelihood methodology 194
 mtDNA data set 170–174, 175–176
 pairwise distance methods 156, 157–159
 parsimony analysis 268, 298
 protein sequences 332
 selection pressures on coding sequences 426
 split networks 640, 644, 646–647
 testing tree topologies 381, 386–387
 bottlenecks, genetic 11, 497, 659
 branch and bound technique 80, 90, 280–282, 655
 branch-by-branch variation estimation 438, 470–475
 branches, phylogenetic trees 20
 branch-swapping 283
 breakpoint detection 507, 512
 Bremer support 299; *see also* decay index
 Buneman index 634, 655, 660
 Buneman tree 646, 647, 655
 split networks 634–636, 639
 burn-in 27, 340, 583, 588–589, 611
 definitions 656, 657
 centering 386
 Chaos–Dialign multiple alignment program 81
 character-based methods 330
 character-state methodology 23–24
 chimaeras, bacterial 494
 cladograms 23, 164, 241, 460
 definitions 656, 666
 Clustal multiple alignment programs 81, 88–89, 93, 95
 file formats/availability 100–101
 protein sequences 325, 327, 330, 336
 sequence data input format 61
 TRIM5 α gene sequence alignment 100–102
 cluster analysis/clusters 20, 24–25, 144–148, 155
 coalescent theory/coalescents 508, 551–552, 592, 656
 advanced topics 561–563
 coalescent-based inference 558–559
 coalescent events 12
 demographic history 556–558
 effective population size 554–555, 656
 Kingman coalescent 552–554
 mutation clock 555–556
 serial coalescent 559–561
 Cobalt multiple alignment program 81, 94
 codon
 substitution models 424–426, 471
 usage bias 317, 656
 volatility 417–418, 656
 coin tossing experiment 182
 cold chains 227, 228, 656
 compatibility matrix 511
 compatibility methodology 506
 complementary hydrogen-bonded pairs 4
 confidence intervals 382, 435–438
 confidence sets 394–395
 conjugation 494, 656; *see also* recombination
 CONSEL software package 202, 397–398, 401–403, 404

- consensus trees 298; *see also* supernetworks
- consistency 426, 615, 656
- consistency-based scoring 89, 656
- constraint trees 300
- contraAlign multiple alignment program 81
- convergence 224–227
- convergent evolution 19, 632, 656, 665
- converse constraints 300
- covarian/covariotide models 232
- credibility intervals 229, 656, 657
- credible sets 229, 657
- criterion-based methods 267–268, 301–302
- cytosine 4

- DAD database 39
- DAMBE software, assessing substitution saturation
 - invertebrateEF1a.FAS example 628–629
 - problem of substitution saturation 615–616
 - simian immunodeficiency genomes example 629–630
 - Steel's method 616–621
 - vertebrate mitochondrial genomes example 624–628
 - Xia's method 621–623
- Darwin, Charles 13
- databases 19, 33, 34, 106–108
 - BLAST searching 62–66
 - composite 39–44
 - ENTREZ 39, 41, 42–43, 55–57
 - errors 38
 - general nucleic acid sequence 35–37
 - general protein sequence 37–39
 - other tools 52–54
 - search strategies 33
 - searching by sequence similarity 45–54
 - sequence data input formats 59–62
 - specialized/secondary 39, 40–41
- Datamonkey software package 454
- dated tips, molecular clock analysis 367–368, 377–380
- DCA (Divide and Conquer) program 82, 91
- DDJB (DNA Data Bank of Japan) 19, 35, 324
- DeBruijn dualization 639
- decay index 299, 657
- delete-half jackknifing 159; *see also* jackknifing
- deletion errors 7
- deoxyribonucleic acid *see* DNA
- deterministic evolution 12
- deterministic forces 11
- DiAlign multiple alignment programs 82, 91–92, 95
- differences in sums of squares methodology 511
- different rates (DR) models 367–368, 377–379
- diploidy 9, 409, 551
 - definitions 657, 663, 670
- directional selection 422, 496, 562, 596, 657
- discrete character states 23
- disruptive selection 16, 657
- distance-matrix methodology 23, 24, 25, 512; *see also* genetic distances
- divergence time 21–22
- diversifying selection *see* positive selection
- Divide and Conquer (DCA) program 82, 91
- DNA (deoxyribonucleic acid) 3–4
 - database of Japan (DDJ) 19; *see also* databases
 - vs. protein analysis 323–324
- DNAML program 191–192
- dN/dS ratio
 - definition 657
 - Influenza A case study 469–470, 473, 475, 480, 486
 - molecular adaptation 413, 415–417
 - and recombination 506
 - selection pressures on coding sequences 429–431
- dot-matrix sequence comparison 74–75, 325–326
- dynamic programming 46, 73, 277
 - definition 657
 - multiple sequence alignment 75–79

- Edman degradation method 37
- effective population size (EPS) 11
 - Bayesian evolutionary analysis 571
 - coalescent theory 554–555
 - definitions 656, 657, 663, 666, 670, 671
 - LAMARC 592, 593
 - molecular adaptation 413
 - molecular clock analysis 363
 - recombination detection 496
- effective sample size (ESS) 226, 437, 583, 611
 - definitions 657, 663
- efficiency 426, 615, 658
- electrophoretic alleles 658
- ELW (expected likelihood weights) 394–395, 400, 404
- EMBL (European Molecular Biology Laboratory) 19, 35, 36, 38, 324
- empirical Bayes approach 443, 658
- enhancers 3
- Ensembl database 40
- ENTREZ databases 39, 41, 42–43, 55–57
- EPS *see* effective population size
- ESS *see* effective sample size
- Essay Towards Solving a Problem in the Doctrine of Chances* (Bayes) 211
- EST (expressed sequence tag) databases 323
- European Molecular Biology Laboratory (EMBL) 19, 35, 36, 38, 324
- evolution
 - deterministic 12
 - fossil record 16
 - implications of recombination on 496–498
 - protein 315–324
 - units of 420
 - viruses 3, 8, 9
- evolutionary
 - distances *see* genetic distances
 - models *see* model of evolution
 - networks 514, 631–633
 - rates 9–10, 157, 421, 658, 664
- exhaustive searches 279

- exons 8
- expected likelihood weights (ELW) 394–395, 400, 404
- exponential distributions 115
- expressed sequence tags (EST) 37, 658
- Expresso multiple alignment program 82, 93
- extant taxa 20
- external nodes, phylogenetic trees 20
- extreme value distributions 49, 658

- FastA software 33, 50–54, 327
 - database searching 66–67
 - definition 658
 - sequence data input format 59
- FASTDNAML program 191–192
- Felsenstein pruning algorithm 188, 189
- finches, Galapagos Island 3, 13
- FigTree software program 166–168, 590
- fit index 639, 643–644, 658
- Fitch algorithm 279
- Fitch–Margoliash method 170, 171, 26, 330, 365
- fitness 409, 419, 665, 658
 - survival of the fittest 13
- fixation/fixation rates 9–10, 409, 658
- fixed effects likelihood (FEL) models 442, 445–446, 478–481, 658
- fluctuate (coalesce) function 598
- fluctuating selection 411, 658
- footprints, molecular 412–413
- fossil record 16
- founder effects 497
- fourfold degenerate sites 7
- four-point metric condition 148, 659
- frequency-dependent selection 411, 659

- G3PDH protein, model of evolution 361
- GA *see* genetic algorithms
- Galapagos Island finches 3, 13
- gap penalties/gap length penalties 45
- gaps problem, multiple sequence alignment 72–74
- GDE multiple alignment editor 97
- GenBank 35, 36, 37, 58, 324
- gene/s
 - database 42
 - duplication 8
 - frequencies 9
 - ontology (GO) 44
 - trees 21
- Genedoc multiple alignment editor 97
- geneology 16, 508, 551, 592; *see also* ancestral recombination graph (ARG)
 - definitions 659, 663, 664
- geneology-based population genetics 565
- general time reversible (GTR) models 121, 122, 174
- generation times 10, 369, 560, 571, 608
 - definition 659
- genetic algorithms (GA)
 - definition 659
 - maximum likelihood methodology 194
 - multiple sequence alignment 90–91
 - recombination detection 513
 - selection pressures on coding sequences 442
- genetic bottlenecks 11, 497, 659
- genetic code, universal 5, 6, 316–318
- genetic distances 24, 111–112
 - choosing among different evolutionary models 140–141
 - definitions 655, 659, 662, 666
 - detecting recombination 504, 513, 525–528
 - estimating standard errors using MEGA4 135–137
 - impact on tree morphology 174–180
 - Jukes and Cantor models 117–118, 128–131
 - Kimura 2-parameters, 131–132
 - Markov models 116–118
 - Markov process derivation 118–121
 - maximum likelihood estimators 113, 121
 - molecular clock analysis 362
 - nucleotide substitution models 112, 121–125
 - nucleotide substitution saturation 137–140
 - observed distance 112–113
 - Poisson process 113–116
 - protein sequences 323, 330
 - Q matrices 116, 117
 - rate heterogeneity 123–125, 133–135
 - selection pressures on coding sequences 426–429
 - software packages 126–128
 - split networks 634
 - substitution saturation 616
 - TN93 model 132–133
- genetic drift 12, 314, 315, 555
 - definition 659
 - molecular sequences 407–408, 410
- genetic exchange 8; *see also* recombination
- genetic hitchhiking (hitchhiking selection) 497, 562, 660
- GenInfo (GI) number 37
- Genious multiple alignment editor 97
- genome 3
- Genome Survey Sequences (GSS) 37, 659
- GenPept database 39, 324–325
- GI (GenInfo) number 37
- global alignment 46
- GO (gene ontology) 44
- goodness-of-fit criterion 24, 267
- graphical recombination detection 511
- greedy stepwise addition 282, 283, 284
- GSS (Genome Survey Sequences) 37, 659
- GTR (general time reversible) models 121, 122, 174
- guanine 4

- HAMAP database 40
- haploidy 409, 551, 571
 - definitions 659, 663, 670
- Hastings ratio 659
- HCV database 40
- heated chains 227, 228, 656
- heterologous (non-homologous) recombination 519, 664
- heterotachy 615, 659

- heterozygosity 9, 555, 659
- heuristic strategies 24, 190, 510
- hidden Markov models (HMMs) 72, 91, 659
- hierarchical likelihood ratio tests (hLRTs) 347, 348–349, 350, 352
- High Quality Automated and Manual Annotation of microbial Proteomes (HAMAP) database 40
- highest posterior density (HPD) intervals 379, 583, 590, 657, 660
- hill-climbing algorithms 288
- hitchhiking selection 497, 562, 660
- HIV data set 9, 40
 - detecting recombination *see below*
 - drug resistance evolution 3
 - envelope gene example, model of evolution 360–361
 - likelihood mapping analysis 207
 - quartet puzzling consensus tree 202, 207–209
 - selection pressures on coding sequences 423
 - SplitsTree program 648–653
- HIV sequences, detecting recombination 521–523
 - analysis of example sequence 533
 - detecting recombination 530–531
 - existing tools for recombination analysis 532
 - RDP3 534–542 546–548
 - SIMPLOT 533–536, 537–538
- HLA database 41
- hLRTs (hierarchical likelihood ratio tests) 347, 348–349, 350, 352
- HMMs (hidden Markov models) 72, 91, 659
- HOGENOM (Homologous sequences from complete Genomes) database 40
- homogeneity 116, 660
- homology/homologous
 - chromosomes 8
 - definition 660
 - genes 551
 - pairs 657
 - phylogeny/homology 17, 19
 - recombination 494, 519, 660; *see also* recombination
 - sequence residues 68; *see also* multiple sequence alignment
 - sequences 33, 40, 318, 421, 654
 - sites 615
- Homologous Vertebrate Genes (HOVERGEN) database 40
- homoplasy 19, 269, 496, 511, 660
- homozygosity 9
- house-keeping proteins 314–315
- HOVERGEN (Homologous Vertebrate Genes) database 40
- HPD (highest posterior density) intervals 379, 583, 590, 657, 660
- HTUs (hypothetical taxonomic units) 20
- human histocompatibility locus A (IGMT/HLA) database 41
- hydrophobicity/hydropathy profiles, protein 329
- hyper-geometric random walks 513
- hyperprior 214, 660
- HyPhy analyses/software package
 - automating choices 487–488
 - codon model, fitting local 471
 - detecting recombination 467
 - estimating global rates 467–470
 - fixed effects likelihood 480
 - importing/manipulating sequence data files 460
 - Influenza A case study 455–456, 487–488
 - inspecting likelihood function 472
 - molecular clock analysis 374
 - previewing sequences 458
 - previewing trees 461
 - random effects likelihood methods 481–483
 - selection pressures on coding sequences 453
 - sequence alignment 462
 - summary of standard analyses 488–490
- hypothetical taxonomic units (HTUs) 20
- identifiers (ID) 35–36
- IGMT/HLA database 41
- Immunogenetics IGMT databases 41
- indels 7, 73
- infinite sites models 413, 496, 660
- Influenza A case study 454–455, 490; *see also* HyPhy software package
 - α/β estimation 477
 - branch-by-branch variation estimation 470–475
 - comparison of internal and terminal branches 474–475
 - detecting recombination 465–467
 - dN/dS ratio estimation 469–470, 473, 475, 480, 486
 - estimating global rates 467–470
 - estimating nucleotide bases 464–465
 - estimating trees 462–464
 - fitting local codon model 471
 - fixed effects likelihood 478–481
 - gene-by-gene variation rate estimation 484
 - HyPhy analyses 455–456, 487–488
 - importing alignments and trees 456–457
 - inferring neighbor-joining tree/estimating branch lengths 464
 - inspecting likelihood function 472
 - interclade variation in substitution rates 473–474
 - nucleotide substitution models, comparison 466
 - preliminary analysis set-up 476
 - previewing sequences 457–459
 - previewing trees 459–461
 - random effects likelihood methods 481–483
 - reassortment 495
 - selection between different genes comparison 485–487
 - selection in different populations comparison 484–485
 - selection pressures between different clades 474

- Influenza A case study (*cont.*)
 selection pressures on coding sequences 421–422
 sequence alignments 461–462
 simulations 488
 single likelihood ancestor counting 477–478, 479, 480
 site-by-site variation estimation 475–476
 summary of standard analyses 488–490
 UPGMA 462
- information criteria 349–351; *see also* BIC
 information entropy 616, 660
 information releases, sequence databases 35
 informative sites 506, 512, 665
 ingroups 20
 insertion errors 7
 INTERALIGN multiple alignment program 82
 internal nodes 20
 internet searches 332–337; *see also* databases;
 web-based aids
 introns 8, 322–323
 invertebrateEF1a.FAS data set 628–629
 IQPNNI software packages 192–193, 199–202, 203–207
 IRMBASE test case collection 93
 island model of migration 557, 660
 isolation index 636, 655, 660, 668
 iterative refinement methods 90
- jackknifing 195, 268, 332, 381
 definition 660
 pairwise distance methods (p-distance) 156, 159
 JalView multiple alignment editor 98
 joint probabilities 219, 661
 Jukes and Cantor (JC69) models 117–118, 128–131, 174
 jumping of recombinant isolates 523, 524
 jumping profiles 513
- Kalign multiple alignment program 82
 keyword searching 33, 44
 Kimura 2-parameters (K80)/F84 genetic distances, 131–132
 Kingman coalescent 552–554
 Kishino–Hasegawa (KH) test 388–389, 400, 403
- Laboratoire d'ImmunoGénétique Moléculaire (IGMT/LIGM) 41
 LAMARC software 592–593, 598, 599, 666
 Bayesian vs likelihood sampling 595
 conclusions 611–612
 confidence intervals 608
 converting data/file converter 604–605
 definition 661
 estimated parameter values 611
 evolutionary model 596
 fluctuate (coalesce) function 598
 maximum likelihood estimators 594, 608
 Metropolis–Hastings MCMC sampler 593–595
 migrate-N function 598–599
 multiple loci 602–603
 output analysis 607–611
 population parameters, estimating 605–607
 population relative to sample size 597
 practice exercise 603–604
 program/package distinctions 600
 random sample 595–596
 rapid growth rates 603
 recombine function 599–600
 run time 597
 sample size considerations 601–602
 sequential samples 603
 space and time 601
 stability 596
 starting values 600–601
 trace plot 610–611
 trees discarded 611
 virus-specific issues 602
- lateral gene transfer 8
 lazy subtree rearrangement 193
 least-squares fit 156
 leaves 20
Leishmania mexicana enzyme analysis 332–337
 likelihood 181, 269, 565; *see also* maximum
 likelihood
 analysis 511
 calculations 27
 definitions 654–655, 661–662, 664, 668, 670
 function 182, 212, 347, 425
 mapping analysis 196–198, 207, 208
 methodology 196–198, 207–208, 327, 594
 scores 422
 values 381, 667
 likelihood ratio test (LRT) 194, 234, 305, 365, 609
 definition 661
 molecular clock analysis 365–367
 and recombination 505, 514
 selection pressures on coding sequences 433–434
 statistic 384, 431, 661
 testing tree topologies 384–386
 lineage sorting 22, 23
 linkage disequilibrium 495–496, 661
 local alignment 46
 local clock model 168, 369, 661
 local-search methods 283
 locus, genome 9
 log odds matrices 45, 327, 662
 long-branch attraction 27, 158, 269, 332, 615
 definition 662
 look-back time 318–320
 LRT *see* likelihood ratio test
- MACAW multiple alignment program 82
 MacClade multiple alignment editor 98
 MAFFT multiple alignment program 82, 90, 95
 majority rule consensus trees 230, 341, 568, 662
 MAP multiple alignment program 83
 marginal density 566
 marginal probability distribution 218, 662
 Markov Chain Monte Carlo (MCMC) sampling 27–28

- Bayesian sampling 220–224, 381, 566–569
- BEAST software 564, 581–582
 - definitions 656, 662, 663, 669, 670
 - geneology-base population genetics 566–569
 - LAMARC 593
 - maximum likelihood methods 193
 - molecular clock analysis 371, 379–380
 - protein sequences 338–342
 - selecting models of evolution 347
 - selection pressures on coding sequence
 - alignments 438
 - testing tree topologies 384
- Markov models 116–118, 174, 660, 662
- Markov process 118–121, 424, 569, 659, 662
 - time-homogenous stationary 116, 118, 120, 140
- mass spectrometry 37
- Match-Box multiple alignment program 83
- MAVID/AMAP multiple alignment program 83
- MAXCHI substitution distribution method 529, 540
- maximum chi-squared test 510
- maximum likelihood 183, 337, 626, 662, 667
- maximum likelihood estimators (MLEs) 183
 - BEAST software 564
 - coalescent theory 558
 - definitions 662, 666
 - distance methods 142
 - genetic 113, 121
 - LAMARC software 594, 608
 - molecular clock analysis 379, 380
 - mtDNA data set 185
 - selecting models of evolution 347
- maximum likelihood methodology 24, 25, 27, 181–184
 - assumptions 345
 - BEAST software 564
 - branch support 194–195, 203–207
 - complex case 185–186
 - computing probability of alignment for fixed tree 186–187
 - distance methodology 154
 - DNAML/FASTDNAML 191–192
 - Felsenstein pruning algorithm 188, 189
 - finding maximum-likelihood tree 189–190
 - full-tree rearrangement 190–191
 - genetic algorithms (GA) 194
 - heuristic strategies 190
 - IQPNNI 192–193, 199–202, 203–207
 - likelihood mapping analysis 196–198, 207
 - ML tree construction example 199–202
 - molecular clock analysis 365
 - Mr Bayes software 234
 - parameter estimation 195
 - PAUP* program 267, 304–313
 - PHYML/PHYML-SPR 192
 - protein sequences 327, 330
 - quartet puzzling 195–196, 203–209
 - RAXML 193
 - selection pressures on coding sequences 420, 428–429
 - simple case: two sequences 184–185
 - simulated annealing 193
 - software packages 199
 - split networks 640
 - testing tree topologies 381
- maximum parsimony (MP) 26–27, 142, 187, 330, 640
 - definitions 657, 662
- McDonald–Kreitman test 417, 662–663
- MCMC *see* Markov Chain Monte Carlo (MCMC)
 - sampling
- MCMCMC *see* Metropolis coupling
- ME *see* minimum evolution
- measurably evolving populations (MEPs) 367, 559, 663
- median networks 633, 638, 640
- MEDLINE 42
- MEGA sequence data input format 61–62
- MEGA4 software package 160, 161
 - genetic distances 135–137, 174–180
 - selection pressures on coding sequences 453
- meiosis 8, 494, 663; *see also* recombination
- MEPs (measurably evolving populations) 367, 559, 663
- Metropolis coupling (MCMCMC) 220, 227–229, 656, 663
- Metropolis–Hastings (MCMC) sampling 220, 564–565, 593–595
 - definitions 659, 661, 663
- microsatellites 602, 663
- midpoint rooting 21, 166, 365, 663
- migrate-N function 598–599
- migration/migration rates 15, 557, 592, 661, 663
- minimum evolution (ME) 149, 268, 617
 - definitions 662, 663, 664
- mixing behavior 224–226, 585
 - definitions 657, 663, 670
- MLEs *see* maximum likelihood estimators
- MMDB (Molecular Modeling Database) 42
- model averaging 353–354, 663
- models of evolution 24, 345–346
 - Bayesian approaches 351–352
 - BEAST software 564
 - coalescents 558
 - definitions 659, 664
 - G3PDH protein example 361
 - genetic distance approaches 140–141, 142
 - hierarchical likelihood ratio tests 347, 348–349, 350, 352
 - HIV-1 envelope gene example 360–361
 - information criteria 349–351
 - LAMARC software 596
 - model averaging 353–354
 - model fit 346–347
 - model selection procedure 355
 - model selection uncertainty 352–353
 - MODELTEST program 355–358, 359
 - PAUP* 351, 356, 357
 - performance-based selection 352
 - PROTTEST 358–359, 361
 - selection pressures on coding sequences 420, 432

- models of evolution (*cont.*)
 - substitution models 217, 320, 660, 664, 670
 - testing tree topologies 381
 - vertebrate mtDNA model selection example 359–360
- model of (sequence/molecular) evolution 181, 594, 664
- MODELTEST program 355–358, 359, 381
- modified Sherman test 511
- molecular adaptation 407, 418; *see also* molecular evolution
 - balancing selection 411
 - basic concepts 407–411
 - codon volatility 417–418
 - comparative methods 413
 - fixation 409
 - genetic drift 407–408, 410
 - molecular footprint of selection 412–413
 - mutations 407, 408
 - positive selection 409, 410
 - summary statistic methods 413, 413–415
- molecular clock analysis 23, 362–363, 371–372; *see also* relaxed molecular clock models
 - BEAST software 565
 - coalescent 555
 - dated tips 367–368, 377–380
 - definitions 661, 664, 670
 - Fitch–Margoliash method 365
 - genetic distances 111, 144
 - likelihood ratio test 365–367
 - Markov Chain Monte Carlo 367, 379–380
 - maximum likelihood estimators/methodology 365, 379, 380
 - NEWICK sequence data input format 376–377
 - PAML 373–375, 377
 - PAUP* 374
 - proteins 363
 - and recombination 505
 - relative rate test 364–365
 - selecting models of evolution 346
 - TREE-PUZZLE software package 374
 - TRIM5 α gene sequence alignment 373–377
 - viral sequence analysis, RSVA data set 377–380
- molecular evolution; *see also* molecular adaptation
 - data, phylogenetic 16–19
 - databases 19
 - genetic information 3–8
 - multifurcating trees 28–29
 - phylogenetic trees 19–23, 23–28
 - population dynamics 9–14
 - rates 565, 667, 670
 - speciation 14–16
 - terminology 9–10, 16, 20
- molecular footprints 412–413
- Molecular Modeling Database (MMDB) 42
- monophyletic taxa 20, 154, 296, 568
 - definitions 655, 661, 664
- Monte Carlo simulation 349, 367, 387, 506; *see also* Markov Chain Monte Carlo sampling
 - definitions 664, 665
- MP *see* maximum parsimony
- MrBayes software 237
 - acquiring/installing software 237
 - analysis based on primates data file 240
 - analyzing partitioned data set 261–266
 - changing the window size 238–239
 - checking model 247–248
 - data entry 241–242
 - data partitioning 261–263
 - entering mixed data 261
 - getting started 238
 - help command 239
 - maximum likelihood analysis 234
 - model specification 242–244
 - quick start version 240–241
 - running analysis 252–254, 265
 - setting priors 244–247
 - setting up analysis 248–252
 - specifying partitioned model 263–265
 - stopping analysis 254–255
 - summarizing samples of substitution model parameters 255–257
 - summarizing samples of trees and branch lengths 257–261
 - useful tips 265–266
- MSA multiple alignment program 83
- mtDNA data set
 - bootstrap analysis 170–174, 175–176
 - Fitch–Margoliash method 170, 171
 - maximum likelihood estimators 185
 - models of evolution 359–360
 - SplitsTree program 642–648
 - substitution saturation, assessing with DAMBE software 624–628
- Multalin multiple alignment program 83
- multifurcating trees, molecular evolution 28–29
- Multi-Lagan multiple alignment program 83
- multi-model inference (model averaging) 353–354, 663
- multiple sequence alignment 68, 69, 87
 - algorithms/programs 79–80, 81–86, 87–92
 - BaliBase test case collection 89, 90, 92, 92–93
 - BLOSUM 62 matrix, 72, 73, 76, 88
 - choosing a program 93–95
 - consistency-based scoring 89
 - dot-matrix sequence comparison 74–75
 - dynamic programming 75–79
 - gaps problem 72–74
 - hidden Markov models software 72, 91
 - iterative refinement methods 90
 - multiple alignment algorithms/programs 79–80, 81–86, 87–92
 - nucleotide sequences vs. amino acid sequences 95–96
 - pairwise sequence alignment 74–79
 - PAM matrix 72, 88
 - practice exercise *see* TRIM5 α gene sequence alignment
 - progressive alignment 80, 87, 87–89
 - protein sequences 327–328
 - repeats problem 68–70

- substitutions problem 70–72
- T-COFFEE multiple alignment program 86, 89–90
- testing multiple alignment methods 92–93
- visualizing alignments/manual editing 96–99
- MUMMALS multiple alignment program 83, 91
- Murlet multiple alignment program 83
- MUSCA multiple alignment program 83
- MUSCLE multiple alignment program 84, 90, 95, 102
- mutation clock 555–556; *see also* molecular clocks
- mutations 6–8, 12, 13, 407, 408
 - synonymous/non-synonymous 7, 10–12, 133, 412, 506; *see also* dN/dS ratio
 - transition/transversion 7, 670
- mutation rates 9–10, 555, 593
 - definitions 663, 664, 670, 671
- National Centre for Biotechnology Information (NCBI) 19, 40
- National Library of Medicine (NLM) 19
- natural selection 12; *see also* molecular adaptation
- NCBI (National Centre for Biotechnology Information) 19
- NCBI Map Viewer 40
- nearest neighbor interchange (NNI) 191, 285, 664
- negative selection 314, 409, 419, 562, 664
- neighbor-joining (NJ) methodology 26, 365
 - definitions 664, 666, 669
 - detecting recombination 527, 528
 - HIV data set 649
 - multiple sequence alignment 87
 - pairwise distance methods 149–154, 155, 163–168, 178–179
 - selection pressures on coding sequences 424
 - split networks 635
- neighbor similarity scores 511
- NeighborNet algorithm 633, 639–640, 649, 651–653
- networks 632; *see also* split networks
- neutral theory of molecular evolution 13
- NEWICK sequence data input format 161, 162, 167, 336, 376–377
- NEXUS sequence data input format 59, 161, 356
 - BEAST software 577
 - Mr Bayes software 257–8
 - parsimony-based inference 291, 304
 - Trypanosomatid phylogeny 339–340
- NJ *see* neighbor-joining methodology
- NLM (National Library of Medicine) 19
- NMR (nuclear magnetic resonance) 38
- nodes 20
- non-degenerate sites 7
- non-homologous recombination 519, 664
- non-parametric bootstrap *see* bootstrap analysis
- non-sense codes 5
- non-synonymous mutations 419
- non-synonymous/synonymous rate ratio *see* dN/dS ratio
- non-synonymous substitutions 320, 664
- nuclear magnetic resonance (NMR) 38
- Nucleotide database 42
- nucleotide sequences 4, 95–96
- nucleotide substitution models 174, 365; *see also* model of evolution
 - bias correction 431
 - genetic distances 112, 121–125
 - Influenza A case study 466
 - Markov process 174
 - TRIM5 α gene sequence alignment 136
- nucleotide substitution saturation *see* substitution saturation
- nuisance parameters 28, 347, 559, 566
 - definitions 664, 666
- null hypothesis 382
- objective function 267
- observed distance 112–113; *see also* pairwise distance methods
- one-sided Kishino–Hasegawa test 389–390, 400, 403
- Online Mendelian Inheritance in Man (OMIM) 42
- ontology 665
- open reading frames (ORFs) 6, 7–8
- operational taxonomic units (OTUs) 20
- optimal alignment 46
- optimality
 - criteria 24, 267
 - principle 75, 77–78
- ORALGEN database 40
- ORFs (open reading frames) 6, 7–8
- orthologous
 - genes 19, 665
 - proteins 314
 - sequences 669
- OTUs (operational taxonomic units) 20
- outer planar networks 638
- outgroup/s 147, 364, 665
 - routing 20, 21, 168–170, 331
 - sequences 414
 - taxa 296
- overdominance 411, 665
- OXBENCH test case collection 93
- pairwise distance methods (p-distance) 142–144, 159–160
 - bootstrap analysis 156, 157–159, 170–174, 175–176
 - cluster analysis (UPGMA/WPGMA) 144–148, 155
 - definition 665
 - editing phylogenetic trees 167
 - Fitch–Margoliash method 170, 171
 - genetic distances, impact on tree morphology 174–180
 - jackknifing 156, 159
 - minimum evolution 149
 - neighbor-joining methodology 149–154, 155, 163–168, 178–179
 - outgroup routing 168–170

- pairwise distance methods (*cont.*)
 - PHYLIP software package 160–163, 170–174
 - programs to display/manipulate phylogenetic trees 161–162
- pairwise homoplasy index 513
- pairwise sequence alignment 74–79
- PAM (point-accepted mutation) 665
 - substitution matrices 45, 72, 88, 662
 - protein sequences 320–322
- PAML (Phylogenetic Analysis by Maximum Likelihood) 373–375, 377, 452
- panmictic populations 552, 665
- parallel evolution 19, 632, 656, 665
- paralogous
 - genes 19, 665
 - isoenzymes 331
- parametric bootstrap *see* Monte Carlo simulation; *see also* Markov Chain Monte Carlo (MCMC) sampling
- parapatric speciation 16, 665
- paraphyletic taxa 20, 665
- parsimony analysis 267–268, 282, 616; *see also* maximum parsimony
 - approximate methods 283–289
 - background 268–270
 - branch and bound technique 280–282
 - calculating length of given tree 270–277
 - exact methods 279–283
 - Fitch's algorithm 279
 - generation of all possible trees 280
 - hill-climbing algorithms 288
 - implicit assumptions 345
 - length 26–27
 - nearest neighbor interchange (NNI) 285
 - practice exercise *see* PAUP*
 - Sankoff's algorithm 278
 - searching for optimal trees 277
 - stepwise addition 282, 283, 284
 - subtree pruning and regrafting 285, 286, 287
 - tree bisection and reconnection 286–287
- parsimony informative sites 506, 512, 665
- partition matrices 511
- pathogenicity in bacteria 494
- patristic distance 156, 665
- PAUP* program 160, 180, 290–293, 330
 - analyzing data through command–line interface 293
 - commands/options 292–293
 - distance optimality criteria 301–304
 - maximum likelihood methodology 267, 304–313
 - models of evolution 355, 356, 357
 - molecular clock analysis 374, 376
 - parsimony analysis/tree searching 294–301
 - running the program 291–292
- PCMA multiple alignment program 84
- PDB (Protein Data Bank) 41
- phenetic distances 634, 666
- phenograms 144, 666
- PHYLIP software package 126–128, 160–163, 170–174, 375
- Phylogenetic Analysis by Maximum Likelihood *see* PAML
- phylogenetic compatibility methodology 513
- phylogenetic hidden Markov model 512
- phylogenetic trees 183
- Phylogeny Inference Package (PHYLIP) 126–128, 160–163, 170–174, 375
- phylograms 23, 164, 241, 296, 460
 - definitions 656, 666
- PhyML software package 192, 336
- PileUp multiple alignment program 84
- PIMA multiple alignment program 84
- POA multiple alignment program 84, 95
- point-accepted mutation *see* PAM
- point mutations 6
- Poisson process 113–116, 174
- polymorphisms 9, 408
- polynucleotide structure 3–4
- polyploidy 8
- polytomies 29, 209, 296, 648, 666
- PopSet database 42
- population dynamics 9–14
- population mutation rate 508, 664; *see also* theta
- population recombination rates 494, 507–509, 561, 592
 - definitions 661, 666
- positional homology 18
- positive selection 314, 363, 419, 423, 496
 - definitions 655, 657, 666
 - molecular sequences 409, 410
 - and recombination 505
- posterior 663
- posterior probability distribution 27, 351
 - Bayesian evolution by sampling trees 565
 - definitions 663, 666
 - maximum likelihood methodology 195
 - molecular clock analysis 351
 - MrBayes software 215
 - protein sequences 332
 - selection pressures on alignments 438
 - Trypanosomatid phylogeny 338
- posterior probability limit 589
- potential scale reduction factor (PSRF) 226, 666
- Praline multiple alignment program 84, 94
- PREFAB test case collection 93
- PRF (Protein Resource Foundation) 39
- primates data set 163–168, 240
- PRIME multiple alignment program 84
- prior probability distributions 27
 - Bayesian phylogenetic analysis 212, 236, 351, 438
 - definition 655, 660, 666
 - molecular clock analysis 371
- priors, vague/uninformative 217
- probabilistic divergence methodology 512
- probability density function 213, 553, 666
- probability mass function 212, 213, 666
- Proalign multiple alignment program 84
- ProbCions multiple alignment program 84, 93, 95
- ProDa multiple alignment program 85, 92

- profile likelihood 436, 437, 609, 666
 progressive alignment 80, 87–89, 666
 Promals multiple alignment program 85, 94
 promoters 3
 PROSITE protein motif database 94
 PROTAL2DNA alignment tool 104–105
 ProtDist program 330
 Protein Data Bank (PDB) 41
 Protein database 42
 protein identifiers 37
 protein sequences in phylogenetic analysis
 314–315
 alignment of two-protein sequences 325–327
 analyzing protein sequences 315
 codon usage bias 317
 DNA or protein choice 323–324
 dot-matrix sequence comparison 325–326
 evolutionary distance, and protein differences
 323
 genetic code, universal 316–318
 inferring protein phylogeny 330–331
 introns/non-coding DNA 322–323
 Leishmania mexicana enzyme analysis 332–337
 look-back time 318–320
 majority rule consensus tree 341
 manual adjustment of protein alignments
 328–329
 Markov Chain Monte Carlo 338–342
 maximum likelihood tree, *Leishmania mexicana*
 337
 molecular clock analysis 363
 multiple sequence alignment 327–328
 phylogenetic trees construction 324–332
 protein evolution 315–324
 protein/nucleic acid search comparison 321
 reliability of inferred trees 332
 retrieval of protein sequences 324–325
 rooting a tree 331
 sequence divergence in proteins 320–322
 tree building 330–332
 Trypanosomatid phylogeny 338–342
 Protein Resource Foundation (PRF) 39
 protein synthesis 3
 PROTEST 358–359, 361
 PRRN/PRRP multiple alignment program 85,
 90
 PSAlign multiple alignment program 85
 PSI-BLAST 94; *see also* BLAST
 PSRF (potential scale reduction factor) 226,
 666
 PubMed database 42
 purifying selection 314, 409, 419, 562, 664
 purines 4
 puzzle support values 196; *see also* quartet puzzling
 pyrimidines 4

 Q matrices 116, 117, 231, 365, 666–667
 QOMA multiple alignment program 85
 quartet puzzling 126, 304, 381
 definition 667
 HIV data set 202, 207–209

 maximum likelihood methodology 195–196,
 203–209
 quartet scanning 513
 quasispecies 9, 667
 QuickJoin 154

 random effects likelihood (REL) methods
 definitions 658, 667
 HyPhy software package 481–483
 Influenza A case study 481–483
 selection pressures on coding sequences
 442–444
 rate/s
 growth or decline 556
 heterogeneity 123–125, 133–135
 RAxML 193
 RDP3 software program 512
 exploratory analysis 538–546
 HIV sequences, detecting recombination
 538–548
 refining recombination hypothesis 546–548
 tree reconstruction 545
 reading frames, genetic 5
 reassortment 495, 519, 667; *see also* recombination
 recombination analysis 8, 29, 493, 506, 513,
 519–520
 average-over-window-based scanning
 approaches 523, 525
 breakpoint detection 507
 detecting 506, 520–523, 530–531; *see also* HIV
 sequences
 detecting recombination, requirements
 520–523
 evolutionary implications 496–498
 genetic distances 504, 513, 525–528
 jumping of recombinant isolates 523, 524
 linkage disequilibrium 495–496
 mechanisms of recombination 493–495
 molecular clocks 505
 Monte Carlo simulation 506
 neighbor-joining methodology 527
 overview of detection tools 509–517
 parametric bootstrap 506
 performance of recombination detection tools
 517–518
 phylogenetic inference, impact on 498–506
 positive selection 505
 rate variation among lineages 504–505
 rate variation among sites 504
 rates *see* population recombination rates
 signals, recombination 523
 similarity/dissimilarity methods 526
 software tools 521, 527–529, 532
 substitution distribution methods 527,
 529
 substitution patterns 496, 497
 theory 523–530
 tree shapes/branch lengths 492
 Refina multiple alignment program 85
 RefSeq database 40
 relative rate test 364–365, 667

- relaxed molecular clock models 23, 369–371, 379
 - Bayesian analysis 230–231
 - BEAST 565
 - coalescent 562
 - definition 667
 - distance methodology 168
- RELL (resampling estimated log likelihoods) 386–387, 667
- repeats problem, multiple sequence alignment 68–70
- replacement/non-synonymous mutations 412
- reproductive isolation 15
- resampling estimated log likelihoods (RELL) 386–387, 667
- restriction fragment length polymorphisms (RFLPs) 16
- reticulation networks 633
- reversible-jump MCMC 235, 351, 574, 655, 668
- RFLPs (restriction fragment length polymorphisms) 16
- RNA (ribonucleic acid) 3
- rooted phylogenetic trees 20, 21
- RSVA data set 377–380

- SAGA multiple alignment program 85
- Salmonella data set 648–653
- SAM multiple alignment program 85
- sampling distributions 382, 668
- sampling importance resampling (SIR) algorithm 436, 668
- sampling variance 668
- Sankoff's algorithm 278
- SATCHMO multiple alignment program 86, 91
- saturation *see* substitution saturation
- Sawyer's statistical test for gene conversion 510
- s-coalescent 668
- Se-Al multiple alignment editor 98
- SeaView multiple alignment editor 98
- segregating sites 414, 496, 668
- selection coefficients 409, 668
- selection pressures on coding sequences 419–420, 450–451
 - asymptomatic normality 435–436
 - Bayesian approaches 438
 - bootstrap analysis 426
 - branch-by-branch variation estimation 438
 - codon substitution models 424–426
 - comparison of rate distributions in different genes 445
 - confidence intervals 435–438
 - data driven branch specification 440–442
 - distance-based approaches 426–428
 - dN/dS estimation 429–431
 - evolutionary models, choosing among different 432
 - fixed effects likelihood models 442, 445–446
 - HIV example 423
 - hypothesis testing 431–435
 - influenza example 421–422
 - likelihood ratio test 433–434
 - local vs. global model 439
 - maximum likelihood methodology 420, 428–429
 - MEGA4 software package 453
 - method choice 447
 - neighbor-joining methodology 424
 - neutral situations 420–421
 - nucleotide substitution bias correction 431
 - phylogenies, effect of 421–422
 - practice exercises *see* Influenza A case study
 - prerequisites 423–424
 - a priori specification 439–440
 - profile likelihood 436, 437
 - random effects likelihood models 442–444
 - rate comparison in different branches 449–450
 - sampling techniques 436–438
 - selection types 422–423
 - simulated data 426
 - single likelihood ancestor counting methods 446–447, 448
 - software packages 452–454
 - synonymous rate variation 449
 - units of evolution 420
- selective pressure, positive/negative 10
- selective sweeps 423, 497, 562, 657, 668
- sense codes 5
- SEQDB 39
- sequence databases *see* databases
- sequence divergence in proteins 320–322; *see also* PAM
- sequence identifiers 33
- Sequence Retrieval System (SRS) 43–44
- sequence version archive 36–37
- serial coalescent 559–561, 668
- Shannon's entropy (information entropy) 616, 660
- Shimodaira-Hasegawa (SH) test 390, 391, 403
- sickle-cell anemia 10
- signals, recombination 523
- silent mutations 412
- simian immunodeficiency (SIV) genomes data set 629–630
- similarity 17–18, 33, 45, 93, 526
- SIMPLOT software package 528, 532, 533–536
 - detecting recombination 533–538
 - mapping recombination 536, 537
 - SIMPLOT groups feature 537–538
- simulated annealing 193, 668
- single likelihood ancestor counting (SLAC) 387
 - Influenza A case study 477–478, 479, 480
 - selection pressures on coding sequences 446–447, 448
- single nucleotide polymorphisms (SNPs) 668
- single rate dated tip (SRDT) models 367–368, 377–379
- SIR (sampling importance resampling) algorithm 436, 668
- sister scanning methodology 512
- sister taxa 168, 617, 662, 668
- site-frequency 412, 668
- SLAC *see* single likelihood ancestor counting
- SNPs (single nucleotide polymorphisms) 668

- software packages
 - genetic/evolutionary distances 126–128
 - maximum likelihood methodology 199
 - recombination events 521, 527–529, 532
 - selection pressures on coding sequences 452–454
 - testing tree topologies 397
- SOWH (Swofford-Olsen-Waddell Hillis) test 393–394
- spatial phylogenetic variation 511
- speciation 14–16
- species trees 21
- spectral analysis 637
- SPeM multiple alignment program 86, 94
- splicing 8
- split decomposition 198, 636–638, 633
 - definitions 658, 668
 - HIV data set 649
 - Salmonella data set 650
- split networks 658, 660
 - Buneman tree 634–636, 639
 - consensus/supernetworks 633, 640–641
 - definition 668
 - evolutionary relationships as networks 631–633
 - median networks 633, 638, 640
 - NeighborNet algorithm 633, 639–640, 649, 651–653
 - origin/introduction of term 633
 - outer planar networks 638
 - practice exercise *see* SplitsTree program
 - split decomposition 636–638
 - theory 633–634
 - weakly compatible splits to networks 638–639
 - weighted networks 636, 639
 - weighted trees 633
- splits 655, 660, 668
- SplitsTree program 642, 644
 - bootstrap analysis 644, 646–647
 - Buneman tree 646, 647
 - computing different networks 646
 - computing trees 646, 647
 - downloading 642
 - fit index 639, 643–644
 - getting started 643
 - HIV data set 648–653
 - laying out split networks 645
 - mtDNA data set, using on 642–648
 - NeighborNet algorithm 651–653
 - printing 647–648
 - recomputing split networks 645–646
 - Salmonella data set 648–653
 - split decomposition tree 650
- SPR (subtree pruning and regrafting) 191, 223, 285, 669
- SRDT (single rate dated tip) models 367–368, 377–379
- SRS (Sequence Retrieval System) 43–44
- SSAHA database search tool 52
- star decomposition 190, 463, 664, 668–669
- star-like radiation 209
- stationarity 116, 187, 222, 424, 615
 - definitions 669, 670
- statistical parsimony networks 633
- Steel's method 616–621
- stepping-stone model of migration 557, 669
- stepwise addition 190, 282, 669
- stepwise recombination detection 513
- stochastic
 - evolution 12, 13
 - forces 11
- strictly bifurcating trees 28
- substitution distribution methodology 506, 527, 529
- substitution models 217, 320, 660, 664, 670; *see also* models of evolution
- substitution patterns, recombination 496, 497
- substitution rates 9–10, 424, 570
- substitution saturation 113, 137–140, 319; *see also* DAMBE software
 - definition 669
 - problem of 70–72, 615–616
- sub-tree pruning and regrafting (SPR) 191, 223, 285, 669
- summary statistic methods 413, 413–415
- supernetworks 633, 640–641
- survival of the fittest 13
- SwissProt 320, 324, 329
 - Leishmania mexicana* enzyme analysis 333–335
 - Uniprot knowledgebase 38
- Swofford–Olsen–Waddell Hillis (SOWH) test 393–394
- symmetrical/non-symmetrical recombination 494
- sympatric speciation 16, 669
- synapomorphy 616, 669
- synonymous rate variation 449
- synonymous substitutions 669
- synonymous/non-synonymous mutations 7, 10–12, 133, 412, 506; *see also* dN/dS ratio
- Tajima's D summary statistic 414, 669
- Taxonomy database 42–43
- TBR (tree bisection and reconnection) 191, 286–287, 670
- T-COFFEE multiple alignment program 86, 89–90, 93, 95, 102
- testing tree topologies 381–382, 387, 395–396
 - approximately unbiased test 392–393, 403, 404
 - confidence sets 394–395
 - CONSEL software package 202, 397–398, 401–403, 404
 - definitions, distributions and testing 382–384
 - expected likelihood weights 394–395, 400, 404
 - Kishino–Hasegawa test 388–389, 403
 - likelihood ratio distribution 385–386
 - likelihood ratio test 384–385
 - non-parametric bootstrap 386–387
 - one-sided Kishino–Hasegawa test 389–390, 400
 - Shimodaira–Hasegawa (SH) test 390, 391, 403
 - software packages 397
 - Swofford–Olsen–Waddell Hillis (SOWH) test 393–394

- testing tree topologies (*cont.*)
 - tree tests, general structure 388
 - TREE-PUZZLE 397–398, 398–401, 403, 404
 - weighted test variants 390–392, 403
- theta 669–671
- third party annotations (TPA) 37
- thymine 4
- time to the most recent common ancestor (TMRCA) 21
- time-homogenous stationary Markov processes
 - 116, 118, 120, 140
- TN93 model 132–133
- topology, tree 20
- TPA (third party annotations) 37
- trace plot
 - BEAST software 584, 585, 586, 587
 - definition 670
 - LAMARC software 610
 - Bayesian phylogenetic analysis 224
- traceback 77–78
- Tracker/Anchored Dialign multiple alignment program 86
- transduction 494, 670; *see also* recombination
- transformation 494, 670; *see also* recombination
- transition mutations 7, 670; *see also*
 - transversion/transition ratios
- transversion 670; *see also* transversion/transition ratios
 - models 360–361
 - mutations 7
- transversion/transition ratios
 - Bayesian analysis 569
 - database searching 45
 - definition 670
 - distance methodology 122, 131–132, 163
 - LAMARC 606
 - maximum likelihood methodology 183
 - recombination detection 536
 - selecting models of evolution 346
 - selection pressures on alignments 421
 - substitution saturation 616
- TreeAnnotator tool 588
- tree bisection and reconnection (TBR) 191, 286–287, 670
- TREECON 160
- TreeEdit algorithm 376
- TREE-PUZZLE software package
 - distance methodology 126–128, 132–133, 140–141, 161
 - likelihood mapping analysis 208
 - maximum likelihood methodology 195, 203–207
 - molecular clock analysis 374
 - parsimony 304
 - protein sequences 330
 - testing tree topologies 397–401, 403, 404
- TreeView algorithm 164–166, 168, 172–173, 376, 532
- TrEMBL 38
- TRIM5 α gene sequence alignment 100–108
- ClustalW multiple alignment program 100–102, 103–104
 - comparison of alignments 103–104
 - databases of alignments 106–108
 - editing/viewing alignments 105–106
 - Jukes and Cantor (JC69) models 128–131
 - molecular clock analysis 373–375, 375–377
 - MUSCLE multiple alignment program 102, 104
 - neighbor-joining (NJ) methodology 163–168, 178–179
 - nucleotide substitution models 136
 - protein to nucleotide alignment 104–105
 - T-COFFEE multiple alignment program 102
- Trypanosoma brucei* 329
- Trypanosoma congolense* 317
- Trypanosomatid phylogeny 317, 338–342
- twofold degenerate sites 7
- UCSC Genome Browser 40
- ultrametric/ultrametricity
 - definitions 664, 670
 - ultrametric trees 144, 147–9, 155, 365, 504
- unrelated relaxed clocks 371, 570, 670
- UniGene database 40, 42
- uninformative priors 217
- Uniprot knowledgebase 38, 320, 324–325, 328
- Uniprot splice variants 38
- UniRef100/90/50 38
- units of evolution 420
- unrooted models 20, 148, 230, 365
 - definitions 661, 670
- UPGMA (unweighted pair group method with arithmetic means) 26
 - definitions 670–671
 - genetic distances 140
 - Influenza A case study 462
 - molecular clock analysis 365
 - pairwise distance methods 144–148, 155
 - parsimony analysis 268
 - protein sequences 330
 - split networks 635
- uracil 4
- vague/uninformative priors 217
- version number 36–37
- vertebrate mtDNA data set *see* mtDNA data set
- viruses
 - evolution 3, 8, 9, 16
 - genetic exchange 494–495, 519–520; *see also* recombination
 - LAMARC software 602
 - molecular clock analysis 377–380
 - visual recombination detection 513
- Watterson estimate 600, 671
- web-based aids 591
- weighbor 154
- weighted
 - neighbor-joining 154

- networks 636, 639
- test variants 390–392, 403
- trees 633
- weighted-pair group method with arithmetic means (WPGMA) 144–148
- whole genome shotgun (WGS) sequences 37
- wild-type genotype 407
- wobble position 316, 318–319
- WPGMA (weighted-pair group method with arithmetic means) 144–148
- Wright–Fisher population model 552, 554, 657, 671
- WSP objective functions 79–80, 90, 91
- Xia’s method, substitution saturation 621–623
- X-ray crystallography 38