

## Index

- absolute fast converging methods, 101, 102, 167–170, 296
- additive matrix, 10, 11, 13–15, 84–89, 93–101, 103, 152–154, 248, 282–283
  - topology invariant neighborhood, 93
- agreement subtrees, 116, 117
  - maximum agreement subtree, 116
  - maximum compatibility subtree, 116
- Aho, Sagiv, Szymanski, and Ullman algorithm, 52, 53, 57, 140, 242, 243
- algorithms, 304, 307–312, 318–320
  - “Big-O” running time analysis, 304, 307–311
  - design techniques, 24, 26, 285, 317–320
  - divide-and-conquer methods, 24, 121, 254, 274, 279, 284–286, 289, 291, 296, 320
  - dynamic programming, 318–320
  - exhaustive search, 311, 312
  - greedy algorithms, 311
  - iteration, 24, 84, 91, 211, 214, 218–221, 254, 285, 286, 288–293
  - pseudocode, 308, 309, 320
  - recursive algorithms, 8, 12–14, 215, 254, 287, 319, 320
- alignment-free phylogeny estimation, 102, 103, 289
- Bayesian phylogenetics, 18, 20, 141, 159–160, 166–167, 172–173, 217, 224, 241, 247, 252, 258, 271, 277–278
  - branch support, 159, 160, 166
  - computational challenges, 17, 159, 160, 277–279
  - detailed balance property, 159
  - maximum *a posteriori* (MAP) tree, 159, 160, 166, 278
  - MCMC techniques, 159, 166, 217, 277, 278
  - MrBayes, 247
  - posterior probability distribution, 160
  - statistical consistency, 160
- binary relation, 305
- bipartition compatibility, 39–41
- bipartition encoding of a tree, 21, 114
- BLAST, 290
- branch support calculation, 165, 299, 300
  - Bayesian support values, 159, 166
  - bootstrapping, 165, 166, 299, 300
- centroid edge, xvii, 214, 218, 221, 233, 287–289, 298
- character compatibility, 63, 69, 70
- character data, 5, 61
- clade compatibility, 33–35
- cladogenesis models, 172, 173
  - birth–death model, 18, 172
  - Yule model, 172
- clustering sets of trees, 117
- coalescent-based species tree estimation, 102, 234–254, 268–270, 292–293
  - \*BEAST, 252–254
  - anomaly zone, 239–241, 244
  - ASTRAL, 246–248, 253, 256, 259, 270
  - ASTRID, 102, 103, 247, 248, 253
  - BBCA, 253, 254
  - BUCKy, 247
  - concatenation analysis, 238, 239
  - fixed-length statistical consistency, 252, 253
  - impact of gene tree estimation error, 248–250
  - MDC, 256
  - METAL, 250, 253
  - MP-EST, 24, 243, 249, 253, 254, 292, 293
  - NJst, 102, 103, 247, 248, 259
  - quartet-tree methods, 129, 244–247, 250, 251
  - site-based methods, 241, 250, 252
  - SMRT-ML, 250, 253
  - SNAPP, 250
  - SRSTE, 243, 244
  - summary methods, 129, 241–250
  - SuperTriplets, 243
  - SUSTE, 244, 245
  - SVDquartets, 250–252
- combinatorial counting, 307
- consensus trees, 41, 44, 109–118, 159, 300
  - asymmetric median tree, 110, 113–115
  - characteristic tree, 115–117
  - compatibility tree, 110, 112, 113

- extended majority consensus, 112
- greedy consensus tree, 110–112, 115, 278
- local consensus tree, 118
- majority consensus tree, 110–115, 122, 159
- strict consensus tree, 41, 44, 110–112, 115, 122
  
- DACTAL, 24, 132, 254, 286, 288–292
- direct optimization, 216, 217
- disk-covering methods (DCMs), 24, 254, 279–288, 290, 293, 296, 297
- dissimilarity matrix, 13, 83
- distance matrix, 83
- distance-based tree estimation, 7–17, 21–23, 83–103, 140–141, 152–154, 167–170, 247–248, 278, 282–285, 296
  - Agarwala et al. algorithm, 94, 96–100, 153, 284
  - balanced minimum evolution, 95, 96, 100
  - BioNJ\*, 248
  - branch length estimation, 94, 95
  - Buneman Tree, 90, 96–100, 284
  - computing distances under statistical models, 14, 83, 102, 103, 152–154, 161
  - error tolerance, 89
  - FastME, 17, 96, 97, 100–103, 140, 153, 248
  - Four Point Condition, 11, 86, 87
  - Four Point Method, 11–15, 87–89, 168–170
  - impact of missing data, 103
  - minimum evolution, 95
  - Naive Quartet Method, 14, 15, 17, 18, 20–22, 24, 89, 90, 97–100, 165, 167, 168
  - neighbor joining, 17, 18, 20–24, 90–92, 97–103, 165, 170, 248, 275, 283–285, 296
  - optimization problems, 94, 95, 100
  - safety radius, 96–101, 153, 154, 248
  - statistical consistency, 92–94, 96–101, 153, 154
  - UPGMA, 9, 16–18, 20–22, 24, 84–86, 91, 210, 215
  - using DCMs, 282, 283, 285
- Dobzhansky, xiii
- dynamic programming, 63–68, 156, 157, 185, 187–192, 200–203, 256, 318–320
  
- evolutionary diameter, 22
  
- false discovery rate, 321
- Felsenstein Zone tree, 161–165, 171
  
- gene duplication and loss, 229, 230, 234, 254–258
  - MixTreEm, 258
  - Phyldog, 258
- gene tree parsimony, 256, 258
- genome rearrangements, 72, 102, 229, 230, 270–272, 296
- genome-scale evolution, xv, 4, 23, 24, 102, 270–272
- graph theory, 304, 305, 309, 312–317
  - adjacency matrix and adjacency list, 309, 315
  - directed acyclic graph (DAG), 196, 224, 225, 305
  - Eulerian graph, 305, 313
  - maximum clique, 312
  - minimum vertex coloring, 249, 312–317
- Hamming distance, 13, 17
- Hasse Diagram, 33–35, 39, 40, 70, 305, 306
- heterotachy, 170, 171
- homology, xvii, 178–183, 194, 204, 205, 212–216, 224, 229, 230
- homoplasy, xvii, 62
- horizontal gene transfer, xv, 4, 23, 24, 61, 109, 234, 235, 259, 300
- hybridization, 4, 23, 24, 61, 234, 300
  
- incomplete lineage sorting, 102, 235–238, 241–254, 270
- ingroup taxa, 44
- insertions and deletions (indels), xvii, 4, 180
  
- long branch attraction, 162, 164
  
- maximum compatibility tree estimation, 62, 63, 69, 70, 72–77, 161, 163
  - compatibility informative characters, 76, 77, 164
  - positively misleading, 161, 163, 164
- maximum likelihood tree estimation, 17, 18, 102, 128, 139–141, 147, 159–164, 171–172, 216, 218, 219, 238, 239, 249, 252, 253, 258, 275, 276, 278, 285, 296
  - FastTree-2, 160, 279, 291
  - heuristics, 158, 275, 276
  - IQTree, 279
  - nhPhyml, 279
  - NP-hard, 17, 158
  - PhyML, 279
  - PoMo, 279
  - RAxML, 139, 279, 291
  - use within supertree methods, 128
  - using DCMs, 296
- maximum parsimony tree estimation, 17, 22, 23, 62–69, 73, 75–77, 127, 161–164, 171, 192, 275, 276, 278
  - Fitch algorithm, 63–66
  - heuristics, 17, 69, 275, 276
  - missing data treatment, 127
  - NP-hard, 17, 68
  - parsimony informative characters, 76, 77, 137, 161
  - positively misleading, 17, 21, 161, 163, 164, 171
  - Sankoff algorithm, 66, 68
  - small parsimony and large parsimony problems, 63, 68, 69
  - statistical inconsistency, 17, 164
  - using DCMs, 296
- minimum spanning tree, 193
- missing data, 71, 75, 103, 127, 128, 178, 248, 257, 334
- multi-species coalescent model, 234–240, 243–246

- multiple sequence alignment, xiii, xv, 4, 19, 24, 102, 178–233, 299  
 aligning alignments, 207–209, 219, 226  
 alignment error measurement, 180, 182, 183  
 BAli-Phy, 217, 220, 224, 226, 229  
 benchmarks, 178, 181  
 Clustal, 210, 217  
 co-estimation of alignments and trees, 211, 215–223, 291  
 consensus alignment, 224–226  
 consistency, 212–214, 224  
 divide-and-conquer methods, 214–216  
 FastSP, 183  
 impact of guide tree, 212  
 impact on tree estimation, 181, 300  
 MAFFT, 211, 213, 215–217, 219, 220  
 MAPGAPS, 214, 215  
 mega-phylogeny, 214, 215  
 Muscle, 211, 217, 219, 220  
 OPAL, 219, 220  
 optimization problems, 190, 191  
 PAGAN, 204, 205, 211, 217, 228, 229  
 PASTA, 24, 212, 214, 215, 218–223, 226, 229, 291  
 phylogeny-aware, 211, 217, 228, 229  
 Prank, 211, 217, 228, 229  
 ProbCons, 211, 213, 217  
 progressive alignment, 209–212, 226, 229  
 PROMALS, 205, 213–215  
 reference-based alignment, 204, 205  
 SATé, 24, 212, 214, 215, 220, 229, 291, 292  
 SATCHMO-JS, 214, 215, 218  
 seed alignment, 206  
 sequence evolution models, 18, 25, 103, 216, 217  
 StatAlign, 217, 224–226  
 statistical models, 194, 196–198, 200–206, 220, 221, 223, 226, 227  
 Sum-of-Pairs Alignment, 190, 191  
 T-Coffee, 217  
 template-based alignment, 205, 206  
 tree alignment, 190–194, 215, 216  
 UPP, 214, 220, 228, 229
- NNI, SPR, and TBR moves, 117, 275–277, 296  
 NP-hard problems, xvii, 17, 68, 69, 72, 74, 122, 158, 274–277, 283, 295, 299, 311–318, 320  
 heuristic search strategies, 274–277, 299  
 Karp reduction, 315–317  
 NP-complete problems, xvii, 314, 316  
 optimization, decision, and construction problems, 311–313
- orthology, 224, 228, 255, 272  
 outgroup taxa, 44
- pairwise sequence alignment, 180, 185–190, 227  
 edit distances, 184
- local alignment, 227  
 Needleman–Wunsch, 185, 187–190, 207, 227  
 Smith–Waterman, 227
- perfect phylogeny, 62, 63, 70–72, 74, 77  
 Pfam, 204
- phylogenetic network, 23, 24, 259–267, 269, 270, 300  
 data-display network, 263–267, 270  
 evolutionary network, 235, 261–264, 267  
 phylogenetic placement, 206, 297  
 phylogenomics, 129, 234–246, 248–261, 263–265, 267, 270–272
- polytomy, xvii, 116  
 post-tree analysis, 228
- profile hidden Markov models, 194, 196–204, 206, 207, 220, 226  
 Baum–Welch method, 204  
 building profile HMMs, 203, 204  
 database searches, 204  
 ensembles of hidden Markov models, 220, 221, 223, 224  
 Forward algorithm, 203  
 HMMER, 205  
 Viterbi algorithm, 201–203, 207
- proof techniques, 307, 308  
 proof by contradiction, 308  
 proof by induction, 307, 308
- quartet trees, 12, 13, 46, 53–58, 87–91, 118, 128–132, 165  
 All Quartets Method, 12, 13, 53–57, 87, 89, 244, 245, 251  
 dyadic closure method, 54–56, 153, 168–170  
 Q(T), 12, 13  
 quartet amalgamation methods, 53, 89, 90, 128, 129, 132, 139, 251, 274, 276  
 quartet-based tree construction methods, 102, 128  
 quartet compatibility, 54, 57  
 Quartets MaxCut, 128, 135, 139  
 short quartets, 54–56, 284, 286
- Robinson–Foulds tree error, 19, 21, 22, 42, 46, 283  
 rogue taxon, 43, 44, 46  
 rooted tree compatibility, 52, 57
- sample complexity, 145, 167  
 sensitivity, specificity, precision, and recall, 321, 322
- sequence evolution models, xiv, xv, 4–8, 17, 19, 21, 24, 26, 29, 102, 140, 145–173, 180, 278, 279  
 amino acid models, 150  
 Cavender–Farris–Neyman (CFN), xiv, 4–11, 13–17, 91–94, 96–98, 102, 128, 145, 148, 153, 156, 158, 161–163  
 codon models, 145  
 computing the probability of a sequence dataset, 154, 156  
 Felsenstein’s Pruning Algorithm, 156–159

- General Markov model, 149, 153, 155, 158, 161  
 Generalised Time Reversible (GTR) model, 19,  
 148, 149, 156, 158, 160  
 heterotachy, 170, 171  
*i.i.d.* assumption, 16, 147, 149  
 Jukes–Cantor (JC69) model, 146–149, 153,  
 156–158, 160, 164, 167–170  
 Jukes–Cantor (JC69) model, 19  
 long branch attraction, 164  
 Markov property, 147  
 model selection, 151, 152, 296  
 No Common Mechanism model, 149, 158, 170, 171  
 rates-across-sites, 16, 149, 150, 172  
 similarity to coin tosses, 6  
 stationary assumption, 147  
 strict molecular clock, 163  
 Thorne–Kishino–Felsenstein 1991 (TKF91), 103  
 time-reversible, 147, 154, 158  
 sequence profiles, 194–198  
 sequencing technologies  
   impact on phylogeny estimation, xiii, 268, 269  
 simulation study, 18–21, 23, 26, 41, 90, 146, 164, 165,  
 167, 173, 178, 180, 181, 238, 258, 267, 283, 328,  
 329  
 statistical binning, 249, 250  
 statistical consistency, xiv, 6, 13–18, 21, 24, 145–147,  
 153, 164, 167, 171, 278  
 statistical identifiability, 145–147  
 statistical significance, 328  
 strict molecular clock, 7–10, 15, 44, 84, 251–253  
 supertree methods, xv, 52, 56, 57, 109, 121–142, 245,  
 258, 279, 281, 285, 286, 288–290, 292  
   Matrix Representation with Parsimony (MRP)  
     missing data treatment, 127  
   Asymmetric Median Supertree, 123, 124, 126  
   compatibility supertree, 52, 56, 57, 122–126, 135  
   distance-based, 140  
   FastRFS, 126, 141, 256  
   guenomu, 258, 270  
   majority rule supertrees, 122  
   Matrix Representation with Likelihood (MRL),  
     122, 128, 135  
     missing data treatment, 128  
   Matrix Representation with Parsimony (MRP), 122,  
     126–128, 135–139, 291  
     MRP matrix, 126, 127  
   maximum likelihood supertree, 141  
   Maximum Quartet Support Supertree, 122, 128,  
     129, 245  
   MinCut Supertree, 122, 140  
   Quartet Median Tree, 122, 128, 129, 245, 246, 254  
   quartet-based methods, 139  
   Robinson–Foulds Supertree, 122, 124–126, 141  
   Split–Constrained Quartet Support Supertree,  
     129–132, 247  
   statistical aspects of, 132, 141  
 Strict Consensus Merger (SCM), 132–139, 285,  
 288–290  
 SuperFine, 132, 135–139, 289–291  
 SuperTriplets, 243  
   use within DCMs, 279, 281, 289–293  
   using rooted source trees, 139, 140  
 taxon sampling, 164, 165  
 transitive closure, 181, 204, 215, 306  
 tree compatibility, 57, 112  
 tree error rates, 18–22, 41, 42, 46  
 trees, 3, 29–46, 57  
   binary tree, 4, 21  
   binary vs. multifurcating, 29, 31, 38, 39, 42  
   bipartitions of unrooted trees, 19, 36–38, 42  
   caterpillar tree, 45, 46, 56, 164  
   clade representations of rooted trees, 31, 32  
   clades of rooted trees, 40  
   comparing two unrooted trees using bipartitions, 38  
   completely balanced tree, 45  
   constructing a rooted tree from its clades, 33  
   constructing an unrooted tree from its bipartition  
     set, 39  
   determining if a set of bipartitions is compatible, 40  
   different graphical ways of representing trees, 30  
   edges and branches, 5, 29  
   fully resolved trees, 38  
   homeomorphic subtree, 12, 45, 46  
   leaves, 29  
   MRCA, 29, 236, 237  
   Newick notation, 30–32, 36, 37  
   number of trees on  $n$  leaves, 43  
   polytomy, 29, 57, 59, 87, 100, 133, 135  
   rooted vs. unrooted, 29, 44, 45  
   star tree, 42, 44, 87–90, 92, 100, 113, 115  
   tree refinement, 38, 111, 112  
   vertices, 29  
 treespace, 43, 117, 270, 296  
 triangle inequality, 14, 83  
 triangulated graphs, 279–283, 285–288, 293–296  
   decompositions of, 280, 281, 287, 288, 294, 296  
   definition, 279, 293  
   perfect elimination ordering, 293–296  
   short subtree graph, 282, 285, 286, 296  
   threshold graph, 282, 283, 296  
   use within DCMs, 280, 285, 289  
 triplet trees, 52, 53, 57, 118, 139, 140  
   maximum triplet support problem, 139  
   triplet tree compatibility problem, 57  
 two-phase phylogeny estimation, 178, 179, 216, 291  
 ultrametric matrix, 8, 9, 84, 85, 140, 141  
 unrooted tree compatibility, 51, 57, 125, 135  
 whole genome alignment, 229, 230  
 writing papers about computational methods,  
 327–330