

- MULTIZ, aligning whole genomes, 225
mutation units, 2
- Nepenthes*, 153
networks: rooting of semidirected, 136–39; species, 124–29; trees versus, 122–24. *See also* species networks
next-generation sequencing (NGS), 150–51
NJMerge, disjoint tree merger (DTM) method, 35, 36
NJst, coalescent-based tree method, 120
Nomascus leucogenys, 78, 79
nonlinear dimensionality reduction (NLDR), 261, 269;
tree set visualization, 261, 269, 270, 271, 275
nonparametric bootstrap (NBS), 240–41. *See also* bootstrap analysis
Nymphaeales, 152, 154, 156
Nyphaea colorata, 152
- omni-gene hypothesis, 247
On the Origin of Species (Darwin), 233
oyster genes, coevolution of, 248
- pairwise disjoint, 34
paleognathous birds: phylogenomic analysis, 216–18, 231. *See also* birds
PAML software, 229, 245, 247
parallel microfluidic PCR + HTS, 8
ParGenes, 33
partitioned coalescence support (PCS), 60
PASTA, software, 38, 225
PAUP*, 27, 28, 68; SVDQuartets, 87; SVDQuartets for quartet sampling and assembly, 75–76
Penstemon, 146; approach to study, 176–79; best maximum pseudolikelihood (ML) networks by SNaQ, 187; bootstrapping, 178; calculation of QCFs, 177–78; candidate hybridization events from rooted triples, 181–82; character evolution and biogeography, 189; gene tree uncertainty, 178; hybridization discussion, 186–90; hypothesis of allopolyploid formation in *P. attenuatus*, 180; materials and methods, 179–82; nuclear amplicon data, 182; phylogenetics of hybrids and polyploids, 189–90; phylogeny of subsections by ASTRAL-III, 183; phylogeny of subsections by qcf and QuartetMaxCut, 184; phylogeny of subsections by RAxML, 185; sample collection, DNA extraction and amplicon sequencing, 180–81; species network inference, 182, 186; species tree inference, 181, 182, 185–86; study system, 179–80; subsections *Humiles* and *Proceri*, 176; taxonomy of subsections *Humiles* and *Proceri*, 188–89; tests for hybridization, 186; validation of QCF estimation, 178–79
PHAST program, 228
PHYLOGDOG, 29, 155
PhyloAcc, 228–29
PhyloG2P paradigm, 215
phylogenetic(s), 1; birds, 211, 216–18; gene expression and epigenetics, 230–31; genome-scale inference, 212; invariants, 13; paleognathous birds as test case for post-genomic, 218; phylogenetic linkage (PL), 255–56; rooted triplets for emu + cassowary, kiwi and rhea lineages, 223, 224; technologies for relationships, 215–16; whole-genome species for ancient radiation of birds, 218–23
phylogenetic analysis, 232, 250; approaches for studying hybridization, 161–63; gene expression and epigenetics, 230–31; gene flow, 120. *See also* phylogenetic heterogeneity; phylogeny
phylogenetic heterogeneity: analysis under, 243–50; coevolution of figs and wasp pollinators, 249–50; coevolution of multiple genes and traits in *Jaltomata*, 248; coevolution of oyster genes with bacterial virulence, 248; gene tree distributions, 239; introgression under uncertainty in birds, 244–45; quantification and visualization of discordance, 238–40; quantification of conflict and tree evaluation, 240–41; selection under heterogeneity in wild tomatoes, 246–47; testing introgression and hybridization, 243–45; testing of coevolution under, 249; testing of selection under, 245–48; testing of traits under, 247–48; theme of studying, 212; visualization of conflict, 241–43
phylogenetic network analysis, 89–90, 213; heuristic search versus MCMC sampling, 93, 94; isomorphic, 113; parameters and identifiability, 92, 93; reading and interpreting, 91–92. *See also* PhyloNet
phylogenomic(s), 14; consequences of gene tree conflict in, 154–60; data types and technologies for generating, 6, 9; discordance, 234; filtering of data, 57–61; goal of analysis, 258–59; pipeline, 56; species tree estimation, 55, 56; studies of per branch quartet support, 65–66; subsampling, 219, 221; SVDQuartets, 81–82. *See also* gene tree conflict
phylogeny, 55, 99, 120, 145, 231, 235–36, 250; of birds, 216, 218–19, 229; branch tests, 246; coalescent theory and, 215–16; consensus, with node concordance, 242; cophylogeny, 250; data generation for species-level, 74; fixed-species, 82–83; flowering plant, 152, 153, 158; gene tree heterogeneity and, 217; hybridization across, 161, 163, 167, 174; introgression, 243, 244–45; of New World kingsnakes, 189; of *Penstemon*, 179, 180, 181, 183, 184, 185, 190; sequence alignment and, 225–26; speciation, 242; species tree estimation, 211–13, 236, 239; traits, 247–48; unlinked character models, 209
PhyloNet, 40, 176, 190, 244; analysis of larger data sets, 106–11; analysis of polyploids, 114–17; ancestral recombination graph, 113; backbone networks, 111; Bayesian inference, 103–5; comparison and summarizing networks, 111–12; computational requirements for methods other than MCMC_SEQ, 107, 108–9; displayed trees, 111; divide-and-conquer approach, 109, 110; hill climbing, 95–96; illustrating inference methods in, 96–106; inference under MDC criterion, 96–98, 99; major trees, 112; Markov chain Monte Carlo (MCMC), 92–93, 94; maximum likelihood inference, 98–102; maximum pseudolikelihood inference, 102–3, 104, 105, 106–7;

- PhyloNet (cont.)
optimization problems, 92; reticulate evolutionary processes in, 112–17; reticulation, 90; running time, 105–6, 107; software, 16–17, 117–19; tree-based augmentation, 107, 109; tree decompositions, 112; tripartitions, 112
- PhyloNetworks, 146, 176; Julia programming language, 133, 143–44; main functions in, 143–44
- Physalis peruviana*, 165, 166. *See also* hybridization; *lochroma*
- Plant and Fungal Tree of Life Project (PAFTOL), 150
- plant classification, 149–50
- plant life, resolution of tree of, 160
- polymerase chain reaction (PCR), 7, 8; DNA amplification, 192, 208; genetic data, 150; microfluidic, 180
- polyploids: analysis of, 114–17; phylogenetics of hybrids and, 189–90
- polyploidy, 146, 180; events, 114, 117; hybridization and, 190. *See also* *Penstemon*
- polytomies, 48, 50, 151; ASTRAL for, 65; tests for, 65
- PoMo method, 41
- Potts model, 263; community detection, 273
- primary concordance tree (PCT), 175
- probability distribution: branch lengths in gene tree histories, 74; gene tree topologies, 4, 5, 95; gene trees, 90; posterior, 17; site pattern, 69; SVDQuartets, 80
- Proceri subsection, 146. *See also* *Penstemon*
- ProgressiveCactus, 225, 227
- pseudolikelihood, 124; composite likelihood, 84; gene trees, 92; measure, 90; methods, 129–30; model, 134, 139. *See also* maximum pseudolikelihood estimation; maximum pseudolikelihood inference
- Python: DendroPy, 193; *ipcoal* package, 252; *msprime* program, 252; *scipy* package in, 256; *toytree* package, 252
- QCF. *See* quartet concordance factors (QCF)
- quartet concordance factors (QCFs), 146; calculation of, 177–78; SNaQ (species networks applying quartets) using, 176; validation of QCF estimation, 178–79
- Quartet MaxCut, 133; species network inference, 182, 184, 185, 186; taxonomy of *Penstemon* subsections, 188
- random genomic sequencing, 6
- RAxML, 57, 167; gene tree inference, 257; introgression under uncertainty in birds, 244; maximum likelihood, 32, 33; phylogeny of *Penstemon* subsections with, 185; species tree inference, 181, 185; species tree method, 219
- Ray, John, 232–33
- regulatory network, genomes, 215
- resolution of tree of life, 160
- restriction site-associated DNA sequencing (RADseq), 6, 8, 9, 39, 80, 146–47
- reticulate evolution, 112; ancestral recombination graph, 113; inference results for *Leucanthemopsis* representatives, 116–17, 118; *Leucanthemopsis* test data, 115; processes in PhyloNet, 112–17
- reticulation: ASTRAL, 51; direction of, 136–37; PhyloNet, 90; reticulation edges, 91; reticulation nodes, 91
- revPoMo method, 40–41
- Robinson-Foulds distance: between genealogies, 255–56; between true and estimated gene trees, 51, 53; computing, 128–29; to guide tree, 35; topological discordance, 249; weighted, 265
- Robinson-Foulds supertree (RFS) problem, 29; FastMulRFS, 29
- running time, 16, 21; ASTRAL, 22–24, 47, 49, 50, 54; ASTRID, 24; Bayesian coestimation method, 25; BBCA analysis, 25–26; Big-O analysis, 41–42; divide-and-conquer approach, 34, 35; phylogenetic analysis, 105–6, 107, 110; PhyloNet, 90; species tree estimation, 28, 30, 31, 33; SVDQuartets, 27
- selection: heterogeneity in wild tomatoes, 246–47; phylogenetic tests of, 245–47
- sequence evolution model, 44
- sequencing technologies, 6, 7, 13, 192, 215
- Ship of Theseus, 234–35, 250
- Simpson, George Gaylord, 236
- simulations: accuracy of ASTRAL in, 52; genealogical variation among chromosomes, 252–56
- single-nucleotide polymorphism (SNP), 146, 192; ABBA-BABA-tests for, 122; analysis with eco-evolity, 195–96, 198, 199, 200, 201, 202, 203, 204, 205; data, 9, 16, 27–28, 70; SVDQuartets, 78–80, 82
- Singular Value Decomposition: SVDQuartets, 27; SVDquest, 27–28
- siskins (*Spinus*), introgression under uncertainty in birds, 244–45
- site pattern frequencies, 10, 13; HyDe tests, 181; quartet concordance factors (QCFs), 176; simplified model, 71, 72; species tree estimation, 68–70, 82–88; SVDQuartets, 80, 82
- sliding window analyses, gene tree inference, 252, 257, 259
- SNAPP, 16, 21, 244
- SNaQ (species networks applying quartets): best networks estimated by, 137–38; choice of number of hybridizations, 139; determining exact hybrid relationships, 172–73; flow chart of procedure, 131; identification of hybrid relationships, 170, 171; maximum pseudolikelihood (ML) networks by, 187; pseudolikelihood method, 129–30; rooting of, 136–38; species network inference, 186; taxonomy of *Penstemon* subsections, 188–89; testing of hybridization, 168; using quartet concordance factors (QCFs), 176
- Society of Systematic Biology, 14
- Solanum* sect. *Lycopersicon*, selection under heterogeneity in wild tomatoes, 246–47

- speciations, 236
- speciation times, 2, 12; algorithm details, 86; application to species relationships among gibbons, 87; composite likelihood, 84, 85; estimation of, 82–87; maximum a posteriori (MAP) estimation, 85–86; maximum composite likelihood estimates (MCLs), 84; pseudolikelihood, 84; recommendations for composite likelihood estimators of, 87; symmetric species tree with, 83; theoretical basis, 83–86; uncertainty quantification, 86
- species: concepts of, 236; divergence times, 12, 146, 228
- species network(s), 124–29; bootstrap analysis, 140–43; candidate networks compatible with known outgroup, 137–39; comparison of, 128–29; displayed trees and subnetworks, 128, 129; explicit versus implicit, 126–27; extended Newick format, 127, 128; extended parenthetical format, 127–28; fast reconstruction of, 129–43; goodness of fit tools, 139–40; identifiability of pseudolikelihood model, 134–36; level-1 and level-2 networks, 126; maximum pseudolikelihood estimation, 130–36; notation for, 124; rooted and semidirected phylogenetic networks, 124, 125; rooted four-taxon network and semidirected version, 130, 132; rooting of semidirected networks, 136–39; SNaQ (species networks applying quartets), 129–33; visual artifacts misleading interpretation, 126
- species network inference: analysis of larger data sets, 106–11; divide-and-conquer, 109; heuristic searches versus MCMC (Markov chain Monte Carlo) sampling, 93, 94; identifiability of network's parameters, 93; illustration of methods in PhyloNet, 96–106; maximum likelihood inference, 98–102; maximum pseudolikelihood inference, 102–3, 104, 105, 106–7; MCMC simulations, 92–93, 95; MDC (minimizing deep coalescence) criterion, 96–98; phylogeny of *Penstemon* subsections using ASTRAL-III, 182, 183; phylogeny of *Penstemon* subsections using QuartetMaxCut, 182, 184; phylogeny of *Penstemon* subsections using RAxML, 182, 185; PhyloNet, 117–19; PhyloNetworks Julia package installation and use, 143–44; reading and interpreting a phylogenetic network, 91–92; reticulate evolutionary processes in PhyloNet, 112–17; summarizing and comparing networks, 111–12; tests for hybridization and, 186; tree-based augmentation, 107, 109; trees versus networks, 122–24. *See also* species network(s)
- species phylogeny, 2, 3. *See also* phylogeny
- species tree(s), 2, 3, 213, 236–38; concept of, 4; controversies in estimation of, 11–12; illustration of model, 193; reconstruction and gene tree conflict, 154–56, 157; three-axon, 5; topology, 253
- species tree estimation, 1, 14, 19–21, 68–69; approaches to, 45; ASTRAL algorithm, 22–24; ASTRAP-MP, 30–31; ASTRID, 24; BBCE, 25–26; coestimation methods, 24–26; divide-and-conquer, 33–36; empirical data in need of exploration, 209–10; empirical performance for method choice, 37–39; evaluation of branch support in species trees, 28; ExaML, 32; future prospects for, 12–13; gene duplication and loss (GDL), 20, 29–30; incomplete lineage sorting (ILS), 20; introduction of site pattern errors, 193–94; maximum likelihood, 20; methods addressing ILS, 21–28; model-based methods of, 191–92; multilocus, using maximum likelihood, 31–33; multispecies coalescent (MSC), 20; overview of current methods for, 9–12; parallel implementations for, 30–33; phylogeny of *Penstemon* subsections with RAxML, 181, 185; quartet sampling and assembly, 75–76; RAxML, 32; recommendations for unlinked-character models, 209; sensitivity to models of sequence evolution, 264–68; simulations of error-free data sets, 192–93; site-based methods, 26–28; statistical consistency for method choice, 36–37; summary, challenges and future directions, 39–41; summary methods, 21–24; SVDQuartets, 27; SVDquest, 27–28; topology using SVDQuartets, 69–82. *See also* ecoevolity; StarBEAST2; SVDQuartets
- Spinus* (siskins), introgression under uncertainty in birds, 244–45
- StarBEAST2: accuracy and precision of divergence time estimates, 196, 198, 199; accuracy and precision of effective size of descendant populations, 195, 203, 204, 205; accuracy and precision of root effective population size, 195, 200, 201, 202; assessing sensitivity to errors, 194; behavior of linked, versus ecoevolity (unlinked) character models, 195; coverage of credible intervals, 197; empirical data in need of exploration, 209–10; MCMC-based method, 16, 17; MCMC convergence and mixing, 197; recommendations for using character models, 209; robustness to character-pattern errors, 207–8; species tree estimation, 21, 25
- STRUCTURE plots, 241
- subnetworks, 119n4: displayed trees and, 128, 129; four-taxon, 130, 132, 139; three-taxon, 111
- summary methods: species tree estimation, 21–24; species tree inference, 10
- supertree methods, 53, 150; Bayesian, 41; divide-and-conquer, 34–35, 39; Robinson-Foulds (RFS), 29
- SVDQuartets, 68; accounting of incomplete lineage sorting (ILS), 74–75; algorithmic details, 76; application to species relationships among gibbons, 78–79; assumptions and misconceptions, 79–81; data filtering, 81–82; data generation process for species-level phylogeny with three species, 74; estimation of speciation times, 82–87; estimation of species tree topology using, 69–82; flattening matrices, 69–70, 73; future work of, 87–88; introgression under uncertainty in birds, 244; method, 15, 16, 40; missing data, 81–82; possible site patterns under simplified model, 71, 72; properties of, 79–82;

- SVDQuartets (cont.)
quartet sampling and assembly, 75–76; rank reductions for flattening matrices, 71–73; recommendations for using, 82; schematic of estimation procedure, 77; site-based methods, 21; species tree inference, 75–76; statistical consistency, 81; SVD-Score, 70, 73, 75, 76, 77; theoretical basis, 69–73; uncertainty quantification, 78
- SVDquest, 27–28, 40
- SVDscore, 70, 73, 75, 76, 77
- Symphalangus syndactylus*, 78, 79
- system, 232, 250; history of, 232–34
- Tanglegrams, 249
- targeted enrichment, sequencing technology, 6, 7
- targeted genotyping-by-sequencing (GBS), 6, 8
- 10kp (Ten Thousand Genomes Project), 150
- terminology, 2–9
- TICR (tree incongruence checking in R), 122, 124, 146, 167
- tomatoes, selection under heterogeneity in wild, 246–47
- traits: coevolution of genes and, in *Jaltomata*, 248; testing, under phylogenetic heterogeneity, 247–48
- transcriptome sequencing technology, 6, 7
- tree-based augmentation, analysis of larger data sets, 107, 109
- TreeMerge, disjoint tree merger method, 35, 36
- TreeScaper software, 213 261
- tree sets, 212–13; bipartition covariance networks, 264, 266, 273–75; community detection methods, 262–64, 273; curvilinear components analysis (CCA), 261, 270; detection of structure in, 262–64; *erato-sara* clade, 272, 274; exploring, 260–64; joint versus independent inference of gene trees, 268–71; Jukes-Cantor (JC) model, 265, 266, 267, 268; maximum a posteriori (MAP), 266, 267, 268; nonlinear dimensionality reduction (NLDR), 261, 269, 270, 271, 275; prospects for future development and application, 275; sequence evolution models, 265, 266, 267, 268; topological affinity network with trees in distinct communities, 263; visualization of, 260–64
- TreeSetViz module, 261
- TreeShrink, 58, 59
- tripartitions, 31; ASTRAL, 49; phylogenetic network, 111, 112; weight of, 47, 48
- unlinked genealogies, 252, 253–56
- Vibrio*, 248; coevolution with oyster larvae, 248
- von Neumann, John, 119n6
- Wallace, Alfred Russel, 233
- whole-genome sequencing, 6, 7, 15, 218
- zones, 233