Phylogenetics

## I. The Basics of Phylogenetics

What is phylogenetics?

Phylogenetics is the study of evolutionary relatedness among a group of organisms. Molecular phylogenetics uses DNA sequence data to infer these relationships. In combination with temporal and geographical data, we can use phylogenetics to track the transmission dynamics of a given pathogen.

Phylogenetic trees

Phylogenetic trees are visualizations of the evolutionary relatedness between organisms. These trees are composed of tips, internal nodes, and branches.

- **Tips**: directly observed samples
- **Internal nodes**: hypothetical common ancestors between samples that are not directly observed
- **Branches**: form connections between nodes and tips or nodes and other nodes

Mutations occur along branches where the parent node does not have a mutation that is present in its descendants. This pattern of shared mutations enables hierarchical clustering which is visualized in the phylogenetic tree. We can obtain information about shared mutations through multiple sequence alignment as shown here:
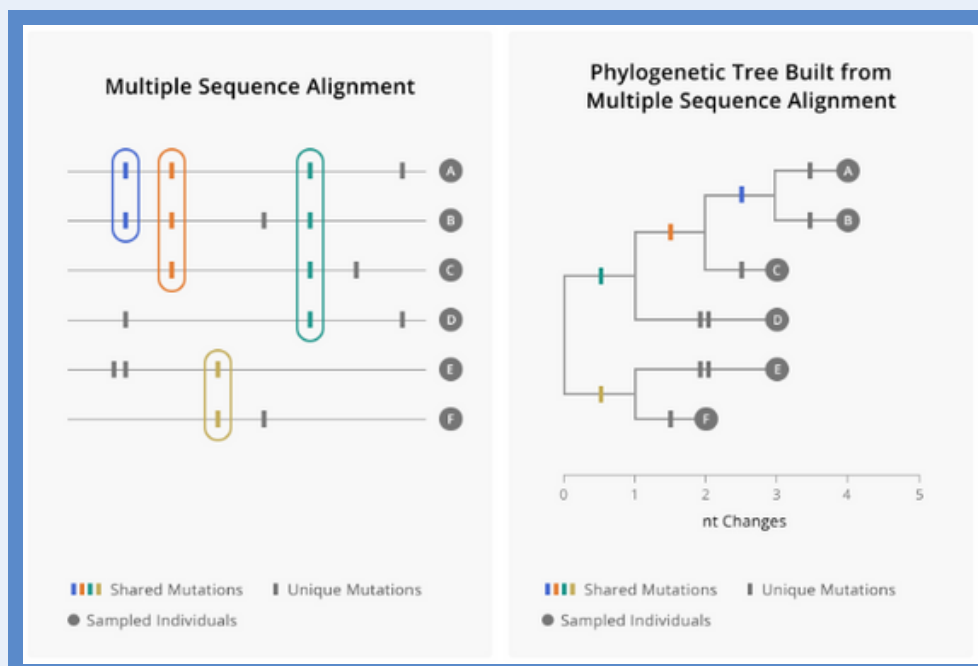
Trees can be both rooted and unrooted. Rooted trees stem from one node and has a unique path to every other node. Unrooted trees only specify the kinship between the samples in a tree, but not the evolutionary path. Trees can also be scaled and unscaled. In an unscaled tree, the branch lengths are not proportional to the number of changes. In a scaled tree, the branch lengths are proportional to the number of changes.

II. Constructing Phylogenetic Trees

Distance Matrix Methods

Distance methods take pairwise distances between sequences and reconstruct a tree the represents similarities between samples. While this is useful, this does not explicitly model how the sequences have evolved.

Estimating genetic relatedness from genome-wide SNPs

Here, we have an equation to determine genetic relatedness from SNP data:

$$A_{jk} = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_{ij} - 2p_i)(x_{ik} - 2pi)}{2p_i(1 - p_i)}$$

x_{ij} = number of copies of the reference allele for the i^{th}SNP of the j^{th}individual

x_{ik} = number of copies of the reference allele for the i^{th}SNP of the k^{th}individual

p_i = the frequency of the reference allele

N = the number of SNPs

A_{jk} = genetic relatedness matrix

For example: Organism A has 2 copies of the reference allele, organism B has 4, and organism C has 1. The frequency of the reference allele is 0.6

$$\text{Relatedness between A and B} \quad = \frac{(2 - 2 \times 0.6)(4 - 2 \times 0.6)}{2 \times 0.6(1 - 0.6)} = 4.67$$

$$\text{Relatedness between B and C} \quad = \frac{(4 - 2 \times 0.6)(1 - 2 \times 0.6)}{2 \times 0.6(1 - 0.6)} = -1.17$$

$$\text{Relatedness between A and C} \quad = \frac{(2 - 2 \times 0.6)(1 - 2 \times 0.6)}{2 \times 0.6(1 - 0.6)} = -0.34$$
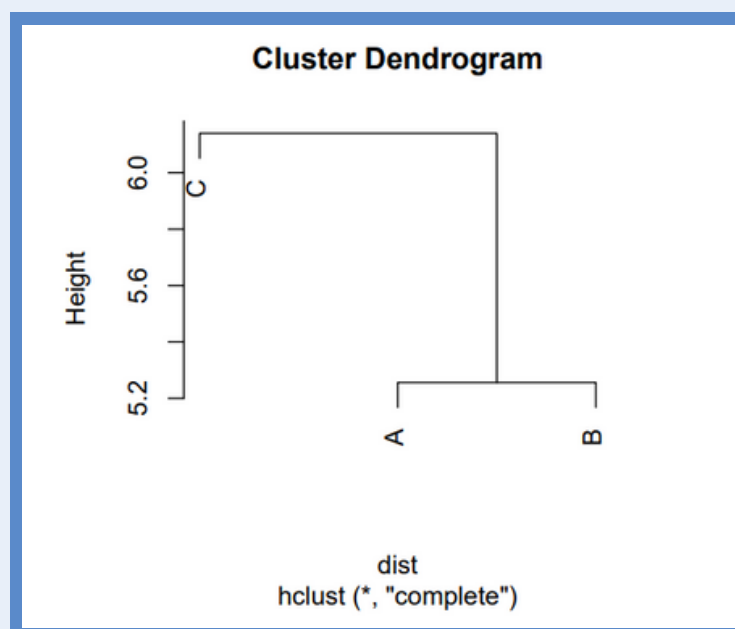
Creating the matrix:

	A	B	C
A	1	4.67	-0.34
B	4.67	1	-1.17
C	-.034	-1.17	1

Using R, we will create a phylogenetic tree from this matrix:

```
# Create Matrix
Mat = matrix(c(1, 4.67, -0.34, 4.67, 1, -1.17, -0.34, -1.17, 1),
nrow = 3, ncol = 3, byrow = TRUE)
rownames(Mat) = c("A", "B", "C")
colnames(Mat) = c("A", "B", "C")

# Create Tree
dist <- dist(Mat, method = "euclidean")
cluster <- hclust(dist)
plot(cluster)
```


Cluster Dendrogram

dist
hclust (*, "complete")

Unweighted pair-group method using an arithmetic average (UPGMA)

This method uses a matrix of pairwise genetic distances, first grouping the pair with the smallest distance, creating a new distance matrix, and continuing on as such. We begin our example with the following pairwise distance matrix:

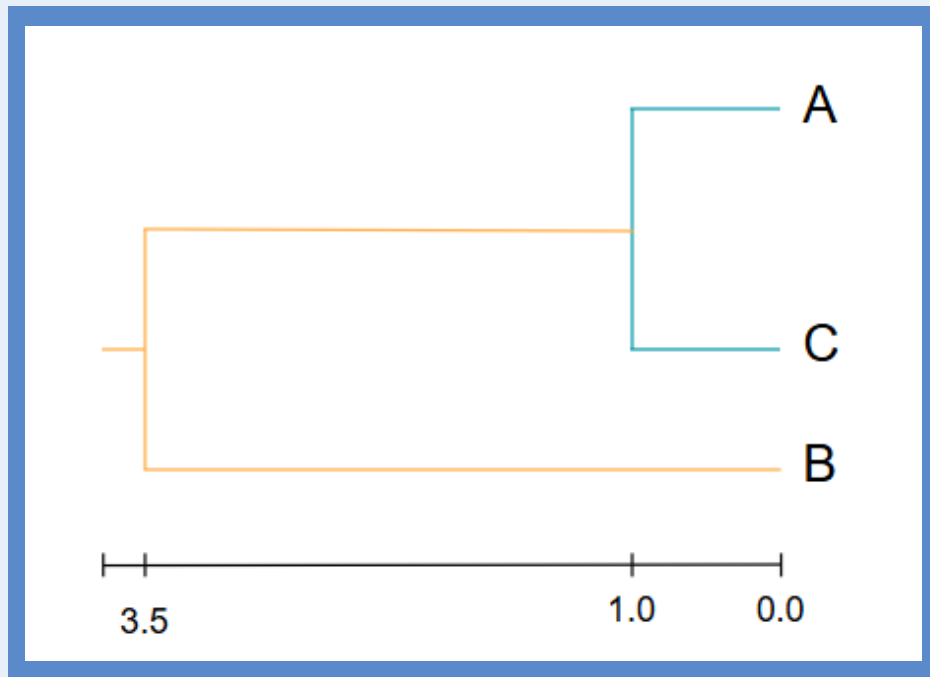|	A	B	C
A			
B	5		
C	2	9	

We find the shortest pairwise distance, here between A and C, which gives us the depth of the first new branch 2/2=1. We then calculate the mean pairwise distances with the other sequences to create a new matrix:

	A	B	C
A			
B	5		
C	2	9	

> $\dfrac{5+9}{2}=7$ >

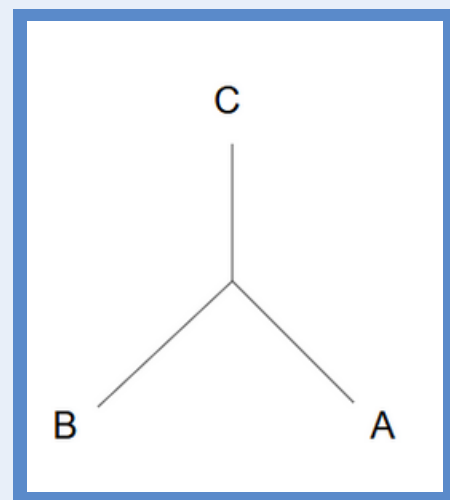|	AC	B
AC		
B	7	

This gives us a new branch length between B and AC of 7/2=3.5 which gives us this tree:

Neighbor-Joining

Neighbor-joining uses an iterative clustering process to determine which leaves are "neighbors" from a given distance matrix. Take the following distance matrix and starting tree:

	A	B	C
A	-	13	6
B		-	9
C			-

Compute $r_i' = \dfrac{\Sigma_j D_{i,j}}{n-2}$ for each terminal node

$$r_A' = \frac{D_{AB} + D_{AC}}{3-2} = \frac{13+6}{1} = 19$$

$$r_B' = \frac{D_{BA} + D_{BC}}{3-2} = \frac{13+9}{1} = 22$$

$$r_C' = \frac{D_{CA} + D_{CB}}{3-2} = \frac{6+9}{1} = 15$$

Compute $d_{i,j}' = d_{i,j} - r_i' - r_j'$ for each terminal node

$$d_{AB}' = 13 - 19 - 22 = -28$$

$$d_{AC}' = 6 - 19 - 15 = {\color{red}-28}$$
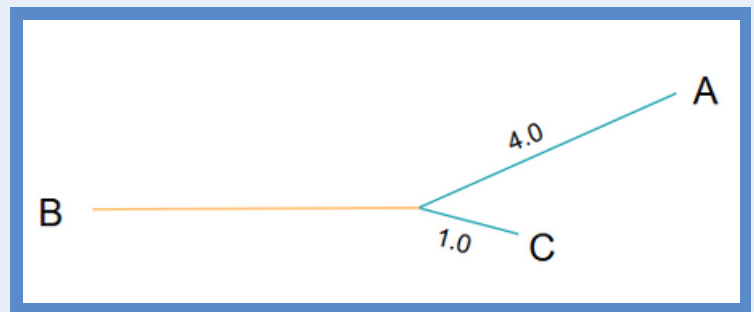
$$d_{BC}' = 9 - 22 - 15 = -28$$

Compute branch lengths for A and C node:

$$V_A = \frac{d_{A,C}}{2} + \frac{(r_A' - r_C')}{2} = \frac{6}{2} + \frac{(11-9)}{2} = 4$$

$$V_C = \frac{d_{C,A}}{2} + \frac{(r_C' - r_A')}{2} = \frac{6}{2} + \frac{9-11}{2} = 1$$

Create new distance matrix and tree:

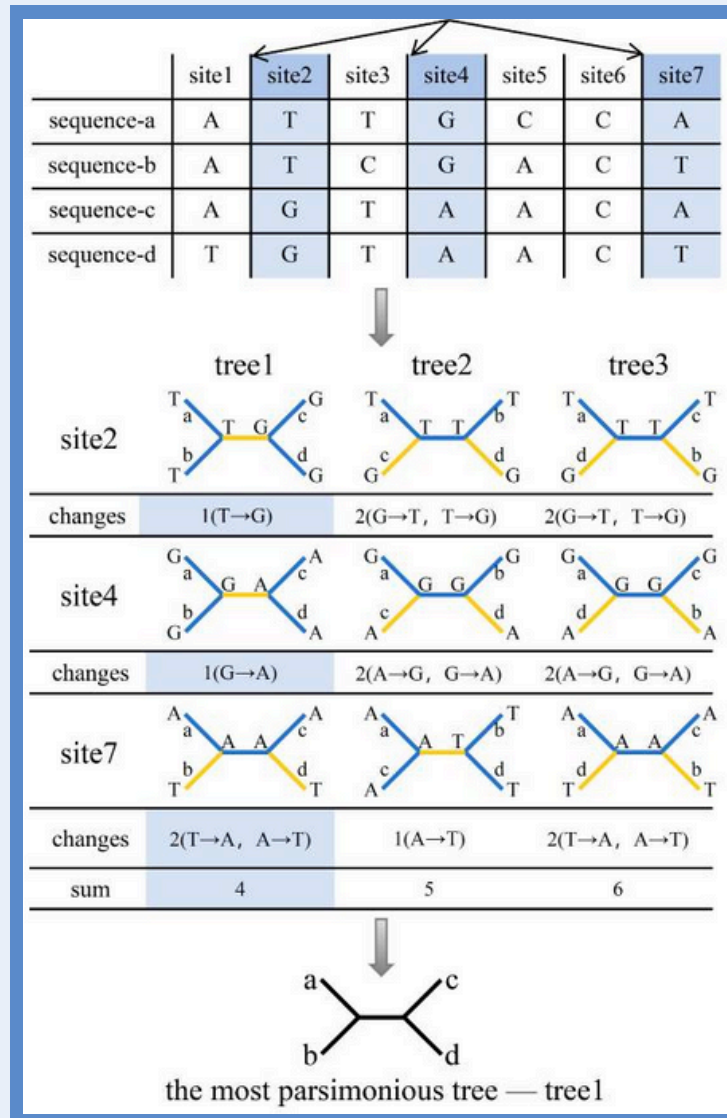|	AC	B
AC	-	{\color{red}16}
B		-

Maximum Parsimony, Maximum Likelihood, and Bayesian Methods

Maximum Parsimony

This method determines the topology of a tree that requires that minimum number of evolutionary changes necessary to create that tree. For example: if we have four sequences and have differing alleles at seven different sites, a tree can be constructed like this:

This method, in which each site is given equal weight, is called unweighted parsimony. There are instances where changes which occur less frequently (like transversions) are given more weight using weighted parsimony. Trees can be derived using an exhaustive search (look at all possible trees and choose the one with the least evolutionary changes), or a heuristic search (begin with a tree from a neighbor-joining algorithm, then search and modify from there).

Maximum Likelihood

The maximum likelihood method constructs a tree with given branch lengths based on the best estimate derived from the construction of many trees. For DNA sequence data, the model is based on the probability of one sequence changing by mutation to another sequence in a specified time.

The following equation can be used to determine the likelihood of the tree from a given point k with descendants i and j:

$$L_{s_k}^{(k)} = (\Sigma_{s_i} P_{s_k s_i}(v_i) L_{s_i}^{(i)})(\Sigma_{s_j} P_{s_k s_j}(v_j) L_{s_j}^{(j)})$$

$$\Sigma_{s_i} \text{ or } \Sigma_{s_j} = \text{Sum of state } i \text{ or } j$$
$$P_{s_k s_i} \text{ or } P_{s_k s_j} = \text{The probability of state } i \text{ or } j \text{ given point } k$$
$$(v_i) \text{ or } (v_j) = \text{The branch length from point } k \text{ to state } i \text{ or } j$$
$$L_{s_i}^{(i)} \text{ or } L_{s_j}^{(j)} = \text{The likelihood of state } i \text{ or } j$$

Computing this equation for all nodes gives you the likelihood of the tree. The likelihood of each tree configuration is computed to determine the tree with the maximum likelihood.

Bayesian Methods

Bayesian methods search for an average tree over sets of plausible trees. In this way, our search for the best tree is limited by prior information in the data. Refer to Bayes theorem for statistical inference:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}$$

$$p(\theta) = \text{Prior belief before observed data}$$
$$p(Y|\theta) = \text{Likelihood of Y}$$
$$p(\theta|Y) = \text{Posterior beliefs}$$
$$p(Y) = \int p(Y|\theta)p(\theta)d\theta$$

While all trees have equal prior probability, the posterior probability is determined via the formula above based on all possible tree configurations. The highest posterior probability is the most likely tree configuration.

The Newick Format

Most computer programs will use the Newick format (Cayley, 1857) which is where trees are represented in a linear form by a series of nested parentheses. The following tree would be described in Newick format at ((A,B),C):
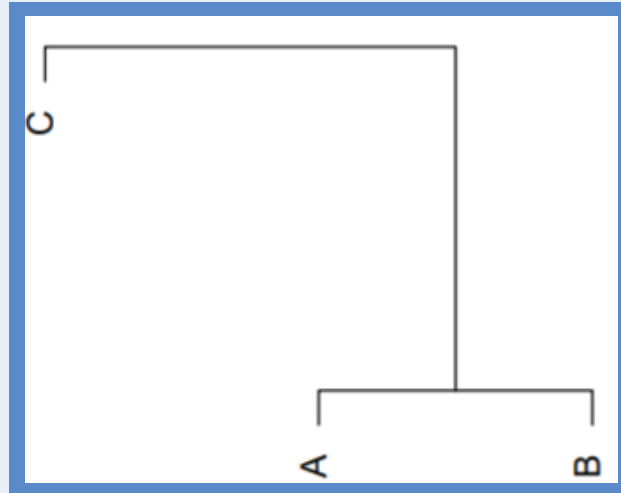
III. Software

Name	Description	Methods	Author
ADMIXTOOLS	R software package that contains the qpGraph, qpAdm, qpWave, and qpDstat programs		Nick Patterson, David Reich
AncesTree	An algorithm for clonal tree reconstruction from multi-sample cancer sequencing data.	Maximum Likelihood, Integer Linear Programming (ILP	M. El-Kebir, L. Oesper, H. Acheson-Field, B. J. Raphael
AliGROOVE	Visualisation of heterogeneous sequence divergence within multiple sequence alignments and detection of inflated branch support	Identification of single taxa which show predominately randomized sequence similarity in comparison with other taxa in a multiple sequence alignment and evaluation of the reliability of node support in a given topology	Patrick Kück, Sandra A Meid, Christian Groß, Bernhard Misof, Johann Wolfgang Wägele.
ape	R-Project package for analysis of phylogenetics and evolution	Provides a large variety of phylogenetics functions	Maintainer: Emmanuel Paradis

Armadillo Workflow Platform	Workflow platform dedicated to phylogenetic and general bioinformatic analysis	Inference of phylogenetic trees using Distance, Maximum Likelihood, Maximum Parsimony, Bayesian methods and related workflows	E. Lord, M. Leclercq, A. Boc, A.B. Diallo and V. Makarenkov
BAli-Phy	Simultaneous Bayesian inference of alignment and phylogeny	Bayesian inference, alignment as well as tree search	M.A. Suchard, B. D. Redelings
BATWING	Simultaneous Bayesian inference of alignment and phylogeny	Bayesian inference, alignment as well as tree search	M.A. Suchard, B. D. Redelings
BayesPhylogenies	Bayesian inference of trees using Markov chain Monte Carlo methods	Bayesian inference, multiple models, mixture model (auto-partitioning)	M. Pagel, A. Meade
BayesTraits	Analyses trait evolution among groups of species for which a phylogeny or sample of phylogenies is available	Trait analysis	M. Pagel, A. Meade

BioNumerics	Universal platform for the management, storage and analysis of all types of biological data, including tree and network inference of sequence data	Neighbor-joining, maximum parsimony, UPGMA, maximum likelihood, distance matrix methods,... Calculation of the reliability of trees/branches using bootstrapping, permutation resampling or error resampling	L. Vauterin & P. Vauterin.
Bosque	Integrated graphical software to perform phylogenetic analyses, from the importing of sequences to the plotting and graphical edition of trees and alignments	Distance and maximum likelihood methods (through PhyML, PHYLIP, Tree-Puzzle)	S. Ramirez, E. Rodriguez.
BUCKy	Bayesian concordance of gene trees	Bayesian concordance using modified greedy consensus of unrooted quartets	C. Ané, B. Larget, D.A. Baum, S.D. Smith, A. Rokas and B. Larget, S.K. Kotha, C.N. Dewey, C. Ané

Canopy	Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing	Maximum Likelihood, Markov Chain Monte Carlo (MCMC) methods	Y. Jiang, Y. Qiu, A. J. Minn, and N. R. Zhang
CGRphylo	CGR method for accurate classification and tracking of rapidly evolving viruses	Chaos Game Representation (CGR) method, based on concepts of statistical physics	Amarinder Singh Thind, Somdatta Sinha
CITUP	Clonality Inference in Tumors Using Phylogeny	Exhaustive search, Quadratic Integer Programming (QIP)	S. Malikic, A.W. McPherson, N. Donmez, C.S. Sahinalp
ClustalW	Progressive multiple sequence alignment	Distance matrix/nearest neighbor	Thompson et al.
CoalEvol	Simulation of DNA and protein evolution along phylogenetic trees (that can also be simulated with the coalescent)	Simulation of multiple sequence alignments of DNA or protein sequences	M. Arenas, D. Posada

CodABC	Coestimation of substitution, recombination and dN/dS in protein sequences	Approximate Bayesian computation	M. Arenas, J.S. Lopes, M.A. Beaumont, D. Posada
Dendroscope	Tool for visualizing rooted trees and calculating rooted networks	Rooted trees, tanglegrams, consensus networks, hybridization networks	Daniel Huson et al.
EXACT	EXACT is based on the perfect phylogeny model, and uses a very fast homotopy algorithm to evaluate the fitness of different trees, and then it brute forces the tree search using GPUs, or multiple CPUs, on the same or on different machines	Brute force search and homotopy algorithm	Jia B., Ray S., Safavi S., Bento J.

EzEditor	EzEditor is a java-based sequence alignment editor for rRNA and protein coding genes. It allows manipulation of both DNA and protein sequence alignments for phylogenetic analysis	Neighbor Joining	Jeon, Y.S. et al.
fastDNAml	Optimized maximum likelihood (nucleotides only)	Maximum likelihood	G.J. Olsen
FastTree	Fast phylogenetic inference for alignments with up to hundreds of thousands of sequences	Approximate maximum likelihood	M.N. Price, P.S. Dehal, A.P. Arkin
fitmodel	Fits branch-site codon models without the need of prior knowledge of clades undergoing positive selection	Maximum likelihood	S. Guindon

Geneious	Geneious provides genome and proteome research tools	Neighbor-joining, UPGMA, MrBayes plugin, PhyML plugin, RAxML plugin, FastTree plugin, GARLi plugin, PAUP* Plugin	A. J. Drummond, M.Suchard, V.Lefort et al.
HyPhy	Hypothesis testing using phylogenies	Maximum likelihood, neighbor-joining, clustering techniques, distance matrices	S.L. Kosakovsky Pond, S.D.W. Frost, S.V. Muse
IQPNNI	Iterative ML treesearch with stopping rule	Maximum likelihood, neighbor-joining	L.S. Vinh, A. von Haeseler, B.Q. Minh
IQ-Tree	An efficient phylogenomic software by maximum likelihood, as successor of IQPNNI and Tree-Puzzle	Maximum likelihood, model selection, partitioning scheme finding, AIC, AICc, BIC, ultrafast bootstrapping,[20] branch tests, tree topology tests, likelihood mapping	Lam-Tung Nguyen, O. Chernomor, H.A. Schmidt, A. von Haeseler, B.Q. Minh
jModelTest	A high-performance computing program to carry out statistical selection of best-fit models of nucleotide substitution	Maximum likelihood, AIC, BIC, DT, hLTR, dLTR	D. Darriba, GL. Taboada, R. Doallo, D. Posada

JolyTree	An alignment-free bioinformatics procedure to infer distance-based phylogenetic trees from genome assemblies, specifically designed to quickly infer trees from genomes belonging to the same genus	MinHash-based pairwise genome distance, Balanced Minimum Evolution (BME), ratchet-based BME tree search, Rate of Elementary Quartets	A. Criscuolo
LisBeth	Three-item analysis for phylogenetics and biogeography	Three-item analysis	J. Ducasse, N. Cao & R. Zaragüeta-Bagils
MEGA	Molecular Evolutionary Genetics Analysis	Distance, Parsimony and Maximum Composite Likelihood Methods	Tamura K, Dudley J, Nei M & Kumar S
MegAlign Pro	MegAlign Pro is part of DNASTAR's Lasergene Molecular Biology package. This application performs multiple and pairwise sequence alignments, provides alignment editing, and generates phylogenetic trees.	Maximum Likelihood (RAxML) and Neighbor-Joining	DNASTAR

Mesquite	Mesquite is software for evolutionary biology, designed to help biologists analyze comparative data about organisms. Its emphasis is on phylogenetic analysis, but some of its modules concern comparative analyses or population genetics, while others do non-phylogenetic multivariate analysis. It can also be used to build timetrees incorporating a geological timescale, with some optional modules.	Maximum parsimony, distance matrix, maximum likelihood	Wayne Maddison and D. R. Maddison
MetaPIGA2	Maximum likelihood phylogeny inference multi-core program for DNA and protein sequences, and morphological data. Analyses can be performed using an extensive and user-friendly graphical interface or by using batch files. It also implements tree visualization tools, ancestral sequences, and automated selection of best substitution model and parameters.	Maximum likelihood, stochastic heuristics (genetic algorithm, metapopulation genetic algorithm, simulated annealing, etc.), discrete Gamma rate heterogeneity, ancestral state reconstruction, model testing	Michel C. Milinkovitch and Raphaël Helaers

MicrobeTrace	MicrobeTrace is a free, browser-based web application.	2D and 3D network visualization tool, Neighbor-joining tree visualization, Gantt charts, bubbles charts, networks visualized on maps, flow diagrams, aggregate tables, epi curves, histograms, alignment viewer, and much more.	Ellsworth M. Campbell, Anthony Boyles, Anupama Shankar, Jay Kim, Sergey Knyazev, Roxana Cintron, William M. Switzer
MNHN-Tree-Tools	MNHN-Tree-Tools is an opensource phylogenetics inference software working on nucleic and protein sequences.	Clustering of DNA or protein sequences and phylogenetic tree inference from a set of sequences. At the core it employs a distance-density based approach.	Thomas Haschka, Loïc Ponger, Christophe Escudé and Julien Mozziconacci
Modelgenerator	Model selection (protein or nucleotide)	Maximum likelihood	Thomas Keane
MOLPHY	Molecular phylogenetics (protein or nucleotide)	Maximum likelihood	J. Adachi and M. Hasegawa
MorphoBank	Web application to organize trait data (morphological characters) for tree building	for use with Maximum Parsimony (via the CIPRES portal), Maximum Likelihood, and Bayesian analysis)	O'Leary, M. A., and S. Kaufman, also K. Alphonse

MrBayes	Posterior probability estimation	Bayesian inference	J. Huelsenbeck, et al.
Network	Free Phylogenetic Network Software	Median Joining, Reduced Median, Steiner Network	A. Roehl
Nona	Phylogenetic inference	Maximum parsimony, implied weighting, ratchet	P. Goloboff
PAML	Phylogenetic analysis by maximum likelihood	Maximum likelihood and Bayesian inference	Z. Yang
ParaPhylo	Computation of gene and species trees based on event-relations (orthology, paralogy)	Cograph-Editing and Triple-Inference	Hellmuth
PartitionFinder	Combined selection of models of molecular evolution and partitioning schemes for DNA and protein alignments	Maximum likelihood, AIC, AICc, BIC	R. Lanfear, B Calcott, SYW Ho, S Guindon
PASTIS	R package for phylogenetic assembly	R, two-stage Bayesian inference using MrBayes 3.2	Thomas et al. 2013

PAUP	Phylogenetic analysis using parsimony (*and other methods)	Maximum parsimony, distance matrix, maximum likelihood	D. Swofford
phangorn	Phylogenetic analysis in R	ML, MP, distance matrix, bootstrap, phylogentic networks, bootstrap, model selection, SH-test, SOWH-test	K. Schliep
Phybase	an R package for species tree analysis	phylogenetics functions, STAR, NJst, STEAC, maxtree, etc	L. Liu & L. Yu
phyclust	Phylogenetic Clustering (Phyloclustering)	Maximum likelihood of Finite Mixture Modes	Wei-Chen Chen
PHYLIP	PHYLogeny Inference Package	Maximum parsimony, distance matrix, maximum likelihood	J. Felsenstein
phyloT	Generates phylogenetic trees in various formats, based on NCBI taxonomy		I. Letunic
PhyloQuart	Quartet implementation (uses sequences or distances)	Quartet method	V. Berry

PhyloWGS	Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors	MCMC	A. G. Deshwar, S. Vembu, C. K. Yung, G. H. Jang, L. Stein, and Q. Morris
PhyML	Fast and accurate estimation of phylogenies using maximum likelihood	Maximum likelihood	S. Guindon & O. Gascuel
phyx	Unix/Linux command line phylogenetic tools	Explore, manipulate, analyze, and simulate phylogenetic objects (alignments, trees, and MCMC logs)	J.W. Brown, J.F. Walker, and S.A. Smith
POY	A phylogenetic analysis program that supports multiple kinds of data and can perform alignment and phylogeny inference. A variety of heuristic algorithms have been developed for this purpose	Maximum parsimony, Maximum likelihood, Chromosome rearrangement, discreet characters, continuous characters, Alignment	A. Varon, N. Lucaroni, L. Hong, W. Wheeler

ProtASR2	Ancestral reconstruction of protein sequences accounting for folding stability	Maximum likelihood, substitution models	M. Arenas, U. Bastolla
ProtEvol	Simulation of protein sequences under structurally constrained substitution models	Simulating sequences, substitution models	M. Arenas, A. Sanchez-Cobos, U. Bastolla U
ProteinEvolver	Simulation of protein sequences along phylogenies under empirical and structurally constrained substitution models of protein evolution	Simulating sequences forward in time, substitution models	M. Arenas, H.G. Dos Santos, D. Posada, U. Bastolla
ProteinEvolver ABC	Coestimation of recombination and substitution rates in protein sequences	Approximate Bayesian computation	M. Arenas
ProtTest3	A high-performance computing program for selecting the model of protein evolution that best fits a given set of aligned sequences	Maximum likelihood, AIC, BIC, DT	D. Darriba, GL. Taboada, R. Doallo, D. Posada

PyCogent	Software library for genomic biology	Simulating sequences, alignment, controlling third party applications, workflows, querying databases, generating graphics and phylogenetic trees	Knight et al.
QuickTree	Tree construction optimized for efficiency	Neighbor-joining	K. Howe, A. Bateman, R. Durbin
RAxML-HPC	Randomized Axelerated Maximum Likelihood for High Performance Computing (nucleotides and aminoacids)	Maximum likelihood, simple Maximum parsimony	A. Stamatakis
RAxML-NG	Randomized Axelerated Maximum Likelihood for High Performance Computing (nucleotides and aminoacids) Next Generation	Maximum likelihood, simple Maximum parsimony	A. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis

SEMPHY	Tree reconstruction using the combined strengths of maximum-likelihood (accuracy) and neighbor-joining (speed). SEMPHY has become outdated. The authors now refer users to RAxML, which is superior in accuracy and speed.	A hybrid maximum-likelihood – neighbor-joining method	M. Ninio, E. Privman, T. Pupko, N. Friedman
SGWE	Simulation of genome-wide evolution along phylogenetic trees	Simulating genome-wide sequences forward time	Arenas M., Posada D.
SimPlot++	Sequence similarity plots (SimPlots), detection of intragenic and intergenic recombination events, bootscan analysis and sequence similarity networks	SimPlot using different nucleotide/protein distance models; Phi, χ^2 and NSS recombination tests; Sequence similarity network analysis	S. Samson, E. Lord, V. Makarenkov
sowhat	Hypothesis testing	SOWH test	Church, Ryan, Dunn

Splatche3	Simulation of genetic data under diverse spatially explicit evolutionary scenarios	Coalescent, molecular evolution, DNA sequences, SNPs, STRs, RFLPs	M. Currat et al.
SplitsTree	Tree and network program	Computation, visualization and exploration of phylogenetic trees and networks	D.H. Huson and D. Bryant
TNT	Phylogenetic inference	Parsimony, weighting, ratchet, tree drift, tree fusing, sectorial searches	P. Goloboff et al.
TOPALi	Phylogenetic inference	Phylogenetic model selection, Bayesian analysis and Maximum Likelihood phylogenetic tree estimation, detection of sites under positive selection, and recombination breakpoint location analysis	Iain Milne, Dominik Lindner et al.
TreeGen	Tree construction given precomputed distance data	Distance matrix	ETH Zurich

TreeAlign	Efficient hybrid method	Distance matrix and approximate parsimony	J. Hein
TreeLine	Tree construction algorithm within the DECIPHER package for R	Maximum likelihood, maximum parsimony, and distance	E. Wright
Treefinder	Fast ML tree reconstruction, bootstrap analysis, model selection, hypothesis testing, tree calibration, tree manipulation and visualization, computation of sitewise rates, sequence simulation, many models of evolution (DNA, protein, rRNA, mixed protein, user-definable), GUI and scripting language	Maximum likelihood, distances, and others	Jobb G, von Haeseler A, Strimmer K
Tree-Puzzle	Maximum likelihood and statistical analysis	Maximum likelihood	Makarenkov

T-REX (Webserver)	Tree inference and visualization, Horizontal gene transfer detection, multiple sequence alignment	Distance (neighbor joining), Parsimony and Maximum likelihood (PhyML, RAxML) tree inference, MUSCLE, MAFFT and ClustalW sequence alignments and related applications	Boc A, Diallo AB, Makarenkov V
UShER	Phylogenetic placement using maximum parsimony for viral genomes	Maximum parsimony	Turakhia Y, Thornlow B, Hinrichs AS, De Maio N, Gozashti L, Lanfear R, Haussler D and Corbett-Detig R
UGENE	Fast and free multiplatform tree editor	GUI with PHYLIP 3.6 and IQTree algorithms	Unipro
VeryFastTree	A highly-tuned tool that uses parallelizing and vectorizing strategies to speed inference of phylogenies for huge alignments	Approximate maximum likelihood	César Piñeiro. José M. Abuín and Juan C. Pichel
Winclada	GUI and tree editor (requires Nona)	Maximum parsimony, ratchet	K. Nixon

Xrate	Phylo-grammar engine	Rate estimation, branch length estimation, alignment annotation	I. Holmes

III. Sources

Estimating genetic relatedness from genome-wide SNPs:
Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011 Jan 7;88(1):76-82. doi: 10.1016/j.ajhg.2010.11.011. Epub 2010 Dec 17. PMID: 21167468; PMCID: PMC3014363

Unweighted pair-group method using arithmetic average:
Fitch WM, Margoliash E. Construction of phylogenetic trees. Science. 1967 Jan 20;155(3760):279-84. doi: 10.1126/science.155.3760.279. PMID: 5334057.

Neighbor-joining:
N Saitou, M Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees., Molecular Biology and Evolution, Volume 4, Issue 4, Jul 1987, Pages 406–425, https://doi.org/10.1093/oxfordjournals.molbev.a040454

Maximum Parsimony:
Zou Y, Zhang Z, Zeng Y, Hu H, Hao Y, Huang S, Li B. Common Methods for Phylogenetic Tree Construction and Their Implementation in R. Bioengineering (Basel). 2024 May 11;11(5):480. doi: 10.3390/bioengineering11050480. PMID: 38790347; PMCID: PMC11117635.

Maximum likelihood:
Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 1981;17(6):368-76. doi: 10.1007/BF01734359. PMID: 7288891.

Bayesian methods:
Z Yang, B Rannala, Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method., Molecular Biology and Evolution, Volume 14, Issue 7, Jul 1997, Pages 717–724, https://doi.org/10.1093/oxfordjournals.molbev.a025811

Software List:
https://en.wikipedia.org/wiki/List_of_phylogenetics_software