Molly Chen, Artem I. Kholodov and Laura A. Hug (2025). The evolution of the tree of life. Phil. Trans. R. Soc. B 380: 20240091. https://doi.org/10.1098/rstb.2024.0091

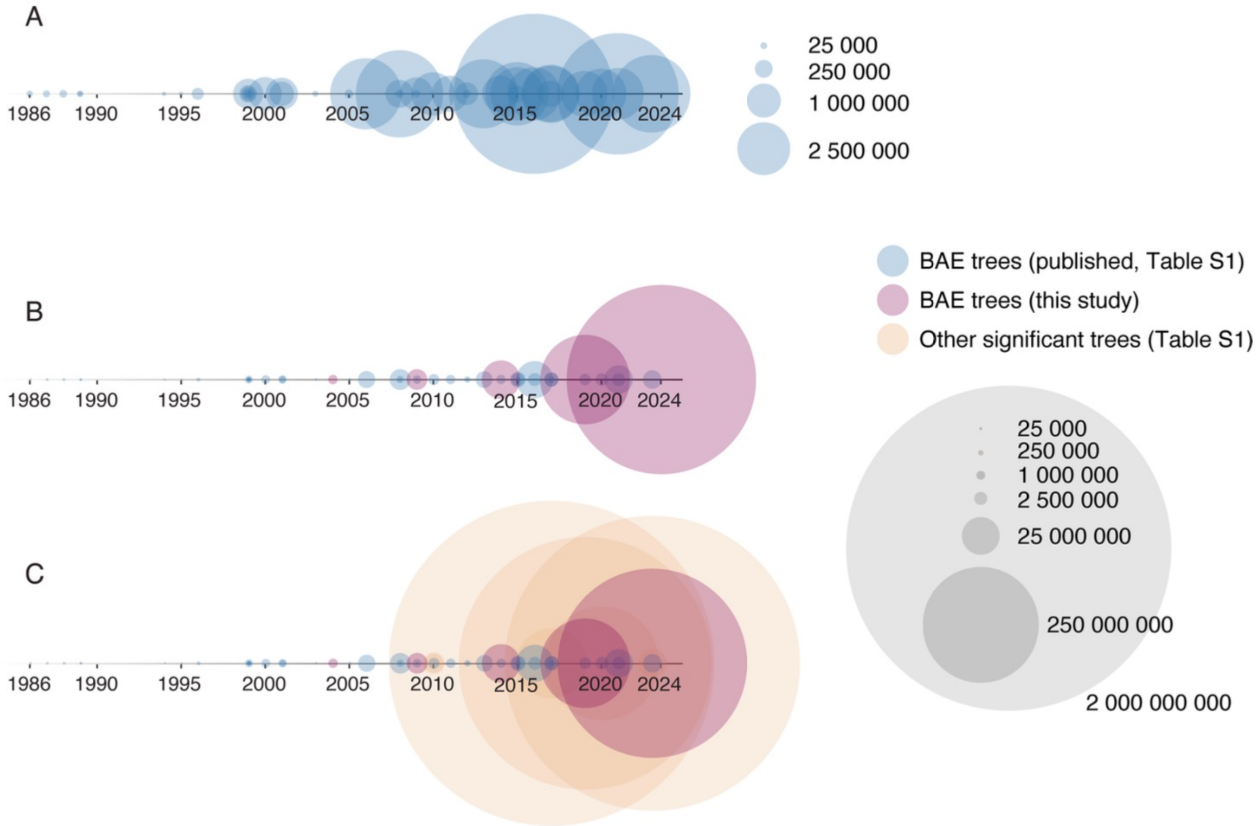
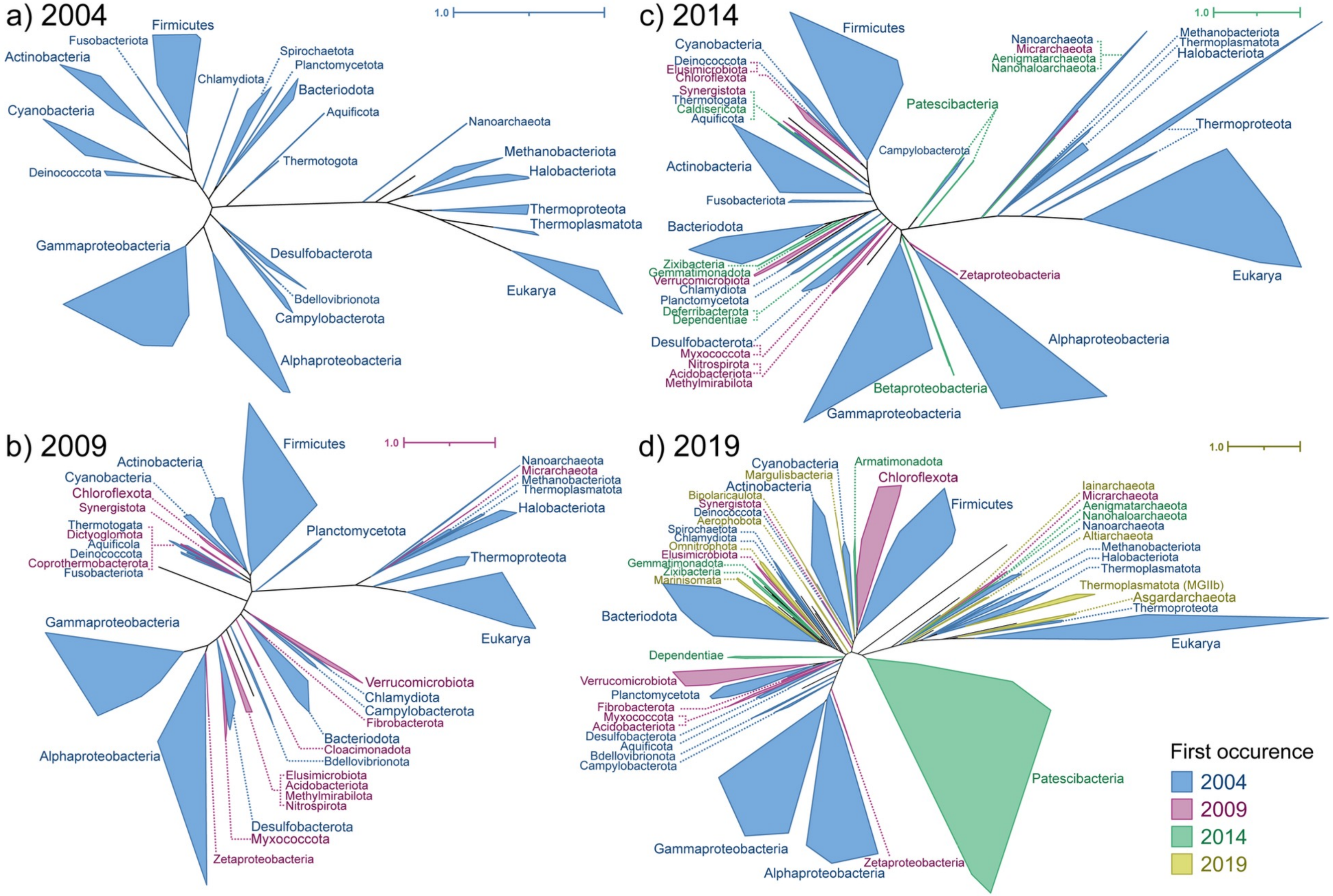


Figure 1. Timeline of universal trees of life, by total information included in the trees. Circles represent individual universal trees of life, with information calculated as # taxa × # alignment positions (×3 for amino acid alignments). (A) Published universal trees of life. (B): Published universal trees of life and reconstructed trees of life from this study (new scale). (C): Timeline in B, with other, large-scale two-domain trees included (orange), same scale as B.

Class focus area: Tree of life and tree metrics

EEB603: Brian O'Meara

All quotes and images from the above paper unless otherwise noted


a) 2004
1.0
Firmicutes
Fusobacteriota
Actinobacteria
Spirochaetota
Planctomycetota
Chlamydiota
Bacteriodota
Aquificota
Cyanobacteria
Thermotogota
Deinococcota
Nanoarchaeota
Methanobacteriota
Halobacteriota
Thermoproteota
Thermoplasmatota
Gammaproteobacteria
Desulfobacterota
Bdellovibrionota
Campylobacterota
Eukarya
Alphaproteobacteria
c) 2014
1.0
Firmicutes
Cyanobacteria
Deinococcota
Elusimicrobiota
Chloroflexota
Synergistota
Thermotogata
Caldisericota
Aquificota
Patescibacteria
Campylobacterota
Actinobacteria
Fusobacteriota
Bacteriodota
Zixibacteria
Gemmatimonadota
Verrucomicrobiota
Chlamydiota
Planctomycetota
Deferribacterota
Dependentiae
Desulfobacterota
Myxococcota
Nitrospirota
Acidobacteriota
Methylmirabilota
Nanoarchaeota
Micrarchaeota
Aenigmatarchaeota
Nanohaloarchaeota
Methanobacteriota
Thermoplasmatota
Halobacteriota
Thermoproteota
Eukarya
Zetaproteobacteria
Alphaproteobacteria
Betaproteobacteria
Gammaproteobacteria
b) 2009
1.0
Firmicutes
Actinobacteria
Cyanobacteria
Chloroflexota
Synergistota
Thermotogata
Dictyoglomota
Aquificola
Deinococcota
Coprothermobacterota
Fusobacteriota
Planctomycetota
Nanoarchaeota
Micrarchaeota
Methanobacteriota
Thermoplasmatota
Halobacteriota
Thermoproteota
Eukarya
Gammaproteobacteria
Verrucomicrobiota
Chlamydiota
Campylobacterota
Fibrobacterota
Bacteriodota
Cloacimonadota
Bdellovibrionota
Elusimicrobiota
Acidobacteriota
Methylmirabilota
Nitrospirota
Alphaproteobacteria
Desulfobacterota
Myxococcota
Zetaproteobacteria
d) 2019
1.0
Cyanobacteria
Armatimonadota
Margulisbacteria
Chloroflexota
Actinobacteria
Firmicutes
Bipolaricaulota
Synergistota
Deinococcota
Aerophobota
Spirochaetota
Chlamydiota
Omnitrophota
Elusimicrobiota
Gemmatimonadota
Zixibacteria
Marinisomata
Bacteriodota
Iainarchaeota
Micrarchaeota
Aenigmatarchaeota
Nanohaloarchaeota
Nanoarchaeota
Altiarchaeota
Methanobacteriota
Halobacteriota
Thermoplasmatota
Thermoplasmatota (MGIIb)
Asgardarchaeota
Thermoproteota
Eukarya
Dependentiae
Verrucomicrobiota
Planctomycetota
Fibrobacterota
Myxococcota
Acidobacteriota
Desulfobacterota
Aquificota
Bdellovibrionota
Campylobacterota
Patescibacteria
Gammaproteobacteria
Alphaproteobacteria
Zetaproteobacteria
First occurence
2004
2009
2014
2019

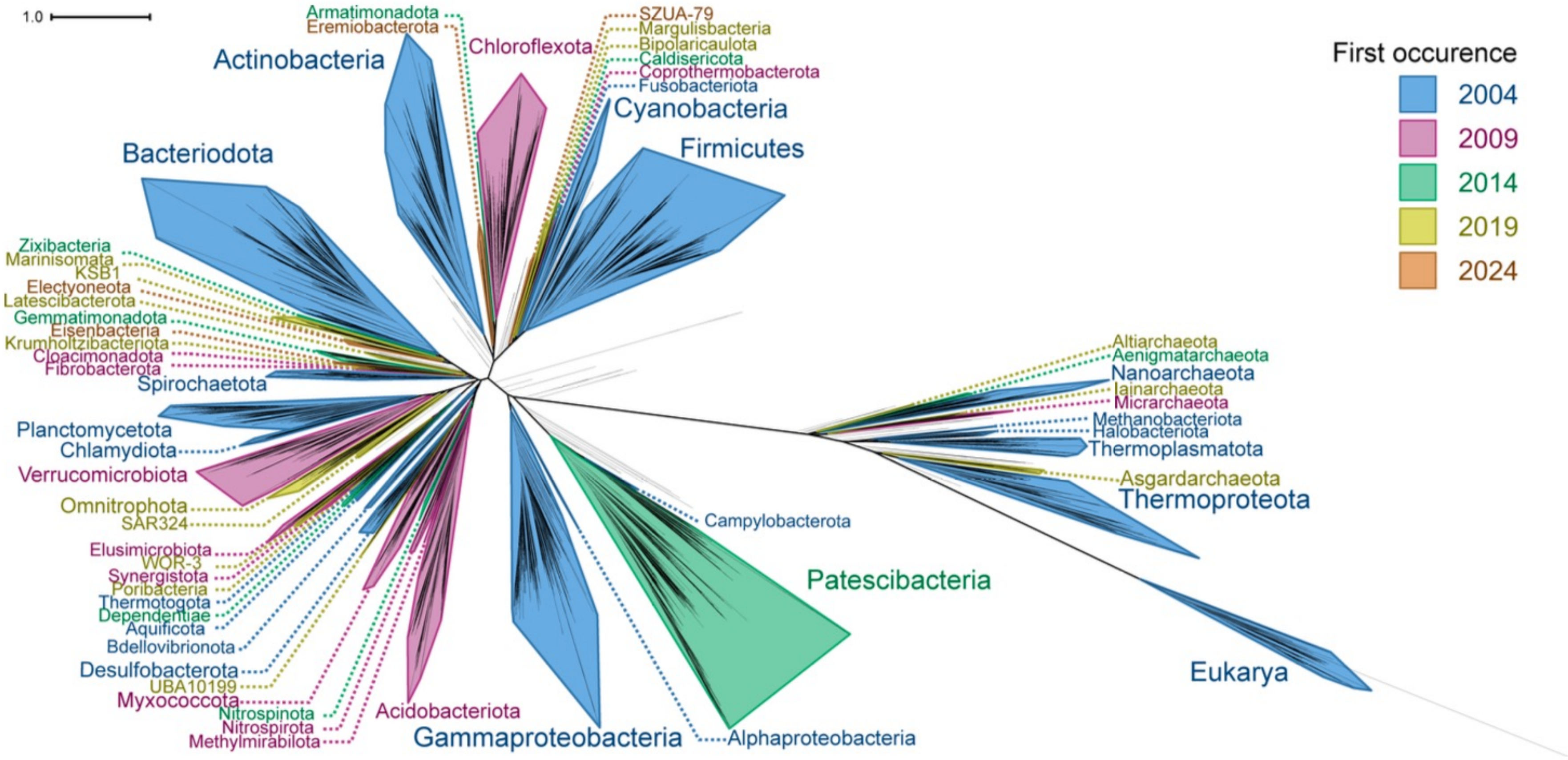


Figure 4. Reconstructed universal tree of life using genome sequencing data from 1999 to 2024**.** The tree was constructed using available sequences in the GenBank sequence database as of August 2024. A total of 69 496 organisms were placed by uDance. Annotations are based on phylum-level taxonomy assigned by GTDB. Phyla with 35 or more representatives are shown, as well as selected phyla present in figure 3. Colours indicate the reconstructed tree that the phylum first appeared in, in chronological order.

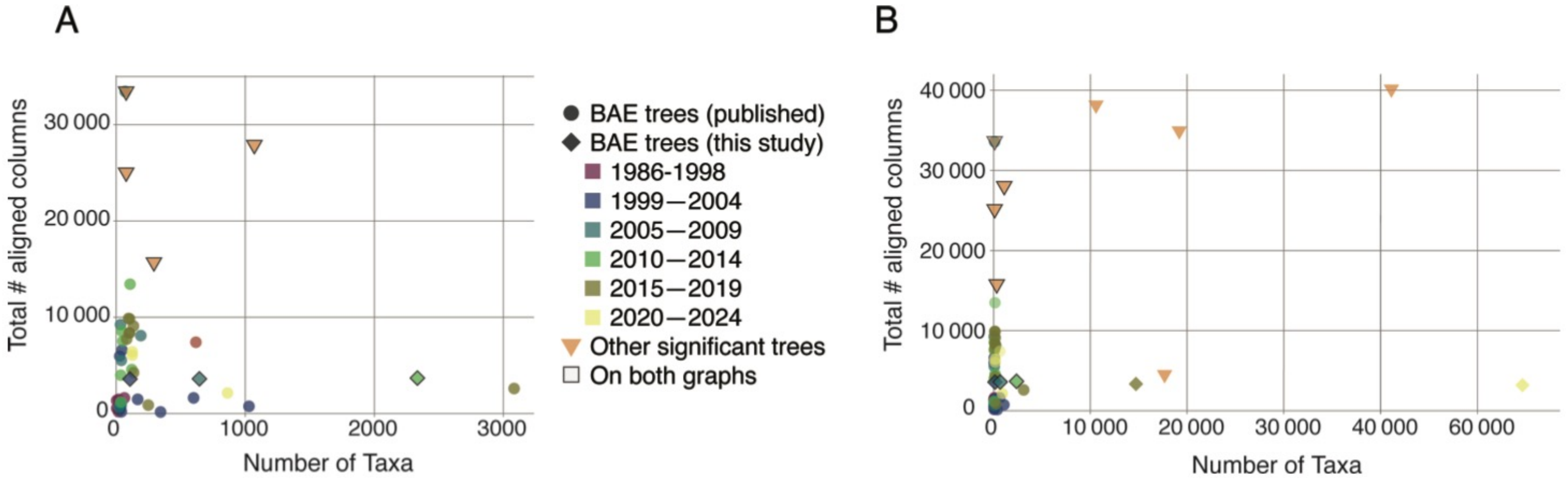
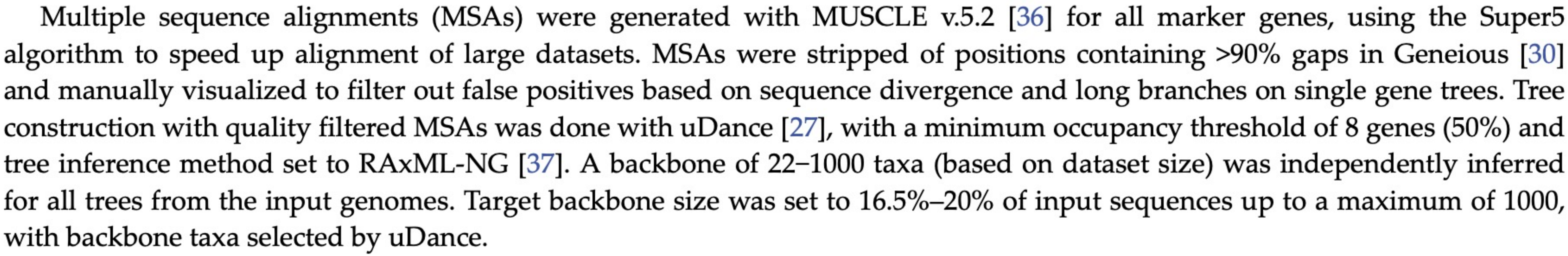


Figure 2. Taxon sampling versus molecular information used to construct universal trees of life. (A) All published universal trees and subsets of reconstructed trees from this study and from other significant two-domain trees. (B): Expanded view, including all reconstructed trees from this study (diamonds) and the full set of other significant two-domain trees included in figure 1 and electronic supplementary material, table S1. All universal trees are coloured by the 5-year window within which they were generated. Reconstructed (diamond) and two-domain (triangle) trees included on both plots are outlined to facilitate connection between the two *x*-axis scales.

You now know what some of this means!
#Learning



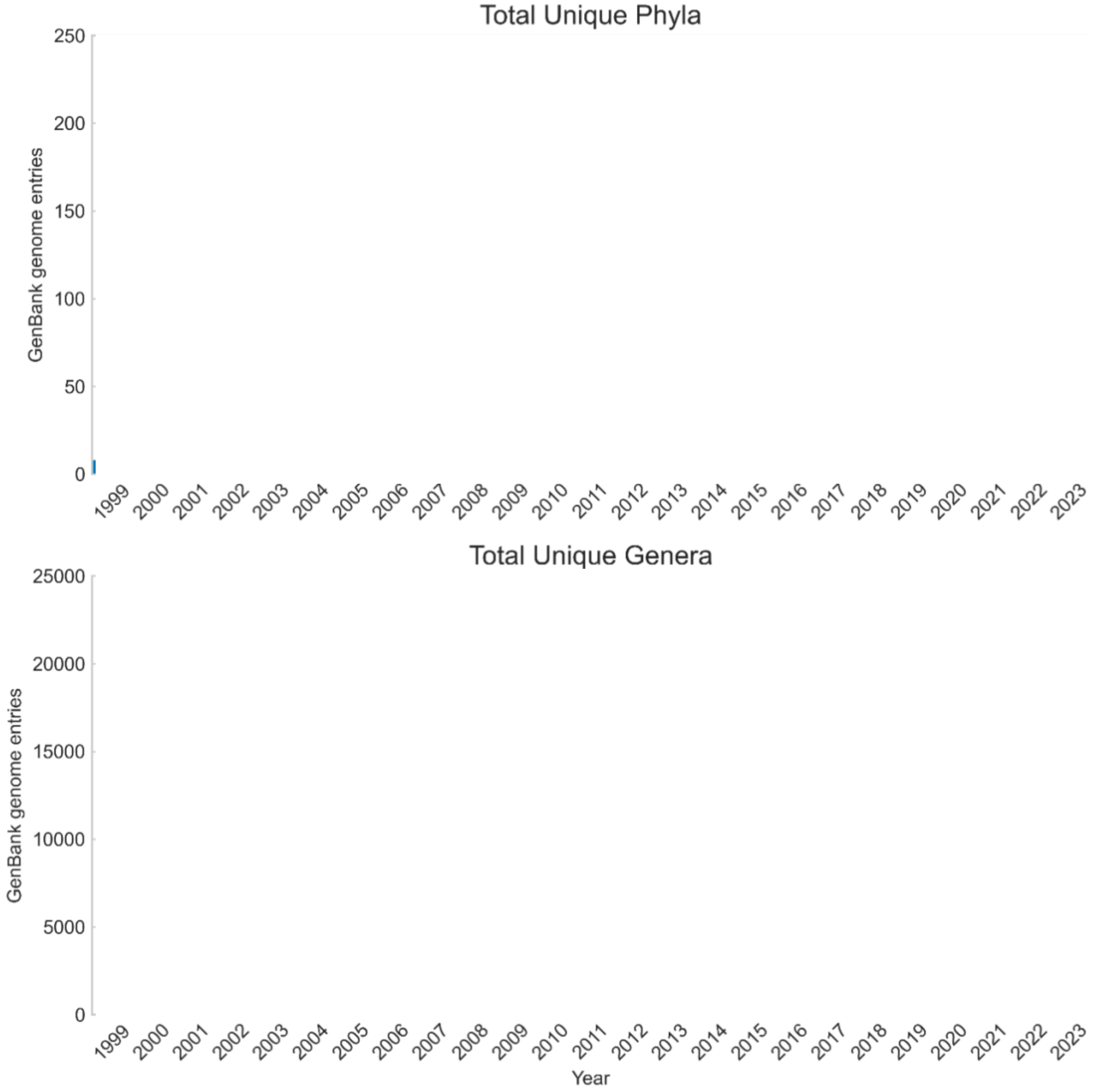
Multiple sequence alignments (MSAs) were generated with MUSCLE v.5.2 [36] for all marker genes, using the Super5 algorithm to speed up alignment of large datasets. MSAs were stripped of positions containing >90% gaps in Geneious [30] and manually visualized to filter out false positives based on sequence divergence and long branches on single gene trees. Tree construction with quality filtered MSAs was done with uDance [27], with a minimum occupancy threshold of 8 genes (50%) and tree inference method set to RAxML-NG [37]. A backbone of 22–1000 taxa (based on dataset size) was independently inferred for all trees from the input genomes. Target backbone size was set to 16.5%–20% of input sequences up to a maximum of 1000, with backbone taxa selected by uDance.

Table 2. Dissimilarity of common taxa between historically reconstructed trees. Pairwise comparisons were made between chronologically adjacent trees. For columns 3–7 ('Overlapping taxa...'), trees were pruned to the 71 (genus-level) taxa that are shared between all five trees. For columns 8 and 9 ('Pairwise overlapping taxa'), trees were pruned to the overlapping taxa that are shared between each successive pair. Only the normalized RF value rather than the Weighted value is reported as the number of common taxa differs in each comparison.

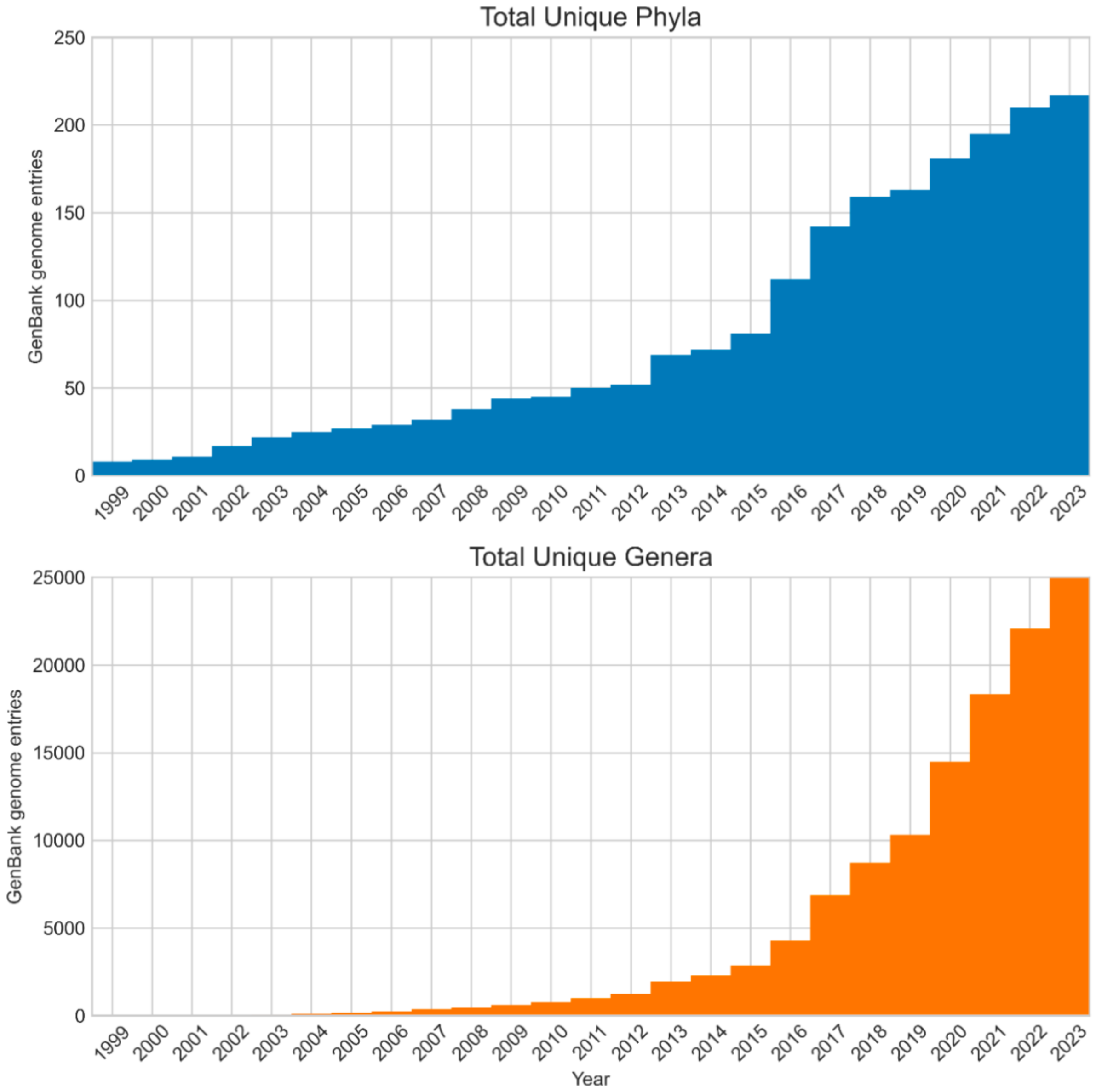
Comparison		Overlapping taxa between all trees (71 total)					Pairwise overlapping taxa	
tree1	tree2	RFWeighted (0.5)	RF (0.5)	MatchingTriplet	MatchingSplit	Normalized RF	Common taxa (genus level)	Normalized RF
2004	2009	6.60	25	17 039	140	0.37	105	0.29
2009	2014	7.83	22	15 665	154	0.33	612	0.31
2014	2019	9.35	35	15 294	153	0.52	1850	0.43
2019	2024	13.83	32	18 447	140	0.47	11 533	0.35

Supplementary Table 2: Summary of GenBank genome data availability. The initial genomes represent a snapshot of how many assembled genomes were available in the GenBank sequence database at the end of five-year intervals from 2004-2024. Genomes were filtered to remove any assembly that was not bacterial, archaeal, or eukaryotic (*e.g.,* metagenome assemblies, synthetic constructs, viruses). Datasets were also filtered and condensed to keep one representative per genus. Then, the NCBI datasets tool was used to download the protein (.faa) files for the filtered dataset, which further excluded genomes without an available protein file.

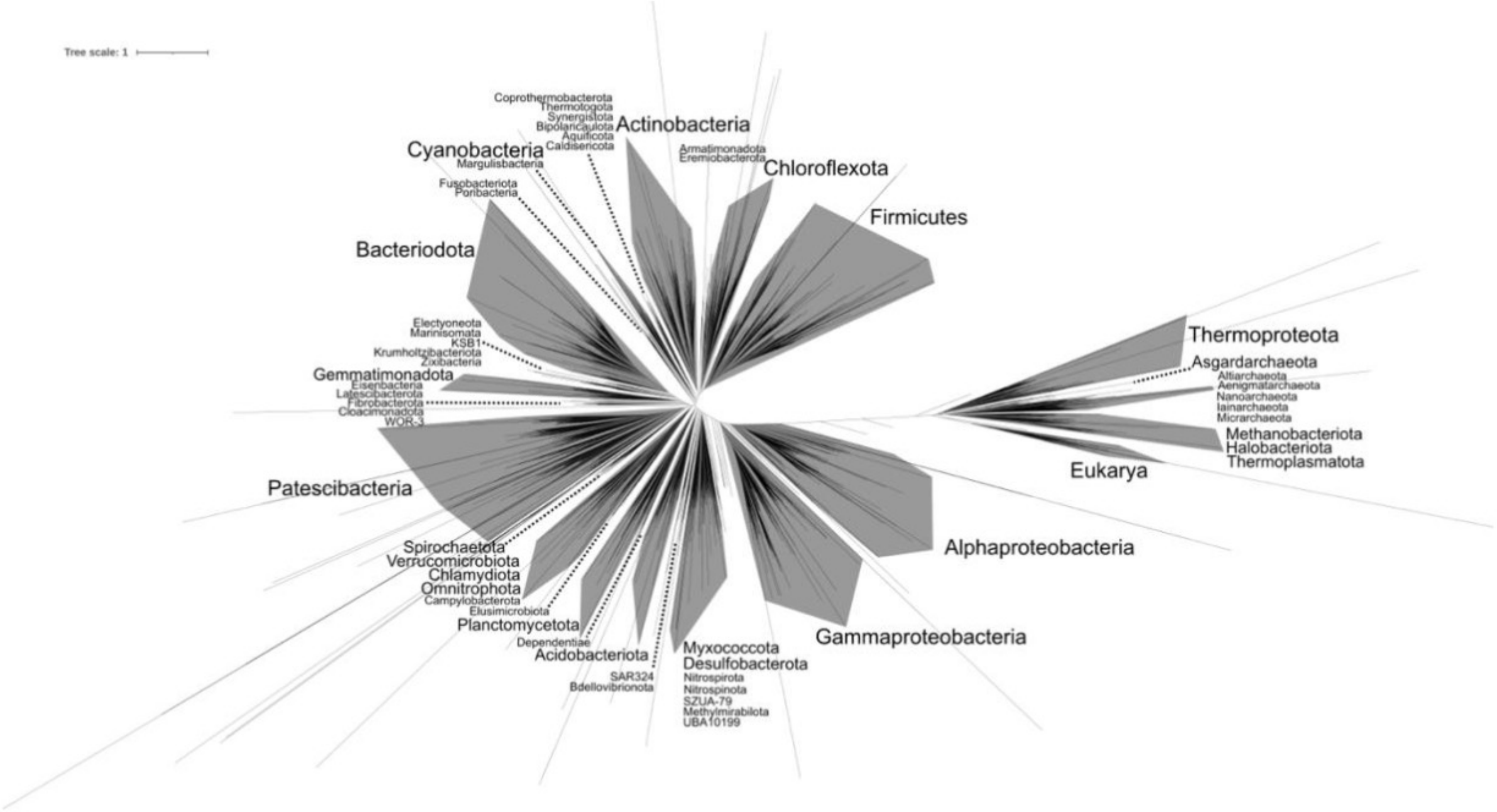
Release date range (inclusive)	Initial genomes	After filtering	Available protein files	Inclusion in uDance tree	Inclusion in FastTree tree
1999-2004	212	124	114	107	114
1999-2009	2,210	835	650	645	650
1999-2014	31,972	3,405	2,420	2,334	2,411
1999-2019	509,290	52,297	22,188	14,706	21,070
1999-2024	2,199,464	333,160	116,819	69,496	104,736



Supplementary Figure 1: Unique phylum and genus-level classifications represented in the GenBank genome database. Taxonomic classifications are based on Genome Taxonomy Database (GTDB) information [51]. Genomes which are mapped to a GTDB entry, or do not have an assigned GTDB phylum or genus, are not counted towards their respective datasets. Data from 1999-2023 are shown, as the 2024 update for GTDB has not yet been finalized.



Supplementary Figure 1: Unique phylum and genus-level classifications represented in the GenBank genome database. Taxonomic classifications are based on Genome Taxonomy Database (GTDB) information [51]. Genomes which are mapped to a GTDB entry, or do not have an assigned GTDB phylum or genus, are not counted towards their respective datasets. Data from 1999-2023 are shown, as the 2024 update for GTDB has not yet been finalized.



Supplementary Figure 3: FastTree reconstructed universal tree of life based on historical genomic sequencing data from 1999-2024. The same data used in Figure 4 is shown, but with FastTree instead of uDance. All GTDB clades present in Figure 4 are also annotated in Supplementary Figure 3. Annotations are approximated due to poor quality placement of many genomes.

Balaban et al.
2024

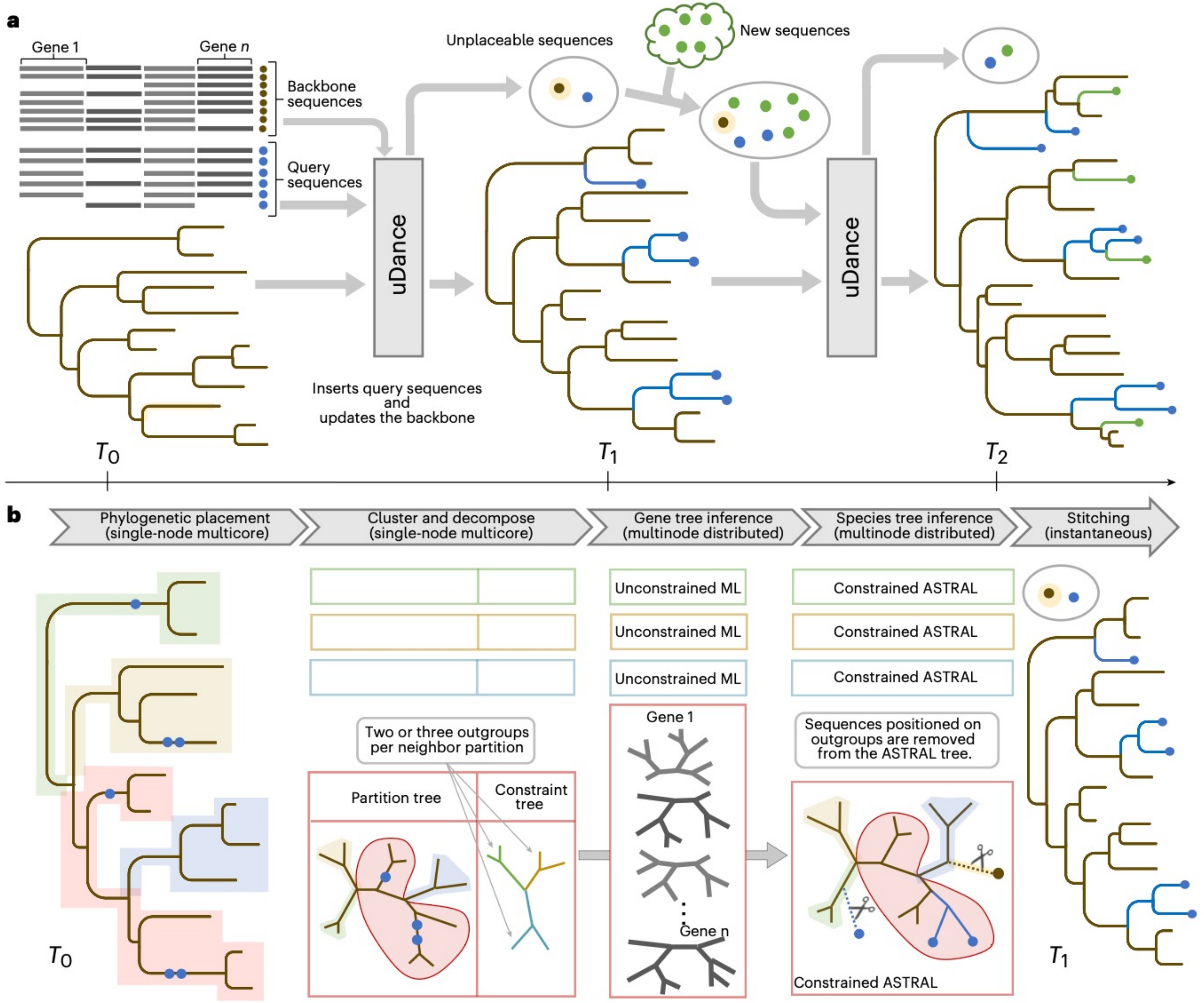
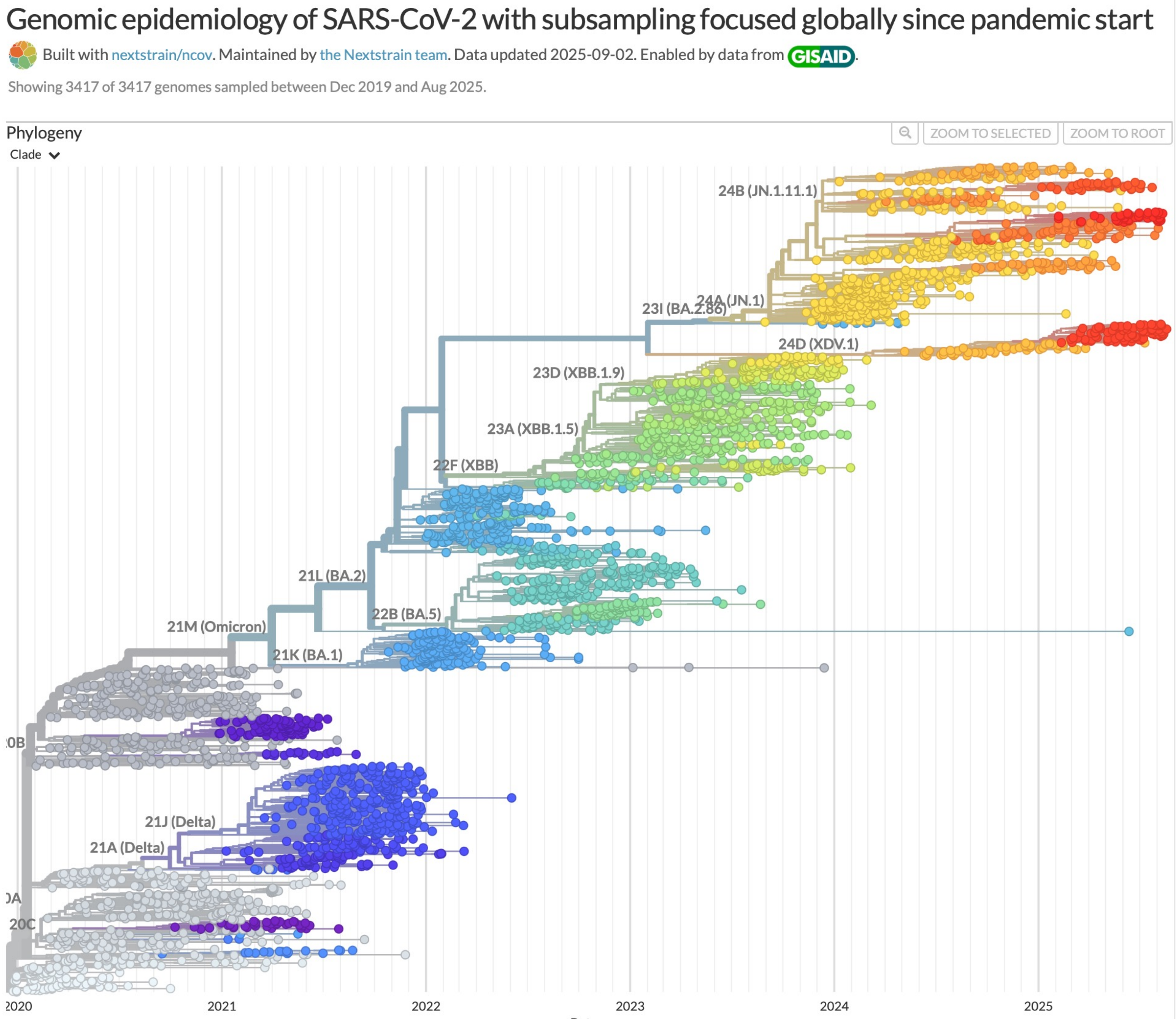


Fig. 1 | The uDance can continuously update trees using divide-and-conquer. **a**, Updates to the phylogenetic tree through time (T_0, T_1 and T_2) with new sequences arriving and each tree used as the backbone tree in the next step. Some sequences may be unplaceable and are added as a query to the next iteration. **b**, Each update involves divide-and-conquer and several steps, most of which can be executed in a distributed fashion. The new sequences are independently added to the tree using phylogenetic placement. Then, the tree with placements is divided into partitions, and each partition plus enough outgroups from other partitions are reanalyzed to infer local gene trees and a local species tree. These species trees are next joined together using a constrained search that makes it possible to stitch back together the subtrees.

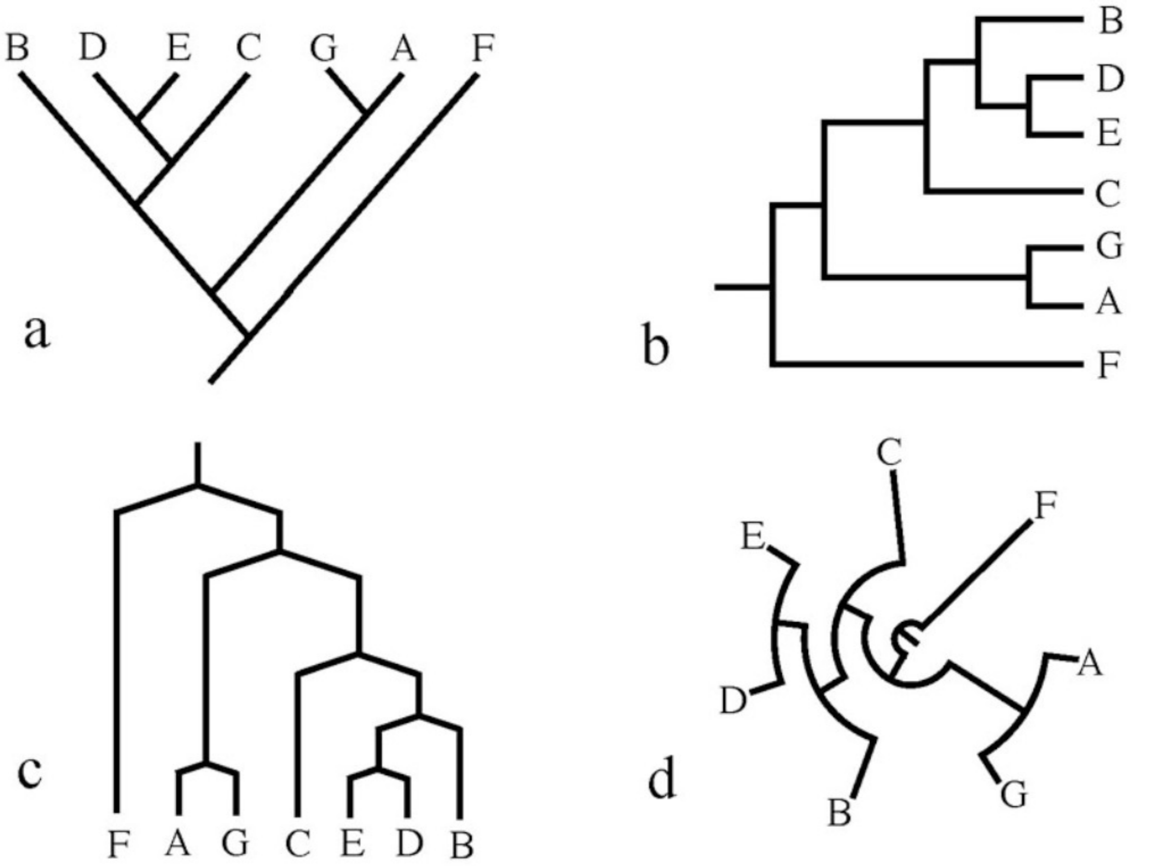


Genomic epidemiology of SARS-CoV-2 with subsampling focused globally since pandemic start

Built with nextstrain/ncov. Maintained by the Nextstrain team. Data updated 2025-09-02. Enabled by data from GISAID.

Showing 3417 of 3417 genomes sampled between Dec 2019 and Aug 2025.

https://nextstrain.org/ncov/gisaid/global/all-time



8) Which of the four trees above depicts a different pattern of relationships than the others?

Baum & Smith 2005

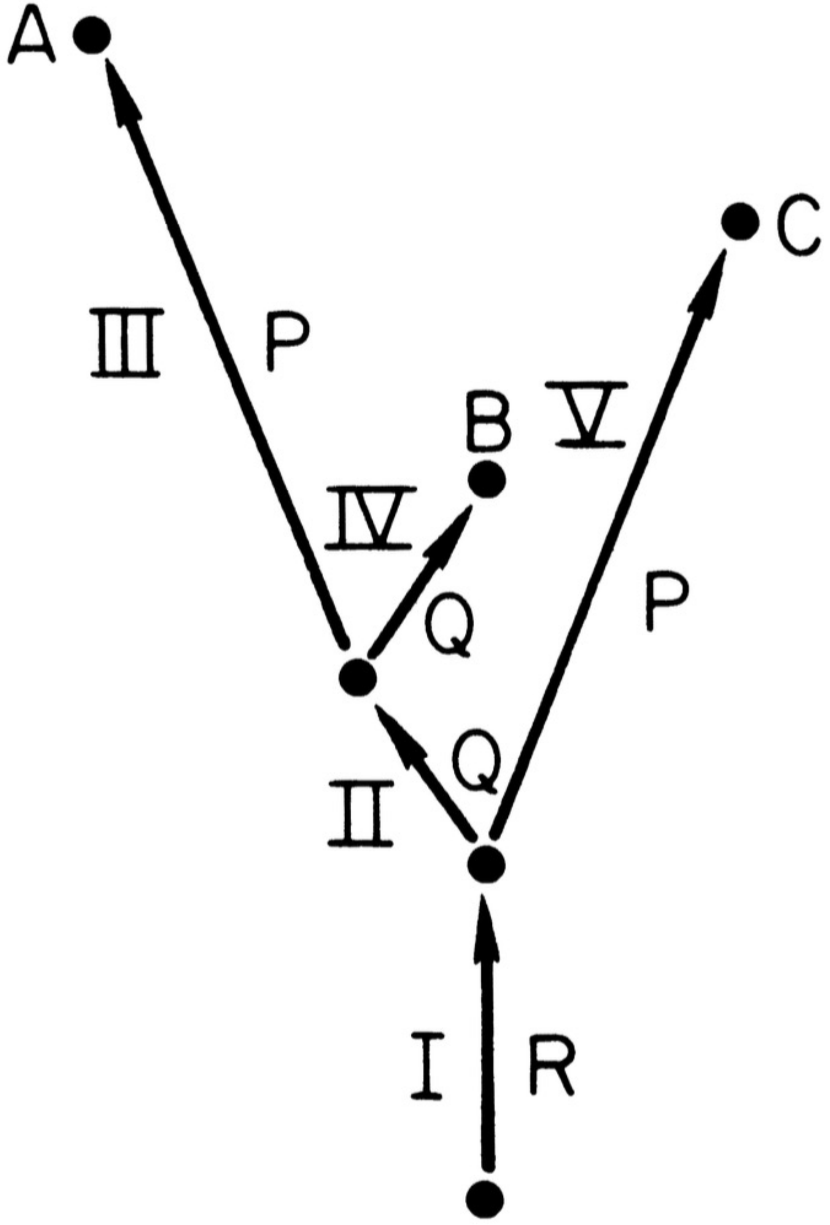


FIG. 1.—An evolutionary tree with three tip species. The segments of the tree are numbered I through V, and next to each is shown the probability of change from state 0 to state 1 in the segment.

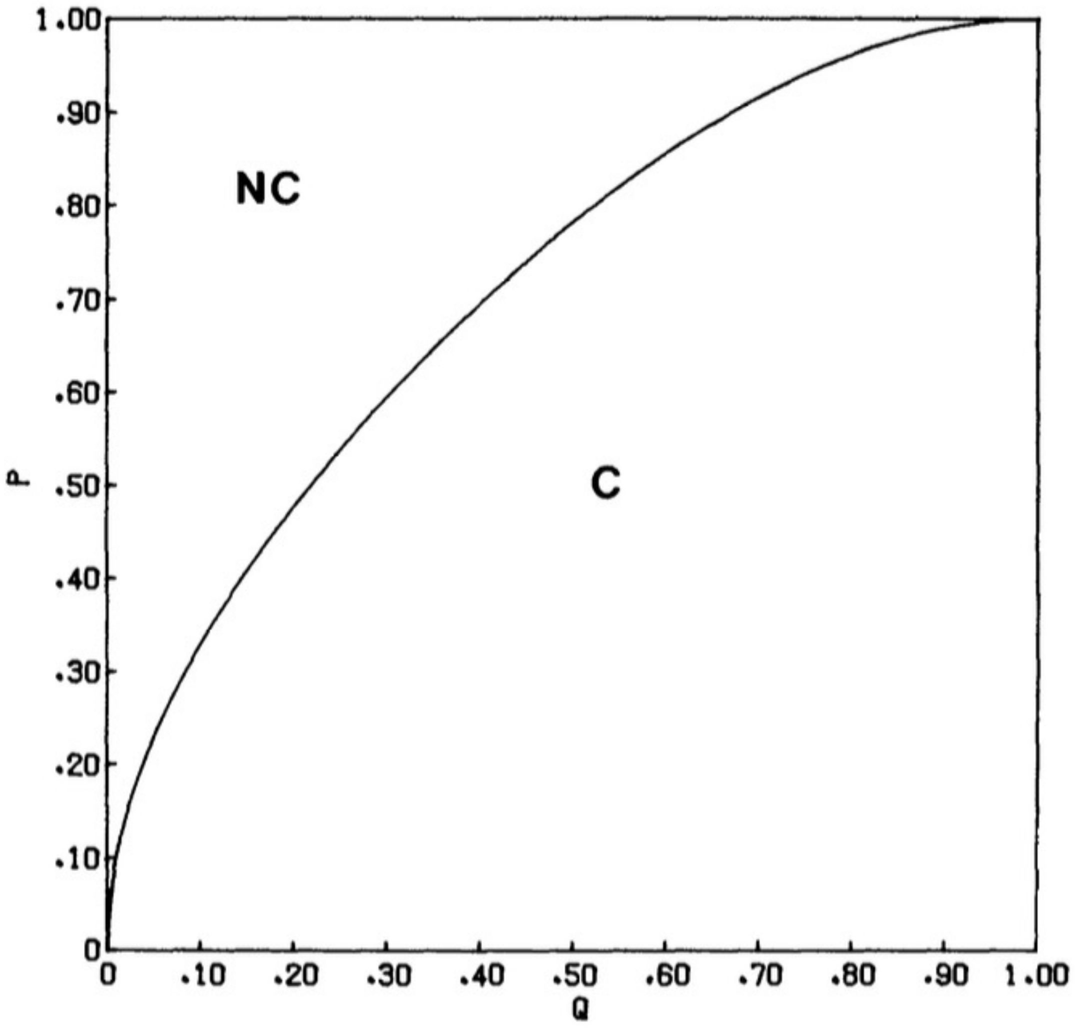


FIG. 2.—Values of P and Q for which the Camin-Sokal method fails to be consistent in the present case. C denotes the region of consistency, NC the region of inconsistency. Their boundary is the curve relating P_1 to Q.

Felsenstein 1978

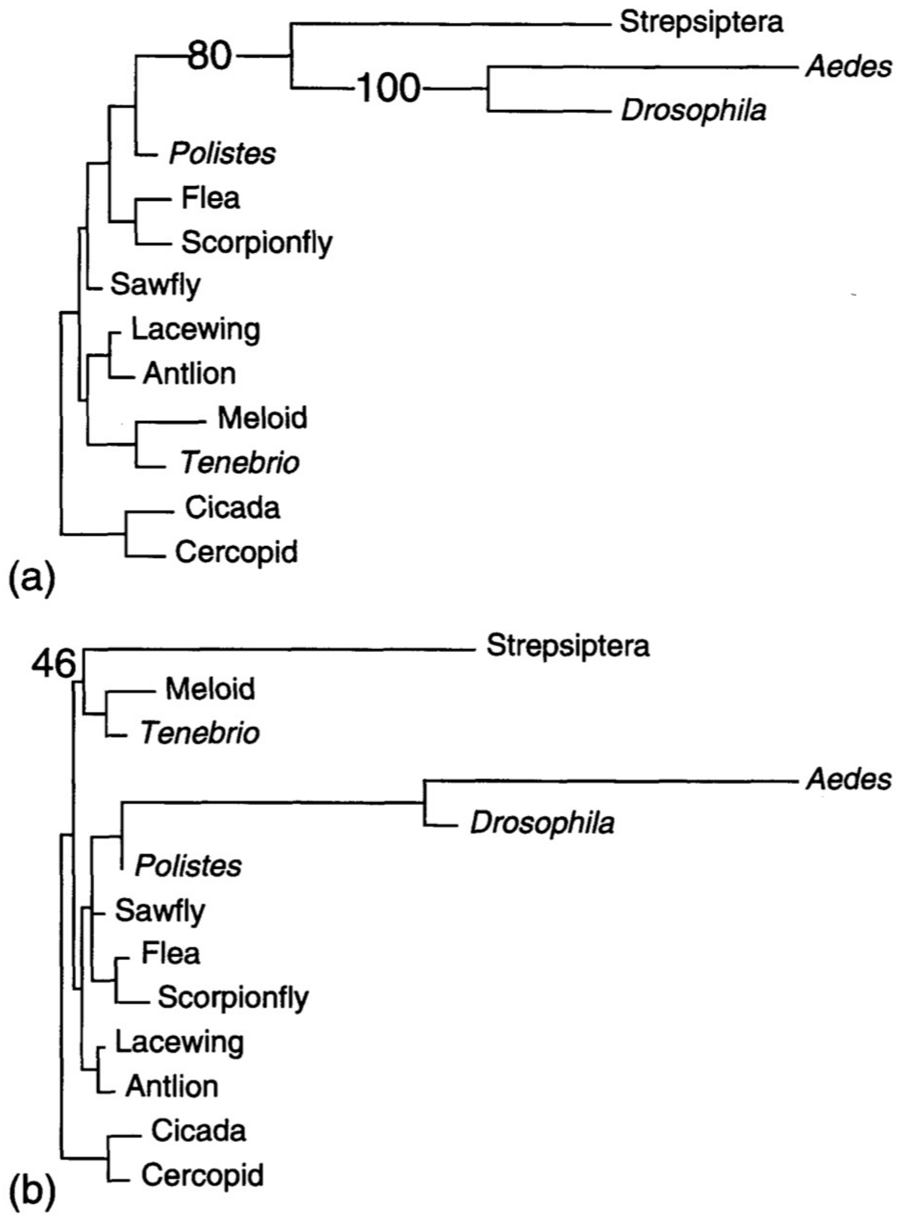


FIGURE 1. Phylogenetic trees for 13 insect groups. Numbers on branches are bootstrap values. (a) One of the trees estimated using the maximum parsimony method with Fitch (1971) optimization. Maximum parsimony estimated 27 trees; each was 364 steps in length. (b) Tree estimated using the maximum likelihood method implemented with the HKY85+Γ_5 model of DNA substitution (Hasegawa et al., 1985; Yang, 1993). The log likelihood of this tree is −2822.86, and the maximum likelihood estimates of κ and α are 3.60 and 0.29, respectively.

Huelsenbeck 1997

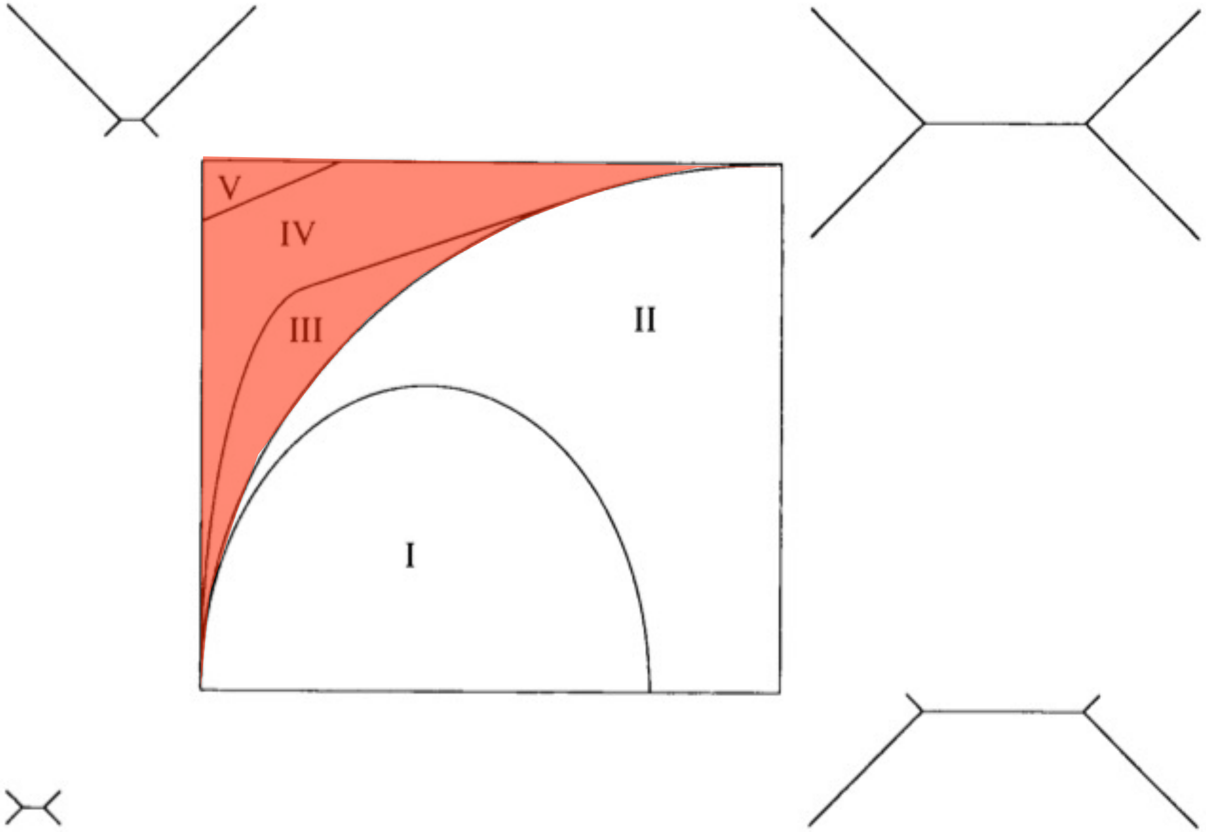


FIGURE 8. Regions of the graph space in which different phylogenetic methods perform best. The x-axis represents the three-branch length and the y-axis represents the two-branch length.

Huelsenbeck and Hillis, 1993