

# An Improved Method For Constructing Mega Phylogenies

Nomi Hadar & Itay Mayrose

Department of Molecular Biology and Ecology of Plants, Tel Aviv University, Tel Aviv Israel

1

Background: Mega Phylogenies

- Reconstructing the "Tree of Life" is one of the enduring goals of evolutionary biology.
- The massive accumulation of sequence data should provide more accurate phylogenies with the hope to resolve the tree of life.
- Recently, the *tree of birds*, *tree of plants* and *tree of reptiles* were presented.
- But, the reconstructed trees can only be as good as their underlying multiple sequence alignment (MSA).

2

Current Approaches For Reconstructing Mega Phylogenies

**Supertree:** compile many individual trees with partially overlapping taxa into a single large phylogeny.

**Supermatrix:** gather sequence data for many loci into a single multiple sequence alignment, which is then used to reconstruct a large phylogeny.

3

Do mega Phylogenies Really Provide Accurate Description Of The True Evolutionary History?

5

The Problem

Aligning rapidly evolving loci for highly diverged taxa leads to poor alignments and inaccurate trees.

**Slowly evolving regions:** Easy to align - good for solving ancient divergence events, cannot be used to solve recent events.

**Rapidly evolving regions:** Hard to align - good for solving recent event, cannot be used to solve ancient divergence events.

4

We Compare Mega Phylogenies to Current Taxonomical Knowledge

If the tree is correct, we expect the species that are part of the same hierarchy family (or other taxonomic hierarchy) to form a monophyletic group in the tree.

Three measures were used to asses monophyly.

E.g.,  
Family = {●●●●●}

Strict = 0

Is family monophyletic?

Relaxed 1 = 4/6

# of species in family

# of species under the last common ancestor

Relaxed 2 = 3/4

# of species in largest monophyletic subgroup

# of species in family

6

A New Approach For Reconstructing mega Phylogenies

7

Divide The Alignment Space To Well-Aligned Regions

	Locus 1 ->	Locus 2 ->	Locus 3 ->	Locus 1	Locus 2a	Locus 2b	Locus 3a	Locus 3b	Locus 3c
Seq1:	ATGGCATATCCCATACAACCTAGGATTCCAAGT	TATCCCGGG ---AAGCTAATT -AATGGAGGAATTTCAAGTATATT	CATTGATGTTTCCTTAATCC - AGATCGGAGGACGTGTTTC	ATGGCATATCCCATACAACCTAGGATTCCAAGT	TATCCCGGGAGGCTAATTAATGGAGGAATTTCAAGTATATT	TATCCCGGGAGGCTAATTAATGGAGGAATTTCAAGTATATT	CATTGATGTTTCCTTAATCCAGATCGGAGGACGTGTTTC	TCGATGACGAAAGCGAA-GTCCTC	TCGATGACGAAAGCGAA-GTCCTC
Seq2:	ATGGCTTACCCATTTCACCTTGGCTTACAAGA	TAT ---GGGCA -----ATGGA --AATTTCAA --TAT	-----TCGATG-----ACGAAAGCGAA--GTCCTC	ATGGCTTACCCATTTCACCTTGGCTTACAAGA	ATGGCTTACCCATTTCACCTTGGCTTACAAGA	ATGGCTTACCCATTTCACCTTGGCTTACAAGA	-----TCGATG-----ACGAAAGCGAA--GTCCTC	TCGATGACGAAAGCGAA-GTCCTC	TCGATGACGAAAGCGAA-GTCCTC
Seq3:	ATGGCAACCACTCCCACTAGGCTTTCAAGT	TATCCCTGG ---AAGGTAA--GAATGGAGGAATTTCAAGTATATT	CATTG ---CGGATTAA-----GTCCTC	ATGGCAACCACTCCCACTAGGCTTTCAAGT	TATCCCTGGAAAGCTAATCAATGGAGGAATTTCAAGTATATT	TATCCCTGGAAAGCTAATCAATGGAGGAATTTCAAGTATATT	-----TCGATG-----ACGAAAGCGAA--GTCCTC	TCGATGACGAAAGCGAA-GTCCTC	TCGATGACGAAAGCGAA-GTCCTC
Seq4:	ATGGCACATCCCAACAATTAGGATTTCCAAGA	TAA ---GGGCA -----ATGGA --AATTGCAA --TAT	-----TCGATG-----AGGATCGCGAA--GTCCTC	ATGGCACATCCCAACAATTAGGATTTCCAAGA	ATGGCACATCCCAACAATTAGGATTTCCAAGA	ATGGCACATCCCAACAATTAGGATTTCCAAGA	-----TCGATG-----AGGATCGCGAA--GTCCTC	TCGATGACGAAAGCGAA-GTCCTC	TCGATGACGAAAGCGAA-GTCCTC
Seq5:	ATGGCCTACCCATTCCAACCTGGTCTACAAGA	TAT ---GCGCA -----ATGCA --AATTTCAA --TAT	CATTG ---CCTACATCC-----CGGCT- CACCAC	ATGGCCTACCCATTCCAACCTGGTCTACAAGA	ATGGCCTACCCATTCCAACCTGGTCTACAAGA	ATGGCCTACCCATTCCAACCTGGTCTACAAGA	-----TCGATG-----AGGATCGCGAA--GTCCTC	TCGATGACGAAAGCGAA-GTCCTC	TCGATGACGAAAGCGAA-GTCCTC

8

Using the resulted alignment to reconstruct the tree

