

# 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).

## 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.

## 6 A New Approach For Reconstructing mega Phylogenies

Aligned	well	Locus 1 ->
badly	o.k.	Seq1: ATGGCATATCCATACACTAGGATTCCAAGT Seq2: ATGGCTAACCATTTCAACTTGCTTCAAGA Seq3: ATGGCAAACACTCCAACTAGGCTTCAGT Seq4: ATGGCACATCCCACACAAATTAGGATTCCAAGA Seq5: ATGGCTAACATTCCAACCTTGCTTCAAGA
Aligned	o.k.	Locus 2 ->
badly	o.k.	Seq1: TATCCCGGGG--AAAGCTTAATT-AATGGAGGAATTCAAGTATATT Seq2: TAT---GGCA-----ATGGA---AATTCA---TAT Seq3: TATCCCTGG--AAAGCTTA-CATGGAGGAATTTCAGTAT Seq4: TAA---GGCA-----ATGGA---AATTCA---TAT Seq5: TAT---GGCCTA-----ATGCA---AATTCA---TAT
Aligned	badly	Locus 3 ->
badly	badly	Seq1: CATTGATTTCTTAACTC-AGAAGCGGA-GTCTC Seq2: TCGATG-----CCCTATTA-CGC-C-GTTCA Seq3: CATTG-----CCCTATTA-CGC-C-GTTCA Seq4: TCGATG-----AGGATCCGGGA-GTCTC Seq5: CATTG-----CCTACCC

## 7 Divide The Alignment Space To Well-Aligned Regions

Locus 1	Locus 2a	Locus 2b	Locus 3a	Locus 3b	Locus 3c
ATGGCATATCCATACACTAGGATTCCAAGT ATGGCTAACCATTTCAACTTGCTTCAAGA ATGGCAAACACTCCAACTAGGCTTCAGT ATGGCACATCCCACACAAATTAGGATTCCAAGA ATGGCTAACATTCCAACCTTGCTTCAAGA	TATCCCGGGGAGCTTAATTAGGAGGAATTCAAGTATATT TATCCCTGGAAAGCTTAATCAATGGAGGAATTCAAGTAT TATGGCAATGGAAATTCAAT TAAGGGCAATGGAAATTCAAT TATGCGCAATGCAAATTCAAT	- - - - -	CATTGATTTCTTAACTCAGATGGAGGACGTGTTTC - - - - -	- - - - -	- - - - -
Seq1: ATGGCATATCCATACACTAGGATTCCAAGT Seq2: ATGGCTAACCATTTCAACTTGCTTCAAGA Seq3: ATGGCAAACACTCCAACTAGGCTTCAGT Seq4: ATGGCACATCCCACACAAATTAGGATTCCAAGA Seq5: ATGGCTAACATTCCAACCTTGCTTCAAGA	Seq1: TATCCCGGGG--AAAGCTTAATT-AATGGAGGAATTCAAGTATATT Seq2: TAT---GGCA-----ATGGA---AATTCA---TAT Seq3: TATCCCTGG--AAAGCTTA-CATGGAGGAATTTCAGTAT Seq4: TAA---GGCA-----ATGGA---AATTCA---TAT Seq5: TAT---GGCCTA-----ATGCA---AATTCA---TAT	Seq1: TATCCCGGGGAGCTTAATTAGGAGGAATTCAAGTATATT Seq2: TATCCCTGGAAAGCTTAATCAATGGAGGAATTCAAGTAT Seq3: TATGGCAATGGAAATTCAAT Seq4: TAAGGGCAATGGAAATTCAAT Seq5: TATGCGCAATGCAAATTCAAT	Seq1: CATTGATTTCTTAACTC-AGAAGCGGA-GTCTC Seq2: TCGATG-----CCCTATTA-CGC-C-GTTCA Seq3: CATTG-----CCCTATTA-CGC-C-GTTCA Seq4: TCGATG-----AGGATCCGGGA-GTCTC Seq5: CATTG-----CCTACCC	Seq1: CATTGATTTCTTAACTCAGATGGAGGACGTGTTTC Seq2: - Seq3: - Seq4: - Seq5: -	Seq1: - Seq2: - Seq3: - Seq4: - Seq5: -

## 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?

Strict = 0  
*Is family monophyletic?*

Relaxed 1 = 4/6  
 $\frac{\# \text{ of species in family}}{\# \text{ of species under the last common ancestor}}$

Relaxed 2 = 3/4  
 $\frac{\# \text{ of species in largest monophyletic subgroup}}{\# \text{ of species in family}}$

## 8 Using the resulted alignment to reconstruct the tree

This study is supported by the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University