



Aggregated Dendrograms for Visual Comparison Between Many Phylogenetic Trees



Zipeng Liu
Shing Hei Zhan
Tamara Munzner
( @tamaramunzner)

Computer Science
Biodiversity Centre
Computer Science

University of British Columbia

Phylogenetic tree

Evolutionary relationships of organisms

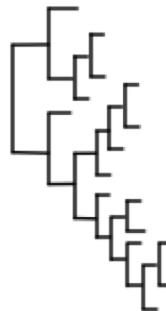
Human	A	T	G	G	A	C	A
Chimpanzee	A	T	G	G	A	C	A
Macaque	A	C	G	G	A	C	A

Genetic information

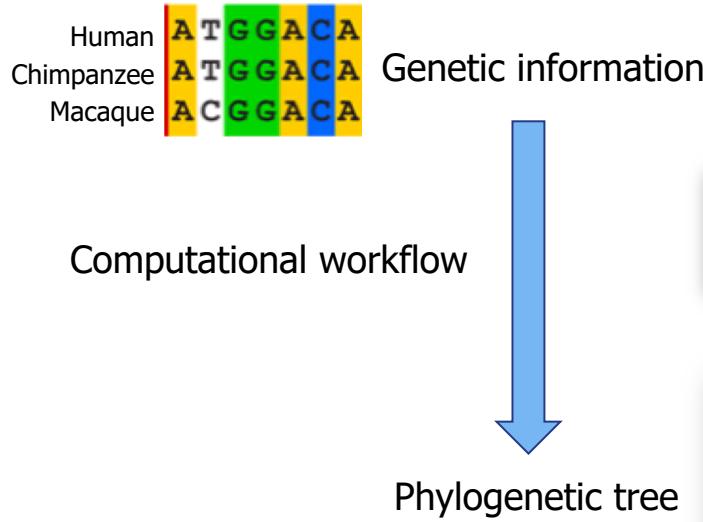
Computational workflow



Phylogenetic tree



Many phylogenetic trees



- Understand relationships between genes and species trees
- Explore trees generated with different methods and data

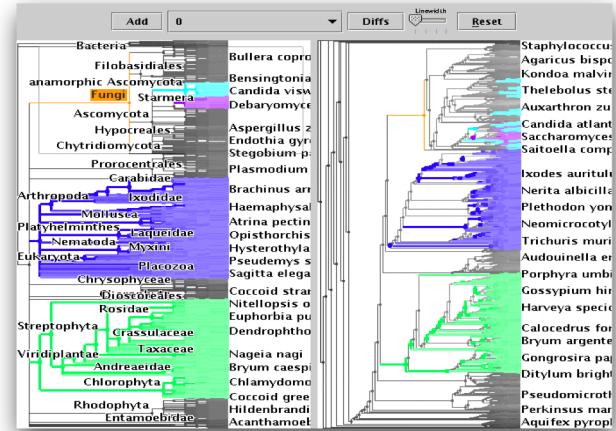
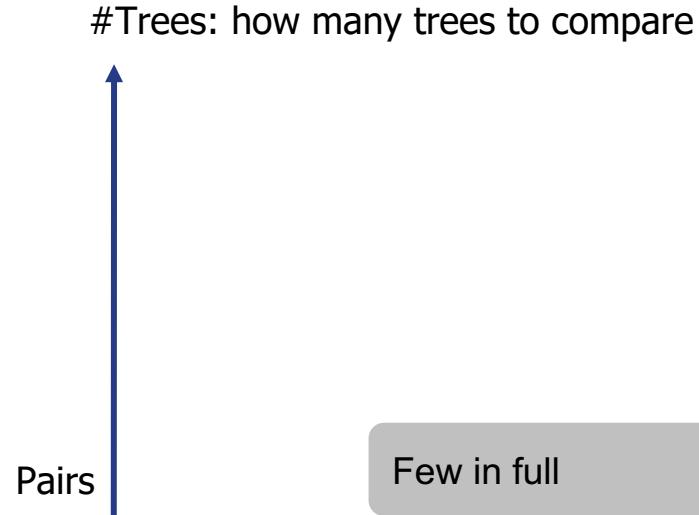
Scalability of Existing Tree Comparison Systems

#Trees: how many trees to compare



Level of detail (LoD):
how much details are visible

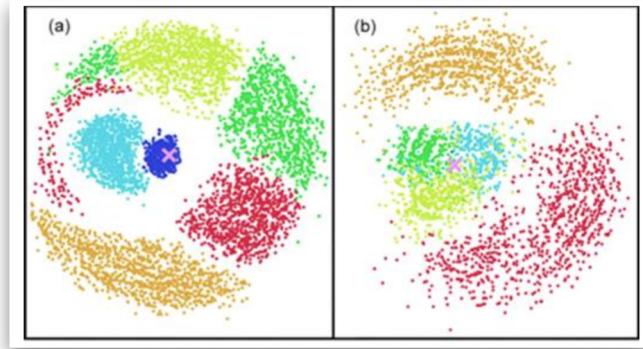
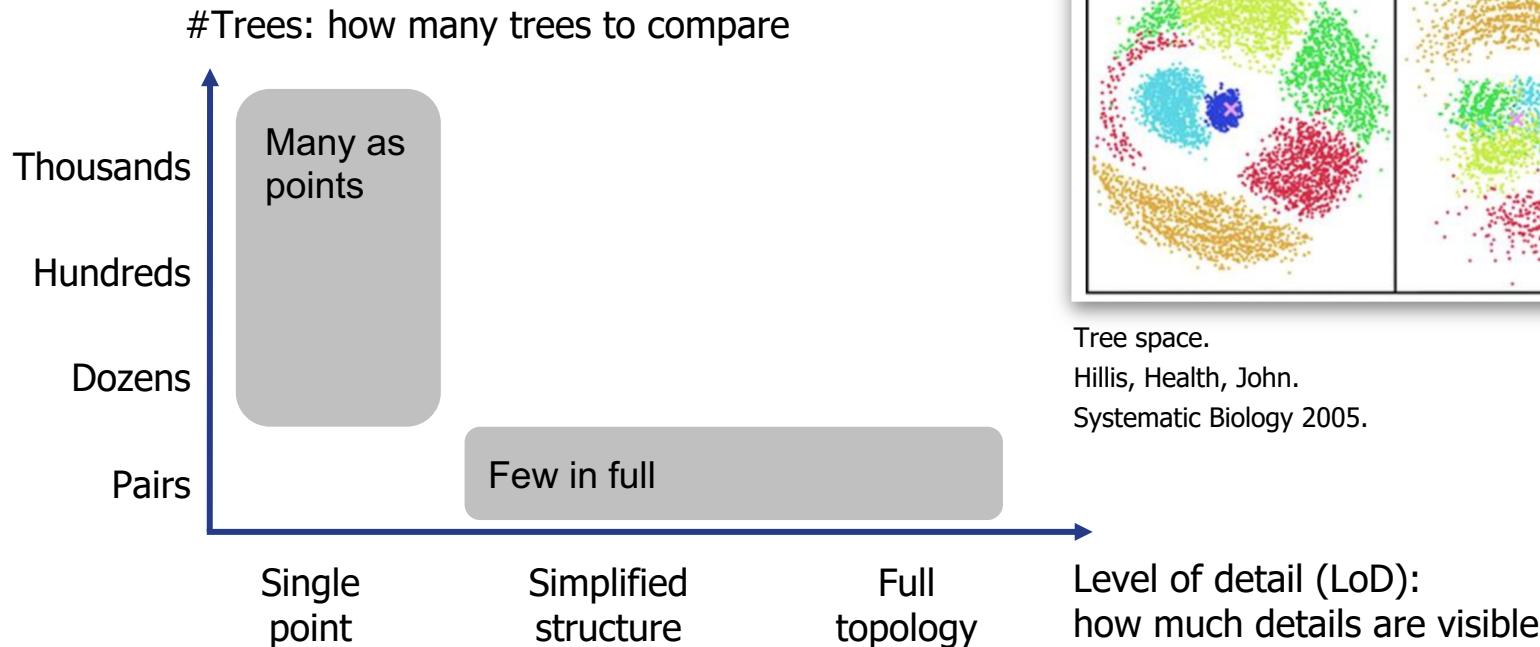
Scalability of Existing Tree Comparison Systems



TreeJuxtaposer.
Munzner, Guimbretière, Zhang, Zhou.
SIGGRAPH 2003

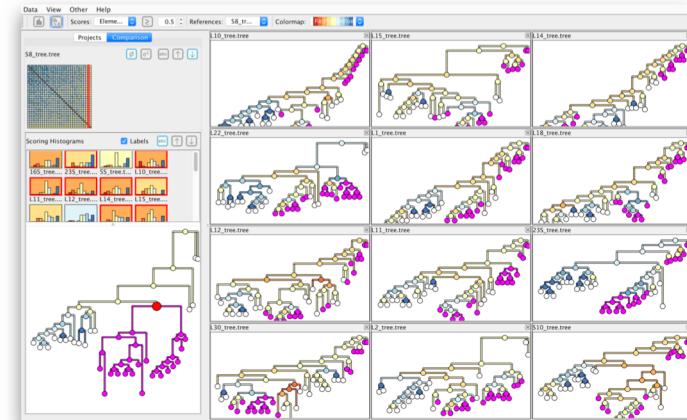
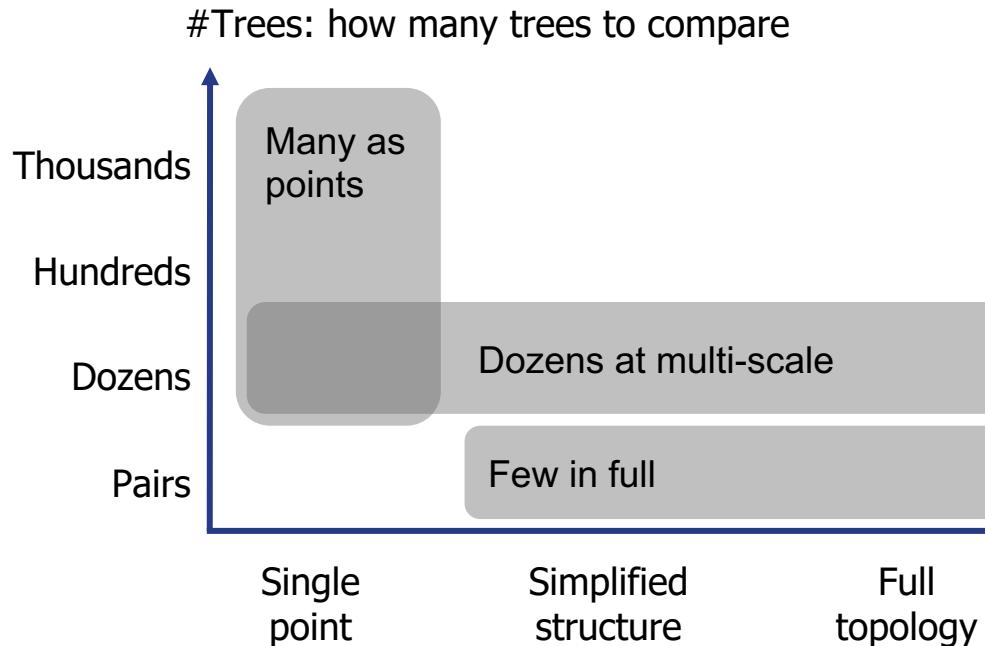
Level of detail (LoD):
how much details are visible

Scalability of Existing Tree Comparison Systems



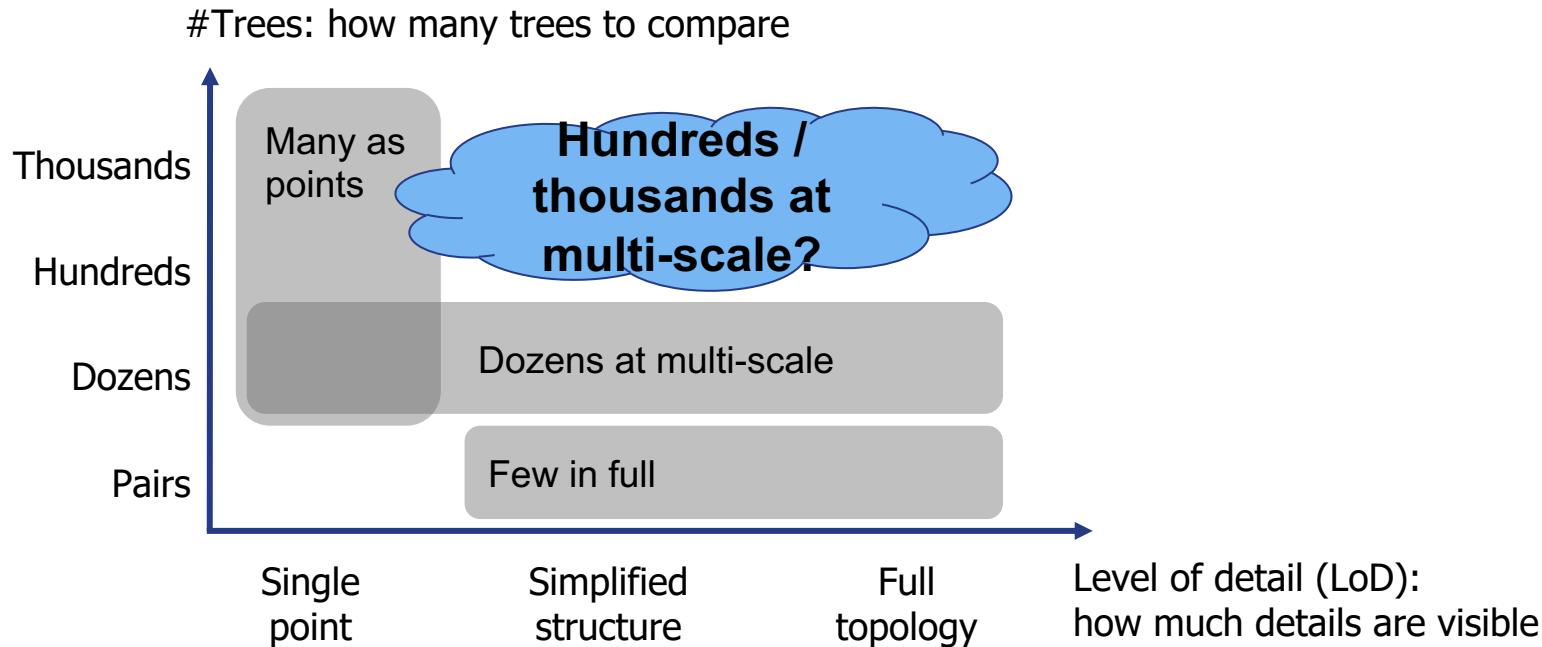
Tree space.
Hillis, Health, John.
Systematic Biology 2005.

Scalability of Existing Tree Comparison Systems



Interactive visual comparison of multiple trees.
Bremm, Landesberger, Heß, Schreck, Weil, Hamacher.
VAST 2011.

Comparing many phylogenetic trees

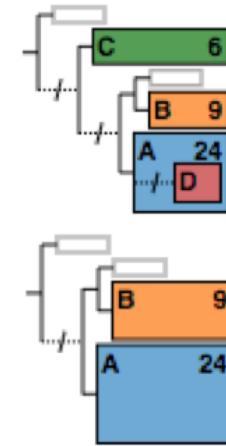


Main contributions

- Data and task abstractions for comparison of phylogenetic trees

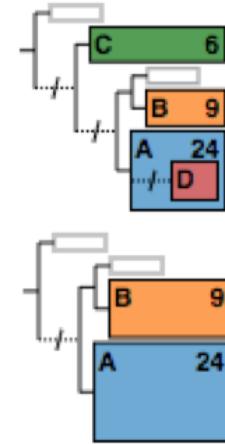
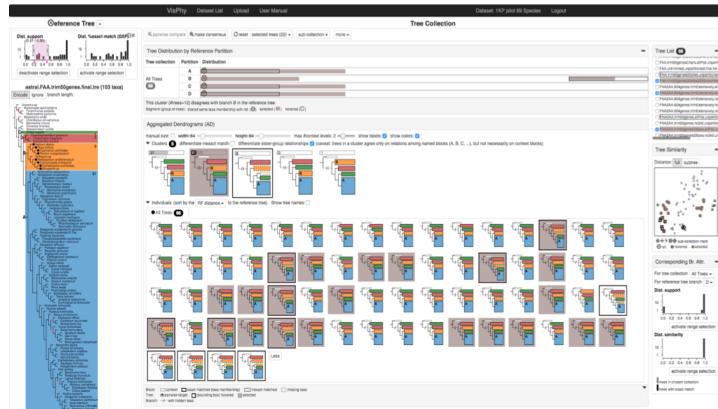
Main contributions

- Data and task abstractions for comparison of phylogenetic trees
- A new visual encoding: **Aggregated Dendrogram**
 - Compact tree representation that focuses on selected subtrees
 - Adapts to available screen space



Main contributions

- Data and task abstractions for comparison of phylogenetic trees
- A new visual encoding: **Aggregated Dendrogram**
 - Compact tree representation that focuses on selected subtrees
 - Adapts to available screen space
- A multi-view interactive tool: **ADView**
 - Covers multiple levels of details for tree comparison





Introduction

Data & Tasks

- Tree data
- Two crucial tasks
- Derived data

Aggregated Dendrogram (AD)

ADView

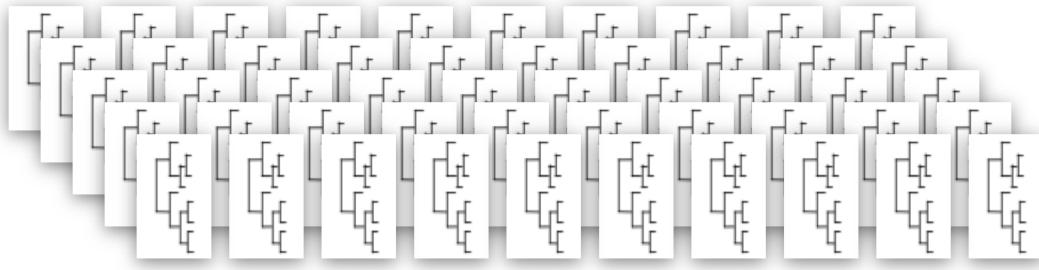
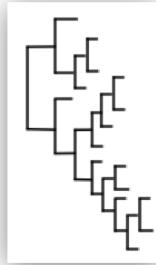
Summary

Tree data

Reference tree

vs.

Tree collection

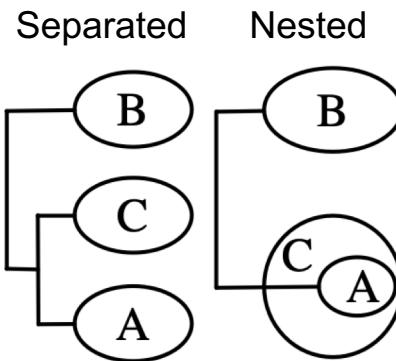


Two crucial tasks

Topological relationships between
subtrees / leaf nodes

Two crucial tasks

Topological relationships between
subtrees / leaf nodes

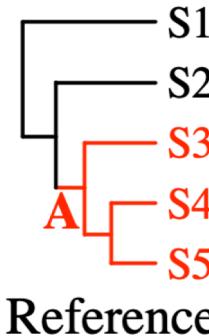
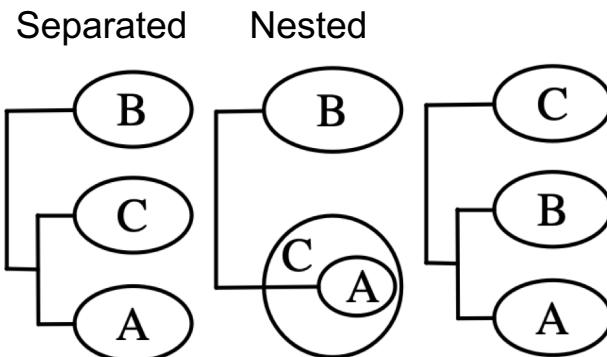


Two crucial tasks

Topological relationships between subtrees / leaf nodes

- Topological distance

Leaf node memberships compared to reference tree

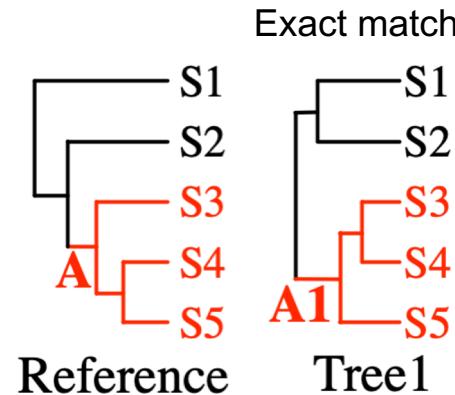
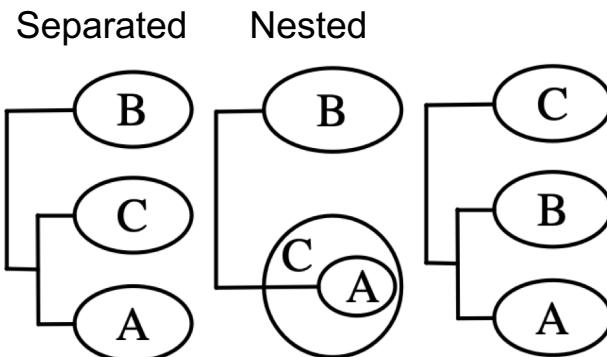


Two crucial tasks

Topological relationships between subtrees / leaf nodes

- Topological distance

Leaf node memberships compared to reference tree

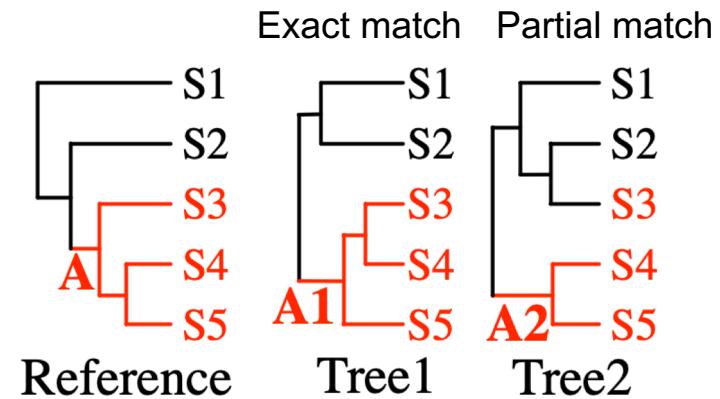
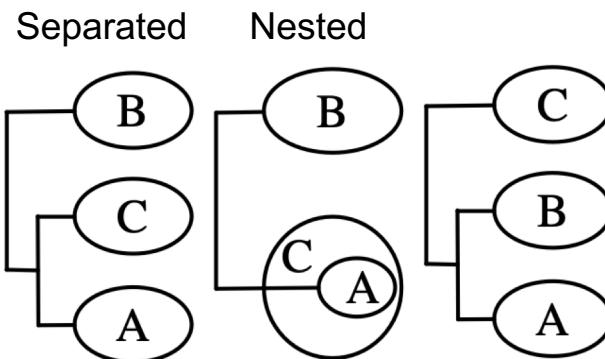


Two crucial tasks

Topological relationships between subtrees / leaf nodes

- Topological distance

Leaf node memberships compared to reference tree



Derived data

Topological relationships between subtrees / leaf nodes

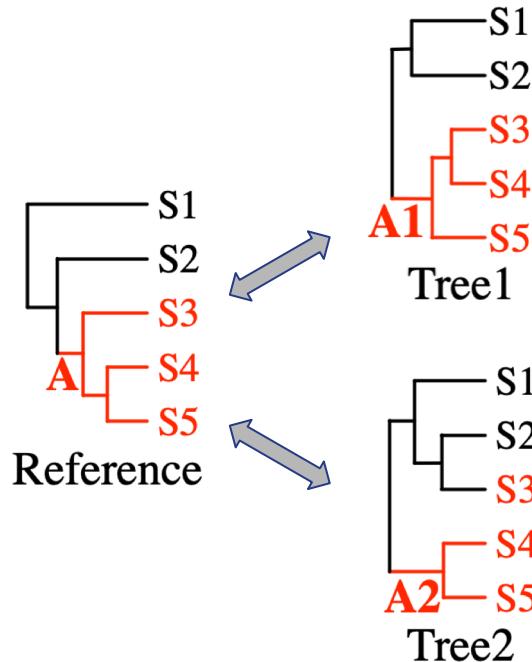
Leaf node memberships compared to reference tree

Finding correspondences between reference tree and other trees

Derive

corresponding branch / subtree

- The most similar subtree in one tree given a subtree in the reference tree





Introduction
Data & Tasks

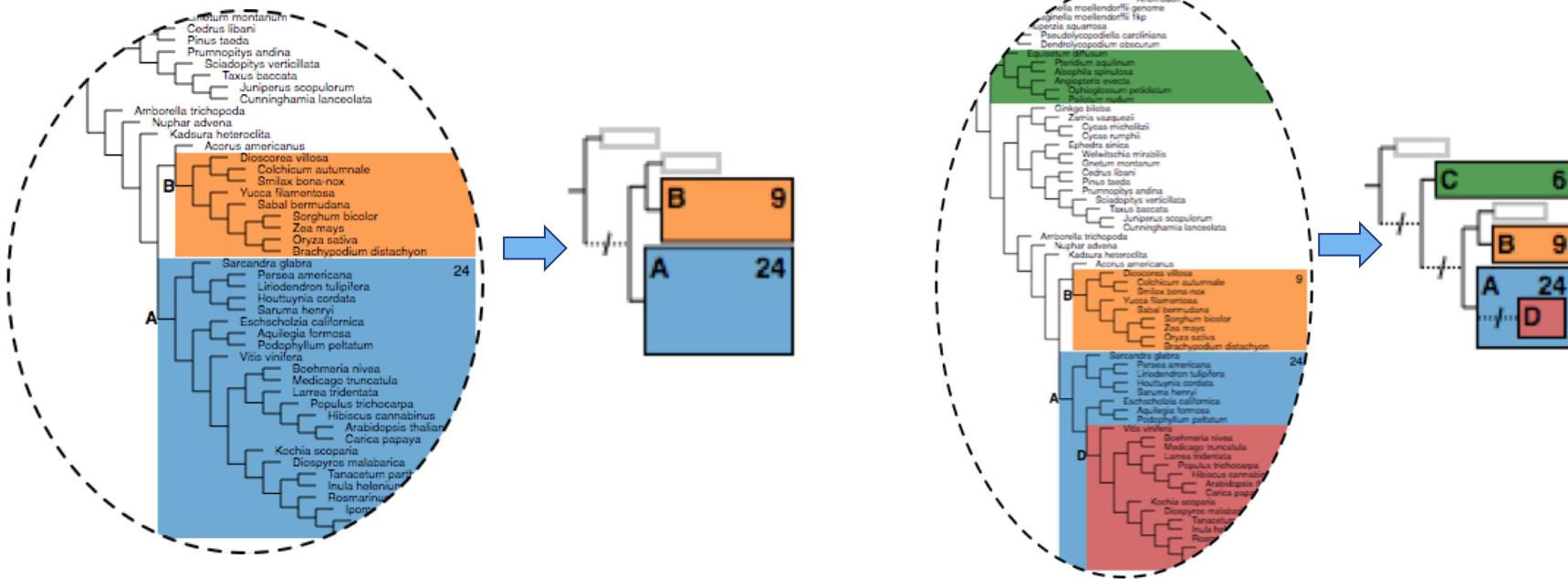
Aggregated Dendrogram (AD)

- Intuition
- Visual design
- Cluster AD

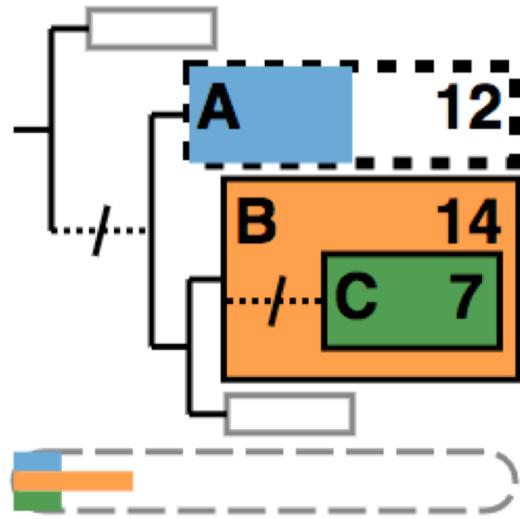
ADView
Summary

Intuition

Use glyphs to compress a tree according to user selections

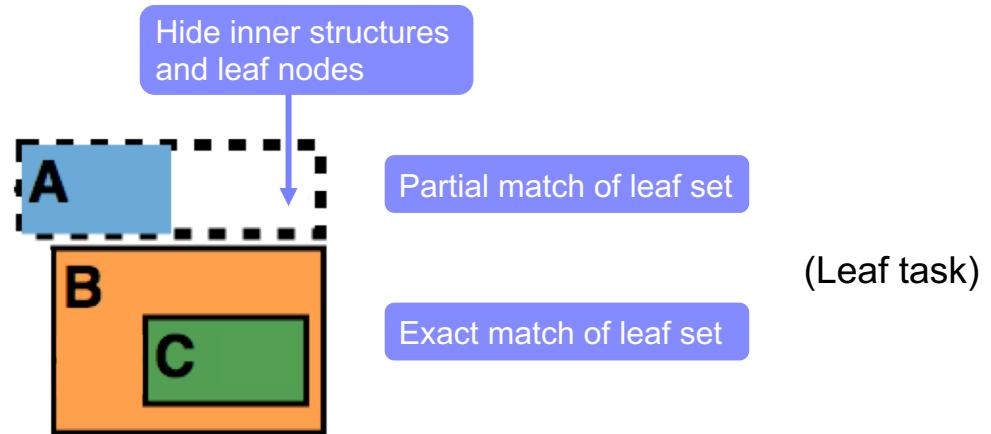


Visual design: focus + context



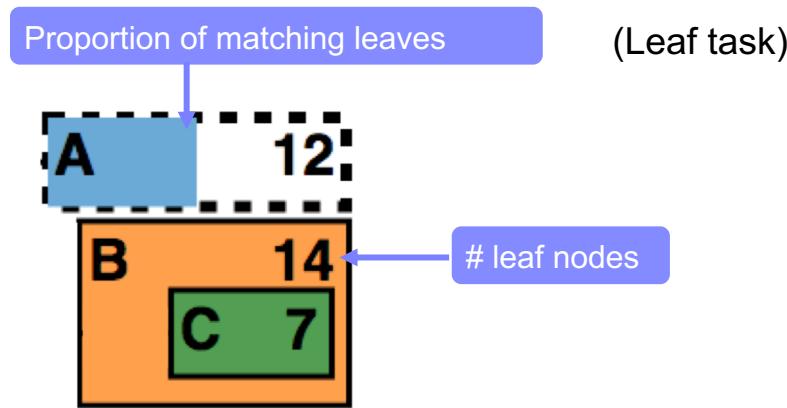
Visual design: focus + context

- Focus
 - Selected subtrees



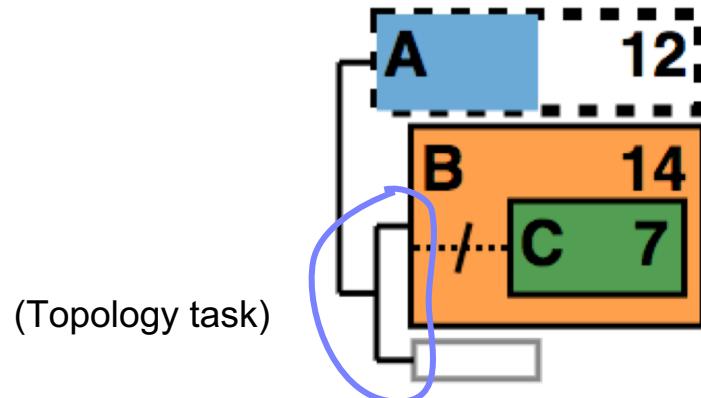
Visual design: focus + context

- Focus
 - Selected subtrees



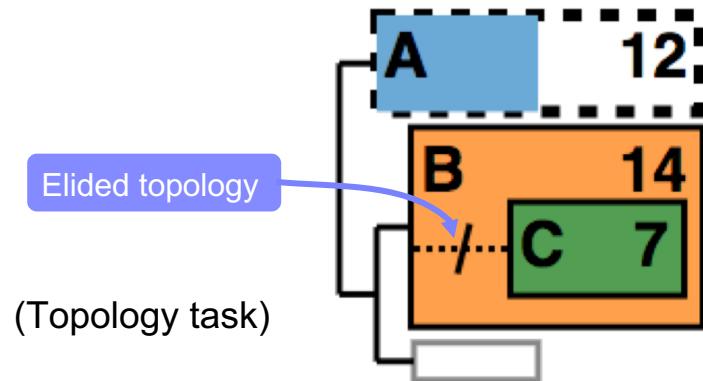
Visual design: focus + context

- Focus
 - Selected subtrees
 - Topological relationships between them



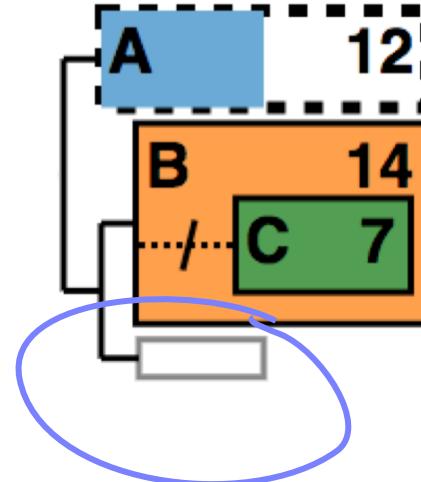
Visual design: focus + context

- Focus
 - Selected subtrees
 - Topological relationships between them



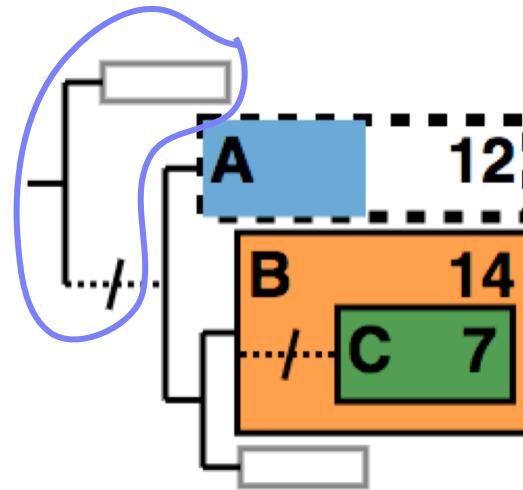
Visual design: focus + context

- Focus
 - Selected subtrees
 - Topological relationships between them
- Context
 - Neighboring subtrees



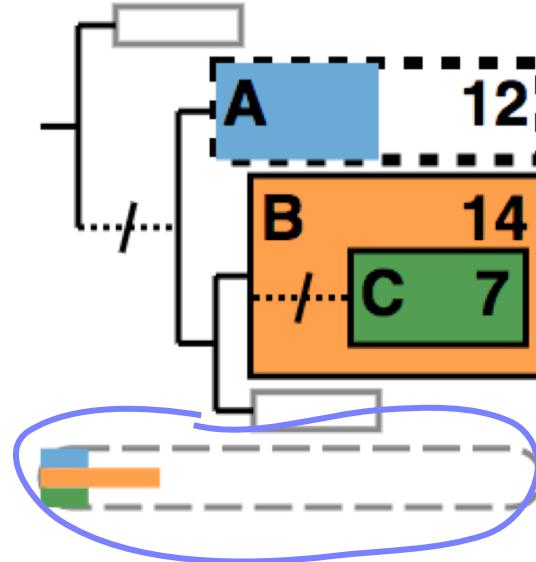
Visual design: focus + context

- Focus
 - Selected subtrees
 - Topological relationships between them
- Context
 - Neighboring subtrees
 - Upstream topology and root



Visual design: focus + context

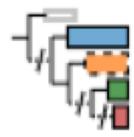
- Focus
 - Selected subtrees
 - Topological relationships between them
- Context
 - Neighboring subtrees
 - Upstream topology and root
 - Missing leaf nodes



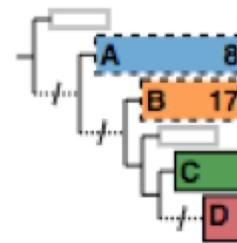
Visual design: algorithm adapts to space

- Show more info when space permitted
 - Labels
 - #leaf nodes
 - Neighboring blocks

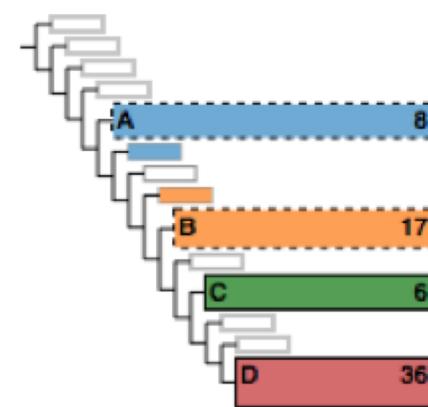
40x40 px



80x80 px

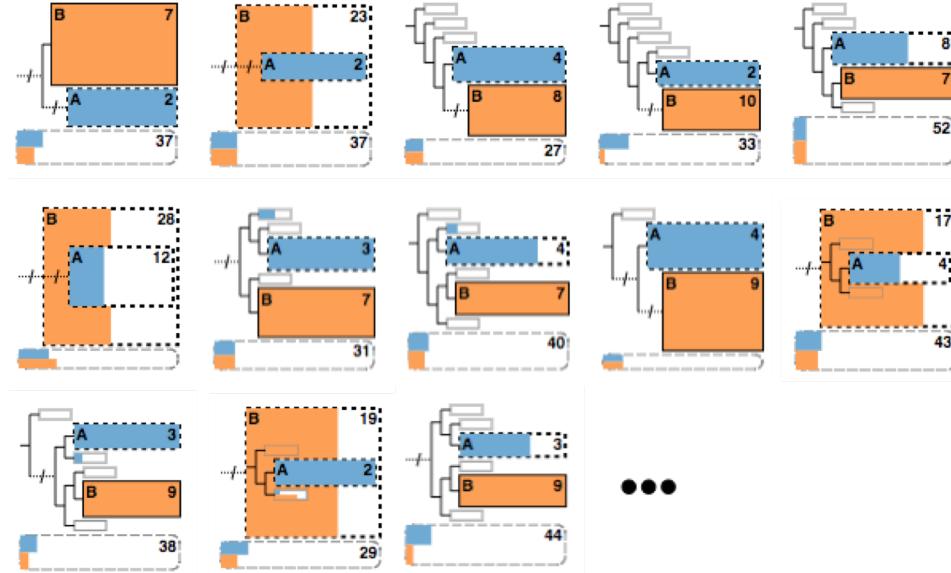


160x160 px



Cluster AD: summary of multiple ADs

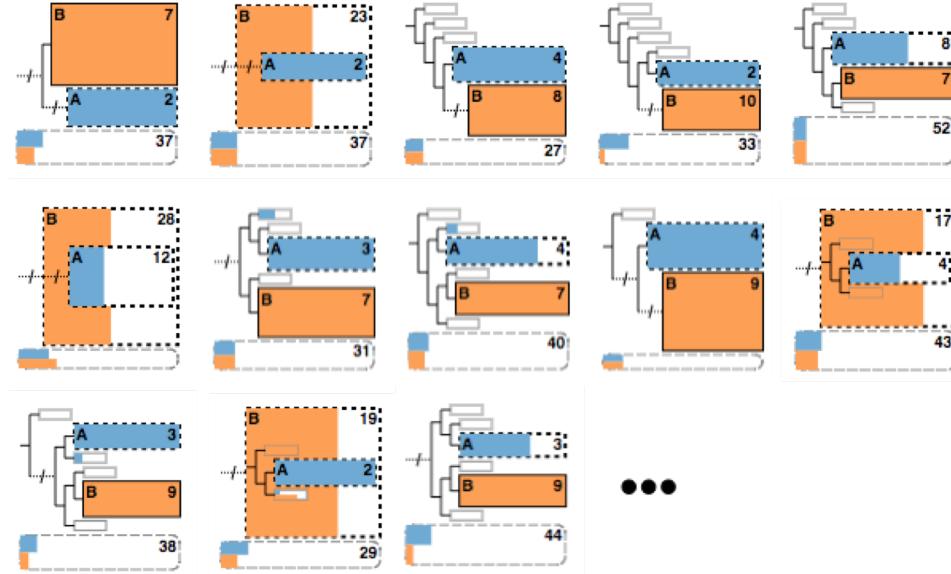
Problem: still many individual ADs



Cluster AD: summary of multiple ADs

Problem: still many individual ADs

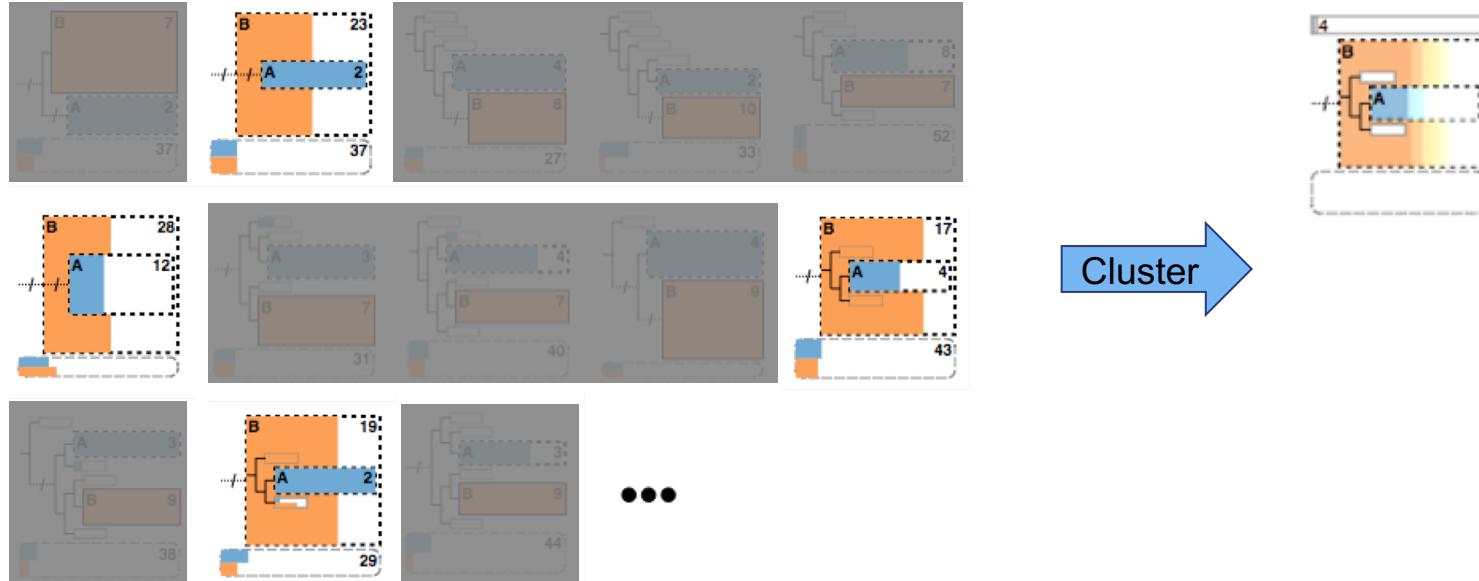
Solution: cluster by topological relationships



Cluster AD: summary of multiple ADs

Problem: still many individual ADs

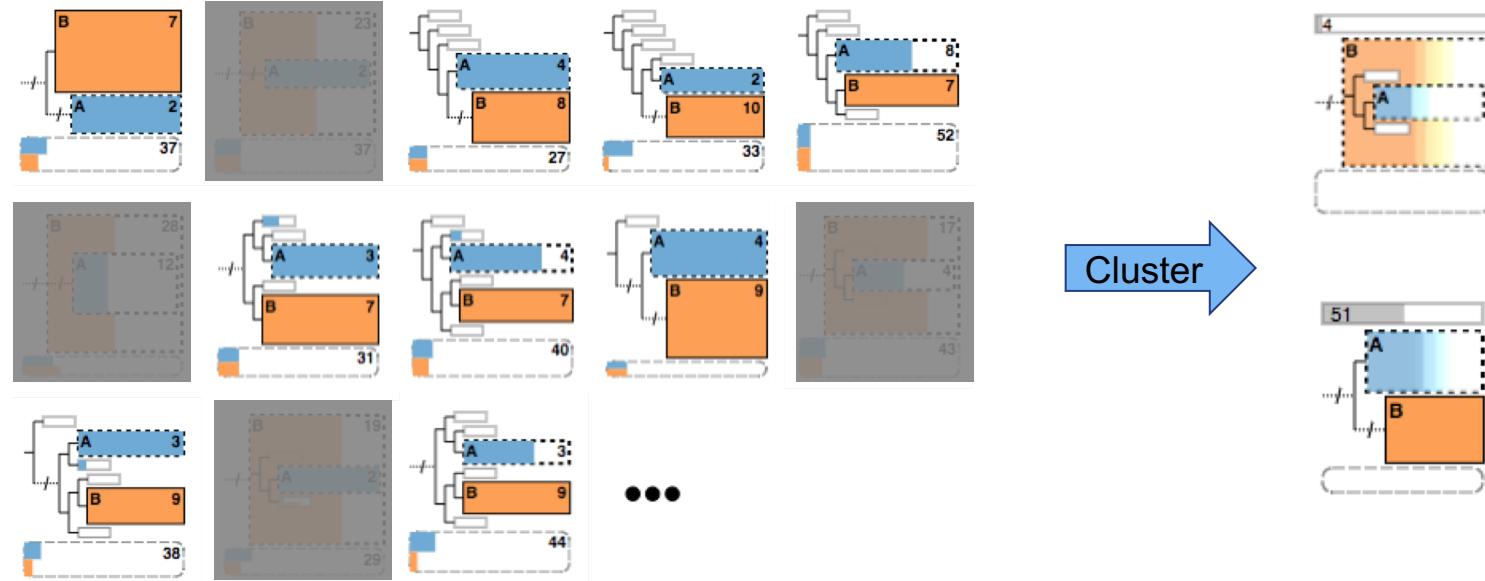
Solution: cluster by topological relationships



Cluster AD: summary of multiple ADs

Problem: still many individual ADs

Solution: cluster by topological relationships





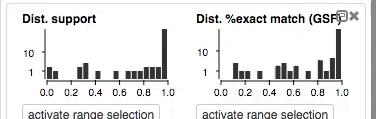
Introduction
Data & Tasks
Aggregated Dendrogram (AD)

ADView

- Demo
- Multi-level structure across views

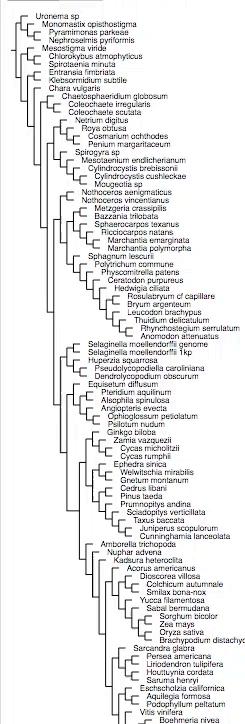
Summary

Reference Tree



astral.FAA.trim50genes.final.tre (103 taxa)

Encode | Ignore | branch length.



Branch: ■ in attribute range ■ hovered ■ for tree similarity
 Taxon: H (highlighted) T (taxon hovered) U (uncertain taxon hovered)
 Monophyletic: ■■■■■ highlight 1-5

Tree Collection

pairwise compare make consensus reset selected trees (0) sub-collection more

Tree Distribution by Reference Partition

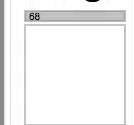
There is no branch of interest yet. Please select from the reference tree (click a branch).

Segment (group of trees): shared same taxa membership with ref. (●) selected (■■■) hovered (□)

Aggregated Dendograms (AD)

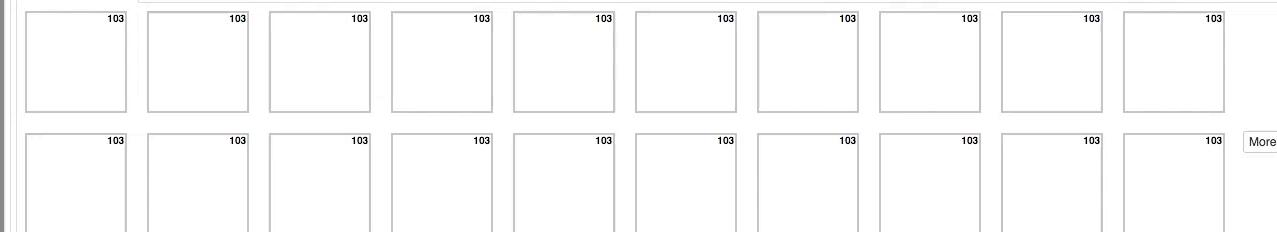
manual size: width: 100 height: 100 max #context levels: 2 show labels: show colors:

▼ Clusters 1 differentiate inexact match: differentiate sister-group relationships: (caveat: trees in a cluster agree only on relations among named blocks (A, B, C, ...), but not necessarily on context blocks)



▼ Individuals (sort by the RF distance to the reference tree). Show tree names:

● All Trees 68

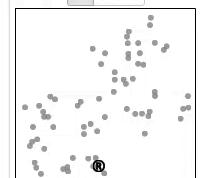


Tree List (68)

- FAA.604genes.trimExtensively.phyloblocks
- FAA.604genes.trimExtensively.phyloblocks
- FAA.604genes.trimExtensively.unpartitioned
- FAA.604genes.trimExtensively.unpartitioned
- FAA.trim50genes3sites.unpartitioned
- FAA.trim50genes3sites3taxa.unpartitioned
- FAA.trim50genes5sites.clustered.partitioned
- FAA.trim50genes5sites.partitioned.gn
- FAA.trim50genes5sites.unpartitioned
- FAA.trim50genes5sites25Xbranches
- FAA.trim50genesChara.allPos.unpartitioned
- FAA.untrimmed.unpartitioned.final.tre
- FNA.trim50genes5sites.unpartitioned

Tree Similarity

Distance: full subtree



●+■+■+■+■ sub-collection mark
 ● ref. ■ hovered ● selected

Corresponding Br. Attr.

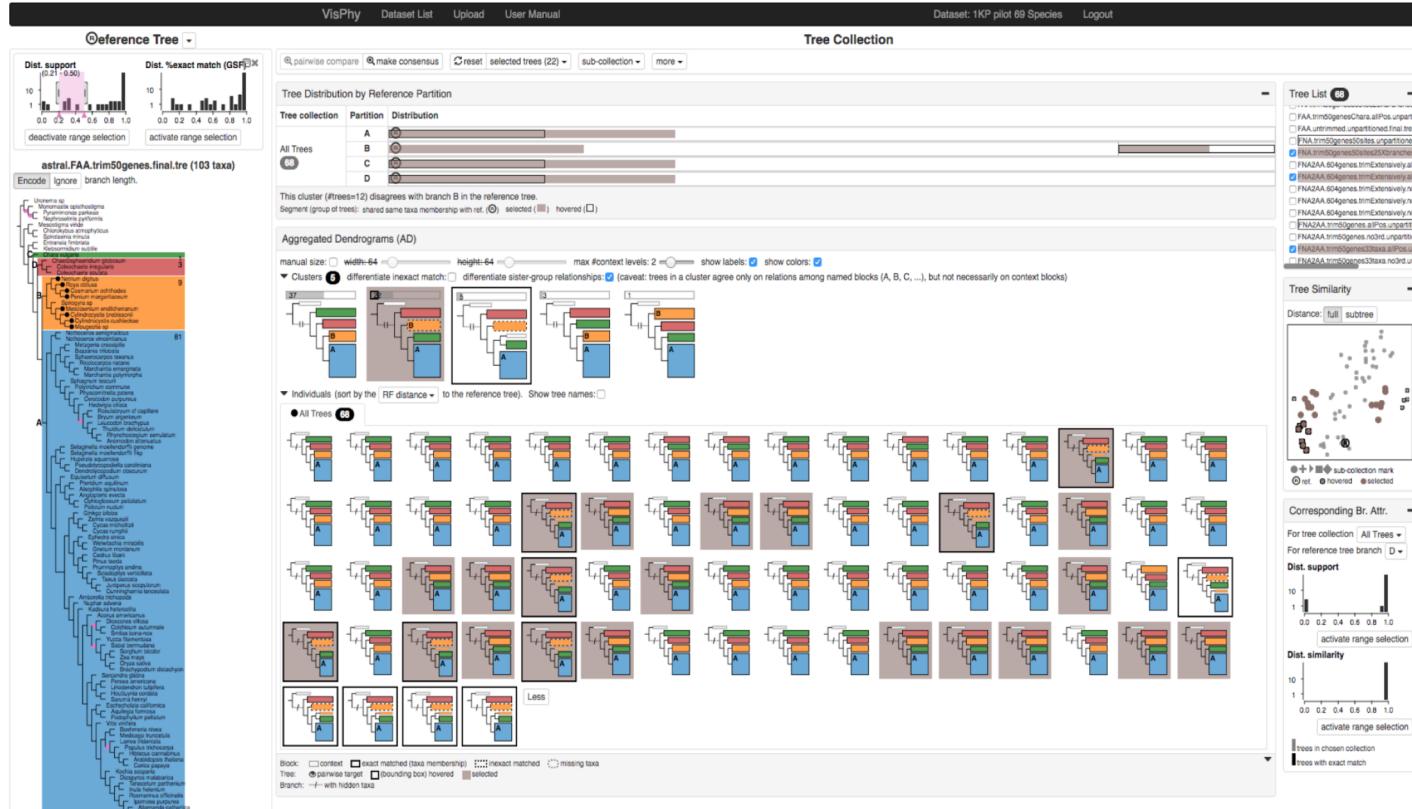
No branch of interest yet.

Block: ■ context ■■ exact matched (taxa membership) ■■■ inexact matched ■■■■ missing taxa

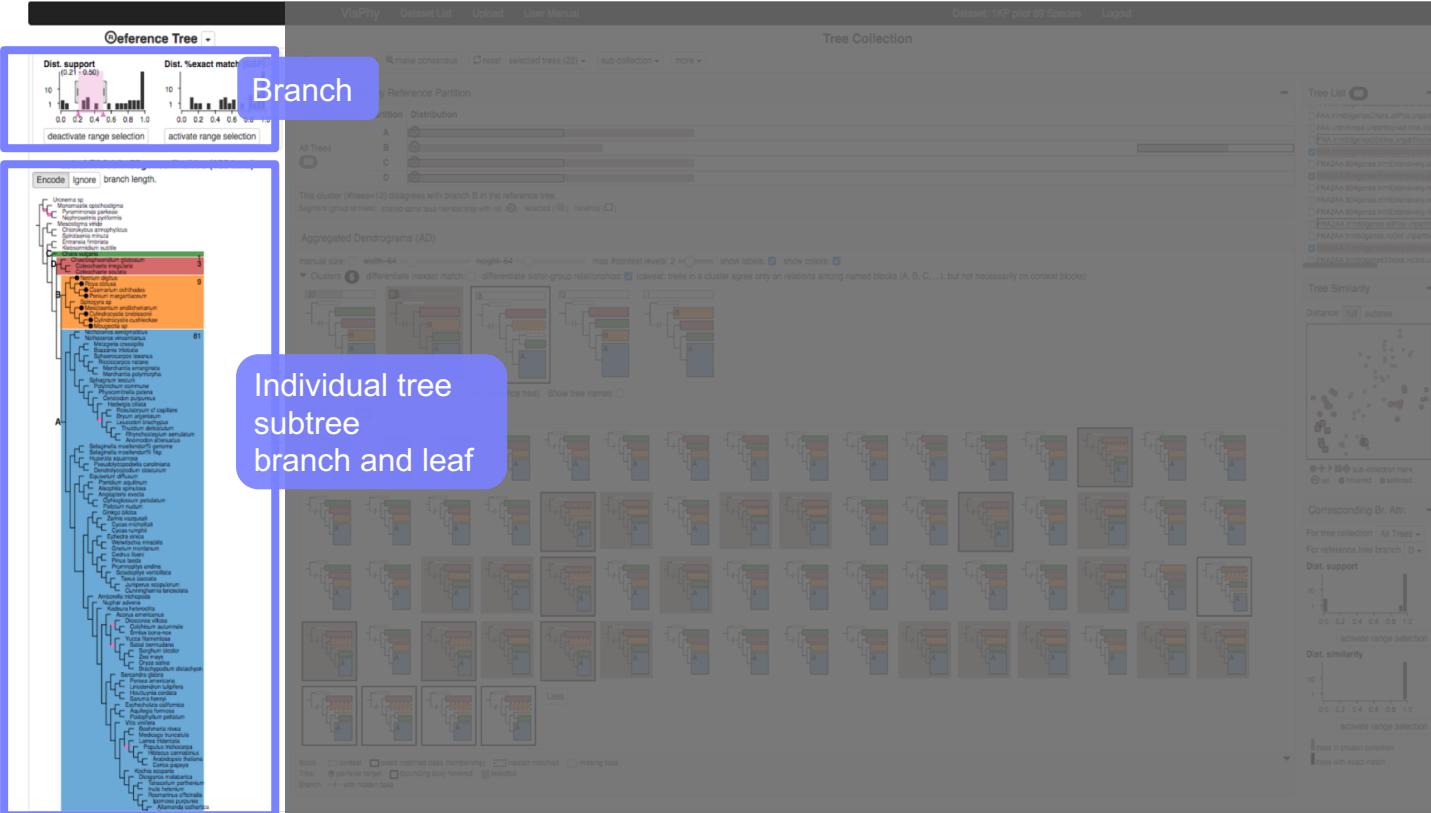
Tree: pairwise target (bounding box) hovered ■■■ selected

Branch: ■—■ with hidden taxa

Multi-level structure across views



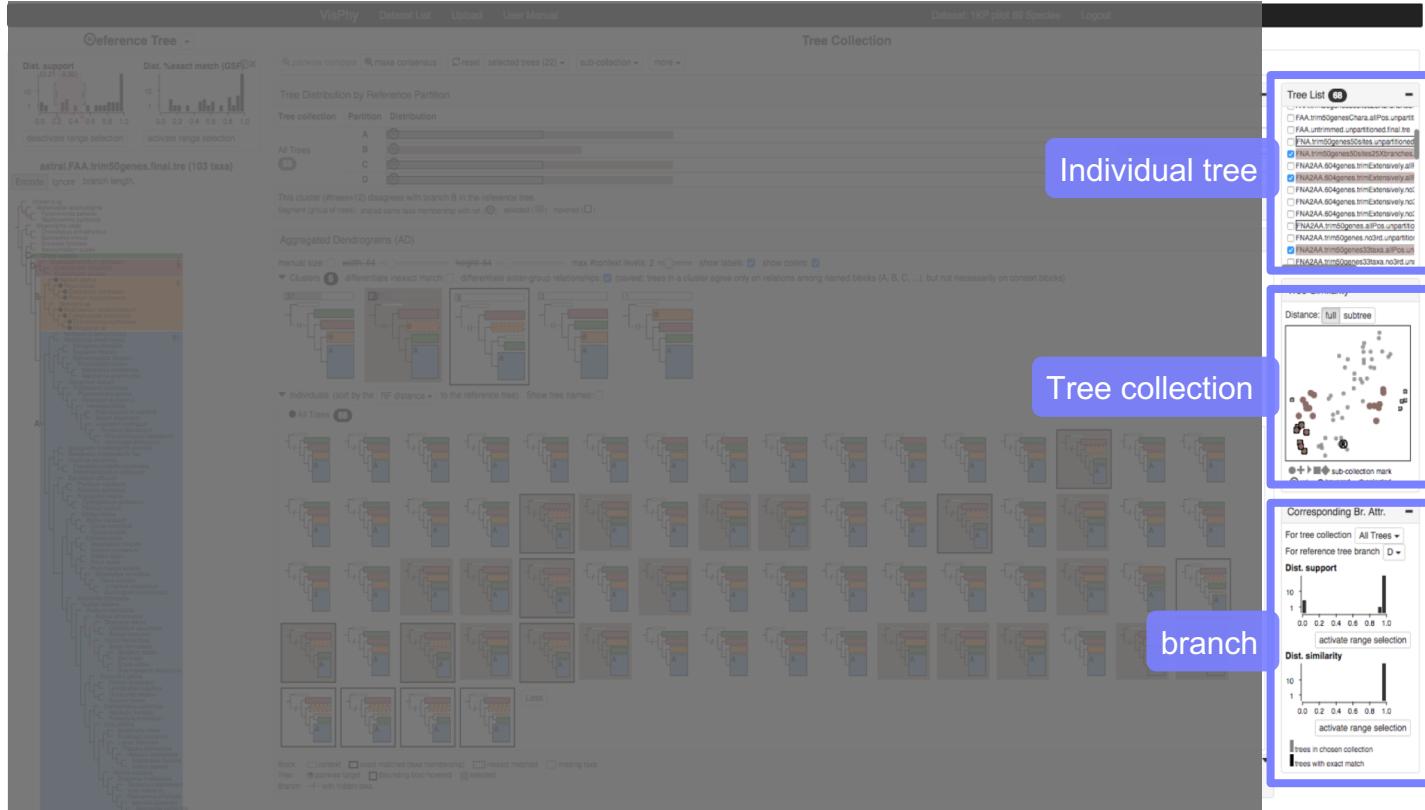
Multi-level structure across views



Interface walkthrough: tree collection main views



Interface walkthrough: tree collection aux. views

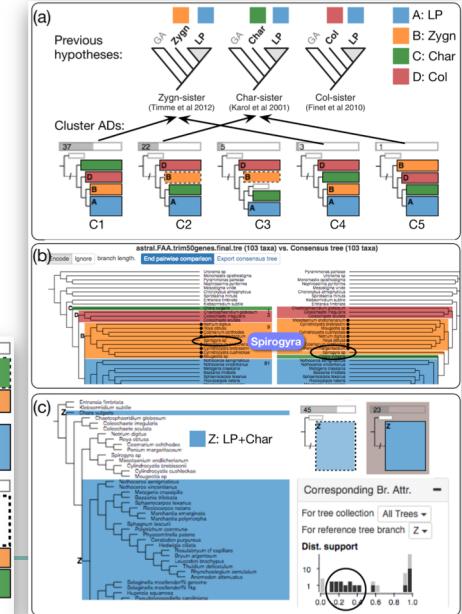
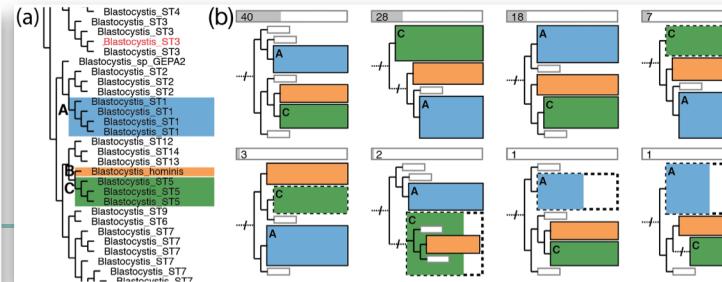
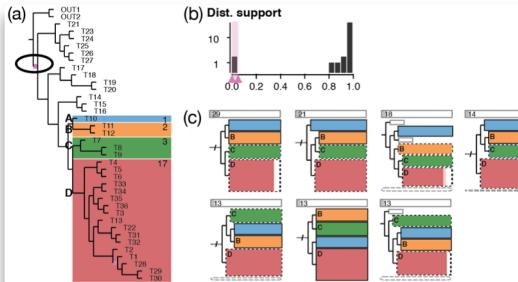


Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
 - 10 biologists at different times throughout project

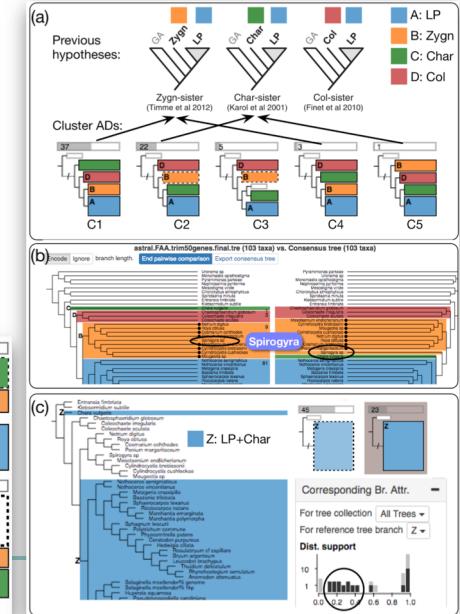
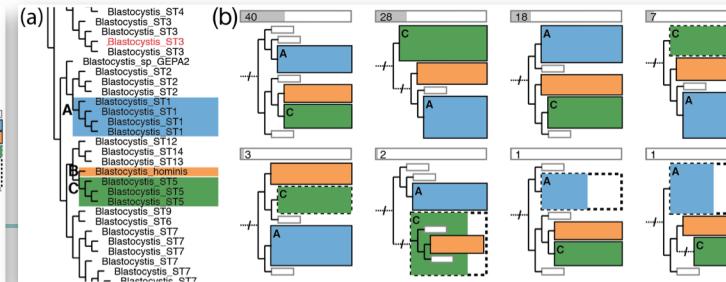
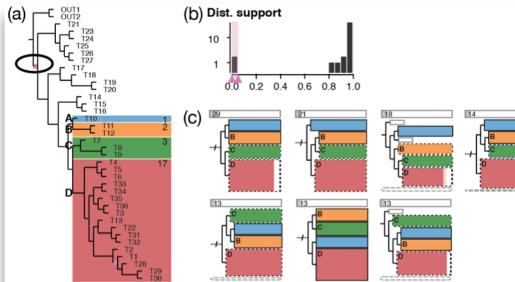
Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
 - 10 biologists at different times throughout project
- User study sessions
 - 5 biologists
 - Using their own datasets



Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
 - 10 biologists at different times throughout project
- User study sessions
 - 5 biologists
 - Using their own datasets
- Biologists confirmed
 - Validity of data and task abstractions
 - Utility of ADView



Conclusion

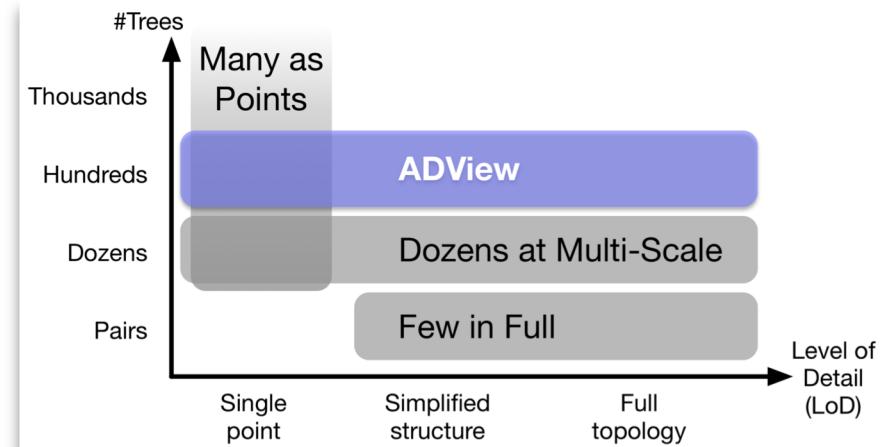
- Data and task abstractions for comparing many phylogenetic trees

Conclusion

- Data and task abstractions for comparing many phylogenetic trees
- Aggregated dendrogram (AD)
 - Compress a full dendrogram to a compact glyph

Conclusion

- Data and task abstractions for comparing many phylogenetic trees
- Aggregated dendrogram (AD)
 - Compress a full dendrogram to a compact glyph
- ADView
 - Compare **hundreds** of trees at **multiple scales**

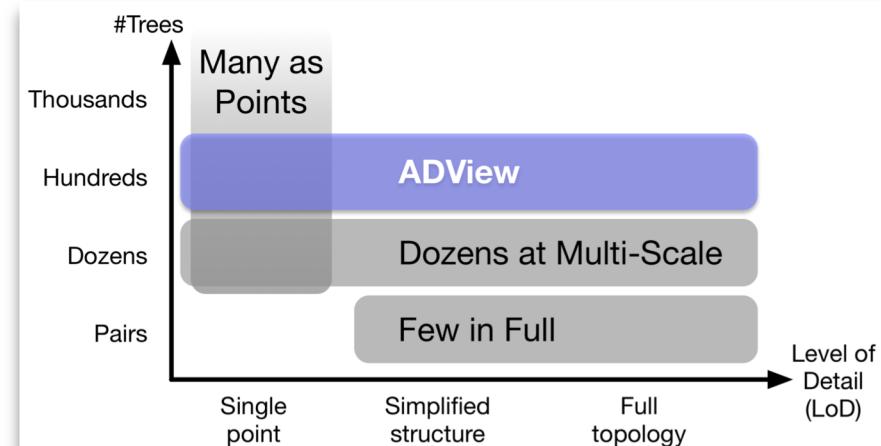


Conclusion

- Data and task abstractions for comparing many phylogenetic trees
- Aggregated dendrogram (AD)
 - Compress a full dendrogram to a compact glyph
- ADView
 - Compare **hundreds** of trees at **multiple scales**

Future work

- Thousands at multi-scale
- Generalize to other domains



Aggregated Dendrograms for Visual Comparison Between Many Phylogenetic Trees

Zipeng Liu (zipeng@cs.ubc.ca), Shing Hei Zhan (shing.zhan@gmail.com), Tamara Munzner ([@tamaramunzner](https://twitter.com/tamaramunzner))

Project page:

<http://www.cs.ubc.ca/labs/imager/tr/2019/adview/>

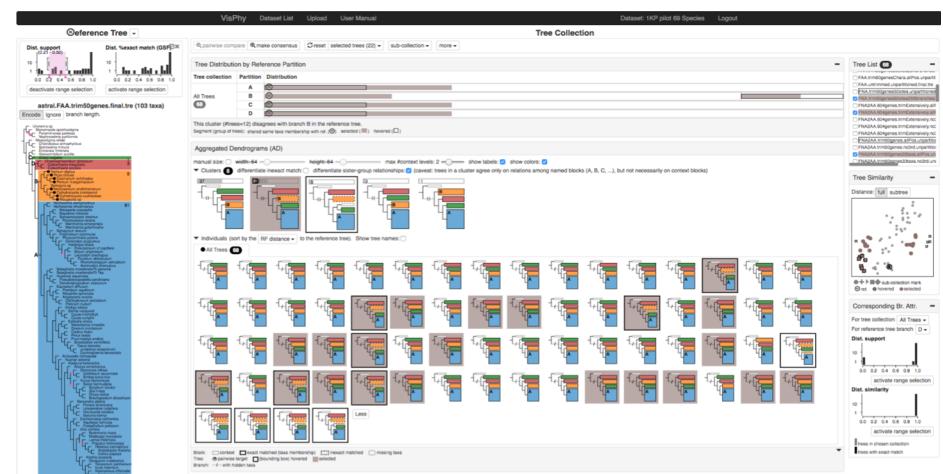
- Paper, supplemental materials, videos, talk slides
 - Biology use cases with real-world data

Open-source code:

<https://github.com/zipengliu/adview>

Live demo:

<http://visphy.cs.ubc.ca>

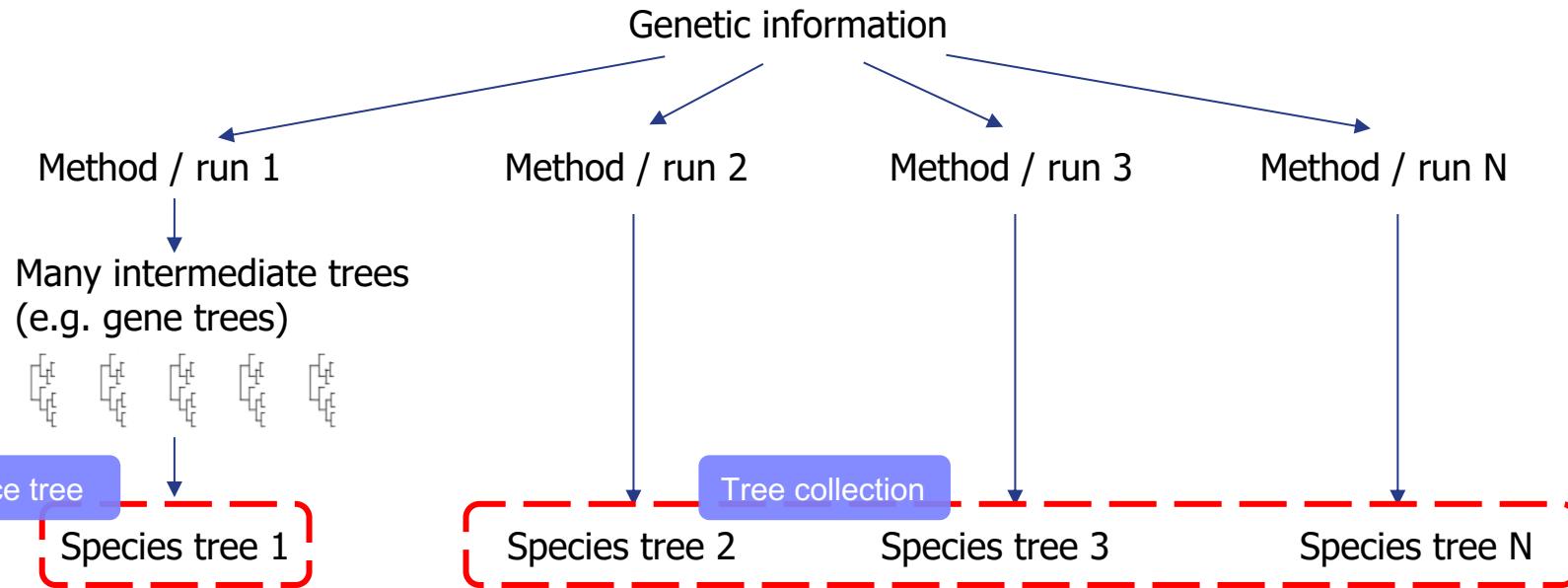




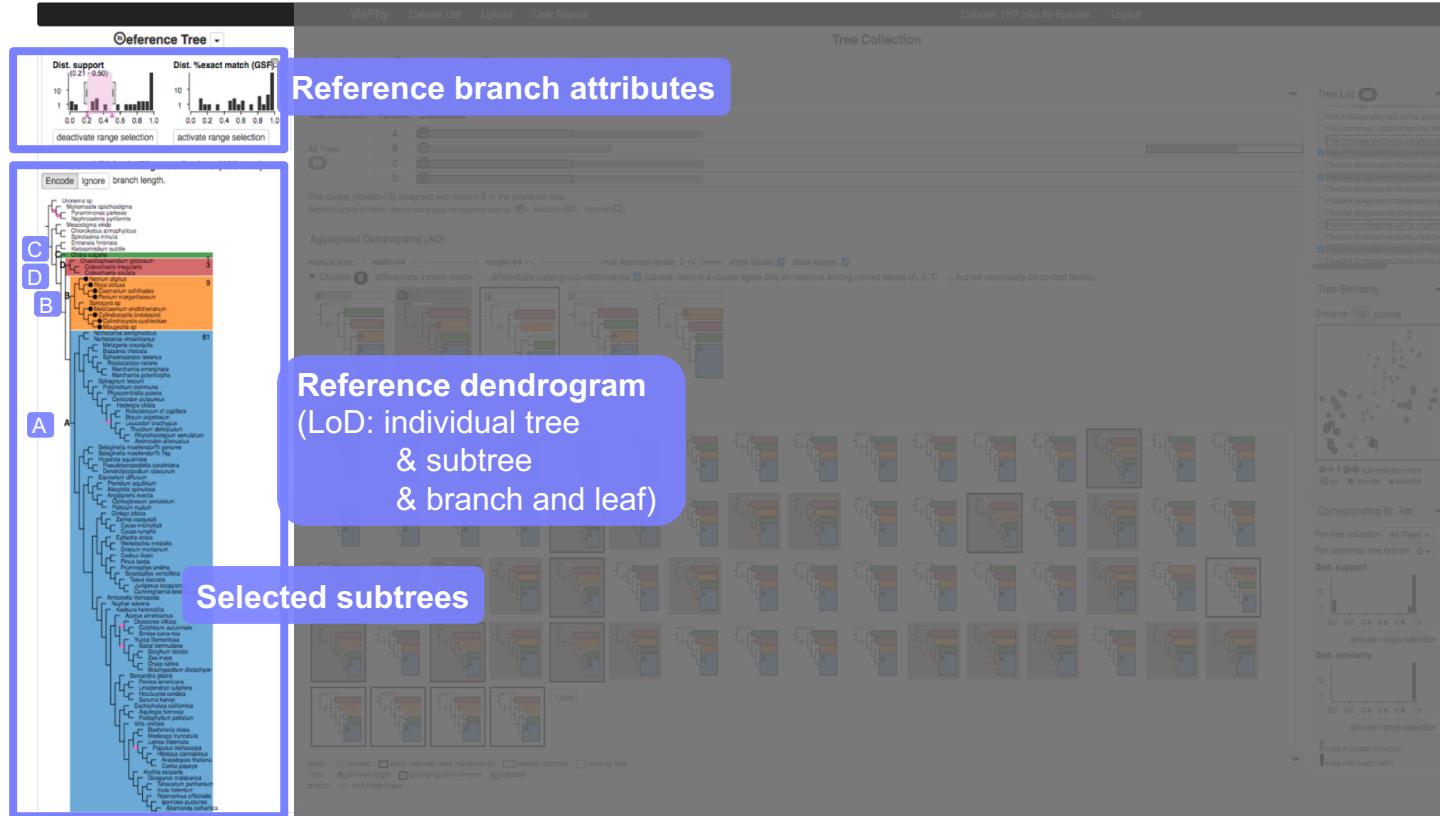
Backup slides

Tree data

- 1 vs N: reference tree vs. tree collection



Multi-level structure across views



Scalability of ADView

- #Trees
 - A few hundred (with cluster AD)
- Level of detail
 - Tree collection
 - Subset of trees
 - Individual tree
 - Subtree
 - Branch and leaf node
- #leaf nodes per tree
 - ~100 if need to show all in one screen
 - Scale up by hiding parts of tree

