

# Pixel-Adaptive Visual Comparison Between Many Phylogenetic Trees

**Tamara Munzner**

Department of Computer Science  
University of British Columbia

*Asilomar Microcomputer Workshop #50  
25 Apr 2024*

**<http://www.cs.ubc.ca/~tmm/talks.html#amw24>**



 **[@tamara@vis.social](https://mstdn.social/@tamaramunzner)**

 **[@tamaramunzner](https://twitter.com/tamaramunzner)**

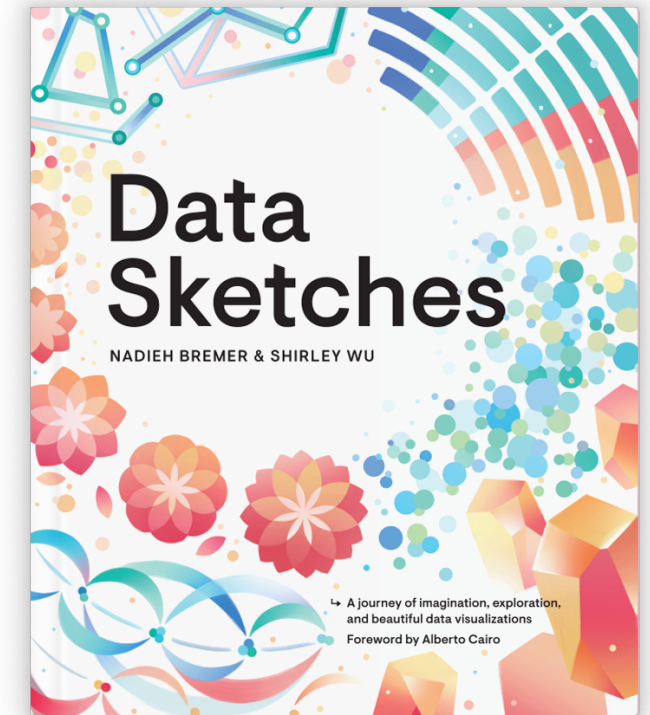
# Hi again, 15 years later!

- still doing data visualization
  - yet more papers / projects / videos / software
  - I edit a book series

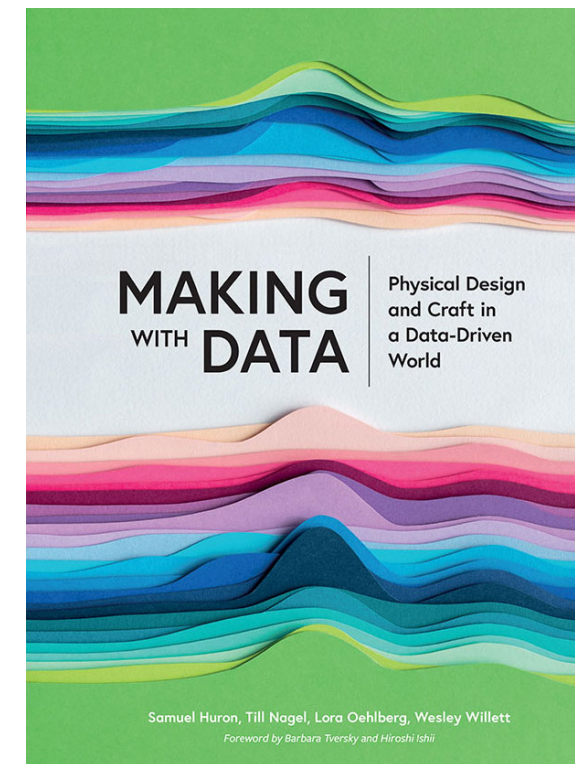
**<http://www.cs.ubc.ca/~tmm/talks.html#amw24>**

# Visualization book series highlights

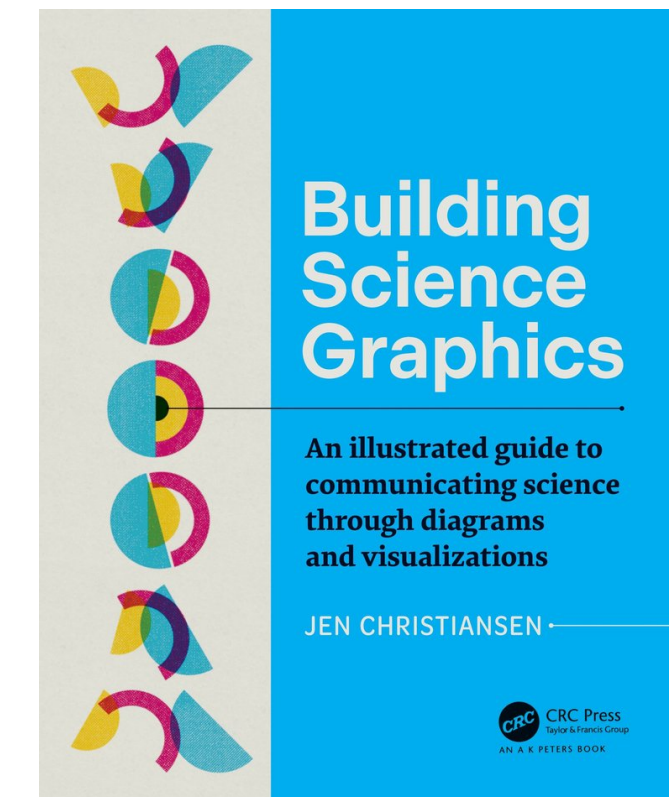
- Data Sketches, by Nadieh Bremer & Shirley Wu



- Making with Data, by multitudes



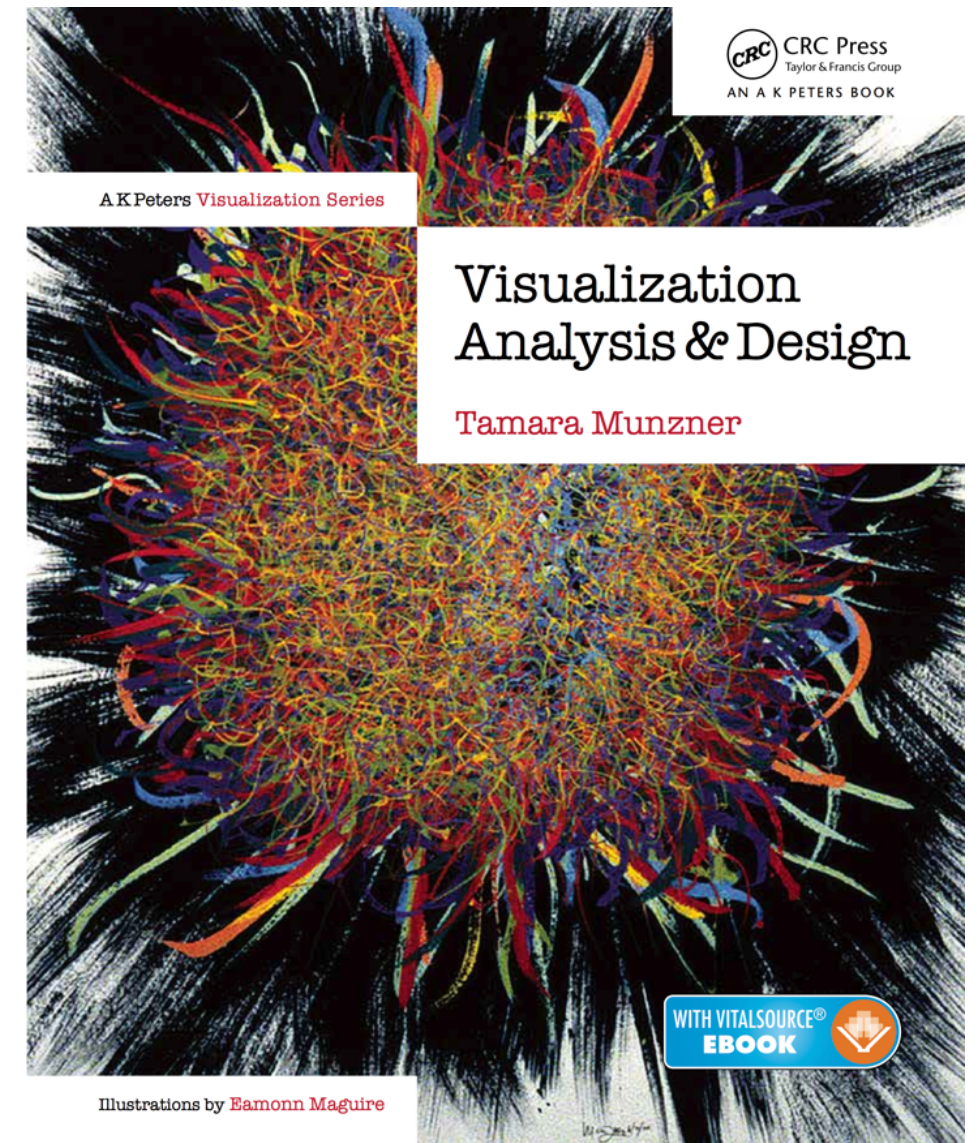
- Building Science Graphics, by Jen Christiansen



<https://www.routledge.com/AK-Peters-Visualization-Series/book-series/CRCVIS>

# Hi again, 15 years later!

- still doing data visualization
  - yet more papers / projects / videos / software
  - I edit a book series
  - I wrote a book



Visualization Analysis and Design. Munzner.  
CRC Press, AK Peters Visualization Series, 2014.

<http://www.cs.ubc.ca/~tmm/talks.html#amw24>

# Visualization Analysis & Design book

What?

Datasets

Attributes

Why?

Actions

Targets

How?

Encode

Manipulate

Facet

Reduce

Arrange

Express

Separate

Map

from **categorical** and **ordered** attributes

Color

→ Hue → Saturation → Luminance

Size, Angle, Curvature, ...

Shape

Motion

Direction, Rate, Frequency, ...

Change

Select

Navigate

Juxtapose

Partition

Superimpose

Filter

Aggregate

Embed

domain

abstraction

idiom

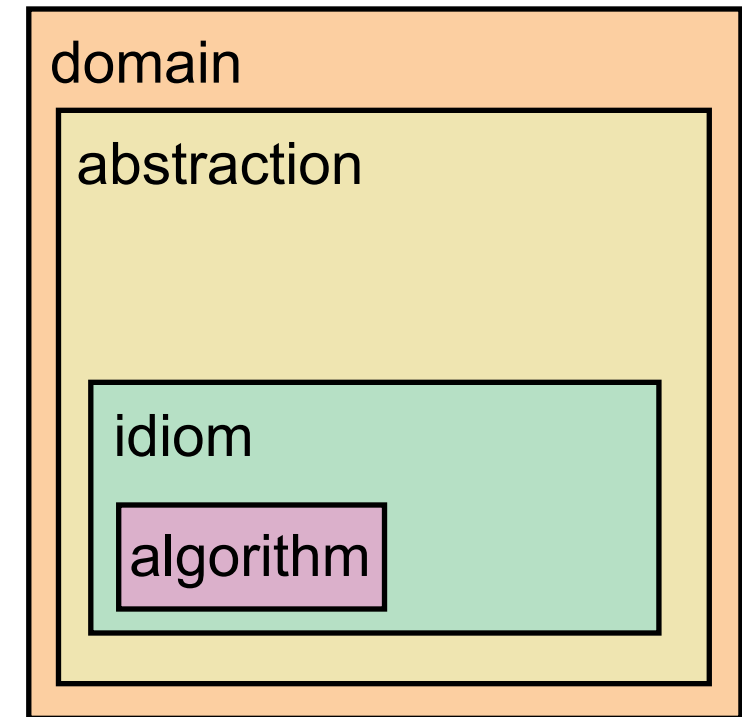
algorithm

What?

Why?

How?

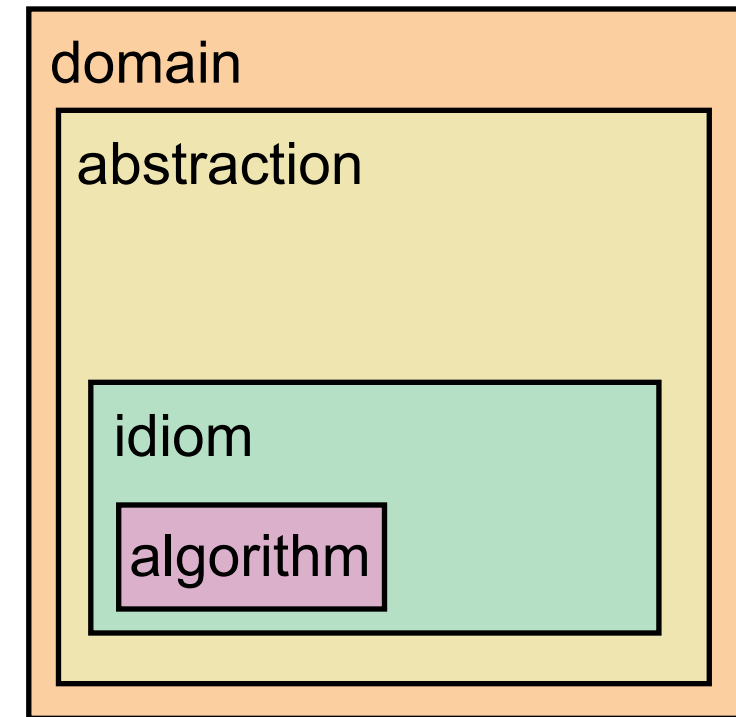
# Nested model: Four levels of visualization concerns



[A Nested Model of Visualization Design and Validation. Munzner. *IEEE TVCG* 15(6):921-928, 2009 (Proc. InfoVis 2009).]

# Nested model: Four levels of visualization concerns

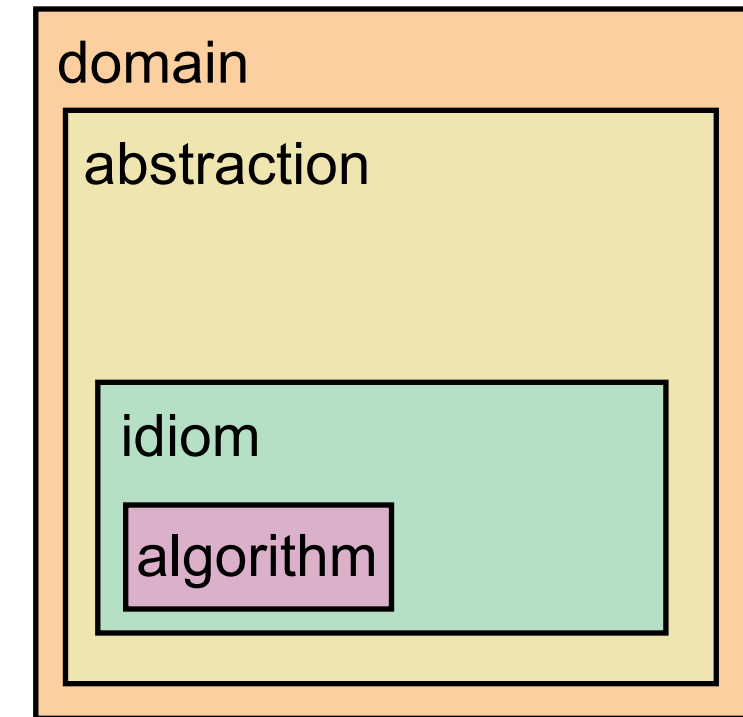
- *domain* situation
  - **who** are the target users?



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

# Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis

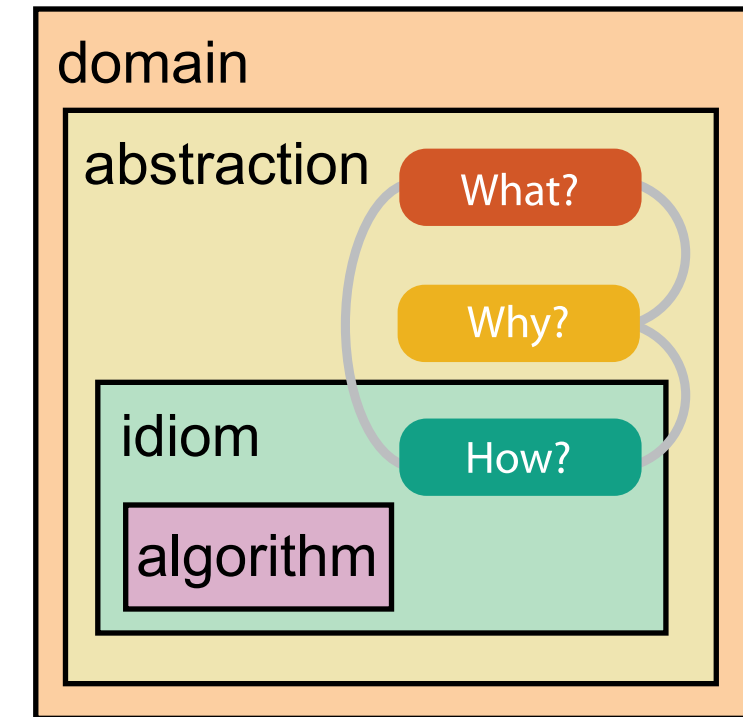


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]



# Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**

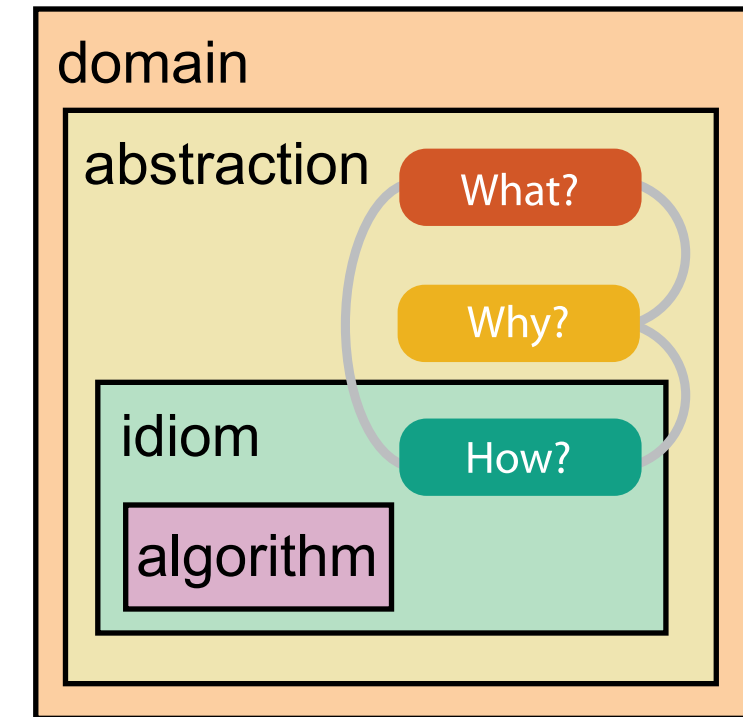


[A Nested Model of Visualization Design and Validation. Munzner. *IEEE TVCG* 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. *IEEE TVCG* 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

# Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form

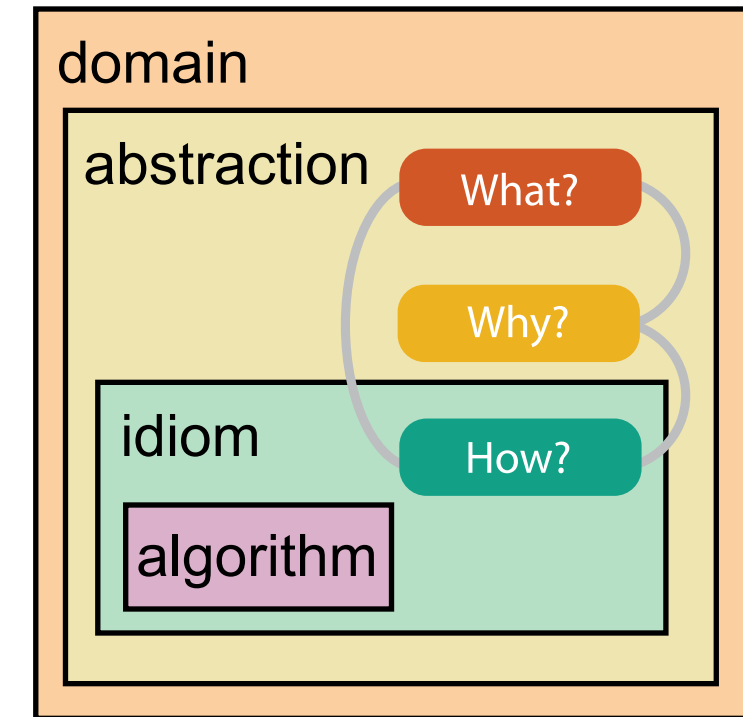


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

# Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**

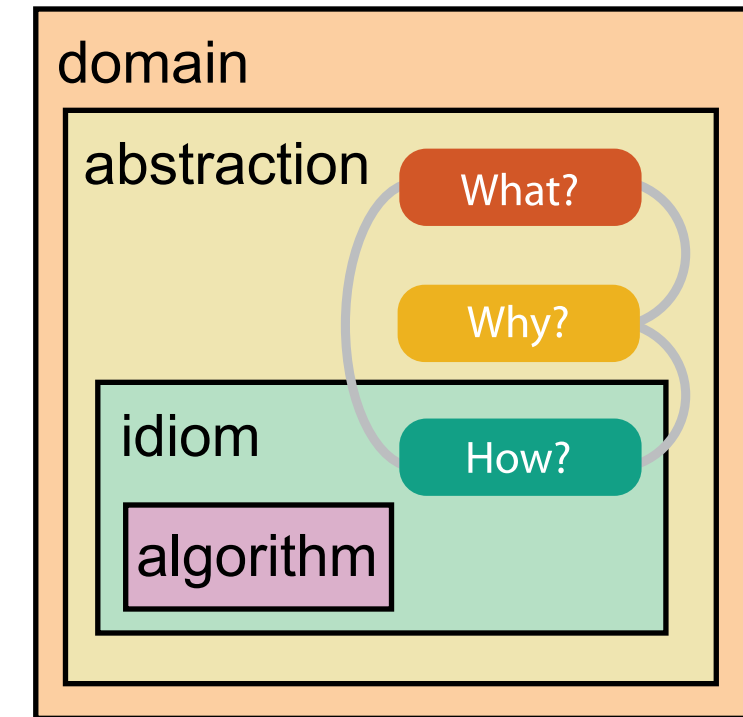


[A Nested Model of Visualization Design and Validation. Munzner. *IEEE TVCG* 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. *IEEE TVCG* 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

# Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- *idiom*
  - **how** is it shown?

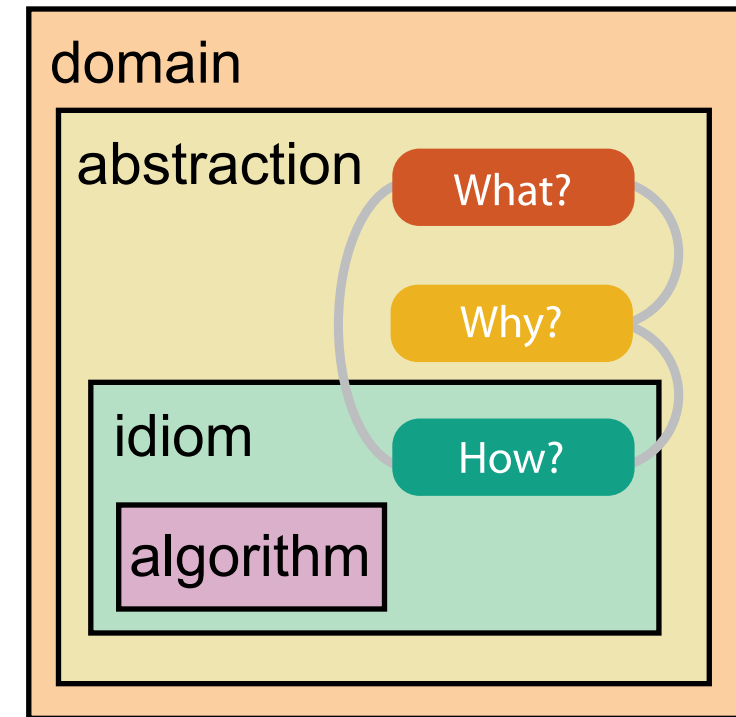


[A Nested Model of Visualization Design and Validation. Munzner. *IEEE TVCG* 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. *IEEE TVCG* 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

# Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- *idiom*
  - **how** is it shown?
    - **visual encoding idiom**: how to draw

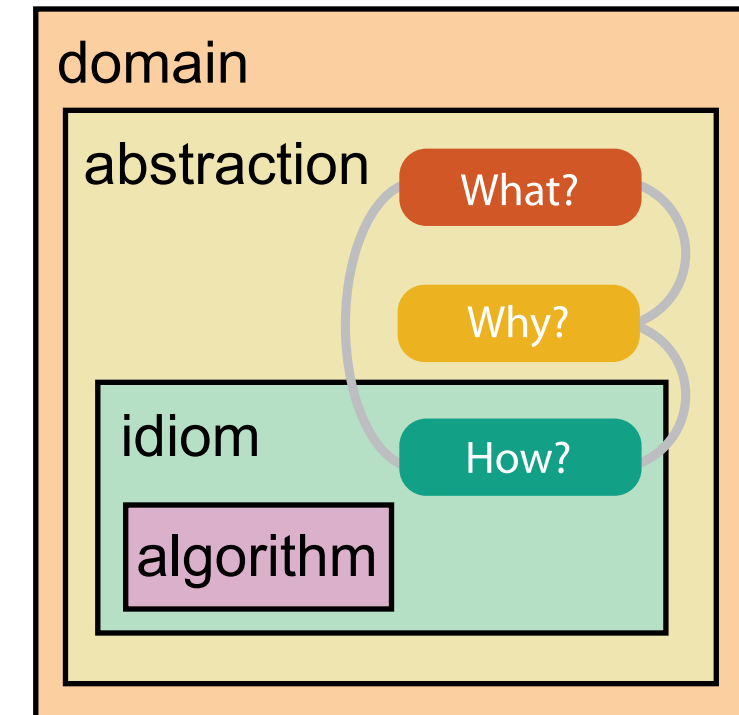


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

# Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- *idiom*
  - **how** is it shown?
    - **visual encoding idiom**: how to draw
    - **interaction idiom**: how to manipulate

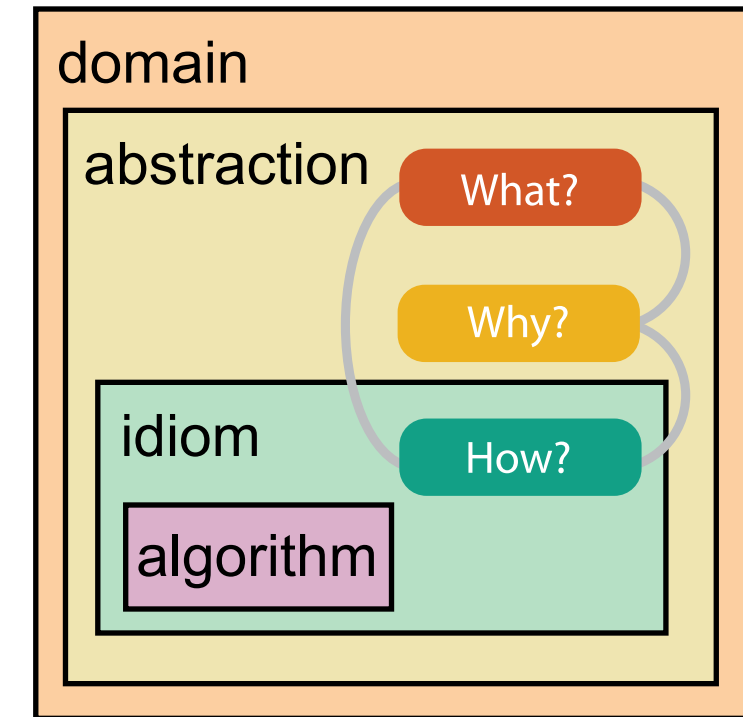


[A Nested Model of Visualization Design and Validation. Munzner. *IEEE TVCG* 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. *IEEE TVCG* 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

# Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- *idiom*
  - **how** is it shown?
    - **visual encoding idiom**: how to draw
    - **interaction idiom**: how to manipulate
- *algorithm*
  - efficient computation



[A Nested Model of Visualization Design and Validation. Munzner. *IEEE TVCG* 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. *IEEE TVCG* 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

# Why is validation difficult?

- different threats to validity at each level

*[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]*



# Why is validation difficult?

- different threats to validity at each level



## Domain situation

You misunderstood their needs

*[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]*

# Why is validation difficult?

- different threats to validity at each level

## **Domain situation**

You misunderstood their needs


## **Data/task abstraction**

You're showing them the wrong thing


*[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]*

# Why is validation difficult?

- different threats to validity at each level

 **Domain situation**  
You misunderstood their needs

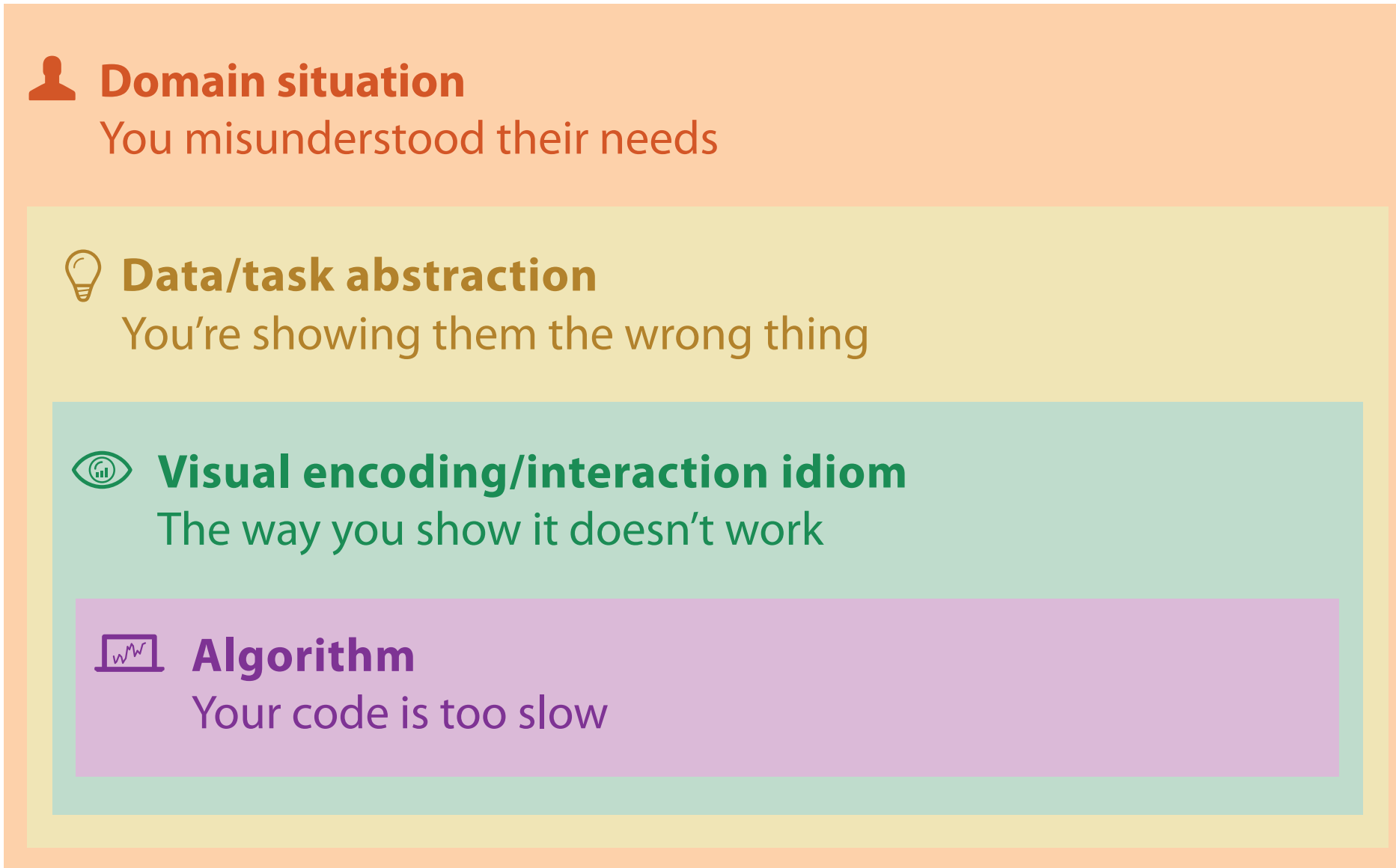
 **Data/task abstraction**  
You're showing them the wrong thing

 **Visual encoding/interaction idiom**  
The way you show it doesn't work

*[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]*

# Why is validation difficult?

- different threats to validity at each level



[A Nested Model of Visualization Design and Validation. Munzner. *IEEE TVCG* 15(6):921-928, 2009 (Proc. InfoVis 2009).]

**Validation solution: Use methods from appropriate fields at each level**

# Validation solution: Use methods from appropriate fields at each level

computer  
science



## Algorithm

Measure system time/memory

Analyze computational complexity

# Validation solution: Use methods from appropriate fields at each level

computer  
science




technique-driven  
work

# Validation solution: Use methods from appropriate fields at each level

design

 **Visual encoding/interaction idiom**  
Justify design with respect to alternatives

computer  
science

 **Algorithm**  
Measure system time/memory  
Analyze computational complexity

cognitive  
psychology

Analyze results qualitatively  
Measure human time with lab experiment (*lab study*)



technique-driven  
work



# Validation solution: Use methods from appropriate fields at each level

anthropology/  
ethnography


 **Domain situation**  
Observe target users using existing tools

 **Data/task abstraction**

design

 **Visual encoding/interaction idiom**  
Justify design with respect to alternatives

computer  
science

 **Algorithm**  
Measure system time/memory  
Analyze computational complexity

cognitive  
psychology

Analyze results qualitatively  
Measure human time with lab experiment (*lab study*)

anthropology/  
ethnography

Observe target users after deployment (*field study*)  
**Measure adoption**

technique-driven  
work

# Validation solution: Use methods from appropriate fields at each level

- avoid mismatches between level and validation

anthropology/  
ethnography


 **Domain situation**  
Observe target users using existing tools

 **Data/task abstraction**

design

 **Visual encoding/interaction idiom**  
Justify design with respect to alternatives

computer  
science

 **Algorithm**  
Measure system time/memory  
Analyze computational complexity

cognitive  
psychology

Analyze results qualitatively  
Measure human time with lab experiment (*lab study*)

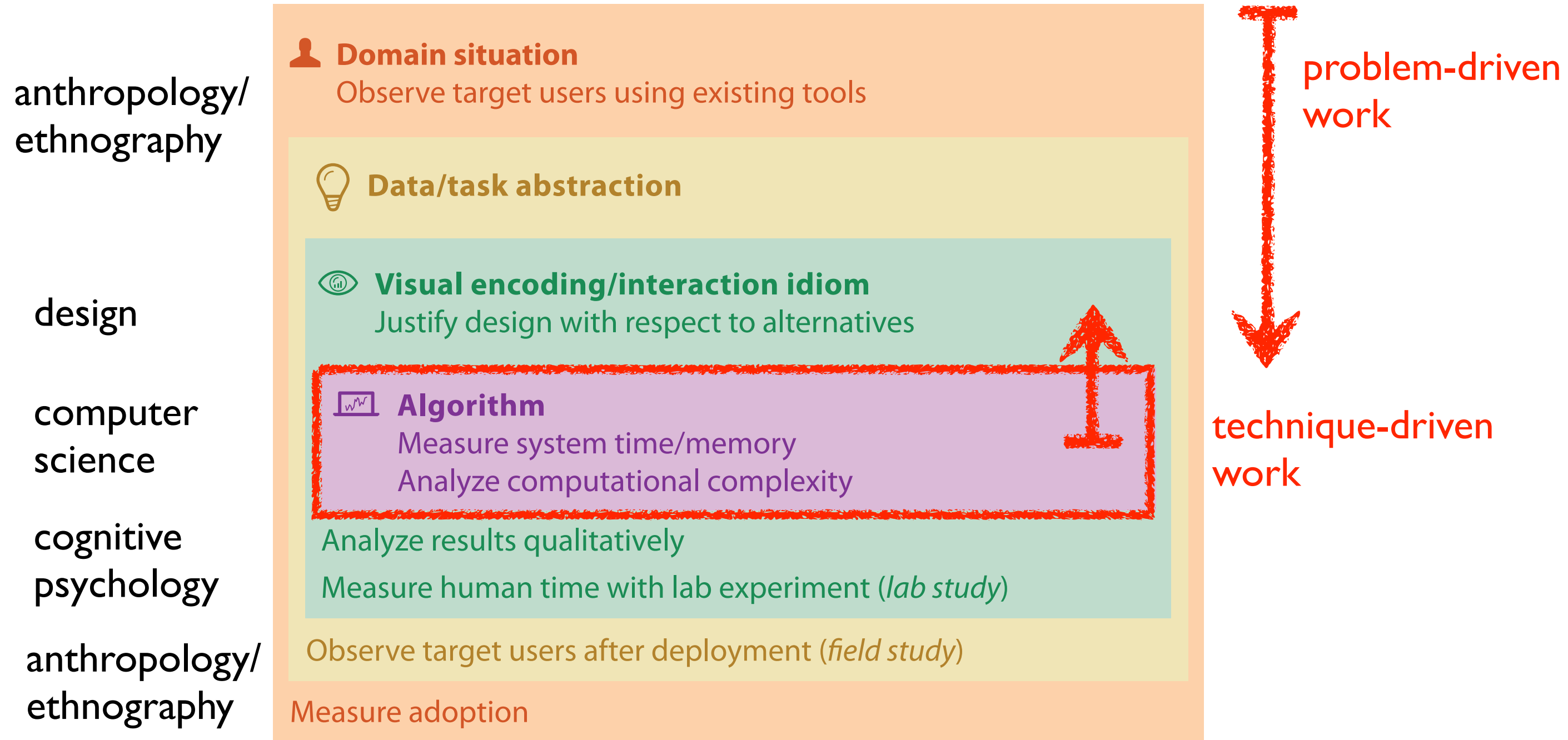
anthropology/  
ethnography

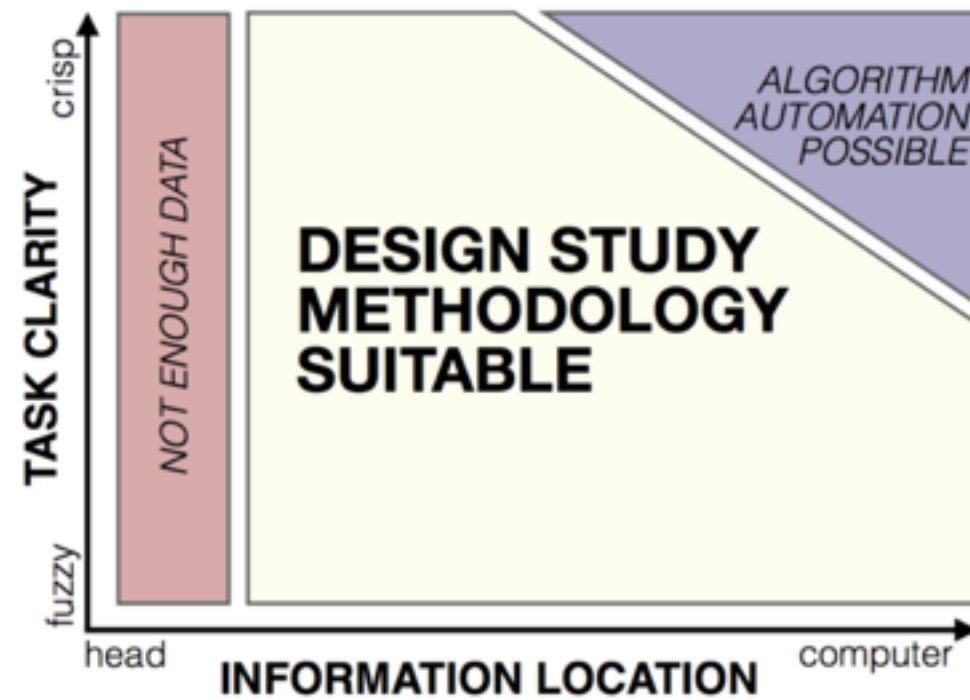
Observe target users after deployment (*field study*)  
**Measure adoption**

technique-driven  
work

# Validation solution: Use methods from appropriate fields at each level

- avoid mismatches between level and validation





Michael Sedlmair



Miriah Meyer



# Design Study Methodology

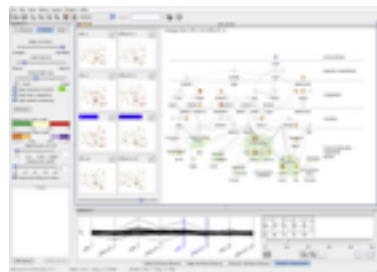
*Reflections from the Trenches and from the Stacks*

<http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/>

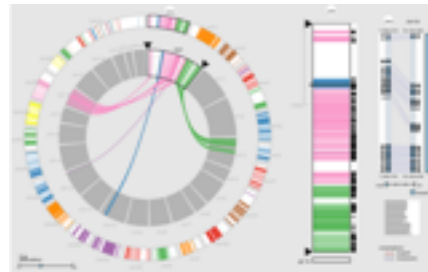
Design Study Methodology: Reflections from the Trenches and from the Stacks.

Sedlmair, Meyer, Munzner. *IEEE Trans. Visualization and Computer Graphics* 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).

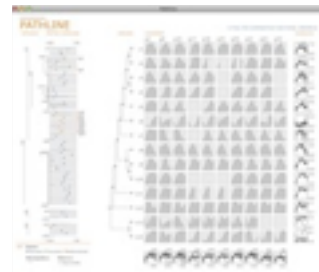
# Lessons learned from the trenches: 20+ between us



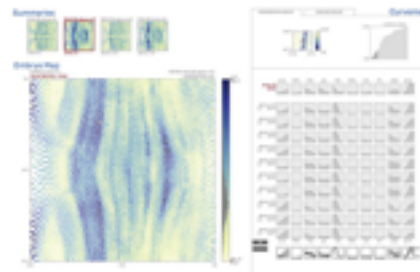
*Cerebral*  
genomics



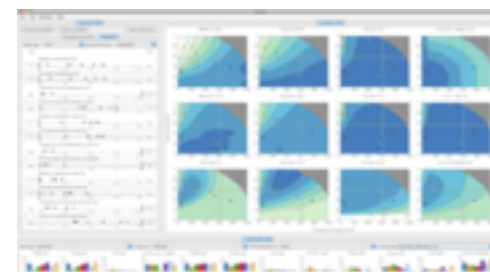
*MizBee*  
genomics



*Pathline*  
genomics



*MulteeSum*  
genomics



*Vismon*  
fisheries management



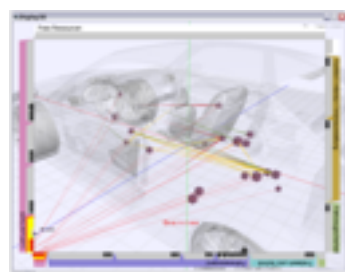
*QuestVis*  
sustainability



*WiKeVis*  
in-car networks



*MostVis*  
in-car networks



*Car-X-Ray*  
in-car networks



*ProgSpy2010*  
in-car networks



*ReEx*  
in-car networks



*Cardiogram*  
in-car networks



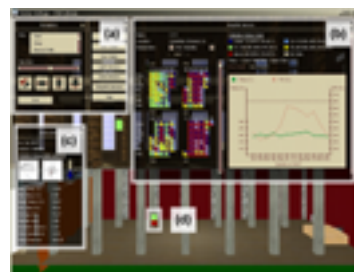
*AutobahnVis*  
in-car networks



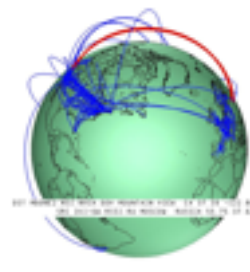
*VisTra*  
in-car networks



*Constellation*  
linguistics



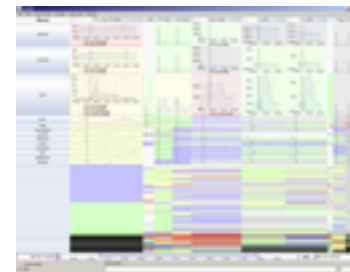
*LibVis*  
cultural heritage



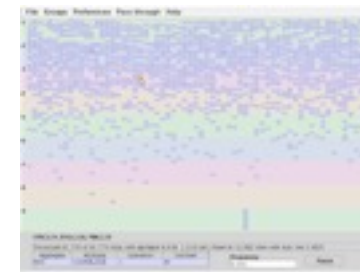
*Caidants*  
multicast



*SessionViewer*  
web log analysis



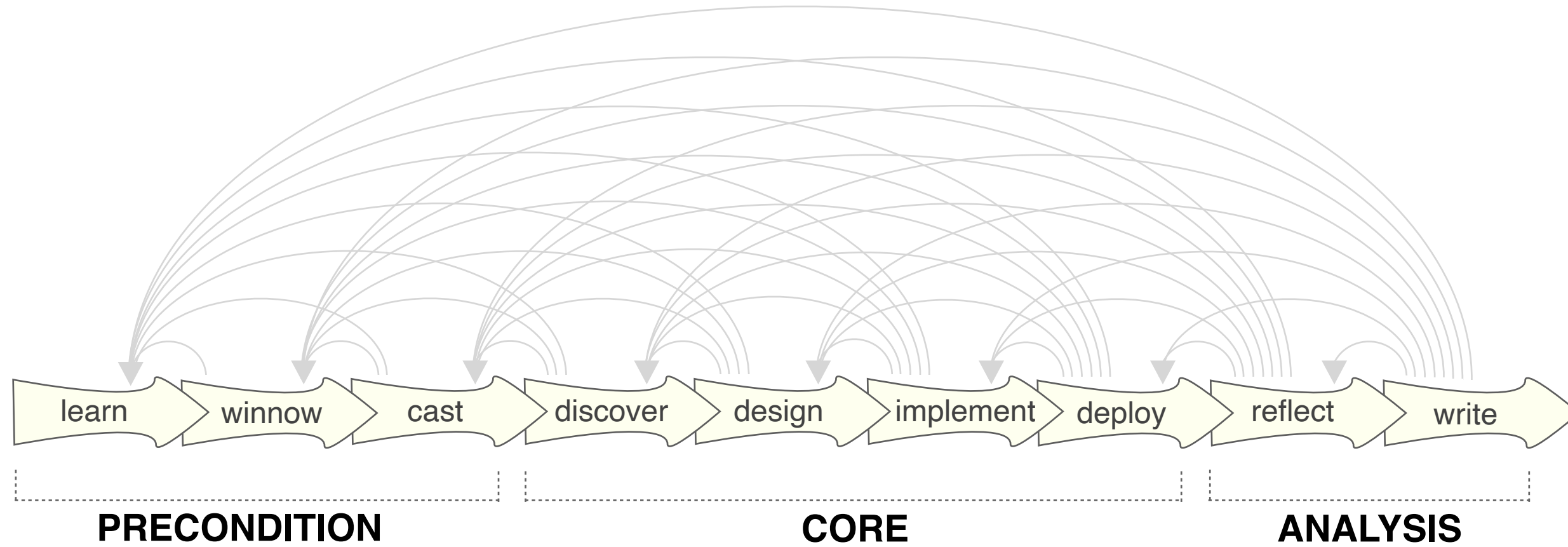
*LiveRAC*  
server hosting



*PowerSetViewer*  
data mining

# 9-stage framework

***iterative***



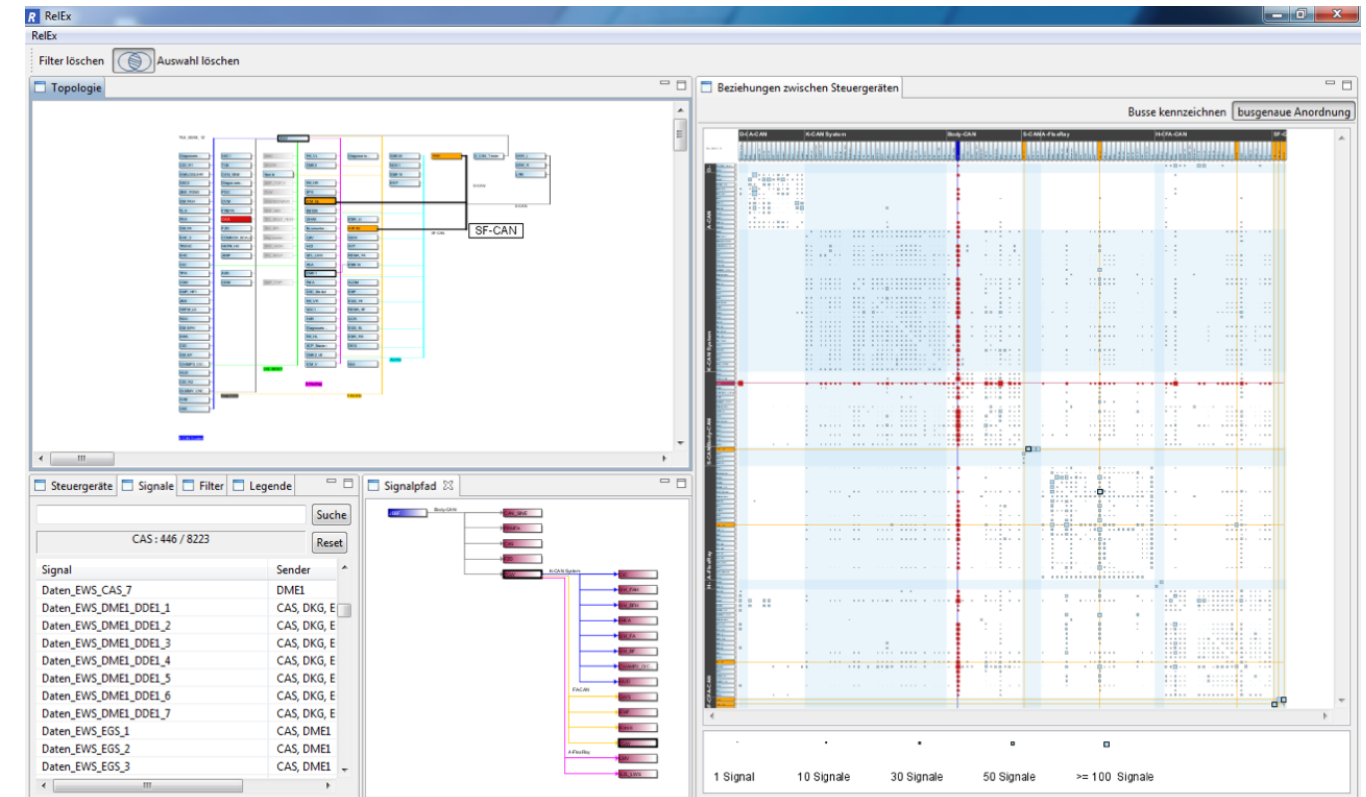
# RelEx

## Visualization for Actively Changing Overlay Network Specifications

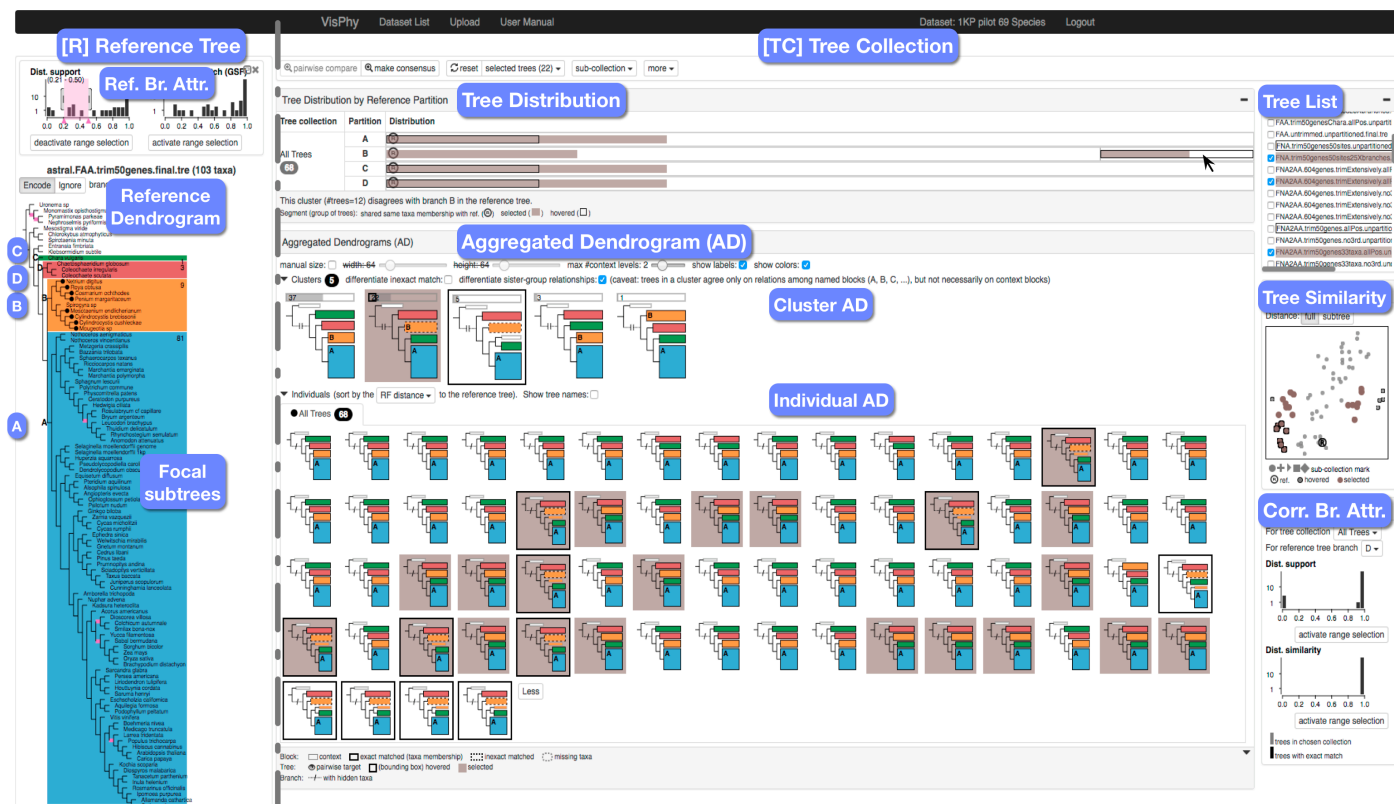
### joint work with:

Michael Sedlmair, Annika Frank, Andreas Butz

<http://www.cs.ubc.ca/labs/imager/tr/2012/relex/>



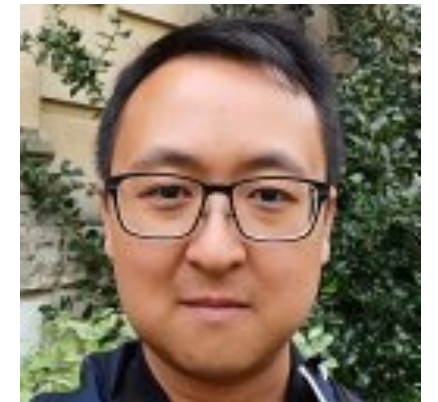
RelEx: Visualization for Actively Changing Overlay Network Specifications.  
Sedlmair, Frank, Butz, Munzner. *IEEE TVCG 18(12): 2729-2738, 2012 (Proc. InfoVis 2012).*



Zipeng Liu



Shing Hei Zhan



# Aggregated Dendrograms

## for Visual Comparison Between Many Phylogenetic Trees

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

Aggregated Dendrograms for Visual Comparison Between Many Phylogenetic Trees.

Liu, Zhan, Munzner. *IEEE Trans. Visualization and Computer Graphics (TVCG)* 26(9):2732-2747, 2019.



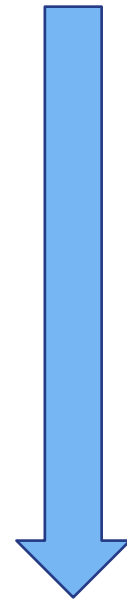
# Phylogenetic tree

Evolutionary relationships of organisms

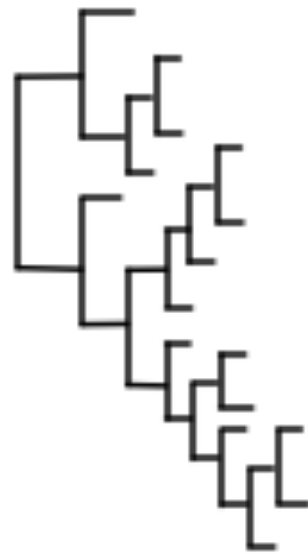
Human	A	T	G	G	A	C	A
Chimpanzee	A	T	G	G	A	C	A
Macaque	A	C	G	G	A	C	A

Genetic information

Computational workflow



Phylogenetic tree



# Many phylogenetic trees

Human  
Chimpanzee  
Macaque

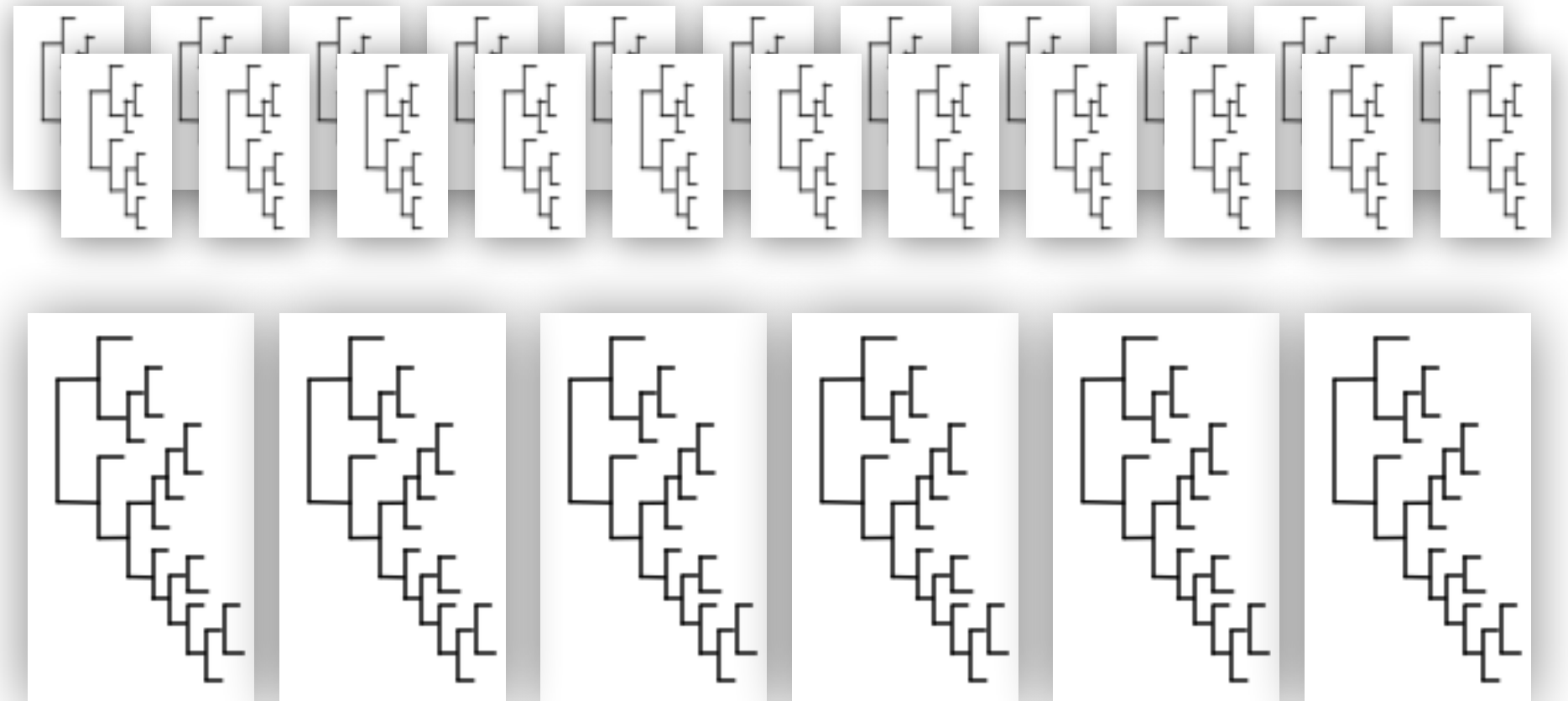
A	T	G	G	A	C	A
A	T	G	G	A	C	A
A	C	G	G	A	C	A

Genetic information

Computational workflow

Phylogenetic tree

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



# Scalability of existing tree comparison systems

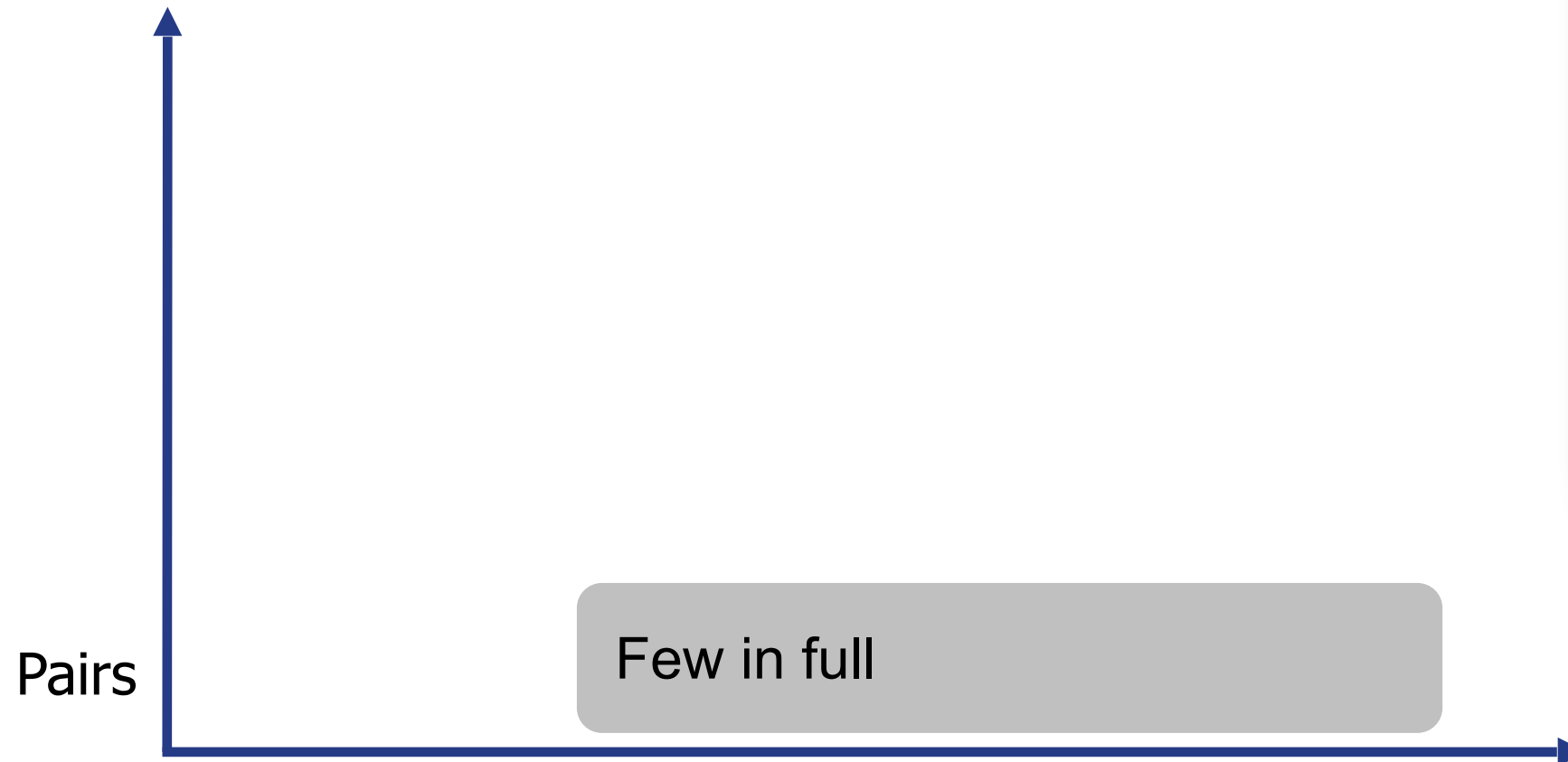
#Trees: how many trees to compare



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

# Scalability of existing tree comparison systems

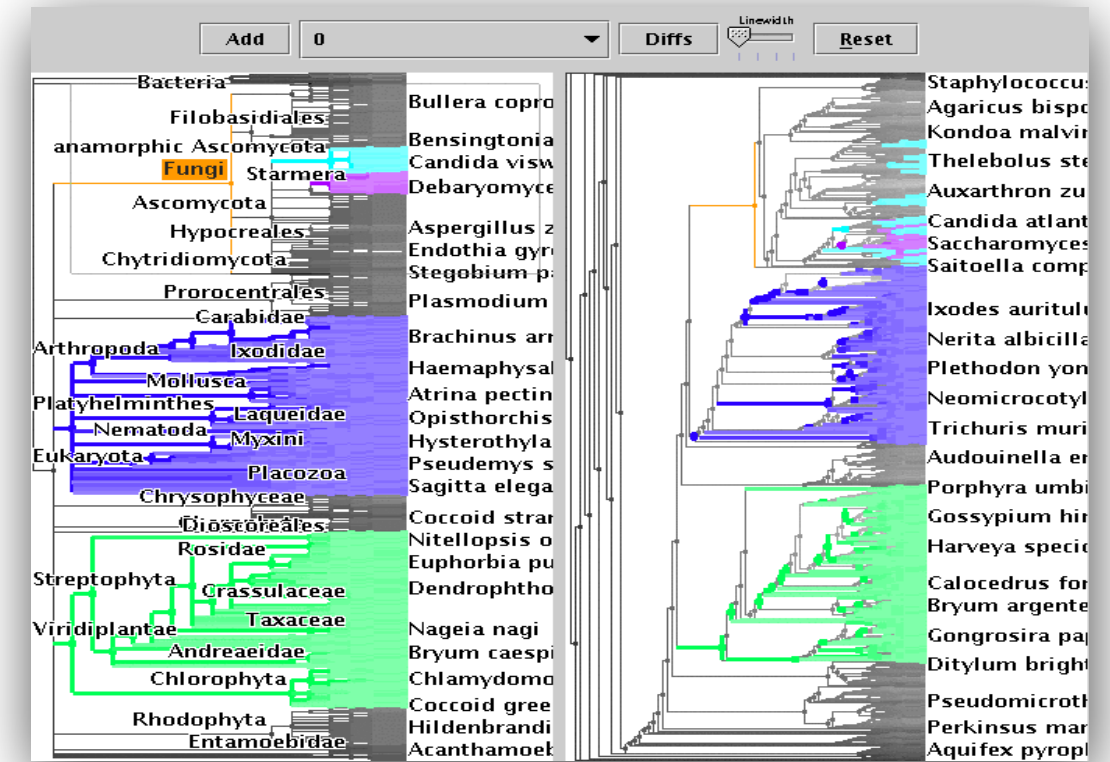
#Trees: how many trees to compare



Simplified structure

Full topology

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

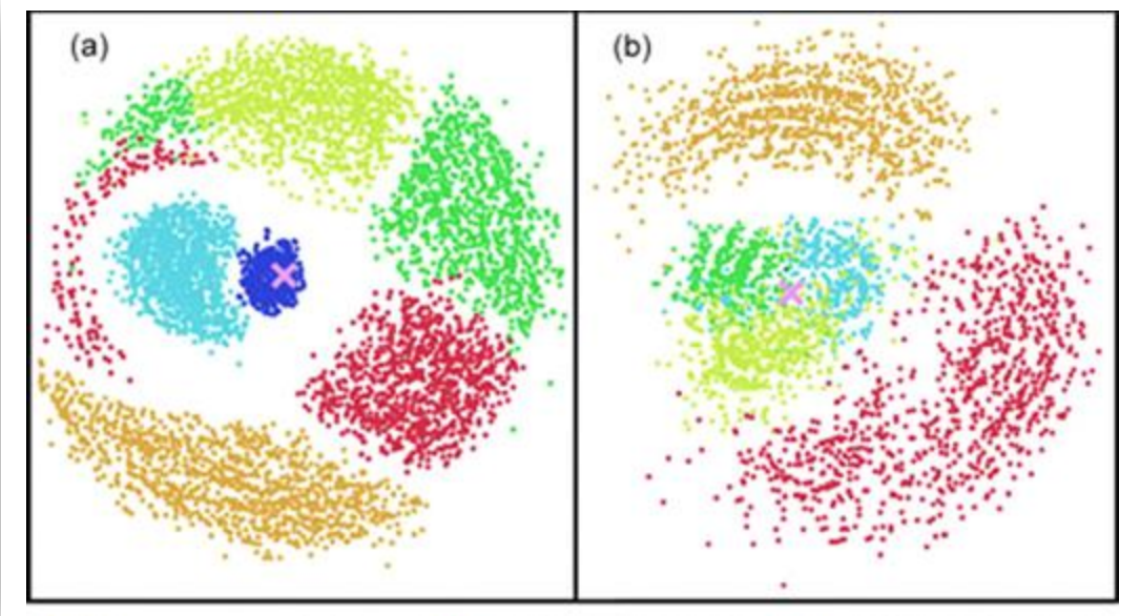
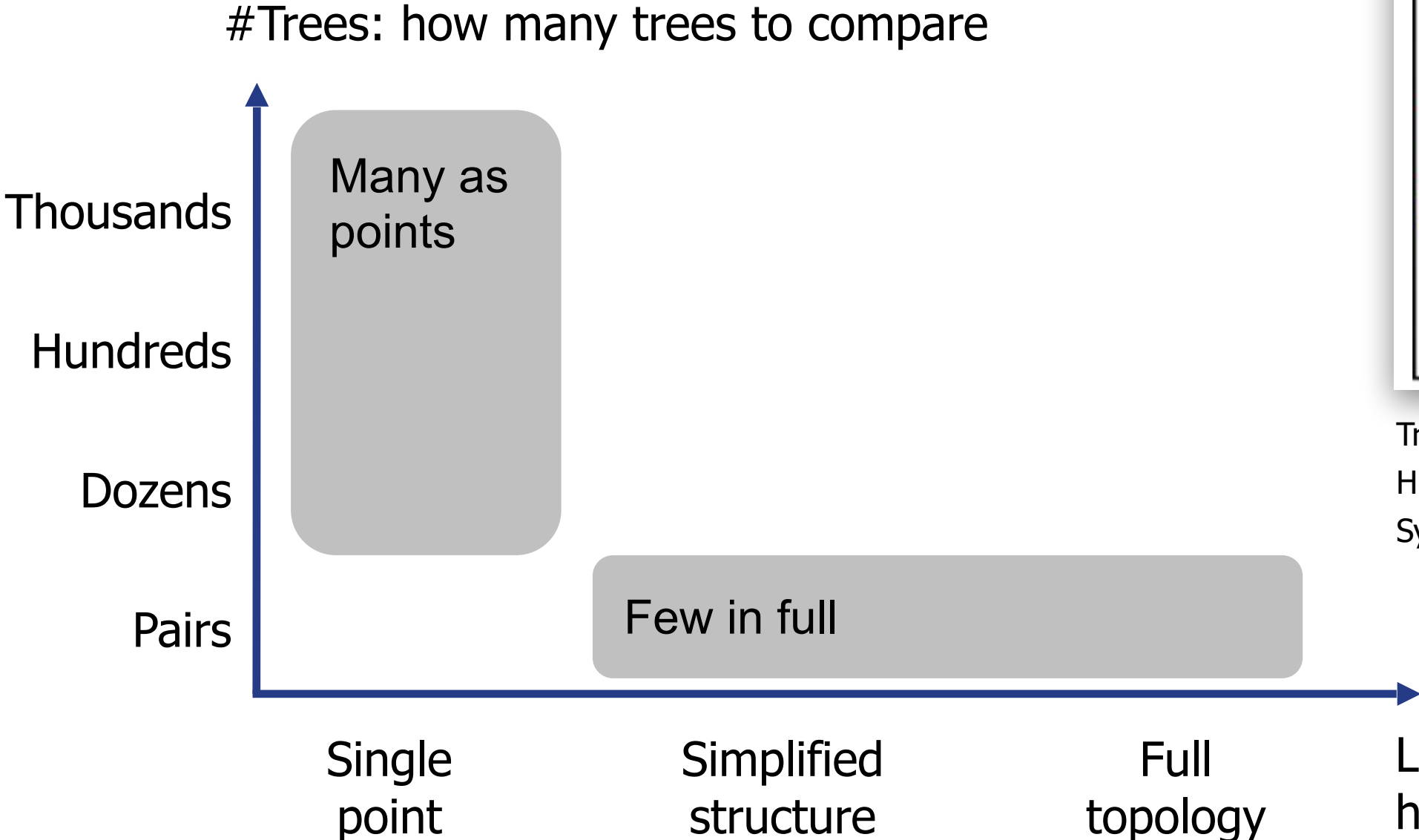


TreeJuxtaposer.

Munzner, Guimbretière, Zhang, Zhou.

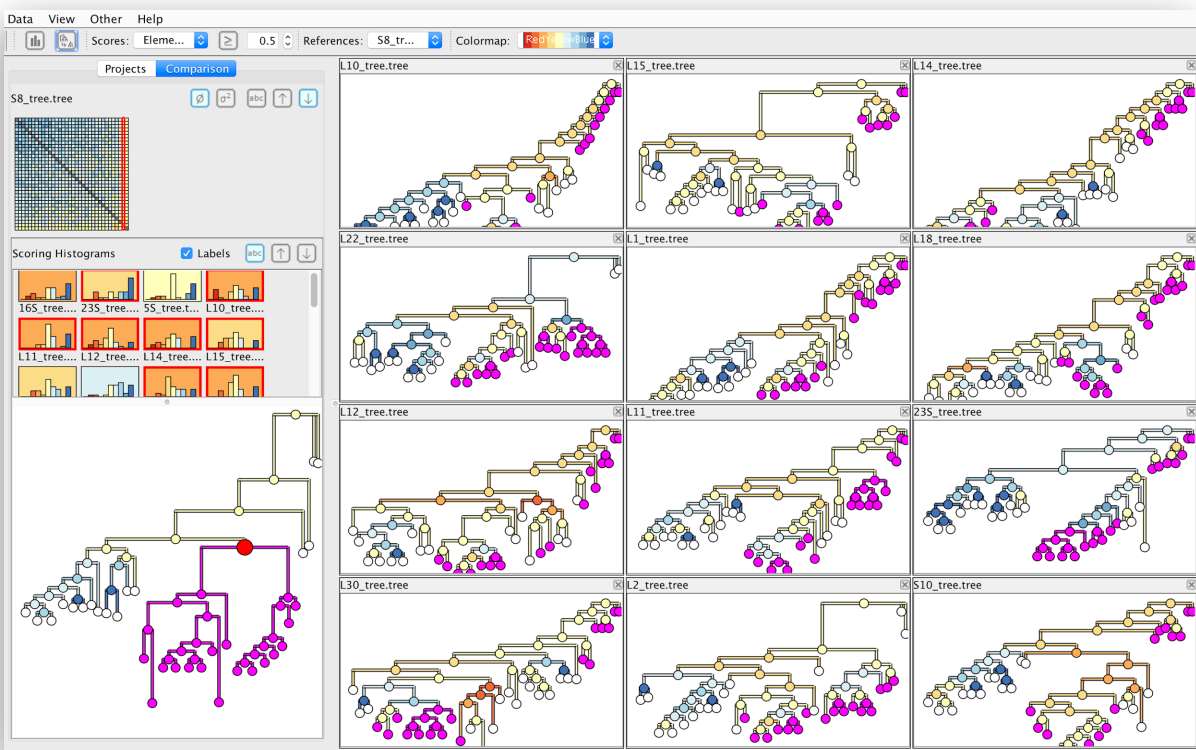
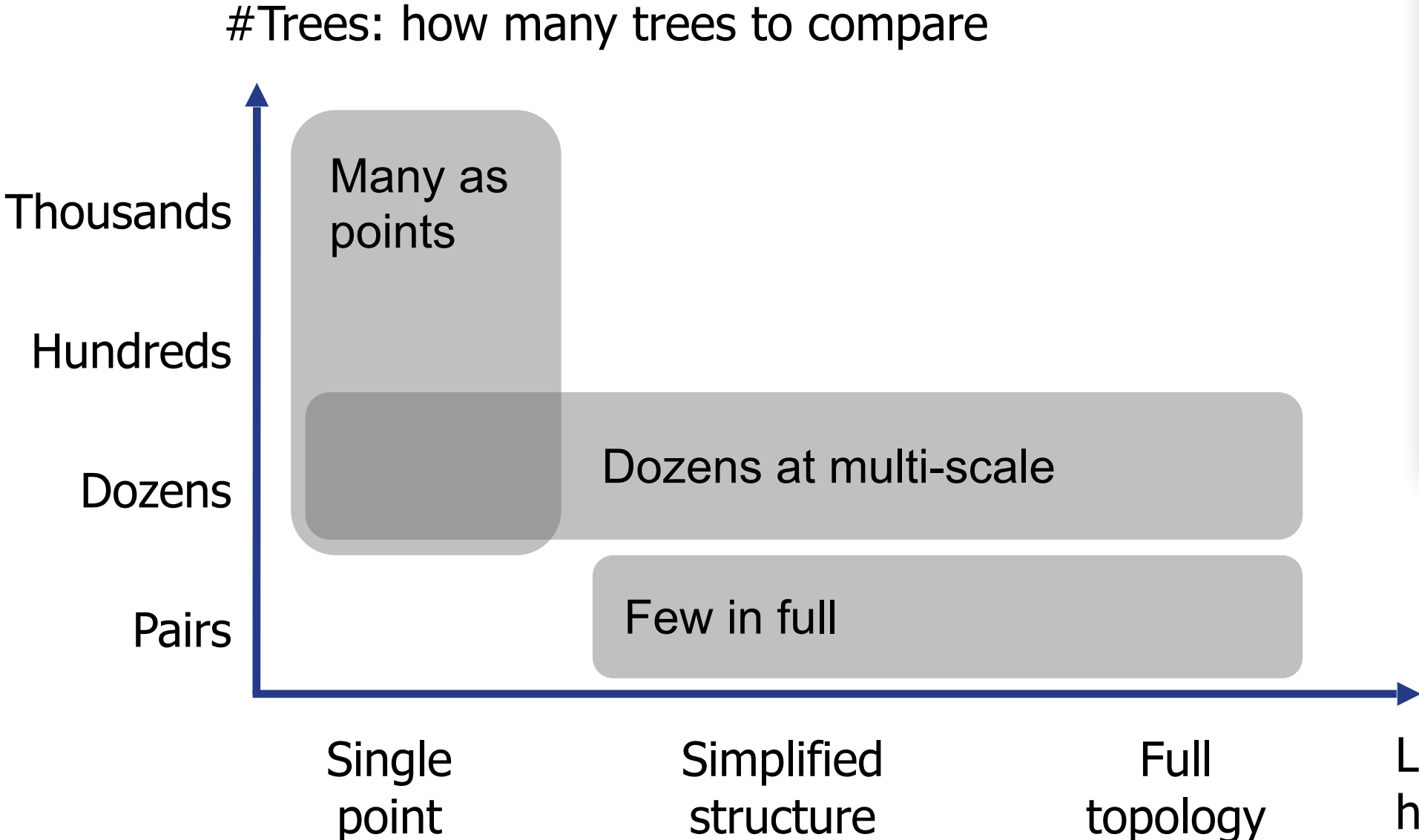
SIGGRAPH 2003

# Scalability of existing tree comparison systems



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

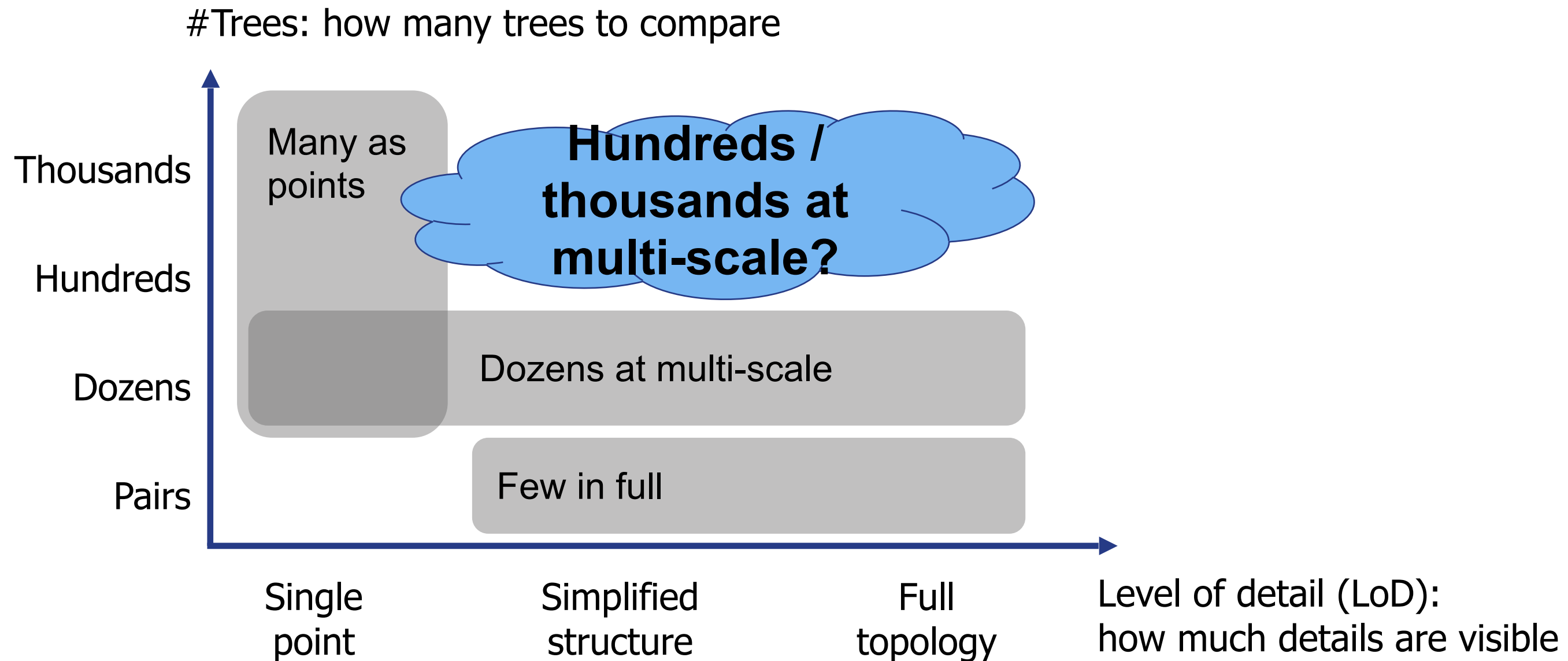
# Scalability of existing tree comparison systems



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

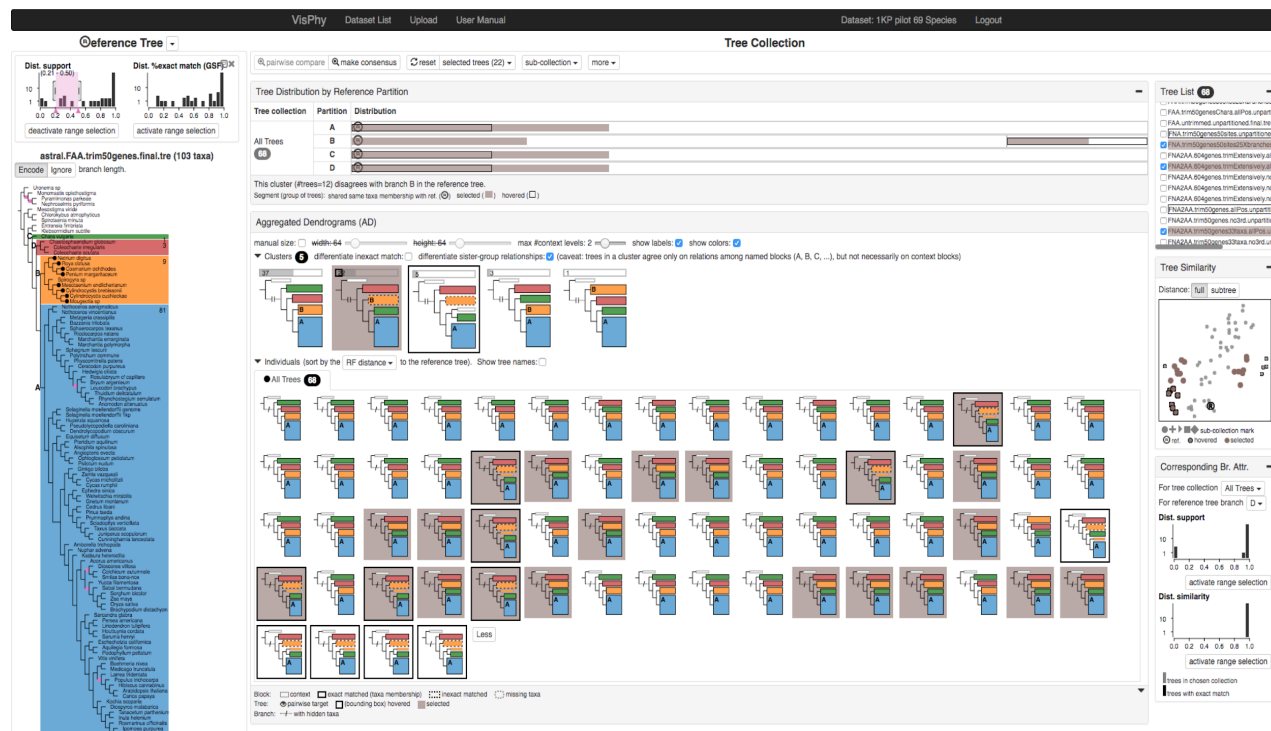
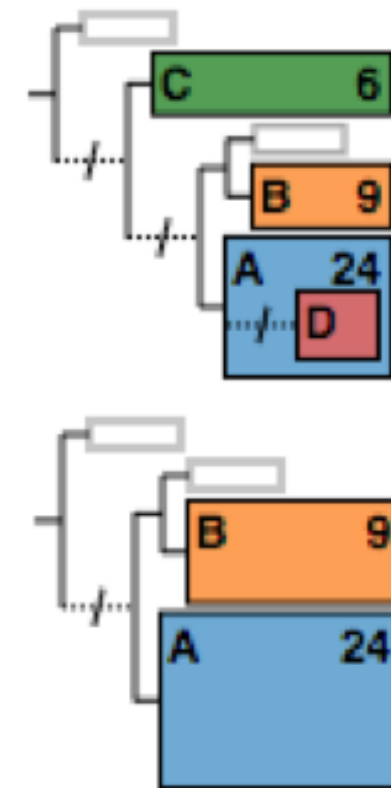
Level of detail (LoD):  
 how much details are visible

# Comparing many phylogenetic trees



# Contributions at abstraction, idiom, & algorithm levels

- data and task **abstractions** for comparison of phylogenetic trees
- new visual encoding **idiom**: Aggregated Dendrogram
  - compact tree representation that focuses on selected subtrees
  - **algorithm** that adapts to available screen space
- interactive multi-view tool: **ADView**
  - covers multiple levels of details for tree comparison



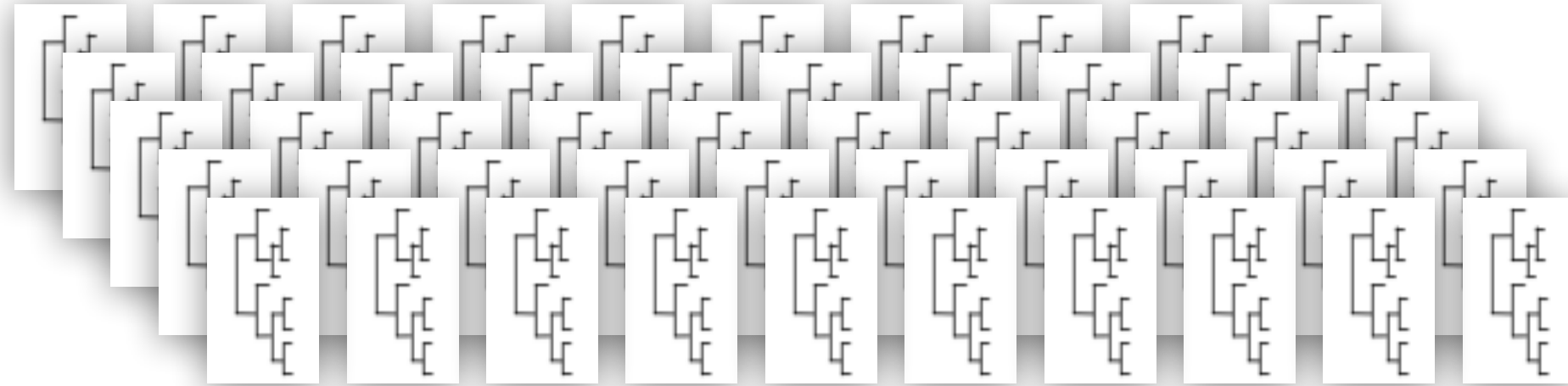
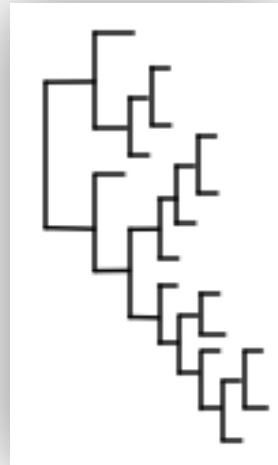


# Data abstraction: Trees

Reference tree

vs.

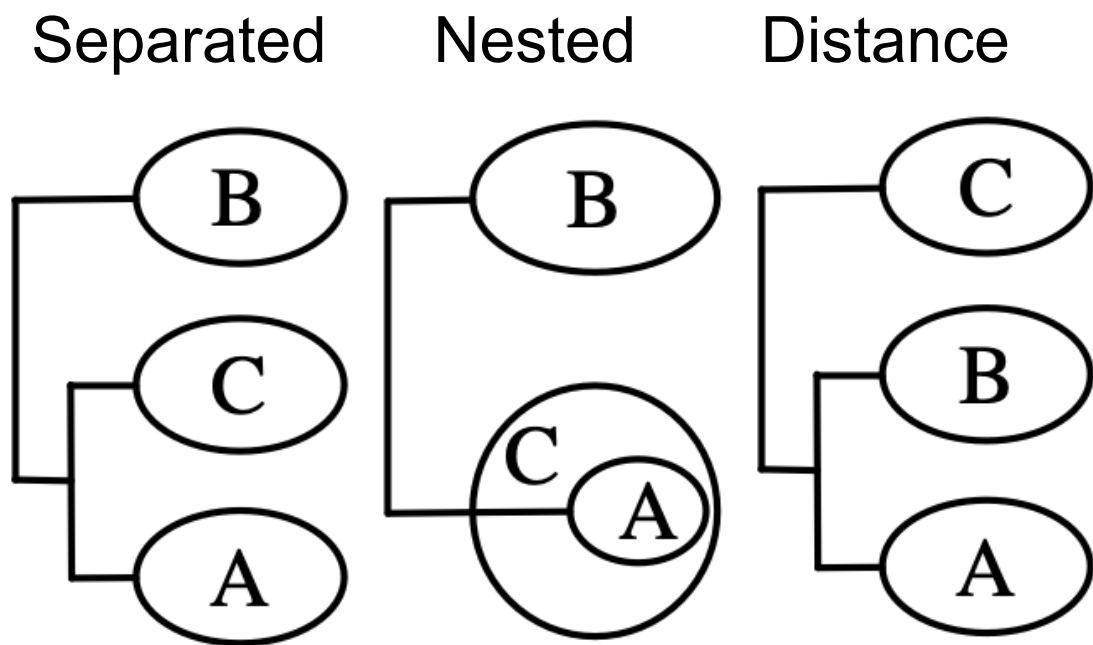
Tree collection



# Task abstraction

# Task abstraction

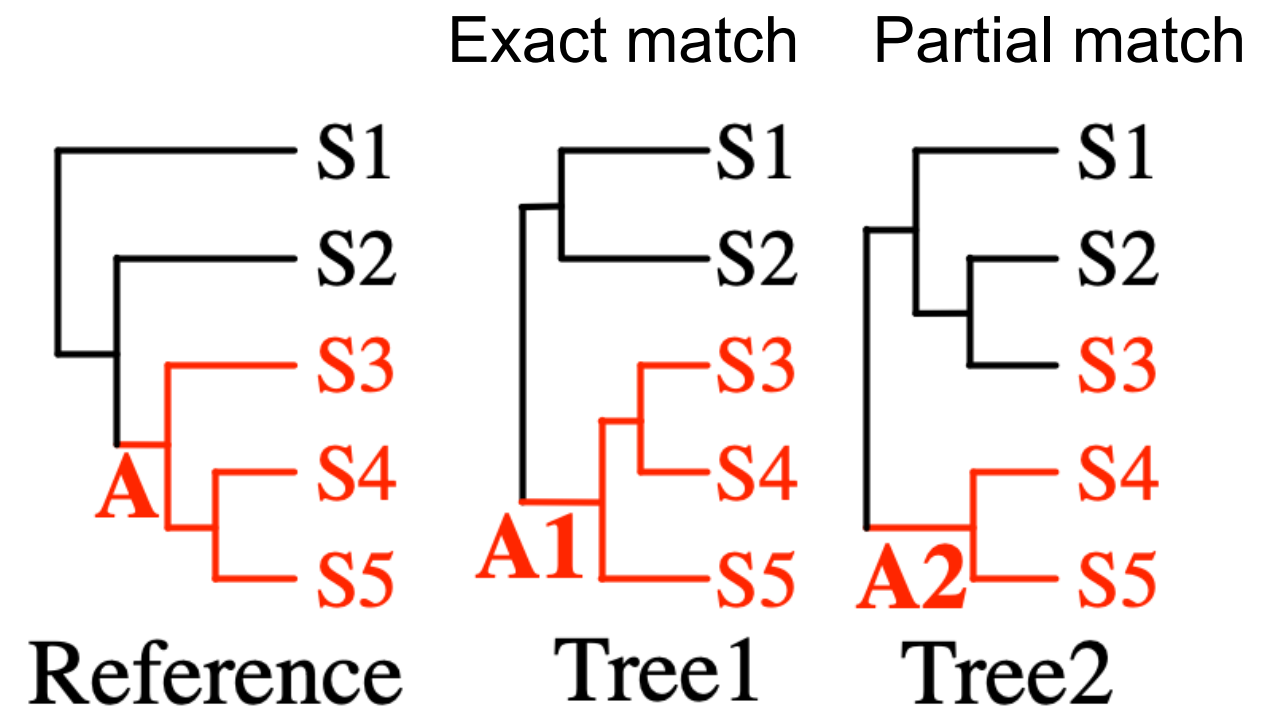
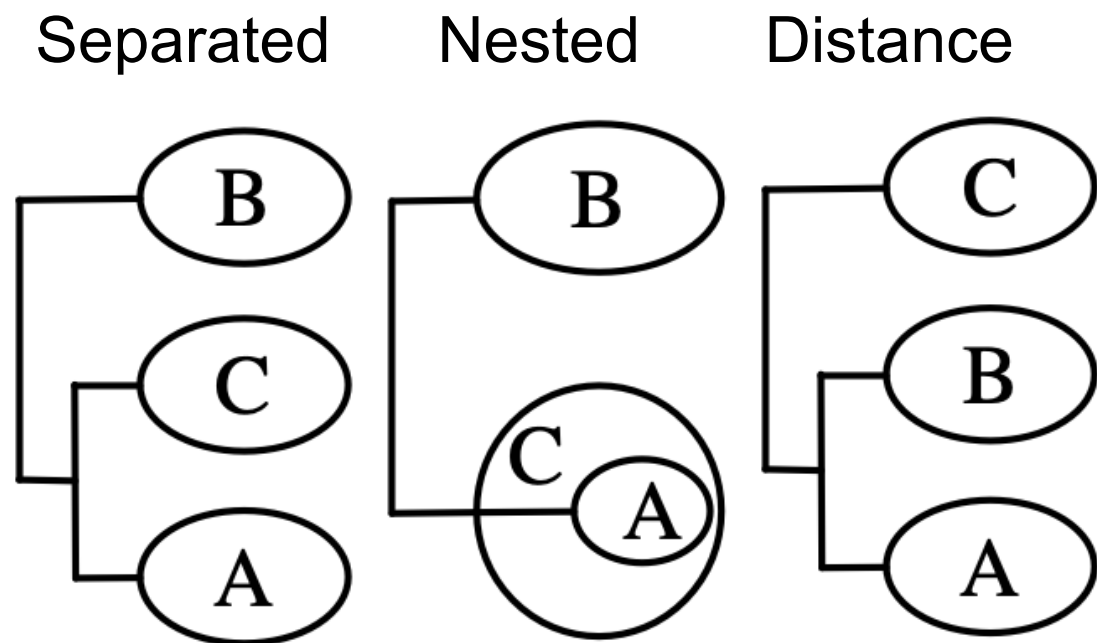
**Topological** relationships & distance  
between subtrees / leaf nodes



# Task abstraction

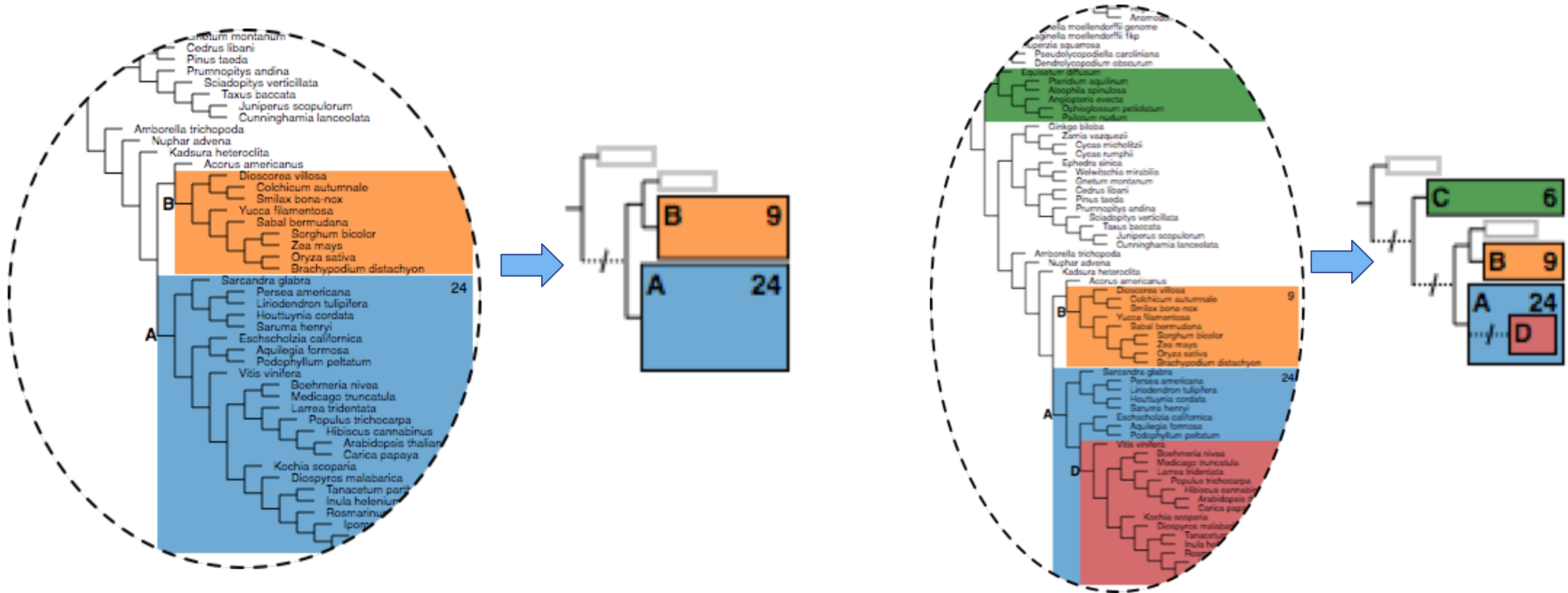
**Topological** relationships & distance between subtrees / leaf nodes

**Leaf** node memberships compared to reference tree



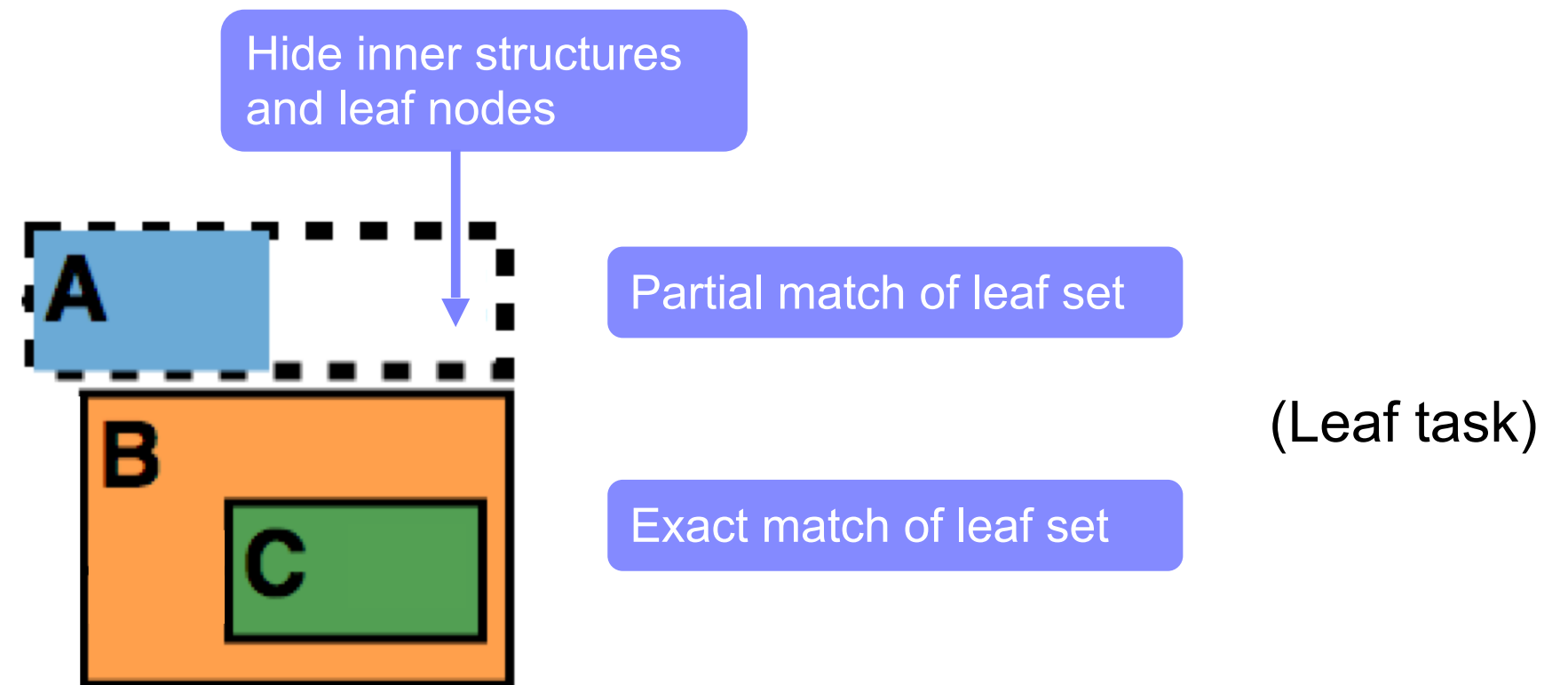
# Aggregated Dendrogram: Intuition

Use glyphs to compress a tree according to user selections



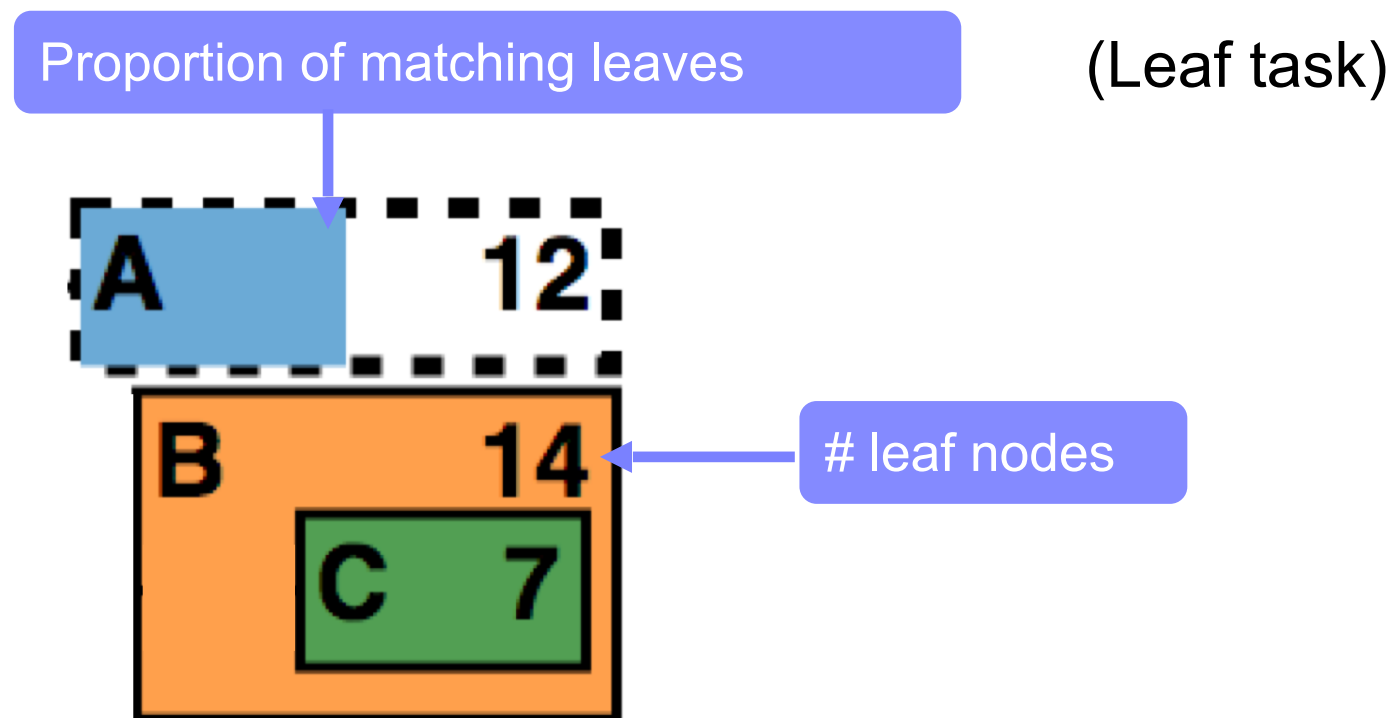
# Visual design: focus + context

- focus
  - selected subtrees



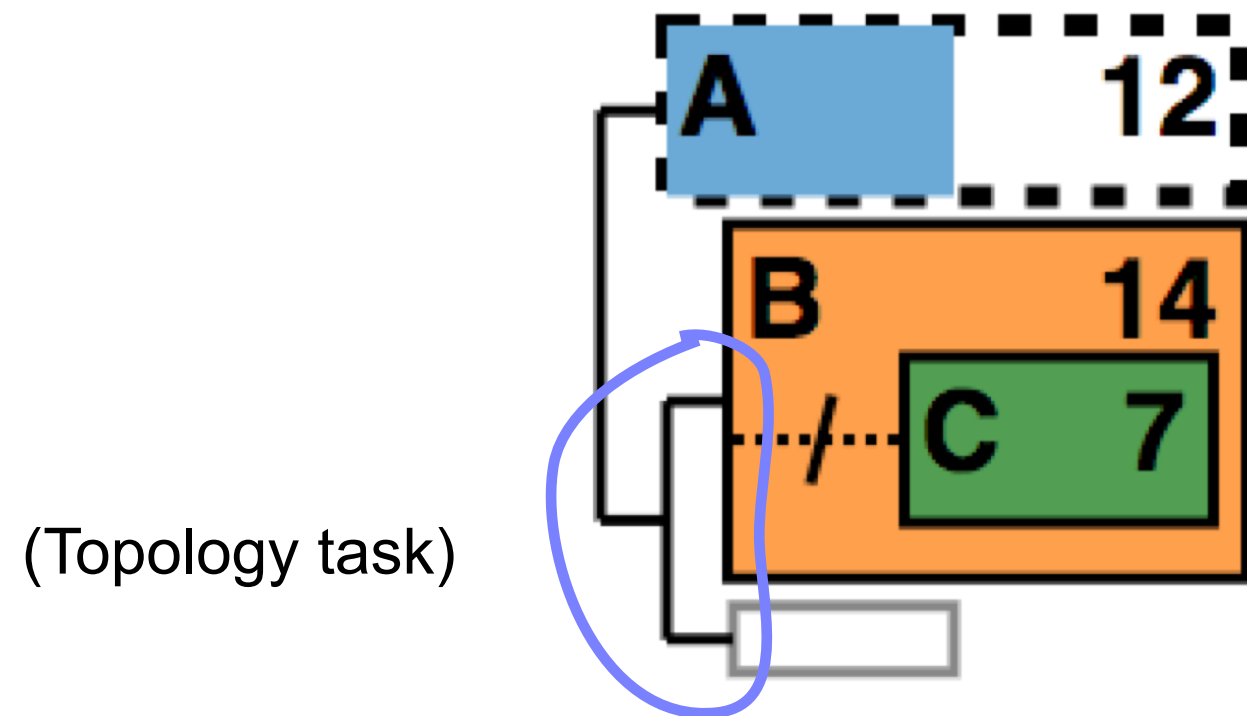
# Visual design: focus + context

- focus
  - selected subtrees



# Visual design: focus + context

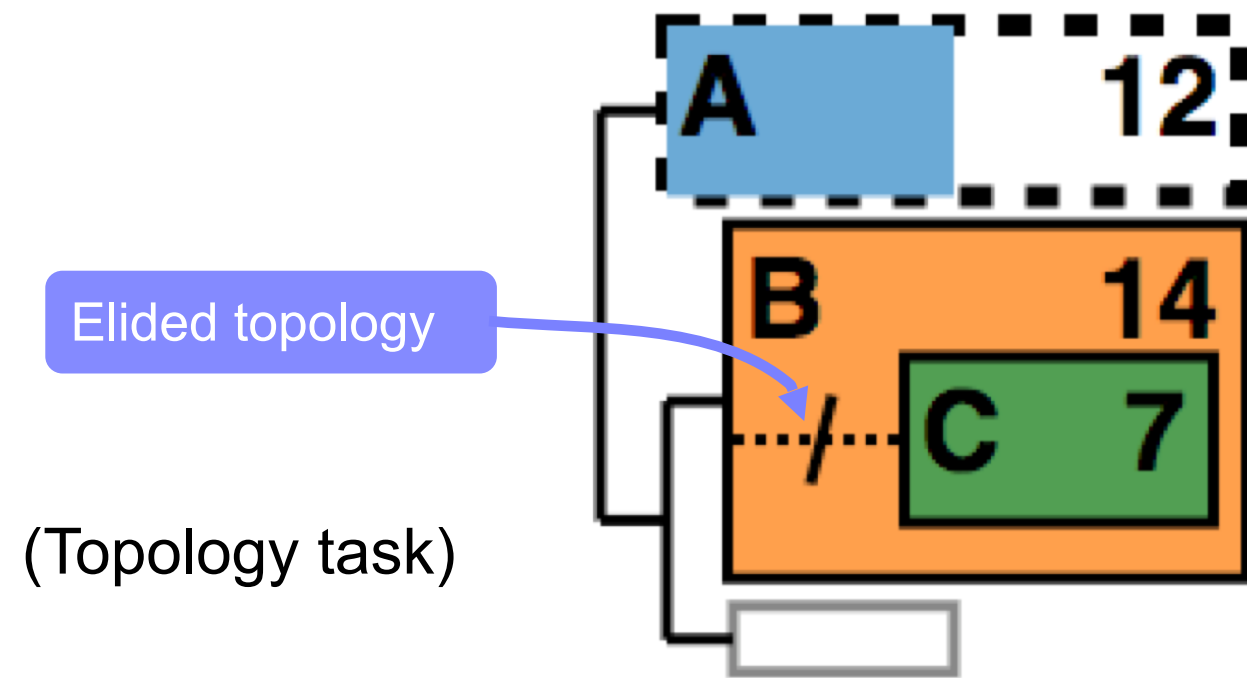
- focus
  - selected subtrees
  - topological relationships between them





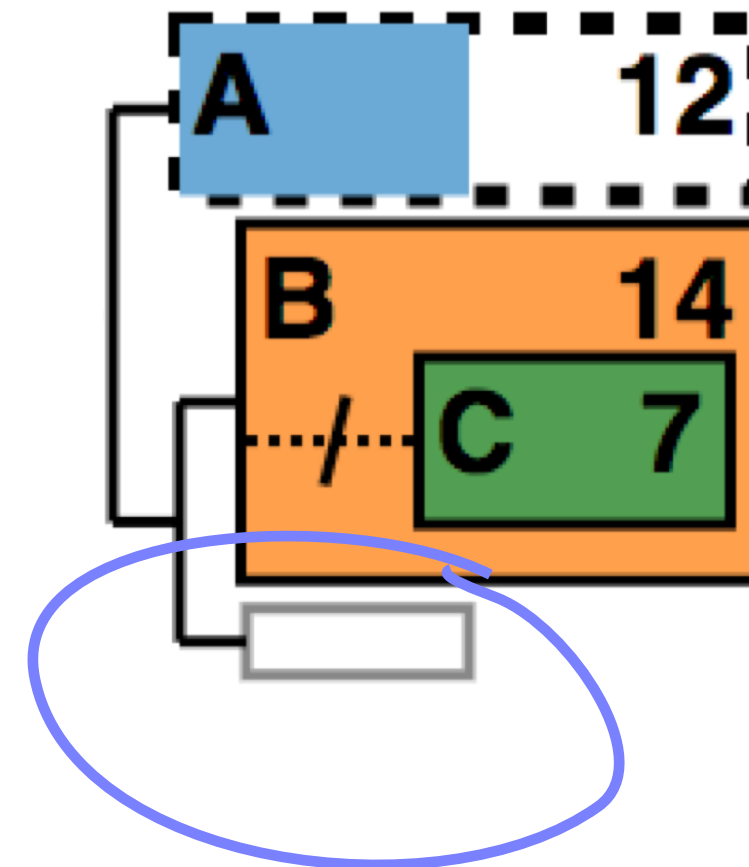
# Visual design: focus + context

- focus
  - selected subtrees
  - topological relationships between them



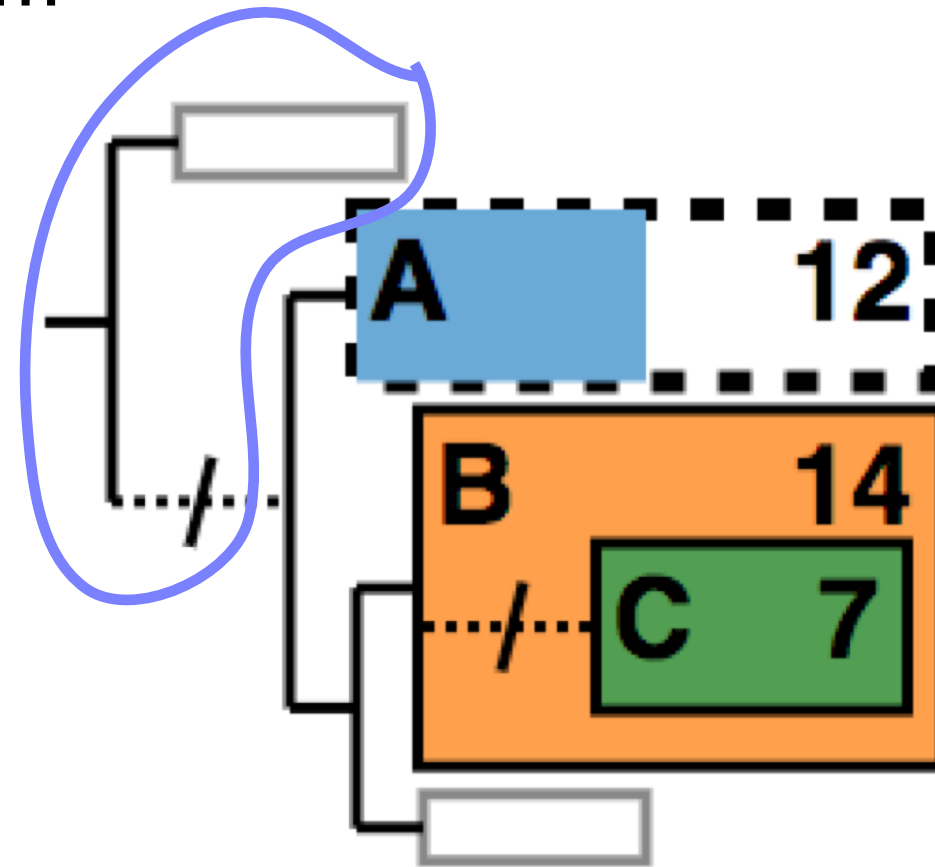
# Visual design: focus + context

- focus
  - selected subtrees
  - topological relationships between them
- context
  - neighboring subtrees



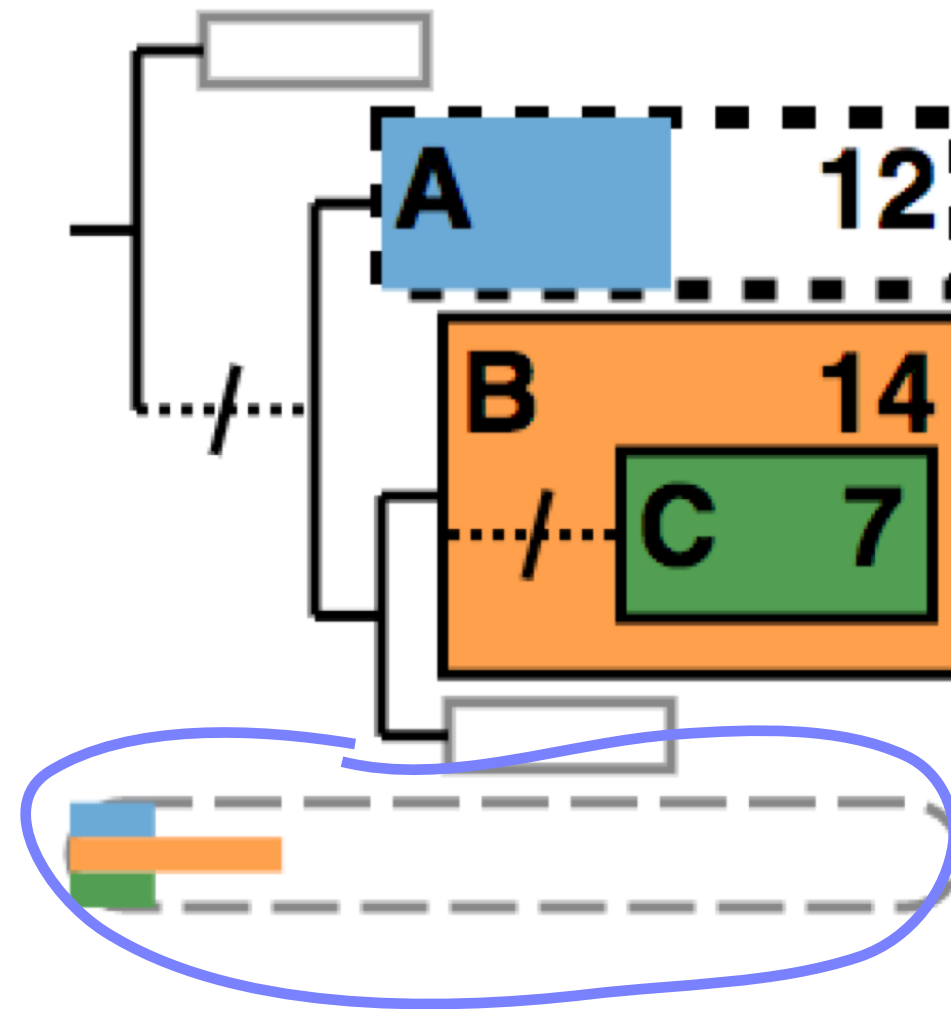
# Visual design: focus + context

- 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



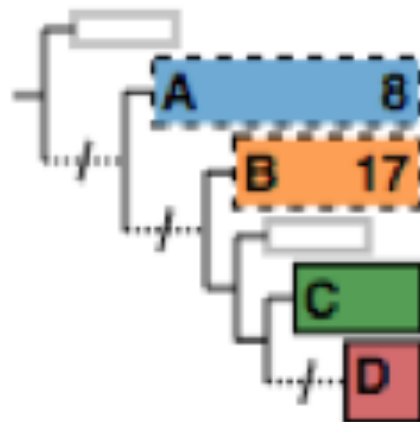
# Visual design: algorithm adapts to space

- show more info when space permitted
  - labels
  - # leaf nodes
  - neighboring blocks

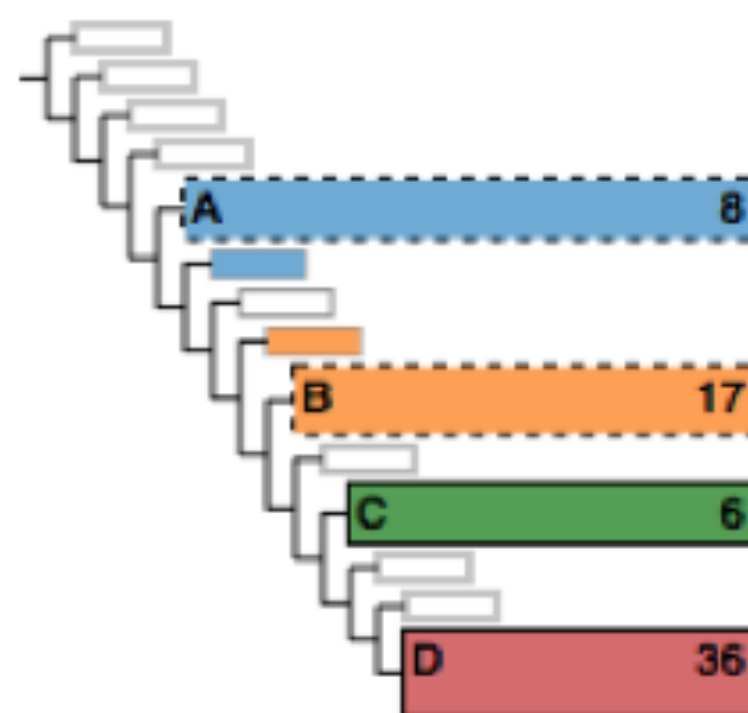
**40x40 px**



**80x80 px**



**160x160 px**



# ADView interface: Multi-level structure across views

VisPhy Dataset List Upload User Manual Dataset: 1KP pilot 69 Species Logout

### Reference Tree

Dist. support (0.21 - 0.50) Dist. %exact match (GSF) X

deactivate range selection activate range selection

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

Encode ignore branch length.

### Tree Collection

pairwise compare make consensus reset selected trees (22) sub-collection more

#### Tree Distribution by Reference Partition

Tree collection	Partition	Distribution
All Trees (68)	A	
	B	
	C	
	D	

This cluster (#trees=12) disagrees with branch B in the reference tree.  
Segment (group of trees): shared same taxa membership with ref. (☉) selected (■) hovered (□)

#### Aggregated Dendrograms (AD)

manual size: width: 64 height: 64 max #context levels: 2 show labels:  show colors:

Clusters: 5 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)

Block:  context  exact matched (taxa membership)  inexact matched  missing taxa  
Tree:  pairwise target  (bounding box) hovered  selected  
Branch:  with hidden taxa

#### Tree List (68)

- FAA.trim50genesChara.allPos.unpartit
- FAA.untrimmed.unpartitioned.final.tre
- FNA.trim50genes50sites.unpartitioned
- FNA.trim50genes50sites25Xbranches
- FNA2AA.604genes.trimExtensively.all
- FNA2AA.604genes.trimExtensively.all
- FNA2AA.604genes.trimExtensively.no
- FNA2AA.604genes.trimExtensively.no
- FNA2AA.604genes.trimExtensively.no
- FNA2AA.604genes.trimExtensively.no
- FNA2AA.trim50genes.allPos.unpartit
- FNA2AA.trim50genes.no3rd.unpartit
- FNA2AA.trim50genes33taxa.allPos.un
- FNA2AA.trim50genes33taxa.no3rd.un

#### Tree Similarity

Distance: full subtree

Corresponding Br. Attr.

For tree collection: All Trees

For reference tree branch: D

#### Dist. support

activate range selection

#### Dist. similarity

activate range selection

Legend:   
■ trees in chosen collection   
■ trees with exact match

# Interface walkthrough: reference tree

The image displays the VisPhy web interface for analyzing tree collections. On the left, a 'Reference Tree' is shown with two histograms: 'Dist. support (0.21 - 0.50)' and 'Dist. %exact match (GSF)'. Below these are controls for 'deactivate range selection' and 'activate range selection'. The reference tree is color-coded by partition (A, B, C, D). A blue box highlights a subtree branch and leaf, with a blue callout bubble containing the text 'Individual tree subtree branch and leaf'. The main interface shows a 'Tree Collection' with a 'Tree Distribution by Reference Partition' table and a grid of 'Aggregated Dendrograms (AD)'. The AD grid shows individual trees from the collection, with a legend at the bottom indicating 'Block: context, exact matched (taxa membership), inexact matched, missing taxa' and 'Tree: pairwise target, bounding box, hovered, selected'. On the right, there is a 'Tree List' with 68 entries, a 'Tree Similarity' plot, and 'Corresponding Br. Attr.' histograms for 'Dist. support' and 'Dist. similarity'. The interface also includes navigation links like 'pairwise compare', 'make consensus', and 'reset', along with a 'Dataset: TKP pilot 89 Species' and 'Logout' options.

# Interface walkthrough: individual & cluster ADs

The screenshot displays the VisPhy Tree Collection interface. At the top, navigation links include 'VisPhy', 'Dataset List', 'Upload', and 'User Manual'. The current dataset is '1KP pilot 69 Species'. The main area is titled 'Tree Collection' and features a search bar with 'pairwise compare', 'make consensus', 'reset', 'selected trees (22)', 'sub-collection', and 'more' options.

**Tree Distribution by Reference Partition**

Tree collection	Partition	Distribution
All Trees (68)	A	100%
	B	0%
	C	0%
	D	0%

This cluster (#trees=12) disagrees with branch B in the reference tree.  
Segment (group of trees): shared same taxa membership with ref. (○) selected (■) hovered (□)

**Annotated Dendrograms (AD)**

Manual size: width: 64 height: 64 max #context levels: 2 show labels:  show colors:

Clusters: 5 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 Similarity**

Distance: full subtree

Corresponding Br. Attr.

For tree collection: All Trees

For reference tree branch: D

Dist. support

Dist. similarity

Legend:

- lock:  context  exact matched (taxa membership)  inexact matched  missing taxa
- ree:  pairwise target  (bounding box) hovered  selected
- branch: --- with hidden taxa



# Interface walkthrough: treespace

The screenshot displays the VisPhy web interface for tree collection analysis. At the top, navigation links include 'VisPhy', 'Dataset List', 'Upload', and 'User Manual'. The current dataset is 'TKP pilot 89 Species'.

**Reference Tree:** Shows two histograms: 'Dist. support (0.21 - 0.99)' and 'Dist. %exact match (GSF)'. Below are controls for 'deactivate range selection' and 'activate range selection'. The reference tree is 'astral.FAA.trim50genes.final.tre (103 taxa)'. A legend indicates 'Encode: ignore branch length'.

**Tree Collection:** Features a 'Tree Distribution by Reference Partition' table with columns for 'Tree collection', 'Partition', and 'Distribution'. A note states: 'This cluster (#trees=12) disagrees with branch B in the reference tree. Segment (group of trees): shared same taxa membership with ref. (○), selected (■), hovered (□)'. Below are 'Aggregated Dendrograms (AD)' with controls for 'manual size', 'width', 'height', 'max #context levels', 'show labels', and 'show colors'. A legend explains the dendrogram symbols: 'Block: context, exact matched (taxa membership), inexact matched, missing taxa; Tree: pairwise target, bounding box, hovered, selected; Branch: -/- with hidden taxa'.

**Tree List:** A list of 68 trees with checkboxes. The 'Tree Similarity' plot is highlighted with a blue box and a callout 'Tree collection'. The plot shows 'Distance: full subtree' and includes a legend for 'sub-collection mark'.

**Corresponding Br. Attr.:** Includes dropdowns for 'For tree collection' (All Trees) and 'For reference tree branch' (D). It contains two histograms: 'Dist. support' and 'Dist. similarity', both with 'activate range selection' controls. A legend indicates 'trees in chosen collection' and 'trees with exact match'.

## Validation with many biologists

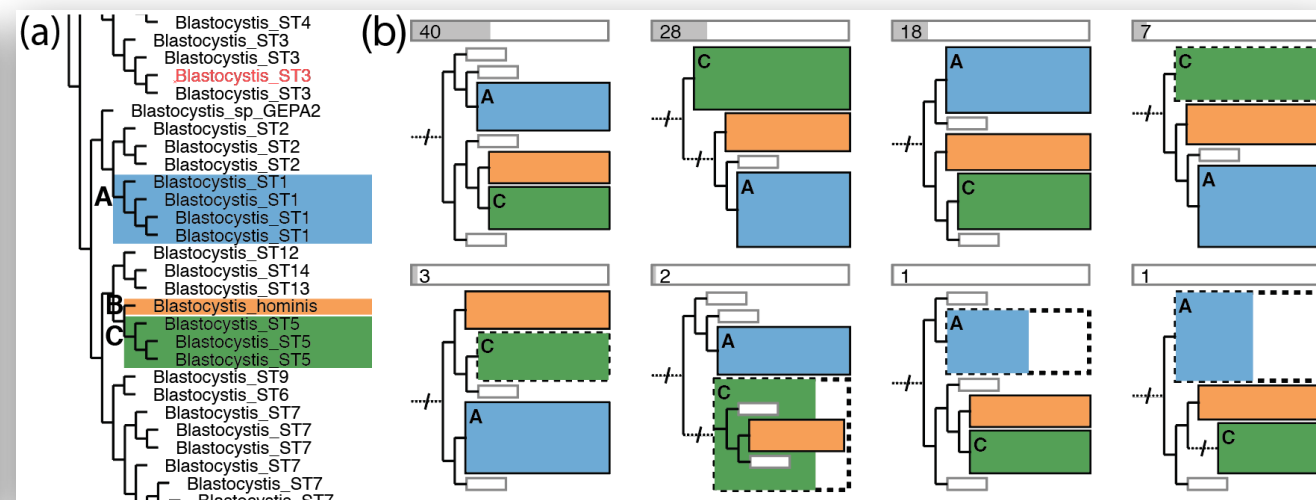
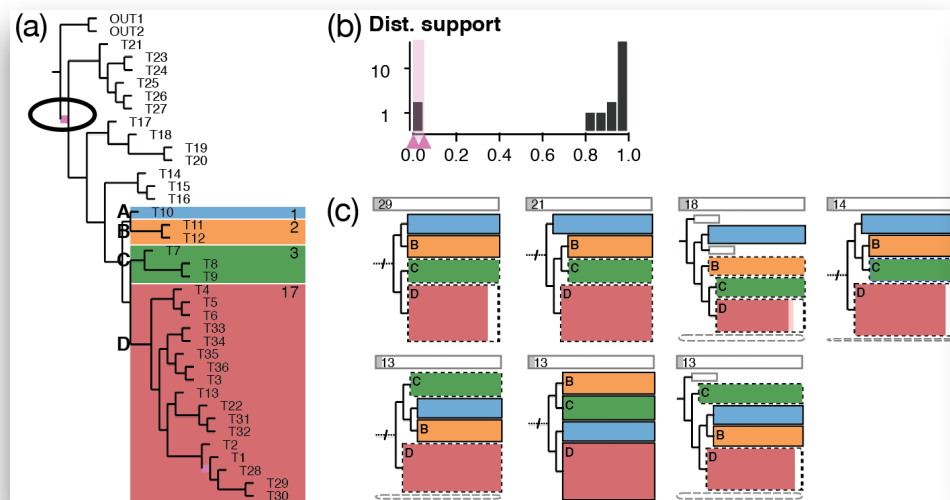
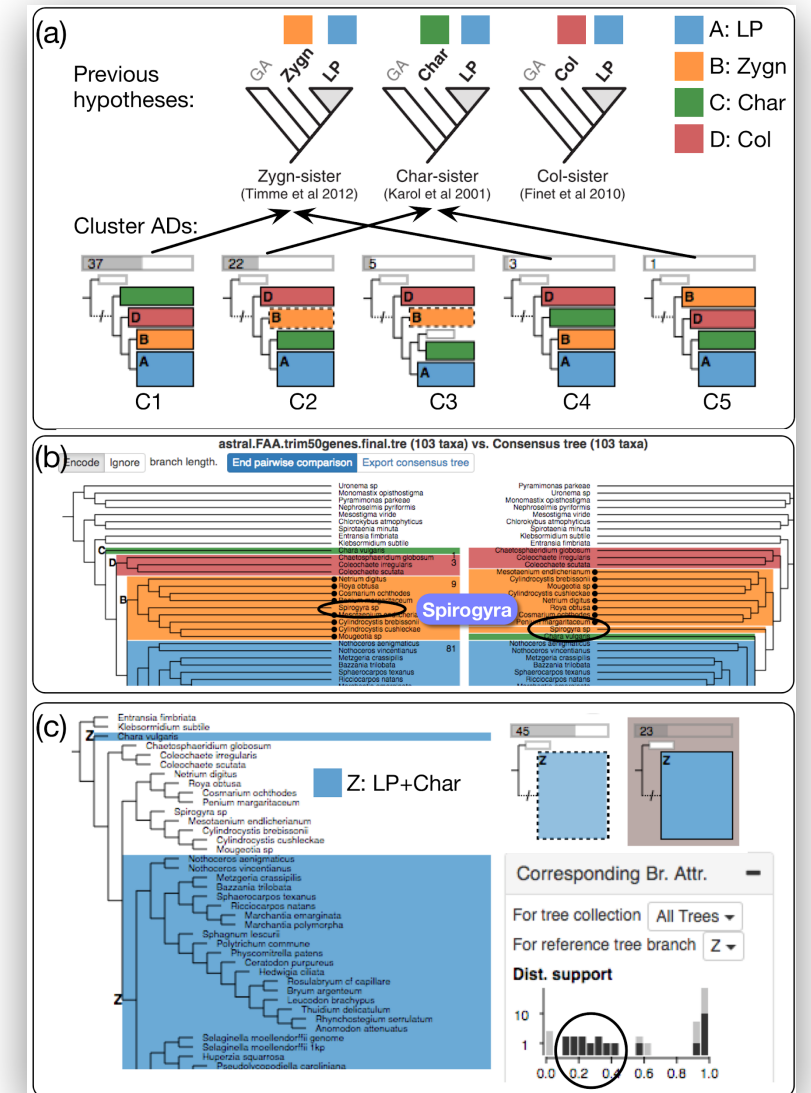
- worked closely with a biology PhD student (second author)

# Validation with many biologists

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

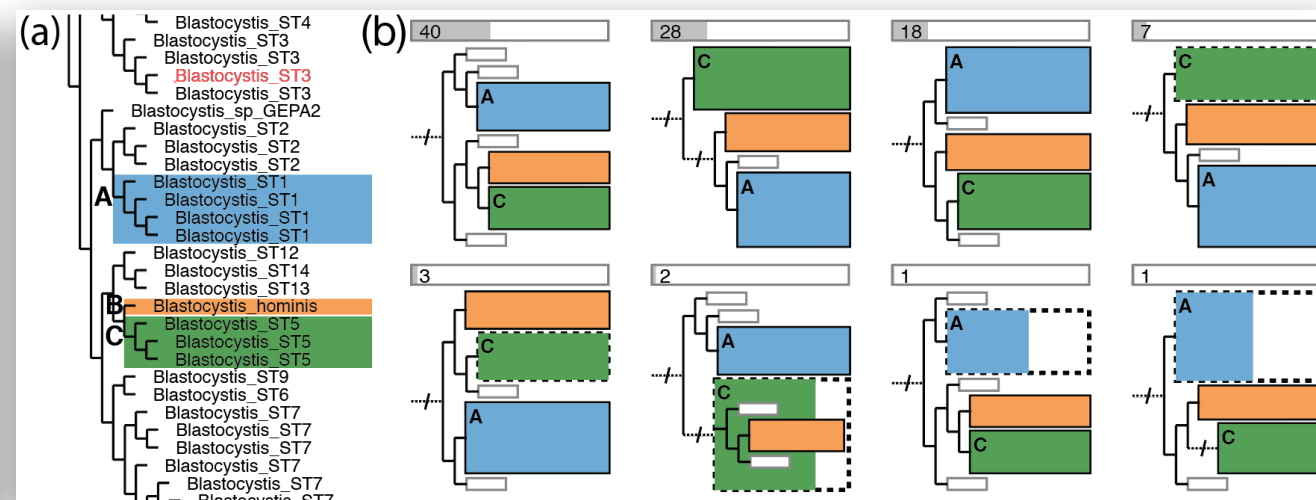
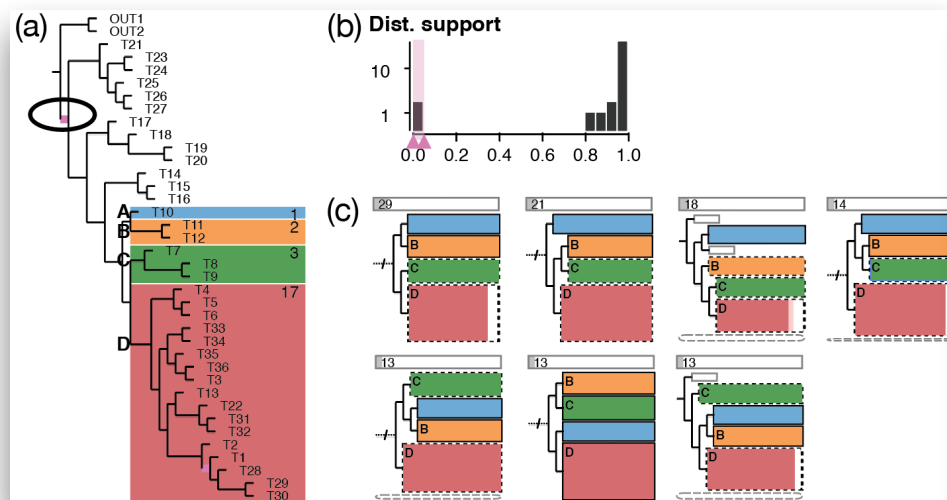
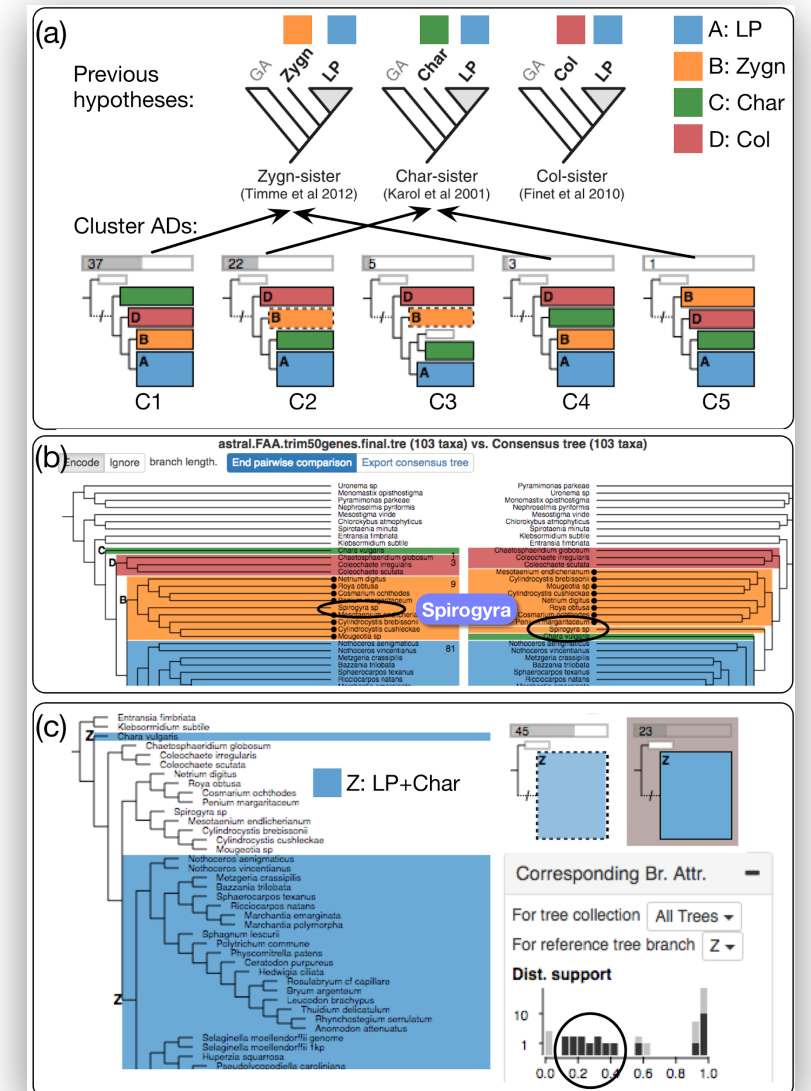
# Validation with many biologists

- worked 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

- worked 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



# Video

VisPhy Dataset List Upload User Manual Dataset: 1KP pilot 69 Species Logout

Reference Tree

support: 100 length: 1 % exact match: 100% label: NA

Dist. %exact match (GSF)

activate range selection

astral.FAA.trim50genes.final.tre (103 taxa) vs. astral.FNA2AA.trim50genes25Xbranches.allPos.final.tre (103 taxa)

Encode ignore branch length. End pairwise comparison

Tree List 424

Tree Similarity

Corresponding Br. Attr.

No branch of interest yet.

Tree Collection

pairwise compare make consensus reset selected trees (1) sub-collection more

Tree Distribution by Reference Partition

Tree collection	Partition	Distribution
All Trees (68)	A	0
	B	0
	C	68
	D	0
	E	0

Segment (group of trees): shared same taxa membership with ref. (0) selected (1) hovered (0)

Aggregated Dendrograms (AD)

manual size: width: 64 height: 64 max #context levels: 2 show labels: show colors:

Clusters 3 differentiate inexact match: differentiate sister-group relationships: (caveat: trees in a cluster agree only on relations among named block necessarily on context blocks)

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

All Trees 68

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

Tree: pairwise target (bounding box) hovered selected

Branch: -- with hidden taxa

Branch: in attribute range hovered for tree similarity

Taxon: taxon hovered uncertain taxon hovered H I 2 3 for partition segments hovered / inspected

Monophyly: highlight 1-5

<https://www.youtube.com/watch?v=2SLcz7KNLjw>

# Problem-driven visualization with design study methodology

- work through all four levels of nested model

- investigate domain

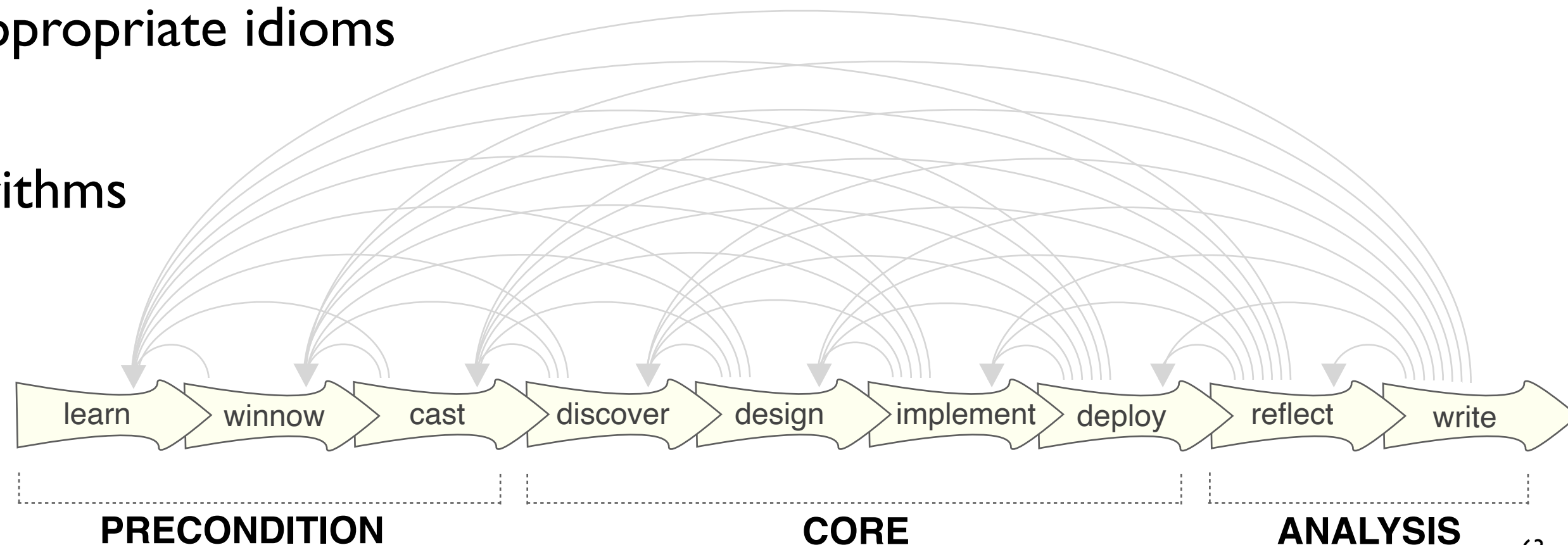
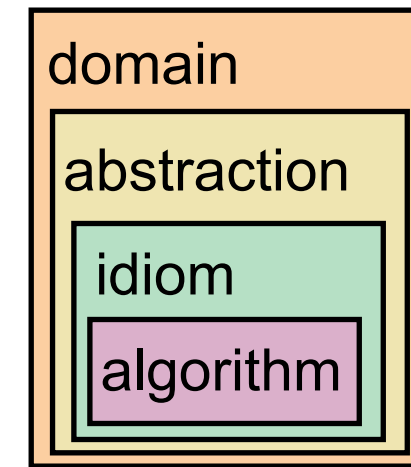
- identify abstractions

- crucial -- & difficult -- iterative process

- select or create appropriate idioms

- develop new algorithms

- if need be



# More information

- this talk

<http://www.cs.ubc.ca/~tmm/talks.html#amw24>

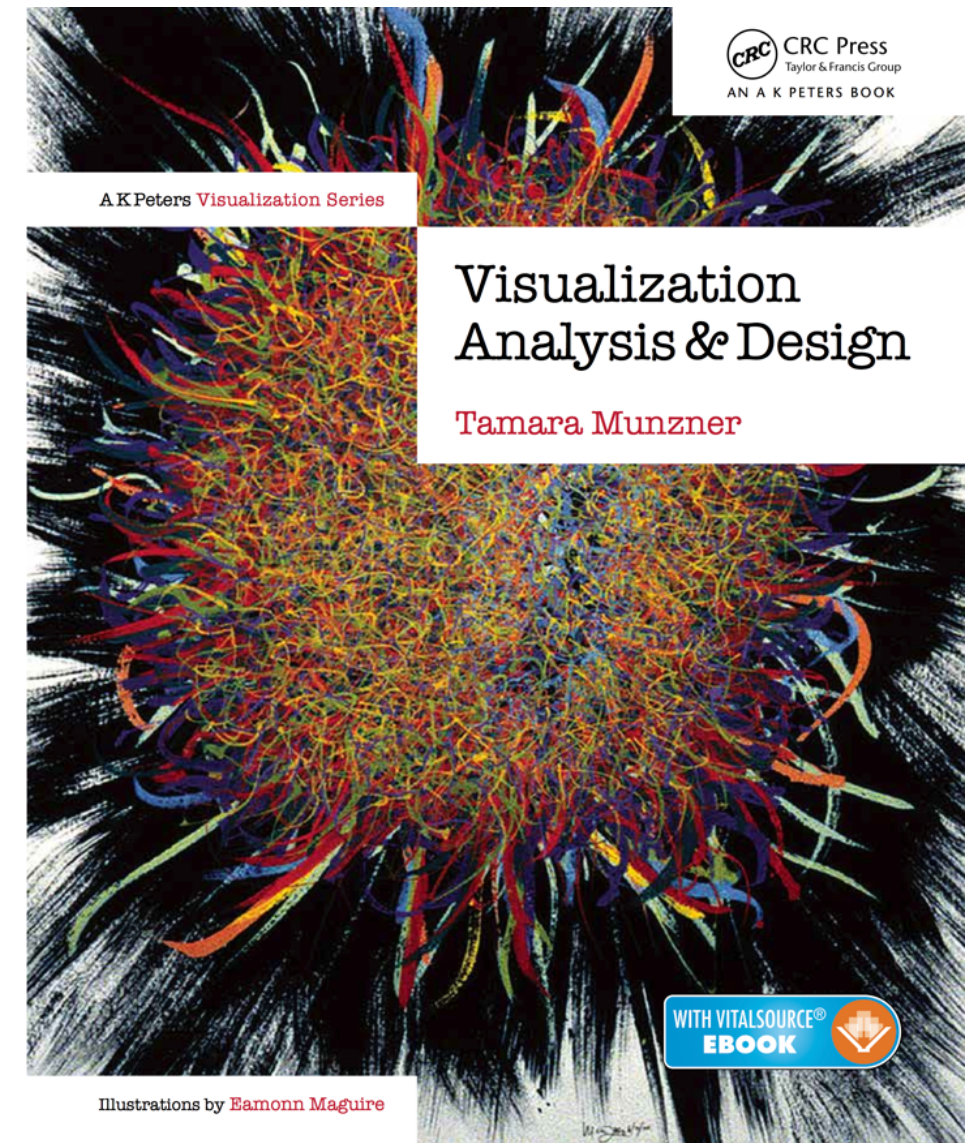
- book

<http://www.cs.ubc.ca/~tmm/vadbook>  
(hardcopy on demo/stuff table)

- full courses, papers, videos, software, talks

<http://www.cs.ubc.ca/group/infovis>

<http://www.cs.ubc.ca/~tmm>



Visualization Analysis and Design. Munzner.  
CRC Press, AK Peters Visualization Series, 2014.

 [@tamara@vis.social](https://medium.com/@tamara@vis.social)

 [@tamaramunzner](https://twitter.com/tamaramunzner)