

Comp/Phys/APSc 715

Bioinformatics Visualization

4/17/2014 Bioinformatics Comp/Phys/APSc: 715 Taylor

Example Videos

- Vis 2013, Schindler
 - [Lagrangian coherent structures in flow](#)
- Matlab bioinformatics toolbox
 - <http://www.mathworks.com/videos/bioinformatics-toolbox-overview-61196.html>

4/17/2014 Bioinformatics Comp/Phys/APSc: 715 Taylor

Administrative

- Presentations next week
 - **Brief** data and goal intro
 - Describe ideal design
 - What perceptual characteristics help user do task?
 - Why parameters chosen (color map, viewpoint)?
 - Consider second-best approach
 - Describe implementation if any (and demo)
 - Evaluation plan or report

4/17/2014 Bioinformatics Comp/Phys/APSc: 715 Taylor

Administrative

- Final Project Turn-in
 - Due 7PM, Tuesday April 29th
 - Written report
 - Described in link from schedule page
 - Example sent out earlier
 - Videos and Paraview State Files
 - Upload to FTP server
 - Or DropBox and tell me where to find
- Demo to me and scientist
 - At or before the final turn-in

4/17/2014 Bioinformatics

Comp/Phys/APS: 715 Taylor

4

Introduction

- Bioinformatics
 - Applying CS algorithms to biological problems
- Examples
 - Protein folding
 - Gene mapping
- Gigantic data sets

4/17/2014 Bioinformatics

Comp/Phys/APS: 715 Taylor

5

What's in this lecture

- IEEE InfoVis special issue on Bioinformatics Visualization
 - 2005, volume 4, no. 3
- Other information from recent pubs/web
- Visualization of:
 - Microarray data (***)
 - Gene sequences
 - Taxonomies
 - Biological pathways

4/17/2014 Bioinformatics

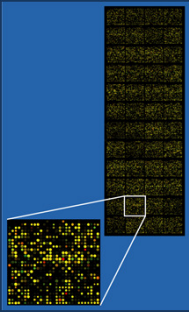
Comp/Phys/APS: 715 Taylor

6

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor

Microarray Data

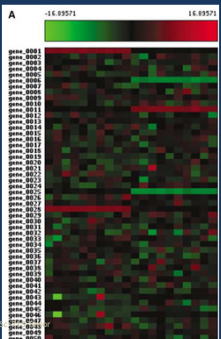
- Warning: IANAB
 - I am not a biologist
- Array of probes (e.g. bits of genes)
- Measure expression level of probes in a sample.
 - relative or absolute
- [Youtube video](#)



4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor

Microarray Data + Score

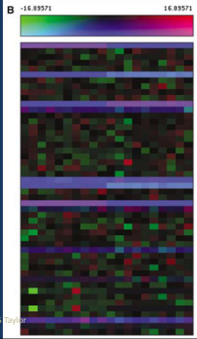
- Gehlenborg et al.
- Default red-black-green map for expression over trial.



4/17/2014 Bioinformatics Comp/Phys/AP

Microarray Data + Score

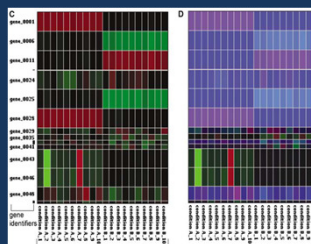
- Gehlenborg et al.
- Default red-black-green map for expression over trial.
- Blue channel for relevance/score
 - Uncertainty vis-ish.



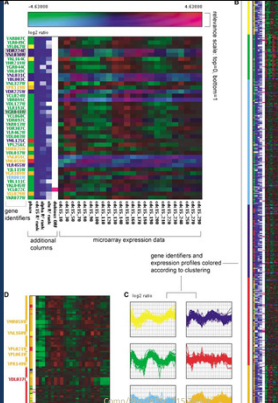
4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor

Microarray Data + Score

- Gehlenborg et al.
- Default red-black-green map for expression vs. condition.
- Blue channel for relevance/score
 - Uncertainty vis-ish.
- Height by gene score.



4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor



A) Extra cols.

B) Overview, color coding for categorization.

C) PC plots

D) Height scaling

4/17/2014 Bioinformatics

Log scaling

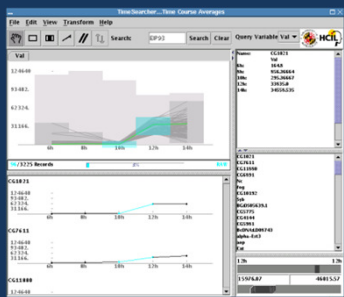
- Most visualizations of microarray data are log-scaled
 - Changes in expression level are smaller for smaller values



Selecting Similar Time Behavior

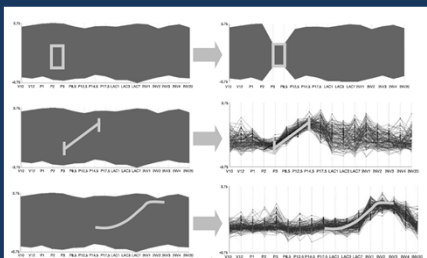
- TimeSearcher

U. Maryland HCI lab



Animated Scatter Plots(1)

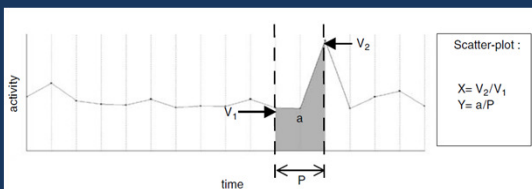
- Parallel Coordinates at one time



4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 16

Animated Scatter Plots(2)

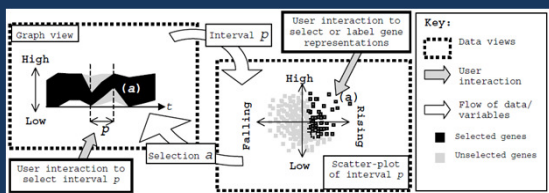
- 2) Pick a time interval
Scatter plot X and Y derived



4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 17

Animated Scatter Plots(3)

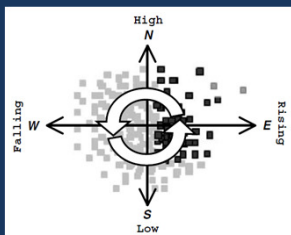
- 3) Compute derivative scatter plot



4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 18

Animated Scatter Plots(4)

4) Animate (move interval)



4/17/2014 Bioinformatics

Comp/Phys/APSc: 715 Taylor

19

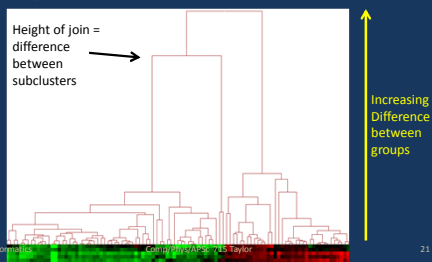


4/17/2014 Bioinformatics

Comp/Phys/APSc: 715 Taylor

Hierarchical Cluster Explorer

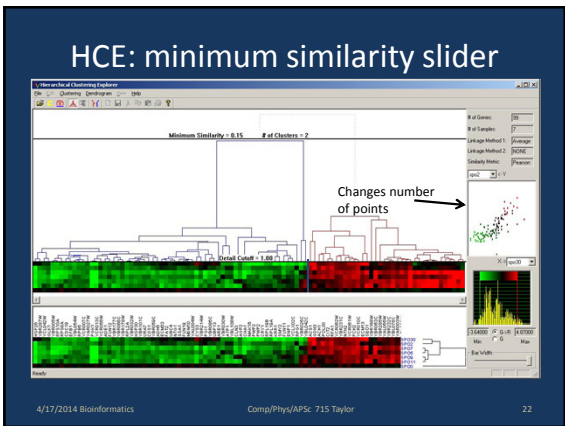
- Seo et al.
 - Find genes that have similar function

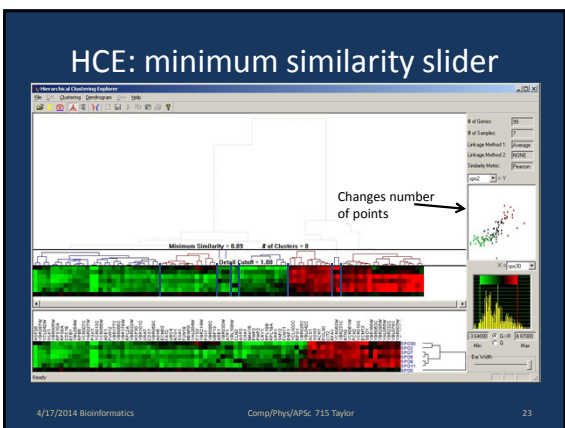


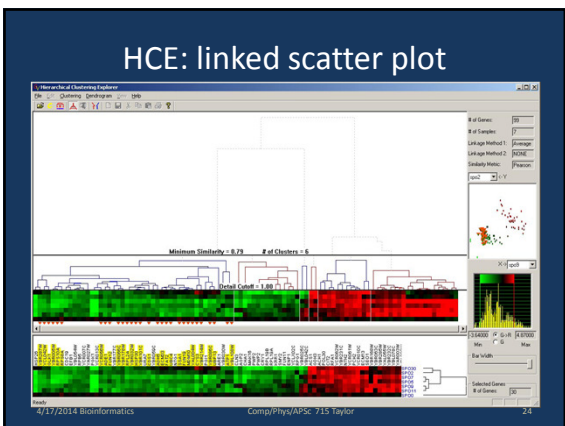
4/17/2014 Bioinformatics

Comp/Phys/APSc: 715 Taylor

21







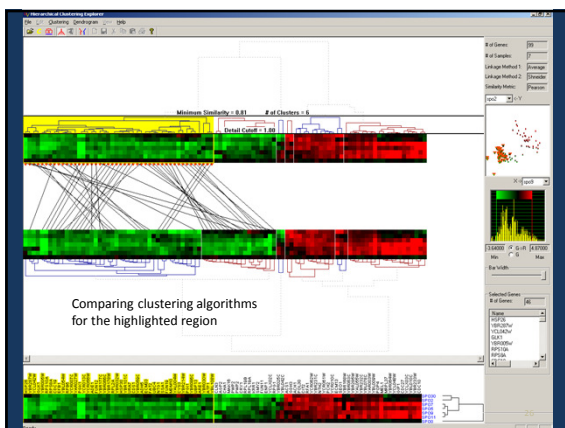
HCE: Detail Cutoff Bar

- How to deal with too much detail?
 - Merge clusters below a size threshold
 - Represent w/ average color

4/17/2014 Bioinformatics

Comp/Phys/APSc: 715 Taylor

25





4/17/2014 Bioinformatics

Comp/Phys/APSc: 715 Taylor

26

aCGH Visualization

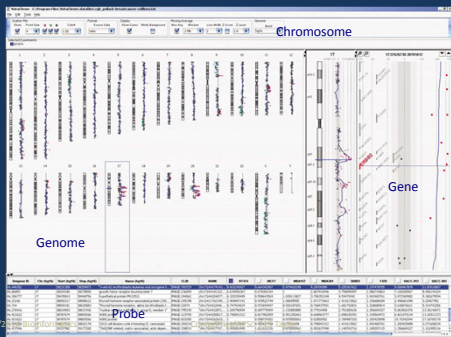
- Array Comparative Genomic Hybridization
- Genome-wide, high resolution copy numbers
- Copy number variation:
 - Segment of DNA with different numbers of copies between genomes.
 - Within patient (two halves of diploid)
 - Between patient (tumor vs. non-tumor)

4/17/2014 Bioinformatics

Comp/Phys/APS: 715 Taylor

28

Visualizing an entire genome

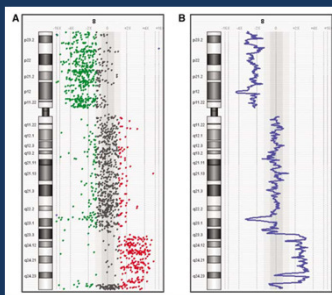


4/17/2014

29

Chromosome View (1)

- Thin = centromeres, variables, cytobands, other
- White = 0-1SD
- Light Gray = 1-2SD
- Dark Gray = 2-3SD
- Dots = samples
 - x=scaled ratio
- Line = windowed moving average



4/17/2014 Bioinformatics

Comp/Phys/APS: 715 Taylor

30

Chromosome View (2)

- Light blue bars are Z scores
 - # SDs from mean
 - ~ # outliers / inliers

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor

Aberration Map

- 17 breast cancer cell lines

Aberration Summary for Chromosome 17
 Window Size: 1 Megabases
 Z Level: 2.0
 Genome: hg16

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor

How do we know if they work?

- Discussion

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 34

Insight User Study

- Count # of “insights” made by users
- Insight:
 - “an individual observation about the data by the participant, a unit of discovery”
- Characteristics:
 - Time, domain value, hypotheses, expectedness, correctness, breadth, category
- Quantification via expert

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 35

Experimental Setup

- 5 Tools
 - Research: Clusterview, TimeSearcher, HCE
 - Commercial: Spotfire, GeneSpring
- 3 Microarray Data sets
 - Timeseries data set—five time-points
 - Virus data set (Categorical)—three viral strains
 - Lupus data set (Multicategorical)—42 healthy, 48 patients
- Participants only used tools they hadn't seen before.

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 36

TABLE 1
Time-Series Data Set Used in the Experiment

GeneName	GenBankId	1.5 Hr	4 hr	6 Hr	8 Hr	12 Hr
aquaporin 4	AA001003	1.54	-0.21	1.49	-0.12	0.96
...

TABLE 2
Viral Data Set Used in the Experiment

Name	Description	wt PR8	NS1 (1-126)	delNS1
ADCY9	adenylate-cyclase-9	0.54	0.91	5.8
...

TABLE 3
Lupus Data Set Used in the Experiment

Accession #	Gene	Ctrl 1	...	Ctrl 42	SLE 1	...	SLE 48
AB008775	Aquaporin 9	-63.7	...	100.1	4418.	...	3433.2
...

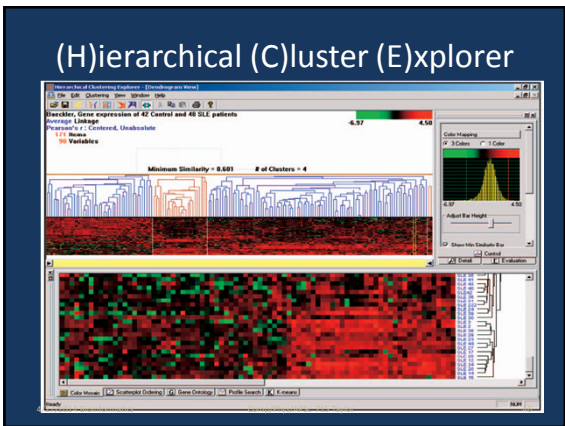
4/17/2014 Bioinformatics 37

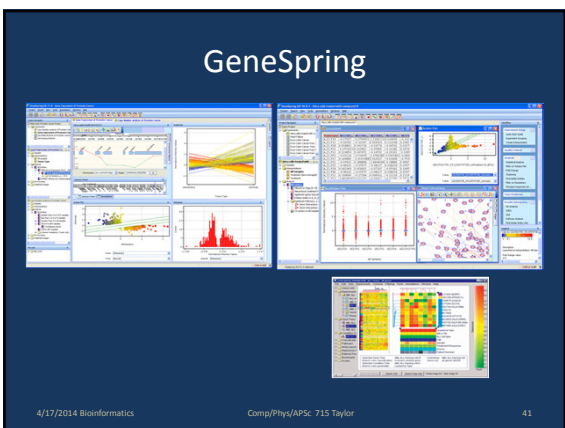
ClusterView

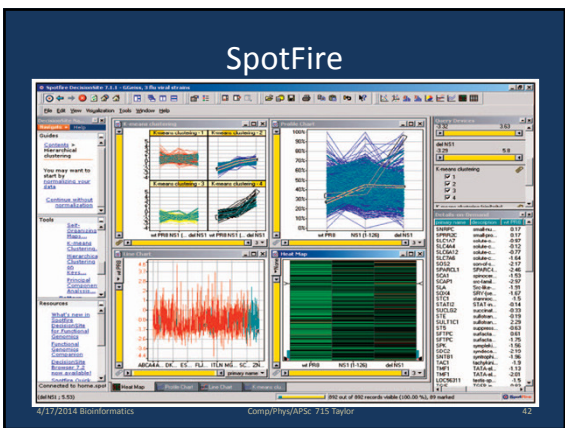
4/17/2014 Bioinformatics 38

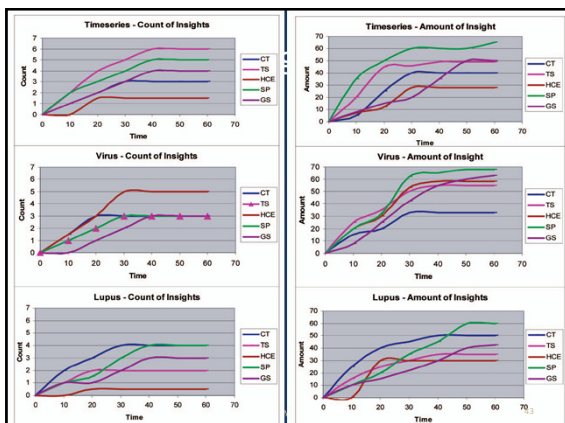
TimeSearcher

4/17/2014 Bioinformatics 39









Anecdotal Results

- Winner was specific to data set
 - Clusterview – Lupus
 - TimeSearcher – time series
 - HCE – viral
 - SpotFire decent for all
- Specific/free vs. general/commercial
 - General == no biological context
 - Tying in literature search is good
- Poor usability can break good visualization
- Motivation!
 - People learn faster if they care.

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 44

Where to go from here

- Lit search +++
- Standardization
- High throughput data
 - Microarray data needs pathway data for context
- Focus+context

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 45



Other topics

- Biological pathway visualization
- Sequence visualization
- Taxonomy visualization

Biological Pathways

- **networks** of complex reactions at the molecular level in living cells

The left diagram is a metabolic map showing the conversion of Glucose to various products. It includes pathways like Glycolysis, Citric Acid Cycle, and various biosynthetic branches leading to amino acids, nucleotides, and other biomolecules. The right diagram is a signaling pathway diagram showing the interaction between Apoptosis, Bcl-2, and Caspases, leading to cell death.

Survey of Popular Techniques

- Saraiya et al.
- Requirements analysis
- Anecdotal system evaluations
- Research agenda (future work)

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 49

General Goals

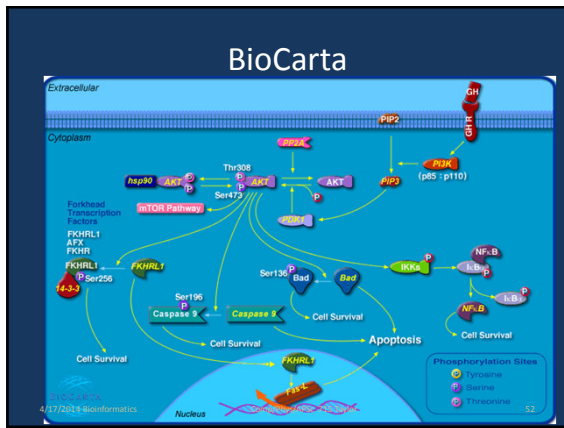
- recognition of changes between *experiment vs control* or between time points
- detection of *changes in relationship* between components of a pathway or between entire pathways
- identification of *global patterns* across a pathway
- mapping *pathway state to phenotype* (observable effects at the physical level in living organisms) or other biological information

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 50

Detailed Requirements

- Construct and update
- Context
- Uncertainty
- Collaboration
- Pathway node and edge info.
- Source
- Spatial information
- Temporal information
- High-throughput data
- Overview
- Interconnectivity
- Multi-scale
- Notebook

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor



GeneMapp

- Building pathways
 - Easy to use
 - But nobody wants to
- Statistical pathway comparison for different treatments
 - microarray data
- Animated node color
 - Different treatments

4/17/2014 Bioinformatics

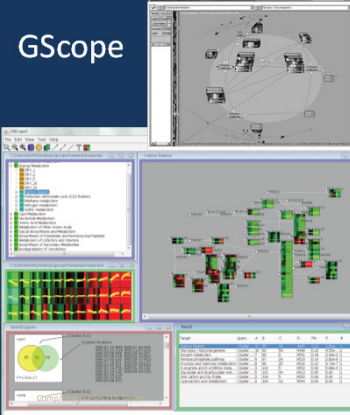
Cytoscape

- Microarray + pathway data
- Customizable everything
- CS-centric
 - Generic network vis
- UI complaints

4/17/2014 Bioinformatics

GScope

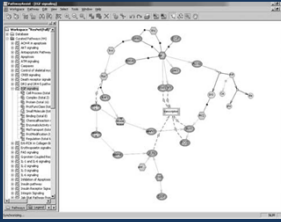
- Fish-eye lens
 - confusing
- Heat-map microarray table icons
- Distortions made condition comparison hard



4/17/2014 Bioinformatics

PathwayAssist: Literature Search

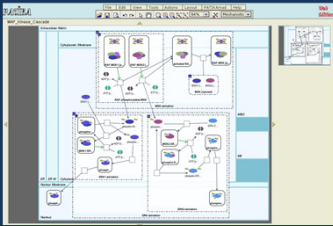
- Manual pathway building
- Automatic pathway building
 - NLP over PubMed or ResNet
 - Requires curation
- Scientific refs.



4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor

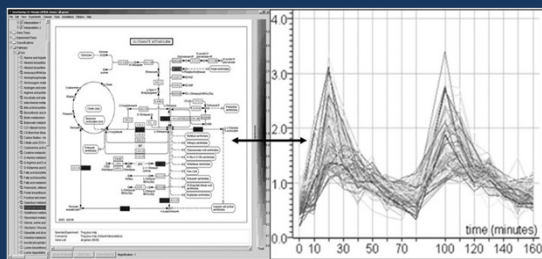
Patika

- Small database
 - Geared toward cells
- Regions make biological sense
 - Nucleus, cytoplasm, etc.



4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor

GeneSpring



4/17/2014 Bioinformatics

Comp/Phys/APS: 715 Taylor

58

Conclusions

- Not enough domain-specific info access
 - important for construction (NLP)
- Context in visualization
 - cell structures, molecular state
- No standardization
- Better microarray incorporation

4/17/2014 Bioinformatics

Comp/Phys/APS: 715 Taylor

59



4/17/2014 Bioinformatics

Comp/Phys/APS: 715 Taylor

60



Taxonomy Visualization

- Graham and Kennedy - Synonymy, Structural Markers

Figure 2 Taxonomies may not always use ranks consistently, even internally.

4/17/2014 Bioinformatics 65

Revisions

4/17/2014 Bioinformatics

Comparing Selections

This node A does not occur in the right-hand side hierarchy. So we must match across the descendant nodes and then track upwards through the ancestor paths to reconstruct a partial tree. Dropping down from the root to the first branch point or selected node gives the inclusive subtree.

- - Selected node after A is chosen in left-hand side hierarchy.
- - Unselected node
- ↗ - Reconstructed tree path
- ↘ - Downward path from root

Finding the least common ancestor by backtracking up through a hierarchy.

4/17/2014 Bioinformatics Comp/Phys/APSC: 715 Taylor 67

data:ITIS:1999Parent:NTSNTSH @ Kingdom:Animalia
Phylum:Chordata
Class:Aves
Family:Psittacidae
Species:Micrositta pusio (Scalater, 1886)

4/17/2014 Bioinformatics 68

Selection % Area Slider

Visualisation after multiple selections.

Selected areas set to just below 10%. As a result the selected, coloured areas are compressed.

Slider set to 70%. Nodes containing selected material expand out at the expense of unselected areas.

Slider set to 100%. Unselected subtrees removed from display or collapsed to leaf nodes.

4/17/2014 Bioinformatics Comp/Phys/APSC: 715 Taylor 69



Multiscale Trajectory Data

- Martin Luboschik, Biovis 2012
 - Compare a signal across time scales
 - Select heterogeneous regions: (small != large) and mark
 - Shows where details differ
 - Enable drill-down

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor

ADVISE: Changes in enzyme annotation

- Sabrina A. Silveira, Biovis 2012
- Annotation difference between releases
 - UniProt/SwissProt database
 - Red/blue is above/below; beige unchanged
 - Area (log) scaled to show value

Image is link to movie

This Course: Range of Topics

- Available Visualization Techniques
 - 2D scalar, 3D scalar, Vector, Tensor, Multivariate
 - InfoVis, Bioinformatics
- Visual Perception
 - How do you pick from all the options?
- Advanced ParaView Techniques
- Working as a team for a client
 - Client feedback, peer feedback, team contracts

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 76

What You Did (for your CV)

- 10 Designs for real-world data sets
 - 5 *team designs* for homework
 - 2 rounds of *team design for a client* (final project)
 - Several in-class designs
- 16 Design critiques
- Formal Evaluation of visualization design

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 77

How Did I Do?

- Evaluations
 - Look for the email with the link soon!
- Which were the most-useful parts?
 - Team design exercises? In-depth final projects?
 - In-class designs? Particular lectures?
- Which seem to be the least-useful parts?
- Suggestions for improvement

4/17/2014 Bioinformatics Comp/Phys/APS: 715 Taylor 78
