Developing Visualizations with Biologists

Miriah Meyer
Postdoctoral Research Fellow
Initiative in Innovative Computing
Harvard University
roadmap

- designing for biology: a methodology
- case study: MizBee
- challenges and lessons learned
designing for biology: a methodology
why biology?

why visualization?
why biology?

CAACATCCGTAAATCCTTTTTTAGCACGAAAAACCGAAGGATTAAAAAGAAGTTGCTAGGAGGAAGGG

why visualization?

replace mental cognition with visual perception
the role of vis research in biology
computer science

visualization

biology

characterization of measured data to data types

abstraction of scientific questions to data-specific tasks
computer science

visualize relationships in visual components

incorporate interaction for data exploration

biology
building effective tools for biological data is an intrinsically multidisciplinary task requiring experts in both biology and information visualization
methodology

1. learn*

2. characterize

3. design*

4. implement

5. validate*

*receive feedback from biologists and reiterate
MizBee
a multiscale synteny browser

design study

- data from the field of comparative genomics
- series of interviews with two biologists
- validate, analyze, and communicate computational results
contributions

1. characterization of the problem domain
2. taxonomy of the design space
3. MizBee, a multiscale synteny browser
4. validation through two case studies
biology concepts

• compare **genomes**

• genomes made of **chromosomes**

• stuff on the same chromosome: **synteny**

• stuff = **features** (genes)

• features grouped into **blocks**

• similar blocks on different chromosomes implies **conservation**
characterization
high level biology questions

**evolution:** How long ago did two species share a common ancestor?

**function:** Which segment of the genome is responsible for a specific function in the cell?

low level data-centric questions

**algorithm refinement:** Are the paired features within a block contiguous?

**refinement:** Which chromosomes share conserved blocks?

Are similarity scores alike within a block?
<table>
<thead>
<tr>
<th>conservation relationships</th>
<th>multiple scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>proximity</td>
<td>genome</td>
</tr>
<tr>
<td>size</td>
<td>chromosome</td>
</tr>
<tr>
<td>orientation</td>
<td>block</td>
</tr>
<tr>
<td>similarity</td>
<td>feature</td>
</tr>
</tbody>
</table>

difficult to answer **multiple questions** across a **range of scales** using computational algorithms alone

visually encode conservation relationships at different scales to validate, analyze, and communicate results
taxonomy
3. design

taxonomy of the design space

- represent chromosomes as segments
- encode conservation

src: [chromosome representation]
dst: [chromosome representation]
3. design

taxonomy of the design space

- represent chromosomes as segments
- encode conservation

src

color

dst

Cinteny [Sinha07]
3. design

taxonomy of the design space

- represent chromosomes as segments
- encode conservation

connection

color

src
dst
3. design

**taxonomy**

of the design space

- represent chromosomes as segments
- encode conservation
3. design

taxonomy of the design space

• represent chromosomes as segments
• encode conservation

color

connection

src
dst

circos [Krzywinski]
3. design

**taxonomy**

of the design space

- represent chromosomes as segments
- encode conservation

MizBee
3. design

taxonomy of the design space

- represent chromosomes as segments
- encode conservation
- layout

<table>
<thead>
<tr>
<th></th>
<th>contiguous</th>
<th>discrete</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td></td>
<td>segregated</td>
</tr>
<tr>
<td>circular</td>
<td></td>
<td>interleaved</td>
</tr>
<tr>
<td>separate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>combined</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. design

taxonomy of the design space

- represent chromosomes as segments
- encode conservation
- layout

Mauve [Darling04]
3. design

**taxonomy** of the design space

- represent chromosomes as
- encode conservation
- layout
3. design

taxonomy of the design space

- represent chromosomes as segments
- encode conservation
- layout

<table>
<thead>
<tr>
<th></th>
<th>contiguous</th>
<th>discrete</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>linear</td>
<td>circular</td>
</tr>
<tr>
<td>separate</td>
<td><img src="image" alt="Separate Linear" /></td>
<td><img src="image" alt="Separate Circular" /></td>
</tr>
<tr>
<td>combined</td>
<td><img src="image" alt="Combined Linear" /></td>
<td><img src="image" alt="Combined Circular" /></td>
</tr>
</tbody>
</table>

*Image sources: [Apollo](Lewis02)*
3. design

taxonomy
of the design space

- represent chromosomes as segments
- encode conservation
- layout
3. design

taxonomy of the design space

- represent chromosomes as segments
- encode conservation
- layout
- encode similarity
3. design

**taxonomy**

- represent chromosomes as segments
- encode conservation
- layout
- encode similarity

MizBee
MizBee novelty

- first synteny browser with side-by-side linked views
  - across the range of scales
  - encoding all four relationship types

- redundantly encodes conservation
  - color and connection
  - edge bundling to enhance trends

- encode similarity in context of other three relationship types
validation
rhizopus
difficult to simultaneously show the location of interesting features and other conservation relationships in a static image
stickleback

pufferfish

http://fish.dnr.cornell.edu/nyfish/Gasterosteidae/stickleback.html

http://win.vergari.com/acquariofilia/salmastro02.asp
“The first time I saw my data in [MizBee] I was totally disappointed. The data was very noisy, and there were many small blocks that went to different chromosomes.”
stickleback and the pufferfish

“The first time I saw my data in [MizBee] I was totally disappointed. The data was very noisy, and there were many small blocks that went to different chromosomes.”

“I don’t think I would ever have gotten here. The noise was very hard to see in the scatter plots while [MizBee] is much more unforgiving.”

Genome-wide synteny through highly sensitive sequence alignment: Satsuma, M. Grabherr et al., submitted.
<table>
<thead>
<tr>
<th><strong>rhizopus</strong></th>
<th><strong>stickleback</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>fungus</td>
<td>fish</td>
</tr>
<tr>
<td>genomic function</td>
<td>evolution</td>
</tr>
<tr>
<td>communication</td>
<td>algorithm refinement</td>
</tr>
<tr>
<td>late stage user</td>
<td>early stage user</td>
</tr>
<tr>
<td>early in MizBee design</td>
<td>late in MizBee design</td>
</tr>
</tbody>
</table>

5. validate
implemented in Processing
processing.org

released open source
mizbee.org

encourage broader user base
research challenges

• language barriers
• interviewing skills
• generalization of specific questions
• communication of abstracted concepts

lessons learned

• picking collaborators is hard
• step one: reimplement the wheel
• think like a scientist
questions?