MizBee
A Multiscale Synteny Browser

Miriah Meyer\textsuperscript{1}
Tamara Munzner\textsuperscript{2}
Hanspeter Pfister\textsuperscript{1}

\textsuperscript{1}Harvard University
\textsuperscript{2}University of British Columbia
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? Which chromosomes share conserved blocks? Are similarity scores alike within a block?
conservation relationships

- proximity
- size
- orientation
- similarity

multiple scales

- genome
- chromosome
- block
- feature

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
taxonomy of the design space

- represent chromosomes as segments
- encode conservation
taxonomy

of the design space

- represent chromosomes as segments
- encode conservation

src

color

dst

Cinteny [Sinha07]
taxonomy of the design space

• represent chromosomes as segments
• encode conservation
taxonomy of the design space

- represent chromosomes as segments
- encode conservation

Mauve [Darling04]
taxonomy
of the design space

• represent chromosomes as segments
• encode conservation

Circos [Krzywinski]
taxonomy of the design space

- represent chromosomes as segments
- encode conservation

connection

src
color
dst
taxonomy of the design space

- represent chromosomes as segments
- encode conservation
- layout

Mauve [Darling04]
taxonomy of the design space

- represent chromosomes as
- encode conservation
- layout


**taxonomy**

of the design space

- represent chromosomes as segments
- encode conservation
- layout

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

[Apollo [Lewis02]]
taxonomy of the design space

- represent chromosomes as segments
- encode conservation
- layout

<table>
<thead>
<tr>
<th>contiguous</th>
<th>discrete</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>circular</td>
</tr>
<tr>
<td>separate</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>combined</th>
<th>segregated</th>
<th>interleaved</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>circular</td>
<td>discrete</td>
</tr>
</tbody>
</table>
taxonomy of the design space

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

[Rasko05]
taxonomy of the design space

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

difficult to simultaneously show the location of interesting features and other conservation relationships in a static image
stickleback

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.”
“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.”

Genome-wide synteny through highly sensitive sequence alignment: Satsuma, M. Grabherr et al., submitted.
“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.”

“Honestly, I don't know. I don't think I would ever have gotten here. The noise was very hard 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.
discussion
and
future work
<table>
<thead>
<tr>
<th>rhizopus</th>
<th>stickleback</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>
implemented in Processing
processing.org

released open source
mizbee.org

encourage broader user base
conclusions

• presented characterization and taxonomy

• **MizBee**
  • first synteny browser to show side-by-side linked views across multiple scales and conservation relationships

• collaboration with two domain experts
questions?

acknowledgements:
data: Manfred Grabherr and Li-Jun Ma
manuscript: Janet Iwasa and Matthew Tobiasz
inspiration: Vis Group at the Broad
funding: IIC at Harvard

visit:
mizbee.org