



UNIVERSITÄT
LEIPZIG

The Supergenome Browser

February 17, 2018

Daniel Wiegrefe (formerly known as Gerighausen)



ScaDS 
DRESDEN LEIPZIG

Motivation

Why building the $n+1$ th genome browser?!

- comparison of different species/genome versions is only possible on a very coarse grained or detailed level
- only coarse grained or local comparison visualizations are available
- so far no inter-species genome browser has been developed

Motivation

Why building the $n+1$ th genome browser?!

- comparison of different species/genome versions is only possible on a very coarse grained or detailed level
- only coarse grained or local comparison methods are available
- so far no inter-species genome browser has been developed
- **Peter doesn't like the other genome browsers!**

Background

Problems with Multiple Sequence Alignments on genome level:

- MSA blocks are not ordered or connected
- to order the blocks a coordinate system is necessary
- how to deal with loops and ramifications in the order?

Background

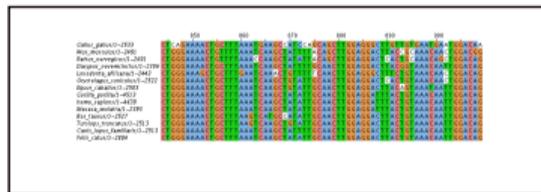
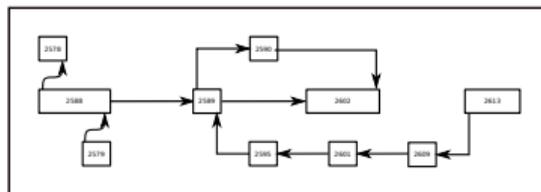
Solving it with the Supergenome Coordinatesystem:

- create an alignment graph based on the MSA blocks
- simplify the graph by local rules (remove loops, etc.)
- use heuristics to solve the underlying betweenness problem to calculate an order of the MSA blocks.

Visualization Model

3 visualization layers:

- global view on chromosome level
- semi-global view in the range of multiple genes
- local view based on the common nucleotide view



Deployment

Managed Hosting

- hosted on the HPC in Dresden
- public access
- storage for a large amount of annotation data and enough computational power

Self Hosting

- local deployment of the server instance
- less privacy issues
- use your own genome versions

Data and Glyph Import

Additional annotation data:

- module system for importing and exporting additional data
- usage of published libraries

Additional glyphs for the visualization:

- API to integrate new glyphs, e.g. for new RNA types
- users can create their own glyphs

Interaction with the Genome Graph

Manipulate and correct the underlying graph:

- choose the reference species for the visualization
- manual manipulation of the block order
- filtering blocks
- recalculate the block order based on the user input

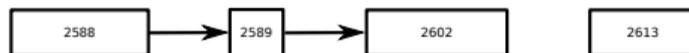
Semi-global View

- display a limited number of vertices of the genome graph
 - in a perfect world: vertices form a chain
 - in reality: lots of ramifications, sinks, and sources
 - different directions for edges of different species
- > How to visualize it?

Semi-global View Algorithm

Calculate the path for the reference species:

1. find all vertices incorporating the requested range of the reference species
2. check if all vertices are connected
3. check ramification events for each vertex



Semi-global View Algorithm

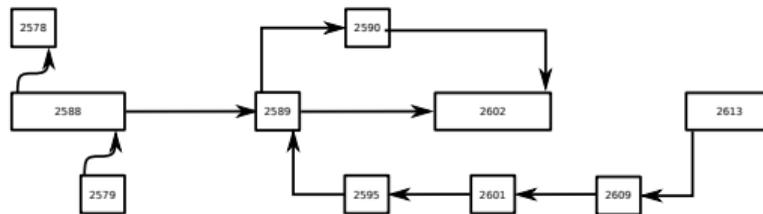
Calculate the other paths:

1. check for each outgoing path if it is a sink:
 - follow the path until it is reunited with the reference path
 - > path is not a sink
 - check the position in the block order for each vertex
 - if the position is smaller the minimum or larger than the maximum of the selected reference vertices, stop it
 - > path is a sink
2. check for unknown incoming paths as sources
3. order the paths by length
4. check for overlapping paths

Semi-global View Algorithm

How to visualize it?

1. draw the reference vertices since they form linear path
2. draw all sinks and sources
3. for each species path:
 - 3.1 pick the shortest from the remaining paths
 - 3.2 check the direction of the path
 - 3.3 check if overlapping path is already drawn
 - 3.4 draw path without unwanted intersections



Conclusion and Feedback

- the Supergenome Browser will be a tool to compare MSAs on different levels of detail
- it will be extended with common genome browser functions
- development of new functions to simplify the maintenance
- different deployment options to avoid privacy issues
- > development is at the beginning, your feedback and ideas are very welcome!

ACKNOWLEDGMENTS

I want to thank:

- The BMBF for funding this project
- Fabian Gärtner, Dirk Zeckzer, and Peter Stadler
- Berni for this wonderful latex template



UNIVERSITÄT
LEIPZIG

Thank You!

Questions?

Daniel Wiegrefe

Image and Signal Processing Group
Department for Computer Science