VG Browser: Interactive Visualization of Genome Variation Graphs

Simon Heumos
Quantitative Biology Center

Biohackathon 2019
1st September, Fukuoka

ZIM
Zentrales Innovationsprogramm Mittelstand

Gefördert durch:
Bundesministerium für Wirtschaft und Energie
aufgrund eines Beschlusses des Deutschen Bundestages

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Order or Chaos?!

Human hg18

chr22

chr22:20,000,000-20,100,000

Possible Data Structure Solution: Genome Variation Graphs ~2017

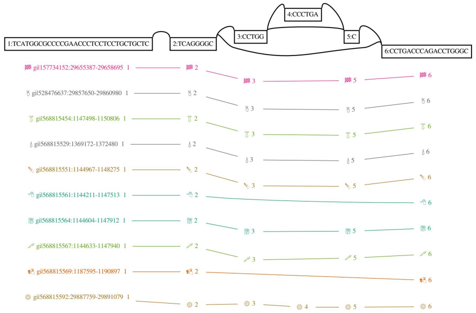
- Population Reference Graph[Dilthey et al. 2015]
  - HLA and MHC data, Mapping only

- GraphTyper[Eggertson et al. 2017]
  - population scale genotyping, Mapping

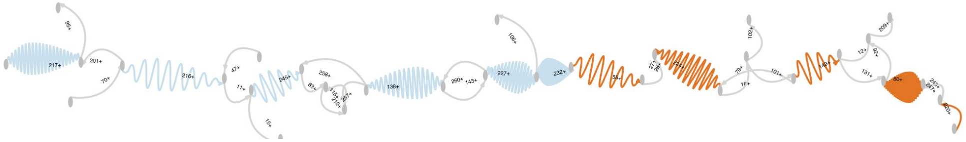
- vg
  - construct
  - map
  - align
  - view
  - call
  - augment
  - index
  - find
  - ...

Overview Visualization Tools ~2017

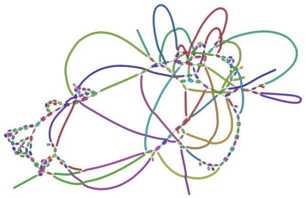
- Graphviz



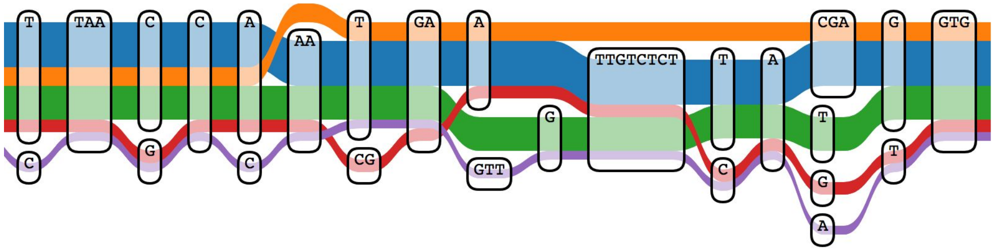
- ABySS-Explorer[Nielsen et al. 2009]



- BANDAGE[Wick et al. 2015]



- Sequence Tube Maps



No interactive genome browser solution available.

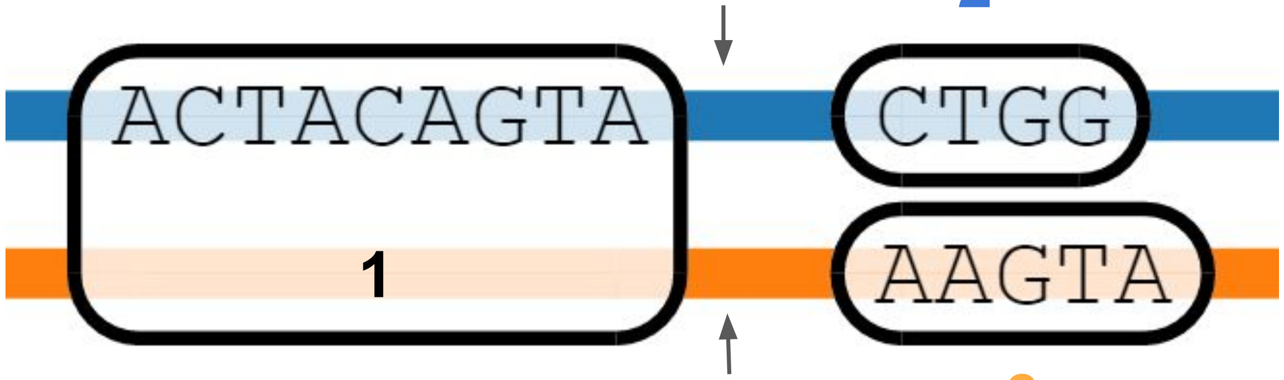
The Graph Data Structure + Visualization Layout

🟦 Genome 1: ACTACAGTA **CTGG** … Path: **1 2**

🟧 Genome 2: ACTACAGTA **AAGTA** .. Path: **1 3**

Genomes are paths through nodes.

Direction: from left to right.



Path Genome 1

2

Number
in/over/below
node: node id

Path Genome 2

3

Node:
- **Contains sequence**
- **Width reflects sequence length**

Master Thesis 2017 - Prototype Implementation

Data Structure

Master Thesis 2017 - Prototype Implementation

Data Structure

Webserver

Master Thesis 2017 - Prototype Implementation

Data Structure

Binding

Webserver

Master Thesis 2017 - Prototype Implementation

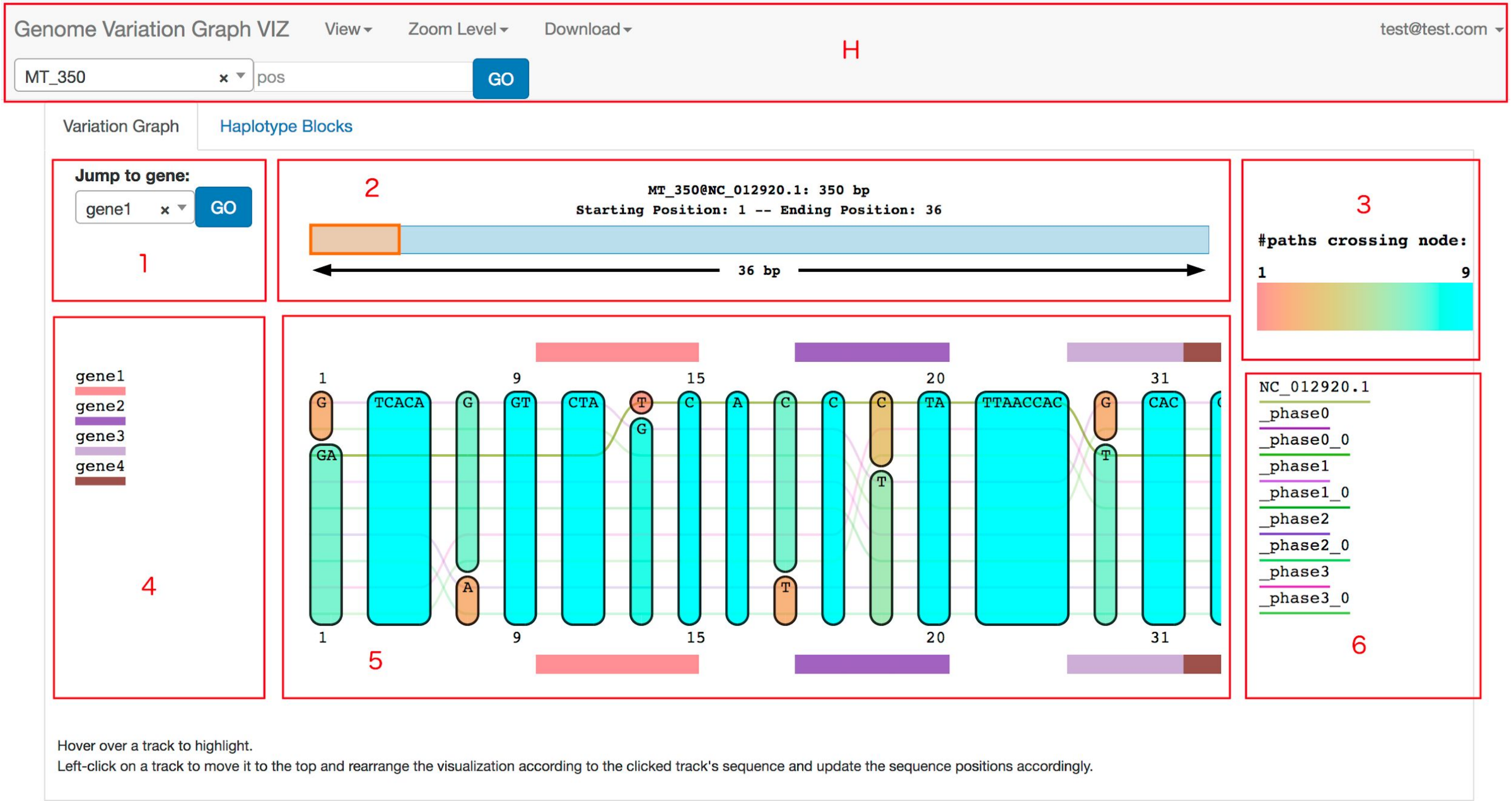
Data Structure

Binding

Webserver

Visualization

Master Thesis 2017 - Prototype Implementation

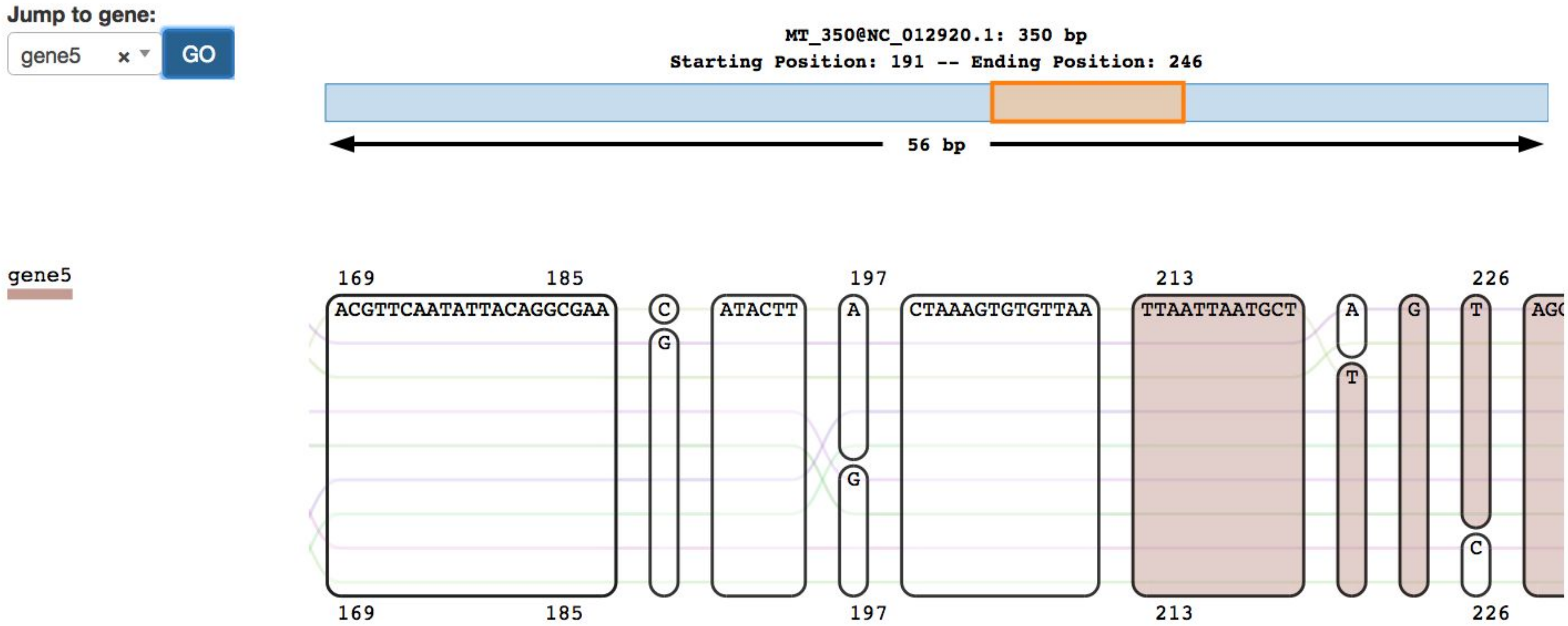


Hover over a track to highlight.

Left-click on a track to move it to the top and rearrange the visualization according to the clicked track's sequence and update the sequence positions accordingly.

H: The header of the page. 1: The jump to gene chart. 2: The genome overview chart. 3: the heatmap chart. 4: The meta annotation chart. 5: The graph chart. 6: The meta path chart.

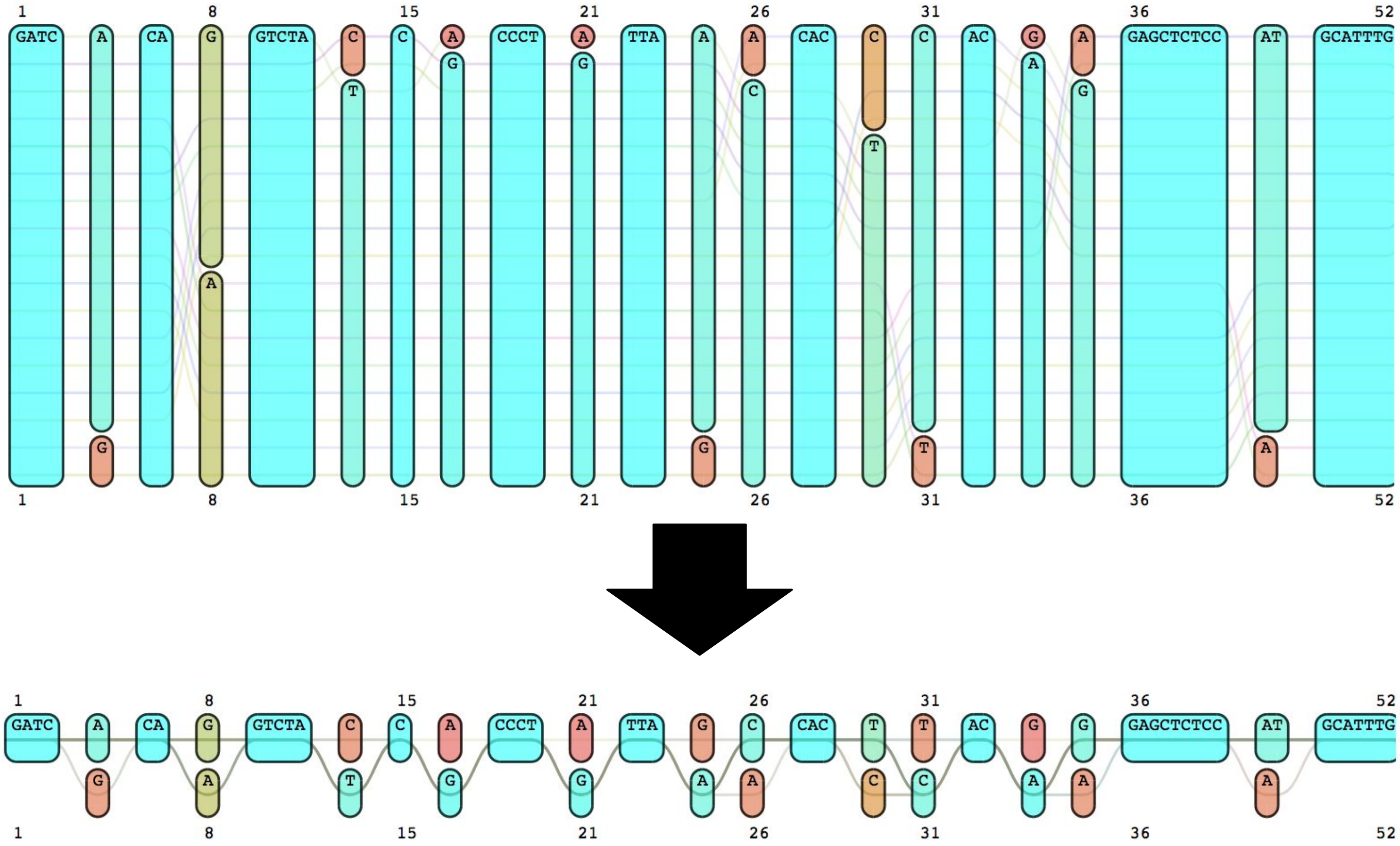
Master Thesis 2017 - Jump to Gene Position



Jump to gene: gene5

MT_350@NC_012920.1: 350 bp
Starting Position: 191 -- Ending Position: 246

Master Thesis 2017 - Thousands of Individuals?!



Visualization concept to visualize variants of thousands of individuals!

Prototype Limitations

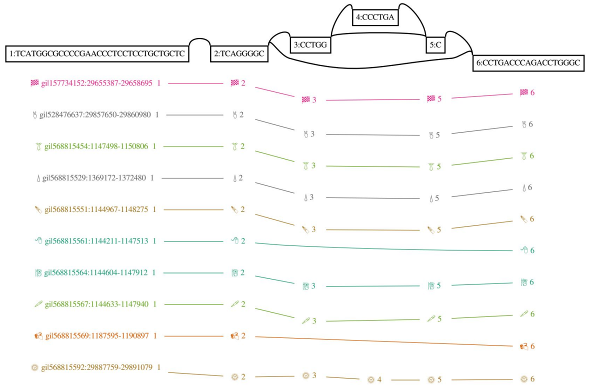
- No sophisticated zooming levels

- High implementation effort due to C++ bindings

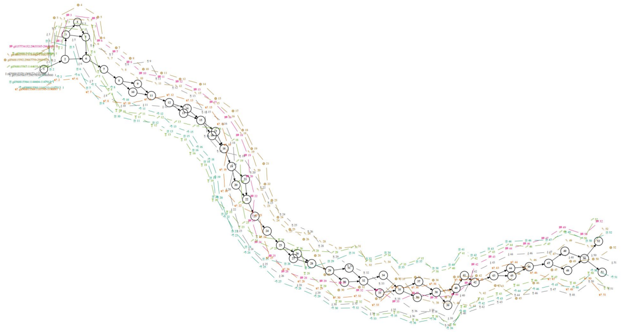
- Project not further pursued until 2019

Current Visualization Approaches - Static:

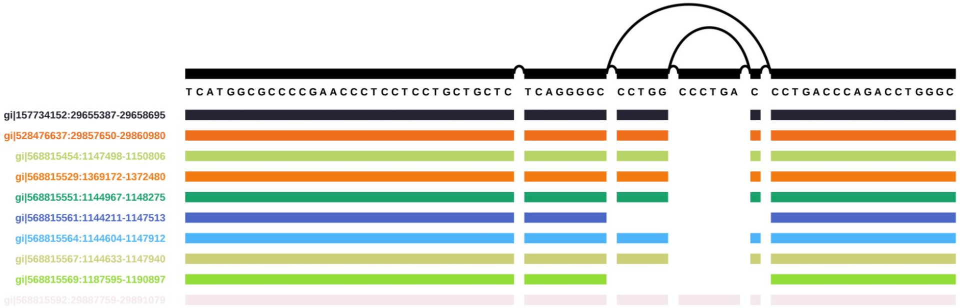
- vg view + Graphviz
  - dot



  - neato

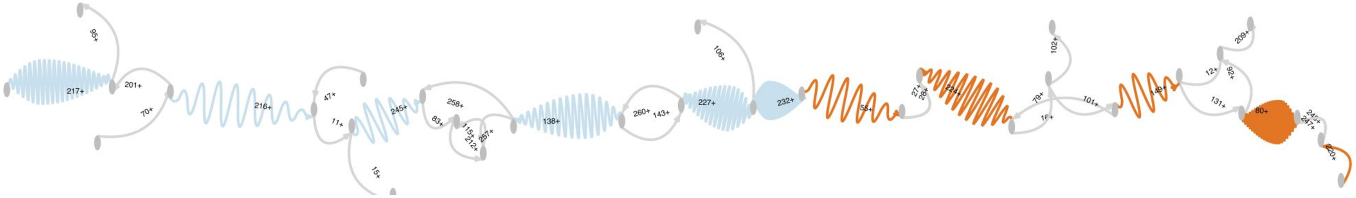


- vg viz

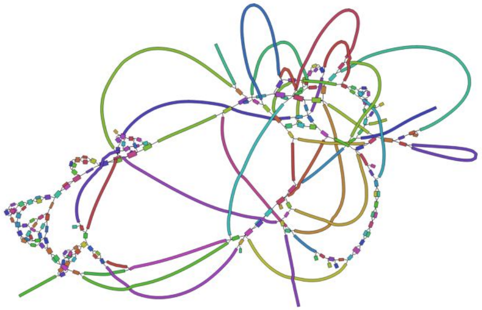


Current Visualization Approaches - Interactive 1:

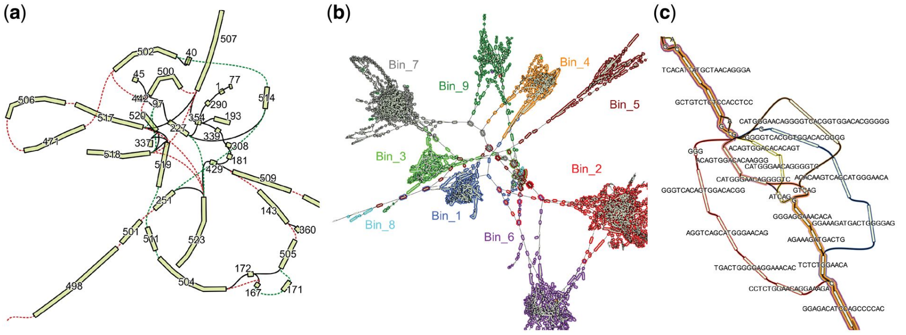
- ABySS-Explorer[Nielsen et al. 2009]



- BANDAGE[Wick et al. 2015]

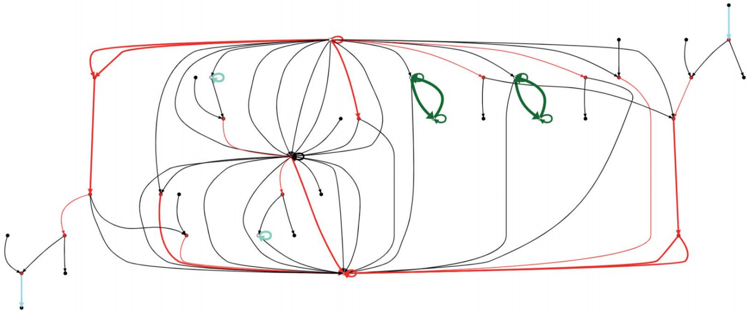


- GfaViz[Gonnella et al. 2018]

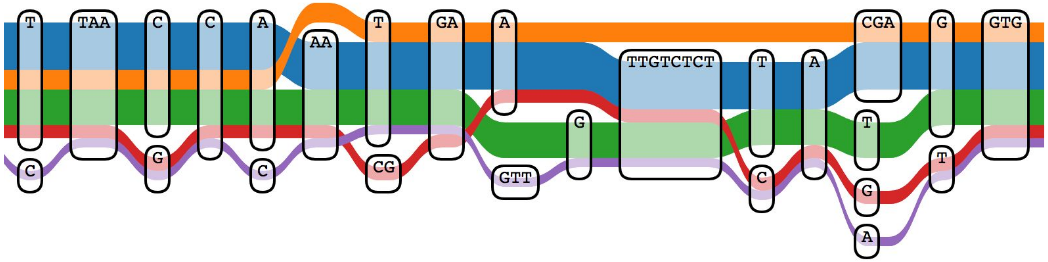


Current Visualization Approaches - Interactive 2:

- Assembly Graph Browser[Mikheenko et al. 2019]



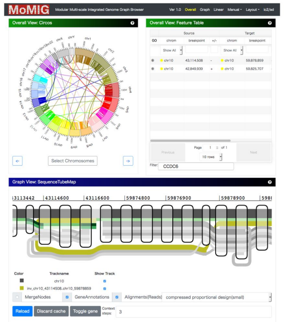
- Sequence Tube Maps[Beyer et al. 2019]



- MoMI-G[Yokoyama et al. 2019]

https://github.com/MoMI-G/MoMI-G

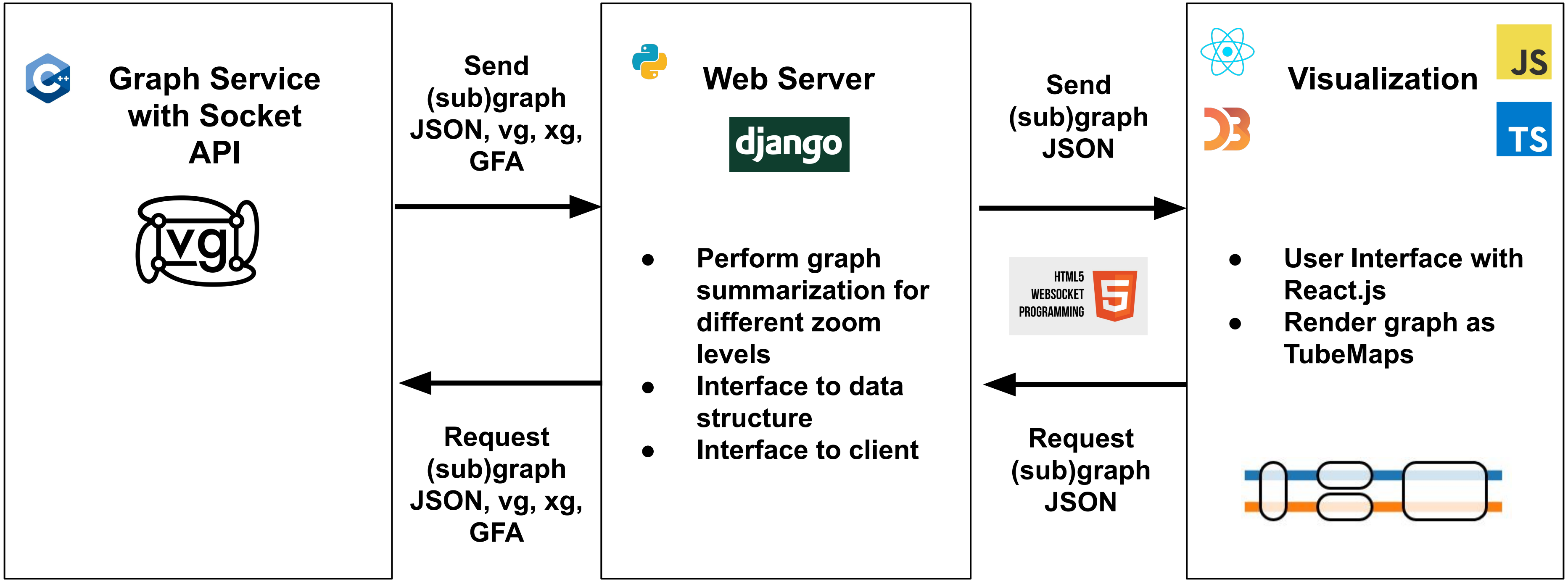
☑ Compare between genomes
☑ Visualize nested SVs
☑ Navigate for every SV
☑ Inspect the authority of SV



Graph Genome Browser - Major Challenges

- Zooming: graph summarization steps
- Display complex variations in a human readable way
- Server-Client communication logic
- Efficiency:
    - Client side: browse fluently, must work on tablet
    - Feasibility for large genomes with abundant complex variants

Graph Genome Browser - Possible Architecture



Graph Service with Socket API

Send (sub)graph JSON, vg, xg, GFA

Request (sub)graph JSON, vg, xg, GFA

Web Server

- **Perform graph summarization for different zoom levels**
- **Interface to data structure**
- **Interface to client**

Send (sub)graph JSON

Request (sub)graph JSON

Visualization

- **User Interface with React.js**
- **Render graph as TubeMaps**

Next Steps

- WIP: Graph summarization steps as zoom levels
    - Discuss file format for Haplotype Blocks?

- Extend frontend to allow continuous browsing of graphs

- Replace SequenceTubeMap Node.js backend with Django backend

- Biohackathon as opening event for project and concept discussions
  → **Prototype implementation**

Thank you!

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Supervisors:

Sven Nahnsen

Michael Krone

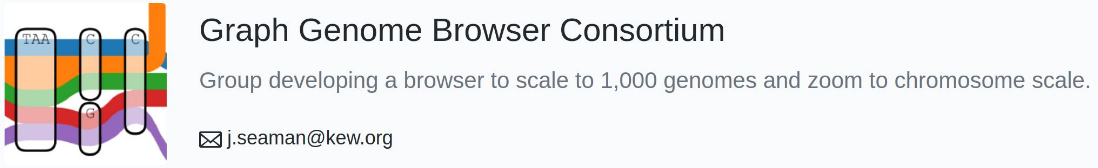
Oliver Kohlbacher

computomics®
molecular data analysis

Sebastian Schultheiss

Björn Geigle

Jörg Hagmann



Graph Genome Browser Consortium

Group developing a browser to scale to 1,000 genomes and zoom to chromosome scale.

✉ j.seaman@kew.org

Josiah Seaman

Toshiyuki Yokoyama

Torsten Pook

Erik Garrison

https://github.com/graph-genome/vgbrowser

References - Visualization Tools

Nielsen et al. **ABySS-Explorer: Visualizing Genome Sequence Assemblies** *IEEE* **volume 15**, **No. 6**, pages 881–888 (2009)

Wick et al. **Bandage: interactive visualization of *de novo* genome assemblies** *Bioinformatics* **volume 31**, **Issue 20**, pages 3350–3352 (2015)

Gonnella et al. **GfaViz: flexible and interactive visualization of GFA sequence graphs** *Bioinformatics* **bty1046,** pages 1–3 (2018)

Mikheenko et al. **Assembly Graph Browser: interactive visualization of assembly graphs** *Bioinformatics* **btz072,** pages 1–3 (2019)

Beyer et al. **Sequence tube maps: making graph genomes intuitive to commuters** *Bioinformatics* (2019)

Yokoyama et al. **MoMI-G: Modular Multi-scale integrated Genome Graph Browser** *bioRxiv* (2019)