



# **Drunken Noodle**

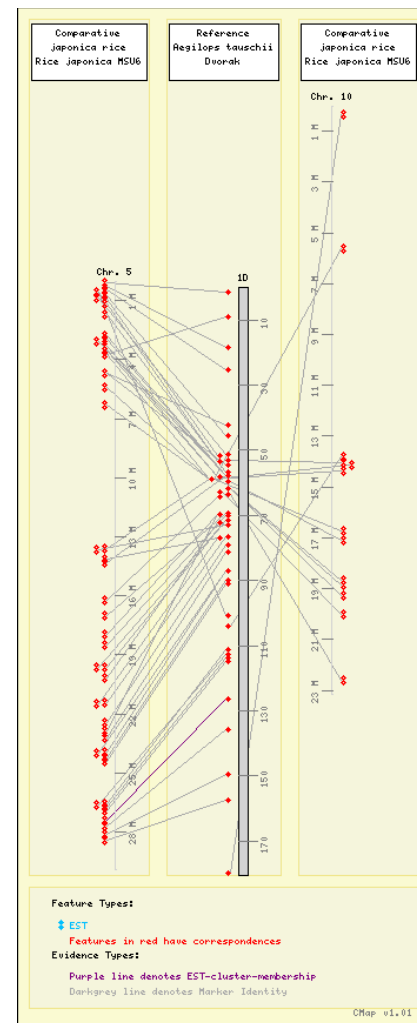
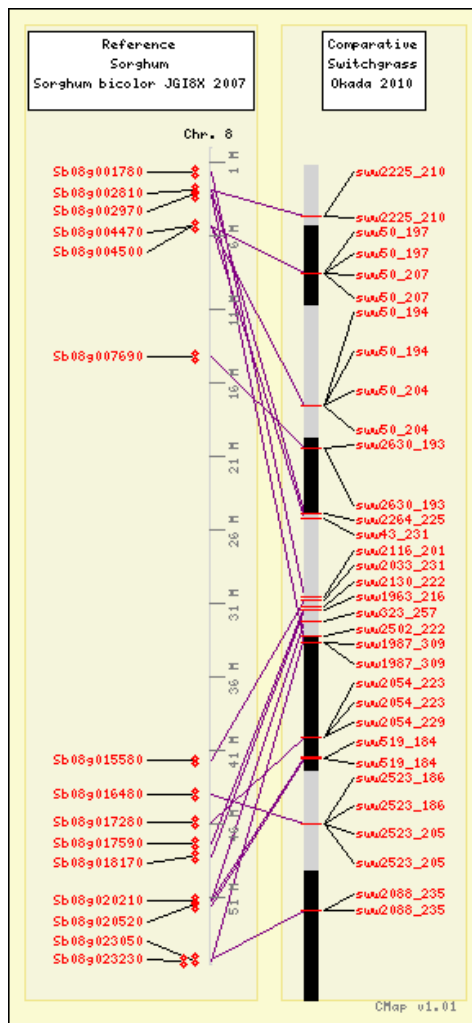
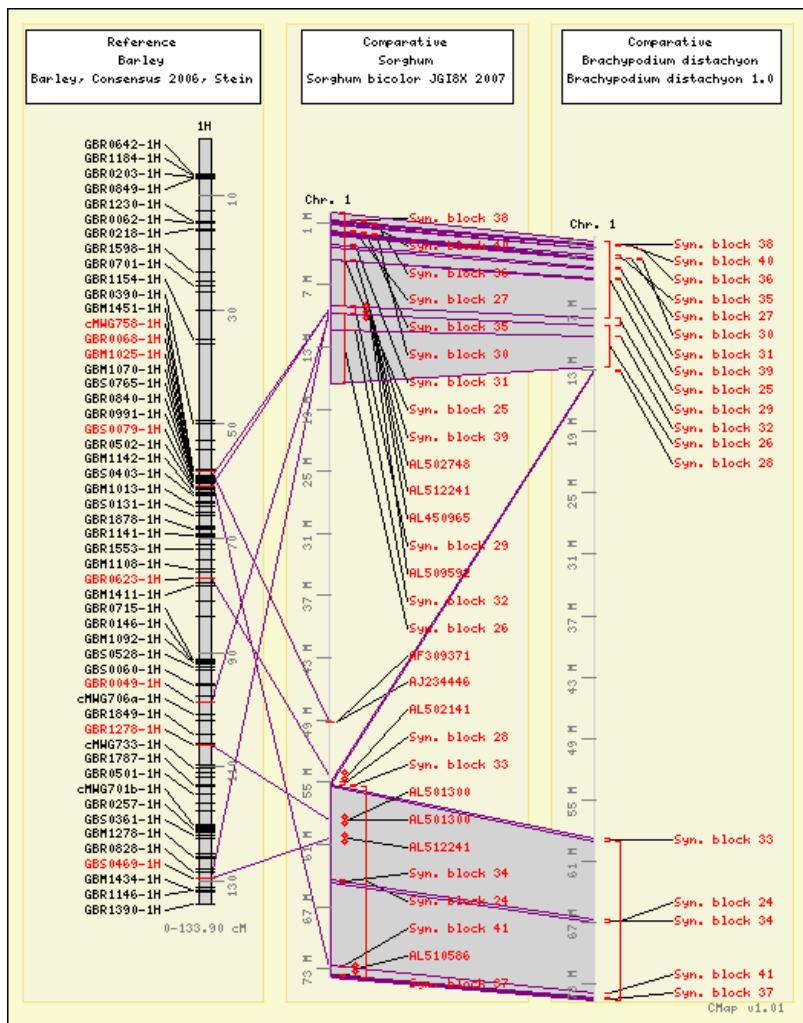
GMOD Meeting  
16 January 2014  
Ken Youens-Clark  
Cold Spring Harbor Lab

# CMap of Olde

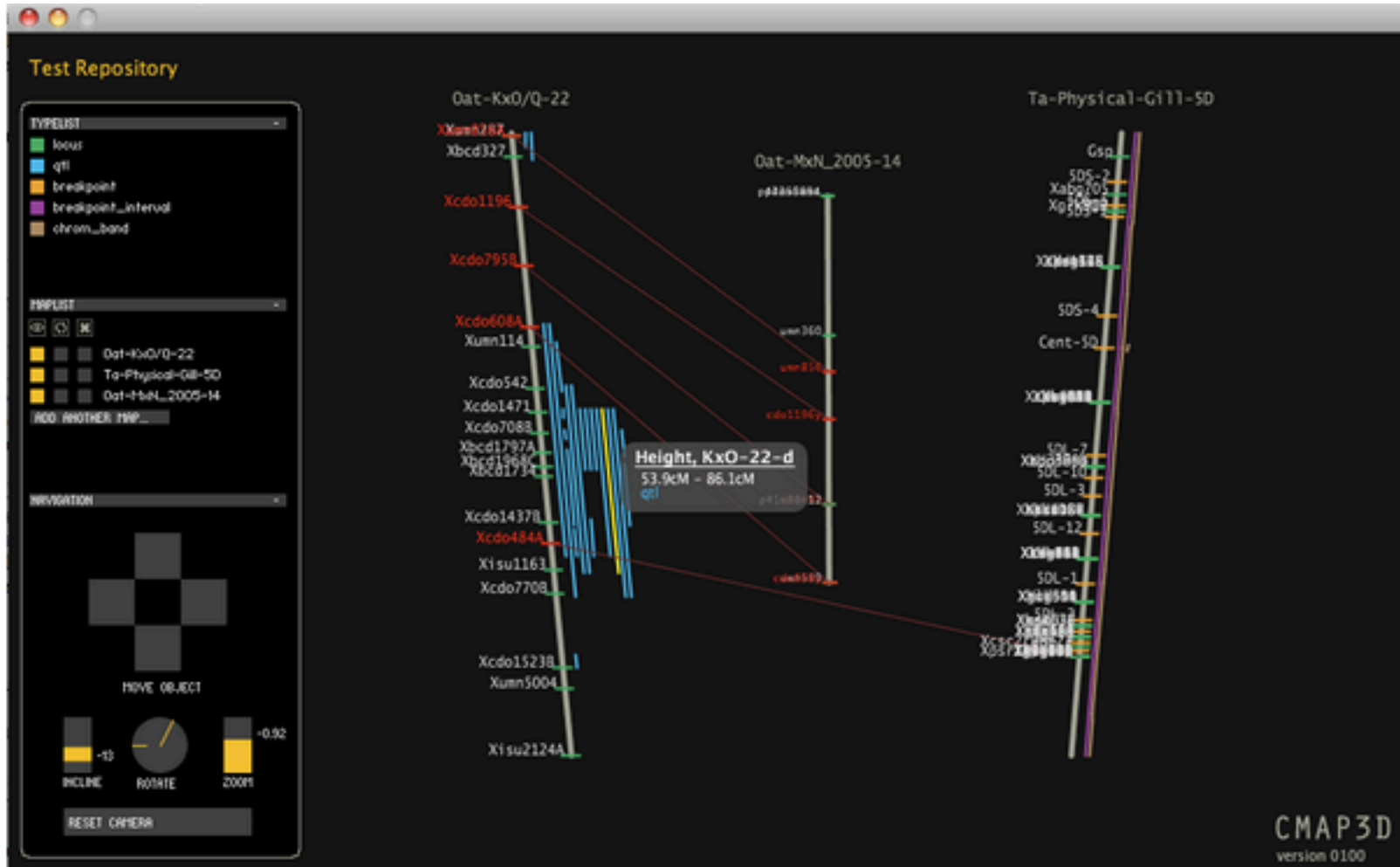
- In development 2001-2008 primarily for Gramene
- By me and Ben Faga
- Entirely Perl/CGI/libgd/RDBM
- <http://gmod.org/wiki/CMap>



# http://archive.gramene.org/cmap/



# CMap3D



CMap3D: a 3D visualization tool for comparative genetic maps. Duran C, Boskovic Z, Imelfort M, Batley J, Hamilton NA, Edwards D. Bioinformatics. 2010 Jan 15;26(2):273-4. Epub 2009 Nov 25.

# Goals

- Faster and more scalable for larger data sets
  - Solr
  - Fastbit
  - Tabix
- More interactive, client-side user interface
- Allow end users to upload data sets (maps, annotations, correspondences, DAS)
- Allow site administrators greater customization

# <https://github.com/warelab/>

- **Sage:** Nodejs REST architecture
- **Iris:** Browser-side library for data visualization widgets written in JavaScript



# Snapdragon

- Snapdragon is a collection of low level tools written in C++ mostly built on top of fastbit.
- You can parse, query, and compare genome annotations.
- There are stand alone command line programs as well as a node module that wraps some Fastbit functions designed to be used by a web service.
- E.g., histogram data for feature type distribution

# Explore visualization libraries

- CMap generates static server-side images (maps)
- Explore D3 (<http://d3js.org/>) and other tools (<http://selection.datavisualization.ch/>)
- “Beauty” is ultimate goal

**Beauty is truth, truth beauty -- that is all  
Ye know on earth, and all ye need to know**

-- Keats



# Data loading formats

- tab-delimited
- GFF
- VCF
- BED

# Map type support

- genetic (QTL, indel, SNP)
- sequence (scaffolds, finished chromosome)
- physical (contigs)

# Data storage

- Previous version used RDBMS (MySQL, etc.); probably not scalable
- May store maps in GFF, use tabix for speed in range queries
- Store map-to-map relationships in pairwise files, possibly use bit vectors (FastBit?) for indexing, querying (e.g., histograms)

# REST interface, server architecture

- If rendering is to be done client-side, then data will be delivered via REST/JSON
- May use Perl server (Mojolicious) as there are good tools for manipulating upload formats, interfaces to FastBit, tabix, etc.; maybe Nodejs

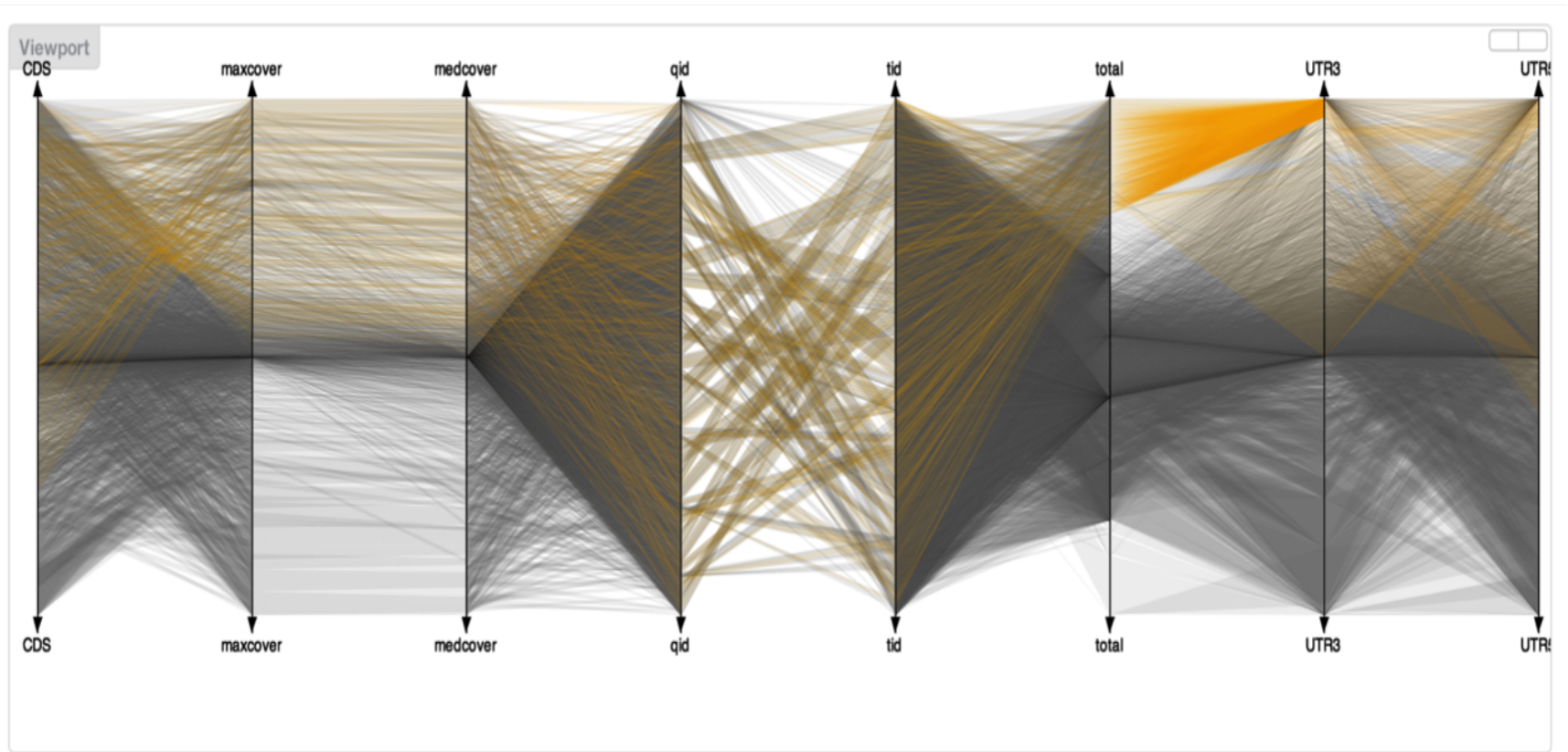
# User data

- Ability to upload user maps comes first
- Public/private option for user maps
- Server-side code to integrate maps, create correspondences, generate syntenic blocks
- User logins to manage maps, update, etc.
- Pulling maps from other sites via DAS, REST

# Performance goals

- Maximum feature density on maps?
- Ways to represent genomic-level sequence maps
- Use of synteny blocks/alpha channels to show high-level views, density of relationships
- Digest data on server, deliver minimum data points (via JSON, shudder at SOAP) to draw

# Parallel Coordinate Plot in Iris



# Manhattan Plots

Iris Manhattan Pcoords Bubble Plot Force-directed Graph

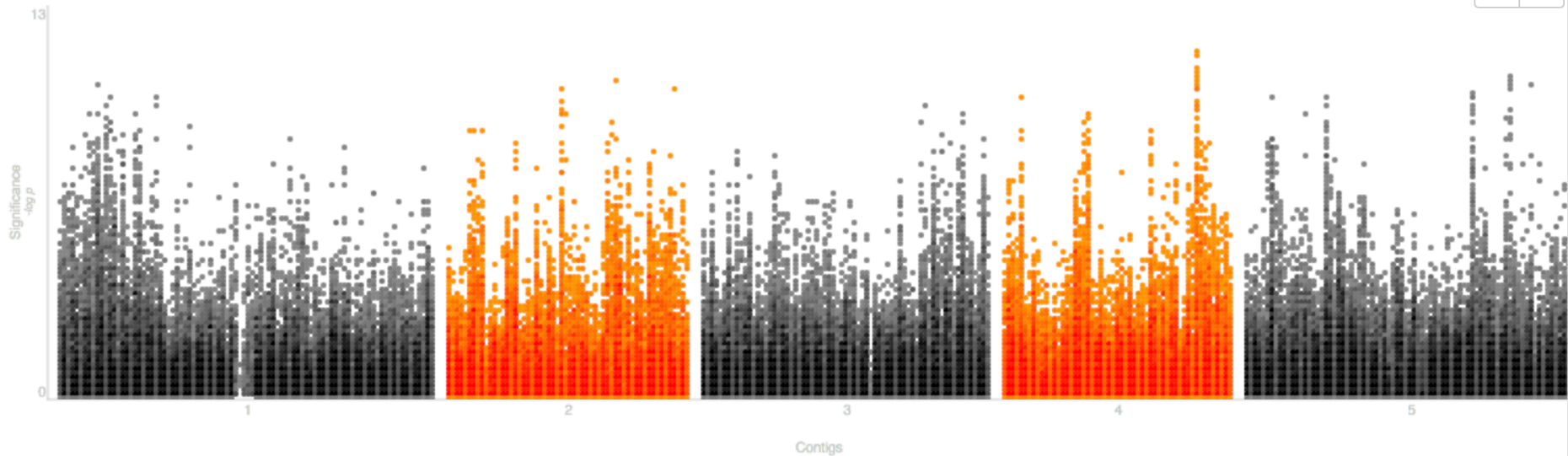
0.2.11

**Widget** Manhattan Plot

**Author** Andrew Olson

**Description**

Viewport

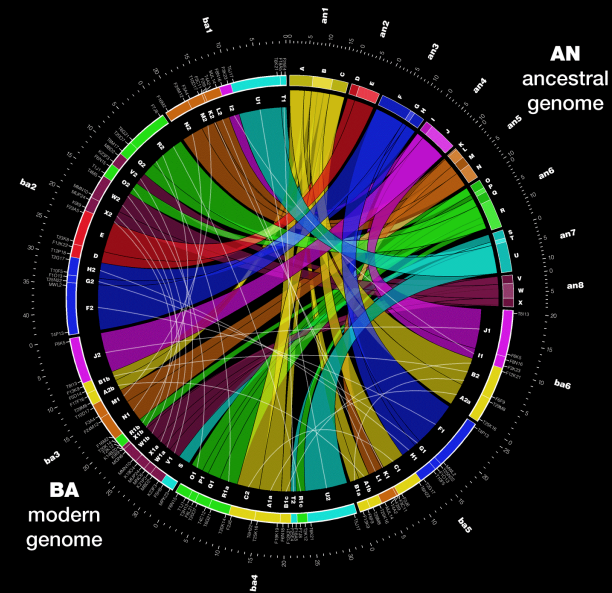
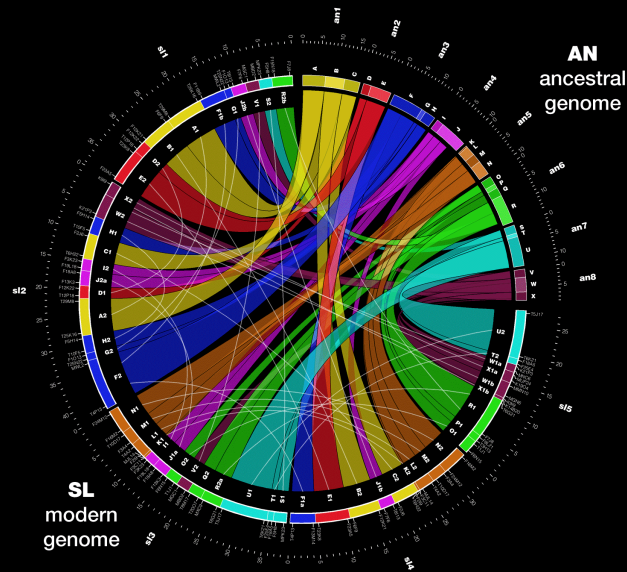
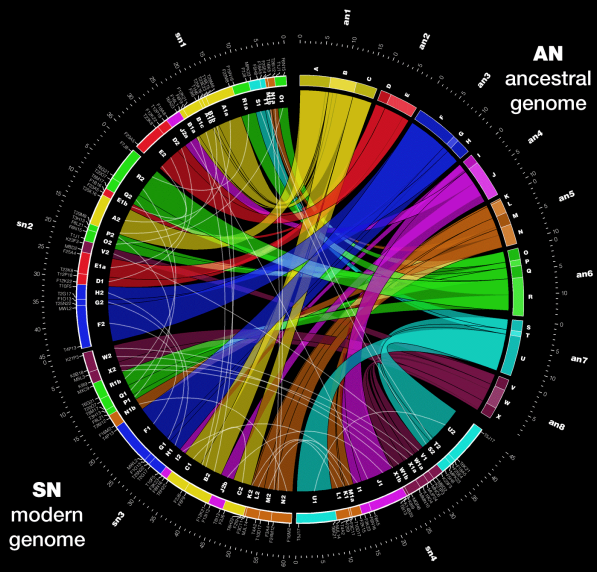




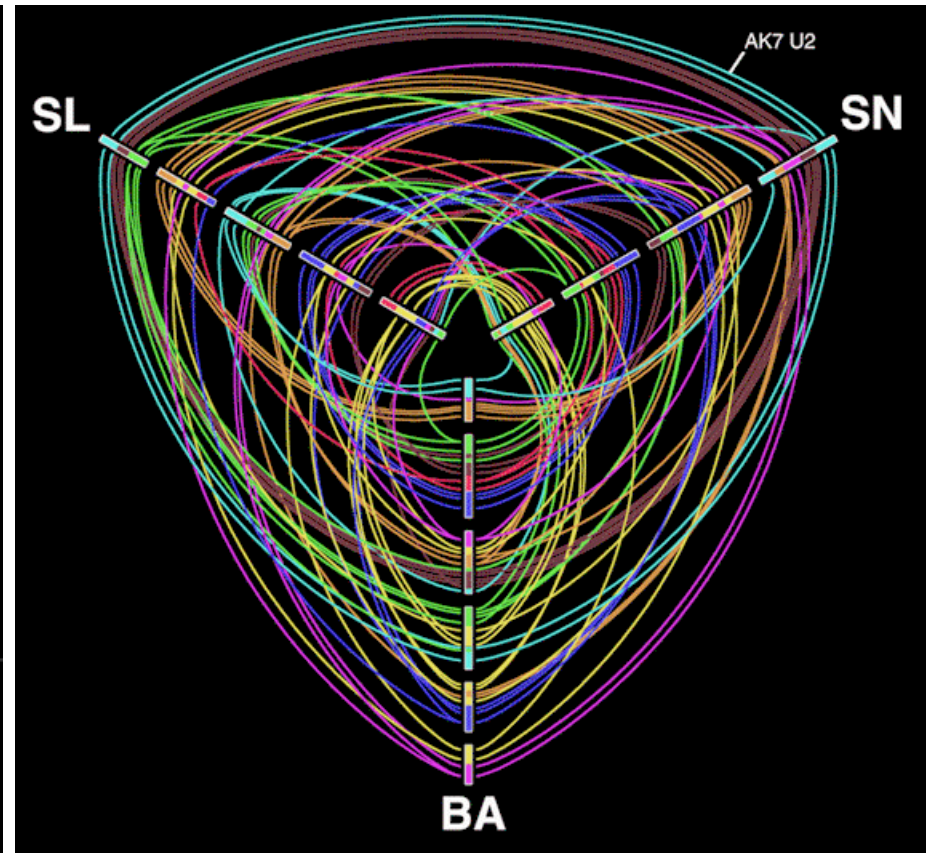
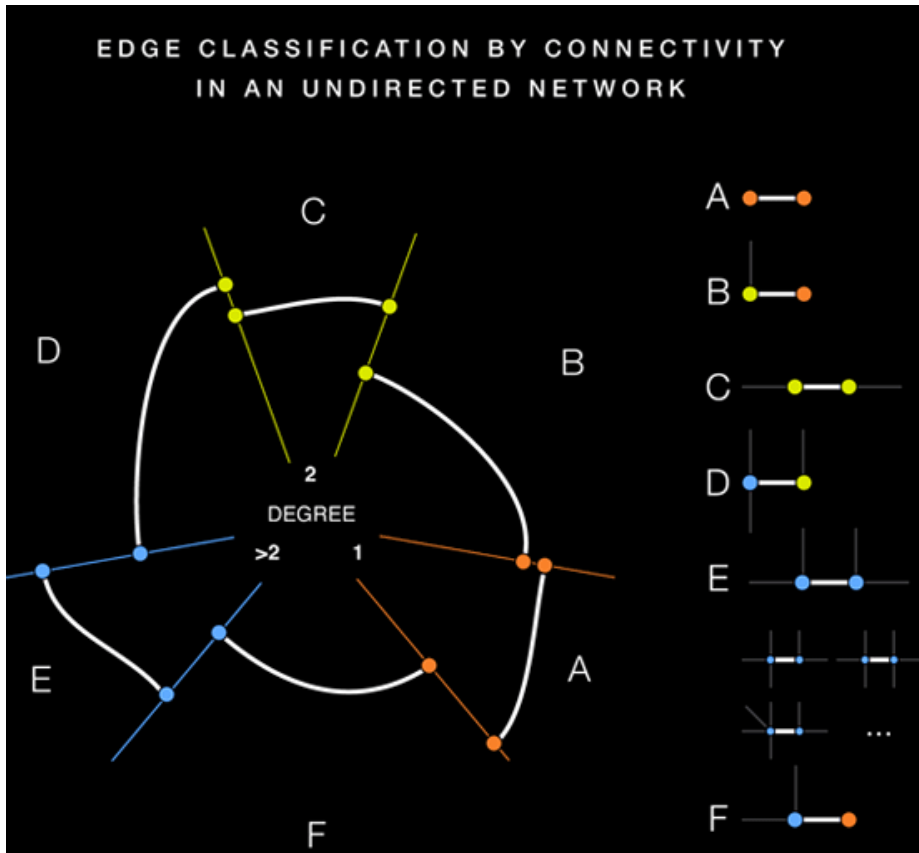
# Visualization options

- One map: vertical, horizontal line; circle
- Two maps: add dot plot
- Three maps: add Hiveplot
- Four maps: ...
- Any number, esp. contig maps: GraphViz  
(see <https://code.google.com/p/canviz>)

# Circos



# Hive Plots



# Gathering Use Cases

- Working group: NCGR, Soybase, Gramene
- Gathering data sets
- <https://github.com/kyclark/drunken-noodle>

# Support



# Acknowledgements

- Doreen Ware
- Steve Cannon
- Rex Nelson
- David Grant
- Andrew Farmer
- Shiran Pasternak
- Andrew Olson
- Jim Thomason

