

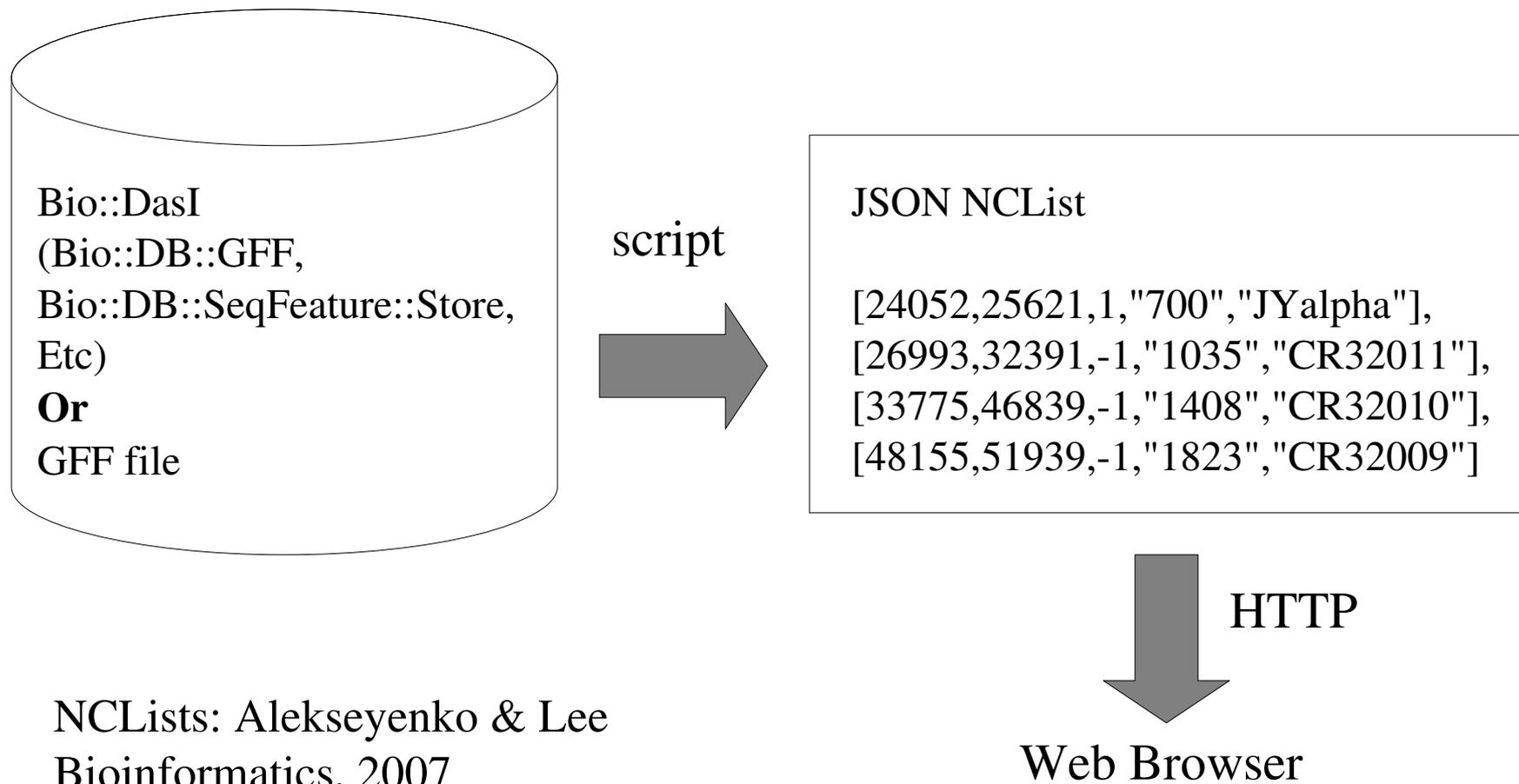
JBrowse

Mitch Skinner
Ian Holmes lab
UC Berkeley

This talk

- How it works; why
- Progress since Toronto
- Walk through setting up browser

How it works (features)



NCLists: Alekseyenko & Lee
Bioinformatics, 2007

Reasons to use pre-generated JSON

- No CGI needed for browsing
 - Easier installation
 - Scale to large genome, large number of users (zoom out to whole chromosome)
- Takes advantage of HTTP caching
 - web browser caches JSON data

Progress since Toronto

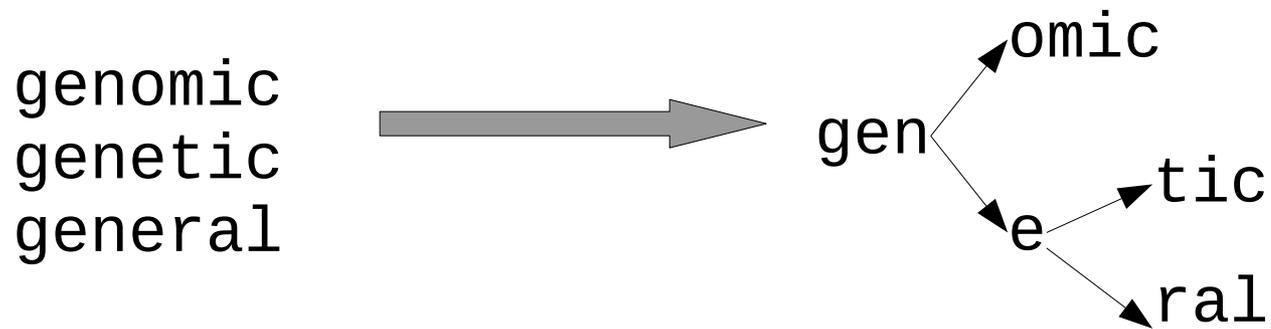
- Name/ID searching
- Quantitative (e.g., wiggle) tracks
- Subfeatures

Name/ID searching

- GBrowse: different search implementation for each data source
- JBrowse: we touch each feature anyway when we generate JSON, so just get names then
 - Afterward, collect name->location mappings into a trie

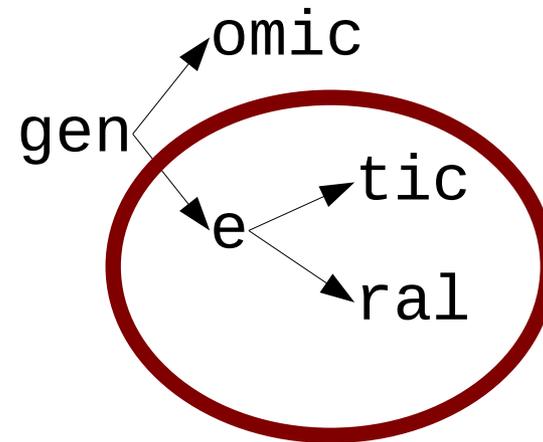
Name/ID searching: Trie

- Shares prefixes among a set of strings



Name/ID searching: Trie

- Subtries are lazily loaded



Quantitative Tracks

- ~12 min to generate tiles for Dmel conservation track (1 data point per base)
- Wiggle tiles compress well
 - ~5 bytes/base, half of which is filesystem overhead

Future

- Lots of things to do
- Question: What's most important to do next?

Thanks

- Ian Holmes, Andrew Uzilov, Lincoln Stein, Scott Cain, Chris Mungall
- NHGRI

Code, demo, wiki:

<http://genome.biowiki.org>