

Composite Datatypes in Galaxy

Ross Lazarus

Introducing SN mapping beta

Galactidulckie = 7

Hiding uninteresting complexity - use cases

- Multi-part datasets
- Complex reports
- Ross provided use cases
- Dan did all the hard work

Multi-part datasets

- EG –bed/bim/fam for genotypes
- Genotypes, map, pedigree
- Must have all 3 components
- Kept together – or broken
- Users generally not interested
- Hide complexity in a subclass
- `lib/galaxy/datatypes/genetics.py`

Subclass the Html datatype

- Derived from Image
- Displays self as web page
- Content can be written by creator
- Legal HTML document
- Links -> files_path
- `The foo plot`
- (extra_files_path in post hook!)

Eg: SNP/WGA lped datatype

The screenshot displays a Galaxy workflow step titled "Analyze Data". The step's output is a list of files generated by the tool "lped_to_pbed_converter.py". A yellow arrow points from the file "CEUYRIchr22small_tdt.pbed" in the left-hand file list to the corresponding entry in the output list.

Navigation: Analyze Data | Workflow | Data Libraries | Visualization | Admin | Help | User

```
## Rgenetics: http://rgenetics.org Galaxy Tools lped_to_pbed_converter.py 22/10/2009 14:13:07
```

1. [hmCEU_YRI_chr22_caco.fam](#)
2. [hmCEU_YRI_chr22_caco.bed](#)
3. [hmCEU_YRI_chr22_caco.log](#)
4. [hmCEU_YRI_chr22_caco.nof](#)
5. [hmCEU_YRI_chr22_caco.bim](#)

File List (Left):

- CEUYRI_null_phenotypes.phe
- 4: CEUYRIchr22small_tdt.pbed
- 3: CEUYRIchr22small_tdt.lped

Benefits

- User sees single history object
- First class datatype
- Tools understand layout
- Look in files_path for components
- Upload, save..should Just Work™
- If not, buy Dan a beer...

Complex outputs => 1 history item

- WGA QC
 - Marker QC
 - Missingness, HWE, freq dist
 - Sample QC
 - Missingness, Fst, (Mendel errors if family)
- (Ancestry PCA, Cryptic relatedness..)
- Goal is to set filter thresholds
- Move to library
- Uniform analysis starting point

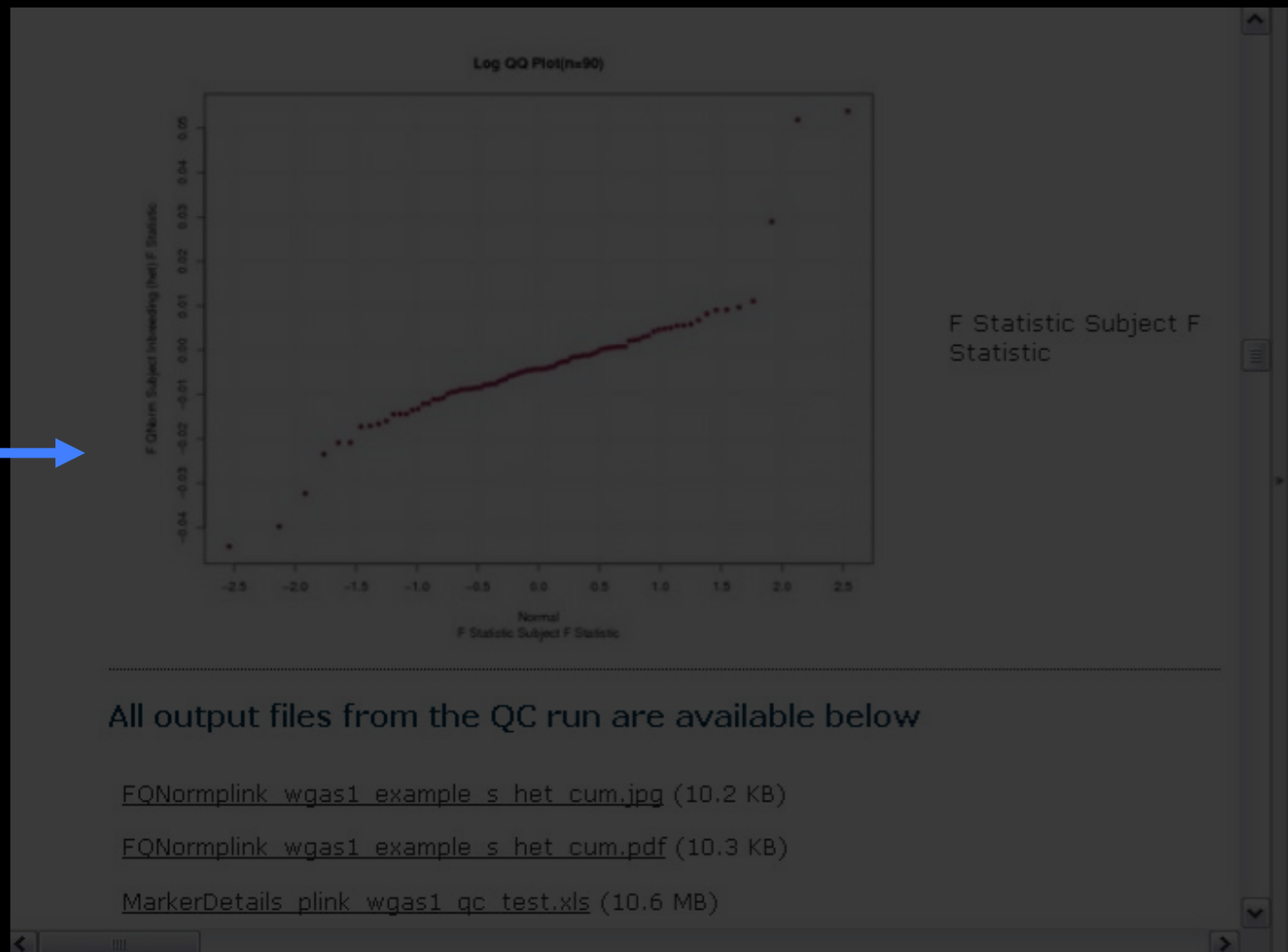
Complex outputs – 1 history item

- Dozens of tab files, pdfs, log outputs
- Individually interesting
- Don't want history cluttered
- Use an html file as output
- Write outputs to files_path
- Generate HTML for Html object content
- Local links -> datasets in files_path

Complexity Hidden

History Options

- 74: http://test.q2.bx.psu.edu/datasets/722f8f9d637b437b/display/MarkerDetails_plink_wgas1_qc
- 73: LD Plots.html
- 68: LD Independent.pbed
- 67: plink wgas1 qc test.html
- 64: LD Plots.html
- 62: LD Plots.html
- 61: Manhattan and QQ plots.html
- 55: http://test.q2.bx.psu.edu/datasets/722f8f9d637b437b/display/MarkerDetails_plink_wgas1_qc





VIVA LA EVOLUCIÓN