

Drosophila Genomics

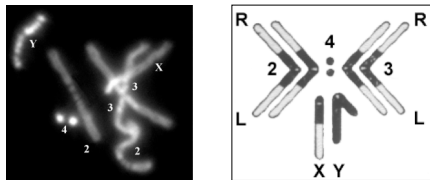
Where are we now?
Where are we going?

Christopher Shaffer, Wilson Leung, Sarah Elgin
Dept of Biology; Washington University in St. Louis

- Inexpensive and easy to culture
- Simple genome, good reference sequence
 - Metazoan: development, behavior, genetic disease
- Many species: population and evolution data

Chris Shaffer; Washington University

Most Drosophila species have a "dot" chromosome



Dot has two parts
 Unbanded Centromere Banded

Chris Shaffer; Washington University

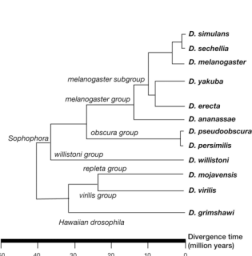
Genomic sequencing

Multiple steps to get to high quality data

1. Produce raw data
 - streamlined; robotic; cheap
 - good coverage; still many gaps; low quality regions
2. Finishing/sequence improvement
 - semi-automated and manual labor
 - expensive, slow

Chris Shaffer; Washington University

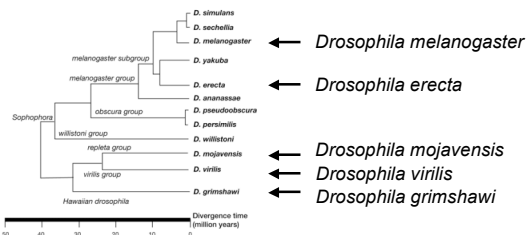
12 Drosophila genomes



12 *Drosophila* genomes have been sequenced. Only *melanogaster* has been "finished". All others have been sequenced to much lower quality using a whole genome shotgun approach without any finishing or sequence improvement; many gaps and low quality regions remain.

Chris Shaffer; Washington University

Research: Focus on Five Species



Chris Shaffer; Washington University

Drosophila grimshawi

- After whole genome shotgun sequencing and assembly:
 - 17440 scaffolds
 - 214,362,671 total bases
 - 5869 gaps
 - 11,750,600 bases
 - Average gap size: 2002 bases

Chris Shaffer, Washington University

D. grimshawi “dot” chromosome

- Need to find which of the scaffolds make up the dot chromosome
 - Observations from our sequencing of *Dvir* suggest most “dot” genes in *Dmel* will stay on the dot through evolution.
 - Do BLAST similarity searches
 - Results identify two scaffolds, 25011 and 24861

Chris Shaffer, Washington University

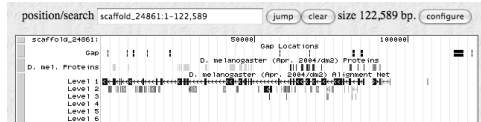
D. grimshawi scaffold 25011



- ~1.0 million bases long
- *Dmel* dot gene similarities in entire scaffold
- This is the first region that will be the focus of the GEP
- 68 gaps; ~3.27% data missing in these gaps
- Low quality? Missassemblies?

Chris Shaffer, Washington University

D. grimshawi scaffold 24861

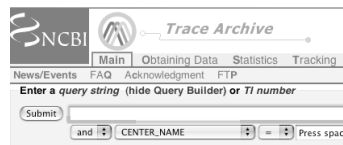


- ~0.1 million bases long
- *Dmel* dot gene similarities in the beginning 94kb
- This is the second region that will be the focus of the GEP
- 11 gaps; ~4.75% data missing in these gaps
- Low quality? Missassemblies?

Chris Shaffer, Washington University

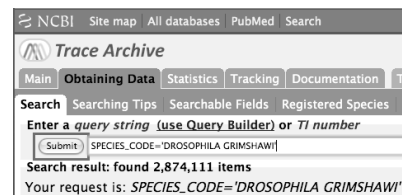
Obtaining data

- All reads generated by the NIH-funded genome centers are deposited in the NCBI trace archive
- <http://www.ncbi.nlm.nih.gov/Traces/>



Chris Shaffer, Washington University

Drosophila grimshawi traces



There are ~3 million reads for *D.grimshawi*

Chris Shaffer, Washington University

The assembly of *D. grimshawi*

- The assembly is also available for download. This includes a file with the position of every read. This is the "reads.placed" file.
- Problem: Trace archive names \neq read names

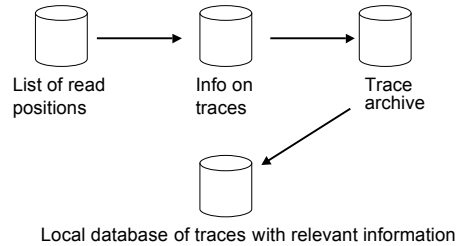
```

Default: ca1Assemblies/dgri (69.25)
ti: 726457678
trace_name: 47303347
contig_name: ABC
species_code: DROSOPHILA GRIMSHAWI
library: WGA
trace_type_code: WGS
submition_type: UPDATE
source_type: GENOMIC
taxid: 7222
center_project: GRANT
clip_quality_left: 11
clip_quality_right: 75
clip_vector_left: 1
clip_vector_right: 75
clone_id: 60028 D.GRIMSHAWI DC-073 P1
target_name: WGS
library_id: D_GRIMSHAWI_RC
.....

```

Chris Shaffer, Washington University

Finding reads for scaffold 25011

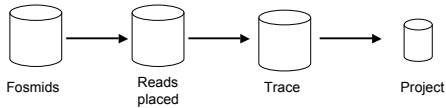


Chris Shaffer, Washington University

Divide and conquer

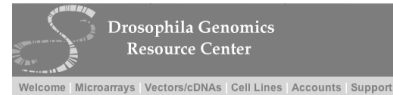
Scaffolds 25011, 24861 assembly of 3kb, 10kb and 40kb clones

41 fosmids (40 kb clones) make a "golden path"



Chris Shaffer, Washington University

Obtain fosmids (DNA)



- Fosmids can be ordered from the Drosophila Genomics Resource Center
- <http://dgrc.cgb.indiana.edu/>

Chris Shaffer, Washington University

Where are we now?

1. Fosmids: DNA ready for sequencing
2. Raw data: "projects" sorted by fosmid
3. Students: learning to finish

Time to do some genomics!

Chris Shaffer, Washington University