Genomics on the Half Shell: Making Science more Open

Steven B. Roberts
Associate Professor
School of Aquatic and Fishery Sciences
University of Washington

robertslab.info

Open Science

- You are free to Share!

- Our lab practices open notebook science

wikispaces wikis for everyone

IP[y]: IPython Interactive Computing

EVERNOTE

- Slides and more available @

oystergen.es/data

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

everything else...

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

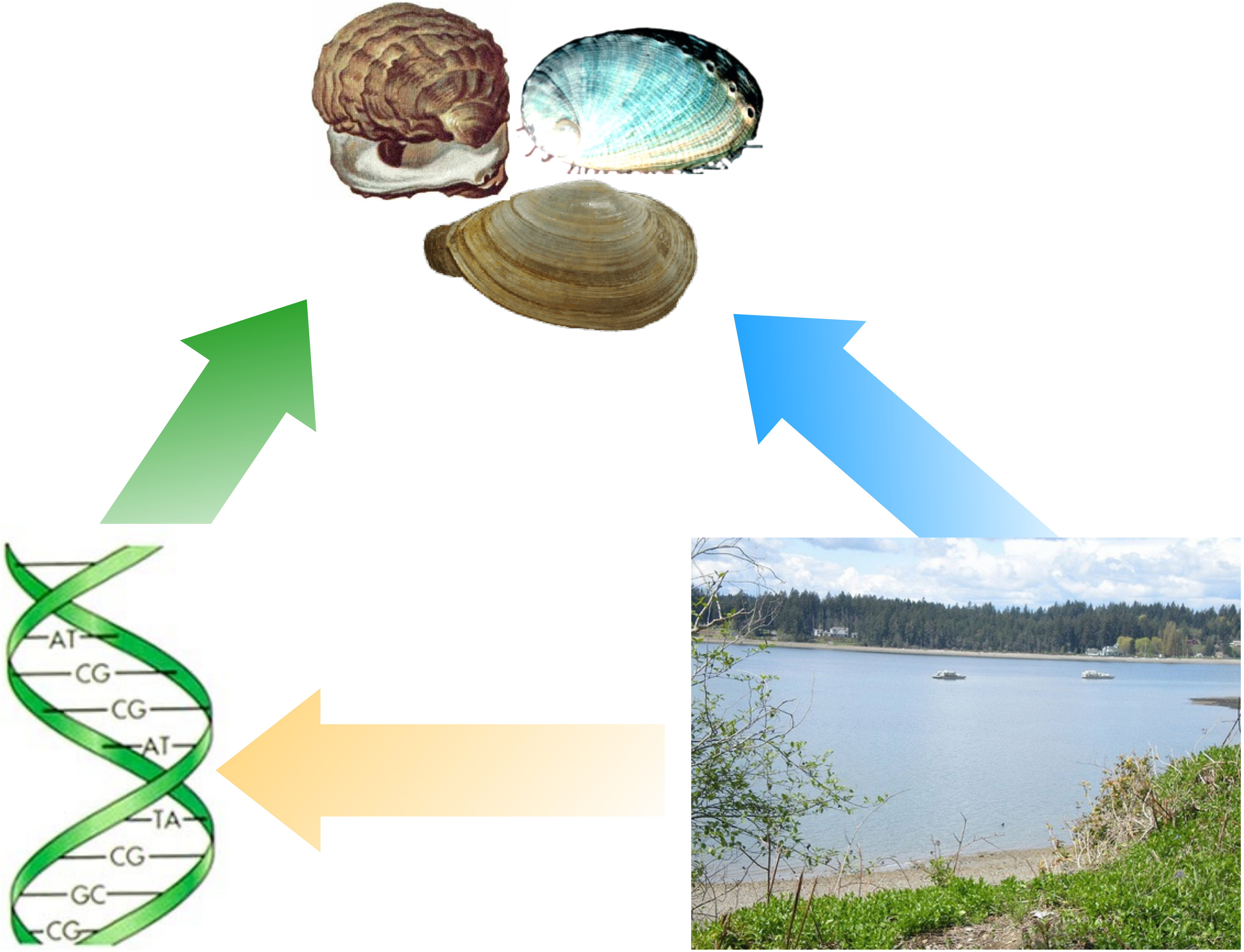
Platforms

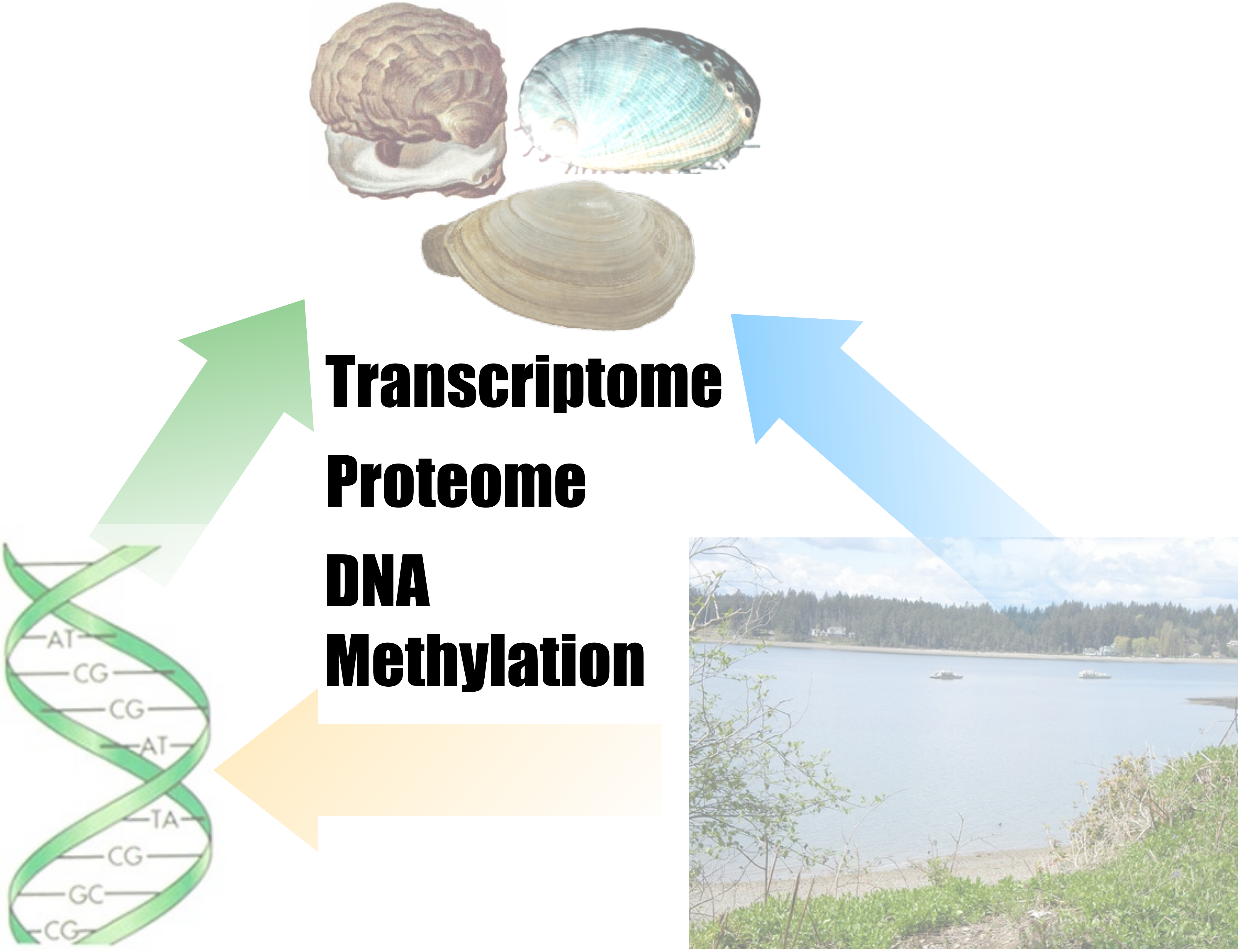
Open Science

Data

everything else...


BIVALVE BANDSTAND
Taylor Shellfish

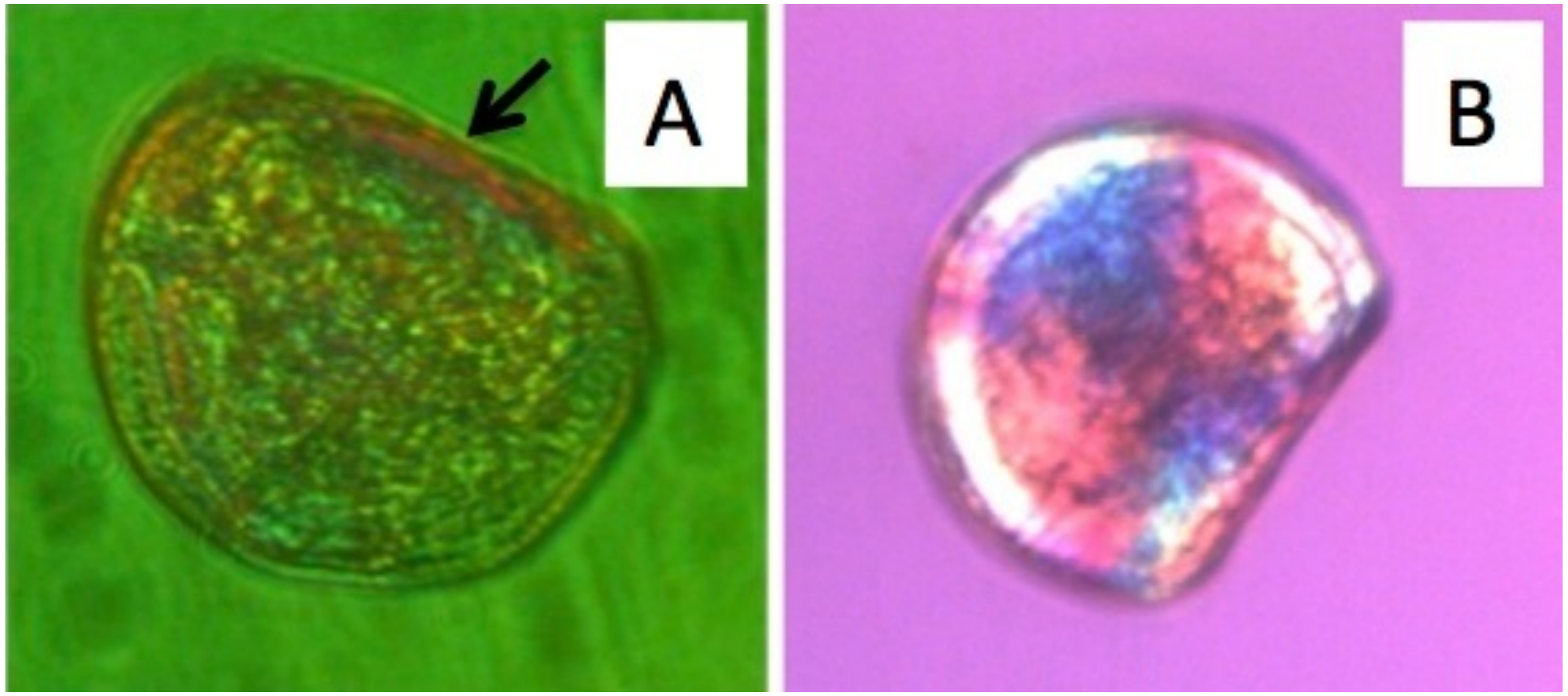

AT
CG
CG
AT
TA
CG
GC
CG



Transcriptome

Proteome

DNA Methylation



Biology

Environment

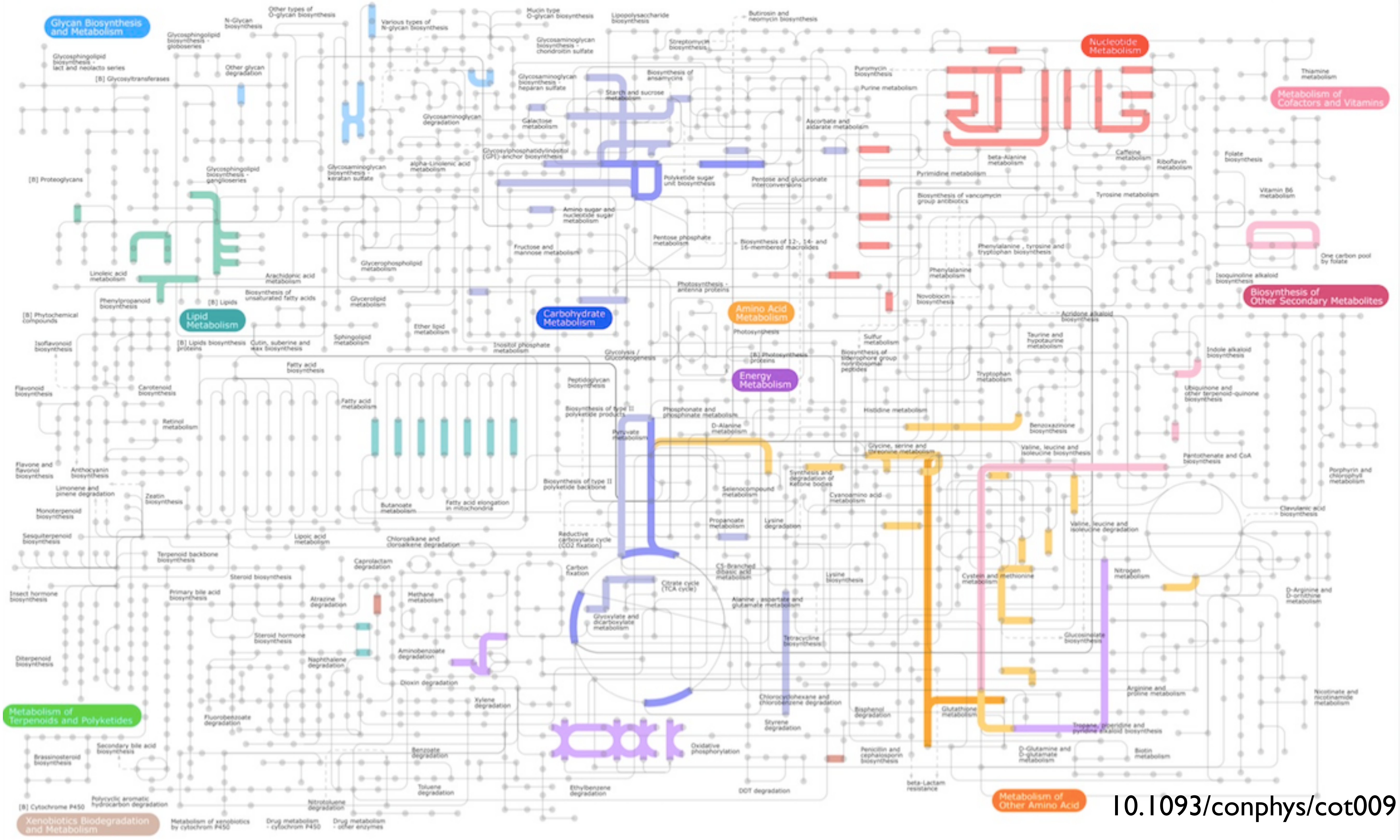
Ocean Acidification

Elevated pCO2 causes developmental delay in early larval Pacific oysters, Crassostrea gigas.
Timmins-Schiffman et al 2012

Shotgun Proteomics

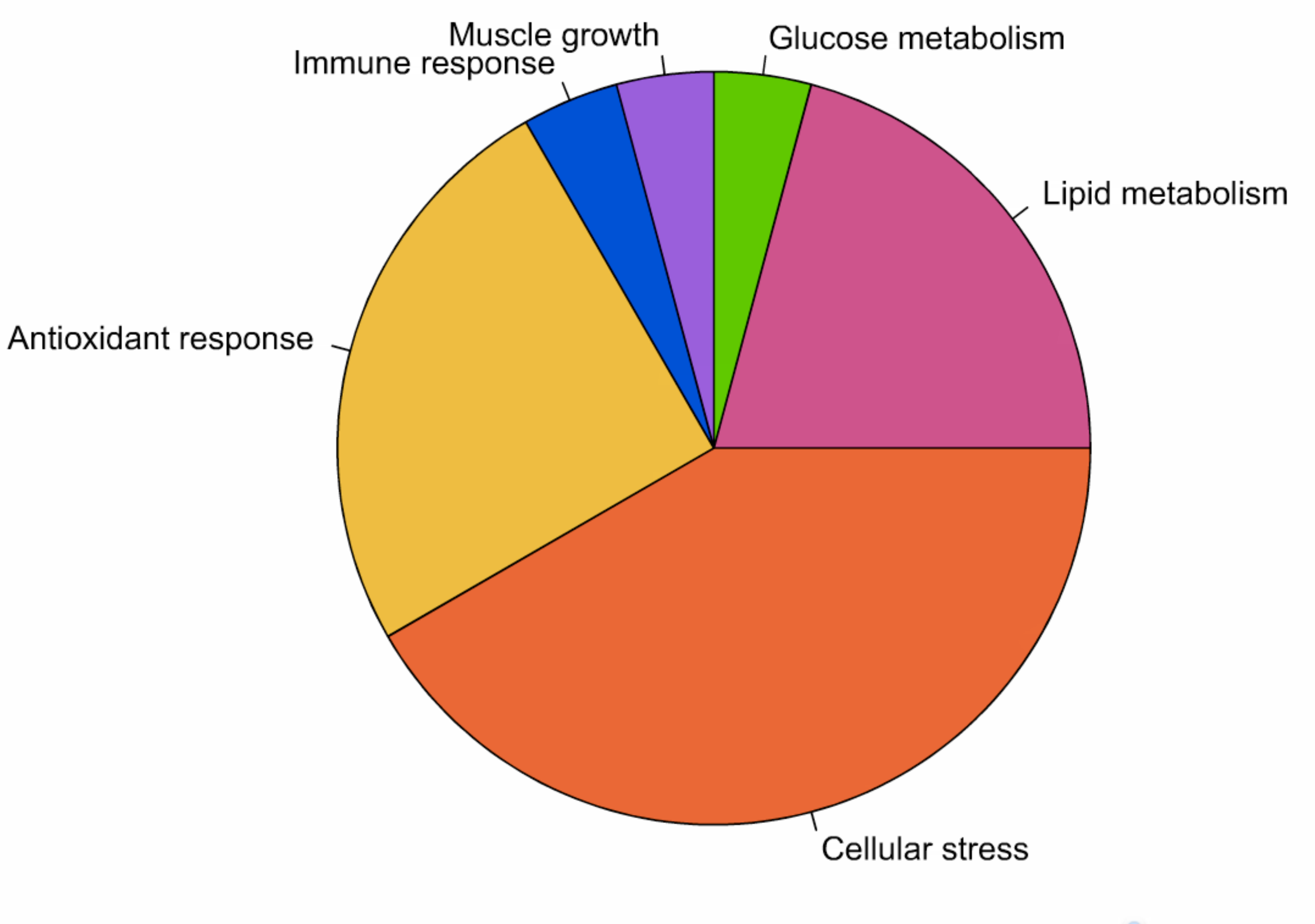
Biology

Environment



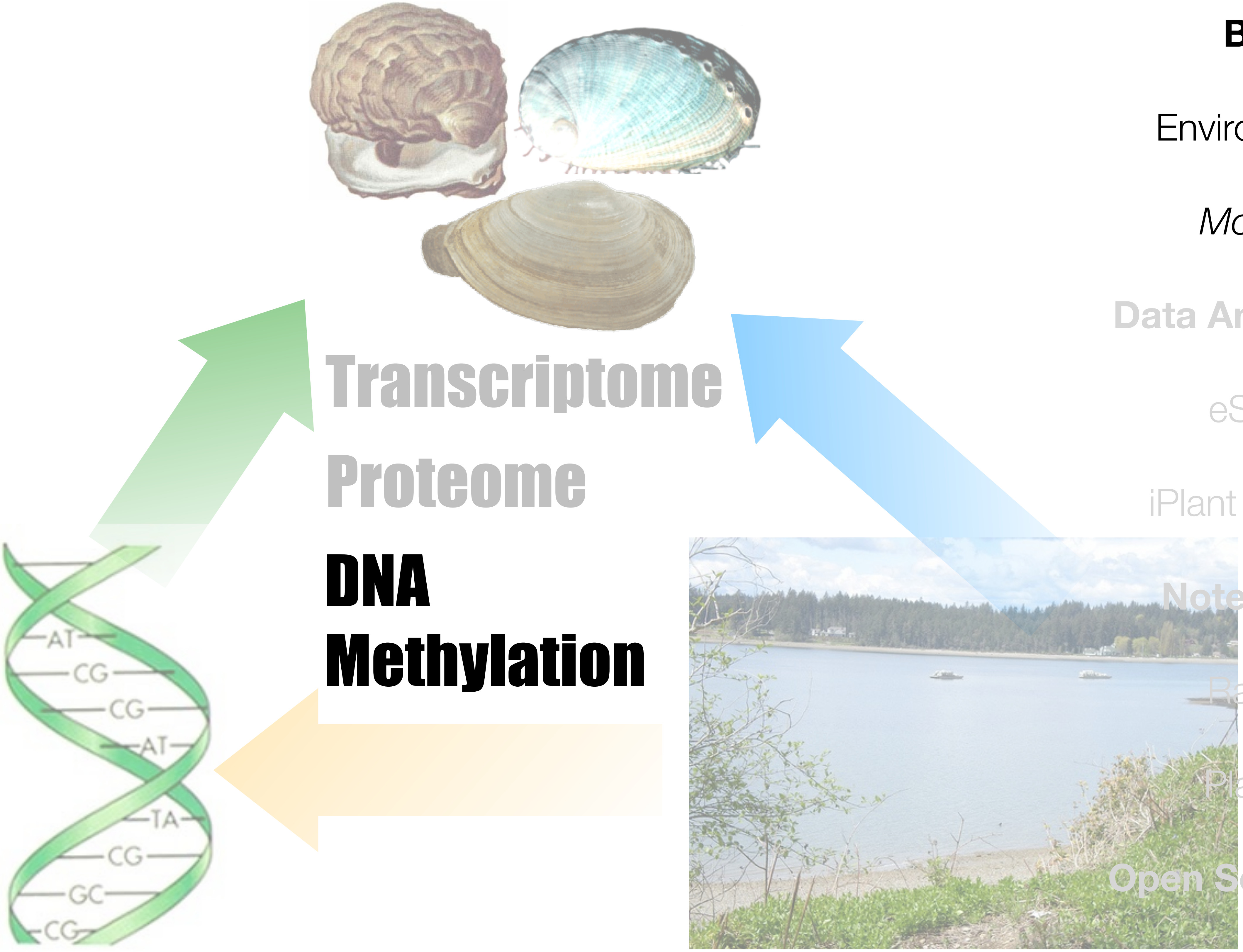
10.1093/conphys/cot009

Shotgun Proteomics



Ocean Acidification

eagle.fish.washington.edu/emma



Transcriptome

Proteome

DNA Methylation

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

everything else...

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

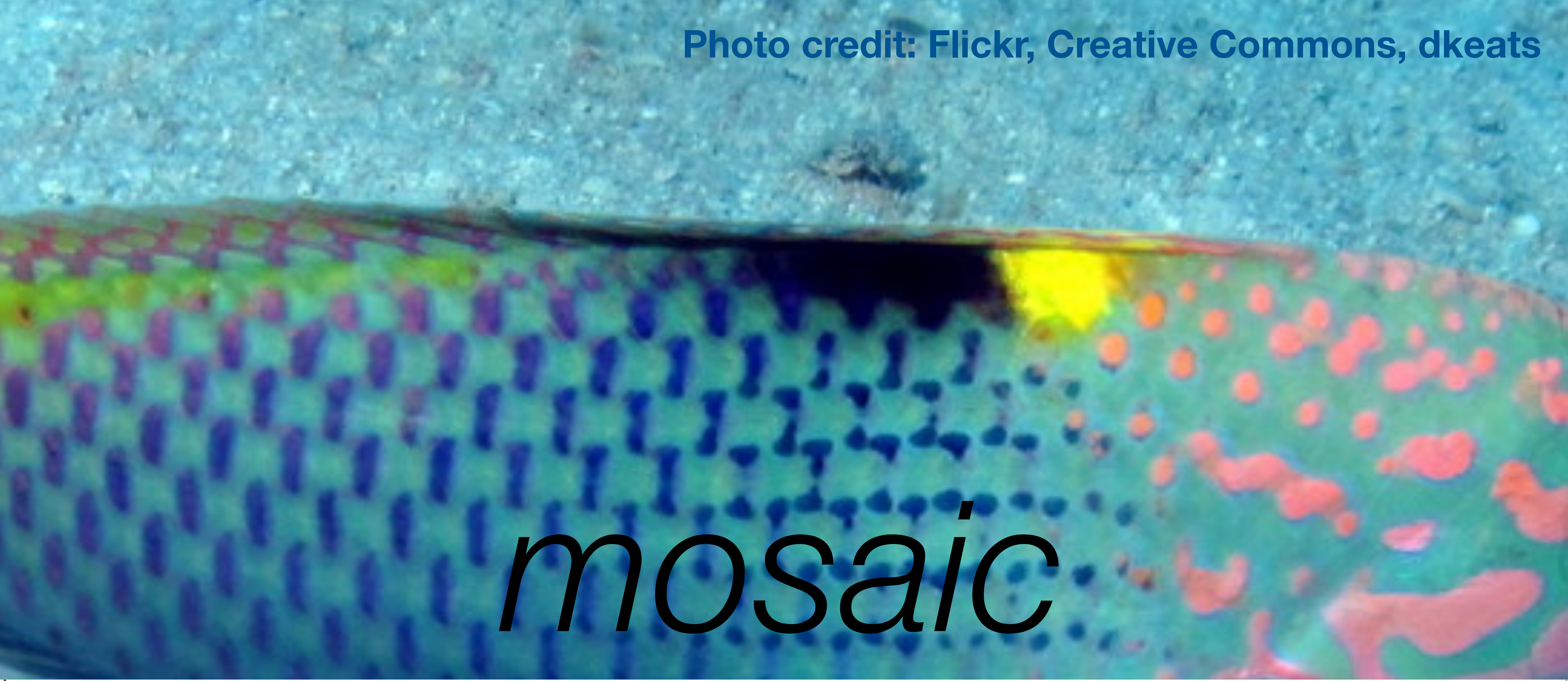
Platforms

Open Science

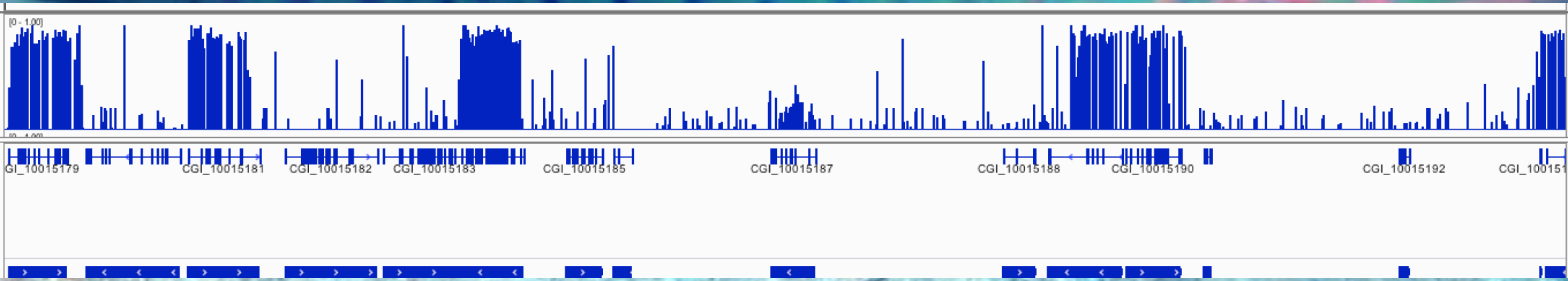
Data

everything else...

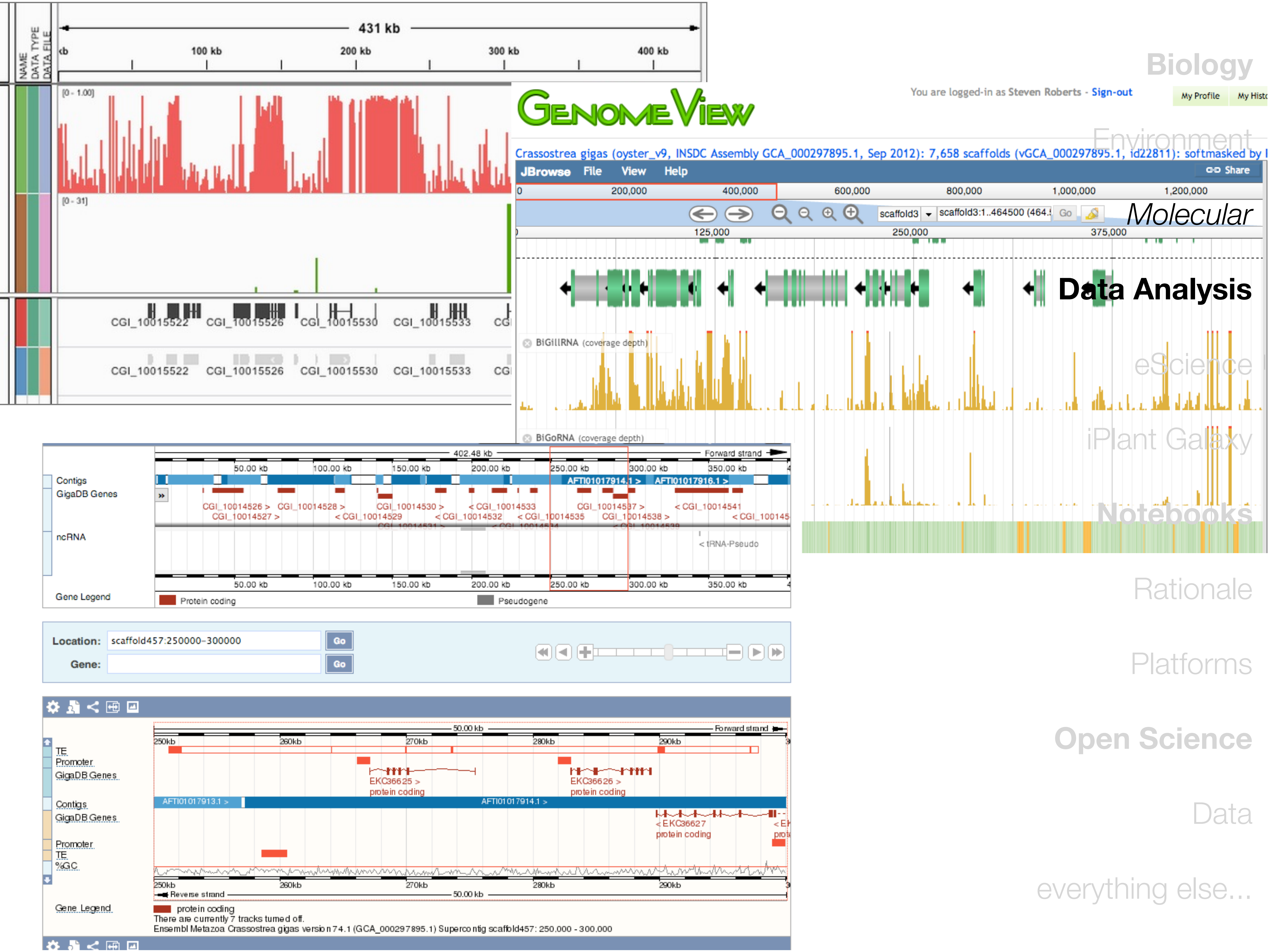
Photo credit: Flickr, Creative Commons, dkeats



mosaic



associated with gene bodies



Crassostrea gigas (oyster_v9, INSDC Assembly GCA_000297895.1, Sep 2012): 7,658 scaffolds (vGCA_000297895.1, id22811): softmasked by I

scaffold3:1..464500 (464.5

Location: scaffold457:250000-300000

There are currently 7 tracks turned off.
Ensembl Metazoa Crassostrea gigas version 74.1 (GCA_000297895.1) Supercontig scaffold457: 250.000 - 300.000

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

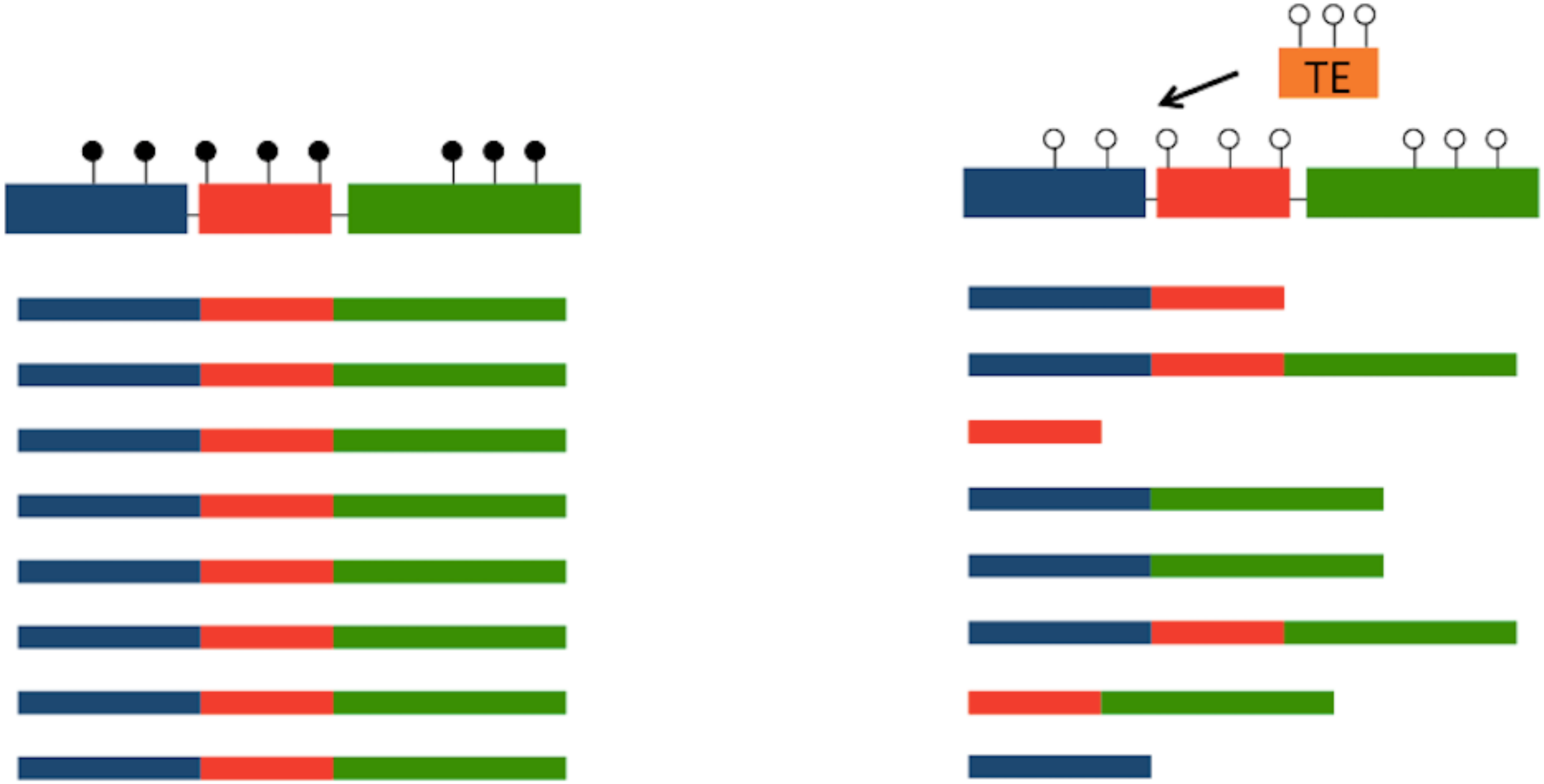
Platforms

Open Science

Data

everything else...

Stochastic Variation



A context dependent role for DNA methylation in bivalves

Mackenzie R. Gavery and Steven B. Roberts

10.1093/bfgp/elt054

10.6084/m9.figshare.880763

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

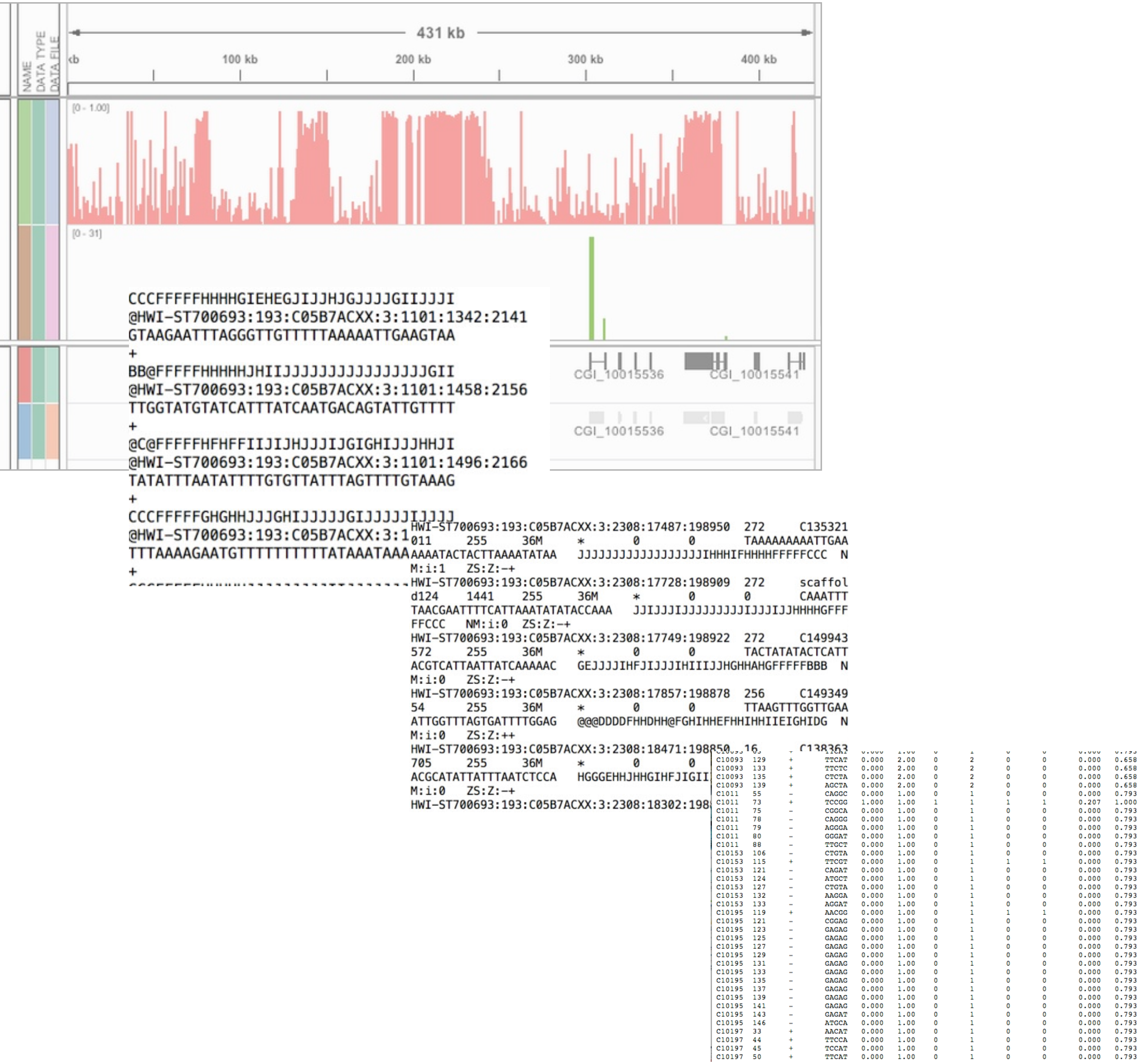
Rationale

Platforms

Open Science

Data

everything else...



```
CCCFFFFFHHHHGIEHEGJIJJHJGJJJJGIIJJJI
@HWI-ST700693:193:C05B7ACXX:3:1101:1342:2141
GTAAGAATTTAGGGTTGTTTTTAAAAATTGAAGTAA
+
BB@FFFFFHHHHHJHIIJJJJJJJJJJJJJJJJGII
@HWI-ST700693:193:C05B7ACXX:3:1101:1458:2156
TTGGTATGTATCATTTATCAATGACAGTATTGTTTT
+
@C@FFFFFHFHFFIIJIJHJJJIJGIGHIJJJHHJI
@HWI-ST700693:193:C05B7ACXX:3:1101:1496:2166
TATATTTAATATTTTGTGTTATTTAGTTTTGTAAAG
+
CCCFFFFFGHGHHJJJGHIJJJJJGIJJJJJIJJJJ
@HWI-ST700693:193:C05B7ACXX:3:1
TTTAAAAGAATGTTTTTTTTTTATAAATAAA
+
```

```
HWI-ST700693:193:C05B7ACXX:3:2308:17487:198950  272     C135321
011     255     36M     *       0       0       TAAAAAAAAATTGAA
AAAATACTACTTAAAATATAA   JJJJJJJJJJJJJJJJJIHHHIFHHHHFFFFFCCC  N
M:i:1   ZS:Z:-+
HWI-ST700693:193:C05B7ACXX:3:2308:17728:198909  272     scaffol
d124    1441    255     36M     *       0       0       CAAATTT
TAACGAATTTTCATTAAATATATACCAAA   JJIJJJIJJJJJJJJIJJJIJJHHHHGFFF
FFCCC   NM:i:0  ZS:Z:-+
HWI-ST700693:193:C05B7ACXX:3:2308:17749:198922  272     C149943
572     255     36M     *       0       0       TACTATATACTCATT
ACGTCATTAATTATCAAAAAC   GEJJJJIHFJIJJJIHIIIJJHGHHAHGFFFFFBBB  N
M:i:0   ZS:Z:-+
HWI-ST700693:193:C05B7ACXX:3:2308:17857:198878  256     C149349
54      255     36M     *       0       0       TTAAGTTTGGTTGAA
ATTGGTTTAGTGATTTTGGAG   @@@DDDDFHHDHH@FGHIHHEFHHIHHIIEIGHIDG  N
M:i:0   ZS:Z:++
HWI-ST700693:193:C05B7ACXX:3:2308:18471:198850  16      C138363
705     255     36M     *       0       0
ACGCATATTATTTAATCTCCA   HGGGEHHJHHGIHFJIGII
M:i:0   ZS:Z:-+
HWI-ST700693:193:C05B7ACXX:3:2308:18302:198
```

```
C10093  129   +   TTCAT  0.000  2.00  0  2  0  0  0.000  0.658
C10093  133   +   TTCTC  0.000  2.00  0  2  0  0  0.000  0.658
C10093  135   +   CTCTA  0.000  2.00  0  2  0  0  0.000  0.658
C10093  139   +   AGCTA  0.000  2.00  0  2  0  0  0.000  0.658
C1011   55    -   CAGGC  0.000  1.00  0  1  0  0  0.000  0.793
C1011   73    +   TCCGG  1.000  1.00  1  1  1  1  0.207  1.000
C1011   75    -   CGGCA  0.000  1.00  0  1  0  0  0.000  0.793
C1011   78    -   CAGGG  0.000  1.00  0  1  0  0  0.000  0.793
C1011   79    -   AGGGA  0.000  1.00  0  1  0  0  0.000  0.793
C1011   80    -   GGGAT  0.000  1.00  0  1  0  0  0.000  0.793
C1011   88    -   TTGCT  0.000  1.00  0  1  0  0  0.000  0.793
C10153  106   -   CTGTA  0.000  1.00  0  1  0  0  0.000  0.793
C10153  115   +   TTCGT  0.000  1.00  0  1  1  1  0.000  0.793
C10153  121   -   CAGAT  0.000  1.00  0  1  0  0  0.000  0.793
C10153  124   -   ATGCT  0.000  1.00  0  1  0  0  0.000  0.793
C10153  127   -   CTGTA  0.000  1.00  0  1  0  0  0.000  0.793
C10153  132   -   AAGGA  0.000  1.00  0  1  0  0  0.000  0.793
C10153  133   -   AGGAT  0.000  1.00  0  1  0  0  0.000  0.793
C10195  119   +   AACGG  0.000  1.00  0  1  1  1  0.000  0.793
C10195  121   -   CGGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  123   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  125   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  127   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  129   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  131   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  133   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  135   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  137   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  139   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  141   -   GAGAG  0.000  1.00  0  1  0  0  0.000  0.793
C10195  143   -   GAGAT  0.000  1.00  0  1  0  0  0.000  0.793
C10195  146   -   ATGCA  0.000  1.00  0  1  0  0  0.000  0.793
C10197  33    +   AACAT  0.000  1.00  0  1  0  0  0.000  0.793
C10197  44    +   TTCCA  0.000  1.00  0  1  0  0  0.000  0.793
C10197  45    +   TCCAT  0.000  1.00  0  1  0  0  0.000  0.793
C10197  50    +   TTCAT  0.000  1.00  0  1  0  0  0.000  0.793
```

Biology
Environment
Molecular
Data Analysis
eScience
iPlant Galaxy
Notebooks
Rationale
Platforms
Open Science
Data
everything else...

```
CCCFFFFFHHHHGIEHEGJI
@HWI-ST700693:193:C0
GTAAGAATTTAGGGTTGTTT
+
BB@FFFFFHHHHHJHIIJJJ
@HWI-ST700693:193:C0
TTGGTATGTATCATTTATCA
+
@C@FFFFFHFHFFIIJIJHJ
@HWI-ST700693:193:C0
TATATTTAATATTTTGTGTT
+
CCCFFFFFGHGHHJJJGHIJ
@HWI-ST700693:193:C0
TTTAAAAGAATGTTTTTTTT
+
```

Peter Cock @pjacock

Or “advanced file format conversion”? RT @pathogenomenick: Bioinformatics… Or ‘advanced file copying’ as I like to call it.

RETWEETS 4 FAVORITES 2

11:53 AM - 29 Jan 2014

John Chilton @jmchilton · Jan 29
@pjacock @pathogenomenick I'm jealous, all I've ever done is convert & copy but I've never succeeded to the point I'd call it advanced.

Nick Loman @pathogenomenick · Jan 29
@jmchilton @pjacock today it was all about the copying .. but I had to use all my skills!

Peter Cock @pjacock · Jan 29
@jmchilton @pathogenomenick if you have to write your own parser & output code, especially for binary formats, then I'd call it advanced ;)

Chris Cole @drchriscole · Jan 30
@pjacock @pathogenomenick I prefer, advanced data compression. Here's your 80gig of fastqs reduced to a bunch R plots.

```
0.000   0.793
0.000   0.658
0.000   0.658
0.000   0.658
0.000   0.658
0.000   0.793
0.207   1.000
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
0.000   0.793
```

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

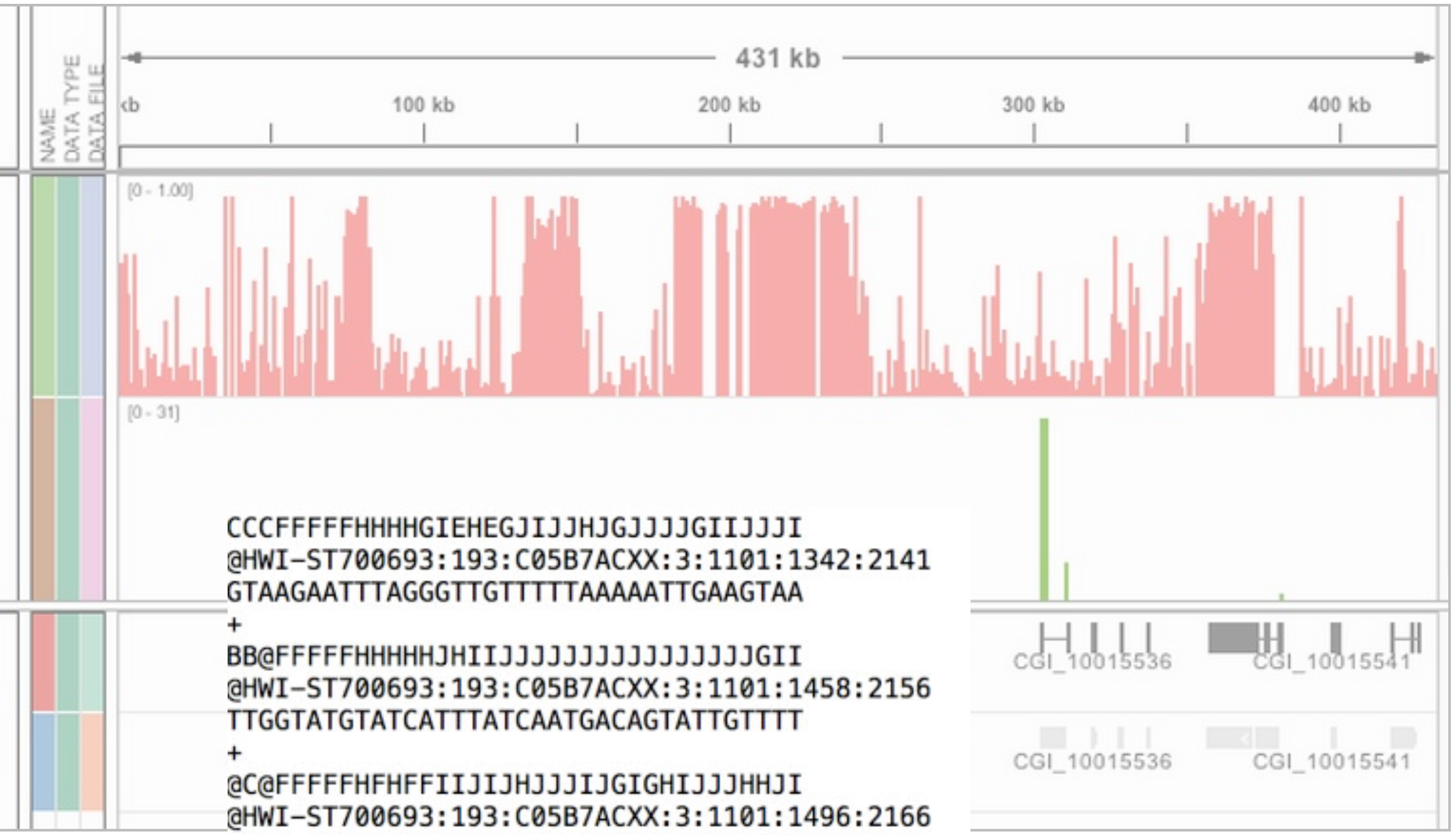
Rationale

Platforms

Open Science

Data

everything else…



```
CCCFFFFFHHHHGIEHEGJIJJHJGJJJGIIJJJI
@HWI-ST700693:193:C05B7ACXX:3:1101:1342:2141
GTAAGAATTTAGGGTTGTTTTTAAAAATTGAAGTAA
+
BB@FFFFFHHHHHJHIIJJJJJJJJJJJJJJJJJGII
@HWI-ST700693:193:C05B7ACXX:3:1101:1458:2156
TTGGTATGTATCATTTATCAATGACAGTATTGTTTT
+
@C@FFFFFHFHFFIIJIJHJJJIJGIGHIJJJHHJI
@HWI-ST700693:193:C05B7ACXX:3:1101:1496:2166
TATATTTAATATTTTGTGTTATTTAGTTTTGTAAAG
+
CCCFFFFFGHGHHJJJGHIJJJJJGIJJJJJIJJJJ
@HWI-ST700693:193:C05B7ACXX:3:1
TTTAAAAGAATGTTTTTTTTTTATAAATAAA
+
```

```
HWI-ST700693:193:C05B7ACXX:3:2308:17487:198950  272     C135321
011     255     36M     *       0       0       TAAAAAAAAATTGAA
AAAATACTACTTAAAATATAA   JJJJJJJJJJJJJJJJJJIHHHIFHHHHFFFFFCCC   N
M:i:1   ZS:Z:-+
HWI-ST700693:193:C05B7ACXX:3:2308:17728:198909  272     scaffol
d124    1441    255     36M     *       0       0       CAAATTT
TAACGAATTTTCATTAAATATATACCAAA   JJIJJJIJJJJJJJJJIJJJIJJHHHHGFFF
FFCCC   NM:i:0  ZS:Z:-+
HWI-ST700693:193:C05B7ACXX:3:2308:17749:198922  272     C149943
572     255     36M     *       0       0       TACTATATACTCATT
ACGTCATTAATTATCAAAAAC   GEJJJJIHFJIJJJIHIIIJJHGHHAHGFFFFFBBB   N
M:i:0   ZS:Z:-+
HWI-ST700693:193:C05B7ACXX:3:2308:17857:198878  256     C149349
54      255     36M     *       0       0       TTAAGTTTGGTTGAA
ATTGGTTTAGTGATTTTGGAG   @@@DDDDFHHDHH@FGHIHHEFHHIHHIIEIGHIDG   N
M:i:0   ZS:Z:++
HWI-ST700693:193:C05B7ACXX:3:2308:18471:198850  16      C138363
705     255     36M     *       0       0
ACGCATATTATTTAATCTCCA   HGGGEHHJHHGIHFJIGII
M:i:0   ZS:Z:-+
HWI-ST700693:193:C05B7ACXX:3:2308:18302:198
```

```
C10093  129     +       TTCAT   0.000   2.00    0       2       0       0       0.000   0.658
C10093  133     +       TTCTC   0.000   2.00    0       2       0       0       0.000   0.658
C10093  135     +       CTCTA   0.000   2.00    0       2       0       0       0.000   0.658
C10093  139     +       AGCTA   0.000   2.00    0       2       0       0       0.000   0.658
C1011   55      -       CAGGC   0.000   1.00    0       1       0       0       0.000   0.793
C1011   73      +       TCCGG   1.000   1.00    1       1       1       1       0.207   1.000
C1011   75      -       CGGCA   0.000   1.00    0       1       0       0       0.000   0.793
C1011   78      -       CAGGG   0.000   1.00    0       1       0       0       0.000   0.793
C1011   79      -       AGGGA   0.000   1.00    0       1       0       0       0.000   0.793
C1011   80      -       GGGAT   0.000   1.00    0       1       0       0       0.000   0.793
C1011   88      -       TTGCT   0.000   1.00    0       1       0       0       0.000   0.793
C10153  106     -       CTGTA   0.000   1.00    0       1       0       0       0.000   0.793
C10153  115     +       TTCGT   0.000   1.00    0       1       1       1       0.000   0.793
C10153  121     -       CAGAT   0.000   1.00    0       1       0       0       0.000   0.793
C10153  124     -       ATGCT   0.000   1.00    0       1       0       0       0.000   0.793
C10153  127     -       CTGTA   0.000   1.00    0       1       0       0       0.000   0.793
C10153  132     -       AAGGA   0.000   1.00    0       1       0       0       0.000   0.793
C10153  133     -       AGGAT   0.000   1.00    0       1       0       0       0.000   0.793
C10195  119     +       AACGG   0.000   1.00    0       1       1       1       0.000   0.793
C10195  121     -       CGGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  123     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  125     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  127     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  129     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  131     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  133     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  135     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  137     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  139     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  141     -       GAGAG   0.000   1.00    0       1       0       0       0.000   0.793
C10195  143     -       GAGAT   0.000   1.00    0       1       0       0       0.000   0.793
C10195  146     -       ATGCA   0.000   1.00    0       1       0       0       0.000   0.793
C10197  33      +       AACAT   0.000   1.00    0       1       0       0       0.000   0.793
C10197  44      +       TTCCA   0.000   1.00    0       1       0       0       0.000   0.793
C10197  45      +       TCCAT   0.000   1.00    0       1       0       0       0.000   0.793
C10197  50      +       TTCAT   0.000   1.00    0       1       0       0       0.000   0.793
```

raw - 70G

mapping - 60G

tables - 40G ....

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

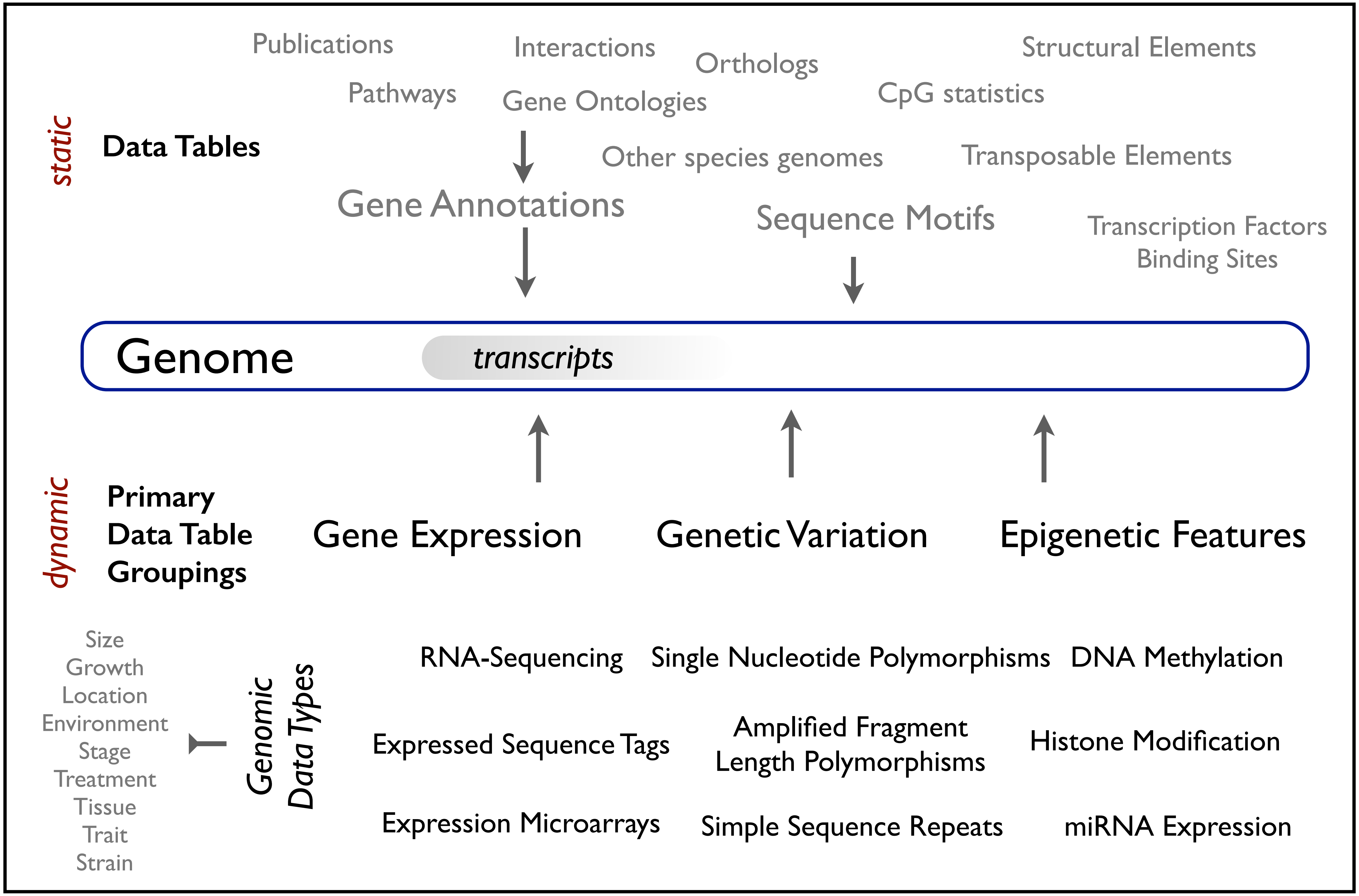
Rationale

Platforms

Open Science

Data

everything else...



static **Data Tables**

Publications
Pathways
Interactions
Gene Ontologies
Orthologs
Other species genomes
Structural Elements
CpG statistics
Transposable Elements
Transcription Factors Binding Sites

Gene Annotations
Sequence Motifs

Genome *transcripts*

dynamic **Primary Data Table Groupings**

Gene Expression
Genetic Variation
Epigenetic Features

Size
Growth
Location
Environment
Stage
Treatment
Tissue
Trait
Strain

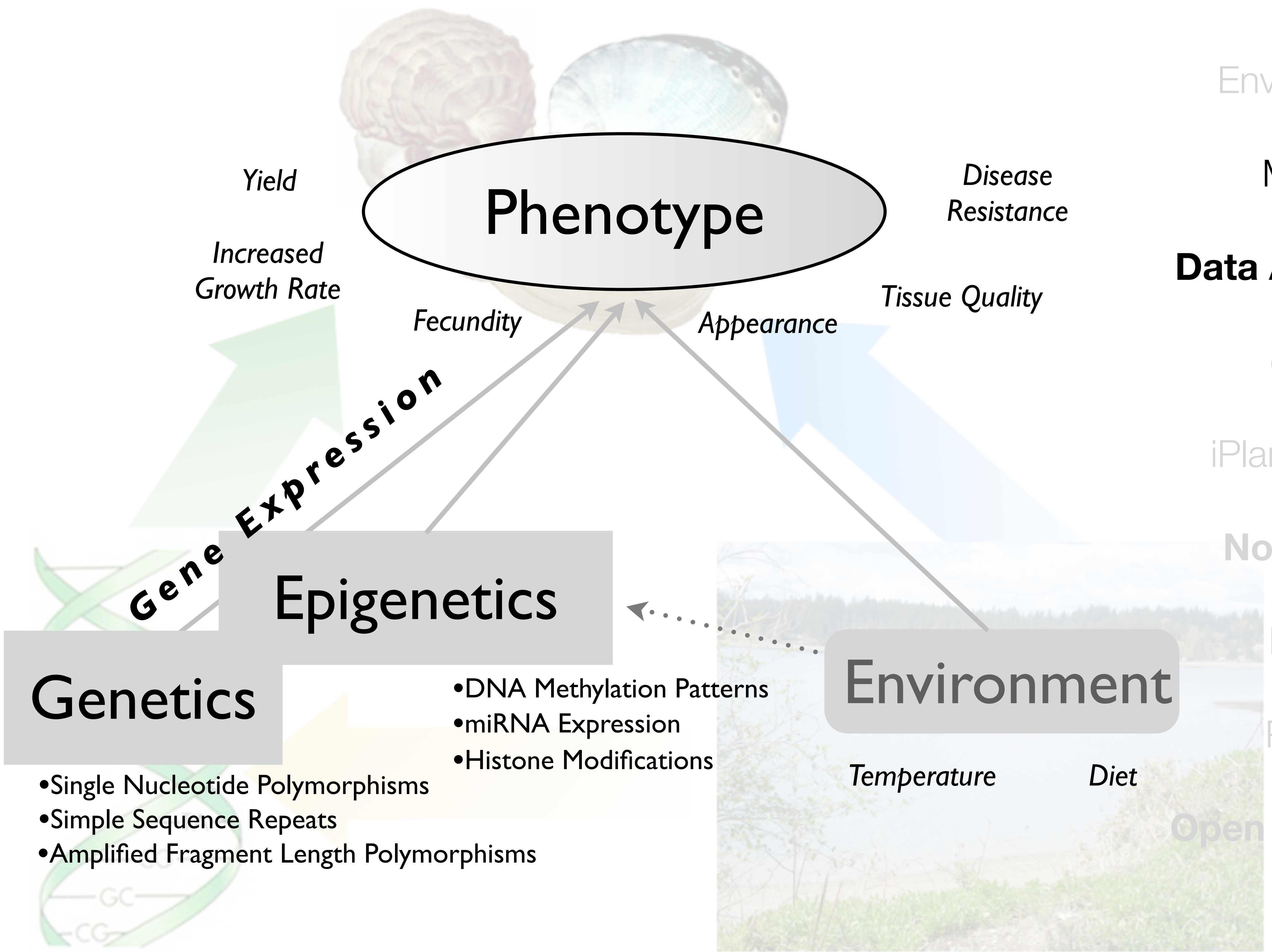
Genomic Data Types

RNA-Sequencing
Expressed Sequence Tags
Expression Microarrays

Single Nucleotide Polymorphisms
Amplified Fragment Length Polymorphisms
Simple Sequence Repeats

DNA Methylation
Histone Modification
miRNA Expression

Biology
Environment
Molecular
Data Analysis
eScience
iPlant Galaxy
Notebooks
Rationale
Platforms
Open Science
Data
everything else...



Phenotype

Yield

Increased Growth Rate

Fecundity

Appearance

Disease Resistance

Tissue Quality

Gene Expression

Epigenetics

- DNA Methylation Patterns
- miRNA Expression
- Histone Modifications

Genetics

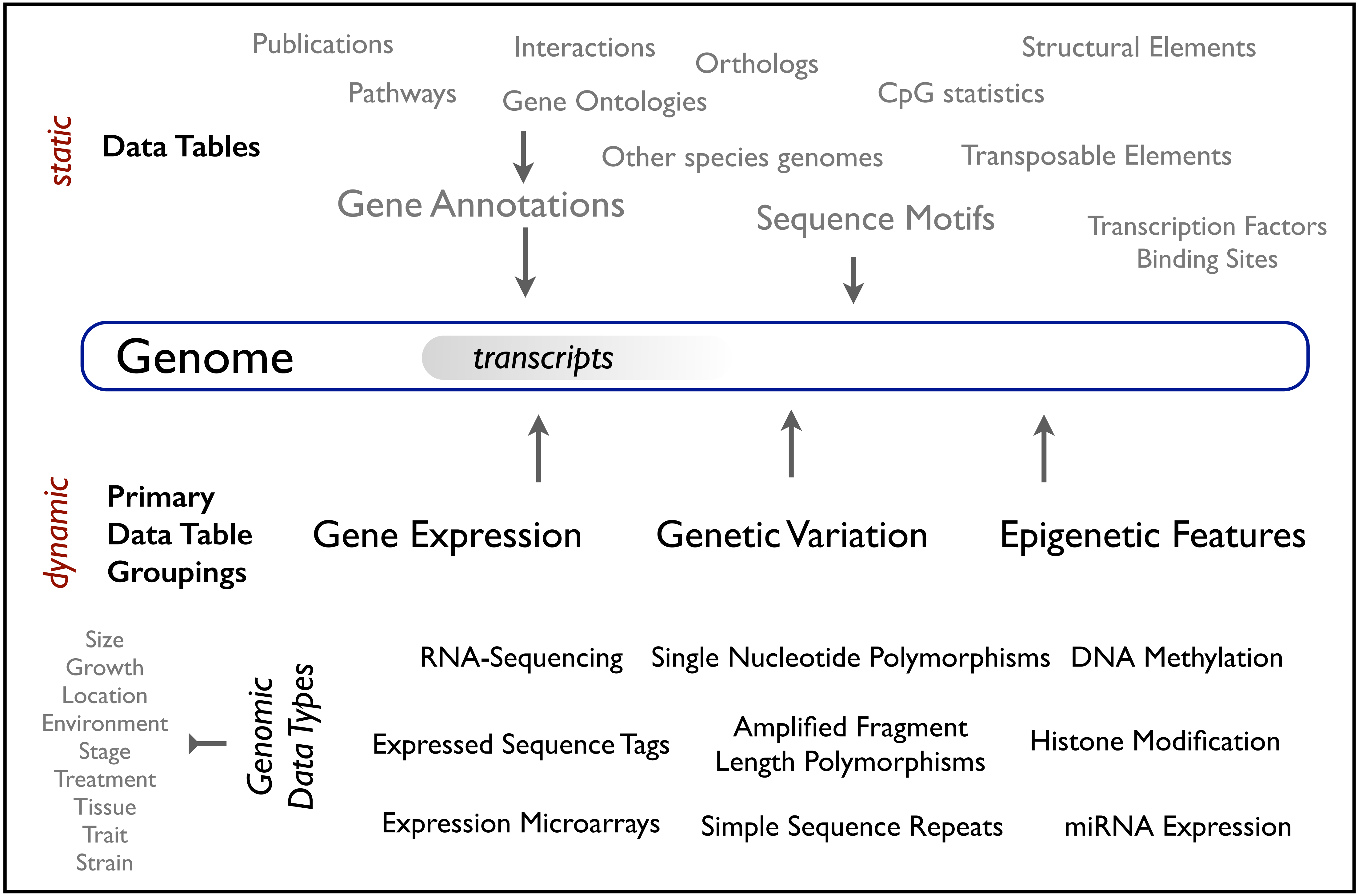
- Single Nucleotide Polymorphisms
- Simple Sequence Repeats
- Amplified Fragment Length Polymorphisms

Environment

Temperature

Diet

Biology
Environment
Molecular
Data Analysis
eScience
iPlant Galaxy
Notebooks
Rationale
Platforms
Open Science
Data
everything else...



Publications

Pathways

Interactions

Gene Ontologies

Orthologs

Structural Elements

CpG statistics

static

Data Tables

Other species genomes

Transposable Elements

Gene Annotations

Sequence Motifs

Transcription Factors
Binding Sites

Genome

transcripts

dynamic

**Primary
Data Table
Groupings**

Gene Expression

Genetic Variation

Epigenetic Features

Size
Growth
Location
Environment
Stage
Treatment
Tissue
Trait
Strain

*Genomic
Data Types*

RNA-Sequencing

Expressed Sequence Tags

Expression Microarrays

Single Nucleotide Polymorphisms

Amplified Fragment
Length Polymorphisms

Simple Sequence Repeats

DNA Methylation

Histone Modification

miRNA Expression

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

everything else...

Querying Disparate Oyster Datasets | qDOD

The goal of this project is to produce a web-based interface for querying and visualizing *Crassostrea gigas* genomic datasets. This site serves as a portal for documenting our efforts, providing user access, as well as a means to gather feedback.

DATA

Preliminary Phase: Aggregating Datasets

Using SQLShare as a platform we have already begun to aggregate and format data. Anyone can view (and contribute) using the tag "qdod". Below is a table describing some of the relevant datasets. "Snapshot" provides you with a screenshot of the data in SQLShare and "Direct Link" brings you directly to the data in SQLShare. You can also open the table in a new webpage.

qDOD online

qDOD_Cgigas_gene_fasta	sequence fasta file. Exon only.	http://goo.gl/ogCxl	https://sqlshare.esc
qDOD_Zhang_Gil_gene_RNA-seq	Gill RNA-seq data (gene based)	http://goo.gl/8oISR	https://sqlshare.esc seq
qDOD_Zhang_Mgo_gene_RNA-seq	Male Gonad RNA-seq data (gene based)	http://goo.gl/6buVz	https://sqlshare.esc seq

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

everything else...

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

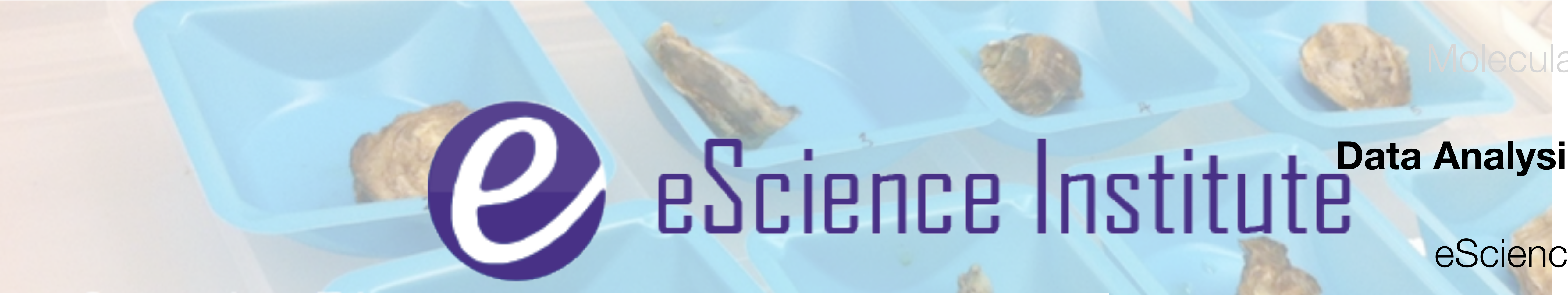
Rationale

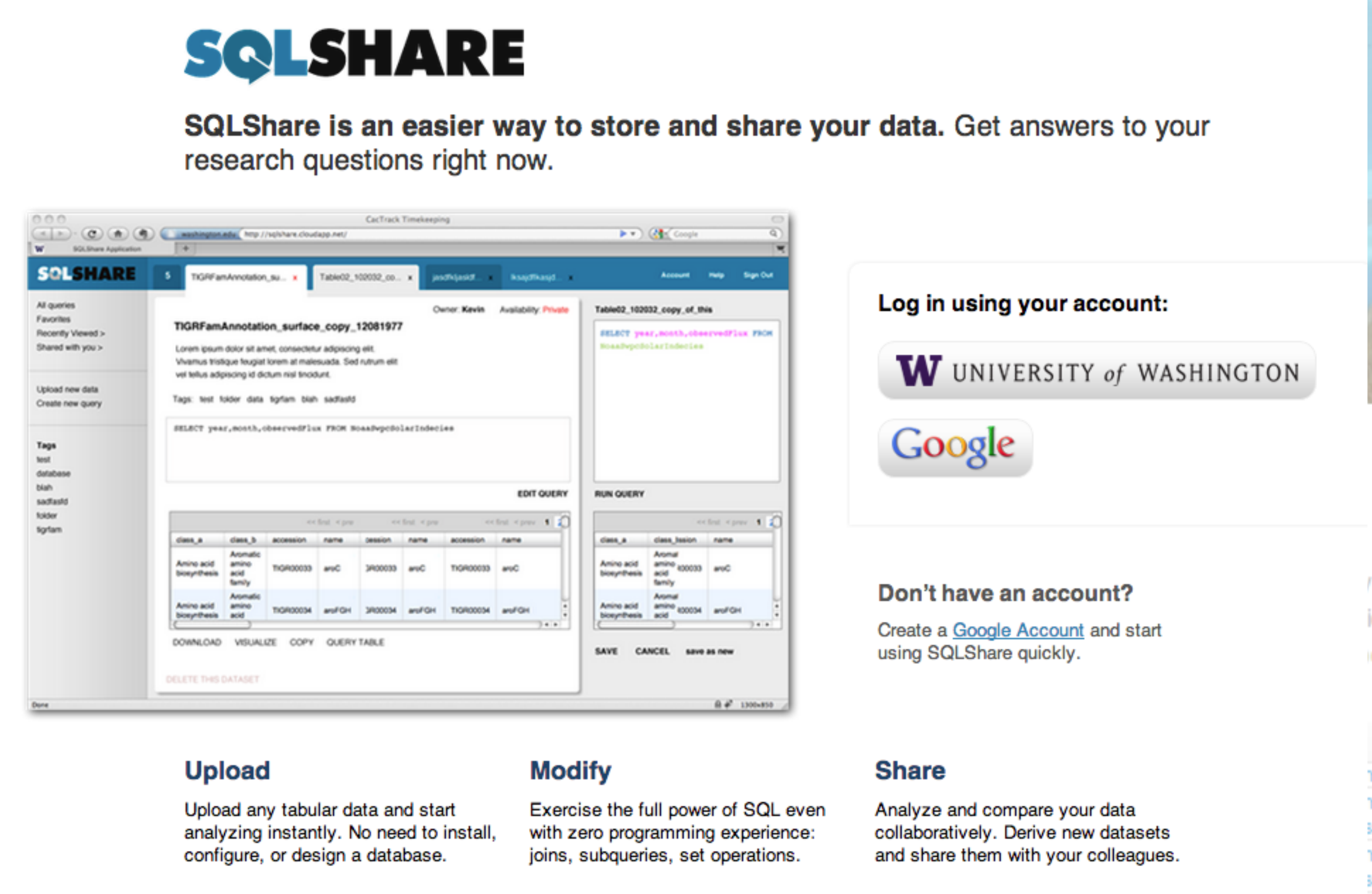
Platforms

Open Science

Data

everything else...





SQLSHARE

SQLShare is an easier way to store and share your data. Get answers to your research questions right now.

Log in using your account:

UNIVERSITY of WASHINGTON

Google

Don't have an account?

Create a Google Account and start using SQLShare quickly.

Upload

Upload any tabular data and start analyzing instantly. No need to install, configure, or design a database.

Modify

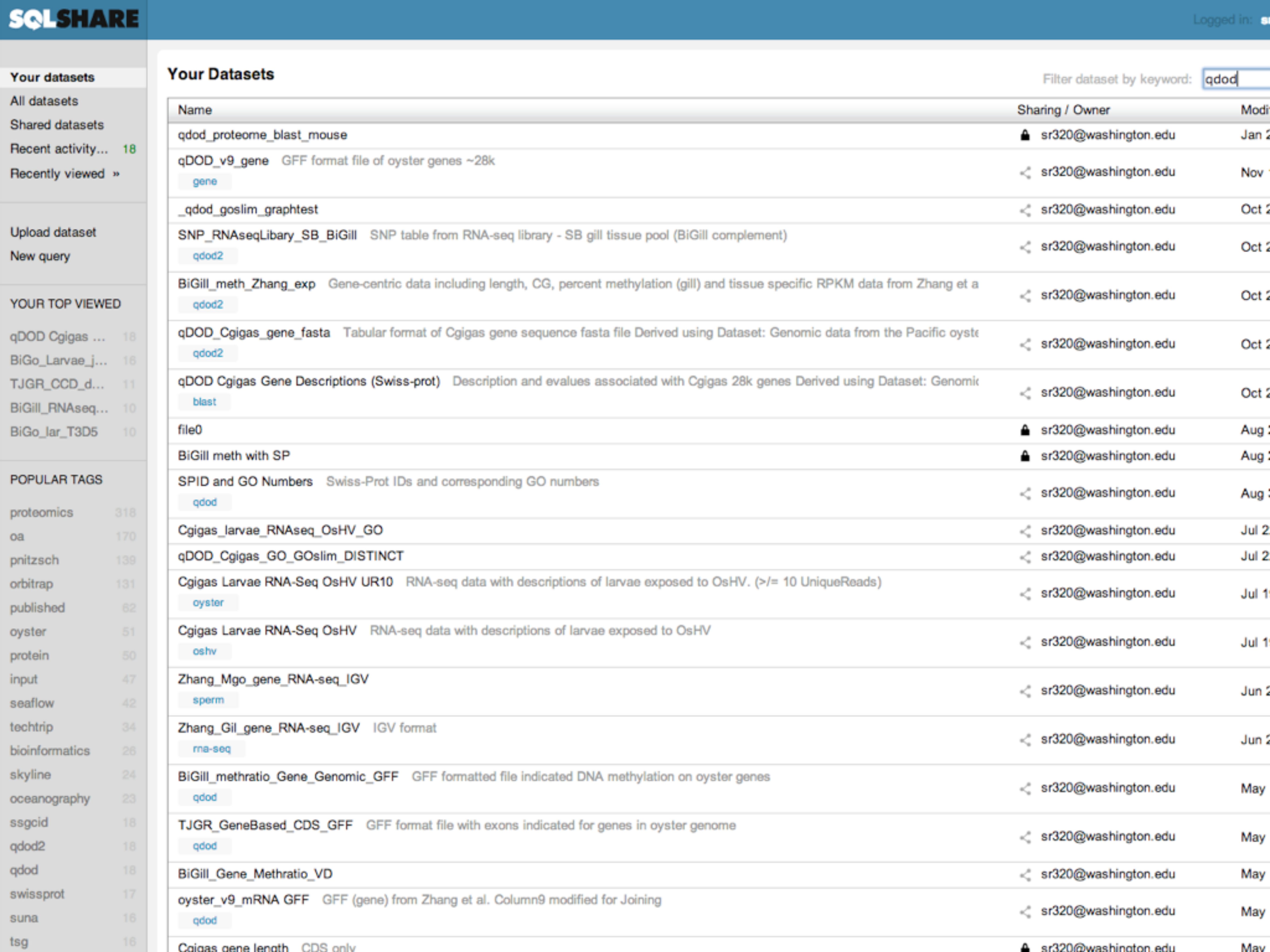
Exercise the full power of SQL even with zero programming experience: joins, subqueries, set operations.

Share

Analyze and compare your data collaboratively. Derive new datasets and share them with your colleagues.

one can view (and contribute) using the
ides you with a screenshot of the data in
en the table in a new webpage.

ttps://sqlshare.esc
ttps://sqlshare.esc
eq
ttps://sqlshare.esc
eq



YOUR TOP VIEWED

qDOD Cgigas ...	18
BiGo_Larvae_j...	16
TJGR_CCD_d...	11
BiGill_RNAseq...	10
BiGo_lar_T3D5	10

POPULAR TAGS

proteomics	318
oa	170
pnitzsch	139
orbitrap	131
published	62
oyster	51
protein	50
input	47
seaflow	42
techtrip	34
bioinformatics	26
skyline	24
oceanography	23
ssgcid	18
qdod2	18
qdod	18
swissprot	17
suna	16
tsg	16

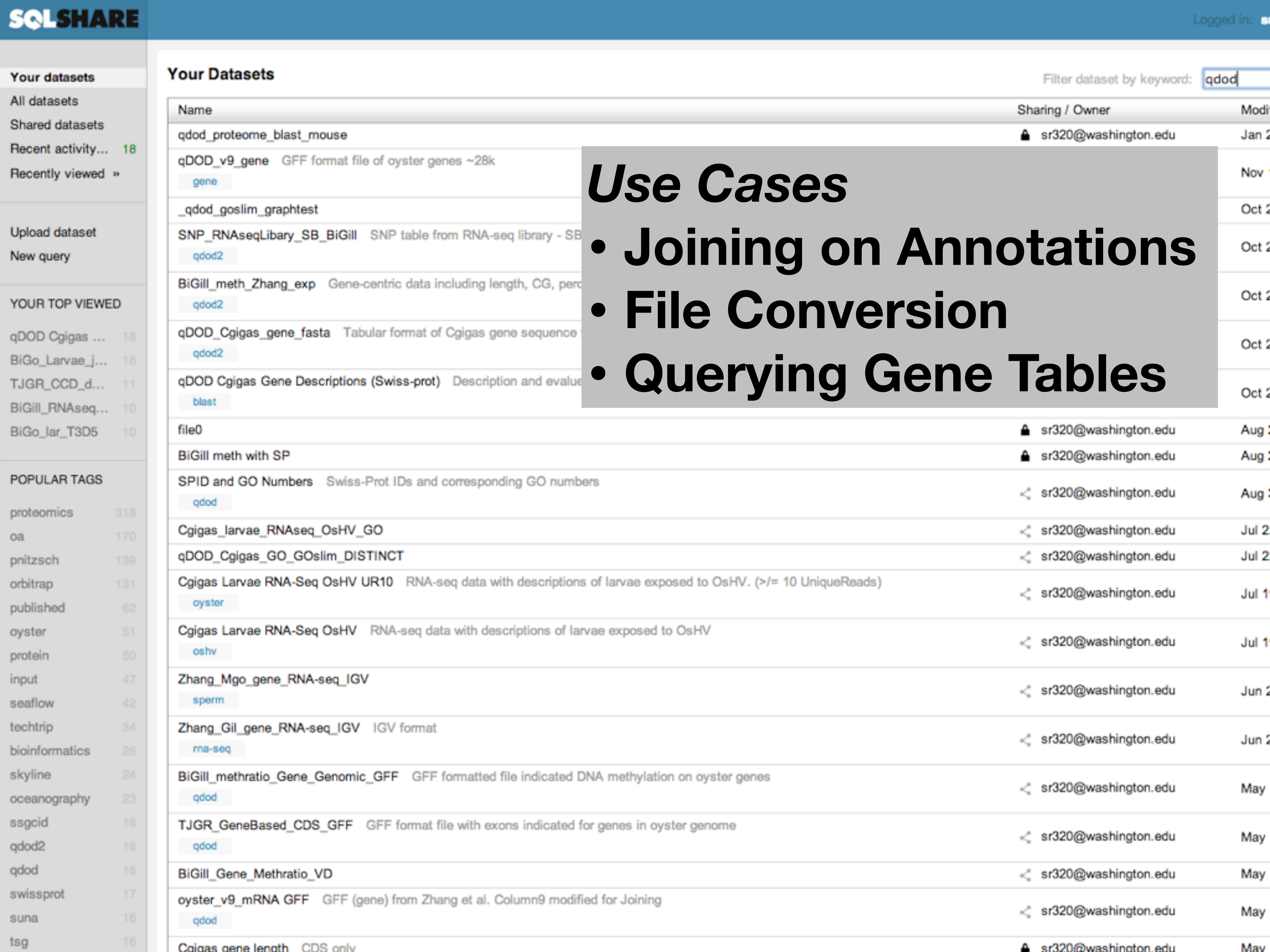
Your Datasets

Filter dataset by keyword: qdod

Name	Sharing / Owner	Modi
qdod_proteome_blast_mouse	🔒 sr320@washington.edu	Jan 2
qDOD_v9_gene GFF format file of oyster genes ~28k gene	< sr320@washington.edu	Nov
_qdod_goslim_graphtest	< sr320@washington.edu	Oct 2
SNP_RNAseqLibary_SB_BiGill SNP table from RNA-seq library - SB gill tissue pool (BiGill complement) qdod2	< sr320@washington.edu	Oct 2
BiGill_meth_Zhang_exp Gene-centric data including length, CG, percent methylation (gill) and tissue specific RPKM data from Zhang et a qdod2	< sr320@washington.edu	Oct 2
qDOD_Cgigas_gene_fasta Tabular format of Cgigas gene sequence fasta file Derived using Dataset: Genomic data from the Pacific oyste qdod2	< sr320@washington.edu	Oct 2
qDOD Cgigas Gene Descriptions (Swiss-prot) Description and evalues associated with Cgigas 28k genes Derived using Dataset: Genomi blast	< sr320@washington.edu	Oct 2
file0	🔒 sr320@washington.edu	Aug 1
BiGill meth with SP	🔒 sr320@washington.edu	Aug 2
SPID and GO Numbers Swiss-Prot IDs and corresponding GO numbers qdod	< sr320@washington.edu	Aug
Cgigas_larvae_RNAseq_OsHV_GO	< sr320@washington.edu	Jul 2
qDOD_Cgigas_GO_GOslim_DISTINCT	< sr320@washington.edu	Jul 2
Cgigas Larvae RNA-Seq OsHV UR10 RNA-seq data with descriptions of larvae exposed to OsHV. (>/= 10 UniqueReads) oyster	< sr320@washington.edu	Jul 1
Cgigas Larvae RNA-Seq OsHV RNA-seq data with descriptions of larvae exposed to OsHV oshv	< sr320@washington.edu	Jul 1
Zhang_Mgo_gene_RNA-seq_IGV sperm	< sr320@washington.edu	Jun 2
Zhang_Gil_gene_RNA-seq_IGV IGV format rna-seq	< sr320@washington.edu	Jun 2
BiGill_methratio_Gene_Genomic_GFF GFF formatted file indicated DNA methylation on oyster genes qdod	< sr320@washington.edu	May
TJGR_GeneBased_CDS_GFF GFF format file with exons indicated for genes in oyster genome qdod	< sr320@washington.edu	May
BiGill_Gene_Methratio_VD	< sr320@washington.edu	May
oyster_v9_mRNA GFF GFF (gene) from Zhang et al. Column9 modified for Joining qdod	< sr320@washington.edu	May
Cgigas gene length CDS only	🔒 sr320@washington.edu	May

Use Cases

- Joining on Annotations
- File Conversion
- Querying Gene Tables



SQLSHARE

Logged in:

Your datasets
All datasets
Shared datasets
Recent activity... 18
Recently viewed »

Upload dataset
New query

YOUR TOP VIEWED

qDOD Cgigas ...	18
BiGo_Larvae_j...	16
TJGR_CCD_d...	11
BiGill_RNAseq...	10
BiGo_lar_T3D5	10

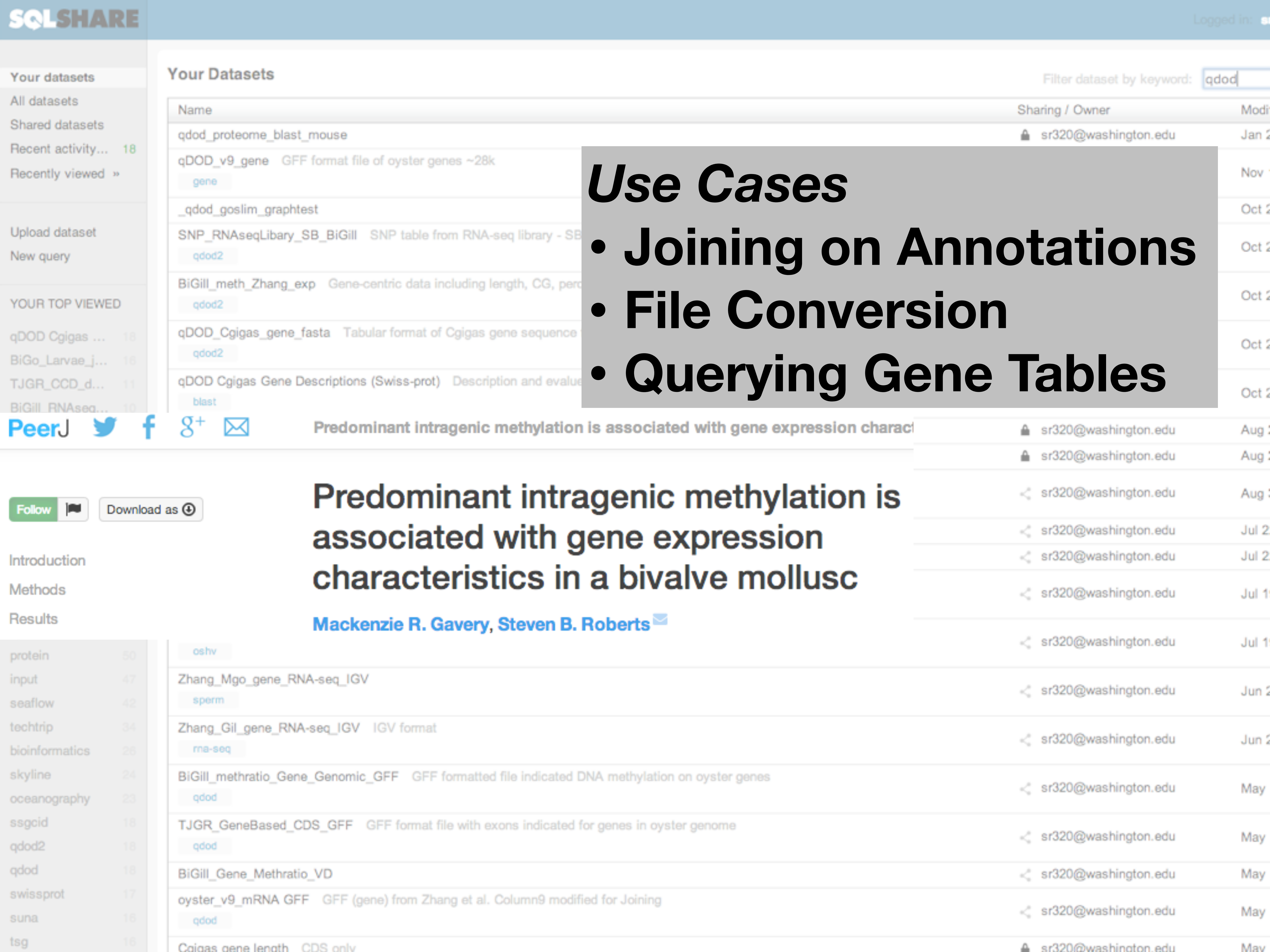
POPULAR TAGS

proteomics	318
oa	170
pnitzsch	139
orbitrap	131
published	62
oyster	51
protein	50
input	47
seaflow	42
techtrip	34
bioinformatics	26
skyline	24
oceanography	23
ssgcid	18
qdod2	18
qdod	18
swissprot	17
suna	16
tsg	16

Your Datasets

Filter dataset by keyword: qdod

Name	Sharing / Owner	Modi
qdod_proteome_blast_mouse	🔒 sr320@washington.edu	Jan 2
qDOD_v9_gene GFF format file of oyster genes ~28k gene		Nov
_qdod_goslim_graphtest		Oct 2
SNP_RNAseqLibary_SB_BiGill SNP table from RNA-seq library - SB qdod2		Oct 2
BiGill_meth_Zhang_exp Gene-centric data including length, CG, perc qdod2		Oct 2
qDOD_Cgigas_gene_fasta Tabular format of Cgigas gene sequence qdod2		Oct 2
qDOD Cgigas Gene Descriptions (Swiss-prot) Description and evalue blast		Oct 2
file0	🔒 sr320@washington.edu	Aug
BiGill meth with SP	🔒 sr320@washington.edu	Aug
SPID and GO Numbers Swiss-Prot IDs and corresponding GO numbers qdod	⋖ sr320@washington.edu	Aug
Cgigas_larvae_RNAseq_OsHV_GO	⋖ sr320@washington.edu	Jul 2
qDOD_Cgigas_GO_GOslim_DISTINCT	⋖ sr320@washington.edu	Jul 2
Cgigas Larvae RNA-Seq OsHV UR10 RNA-seq data with descriptions of larvae exposed to OsHV. (>/= 10 UniqueReads) oyster	⋖ sr320@washington.edu	Jul 1
Cgigas Larvae RNA-Seq OsHV RNA-seq data with descriptions of larvae exposed to OsHV oshv	⋖ sr320@washington.edu	Jul 1
Zhang_Mgo_gene_RNA-seq_IGV sperm	⋖ sr320@washington.edu	Jun 2
Zhang_Gil_gene_RNA-seq_IGV IGV format rna-seq	⋖ sr320@washington.edu	Jun 2
BiGill_methratio_Gene_Genomic_GFF GFF formatted file indicated DNA methylation on oyster genes qdod	⋖ sr320@washington.edu	May
TJGR_GeneBased_CDS_GFF GFF format file with exons indicated for genes in oyster genome qdod	⋖ sr320@washington.edu	May
BiGill_Gene_Methratio_VD	⋖ sr320@washington.edu	May
oyster_v9_mRNA GFF GFF (gene) from Zhang et al. Column9 modified for Joining qdod	⋖ sr320@washington.edu	May
Cgigas gene length CDS only	🔒 sr320@washington.edu	May



Use Cases

- Joining on Annotations
- File Conversion
- Querying Gene Tables

Your Datasets

Name	Sharing / Owner	Modi
qdod_proteome_blast_mouse	sr320@washington.edu	Jan
qDOD_v9_gene GFF format file of oyster genes ~28k gene		Nov
_qdod_goslim_graphtest		Oct
SNP_RNAseqLibary_SB_BiGill SNP table from RNA-seq library - SB qdod2		Oct
BiGill_meth_Zhang_exp Gene-centric data including length, CG, perc qdod2		Oct
qDOD_Cgigas_gene_fasta Tabular format of Cgigas gene sequence qdod2		Oct
qDOD Cgigas Gene Descriptions (Swiss-prot) Description and evalue blast		Oct
	sr320@washington.edu	Aug
	sr320@washington.edu	Aug
	sr320@washington.edu	Aug
	sr320@washington.edu	Jul
	sr320@washington.edu	Jul
	sr320@washington.edu	Jul
oshv	sr320@washington.edu	Jul
Zhang_Mgo_gene_RNA-seq_IGV sperm	sr320@washington.edu	Jun
Zhang_Gil_gene_RNA-seq_IGV IGV format rna-seq	sr320@washington.edu	Jun
BiGill_methratio_Gene_Genomic_GFF GFF formatted file indicated DNA methylation on oyster genes qdod	sr320@washington.edu	May
TJGR_GeneBased_CDS_GFF GFF format file with exons indicated for genes in oyster genome qdod	sr320@washington.edu	May
BiGill_Gene_Methratio_VD	sr320@washington.edu	May
oyster_v9_mRNA GFF GFF (gene) from Zhang et al. Column9 modified for Joining qdod	sr320@washington.edu	May
Cgigas gene length CDS only	sr320@washington.edu	May

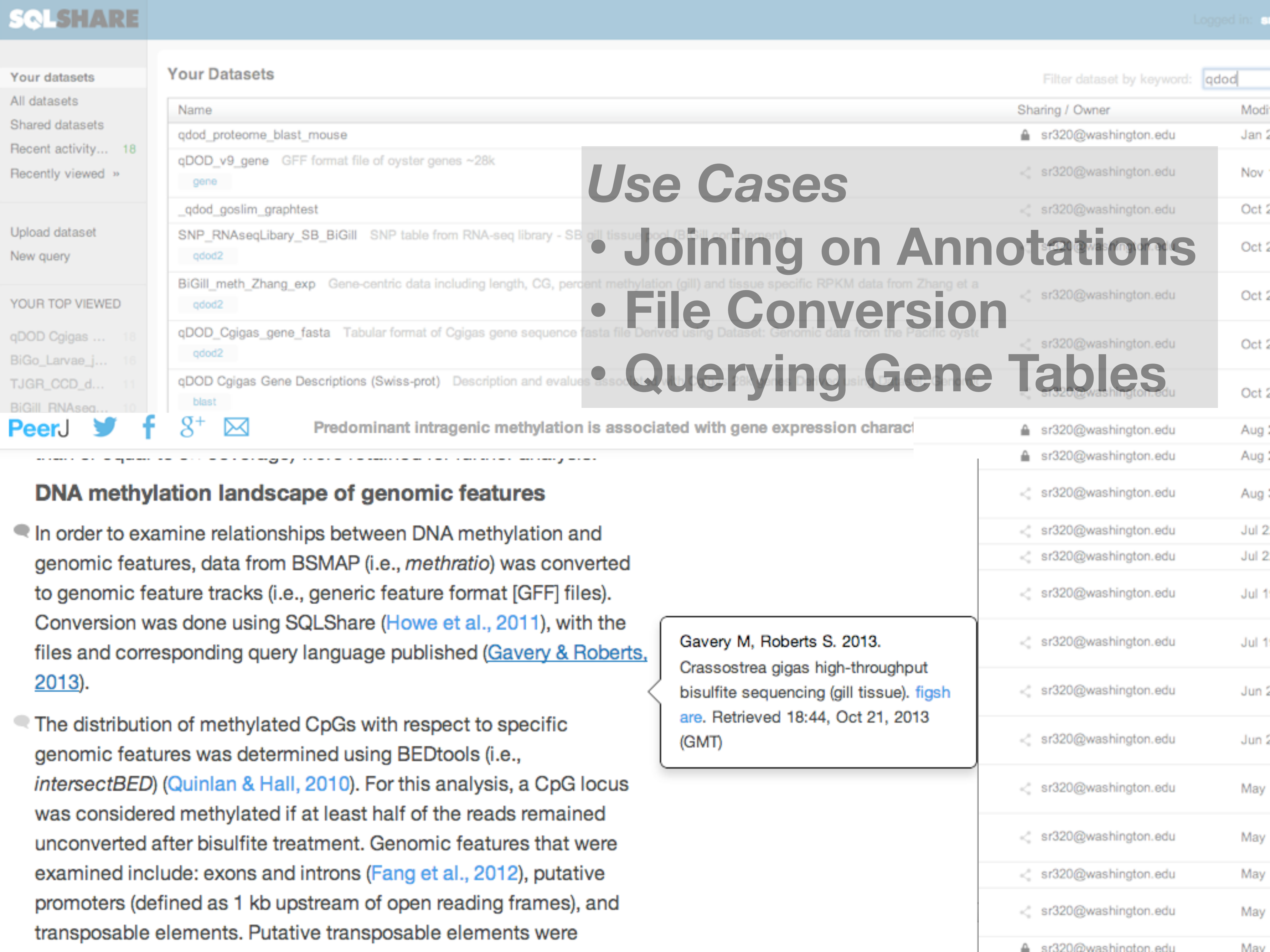
Predominant intragenic methylation is associated with gene expression charact

Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc

Mackenzie R. Gavery, Steven B. Roberts

Use Cases

- Joining on Annotations
- File Conversion
- Querying Gene Tables



Your Datasets

Name	Sharing / Owner	Modi
qdod_proteome_blast_mouse	sr320@washington.edu	Jan
qDOD_v9_gene GFF format file of oyster genes ~28k gene	sr320@washington.edu	Nov
_qdod_goslim_graphtest	sr320@washington.edu	Oct
SNP_RNAseqLibary_SB_BiGill SNP table from RNA-seq library - SB gill tissue pool (BiGill complement) qdod2	sr320@washington.edu	Oct
BiGill_meth_Zhang_exp Gene-centric data including length, CG, percent methylation (gill) and tissue specific RPKM data from Zhang et a qdod2	sr320@washington.edu	Oct
qDOD_Cgigas_gene_fasta Tabular format of Cgigas gene sequence fasta file Derived using Dataset: Genomic data from the Pacific oyste qdod2	sr320@washington.edu	Oct
qDOD Cgigas Gene Descriptions (Swiss-prot) Description and evalues associated with Cgigas 28k genes Derived using Dataset: Genom blast	sr320@washington.edu	Oct
	sr320@washington.edu	Aug
	sr320@washington.edu	Aug
	sr320@washington.edu	Aug
	sr320@washington.edu	Jul
	sr320@washington.edu	Jul
	sr320@washington.edu	Jul
	sr320@washington.edu	Jul
	sr320@washington.edu	Jun
	sr320@washington.edu	Jun
	sr320@washington.edu	May
	sr320@washington.edu	May
	sr320@washington.edu	May
	sr320@washington.edu	May
	sr320@washington.edu	May

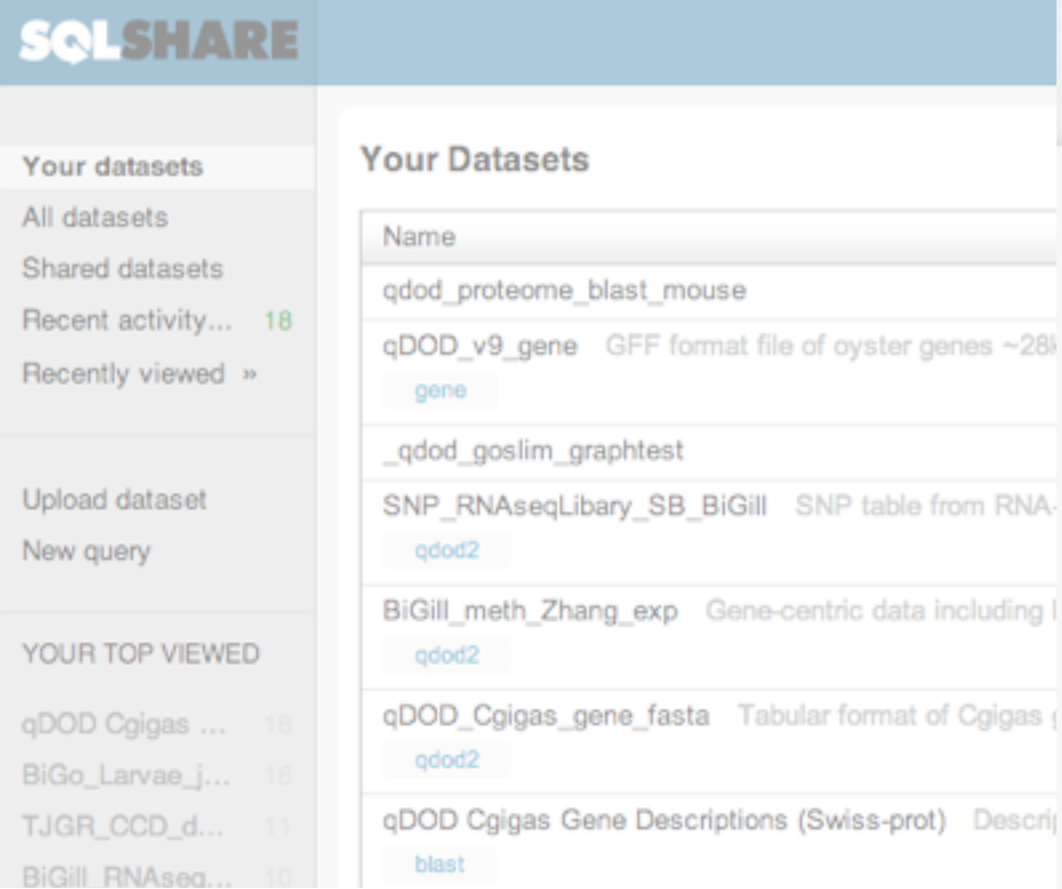
Predominant intragenic methylation is associated with gene expression charact

DNA methylation landscape of genomic features

In order to examine relationships between DNA methylation and genomic features, data from BSMAP (i.e., *methratio*) was converted to genomic feature tracks (i.e., generic feature format [GFF] files). Conversion was done using SQLShare (Howe et al., 2011), with the files and corresponding query language published (Gavery & Roberts, 2013).

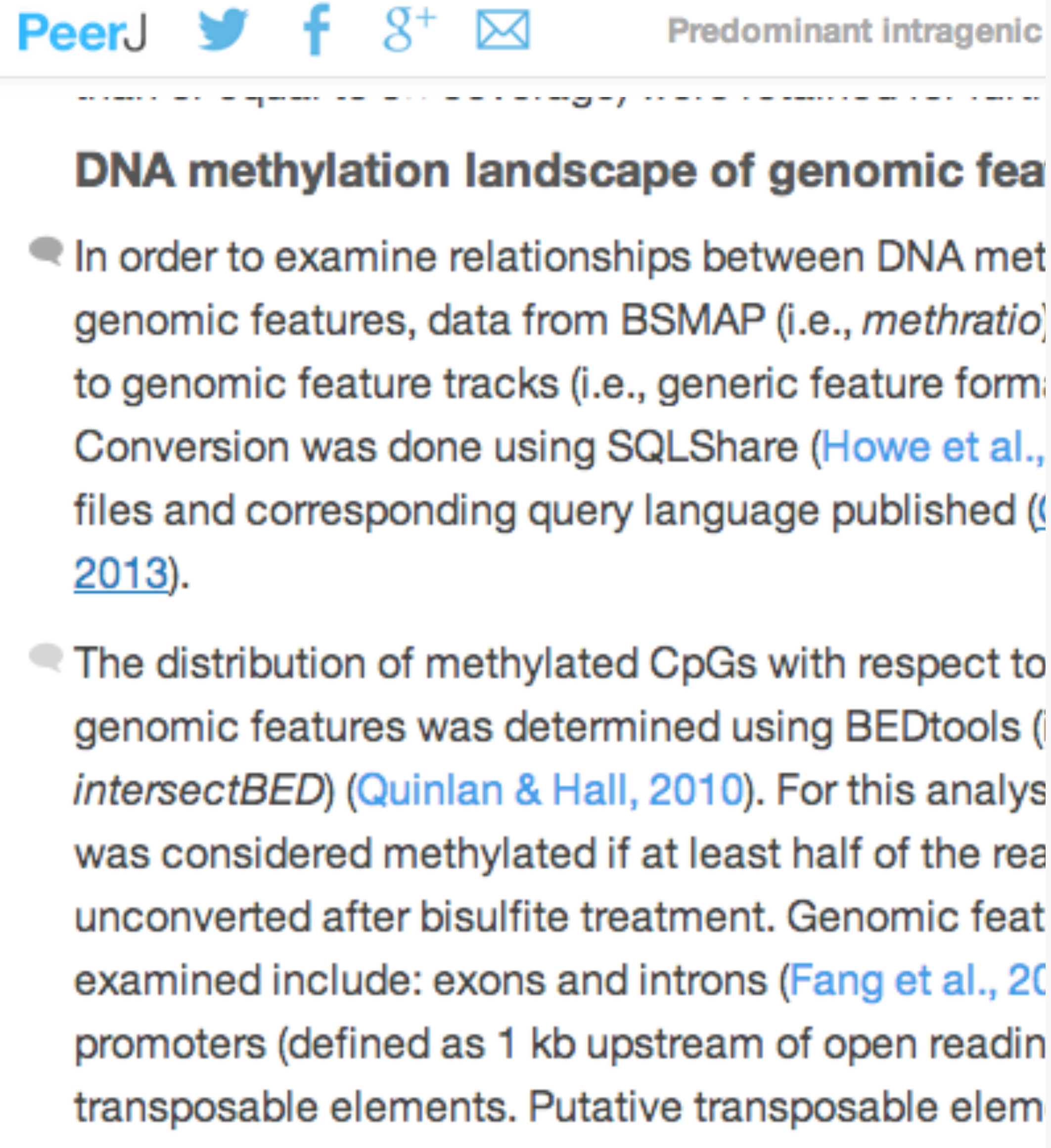
The distribution of methylated CpGs with respect to specific genomic features was determined using BEDtools (i.e., *intersectBED*) (Quinlan & Hall, 2010). For this analysis, a CpG locus was considered methylated if at least half of the reads remained unconverted after bisulfite treatment. Genomic features that were examined include: exons and introns (Fang et al., 2012), putative promoters (defined as 1 kb upstream of open reading frames), and transposable elements. Putative transposable elements were

Gavery M, Roberts S. 2013. Crassostrea gigas high-throughput bisulfite sequencing (gill tissue). figshare. Retrieved 18:44, Oct 21, 2013 (GMT)



Your Datasets

Name
qdod_proteome_blast_mouse
qDOD_v9_gene GFF format file of oyster genes ~28 gene
_qdod_goslim_graphtest
SNP_RNAseqLibary_SB_BiGill SNP table from RNA- qdod2
BiGill_meth_Zhang_exp Gene-centric data including qdod2
qDOD_Cgigas_gene_fasta Tabular format of Cgigas qdod2
qDOD Cgigas Gene Descriptions (Swiss-prot) Descri blast

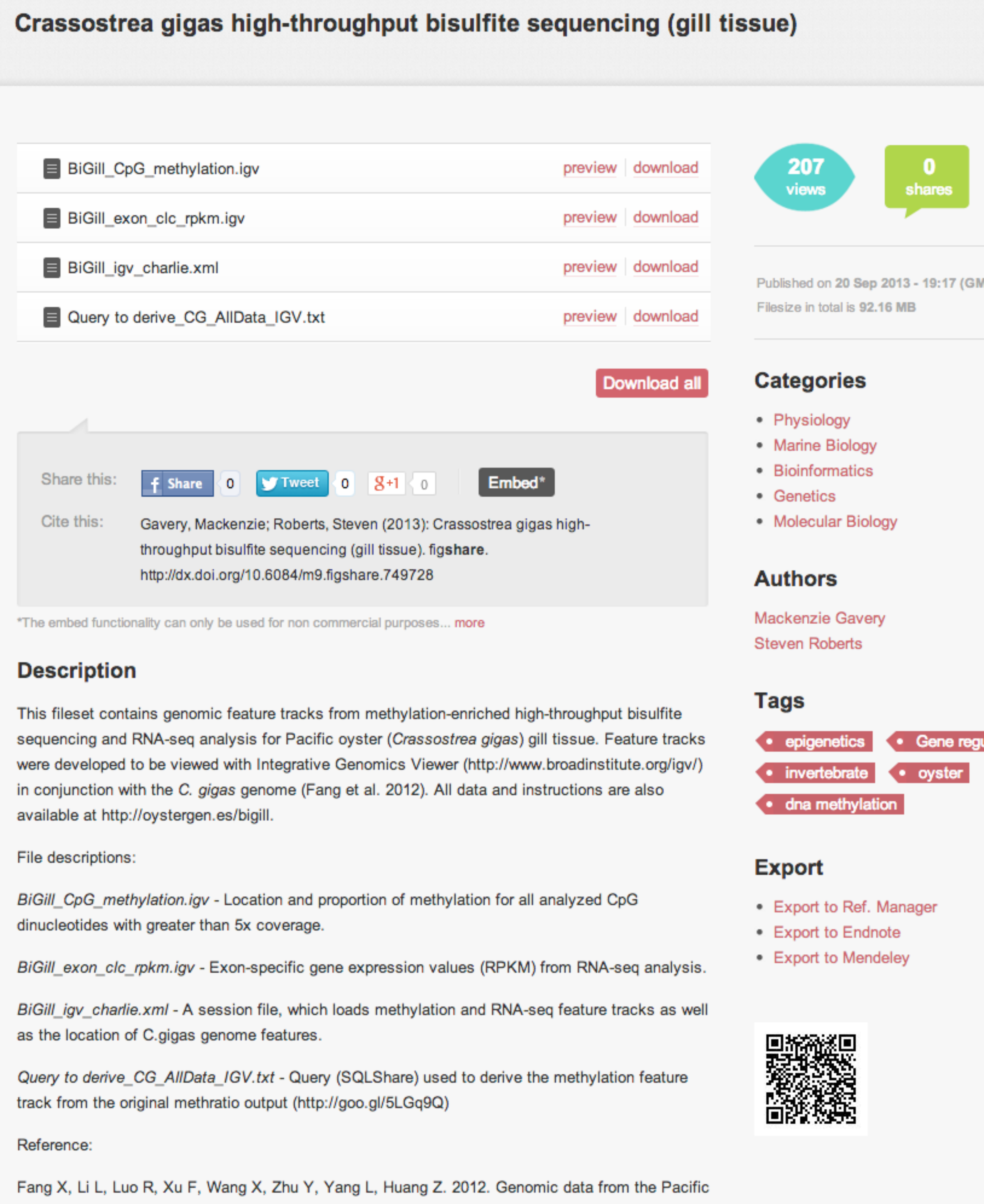


Predominant intragenic

DNA methylation landscape of genomic fea

In order to examine relationships between DNA met genomic features, data from BSMAP (i.e., *methratio* to genomic feature tracks (i.e., generic feature form Conversion was done using SQLShare (Howe et al., files and corresponding query language published (2013).

The distribution of methylated CpGs with respect to genomic features was determined using BEDtools (*intersectBED*) (Quinlan & Hall, 2010). For this analys was considered methylated if at least half of the rea unconverted after bisulfite treatment. Genomic feat examined include: exons and introns (Fang et al., 20 promoters (defined as 1 kb upstream of open readin transposable elements. Putative transposable elem



Crassostrea gigas high-throughput bisulfite sequencing (gill tissue)

BiGill_CpG_methylation.igv	preview	download
BiGill_exon_clc_rpkm.igv	preview	download
BiGill_igv_charlie.xml	preview	download
Query to derive_CG_AllData_IGV.txt	preview	download

Published on 20 Sep 2013 - 19:17 (GM

Filesize in total is 92.16 MB

Share this:

Cite this: Gavery, Mackenzie; Roberts, Steven (2013): Crassostrea gigas high-throughput bisulfite sequencing (gill tissue). figshare. http://dx.doi.org/10.6084/m9.figshare.749728

*The embed functionality can only be used for non commercial purposes... more

Categories

- Physiology
- Marine Biology
- Bioinformatics
- Genetics
- Molecular Biology

Authors

Mackenzie Gavery
Steven Roberts

Tags

- epigenetics
- Gene regu
- invertebrate
- oyster
- dna methylation

Export

- Export to Ref. Manager
- Export to Endnote
- Export to Mendeley

Description

This fileset contains genomic feature tracks from methylation-enriched high-throughput bisulfite sequencing and RNA-seq analysis for Pacific oyster (*Crassostrea gigas*) gill tissue. Feature tracks were developed to be viewed with Integrative Genomics Viewer (http://www.broadinstitute.org/igv/) in conjunction with the *C. gigas* genome (Fang et al. 2012). All data and instructions are also available at http://oystergen.es/bigill.

File descriptions:

BiGill_CpG_methylation.igv - Location and proportion of methylation for all analyzed CpG dinucleotides with greater than 5x coverage.

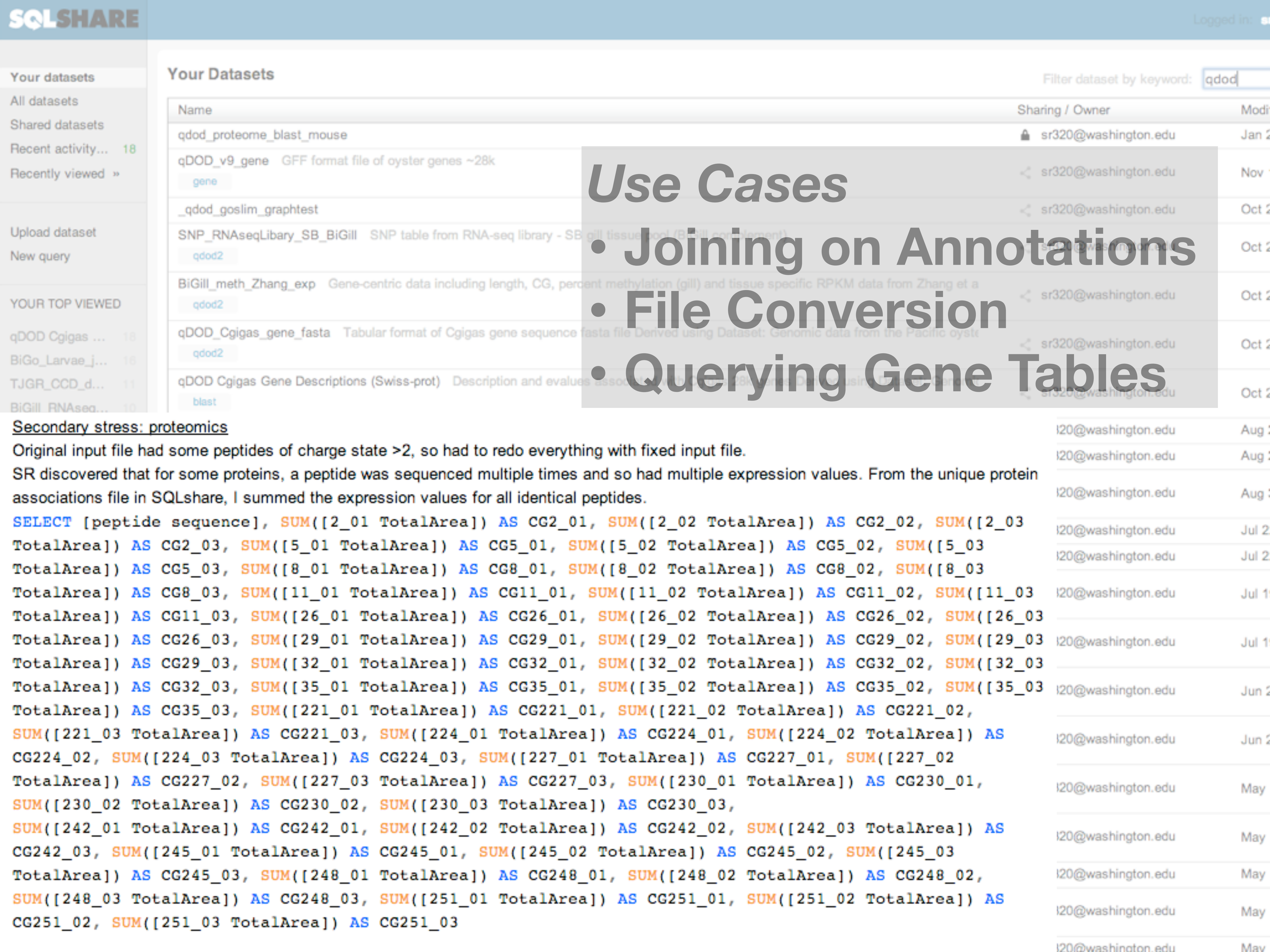
BiGill_exon_clc_rpkm.igv - Exon-specific gene expression values (RPKM) from RNA-seq analysis.

BiGill_igv_charlie.xml - A session file, which loads methylation and RNA-seq feature tracks as well as the location of C.gigas genome features.

Query to derive_CG_AllData_IGV.txt - Query (SQLShare) used to derive the methylation feature track from the original methratio output (http://goo.gl/5LGq9Q)

Reference:

Fang X, Li L, Luo R, Xu F, Wang X, Zhu Y, Yang L, Huang Z. 2012. Genomic data from the Pacific



Name	Sharing / Owner	Modi
qdod_proteome_blast_mouse	sr320@washington.edu	Jan 2
qDOD_v9_gene GFF format file of oyster genes ~28k gene	sr320@washington.edu	Nov
_qdod_goslim_graphtest	sr320@washington.edu	Oct 2
SNP_RNAseqLibary_SB_BiGill SNP table from RNA-seq library - SB gill tissue pool (BiGill complement) qdod2	sr320@washington.edu	Oct 2
BiGill_meth_Zhang_exp Gene-centric data including length, CG, percent methylation (gill) and tissue specific RPKM data from Zhang et a qdod2	sr320@washington.edu	Oct 2
qDOD_Cgigas_gene_fasta Tabular format of Cgigas gene sequence fasta file Derived using Dataset: Genomic data from the Pacific oyste qdod2	sr320@washington.edu	Oct 2
qDOD Cgigas Gene Descriptions (Swiss-prot) Description and evalues associated with Cgigas 28k genes Derived using Dataset: Genom blast	sr320@washington.edu	Oct 2
	20@washington.edu	Aug
	20@washington.edu	Aug
	20@washington.edu	Aug
	20@washington.edu	Jul 2
	20@washington.edu	Jul 2
	20@washington.edu	Jul 1
	20@washington.edu	Jul 1
	20@washington.edu	Jun 2
	20@washington.edu	Jun 2
	20@washington.edu	May
	20@washington.edu	May
	20@washington.edu	May
	20@washington.edu	May
	20@washington.edu	May

Use Cases

- Joining on Annotations
- File Conversion
- Querying Gene Tables

Secondary stress: proteomics

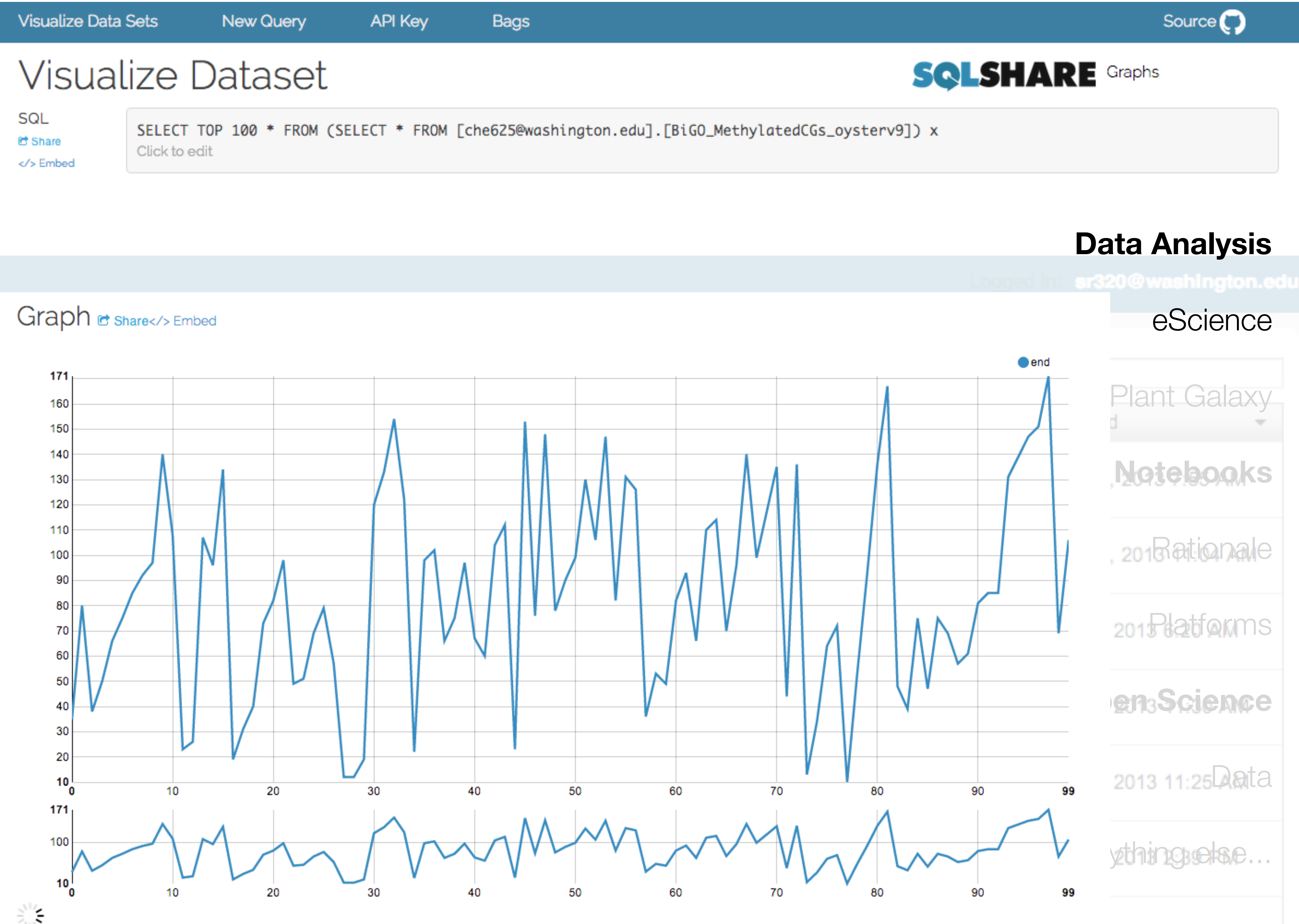
Original input file had some peptides of charge state >2, so had to redo everything with fixed input file.

SR discovered that for some proteins, a peptide was sequenced multiple times and so had multiple expression values. From the unique protein associations file in SQLshare, I summed the expression values for all identical peptides.

```
SELECT [peptide sequence], SUM([2_01 TotalArea]) AS CG2_01, SUM([2_02 TotalArea]) AS CG2_02, SUM([2_03
TotalArea]) AS CG2_03, SUM([5_01 TotalArea]) AS CG5_01, SUM([5_02 TotalArea]) AS CG5_02, SUM([5_03
TotalArea]) AS CG5_03, SUM([8_01 TotalArea]) AS CG8_01, SUM([8_02 TotalArea]) AS CG8_02, SUM([8_03
TotalArea]) AS CG8_03, SUM([11_01 TotalArea]) AS CG11_01, SUM([11_02 TotalArea]) AS CG11_02, SUM([11_03
TotalArea]) AS CG11_03, SUM([26_01 TotalArea]) AS CG26_01, SUM([26_02 TotalArea]) AS CG26_02, SUM([26_03
TotalArea]) AS CG26_03, SUM([29_01 TotalArea]) AS CG29_01, SUM([29_02 TotalArea]) AS CG29_02, SUM([29_03
TotalArea]) AS CG29_03, SUM([32_01 TotalArea]) AS CG32_01, SUM([32_02 TotalArea]) AS CG32_02, SUM([32_03
TotalArea]) AS CG32_03, SUM([35_01 TotalArea]) AS CG35_01, SUM([35_02 TotalArea]) AS CG35_02, SUM([35_03
TotalArea]) AS CG35_03, SUM([221_01 TotalArea]) AS CG221_01, SUM([221_02 TotalArea]) AS CG221_02,
SUM([221_03 TotalArea]) AS CG221_03, SUM([224_01 TotalArea]) AS CG224_01, SUM([224_02 TotalArea]) AS
CG224_02, SUM([224_03 TotalArea]) AS CG224_03, SUM([227_01 TotalArea]) AS CG227_01, SUM([227_02
TotalArea]) AS CG227_02, SUM([227_03 TotalArea]) AS CG227_03, SUM([230_01 TotalArea]) AS CG230_01,
SUM([230_02 TotalArea]) AS CG230_02, SUM([230_03 TotalArea]) AS CG230_03,
SUM([242_01 TotalArea]) AS CG242_01, SUM([242_02 TotalArea]) AS CG242_02, SUM([242_03 TotalArea]) AS
CG242_03, SUM([245_01 TotalArea]) AS CG245_01, SUM([245_02 TotalArea]) AS CG245_02, SUM([245_03
TotalArea]) AS CG245_03, SUM([248_01 TotalArea]) AS CG248_01, SUM([248_02 TotalArea]) AS CG248_02,
SUM([248_03 TotalArea]) AS CG248_03, SUM([251_01 TotalArea]) AS CG251_01, SUM([251_02 TotalArea]) AS
CG251_02, SUM([251_03 TotalArea]) AS CG251_03
```

Data Analysis

eScience



Visualize Dataset

SQL

```
SELECT TOP 100 * FROM (SELECT * FROM [che625@washington.edu].[BiGO_MethylatedCGs_oysterv9]) x
```

Click to edit

Logged in: sr320@washington.edu

Graph

Plant Galaxy

Notebooks

Rationale

Platforms

en Science

Data

ything else…

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

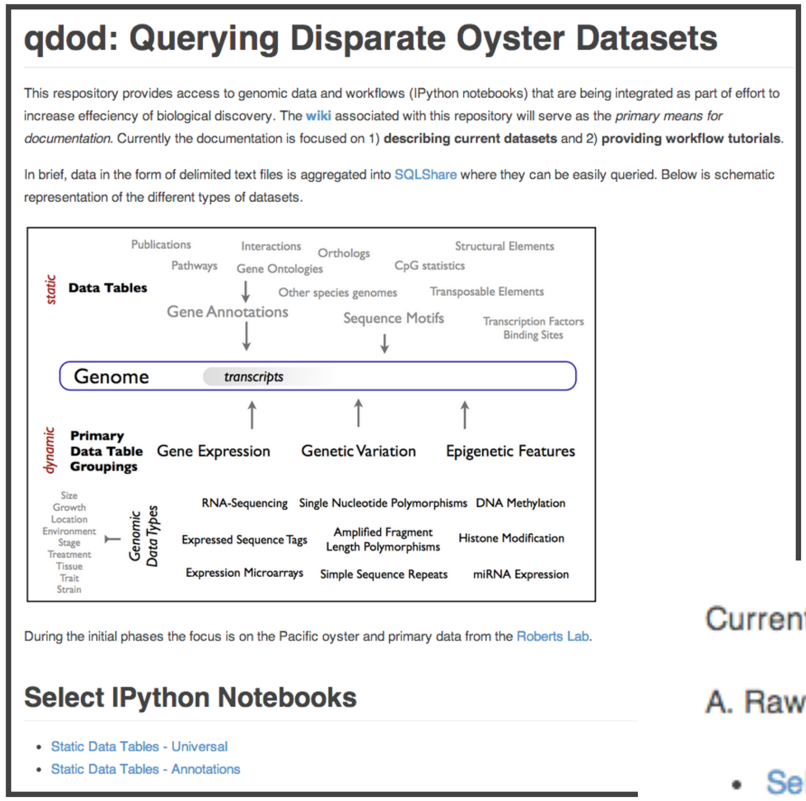
Rationale

Platforms

Open Science

Data

everything else...



qdod: Querying Disparate Oyster Datasets

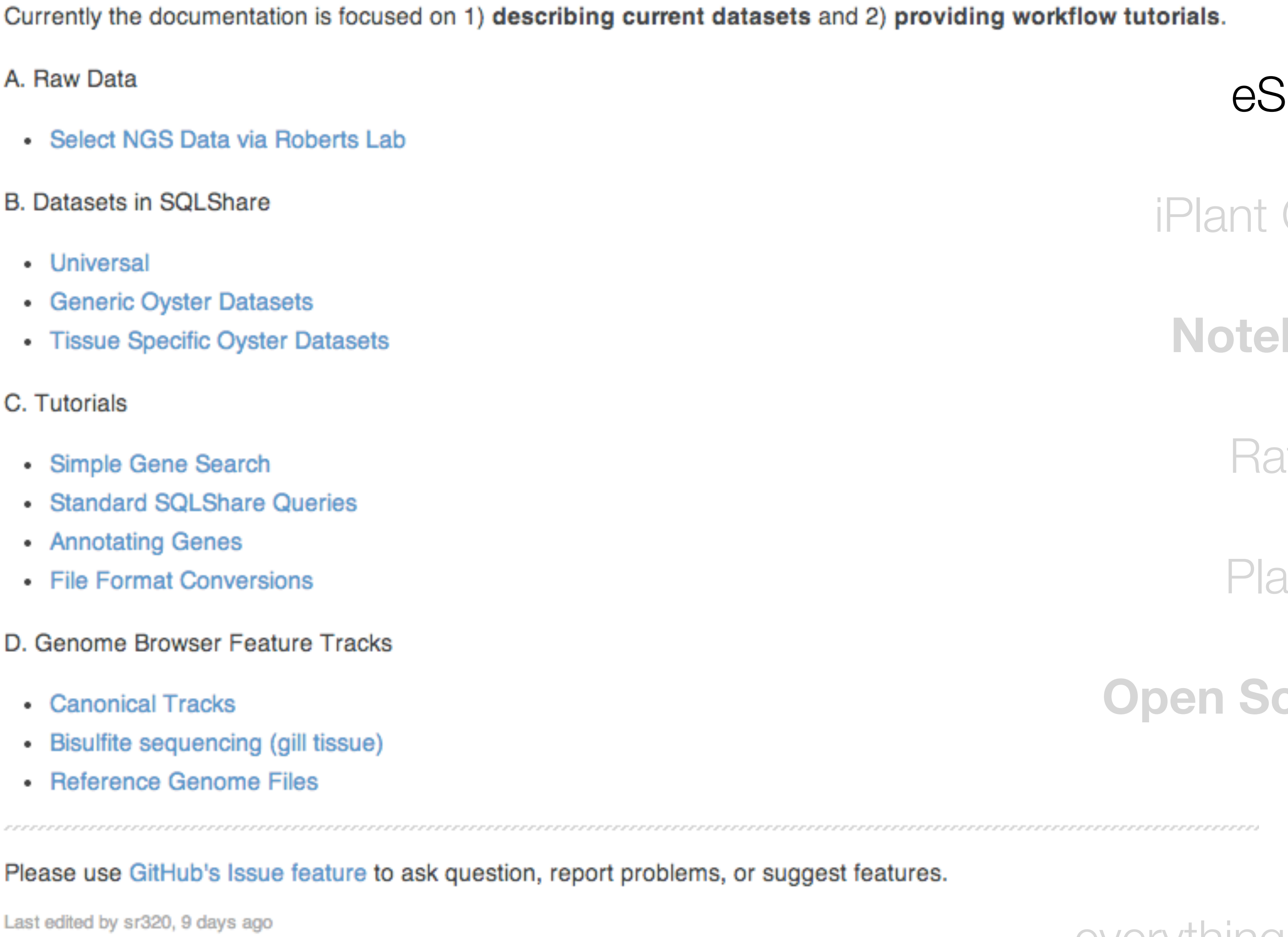
This respository provides access to genomic data and workflows (IPython notebooks) that are being integrated as part of effort to increase effeciency of biological discovery. The **wiki** associated with this repository will serve as the *primary means for documentation*. Currently the documentation is focused on 1) **describing current datasets** and 2) **providing workflow tutorials**.

In brief, data in the form of delimited text files is aggregated into SQLShare where they can be easily queried. Below is schematic representation of the different types of datasets.

During the initial phases the focus is on the Pacific oyster and primary data from the Roberts Lab.

Select IPython Notebooks

- Static Data Tables - Universal
- Static Data Tables - Annotations



Currently the documentation is focused on 1) **describing current datasets** and 2) **providing workflow tutorials**.

A. Raw Data

- Select NGS Data via Roberts Lab

B. Datasets in SQLShare

- Universal
- Generic Oyster Datasets
- Tissue Specific Oyster Datasets

C. Tutorials

- Simple Gene Search
- Standard SQLShare Queries
- Annotating Genes
- File Format Conversions

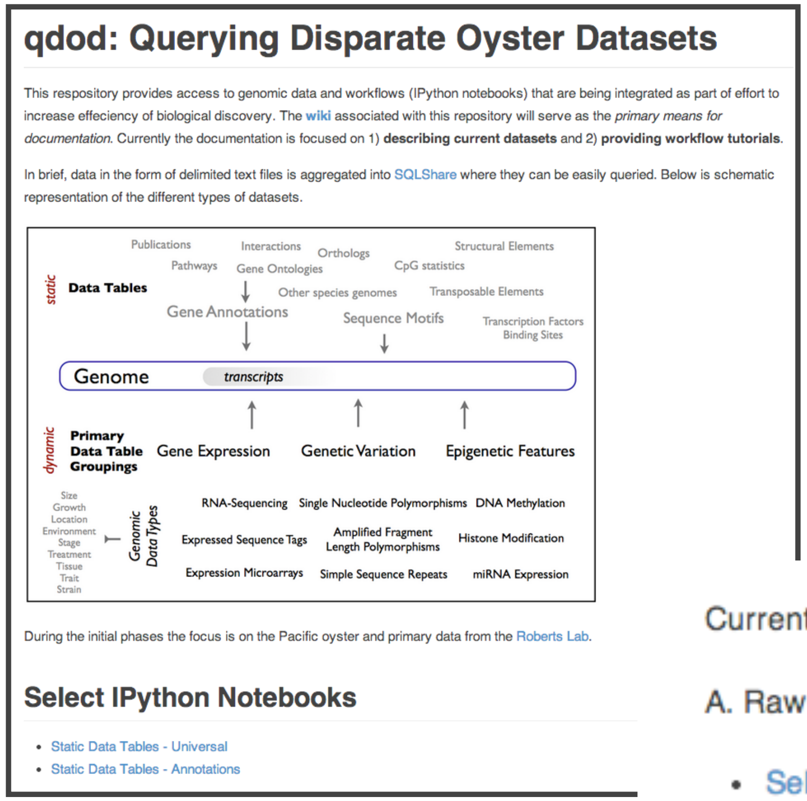
D. Genome Browser Feature Tracks

- Canonical Tracks
- Bisulfite sequencing (gill tissue)
- Reference Genome Files

Please use GitHub's Issue feature to ask question, report problems, or suggest features.

Last edited by sr320, 9 days ago

github.com/sr320/qdod/wiki



qdod: Querying Disparate Oyster Datasets

This respository provides access to genomic data and workflows (IPython notebooks) that are being integrated as part of effort to increase effeciency of biological discovery. The **wiki** associated with this repository will serve as the *primary means for documentation*. Currently the documentation is focused on 1) **describing current datasets** and 2) **providing workflow tutorials**.

In brief, data in the form of delimited text files is aggregated into SQLShare where they can be easily queried. Below is schematic representation of the different types of datasets.

During the initial phases the focus is on the Pacific oyster and primary data from the Roberts Lab.

Select IPython Notebooks

- Static Data Tables - Universal
- Static Data Tables - Annotations

Currently the documentation is focused on

A. Raw Data

- Select NGS Data via Roberts Lab

B. Datasets in SQLShare

- Universal
- Generic Oyster Datasets
- Tissue Specific Oyster Datasets

C. Tutorials

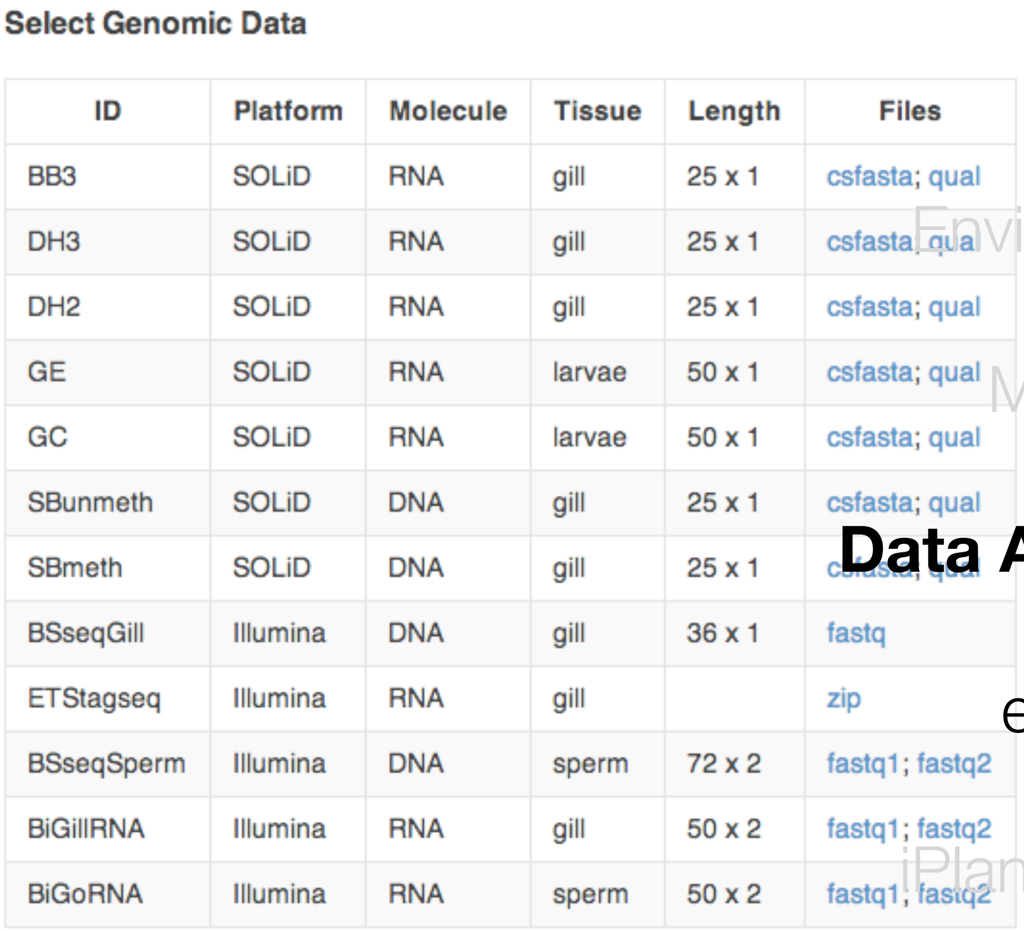
- Simple Gene Search
- Standard SQLShare Queries
- Annotating Genes
- File Format Conversions

D. Genome Browser Feature Tracks

- Canonical Tracks
- Bisulfite sequencing (gill tissue)
- Reference Genome Files

Please use GitHub's Issue feature to ask question, report problems, or suggest features.

Last edited by sr320, 9 days ago

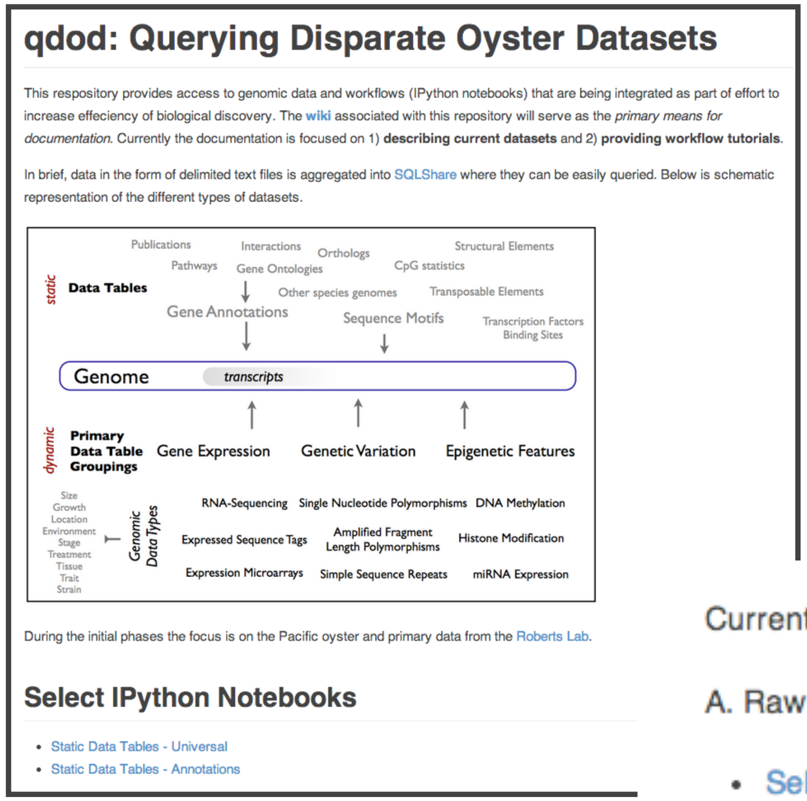


Select Genomic Data

ID	Platform	Molecule	Tissue	Length	Files
BB3	SOLiD	RNA	gill	25 x 1	csfasta; qual
DH3	SOLiD	RNA	gill	25 x 1	csfasta; qual
DH2	SOLiD	RNA	gill	25 x 1	csfasta; qual
GE	SOLiD	RNA	larvae	50 x 1	csfasta; qual
GC	SOLiD	RNA	larvae	50 x 1	csfasta; qual
SBunmeth	SOLiD	DNA	gill	25 x 1	csfasta; qual
SBmeth	SOLiD	DNA	gill	25 x 1	csfasta; qual
BSseqGill	Illumina	DNA	gill	36 x 1	fastq
ETStagseq	Illumina	RNA	gill		zip
BSseqSperm	Illumina	DNA	sperm	72 x 2	fastq1; fastq2
BiGillRNA	Illumina	RNA	gill	50 x 2	fastq1; fastq2
BiGoRNA	Illumina	RNA	sperm	50 x 2	fastq1; fastq2

Biology
Environment
Molecular
Data Analysis
eScience
iPlant Galaxy
Notebooks
Rationale
Platforms
Open Science
Data
everything else...

github.com/sr320/qdod/wiki



qdod: Querying Disparate Oyster Datasets

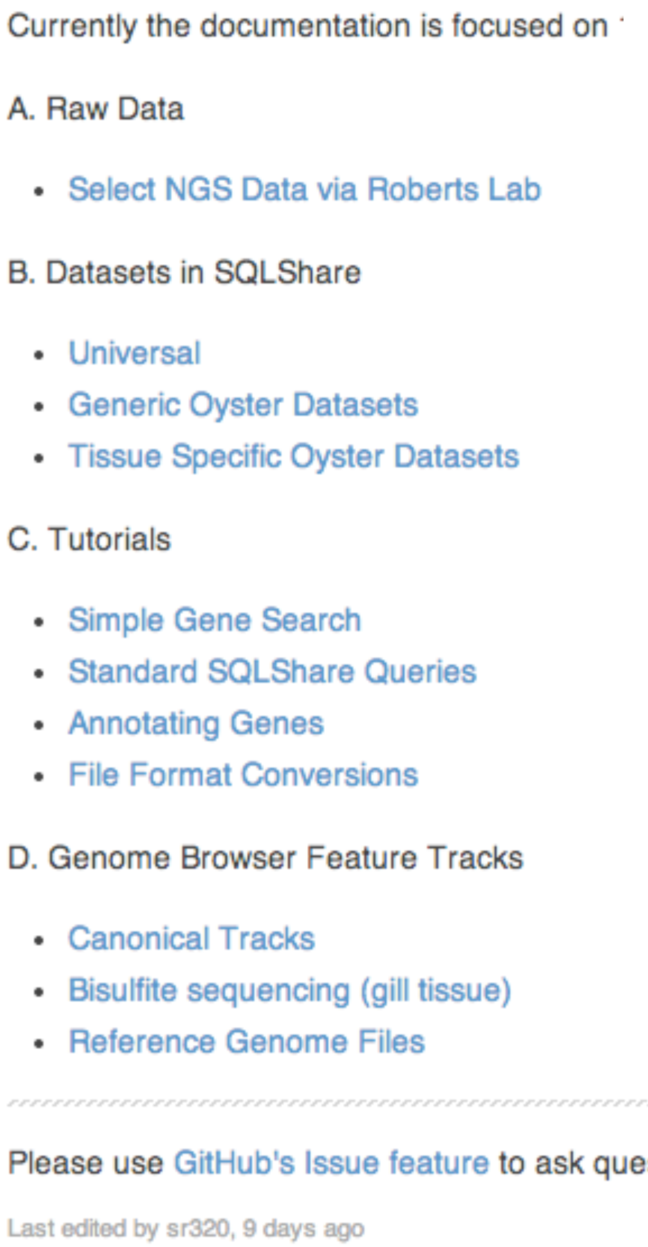
This respository provides access to genomic data and workflows (IPython notebooks) that are being integrated as part of effort to increase effeciency of biological discovery. The **wiki** associated with this repository will serve as the *primary means for documentation*. Currently the documentation is focused on 1) **describing current datasets** and 2) **providing workflow tutorials**.

In brief, data in the form of delimited text files is aggregated into SQLShare where they can be easily queried. Below is schematic representation of the different types of datasets.

During the initial phases the focus is on the Pacific oyster and primary data from the Roberts Lab.

Select IPython Notebooks

- Static Data Tables - Universal
- Static Data Tables - Annotations



Currently the documentation is focused on

A. Raw Data

- Select NGS Data via Roberts Lab

B. Datasets in SQLShare

- Universal
- Generic Oyster Datasets
- Tissue Specific Oyster Datasets

C. Tutorials

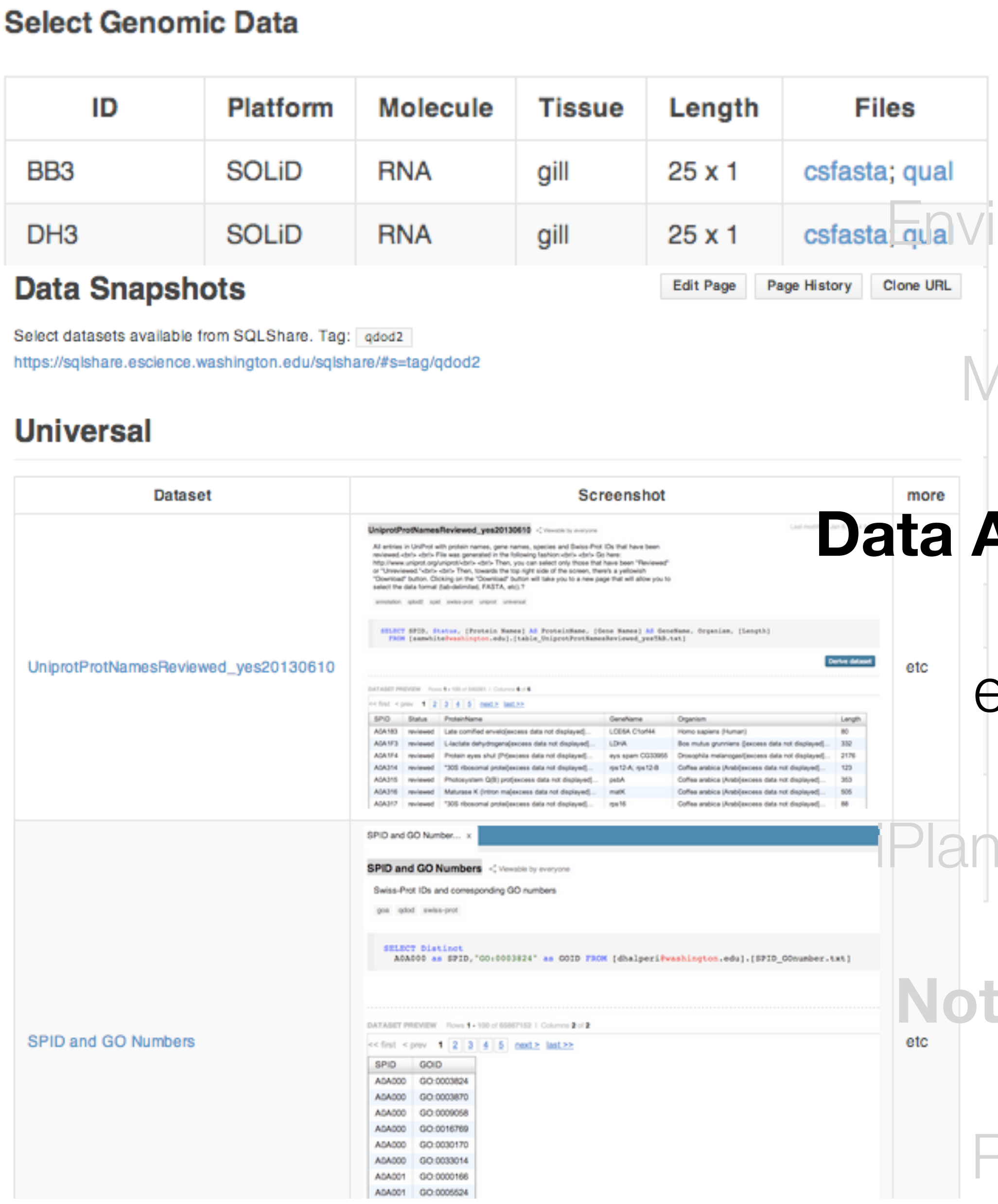
- Simple Gene Search
- Standard SQLShare Queries
- Annotating Genes
- File Format Conversions

D. Genome Browser Feature Tracks

- Canonical Tracks
- Bisulfite sequencing (gill tissue)
- Reference Genome Files

Please use GitHub's Issue feature to ask question, report problems, or suggest features.

Last edited by sr320, 9 days ago



Select Genomic Data

ID	Platform	Molecule	Tissue	Length	Files
BB3	SOLiD	RNA	gill	25 x 1	csfasta; qual
DH3	SOLiD	RNA	gill	25 x 1	csfasta; qual

Data Snapshots

Select datasets available from SQLShare. Tag: `qdod2`
https://sqlshare.escience.washington.edu/sqlshare/#s=tag/qdod2

Universal

Dataset	Screenshot	more
UniprotProtNamesReviewed_yes20130610		etc
SPID and GO Numbers		etc

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

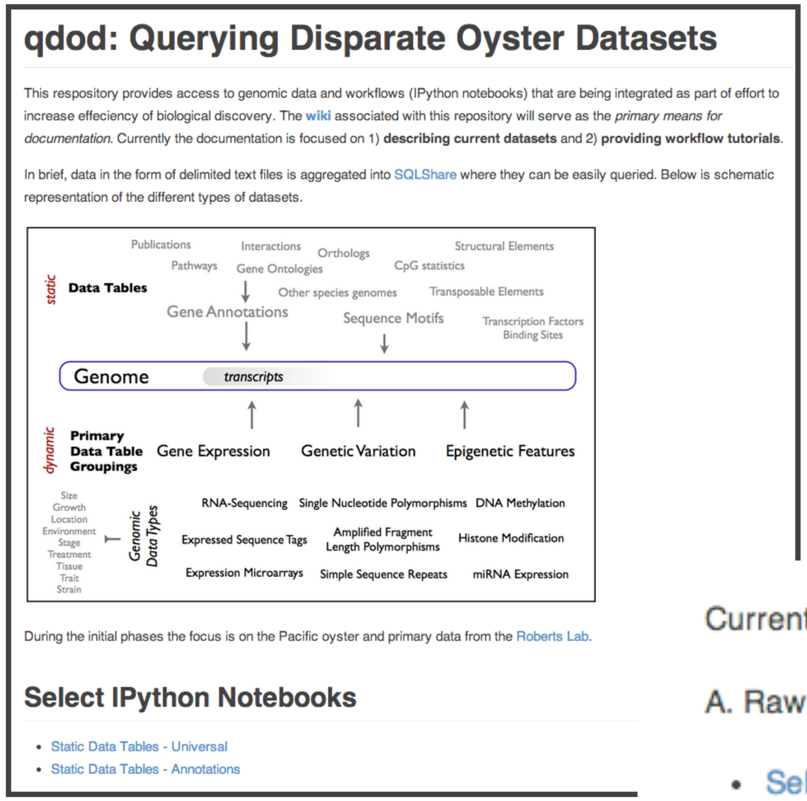
Platforms

Open Science

Data

everything else...

github.com/sr320/qdod/wiki



qdod: Querying Disparate Oyster Datasets

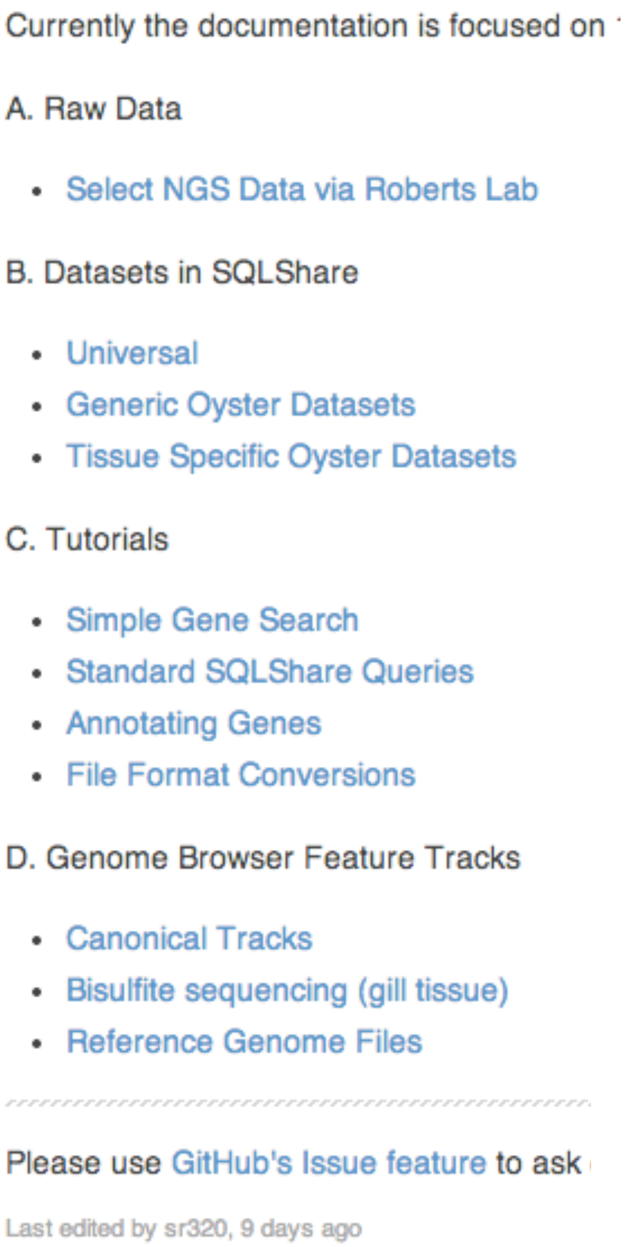
This respository provides access to genomic data and workflows (IPython notebooks) that are being integrated as part of effort to increase effeciency of biological discovery. The **wiki** associated with this repository will serve as the *primary means for documentation*. Currently the documentation is focused on 1) **describing current datasets** and 2) **providing workflow tutorials**.

In brief, data in the form of delimited text files is aggregated into SQLShare where they can be easily queried. Below is schematic representation of the different types of datasets.

During the initial phases the focus is on the Pacific oyster and primary data from the Roberts Lab.

Select IPython Notebooks

- Static Data Tables - Universal
- Static Data Tables - Annotations



Currently the documentation is focused on

A. Raw Data

- Select NGS Data via Roberts Lab

B. Datasets in SQLShare

- Universal
- Generic Oyster Datasets
- Tissue Specific Oyster Datasets

C. Tutorials

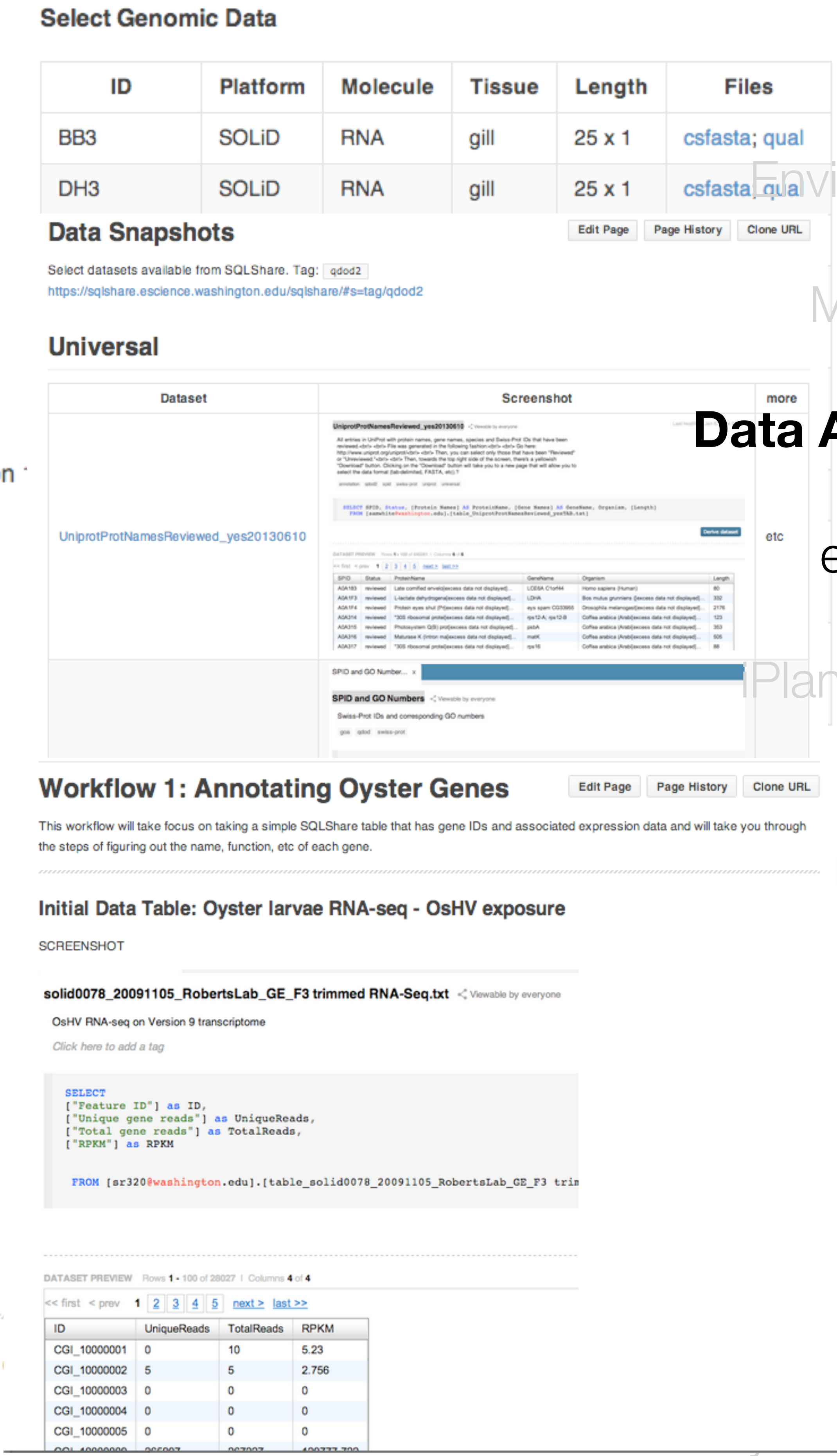
- Simple Gene Search
- Standard SQLShare Queries
- Annotating Genes
- File Format Conversions

D. Genome Browser Feature Tracks

- Canonical Tracks
- Bisulfite sequencing (gill tissue)
- Reference Genome Files

Please use GitHub's Issue feature to ask

Last edited by sr320, 9 days ago



Select Genomic Data

ID	Platform	Molecule	Tissue	Length	Files
BB3	SOLiD	RNA	gill	25 x 1	csfasta; qual
DH3	SOLiD	RNA	gill	25 x 1	csfasta; qua

Data Snapshots

Select datasets available from SQLShare. Tag: `qdod2`
https://sqlshare.escience.washington.edu/sqlshare/#s=tag/qdod2

Universal

Dataset	Screenshot	more
UniprotProtNamesReviewed_yes20130610		etc

Workflow 1: Annotating Oyster Genes

This workflow will take focus on taking a simple SQLShare table that has gene IDs and associated expression data and will take you through the steps of figuring out the name, function, etc of each gene.

Initial Data Table: Oyster larvae RNA-seq - OsHV exposure

SCREENSHOT

solid0078_20091105_RobertsLab_GE_F3 trimmed RNA-Seq.txt Viewable by everyone

OsHV RNA-seq on Version 9 transcriptome

Click here to add a tag

```
SELECT
["Feature ID"] as ID,
["Unique gene reads"] as UniqueReads,
["Total gene reads"] as TotalReads,
["RPKM"] as RPKM

 FROM [sr320@washington.edu].[table_solid0078_20091105_RobertsLab_GE_F3 trim
```

DATASET PREVIEW Rows 1 - 100 of 28027 | Columns 4 of 4

<< first < prev 1 2 3 4 5 next > last >>

ID	UniqueReads	TotalReads	RPKM
CGI_10000001	0	10	5.23
CGI_10000002	5	5	2.756
CGI_10000003	0	0	0
CGI_10000004	0	0	0
CGI_10000005	0	0	0

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

ebooks

ationale

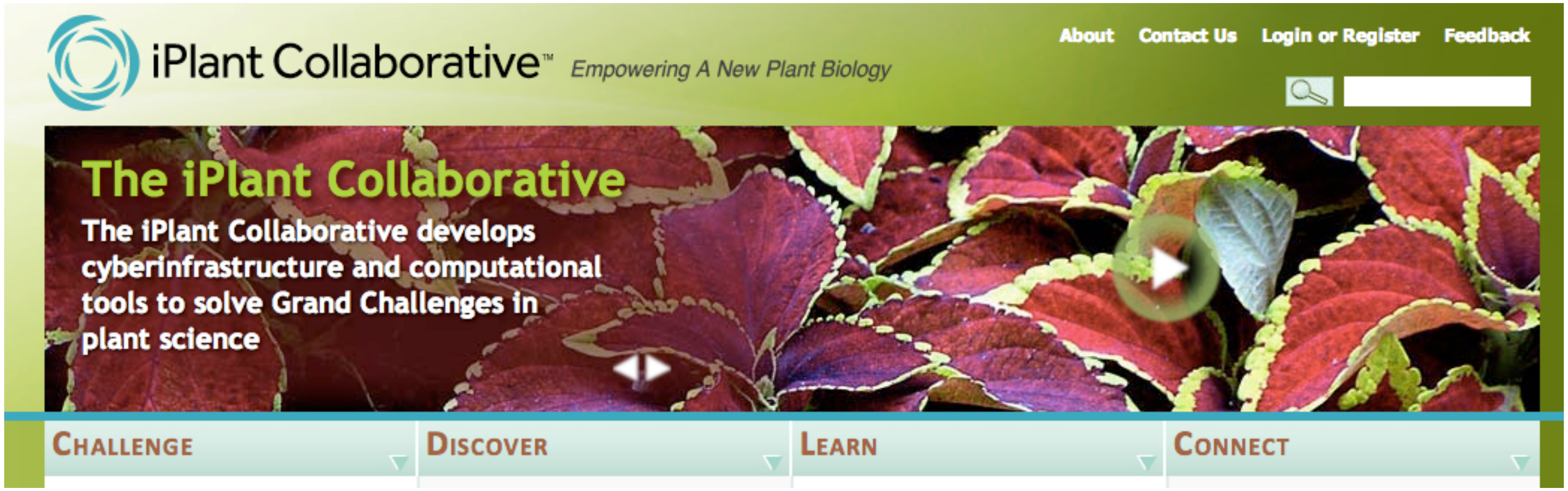
latforms

Science

Data

g else...

github.com/sr320/qdod/wiki



The iPlant Collaborative

The iPlant Collaborative develops cyberinfrastructure and computational tools to solve Grand Challenges in plant science

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

everything else...

The iPlant Collaborative

The iPlant Collaborative develops cyberinfrastructure and computational tools to solve Grand Challenges in plant science

Biology

Environment

Molecular

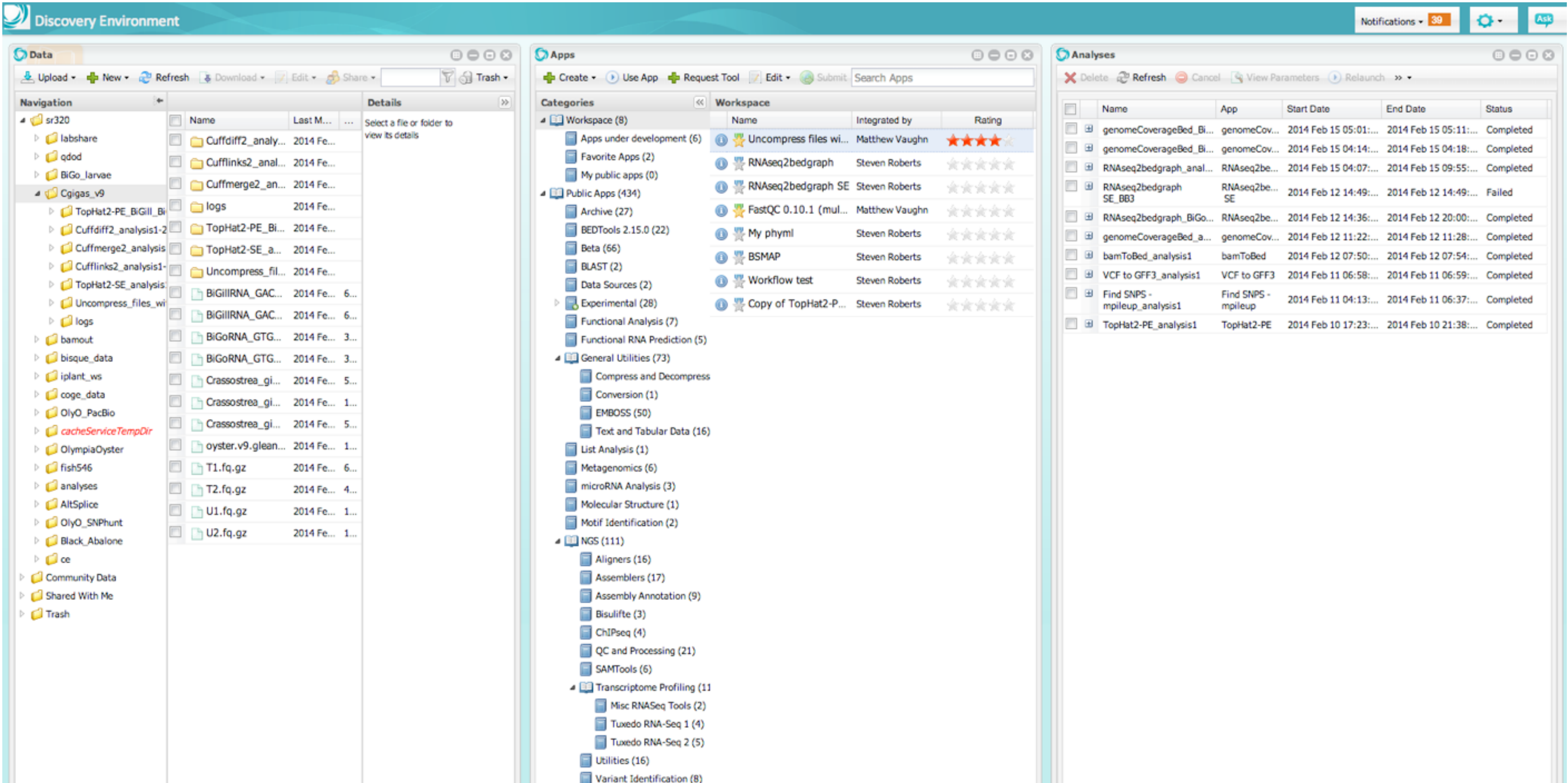
Data Analysis

Biology

Environment

Molecular

Data Analysis



- sr320
  - labshare
  - qdod
  - BiGo_larvae
  - Cgigas_v9
    - TopHat2-PE_BiGill_Bi
    - Cuffdiff2_analysis1-2
    - Cuffmerge2_analysis
    - Cufflinks2_analysis1-
    - TopHat2-SE_analysis
    - Uncompress_files_wi
    - logs
  - bamout
  - bisque_data
  - iplant_ws
  - coge_data
  - OlyO_PacBio
  - *cacheServiceTempDir*
  - OlympiaOyster
  - fish546
  - analyses
  - AltSplice
  - OlyO_SNPhunt
  - Black_Abalone
  - ce
- Community Data
- Shared With Me
- Trash

☐	Name	Last M...	...
☐	Cuffdiff2_analy...	2014 Fe...	
☐	Cufflinks2_anal...	2014 Fe...	
☐	Cuffmerge2_an...	2014 Fe...	
☐	logs	2014 Fe...	
☐	TopHat2-PE_Bi...	2014 Fe...	
☐	TopHat2-SE_a...	2014 Fe...	
☐	Uncompress_fil...	2014 Fe...	
☐	BiGillRNA_GAC...	2014 Fe...	6...
☐	BiGillRNA_GAC...	2014 Fe...	6...
☐	BiGoRNA_GTG...	2014 Fe...	3...
☐	BiGoRNA_GTG...	2014 Fe...	3...
☐	Crassostrea_gi...	2014 Fe...	5...
☐	Crassostrea_gi...	2014 Fe...	1...
☐	Crassostrea_gi...	2014 Fe...	5...
☐	oyster.v9.glean...	2014 Fe...	1...
☐	T1.fq.gz	2014 Fe...	6...
☐	T2.fq.gz	2014 Fe...	4...
☐	U1.fq.gz	2014 Fe...	1...
☐	U2.fq.gz	2014 Fe...	1...

Select a file or folder to view its details

- Workspace (8)
  - Apps under development (6)
  - Favorite Apps (2)
  - My public apps (0)
- Public Apps (434)
  - Archive (27)
  - BEDTools 2.15.0 (22)
  - Beta (66)
  - BLAST (2)
  - Data Sources (2)
  - Experimental (28)
  - Functional Analysis (7)
  - Functional RNA Prediction (5)
  - General Utilities (73)
    - Compress and Decompress
    - Conversion (1)
    - EMBOSS (50)
    - Text and Tabular Data (16)
  - List Analysis (1)
  - Metagenomics (6)
  - microRNA Analysis (3)
  - Molecular Structure (1)
  - Motif Identification (2)
  - NGS (111)
    - Aligners (16)
    - Assemblers (17)
    - Assembly Annotation (9)
    - Bisulifte (3)
    - ChIPseq (4)
    - QC and Processing (21)
    - SAMTools (6)
    - Transcriptome Profiling (11
      - Misc RNASeq Tools (2)
      - Tuxedo RNA-Seq 1 (4)
      - Tuxedo RNA-Seq 2 (5)
    - Utilities (16)
    - Variant Identification (8)

Name	Integrated by	Rating
Uncompress files wi...	Matthew Vaughn	★★★★☆
RNAseq2bedgraph	Steven Roberts	☆☆☆☆☆
RNAseq2bedgraph SE	Steven Roberts	☆☆☆☆☆
FastQC 0.10.1 (mul...	Matthew Vaughn	☆☆☆☆☆
My phyml	Steven Roberts	☆☆☆☆☆
BSMAP	Steven Roberts	☆☆☆☆☆
Workflow test	Steven Roberts	☆☆☆☆☆
Copy of TopHat2-P...	Steven Roberts	☆☆☆☆☆

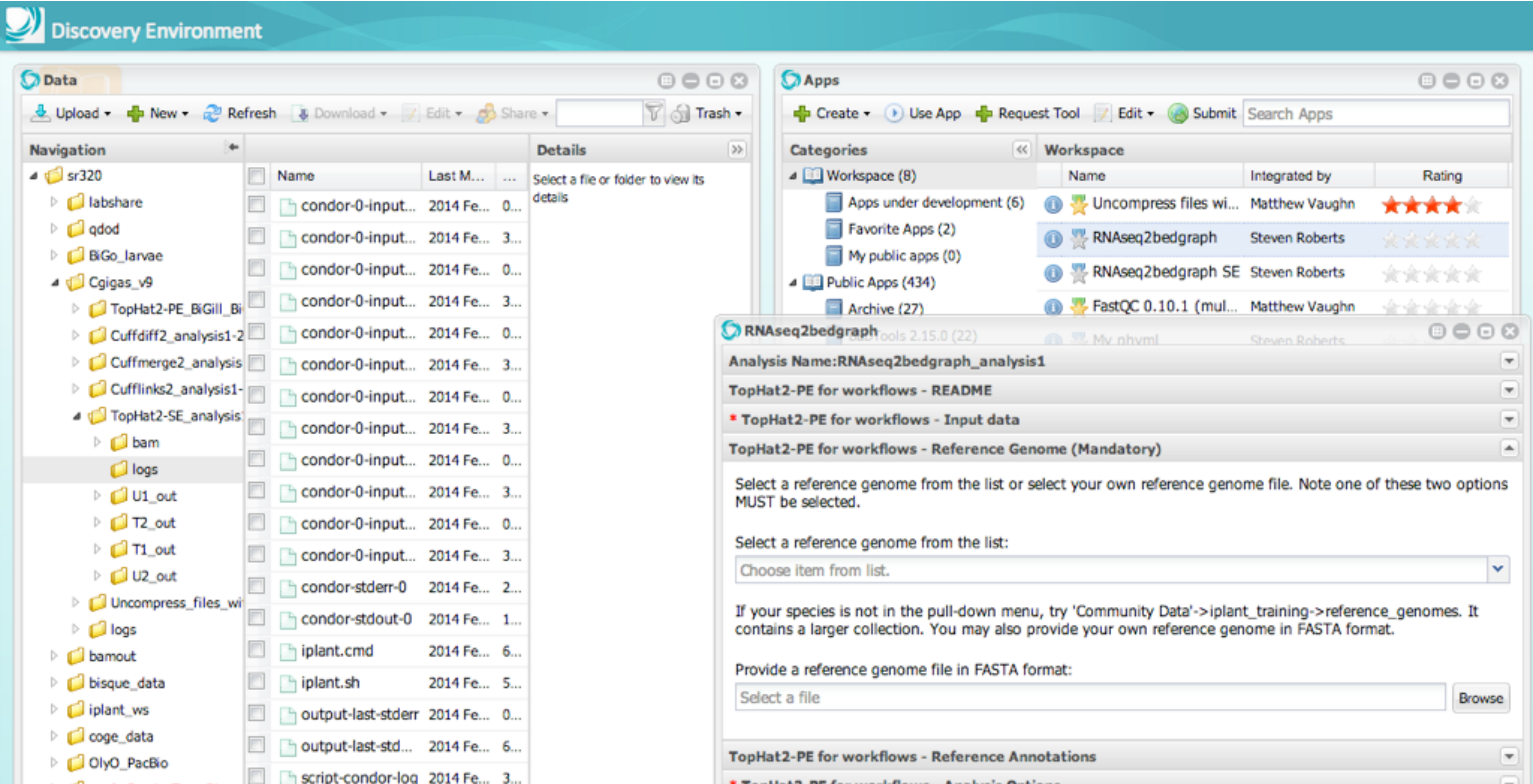
☐		Name	App	Start Date	End Date	Status
☐	⊞	genomeCoverageBed_Bi...	genomeCov...	2014 Feb 15 05:01:...	2014 Feb 15 05:11:...	Completed
☐	⊞	genomeCoverageBed_Bi...	genomeCov...	2014 Feb 15 04:14:...	2014 Feb 15 04:18:...	Completed
☐	⊞	RNAseq2bedgraph_anal...	RNAseq2be...	2014 Feb 15 04:07:...	2014 Feb 15 09:55:...	Completed
☐	⊞	RNAseq2bedgraph SE_BB3	RNAseq2be... SE	2014 Feb 12 14:49:...	2014 Feb 12 14:49:...	Failed
☐	⊞	RNAseq2bedgraph_BiGo...	RNAseq2be...	2014 Feb 12 14:36:...	2014 Feb 12 20:00:...	Completed
☐	⊞	genomeCoverageBed_a...	genomeCov...	2014 Feb 12 11:22:...	2014 Feb 12 11:28:...	Completed
☐	⊞	bamToBed_analysis1	bamToBed	2014 Feb 12 07:50:...	2014 Feb 12 07:54:...	Completed
☐	⊞	VCF to GFF3_analysis1	VCF to GFF3	2014 Feb 11 06:58:...	2014 Feb 11 06:59:...	Completed
☐	⊞	Find SNPS - mpileup_analysis1	Find SNPS - mpileup	2014 Feb 11 04:13:...	2014 Feb 11 06:37:...	Completed
☐	⊞	TopHat2-PE_analysis1	TopHat2-PE	2014 Feb 10 17:23:...	2014 Feb 10 21:38:...	Completed

Biology

Environment

Molecular

Data Analysis



Select a file or folder to view its details

Name	Last M...	...
condor-0-input...	2014 Fe...	0...
condor-0-input...	2014 Fe...	3...
condor-0-input...	2014 Fe...	0...
condor-0-input...	2014 Fe...	3...
condor-0-input...	2014 Fe...	0...
condor-0-input...	2014 Fe...	3...
condor-0-input...	2014 Fe...	0...
condor-0-input...	2014 Fe...	3...
condor-0-input...	2014 Fe...	0...
condor-0-input...	2014 Fe...	3...
condor-0-input...	2014 Fe...	0...
condor-0-input...	2014 Fe...	3...
condor-stderr-0	2014 Fe...	2...
condor-stdout-0	2014 Fe...	1...
iplant.cmd	2014 Fe...	6...
iplant.sh	2014 Fe...	5...
output-last-stderr	2014 Fe...	0...
output-last-std...	2014 Fe...	6...
script-condor-log	2014 Fe...	3...

Name	Integrated by	Rating
Uncompress files wi...	Matthew Vaughn	★★★★☆
RNAseq2bedgraph	Steven Roberts	☆☆☆☆☆
RNAseq2bedgraph SE	Steven Roberts	☆☆☆☆☆
FastQC 0.10.1 (mul...	Matthew Vaughn	☆☆☆☆☆

Analysis Name:RNAseq2bedgraph_analysis1

TopHat2-PE for workflows - README

* TopHat2-PE for workflows - Input data

TopHat2-PE for workflows - Reference Genome (Mandatory)

Select a reference genome from the list or select your own reference genome file. Note one of these two options MUST be selected.

Select a reference genome from the list:

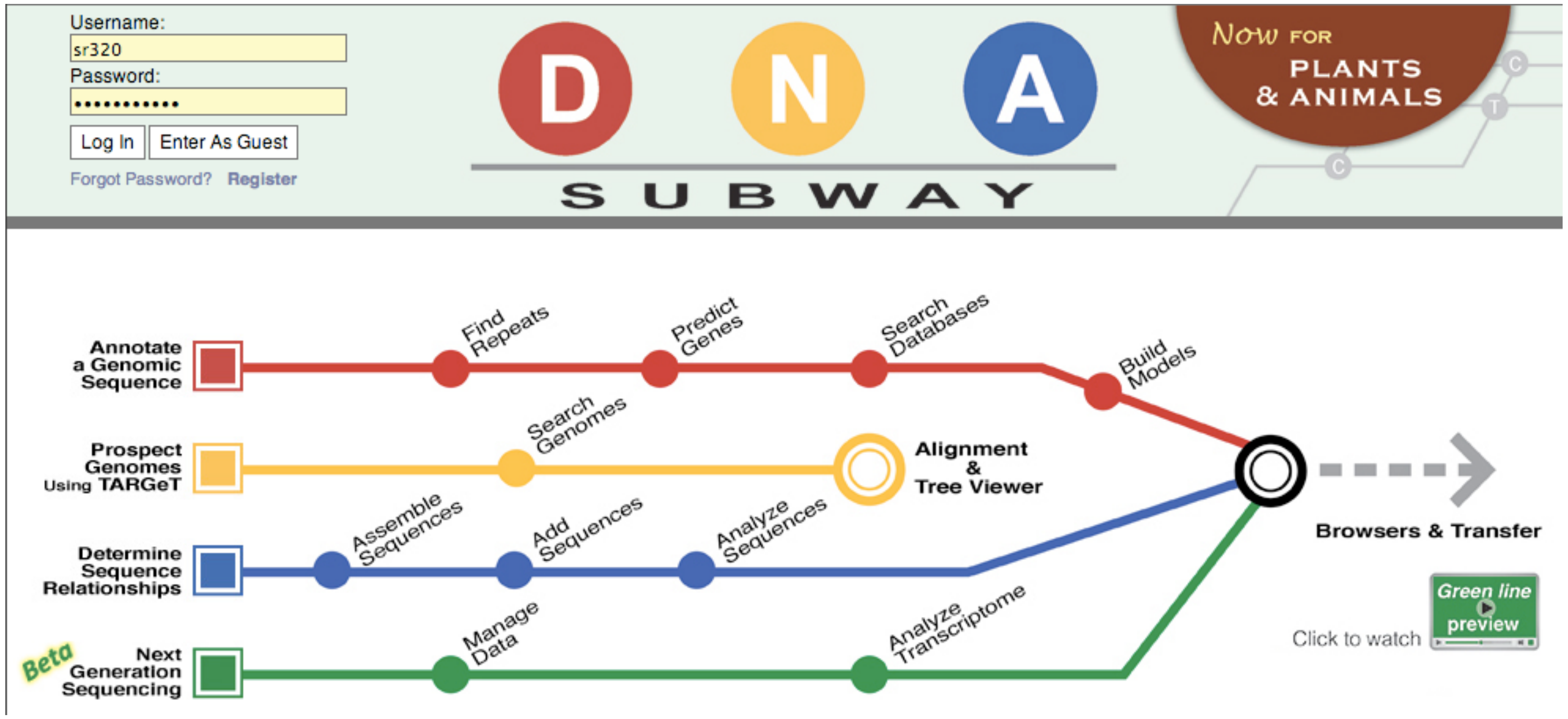
Choose item from list.

If your species is not in the pull-down menu, try 'Community Data'->iplant_training->reference_genomes. It contains a larger collection. You may also provide your own reference genome in FASTA format.

Provide a reference genome file in FASTA format:

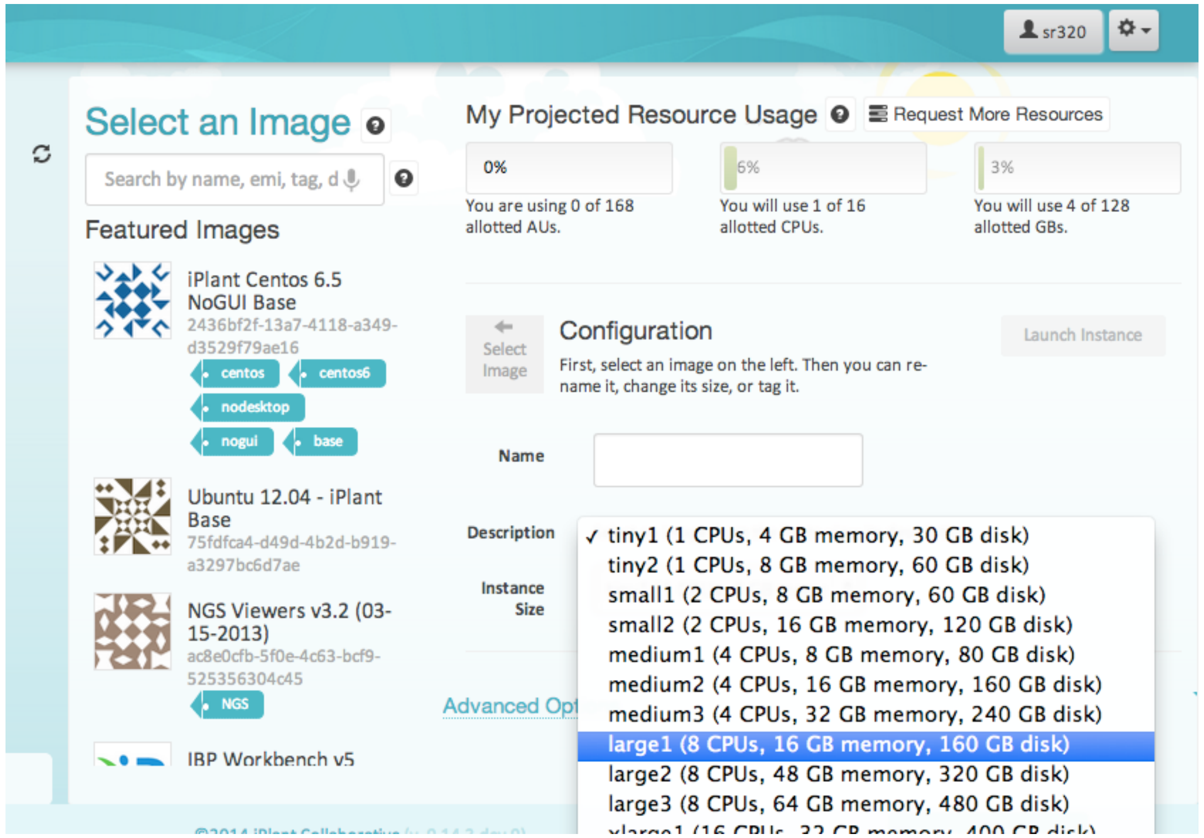
Select a file

TopHat2-PE for workflows - Reference Annotations



Username: sr320

Password: •••••••••••



You are using 0 of 168 allotted AUs.

You will use 1 of 16 allotted CPUs.

You will use 4 of 128 allotted GBs.

Configuration

First, select an image on the left. Then you can re-name it, change its size, or tag it.

Name

Description

Instance Size

- ✓ tiny1 (1 CPUs, 4 GB memory, 30 GB disk)
- tiny2 (1 CPUs, 8 GB memory, 60 GB disk)
- small1 (2 CPUs, 8 GB memory, 60 GB disk)
- small2 (2 CPUs, 16 GB memory, 120 GB disk)
- medium1 (4 CPUs, 8 GB memory, 80 GB disk)
- medium2 (4 CPUs, 16 GB memory, 160 GB disk)
- medium3 (4 CPUs, 32 GB memory, 240 GB disk)
- large1 (8 CPUs, 16 GB memory, 160 GB disk)
- large2 (8 CPUs, 48 GB memory, 320 GB disk)
- large3 (8 CPUs, 64 GB memory, 480 GB disk)

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

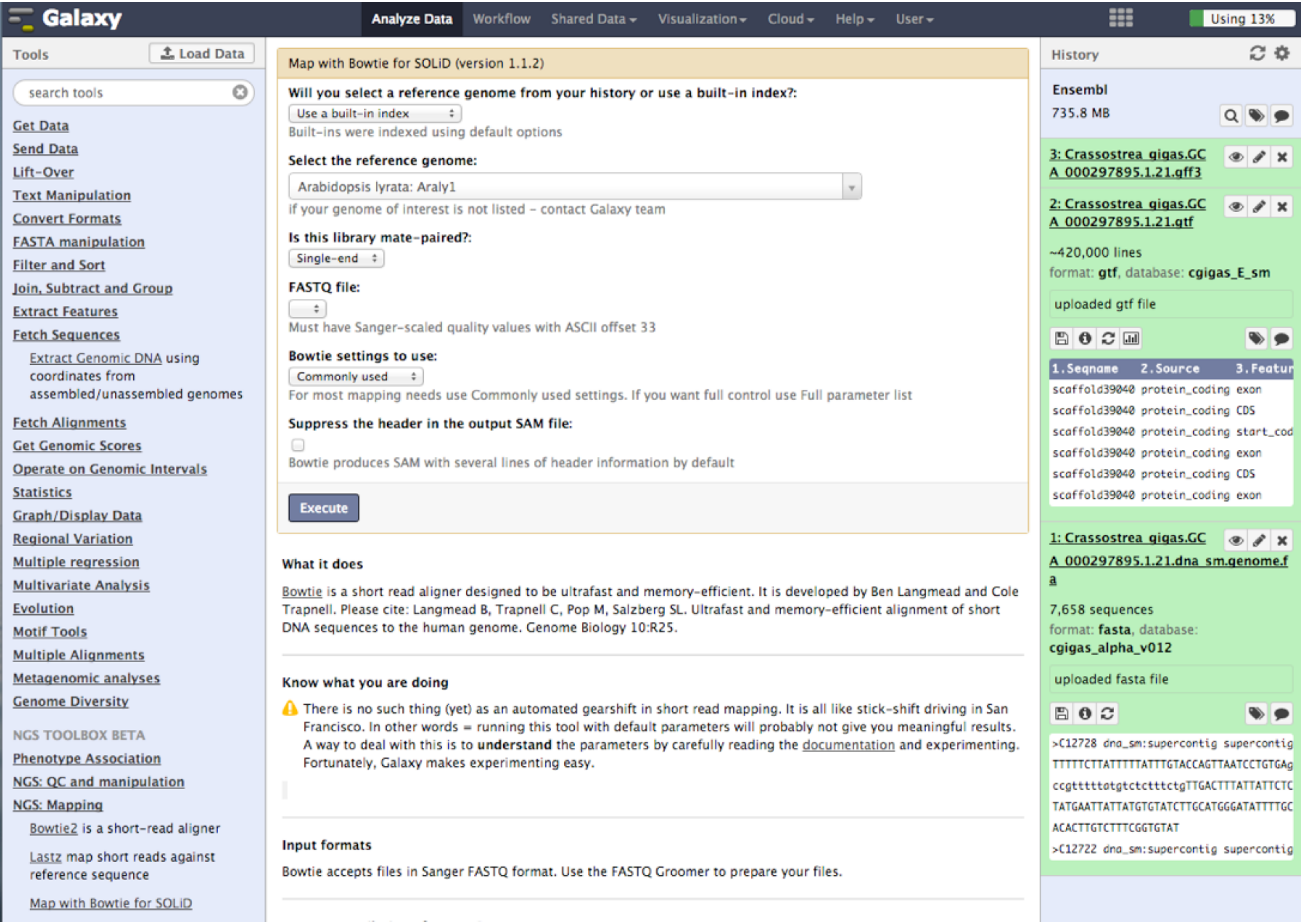
Rationale

Platforms

Open Science

Data

everything else...



- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract and Group
- Extract Features
- Fetch Sequences
  - Extract Genomic DNA using coordinates from assembled/unassembled genomes
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- Genome Diversity

NGS TOOLBOX BETA

- Phenotype Association
- NGS: QC and manipulation
- NGS: Mapping
  - Bowtie2 is a short-read aligner
  - Lastz map short reads against reference sequence
  - Map with Bowtie for SOLiD

Map with Bowtie for SOLiD (version 1.1.2)

Will you select a reference genome from your history or use a built-in index?:
Use a built-in index
Built-ins were indexed using default options

Select the reference genome:
Arabidopsis lyrata: Araly1
If your genome of interest is not listed - contact Galaxy team

Is this library mate-paired?:
Single-end

FASTQ file:
Must have Sanger-scaled quality values with ASCII offset 33

Bowtie settings to use:
Commonly used
For most mapping needs use Commonly used settings. If you want full control use Full parameter list

Suppress the header in the output SAM file:
☐
Bowtie produces SAM with several lines of header information by default

What it does

Bowtie is a short read aligner designed to be ultrafast and memory-efficient. It is developed by Ben Langmead and Cole Trapnell. Please cite: Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

Know what you are doing

⚠ There is no such thing (yet) as an automated gearshift in short read mapping. It is all like stick-shift driving in San Francisco. In other words = running this tool with default parameters will probably not give you meaningful results. A way to deal with this is to **understand** the parameters by carefully reading the documentation and experimenting. Fortunately, Galaxy makes experimenting easy.

Input formats

Bowtie accepts files in Sanger FASTQ format. Use the FASTQ Groomer to prepare your files.

Ensembl
735.8 MB

3: Crassostrea_gigas.GCA_000297895.1.21.gff3

2: Crassostrea_gigas.GCA_000297895.1.21.gtf

~420,000 lines
format: **gtf**, database: **cgigas_E_sm**

uploaded gtf file

1.Seqname	2.Source	3.Featur
scaffold39040	protein_coding	exon
scaffold39040	protein_coding	CDS
scaffold39040	protein_coding	start_cod
scaffold39040	protein_coding	exon
scaffold39040	protein_coding	CDS
scaffold39040	protein_coding	exon

1: Crassostrea_gigas.GCA_000297895.1.21.dna_sm.genome.fa

7,658 sequences
format: **fasta**, database: **cgigas_alpha_v012**

uploaded fasta file

```
>C12728 dna_sm:supercontig supercontig
TTTTTCTTATTTTTATTTGTACCAGTTAATCCTGTGAg
ccgttttttatgtctctttctgTTGACTTTATTATTCTC
TATGAATTATTATGTGTATCTTGCATGGGATATTTTGC
ACACTTGTCTTTCGGTGTAT
>C12722 dna_sm:supercontig supercontig
```

Biology

Environment

Molecular

Data Analysis

eScience

iPlant **Galaxy**

Notebooks

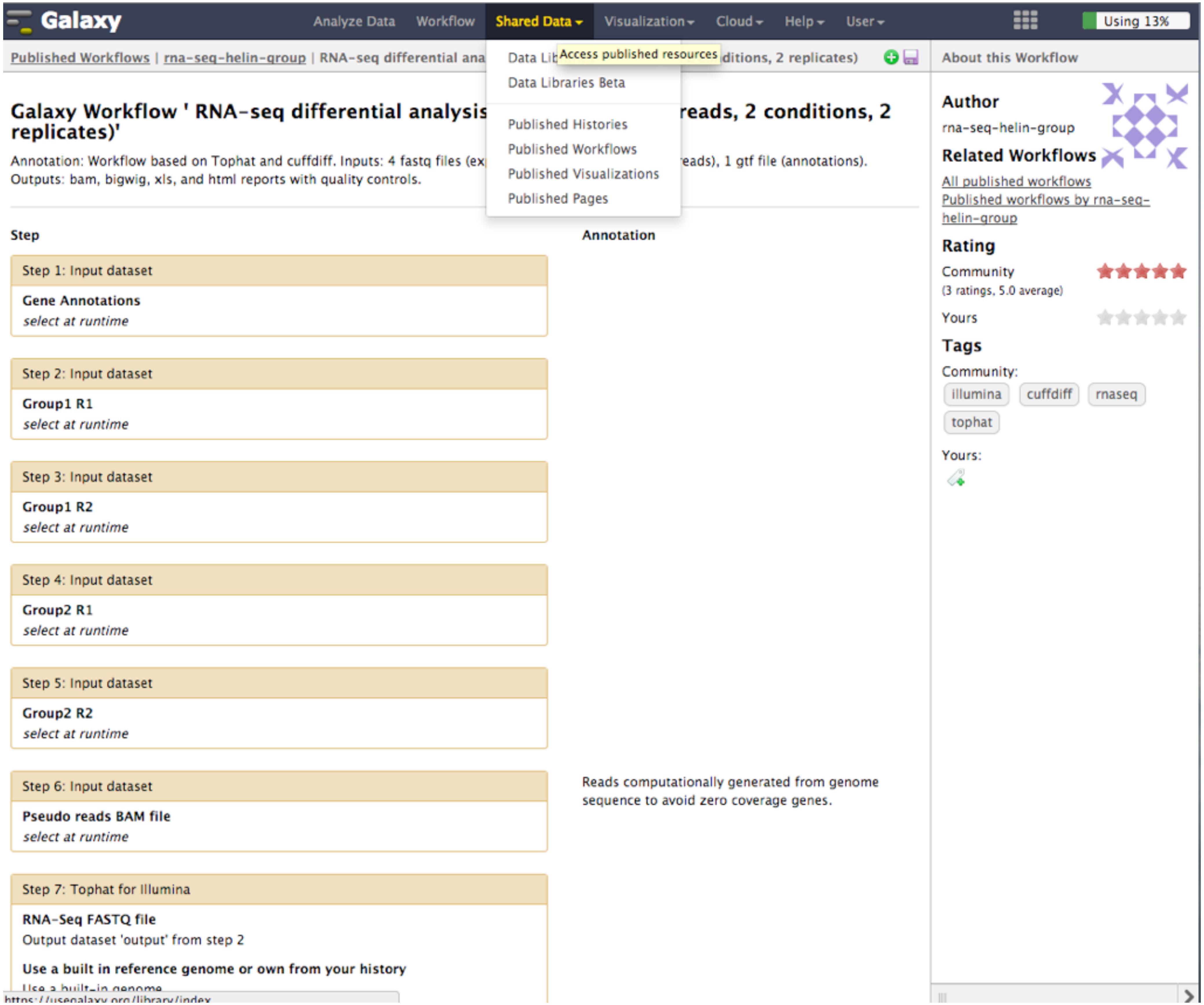
Rationale

Platforms

Open Science

Data

everything else...



Galaxy Workflow ' RNA-seq differential analysis … reads, 2 conditions, 2 replicates)'

Annotation: Workflow based on Tophat and cuffdiff. Inputs: 4 fastq files (ex… eads), 1 gtf file (annotations). Outputs: bam, bigwig, xls, and html reports with quality controls.

Step	Annotation
Step 1: Input dataset **Gene Annotations** *select at runtime*	
Step 2: Input dataset **Group1 R1** *select at runtime*	
Step 3: Input dataset **Group1 R2** *select at runtime*	
Step 4: Input dataset **Group2 R1** *select at runtime*	
Step 5: Input dataset **Group2 R2** *select at runtime*	
Step 6: Input dataset **Pseudo reads BAM file** *select at runtime*	Reads computationally generated from genome sequence to avoid zero coverage genes.
Step 7: Tophat for Illumina **RNA-Seq FASTQ file** Output dataset 'output' from step 2 **Use a built in reference genome or own from your history** Use a built-in genome	

https://usegalaxy.org/library/index

Author

rna-seq-helin-group

Related Workflows

All published workflows

Published workflows by rna-seq-helin-group

Rating

Community ★★★★★

(3 ratings, 5.0 average)

Yours ☆☆☆☆☆

Tags

Community:

illumina cuffdiff rnaseq tophat

Yours:

Biology

Environment

Molecular

Data Analysis

eScience

iPlant **Galaxy**

Notebooks

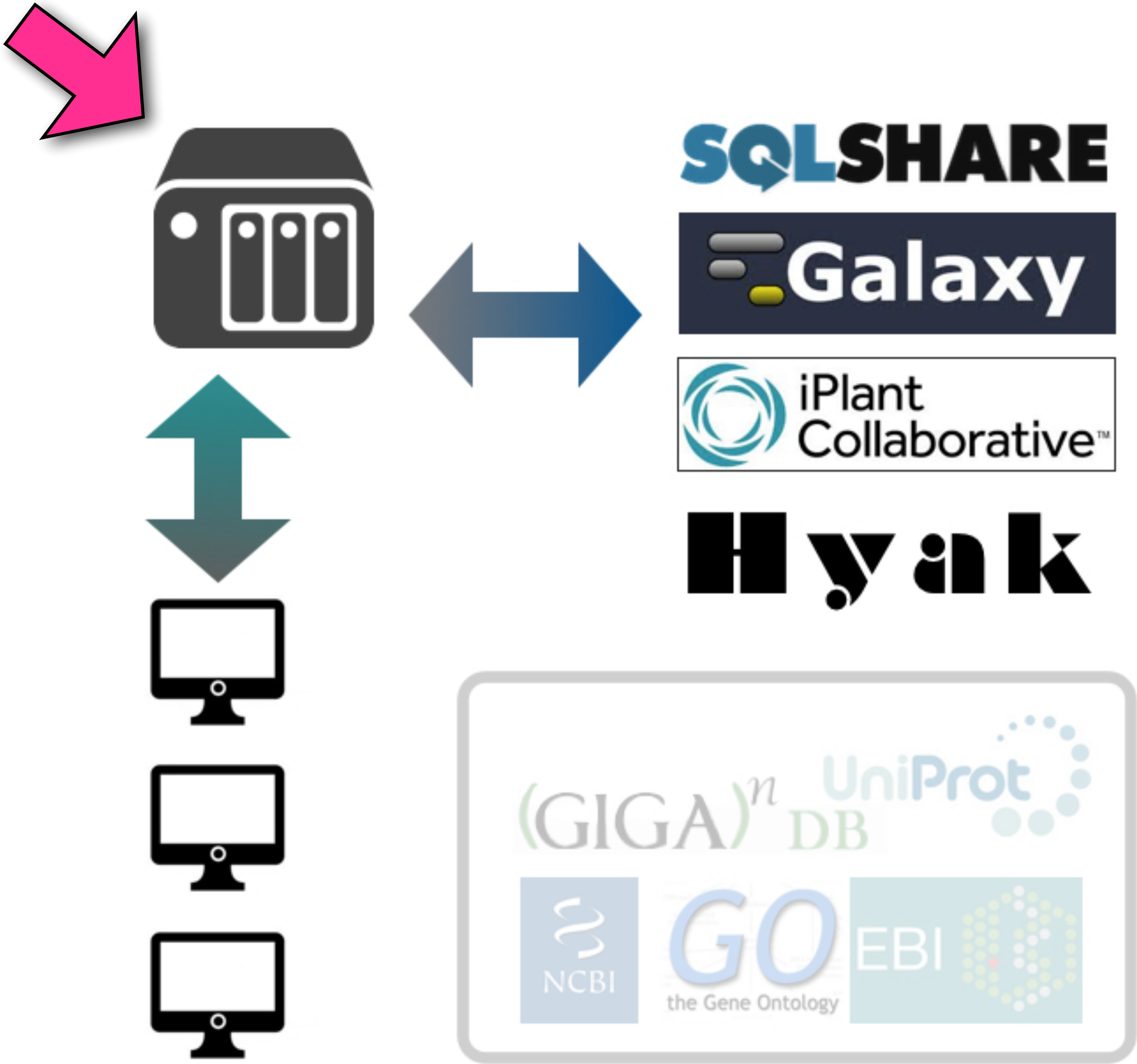
Rationale

Platforms

Open Science

Data

everything else...



Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

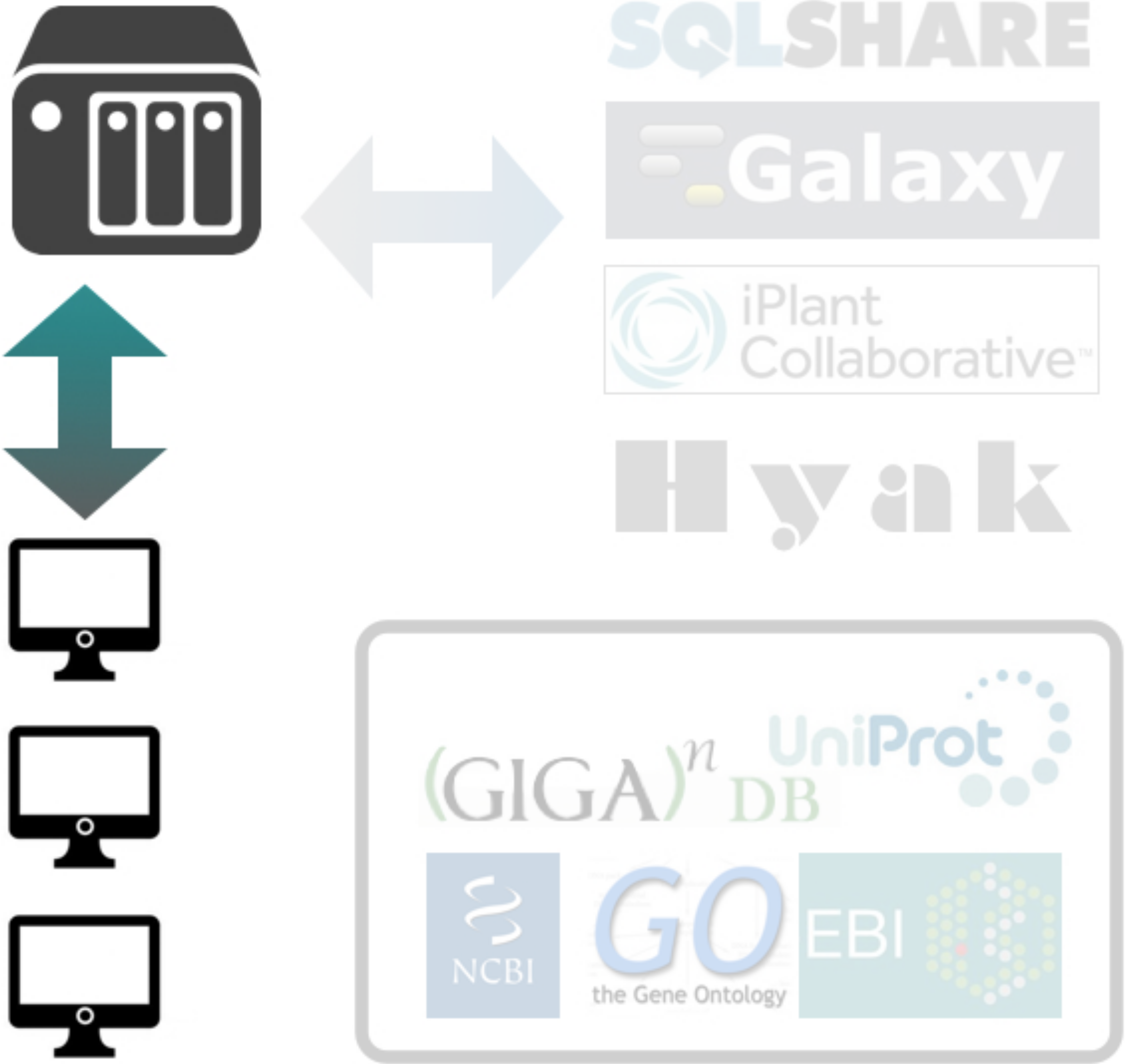
Rationale

Platforms

Open Science

Data

everything else...



Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

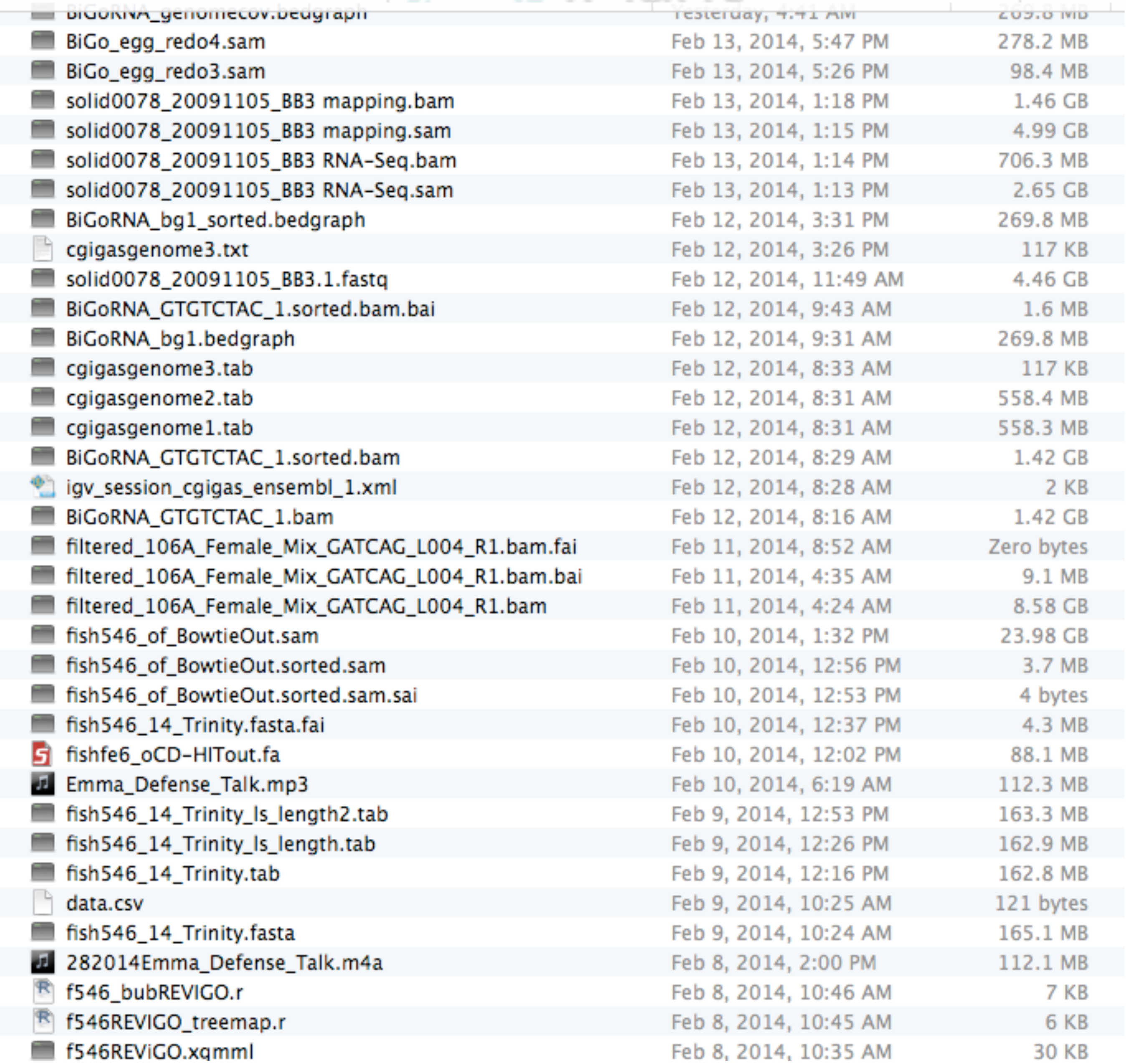
Data

everything else...

SQLSHARE

Galaxy

iPlant



BiGoRNA_genomecov.bedgraph	Yesterday, 4:41 AM	269.8 MB
BiGo_egg_redo4.sam	Feb 13, 2014, 5:47 PM	278.2 MB
BiGo_egg_redo3.sam	Feb 13, 2014, 5:26 PM	98.4 MB
solid0078_20091105_BB3 mapping.bam	Feb 13, 2014, 1:18 PM	1.46 GB
solid0078_20091105_BB3 mapping.sam	Feb 13, 2014, 1:15 PM	4.99 GB
solid0078_20091105_BB3 RNA-Seq.bam	Feb 13, 2014, 1:14 PM	706.3 MB
solid0078_20091105_BB3 RNA-Seq.sam	Feb 13, 2014, 1:13 PM	2.65 GB
BiGoRNA_bg1_sorted.bedgraph	Feb 12, 2014, 3:31 PM	269.8 MB
cgigasgenome3.txt	Feb 12, 2014, 3:26 PM	117 KB
solid0078_20091105_BB3.1.fastq	Feb 12, 2014, 11:49 AM	4.46 GB
BiGoRNA_GTGTCTAC_1.sorted.bam.bai	Feb 12, 2014, 9:43 AM	1.6 MB
BiGoRNA_bg1.bedgraph	Feb 12, 2014, 9:31 AM	269.8 MB
cgigasgenome3.tab	Feb 12, 2014, 8:33 AM	117 KB
cgigasgenome2.tab	Feb 12, 2014, 8:31 AM	558.4 MB
cgigasgenome1.tab	Feb 12, 2014, 8:31 AM	558.3 MB
BiGoRNA_GTGTCTAC_1.sorted.bam	Feb 12, 2014, 8:29 AM	1.42 GB
igv_session_cgigas_ensembl_1.xml	Feb 12, 2014, 8:28 AM	2 KB
BiGoRNA_GTGTCTAC_1.bam	Feb 12, 2014, 8:16 AM	1.42 GB
filtered_106A_Female_Mix_GATCAG_L004_R1.bam.fai	Feb 11, 2014, 8:52 AM	Zero bytes
filtered_106A_Female_Mix_GATCAG_L004_R1.bam.bai	Feb 11, 2014, 4:35 AM	9.1 MB
filtered_106A_Female_Mix_GATCAG_L004_R1.bam	Feb 11, 2014, 4:24 AM	8.58 GB
fish546_of_BowtieOut.sam	Feb 10, 2014, 1:32 PM	23.98 GB
fish546_of_BowtieOut.sorted.sam	Feb 10, 2014, 12:56 PM	3.7 MB
fish546_of_BowtieOut.sorted.sam.sai	Feb 10, 2014, 12:53 PM	4 bytes
fish546_14_Trinity.fasta.fai	Feb 10, 2014, 12:37 PM	4.3 MB
fishfe6_oCD-HITout.fa	Feb 10, 2014, 12:02 PM	88.1 MB
Emma_Defense_Talk.mp3	Feb 10, 2014, 6:19 AM	112.3 MB
fish546_14_Trinity_ls_length2.tab	Feb 9, 2014, 12:53 PM	163.3 MB
fish546_14_Trinity_ls_length.tab	Feb 9, 2014, 12:26 PM	162.9 MB
fish546_14_Trinity.tab	Feb 9, 2014, 12:16 PM	162.8 MB
data.csv	Feb 9, 2014, 10:25 AM	121 bytes
fish546_14_Trinity.fasta	Feb 9, 2014, 10:24 AM	165.1 MB
282014Emma_Defense_Talk.m4a	Feb 8, 2014, 2:00 PM	112.1 MB
f546_bubREVIGO.r	Feb 8, 2014, 10:46 AM	7 KB
f546REVIGO_treemap.r	Feb 8, 2014, 10:45 AM	6 KB
f546REViGO.xgmml	Feb 8, 2014, 10:35 AM	30 KB

eagle.fish.washington.edu

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

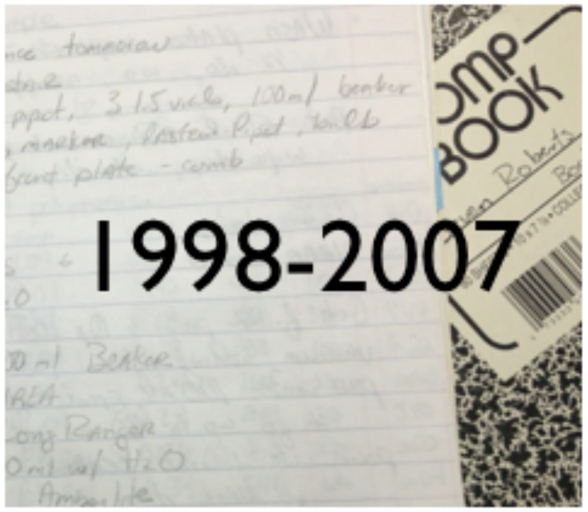
Platforms

Open Science

Data

everything else...

The Evolution of My Lab Notebook



1998-2007

2008-2009

2010-2013

now

Open Notebook Science

> ... there is a URL to a laboratory notebook that is freely available and indexed on common search engines. It does not necessarily have to look like a paper notebook but it is essential that all of the information available to the researchers to make their conclusions is equally available to the rest of the world.
>
> —Jean-Claude Bradley

Open Notebook Science

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

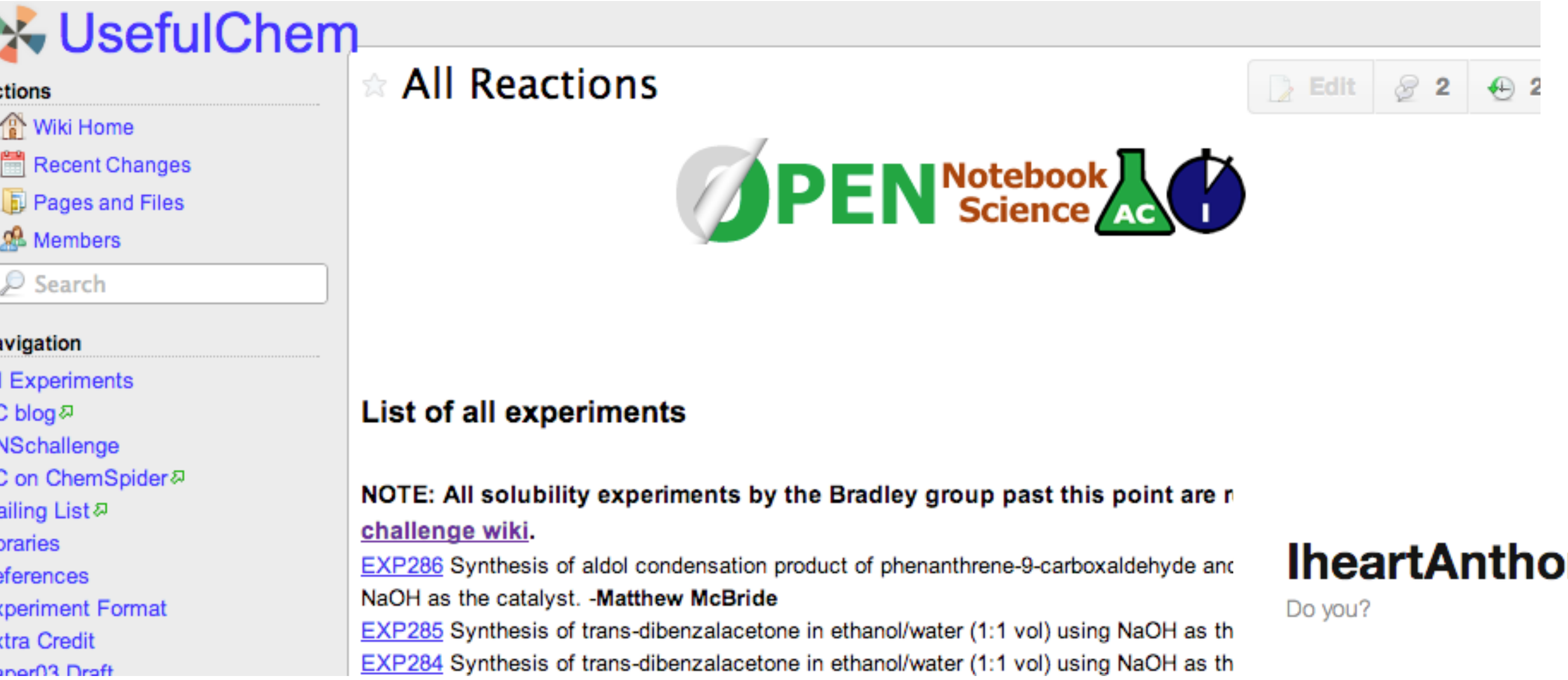
Platforms

Open Science

Data

everything else...

Open Notebook Science



All Reactions

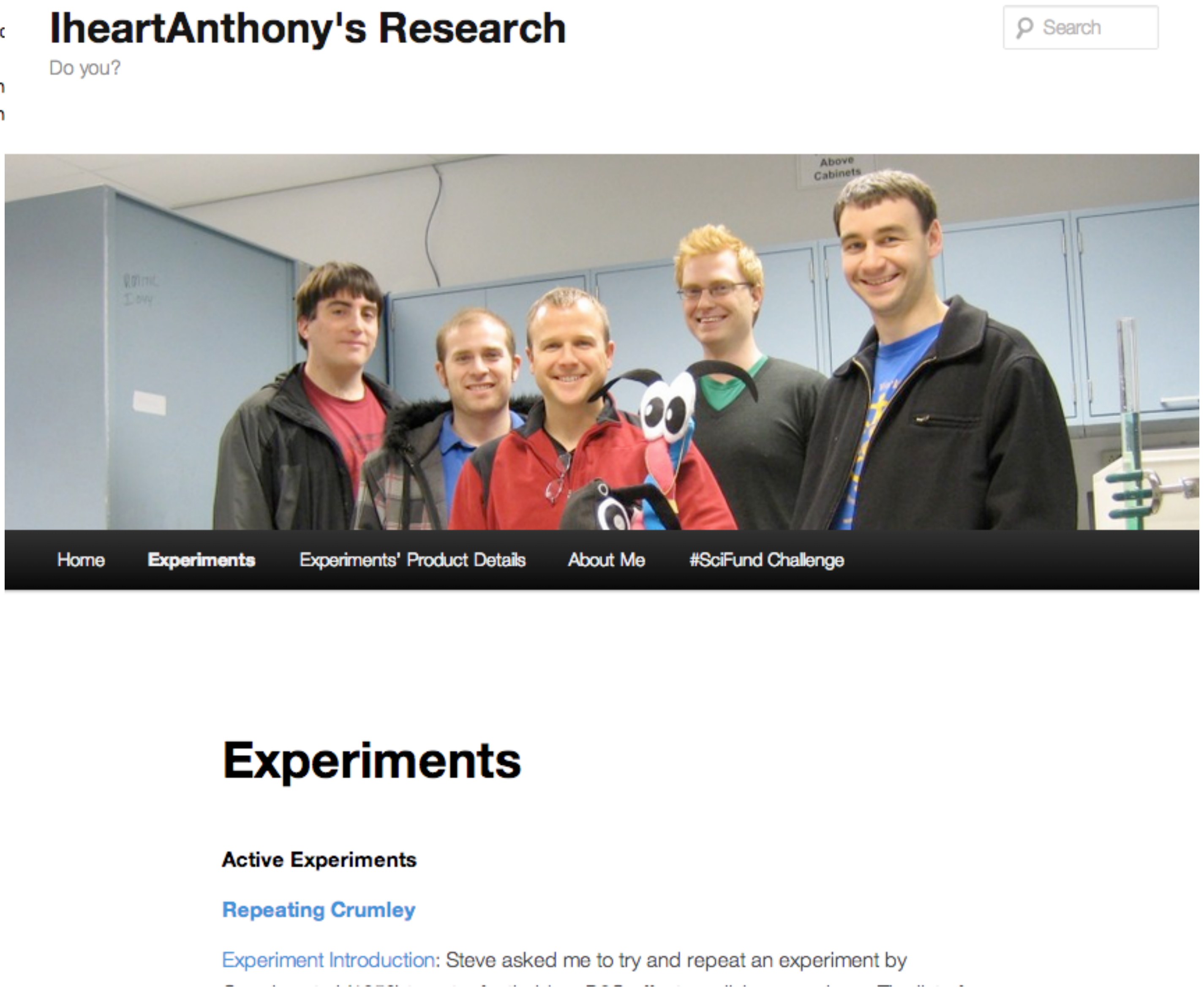
List of all experiments

NOTE: All solubility experiments by the Bradley group past this point are r **challenge wiki**.

EXP286 Synthesis of aldol condensation product of phenanthrene-9-carboxaldehyde anc NaOH as the catalyst. -**Matthew McBride**

EXP285 Synthesis of trans-dibenzalacetone in ethanol/water (1:1 vol) using NaOH as th

EXP284 Synthesis of trans-dibenzalacetone in ethanol/water (1:1 vol) using NaOH as th



IheartAnthony's Research

Do you?

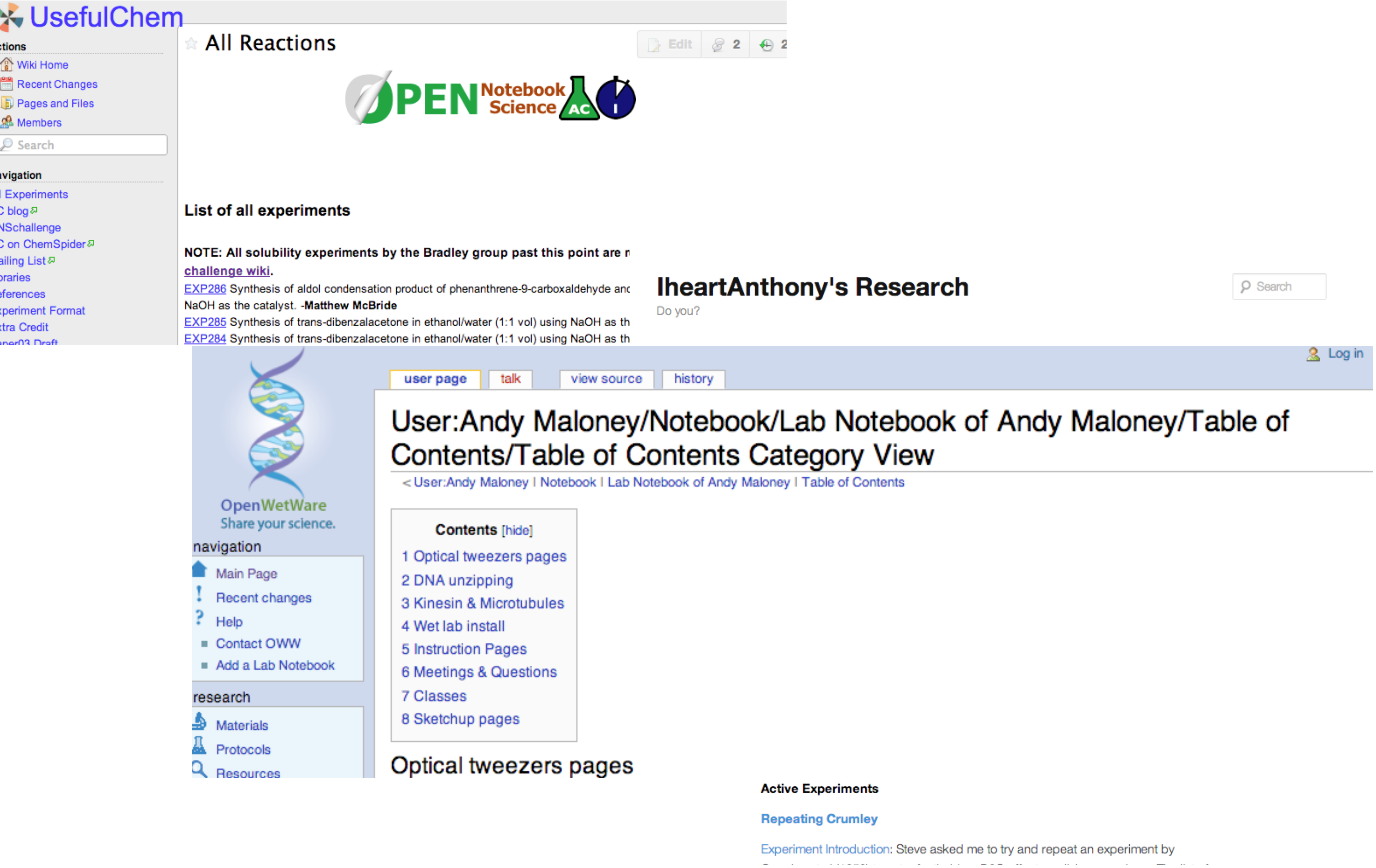
Experiments

Active Experiments

Repeating Crumley

Experiment Introduction: Steve asked me to try and repeat an experiment by

Open Notebook Science



UsefulChem

ctions
- Wiki Home
- Recent Changes
- Pages and Files
- Members

Search

vigation
- Experiments
- C blog
- NSchallenge
- C on ChemSpider
- ailing List
- raries
- eferences
- xperiment Format
- xtra Credit
- per03 Draft

All Reactions

List of all experiments

NOTE: All solubility experiments by the Bradley group past this point are r **challenge wiki**.

EXP286 Synthesis of aldol condensation product of phenanthrene-9-carboxaldehyde anc NaOH as the catalyst. -**Matthew McBride**

EXP285 Synthesis of trans-dibenzalacetone in ethanol/water (1:1 vol) using NaOH as th

EXP284 Synthesis of trans-dibenzalacetone in ethanol/water (1:1 vol) using NaOH as th

IheartAnthony's Research

Do you?

navigation
- Main Page
- Recent changes
- Help
- Contact OWW
- Add a Lab Notebook

research
- Materials
- Protocols
- Resources

User:Andy Maloney/Notebook/Lab Notebook of Andy Maloney/Table of Contents/Table of Contents Category View

< User:Andy Maloney | Notebook | Lab Notebook of Andy Maloney | Table of Contents

Contents [hide]

Optical tweezers pages

Active Experiments

Repeating Crumley

Experiment Introduction: Steve asked me to try and repeat an experiment by

Open Notebook Science

Carl Boettiger

Theoretical Ecology and Evolution

I'm Carl Boettiger , a post-doctoral researcher with Marc Mangel and Steve Munch in the Center for Stock Assessment Research, at UC Santa Cruz, working on regime shifts in ecology and evolution. Such shifts mark the most dramatic events in complex systems. What causes these transitions to occur? Can we detect or forecast these changes? How do we robustly manage ecosystems that experience such tipping points? Understanding and forecasting these events challenges the capacity of both our statistical methods and the data available. My research aims to extend and quantify the limits of understanding through nonparametric Bayesian inference, high performance computing and big data

Lab Notebook

(Introduction)

Entries

Notes

31 Jan 2014

pageviews: (not calculated)

Scratch notes on uncertainty.

Function definition: multiple_uncertainty.m

Settings:

```
f=@(x, h) max( (x-h) * (1 - (x-h) ./ 100)
 + (x-h), 0);
x_grid = [0:5:100];
h_grid = x_grid; % Must be same dimension
s as x_grid, or L91 errors...
Tmax = 5;
sigma g = 0.1;
```

Read more

Notes

09 Jan 2014

pageviews: 8

- fishbase API discussions continue
- some discussion / work on rfigshare continues (As Karthik writes comprehensive test suite, yay).
- Possible ontological term for simulated data? modeling and simulation operation looking for example using this term now…
- Request for term on OBI
- Moore DDD investigators, full call

.@NOAA gives a glimpse of what the future of

Read more

Notes

07 Jan 2014

pageviews: 6

Coding

- fishbase API discussions
- rfishbase updates pushed to CRAN
- Merges and minor bug fixes for rfigshare
- RNeXML: After some cooperative sleuth-work, we successfully resolved issue #23, uncovering a bug (missing feature really) in xmllint.

as usual, detailed activity on github.

Read more

Notes

06 Jan 2014

pageviews: 3

- fishbase API discussions
- knitcitations (push previous updates to CRAN)
- data extraction from sardine collapse story (see EML in data/), news article

Open Science Literature Highlights

02 Jan 2014

pageviews: 40

A student recently asked me for some recommendations for an article on an open science topic for a journal club. Since I haven't jotted these down in one place before, though I might copy my reply here for

Notes

10 Dec 2013

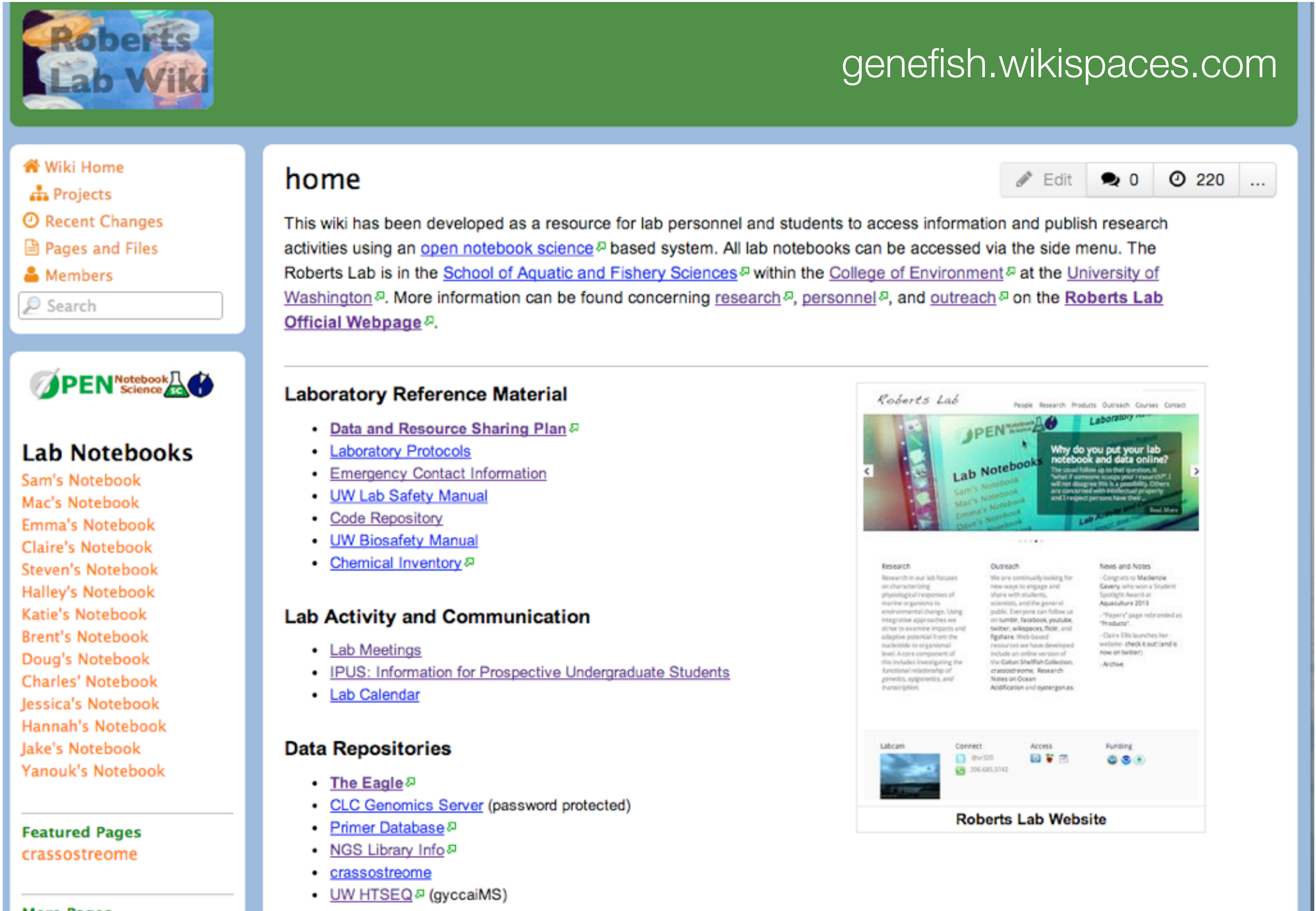
pageviews: 64

Consider the model

$$dX_t = \alpha(\theta - X_t)dt + \sigma dB_t \quad (1)$$

carlboettiger.info/lab-notebook

Open Notebook Science



home

This wiki has been developed as a resource for lab personnel and students to access information and publish research activities using an open notebook science based system. All lab notebooks can be accessed via the side menu. The Roberts Lab is in the School of Aquatic and Fishery Sciences within the College of Environment at the University of Washington. More information can be found concerning research, personnel, and outreach on the **Roberts Lab Official Webpage**.

Laboratory Reference Material

- **Data and Resource Sharing Plan**
- Laboratory Protocols
- Emergency Contact Information
- UW Lab Safety Manual
- Code Repository
- UW Biosafety Manual
- Chemical Inventory

Lab Activity and Communication

- Lab Meetings
- IPUS: Information for Prospective Undergraduate Students
- Lab Calendar

Data Repositories

- **The Eagle**
- CLC Genomics Server (password protected)
- Primer Database
- NGS Library Info
- crassostreome
- UW HTSEQ (gyccaiMS)

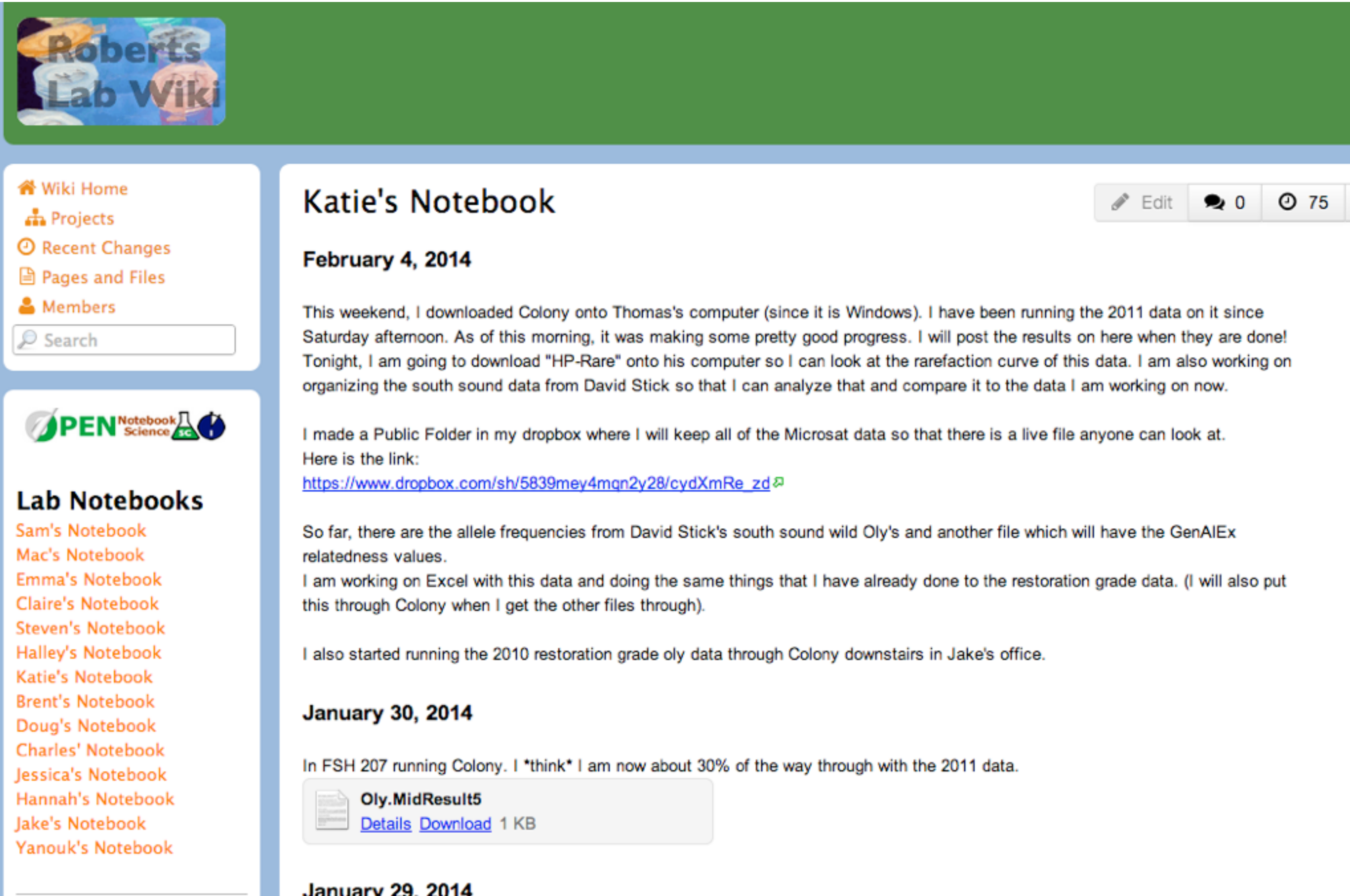
Roberts Lab Website

Lab Notebooks

Sam's Notebook
Mac's Notebook
Emma's Notebook
Claire's Notebook
Steven's Notebook
Halley's Notebook
Katie's Notebook
Brent's Notebook
Doug's Notebook
Charles' Notebook
Jessica's Notebook
Hannah's Notebook
Jake's Notebook
Yanouk's Notebook

Featured Pages
crassostreome

Open Notebook Science



Lab Notebooks

- Sam's Notebook
- Mac's Notebook
- Emma's Notebook
- Claire's Notebook
- Steven's Notebook
- Halley's Notebook
- Katie's Notebook
- Brent's Notebook
- Doug's Notebook
- Charles' Notebook
- Jessica's Notebook
- Hannah's Notebook
- Jake's Notebook
- Yanouk's Notebook

Katie's Notebook

February 4, 2014

This weekend, I downloaded Colony onto Thomas's computer (since it is Windows). I have been running the 2011 data on it since Saturday afternoon. As of this morning, it was making some pretty good progress. I will post the results on here when they are done! Tonight, I am going to download "HP-Rare" onto his computer so I can look at the rarefaction curve of this data. I am also working on organizing the south sound data from David Stick so that I can analyze that and compare it to the data I am working on now.

I made a Public Folder in my dropbox where I will keep all of the Microsat data so that there is a live file anyone can look at.
Here is the link:
https://www.dropbox.com/sh/5839mey4mqn2y28/cydXmRe_zd

So far, there are the allele frequencies from David Stick's south sound wild Oly's and another file which will have the GenAlEx relatedness values.
I am working on Excel with this data and doing the same things that I have already done to the restoration grade data. (I will also put this through Colony when I get the other files through).

I also started running the 2010 restoration grade oly data through Colony downstairs in Jake's office.

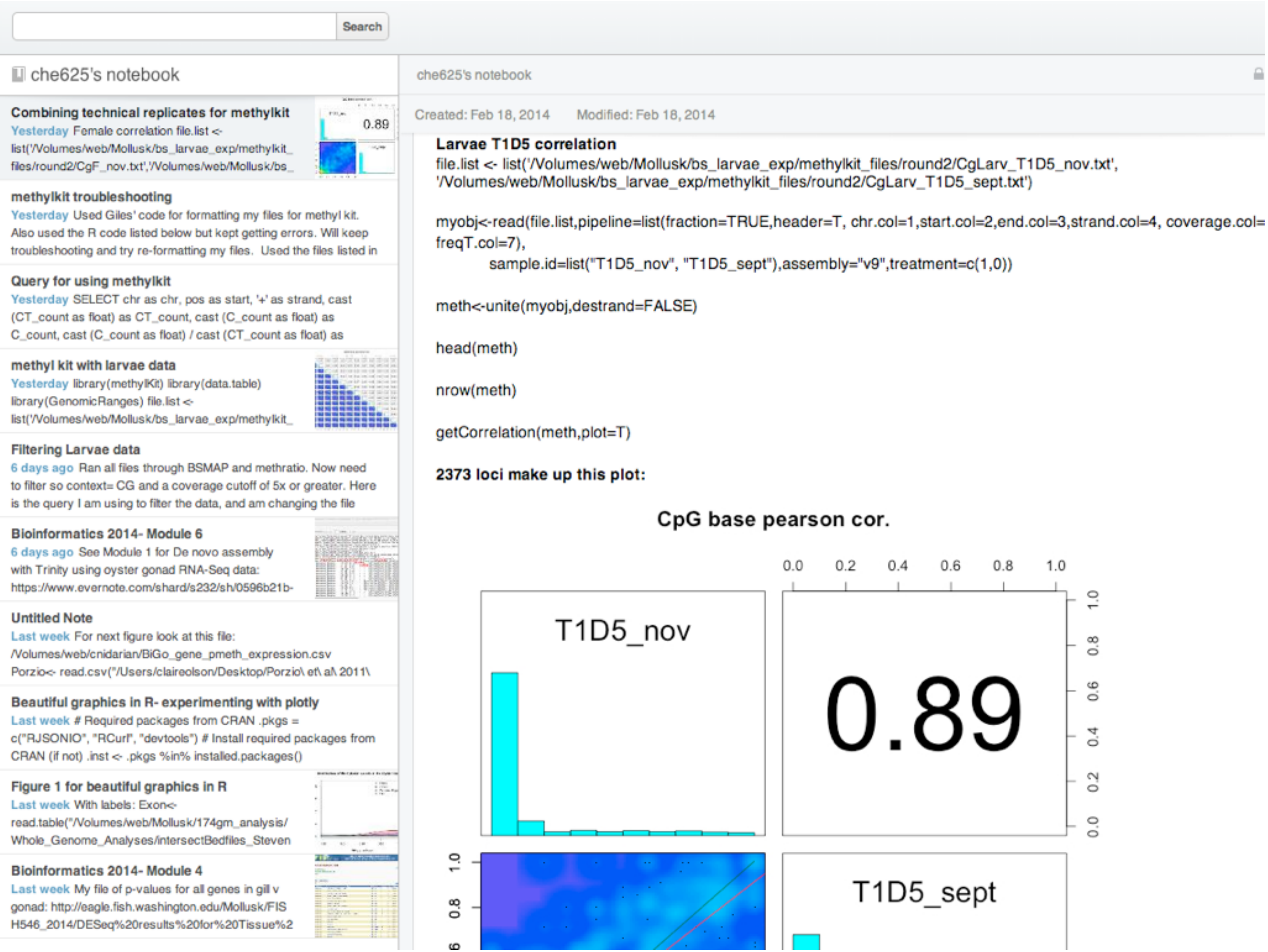
January 30, 2014

In FSH 207 running Colony. I *think* I am now about 30% of the way through with the 2011 data.

Oly.MidResult5
Details Download 1 KB

January 29, 2014

Open Notebook Science



Combining technical replicates for methylkit
Yesterday Female correlation file.list <- list('/Volumes/web/Mollusk/bs_larvae_exp/methylkit_files/round2/CgF_nov.txt','/Volumes/web/Mollusk/bs_

methylkit troubleshooting
Yesterday Used Giles' code for formatting my files for methyl kit. Also used the R code listed below but kept getting errors. Will keep troubleshooting and try re-formatting my files. Used the files listed in

Query for using methylkit
Yesterday SELECT chr as chr, pos as start, '+' as strand, cast (CT_count as float) as CT_count, cast (C_count as float) as C_count, cast (C_count as float) / cast (CT_count as float) as

methyl kit with larvae data
Yesterday library(methylKit) library(data.table) library(GenomicRanges) file.list <- list('/Volumes/web/Mollusk/bs_larvae_exp/methylkit_

Filtering Larvae data
6 days ago Ran all files through BSMAP and methratio. Now need to filter so context= CG and a coverage cutoff of 5x or greater. Here is the query I am using to filter the data, and am changing the file

Bioinformatics 2014- Module 6
6 days ago See Module 1 for De novo assembly with Trinity using oyster gonad RNA-Seq data: https://www.evernote.com/shard/s232/sh/0596b21b-

Untitled Note
Last week For next figure look at this file: /Volumes/web/cnidarian/BiGo_gene_pmeth_expression.csv Porzio<- read.csv("/Users/claireolson/Desktop/Porzio\ et\ al\ 2011\

Beautiful graphics in R- experimenting with plotly
Last week # Required packages from CRAN .pkgs = c("RJSONIO", "RCurl", "devtools") # Install required packages from CRAN (if not) .inst <- .pkgs %in% installed.packages()

Figure 1 for beautiful graphics in R
Last week With labels: Exon<- read.table("/Volumes/web/Mollusk/174gm_analysis/Whole_Genome_Analyses/intersectBedfiles_Steven

Bioinformatics 2014- Module 4
Last week My file of p-values for all genes in gill v gonad: http://eagle.fish.washington.edu/Mollusk/FISH546_2014/DESeq%20results%20for%20Tissue%2

Created: Feb 18, 2014 Modified: Feb 18, 2014

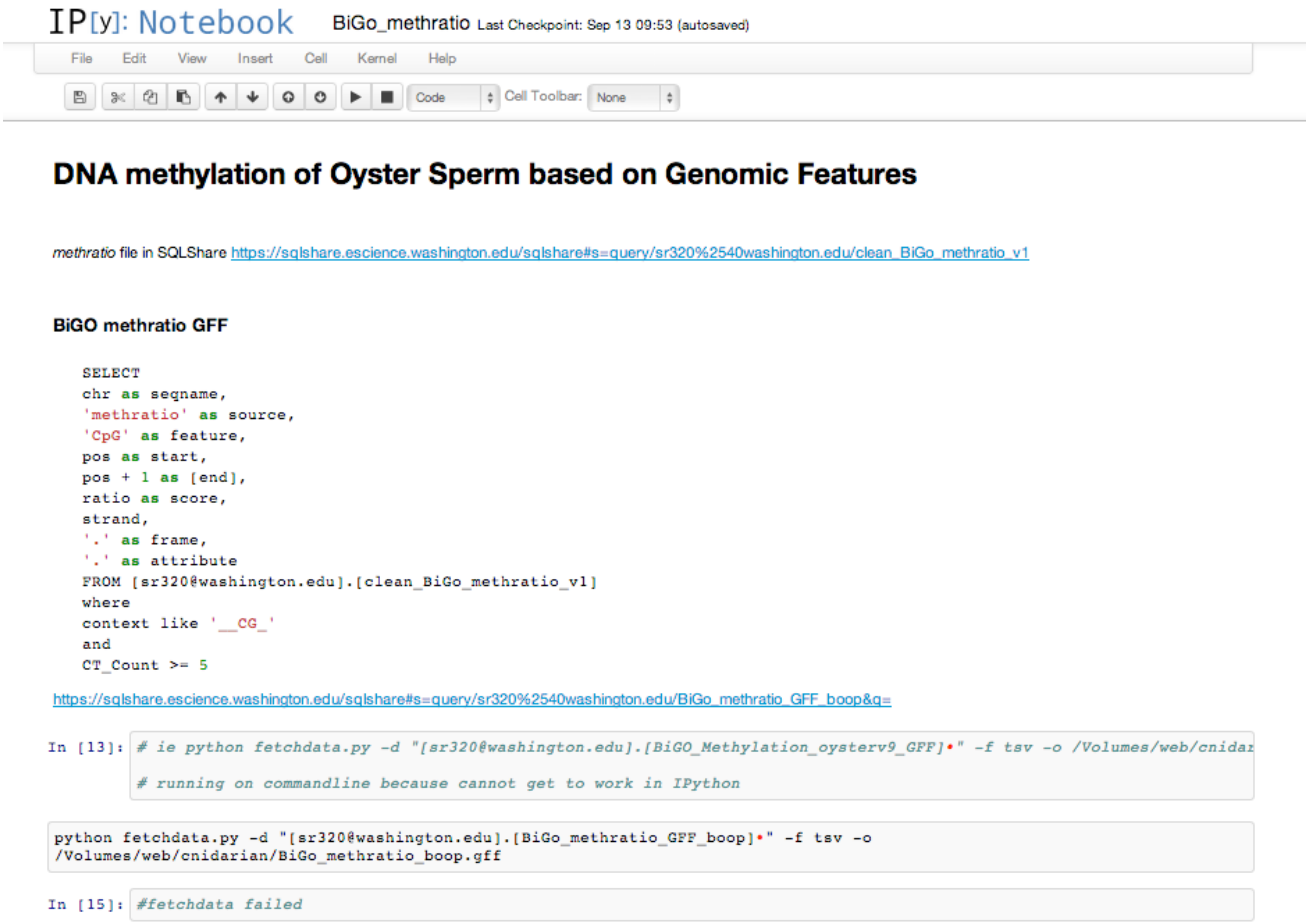
Larvae T1D5 correlation

```
file.list <- list('/Volumes/web/Mollusk/bs_larvae_exp/methylkit_files/round2/CgLarv_T1D5_nov.txt',
'/Volumes/web/Mollusk/bs_larvae_exp/methylkit_files/round2/CgLarv_T1D5_sept.txt')

myobj<-read(file.list,pipeline=list(fraction=TRUE,header=T, chr.col=1,start.col=2,end.col=3,strand.col=4, coverage.col=
freqT.col=7),
        sample.id=list("T1D5_nov", "T1D5_sept"),assembly="v9",treatment=c(1,0))

meth<-unite(myobj,destrand=FALSE)

head(meth)

nrow(meth)

getCorrelation(meth,plot=T)
```

2373 loci make up this plot:

evernote.com/pub/che625/che625snotebook

Open Notebook Science



DNA methylation of Oyster Sperm based on Genomic Features

methratio file in SQLShare https://sqlshare.escience.washington.edu/sqlshare#s=query/sr320%2540washington.edu/clean_BiGo_methratio_v1

BiGO methratio GFF

```
SELECT
chr as seqname,
'methratio' as source,
'CpG' as feature,
pos as start,
pos + 1 as [end],
ratio as score,
strand,
'.' as frame,
'.' as attribute
FROM [sr320@washington.edu].[clean_BiGo_methratio_v1]
where
context like '__CG_'
and
CT_Count >= 5
```

https://sqlshare.escience.washington.edu/sqlshare#s=query/sr320%2540washington.edu/BiGo_methratio_GFF_boop&q=

```
In [13]: # ie python fetchdata.py -d "[sr320@washington.edu].[BiGO_Methylation_oysterv9_GFF]•" -f tsv -o /Volumes/web/cnidar

         # running on commandline because cannot get to work in IPython
```

```
python fetchdata.py -d "[sr320@washington.edu].[BiGo_methratio_GFF_boop]•" -f tsv -o
/Volumes/web/cnidarian/BiGo_methratio_boop.gff
```

```
In [15]: #fetchdata failed
```

Open Notebook Science

Automating a Workflow: Beyond Blast - to GO Slim

The concept is that you can take a fasta file in a working directory and end up with GO slim information all within a single notebook that is automated. Currently this work by writing (and overwriting) as scracth file to SQLShare. Assumptions are that you are working in a directory with fasta file named `query.fa`. And blast algorithms are in PATH.

```
In [13]: #allows plots to be shown inline
         %pylab inline

         Populating the interactive namespace from numpy and matplotlib

In [4]:  #Setting Working Directory
         wd="/Volumes/web/whale/fish546/qpx_go_val"
         #Setting directory of Blast Databases
         dbd="/Volumes/Bay3/Software/ncbi-blast-2.2.29\+/db/"
         #Database name
         dbn="uniprot_sprot_r2013_12"
         #Blast algorithim
         ba="blastx"
         #Location of SQLShare python tools: you can empty ("") if tools are in PATH
         spd="/Users/sr320/sqlshare-pythonclient/tools/"

In [5]:  cd {wd}

         /Volumes/web/whale/fish546/qpx_go_val

In [5]:  !{ba} -query query.fa -db {dbd}{dbn} -out {dbn}_{ba}_out.tab -evalue 1E-50 -num_threads 4 -max_hsps_per_subject 1 -

         BLAST Database error: No alias or index file found for protein database [/Volumes/Bay3/Software/ncbi-blast-2.2.29+
         /db/uniprot_sprot_r2013_12] in search path [/Volumes/web/whale/fish546/pipeline_test_dir4::]

In [6]:  !head -1 {dbn}_{ba}_out.tab

         QPX_transcriptome_v1_Contig_2   sp|P52712|CBPX_ORYSJ    43.75   416     213     12      2095    869     6       40
         7       3e-98   326

In [17]: #Translate pipes to tab so SPID is in separate column for Joining
         !tr '|' "\t" <{dbn}_{ba}_out.tab> {dbn}_{ba}_out2.tab

In [18]: !head -1 {dbn}_{ba}_out2.tab

In [8]:  #Uploads formatted blast table to SQLshare; currently has generic name and meant to be temporary: Warning will over
         !python {spd}singleupload.py -d scratchblast_out {dbn}_{ba}_out2.tab

                                                            ...

In [9]:  !python {spd}fetchdata.py -s "SELECT * FROM [sr320@washington.edu].[scratchblast_out]blast Left Join [sr320@washing

In [10]: !head -2 {dbn}_join2goslim.txt

                                                            ...

In [11]: !python {spd}singleupload.py -d scratchjoin_slim {dbn}_join2goslim.txt

         processing chunk line 0 to 18037 (0.0718240737915 s elapsed)
         pushing uniprot_sprot_r2013_12_join2goslim.txt...
         parsing 9A18D989...
         finished scratchjoin_slim

In [12]: #Sets GO aspect
         !python {spd}fetchdata.py -s "SELECT Distinct Column1 as query, Column3 as SPID, GOSlim_bin FROM [sr320@washington.

In [13]: !head justslim.txt

                                                            ...

In [15]: from pandas import *
```

Set some variables

blast

convert file format

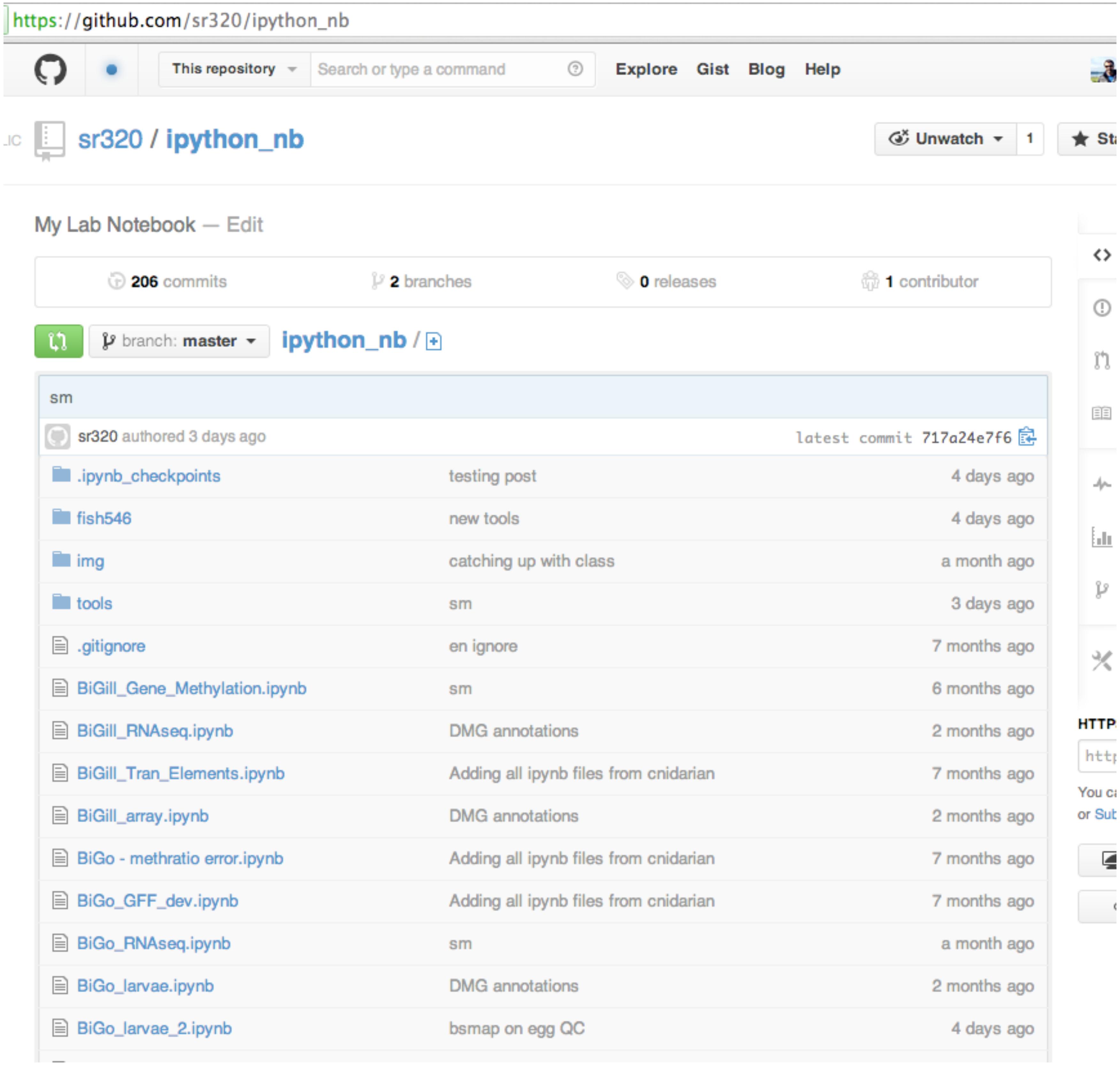
upload to SQLShare (python client)

join in SQLShare - download

read in pandas

matplotlib generates graph of GOsllim

Open Notebook Science



sr320 / ipython_nb

My Lab Notebook — Edit

206 commits · **2** branches · **0** releases · **1** contributor

ipython_nb / +

sm

sr320 authored 3 days ago — latest commit 717a24e7f6

.ipynb_checkpoints	testing post	4 days ago
fish546	new tools	4 days ago
img	catching up with class	a month ago
tools	sm	3 days ago
.gitignore	en ignore	7 months ago
BiGill_Gene_Methylation.ipynb	sm	6 months ago
BiGill_RNAseq.ipynb	DMG annotations	2 months ago
BiGill_Tran_Elements.ipynb	Adding all ipynb files from cnidarian	7 months ago
BiGill_array.ipynb	DMG annotations	2 months ago
BiGo - methratio error.ipynb	Adding all ipynb files from cnidarian	7 months ago
BiGo_GFF_dev.ipynb	Adding all ipynb files from cnidarian	7 months ago
BiGo_RNAseq.ipynb	sm	a month ago
BiGo_larvae.ipynb	DMG annotations	2 months ago
BiGo_larvae_2.ipynb	bsmap on egg QC	4 days ago

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

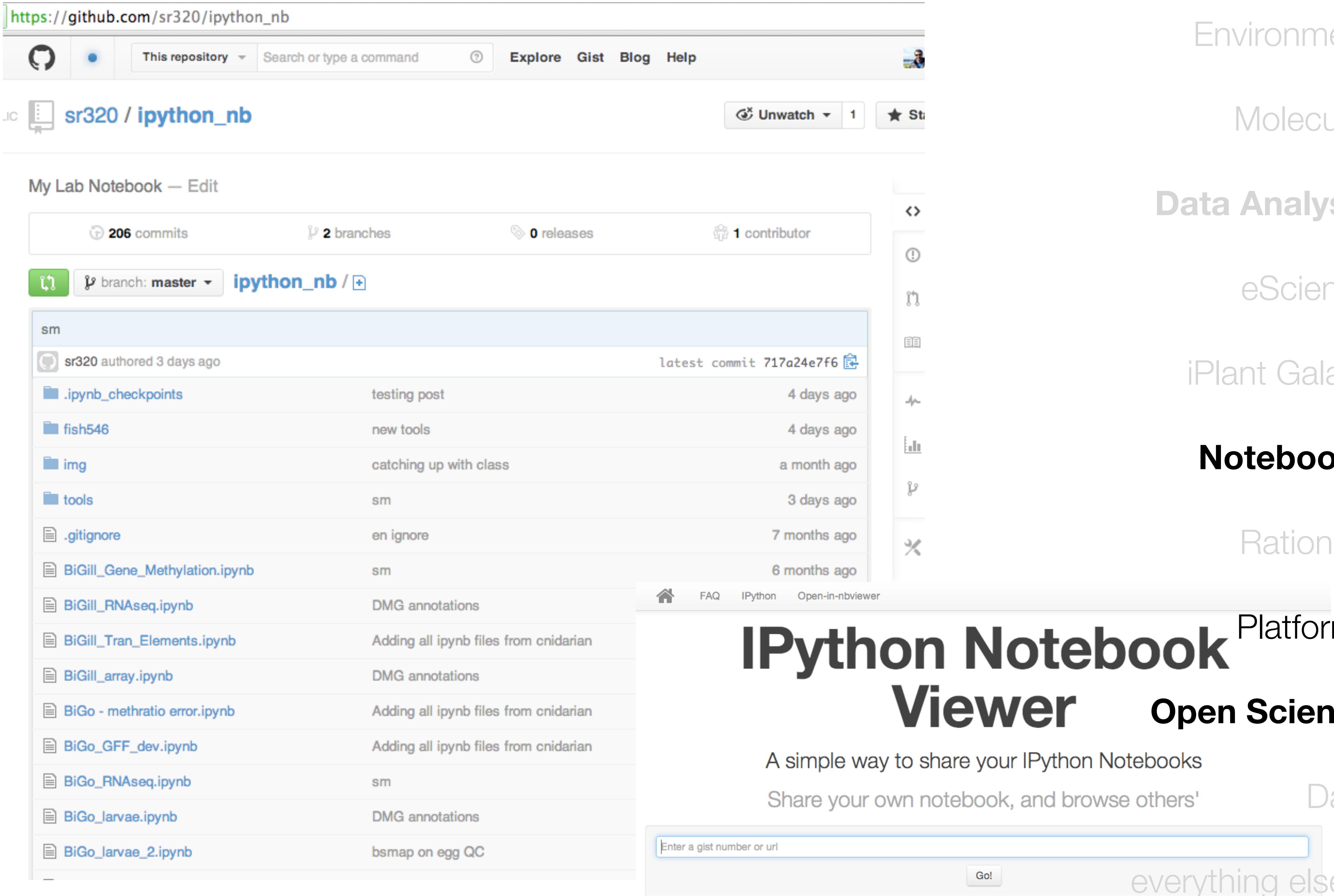
Platforms

Open Science

Data

everything else...

Open Notebook Science



https://github.com/sr320/ipython_nb

sr320 / ipython_nb

My Lab Notebook — Edit

206 commits | **2** branches | **0** releases | **1** contributor

branch: **master** ipython_nb / +

sm

sr320 authored 3 days ago — latest commit 717a24e7f6

.ipynb_checkpoints	testing post	4 days ago
fish546	new tools	4 days ago
img	catching up with class	a month ago
tools	sm	3 days ago
.gitignore	en ignore	7 months ago
BiGill_Gene_Methylation.ipynb	sm	6 months ago
BiGill_RNAseq.ipynb	DMG annotations	
BiGill_Tran_Elements.ipynb	Adding all ipynb files from cnidarian	
BiGill_array.ipynb	DMG annotations	
BiGo - methratio error.ipynb	Adding all ipynb files from cnidarian	
BiGo_GFF_dev.ipynb	Adding all ipynb files from cnidarian	
BiGo_RNAseq.ipynb	sm	
BiGo_larvae.ipynb	DMG annotations	
BiGo_larvae_2.ipynb	bsmap on egg QC	

IPython Notebook Viewer

A simple way to share your IPython Notebooks

Share your own notebook, and browse others'

Enter a gist number or url

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

everything else...

Open Notebook Science

a very new experiment

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

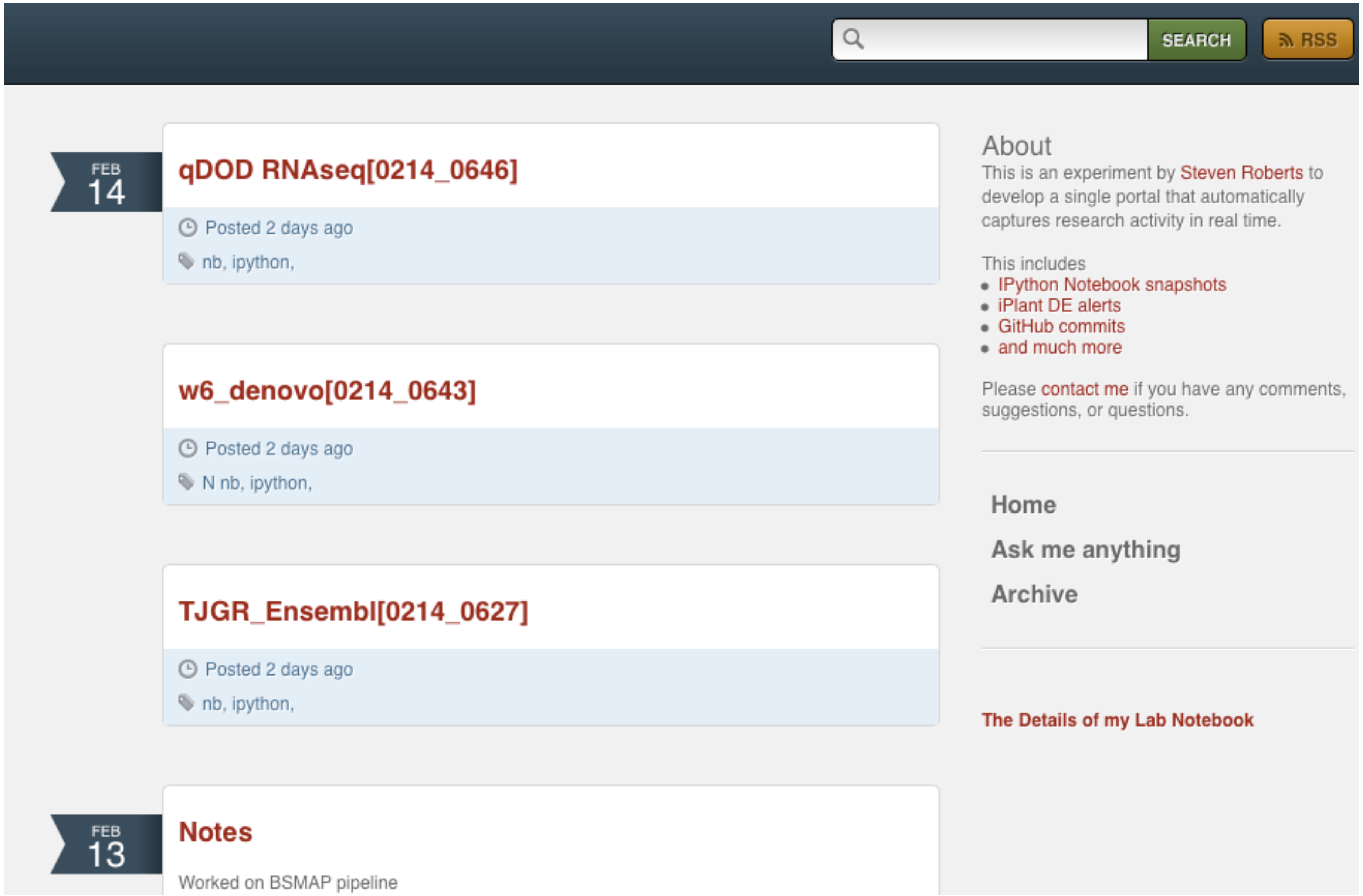
Open Science

Data

everything else...

Open Notebook Science

a very new experiment



FEB 14

qDOD RNAseq[0214_0646]

Posted 2 days ago

nb, ipython,

w6_denovo[0214_0643]

Posted 2 days ago

N nb, ipython,

TJGR_Ensembl[0214_0627]

Posted 2 days ago

nb, ipython,

FEB 13

Notes

Worked on BSMAP pipeline

About

This is an experiment by Steven Roberts to develop a single portal that automatically captures research activity in real time.

This includes
- IPython Notebook snapshots
- iPlant DE alerts
- GitHub commits
- and much more

Please contact me if you have any comments, suggestions, or questions.

Home

Ask me anything

Archive

The Details of my Lab Notebook

sr320.info

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

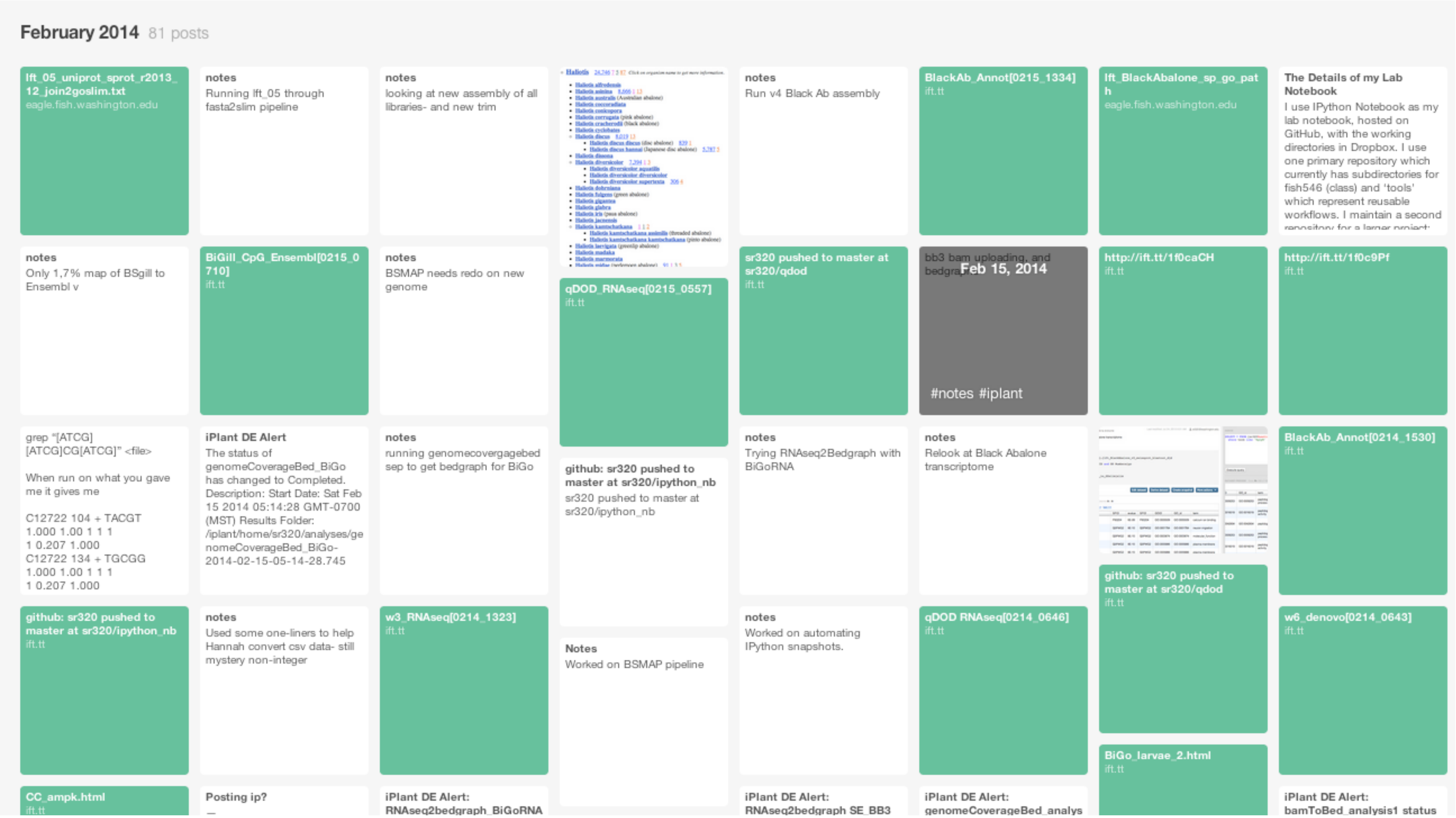
Open Science

Data

everything else...

Open Notebook Science

a very new experiment



February 2014 81 posts

lft_05_uniprot_sprot_r2013_12_join2goslim.txt
eagle.fish.washington.edu

notes
Only 1,7% map of BSgill to Ensembl v

grep "[ATCG][ATCG]CG[ATCG]" <file>

When run on what you gave me it gives me

C12722 104 + TACGT
1.000 1.00 1 1 1
1 0.207 1.000
C12722 134 + TGCGG
1.000 1.00 1 1 1
1 0.207 1.000

github: sr320 pushed to master at sr320/ipython_nb
ift.tt

CC_ampk.html
ift.tt

notes
Running lft_05 through fasta2slim pipeline

BiGill_CpG_Ensembl[0215_0710]
ift.tt

iPlant DE Alert
The status of genomeCoverageBed_BiGo has changed to Completed. Description: Start Date: Sat Feb 15 2014 05:14:28 GMT-0700 (MST) Results Folder: /iplant/home/sr320/analyses/genomeCoverageBed_BiGo-2014-02-15-05-14-28.745

notes
Used some one-liners to help Hannah convert csv data- still mystery non-integer

Posting ip?
—

notes
looking at new assembly of all libraries- and new trim

notes
BSMAP needs redo on new genome

notes
running genomecovergagebed sep to get bedgraph for BiGo

w3_RNAseq[0214_1323]
ift.tt

iPlant DE Alert:
RNAseq2bedgraph_BiGoRNA

qDOD_RNAseq[0215_0557]
ift.tt

github: sr320 pushed to master at sr320/ipython_nb
sr320 pushed to master at sr320/ipython_nb

Notes
Worked on BSMAP pipeline

notes
Run v4 Black Ab assembly

sr320 pushed to master at sr320/qdod
ift.tt

notes
Trying RNAseq2Bedgraph with BiGoRNA

notes
Worked on automating IPython snapshots.

iPlant DE Alert:
RNAseq2bedgraph SE_BB3

BlackAb_Annot[0215_1334]
ift.tt

bb3 bam uploading, and bedgra**Feb 15, 2014**

#notes #iplant

notes
Relook at Black Abalone transcriptome

qDOD RNAseq[0214_0646]
ift.tt

iPlant DE Alert:
genomeCoverageBed_analys

lft_BlackAbalone_sp_go_path
eagle.fish.washington.edu

http://ift.tt/1f0caCH
ift.tt

github: sr320 pushed to master at sr320/qdod
ift.tt

BiGo_larvae_2.html
ift.tt

The Details of my Lab Notebook
I use IPython Notebook as my lab notebook, hosted on GitHub, with the working directories in Dropbox. I use one primary repository which currently has subdirectories for fish546 (class) and 'tools' which represent reusable workflows. I maintain a second repository for a larger project:

http://ift.tt/1f0c9Pf
ift.tt

BlackAb_Annot[0214_1530]
ift.tt

w6_denovo[0214_0643]
ift.tt

iPlant DE Alert:
bamToBed_analysis1 status

sr320.info

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Platforms

Open Science

Data

everything else...

Open Science

Open Science

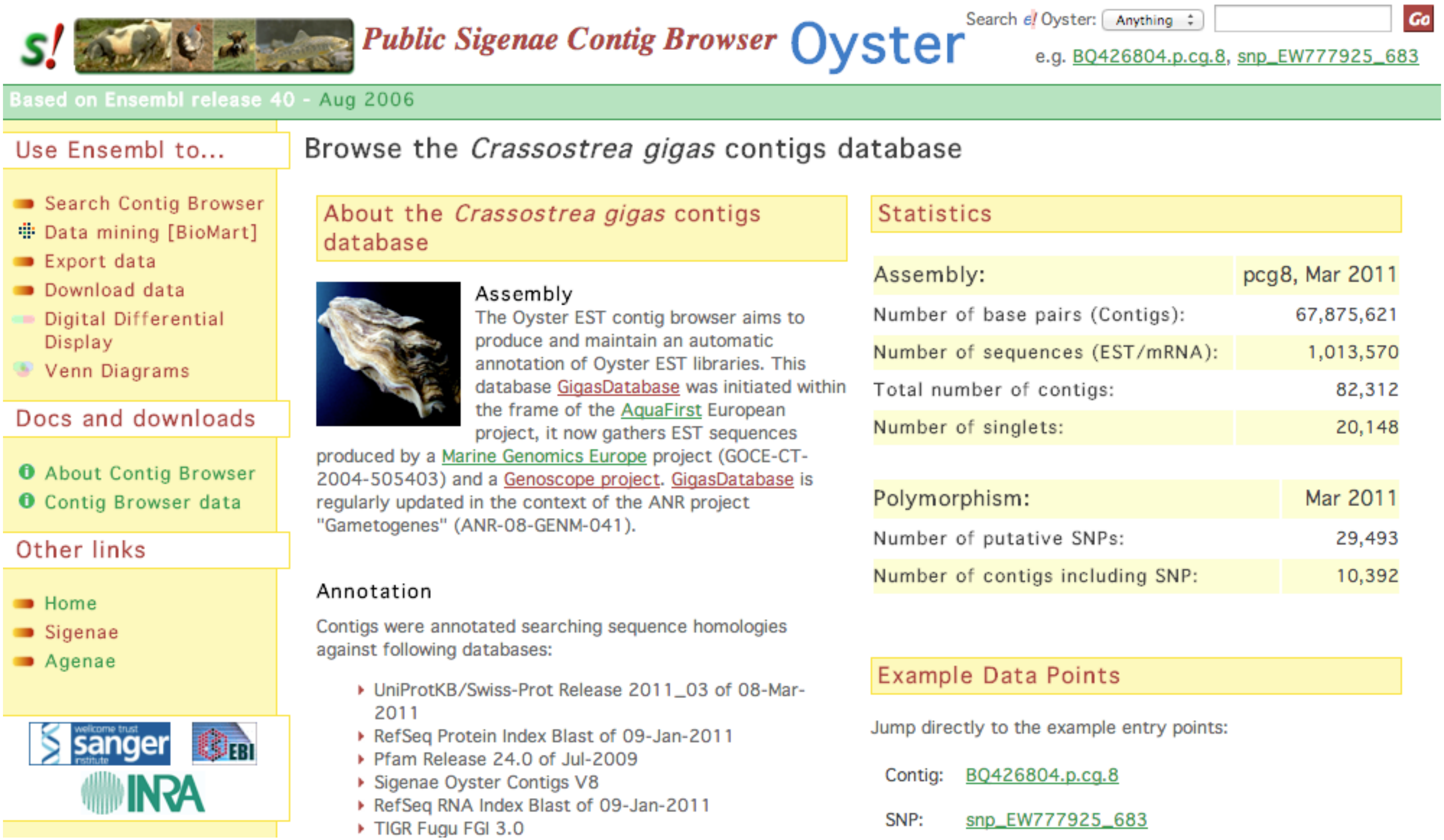
web-native scholarship

Biology
Environment
Molecular
Data Analysis
eScience
iPlant Galaxy
Notebooks
Rationale
Platforms
Open Science
Data
everything else...

Sharing

Photo credit: Flickr, Creative Commons, speechless

Example



e.g. BQ426804.p.cg.8, snp_EW777925_683

Based on Ensembl release 40 - Aug 2006

Browse the *Crassostrea gigas* contigs database

About the *Crassostrea gigas* contigs database

Assembly

The Oyster EST contig browser aims to produce and maintain an automatic annotation of Oyster EST libraries. This database GigasDatabase was initiated within the frame of the AquaFirst European project, it now gathers EST sequences produced by a Marine Genomics Europe project (GOCE-CT-2004-505403) and a Genoscope project. GigasDatabase is regularly updated in the context of the ANR project "Gametogenes" (ANR-08-GENM-041).

Annotation

Contigs were annotated searching sequence homologies against following databases:

- UniProtKB/Swiss-Prot Release 2011_03 of 08-Mar-2011
- RefSeq Protein Index Blast of 09-Jan-2011
- Pfam Release 24.0 of Jul-2009
- Sigenae Oyster Contigs V8
- RefSeq RNA Index Blast of 09-Jan-2011
- TIGR Fugu FGI 3.0

Statistics

Assembly:	pcg8, Mar 2011
Number of base pairs (Contigs):	67,875,621
Number of sequences (EST/mRNA):	1,013,570
Total number of contigs:	82,312
Number of singlets:	20,148

Polymorphism:	Mar 2011
Number of putative SNPs:	29,493
Number of contigs including SNP:	10,392

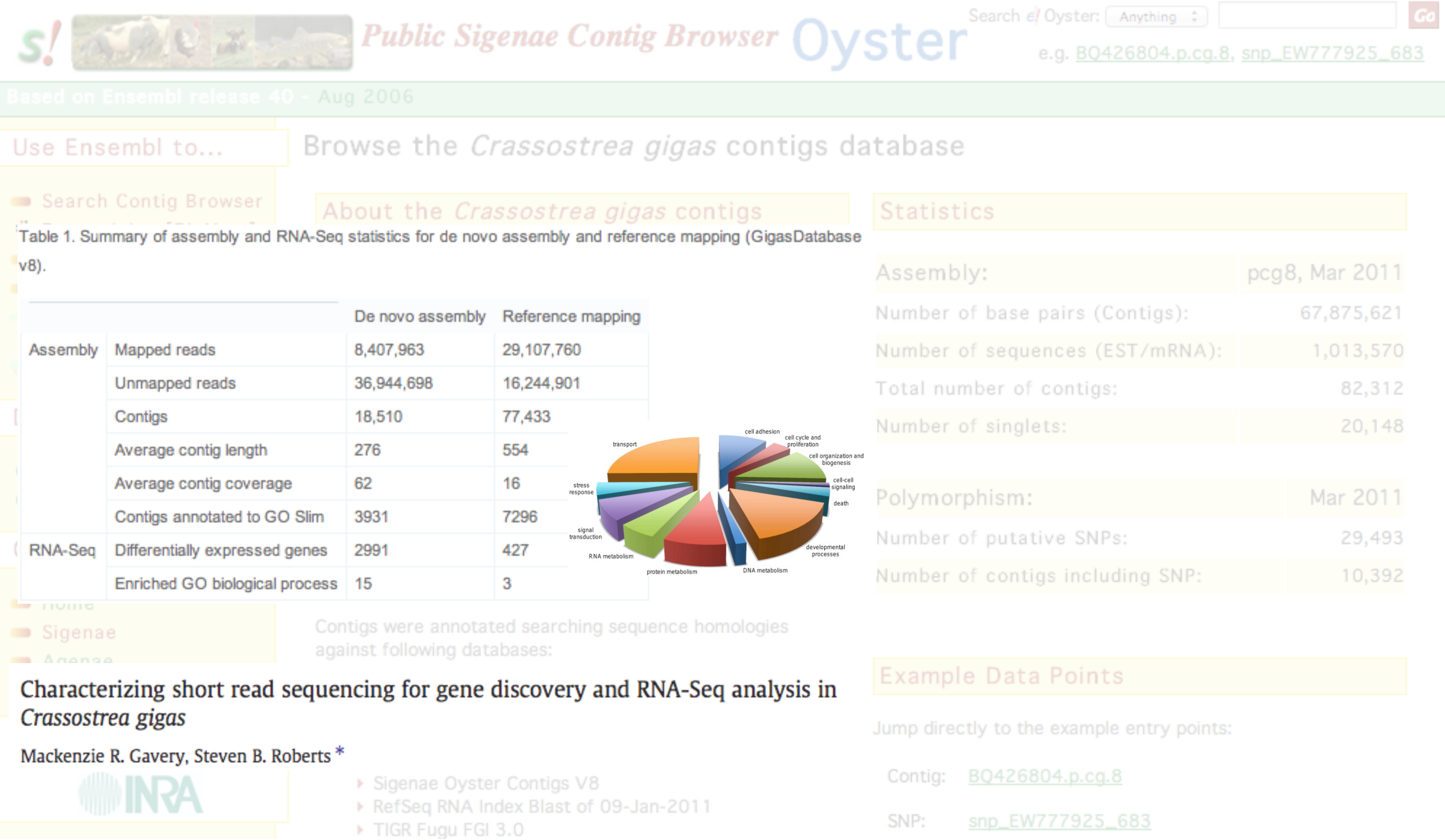
Example Data Points

Jump directly to the example entry points:

Contig: BQ426804.p.cg.8

SNP: snp_EW777925_683

Example



Public Sigenae Contig Browser Oyster

e.g. BQ426804.p.cg.8, snp_EW777925_683

Based on Ensembl release 40 - Aug 2006

Use Ensembl to...

- Search Contig Browser

Browse the *Crassostrea gigas* contigs database

About the *Crassostrea gigas* contigs

Table 1. Summary of assembly and RNA-Seq statistics for de novo assembly and reference mapping (GigasDatabase v8).

		De novo assembly	Reference mapping
Assembly	Mapped reads	8,407,963	29,107,760
	Unmapped reads	36,944,698	16,244,901
	Contigs	18,510	77,433
	Average contig length	276	554
	Average contig coverage	62	16
	Contigs annotated to GO Slim	3931	7296
RNA-Seq	Differentially expressed genes	2991	427
	Enriched GO biological process	15	3

Statistics

Assembly:	pcg8, Mar 2011
Number of base pairs (Contigs):	67,875,621
Number of sequences (EST/mRNA):	1,013,570
Total number of contigs:	82,312
Number of singlets:	20,148

Polymorphism:	Mar 2011
Number of putative SNPs:	29,493
Number of contigs including SNP:	10,392

- Home
- Sigenae
- Agenae

Contigs were annotated searching sequence homologies against following databases:

- Sigenae Oyster Contigs V8
- RefSeq RNA Index Blast of 09-Jan-2011
- TIGR Fugu FGI 3.0

Example Data Points

Jump directly to the example entry points:

Contig: BQ426804.p.cg.8

SNP: snp_EW777925_683

Characterizing short read sequencing for gene discovery and RNA-Seq analysis in *Crassostrea gigas*

Mackenzie R. Gavery, Steven B. Roberts *

Example

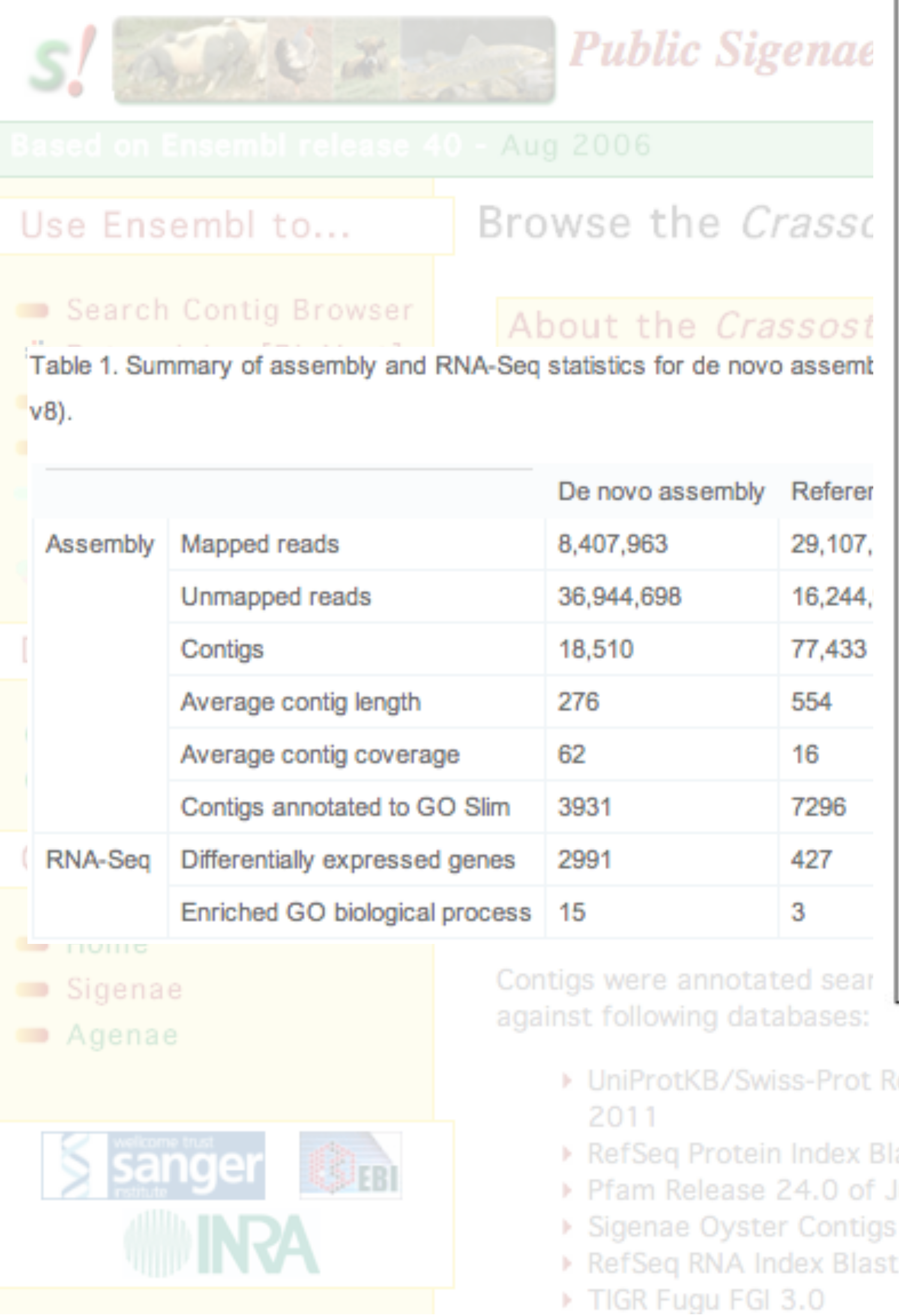
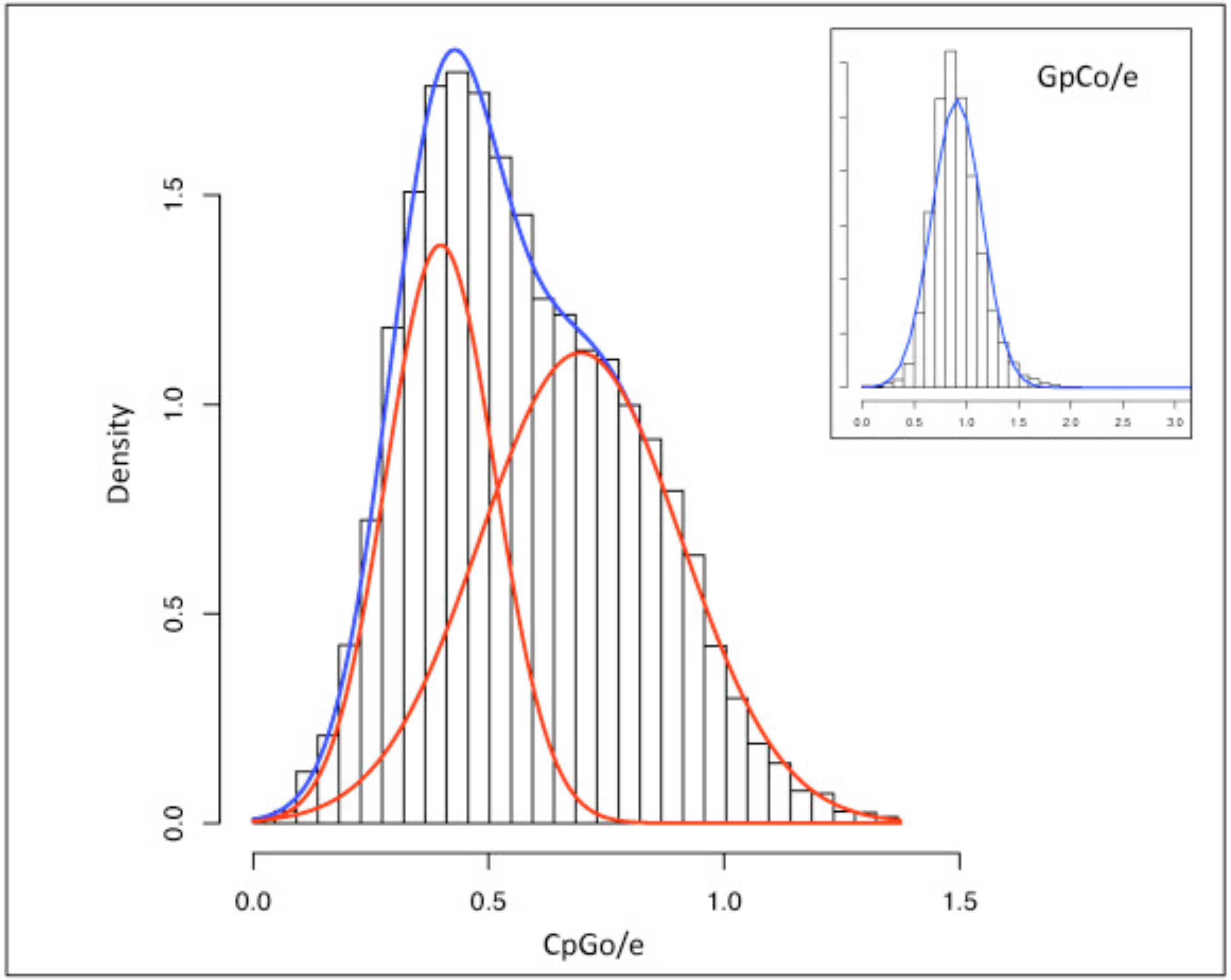


Table 1. Summary of assembly and RNA-Seq statistics for de novo assemb v8).

		De novo assembly	Referer
Assembly	Mapped reads	8,407,963	29,107,
	Unmapped reads	36,944,698	16,244,
	Contigs	18,510	77,433
	Average contig length	276	554
	Average contig coverage	62	16
	Contigs annotated to GO Slim	3931	7296
RNA-Seq	Differentially expressed genes	2991	427
	Enriched GO biological process	15	3

Contigs were annotated sear against following databases:

- UniProtKB/Swiss-Prot Re 2011
- RefSeq Protein Index Bla
- Pfam Release 24.0 of Ju
- Sigenae Oyster Contigs
- RefSeq RNA Index Blast
- TIGR Fugu FGI 3.0



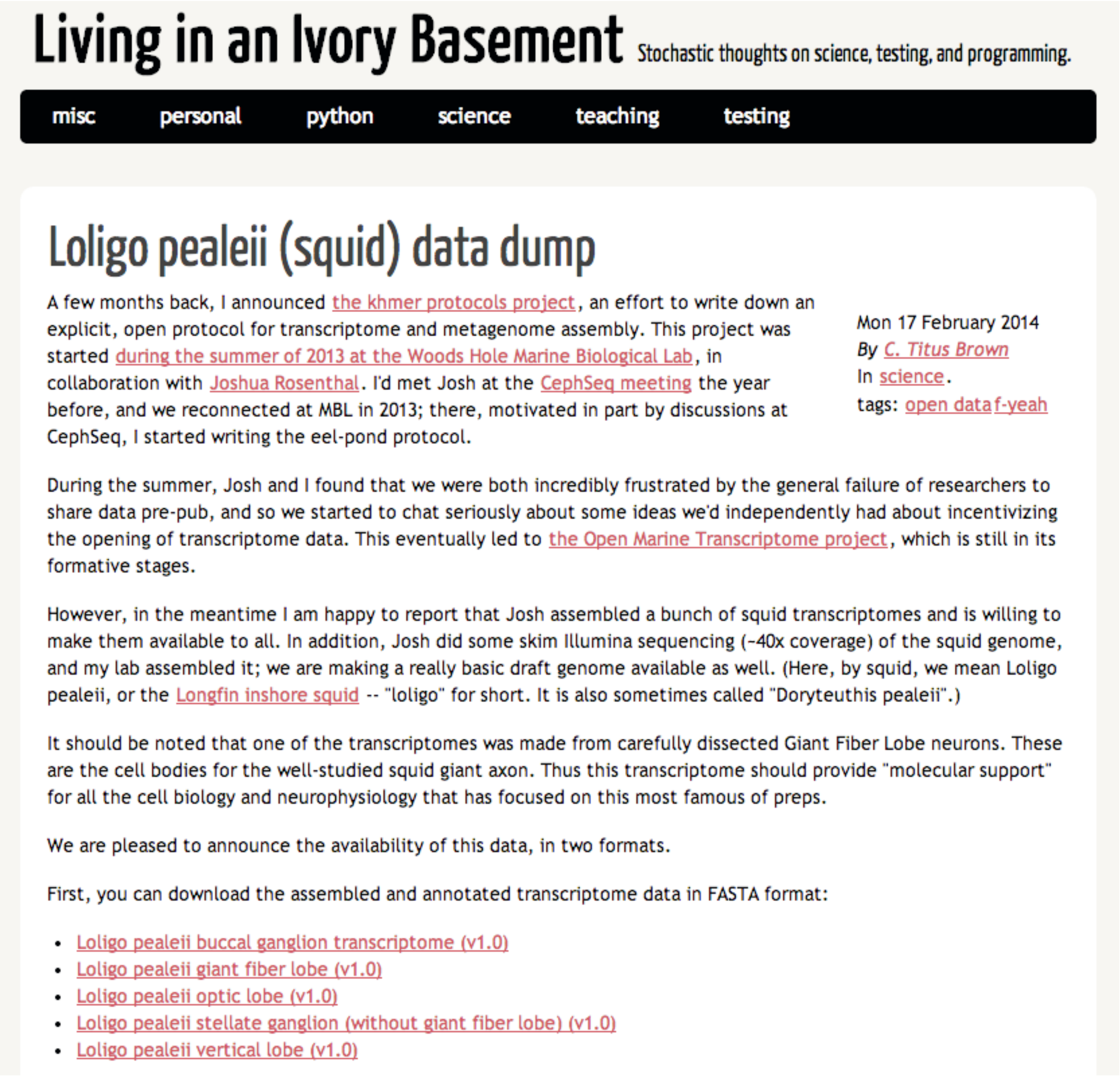
Research article

Highly accessed

Open Access

DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*)

Mackenzie R Gavery and **Steven B Roberts***



Living in an Ivory Basement

Stochastic thoughts on science, testing, and programming.

Loligo pealeii (squid) data dump

Mon 17 February 2014
By C. Titus Brown
In science.
tags: open data f-yeah

A few months back, I announced the khmer protocols project, an effort to write down an explicit, open protocol for transcriptome and metagenome assembly. This project was started during the summer of 2013 at the Woods Hole Marine Biological Lab, in collaboration with Joshua Rosenthal. I'd met Josh at the CephSeq meeting the year before, and we reconnected at MBL in 2013; there, motivated in part by discussions at CephSeq, I started writing the eel-pond protocol.

During the summer, Josh and I found that we were both incredibly frustrated by the general failure of researchers to share data pre-pub, and so we started to chat seriously about some ideas we'd independently had about incentivizing the opening of transcriptome data. This eventually led to the Open Marine Transcriptome project, which is still in its formative stages.

However, in the meantime I am happy to report that Josh assembled a bunch of squid transcriptomes and is willing to make them available to all. In addition, Josh did some skim Illumina sequencing (~40x coverage) of the squid genome, and my lab assembled it; we are making a really basic draft genome available as well. (Here, by squid, we mean Loligo pealeii, or the Longfin inshore squid -- "loligo" for short. It is also sometimes called "Doryteuthis pealeii".)

It should be noted that one of the transcriptomes was made from carefully dissected Giant Fiber Lobe neurons. These are the cell bodies for the well-studied squid giant axon. Thus this transcriptome should provide "molecular support" for all the cell biology and neurophysiology that has focused on this most famous of preps.

We are pleased to announce the availability of this data, in two formats.

First, you can download the assembled and annotated transcriptome data in FASTA format:

- Loligo pealeii buccal ganglion transcriptome (v1.0)
- Loligo pealeii giant fiber lobe (v1.0)
- Loligo pealeii optic lobe (v1.0)
- Loligo pealeii stellate ganglion (without giant fiber lobe) (v1.0)
- Loligo pealeii vertical lobe (v1.0)

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

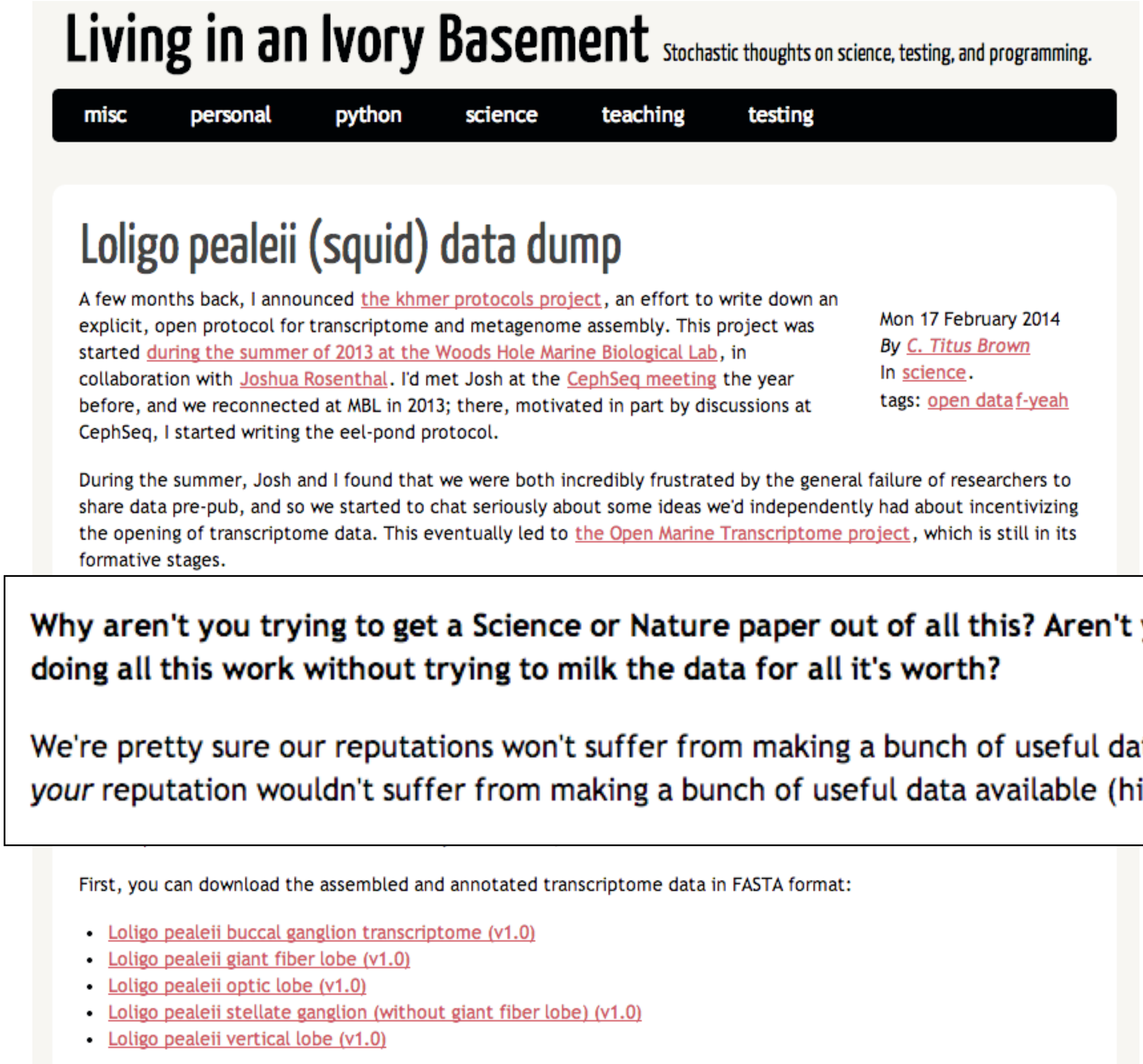
Platforms

Open Science

Data

everything else...

http://ivory.idyll.org/blog/



Living in an Ivory Basement

Stochastic thoughts on science, testing, and programming.

misc | personal | python | science | teaching | testing

Loligo pealeii (squid) data dump

A few months back, I announced the khmer protocols project, an effort to write down an explicit, open protocol for transcriptome and metagenome assembly. This project was started during the summer of 2013 at the Woods Hole Marine Biological Lab, in collaboration with Joshua Rosenthal. I'd met Josh at the CephSeq meeting the year before, and we reconnected at MBL in 2013; there, motivated in part by discussions at CephSeq, I started writing the eel-pond protocol.

Mon 17 February 2014
By *C. Titus Brown*
In science.
tags: open data f-yeah

During the summer, Josh and I found that we were both incredibly frustrated by the general failure of researchers to share data pre-pub, and so we started to chat seriously about some ideas we'd independently had about incentivizing the opening of transcriptome data. This eventually led to the Open Marine Transcriptome project, which is still in its formative stages.

Why aren't you trying to get a Science or Nature paper out of all this? Aren't your reputations going to suffer for doing all this work without trying to milk the data for all it's worth?

We're pretty sure our reputations won't suffer from making a bunch of useful data available. Heck, we're pretty sure *your* reputation wouldn't suffer from making a bunch of useful data available (hint).

First, you can download the assembled and annotated transcriptome data in FASTA format:

- Loligo pealeii buccal ganglion transcriptome (v1.0)
- Loligo pealeii giant fiber lobe (v1.0)
- Loligo pealeii optic lobe (v1.0)
- Loligo pealeii stellate ganglion (without giant fiber lobe) (v1.0)
- Loligo pealeii vertical lobe (v1.0)

http://ivory.idyll.org/blog/

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Open Science

Data

everything else...

OYSTERGEN.ES

Olympia

Ostrea lurida

The Olympia oyster (*Ostrea lurida*) is the only native oyster on the west coast of the US.

Alleviating Regulatory Impediments To Native Shellfish Aquaculture

Our approach is to simultaneously address local adaptation in three genetically differentiated populations of Olympia oysters by evaluating genotype-by-environment interactions. We will reciprocally transplant seed produced from wild parents collected from contrasting environments into all environments. This very large reciprocal transplant experiment can test for a home field advantage in survival, maturation and growth in Olympia oysters. The overall goals of this project are to increase our knowledge of local adaptation in Olympia oysters to address concerns that interbreeding between potentially maladapted cultured and wild stocks could negatively impact wild populations. Accordingly, in order to attain these goals, the specific objectives of this proposal are to 1) Evaluate fitness components and performance of seed from different origins in a reciprocal transplant experiment and 2) Characterize genetic and epigenetic

Check out our wiki for up to date data and more details.

Genomic Resources

Transcriptomes

O.lurida transcriptome version3 *[10/13]*

O.lurida transcriptome version2

O.lurida transcriptome version1

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

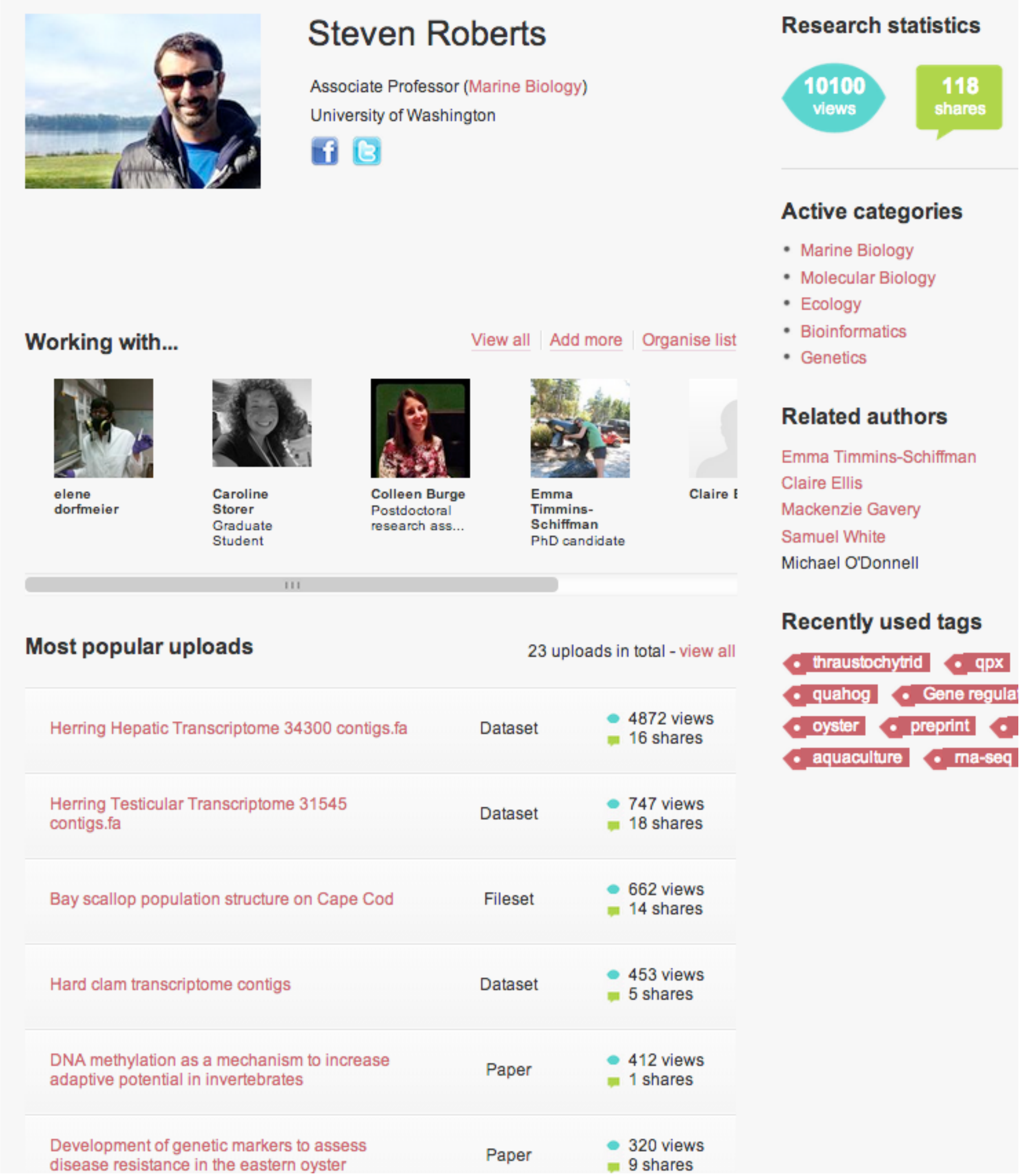
Rationale

Platforms

Open Science

Data

everything else...



Steven Roberts

Associate Professor (Marine Biology)
University of Washington

Research statistics

10100 views
118 shares

Active categories

- Marine Biology
- Molecular Biology
- Ecology
- Bioinformatics
- Genetics

Working with...

View all | Add more | Organise list

elene dorfmeier

Caroline Storer
Graduate Student

Colleen Burge
Postdoctoral research ass...

Emma Timmins-Schiffman
PhD candidate

Claire E

Related authors

Emma Timmins-Schiffman
Claire Ellis
Mackenzie Gavery
Samuel White
Michael O'Donnell

Recently used tags

thraustochytrid
qpx
quahog
Gene regula
oyster
preprint
aquaculture
rna-seq

Most popular uploads

23 uploads in total - view all

Herring Hepatic Transcriptome 34300 contigs.fa	Dataset	4872 views 16 shares
Herring Testicular Transcriptome 31545 contigs.fa	Dataset	747 views 18 shares
Bay scallop population structure on Cape Cod	Fileset	662 views 14 shares
Hard clam transcriptome contigs	Dataset	453 views 5 shares
DNA methylation as a mechanism to increase adaptive potential in invertebrates	Paper	412 views 1 shares
Development of genetic markers to assess disease resistance in the eastern oyster	Paper	320 views 9 shares

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

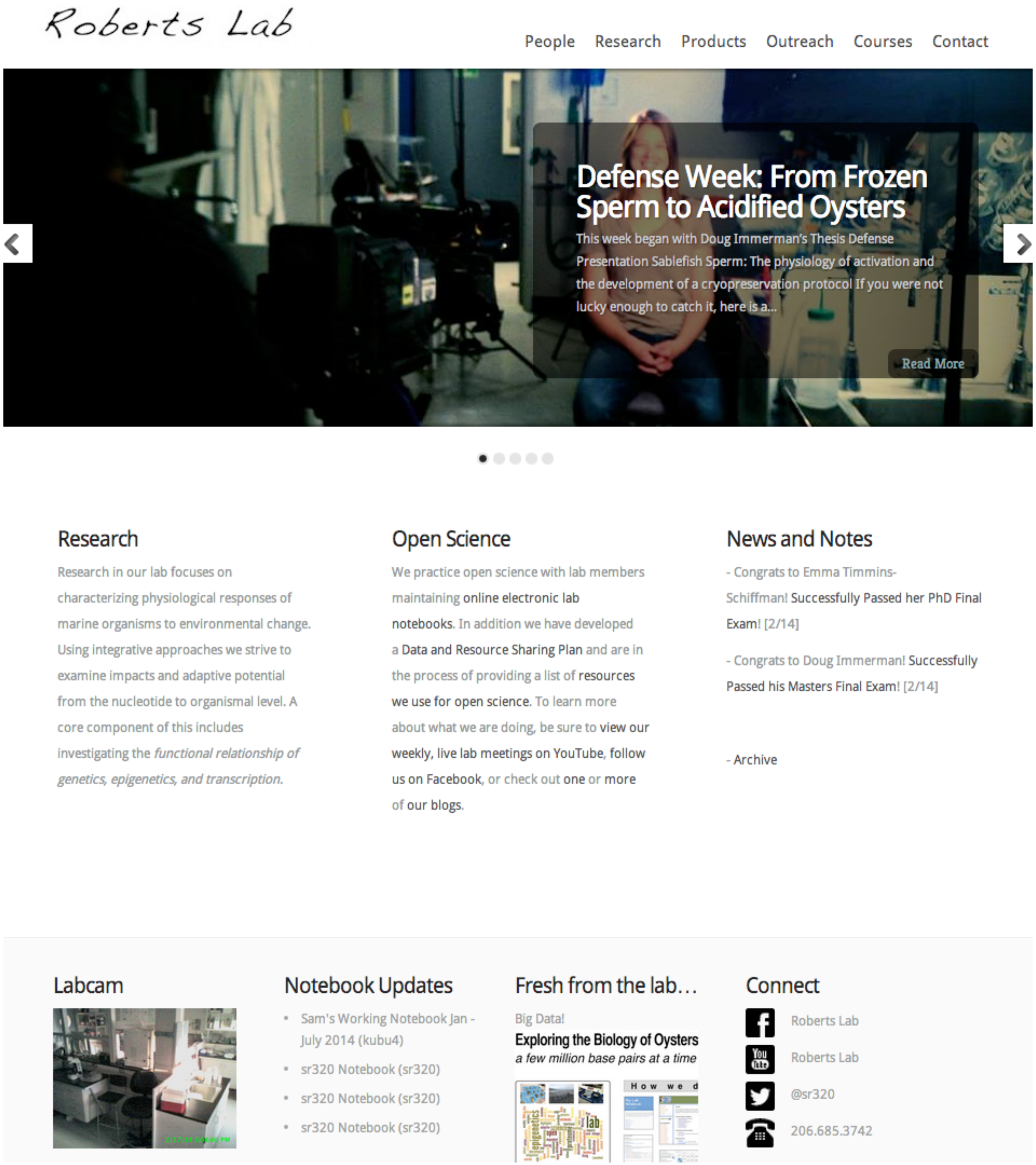
Rationale

Platforms

Open Science

Data

everything else...


Roberts Lab

Defense Week: From Frozen Sperm to Acidified Oysters

This week began with Doug Immerman's Thesis Defense Presentation Sablefish Sperm: The physiology of activation and the development of a cryopreservation protocol If you were not lucky enough to catch it, here is a...

Research

Research in our lab focuses on characterizing physiological responses of marine organisms to environmental change. Using integrative approaches we strive to examine impacts and adaptive potential from the nucleotide to organismal level. A core component of this includes investigating the *functional relationship of genetics, epigenetics, and transcription.*

Open Science

We practice open science with lab members maintaining online electronic lab notebooks. In addition we have developed a Data and Resource Sharing Plan and are in the process of providing a list of resources we use for open science. To learn more about what we are doing, be sure to view our weekly, live lab meetings on YouTube, follow us on Facebook, or check out one or more of our blogs.

News and Notes

- Congrats to Emma Timmins-Schiffman! Successfully Passed her PhD Final Exam! [2/14]

- Congrats to Doug Immerman! Successfully Passed his Masters Final Exam! [2/14]

- Archive

Labcam

Notebook Updates

- Sam's Working Notebook Jan - July 2014 (kubu4)
- sr320 Notebook (sr320)
- sr320 Notebook (sr320)
- sr320 Notebook (sr320)

Fresh from the lab…

Big Data!

Exploring the Biology of Oysters
a few million base pairs at a time

Connect

Roberts Lab
Roberts Lab
@sr320
206.685.3742

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

everything else...

robertslab.info

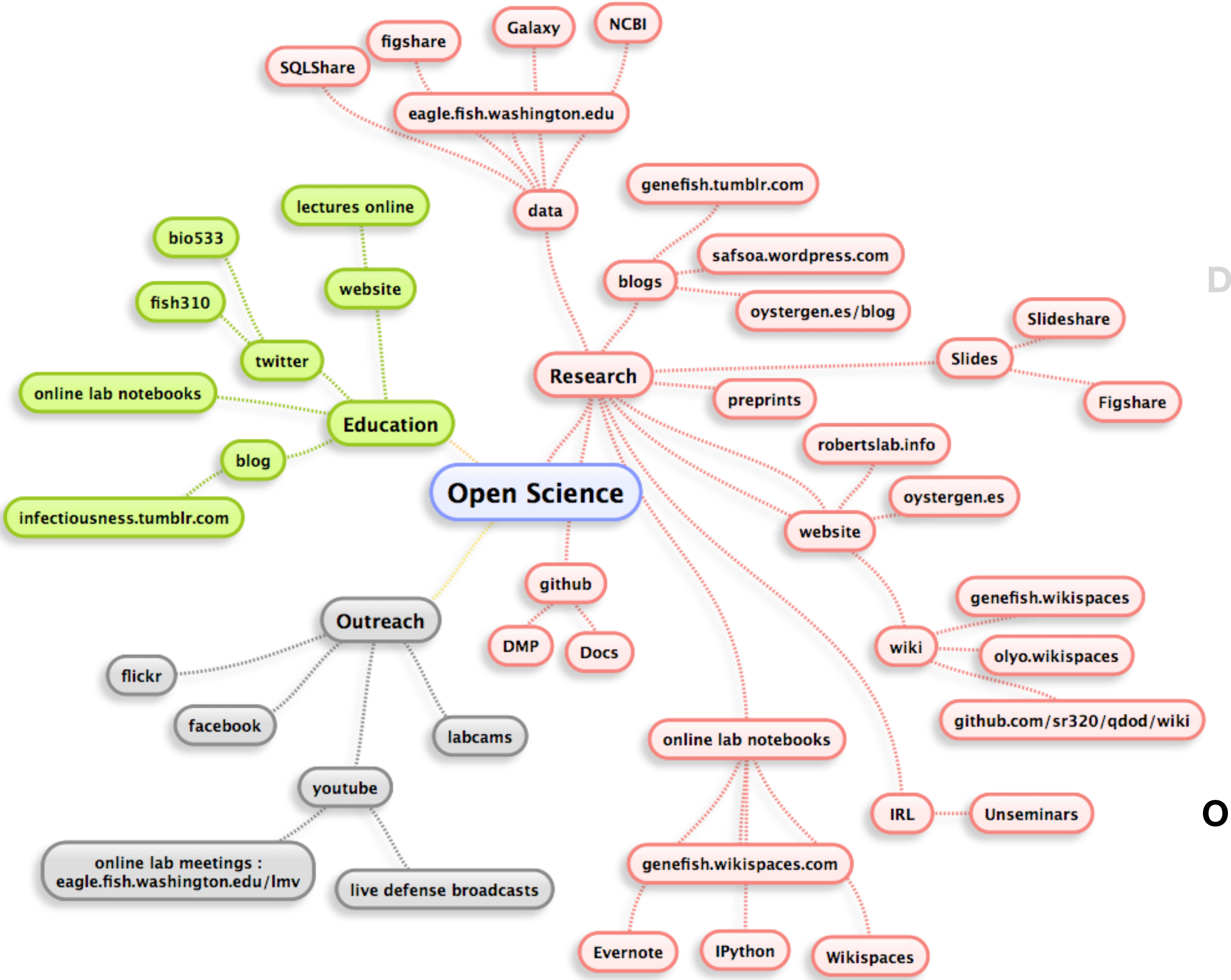
Open Science Philosophy

Transparency with limited effort

Open Science Philosophy

Transparency with limited effort

will try just about anything



- Open Science
  - Research
    - data
      - eagle.fish.washington.edu
        - figshare
        - Galaxy
        - NCBI
      - SQLShare
    - blogs
      - genefish.tumblr.com
      - safsoa.wordpress.com
      - oystergen.es/blog
    - Slides
      - Slideshare
      - Figshare
    - preprints
    - website
      - robertslab.info
      - oystergen.es
      - wiki
        - genefish.wikispaces
        - olyo.wikispaces
        - github.com/sr320/qdod/wiki
    - online lab notebooks
      - genefish.wikispaces.com
      - Evernote
      - IPython
      - Wikispaces
    - IRL
      - Unseminars
  - github
    - DMP
    - Docs
  - Education
    - website
      - lectures online
    - twitter
      - bio533
      - fish310
    - online lab notebooks
    - blog
      - infectiousness.tumblr.com
  - Outreach
    - flickr
    - facebook
    - labcams
    - youtube
      - online lab meetings : eagle.fish.washington.edu/lmv
      - live defense broadcasts

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

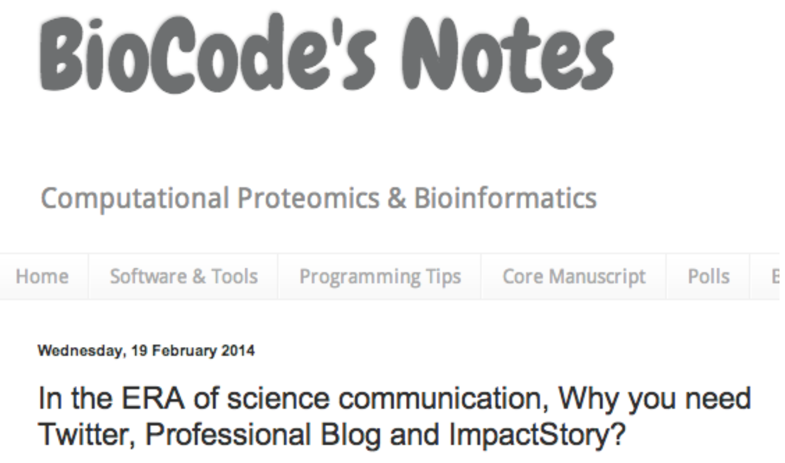
Rationale

Platforms

Open Science

Data

everything else...



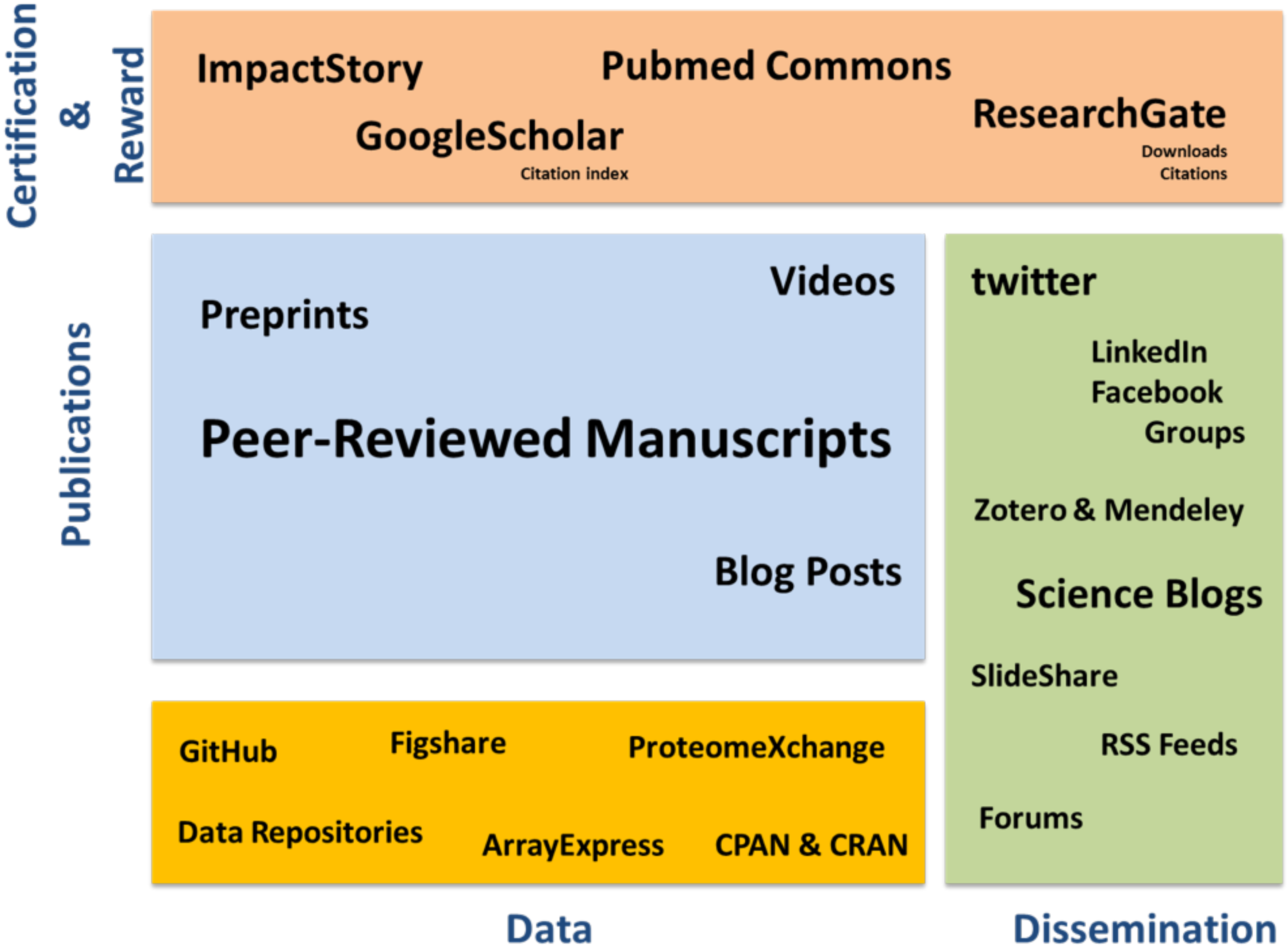
BioCode's Notes

Computational Proteomics & Bioinformatics

Wednesday, 19 February 2014

In the ERA of science communication, Why you need Twitter, Professional Blog and ImpactStory?

Yasset Perez-Riverol en Wednesday, February 19, 2014



Certification & Reward

ImpactStory

Pubmed Commons

GoogleScholar
Citation index

ResearchGate
Downloads
Citations

Publications

Preprints

Videos

Peer-Reviewed Manuscripts

Blog Posts

Data

GitHub

Figshare

ProteomeXchange

Data Repositories

ArrayExpress

CPAN & CRAN

Dissemination

twitter

LinkedIn
Facebook
Groups

Zotero & Mendeley

Science Blogs

SlideShare

RSS Feeds

Forums

computationalproteomic.blogspot.com

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

everything else...

Start them early

Acknowledgements

Emma Timmins-Schiffman

acidification

Saltonstall-Kennedy

Mackenzie Gavery
Claire Olson

DNA methylation

EPA
STAR

Sam White
Brent Vadopalas
Jake Heare

Aquaculture Program

eScience Institute

USDA

Bill Howe
Dan Halperin

oystergen.es/data