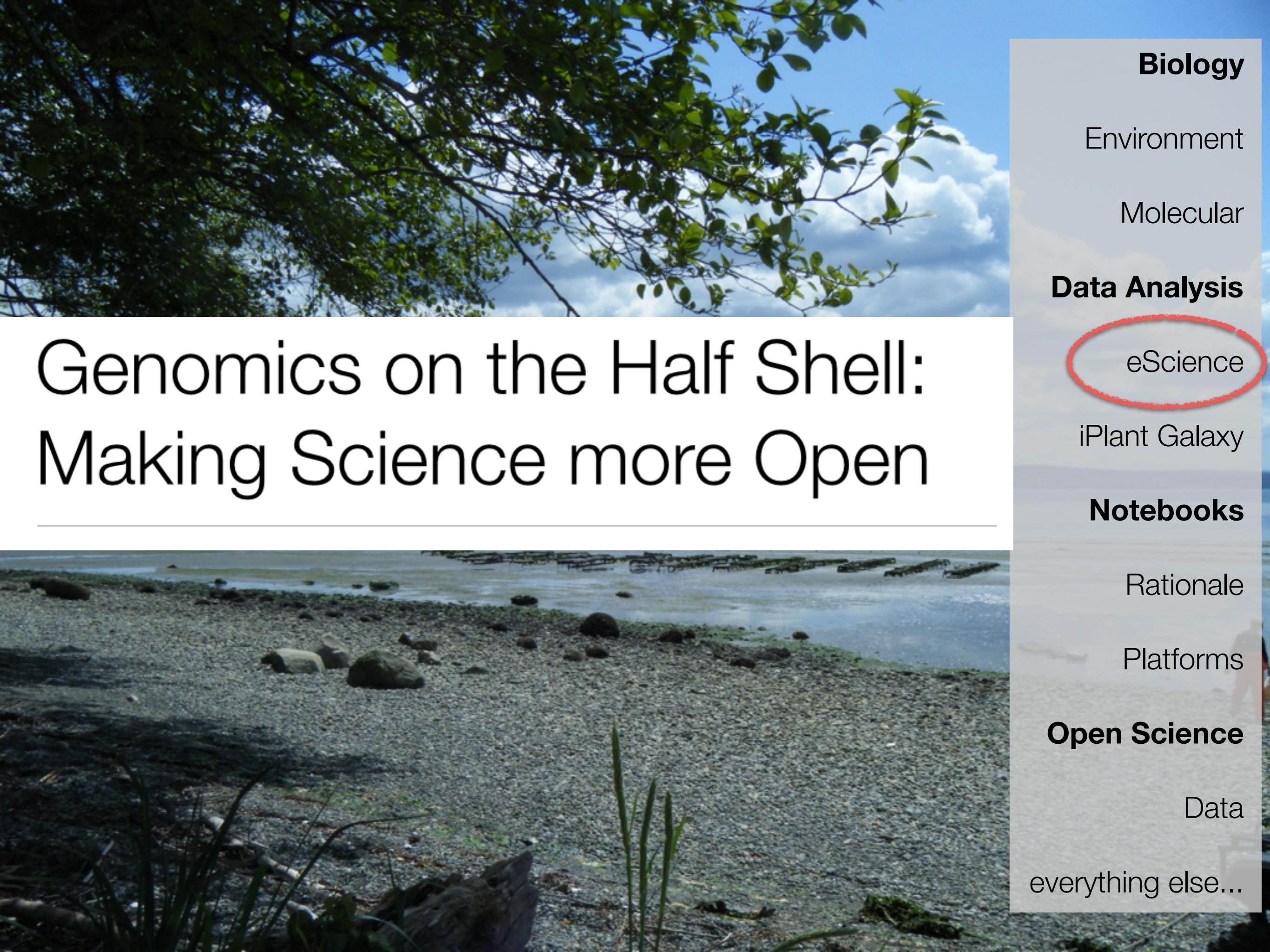


Integrating SQLShare into Genomic Workflows

Steven Roberts
Associate Professor
School of Aquatic and Fishery Sciences
sr320@uw.edu
[@sr320](https://twitter.com/sr320)

A photograph of a rocky beach. In the background, a line of oyster shells is visible in the shallow water. The foreground shows a mix of dark pebbles and some green plants growing out of the rocks.

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

everything else...

Genomics on the Half Shell: Making Science more Open

Open Science

- You are free to Share!
- Our lab practices open notebook science



<p>Steven Roberts into IPython notebook rss</p> <p>http://sr320.tumblr.com/</p> <ul style="list-style-type: none">_nb(nbyviewer) Yesterday_nb(nbyviewer) Fridayeimd_ss_manuscript(nbyviewer) Oct 8, 2014sr320 pushed to master at sr320/eimd Oct 8, 2014eimd_ss_manuscript(nbyviewer) Oct 3, 2014sr320 pushed to master at sr320/eimd Oct 3, 2014HIF_clup(nbyviewer) Oct 3, 2014sr320 pushed to master at sr320/eimd Oct 1, 2014sr320 pushed to master at sr320/a...ss Oct 1, 2014_DGU_blast(nbyviewer) Oct 2, 2014	<p>Sam White keeping it real notebook rss</p> <p>20140926</p> <p>Package - Received Package from Jerome LaPeyre from LSU</p> <p>Oyster (<i>C.virginica</i>) gill samples exposed to "no oil" and "highest level of oil." Samples were stored in Rack #2 in the -80C. Images of the box label and included sample below.</p>	<p>Claire Olson larval methylation notebook rss</p> <p>Claire's lab notebook</p> <ul style="list-style-type: none">Stacked bar plot code- DMLs vs all CpG Oct 7, 2014```DMLs````head(DMLs)````par(mar=...Stats for DMLs vs all CGs Oct 7, 2014#DMLs versus all CGs Oct 7, 2014#DMLs versus all CGs Oct 1, 2014Stacked bar graphs- DMLs vs all CGs Oct 1, 2014File location:&nbsp; http://eagle.fish...Similarities between two males Sep 19, 2014Similar:SELECT &nbsp;*&nbsp;(M1ratio...Enrichment analysis for DML genes Aug 21, 2014```library(dplyr) &lt;- dplyr::library(...Lineage_sp and developmentally different barplot revisited Jul 28, 2014library(plotly) &lt;- plotly(username...
<p>Jake Heare Oly oysters notebook rss</p> <p>OLYMPIA</p> <p>Jake Heare Research Central</p> <ul style="list-style-type: none">10.3.2014 How to Convert and Plot Hobo logger Temp Data Oct 3, 20149.19.2014 oyster bay fall maintenance Sep 19, 2014Kamliche waHigh 60s low 70s cloudy to...9.14.2014 Kaplan Meier Graphs Improved Sep 14, 20149.12.2014 Brooder Numbers Graphs Sep 12, 20149.9.2014 Distribution Overlay Graphs in R for Size Data Sep 9, 20149.8.2014 Kaplan Meier Survival Graphs Sep 8, 2014Seattle WAMir 60s to low 70s Mornly	<p>Katie Jackson satellites and sleep notebook rss</p> <p>September 8, 2014</p> <p>I have finally gotten some of my rarefaction data completed. Interestingly, the only significant differences are between Wild vs. Commercial and 2010 vs. 2011 restoration. Commercial and 2010 both having</p>	<p>Jay Dimond Corals notebook rss</p>

<http://genefish.wikispaces.com>

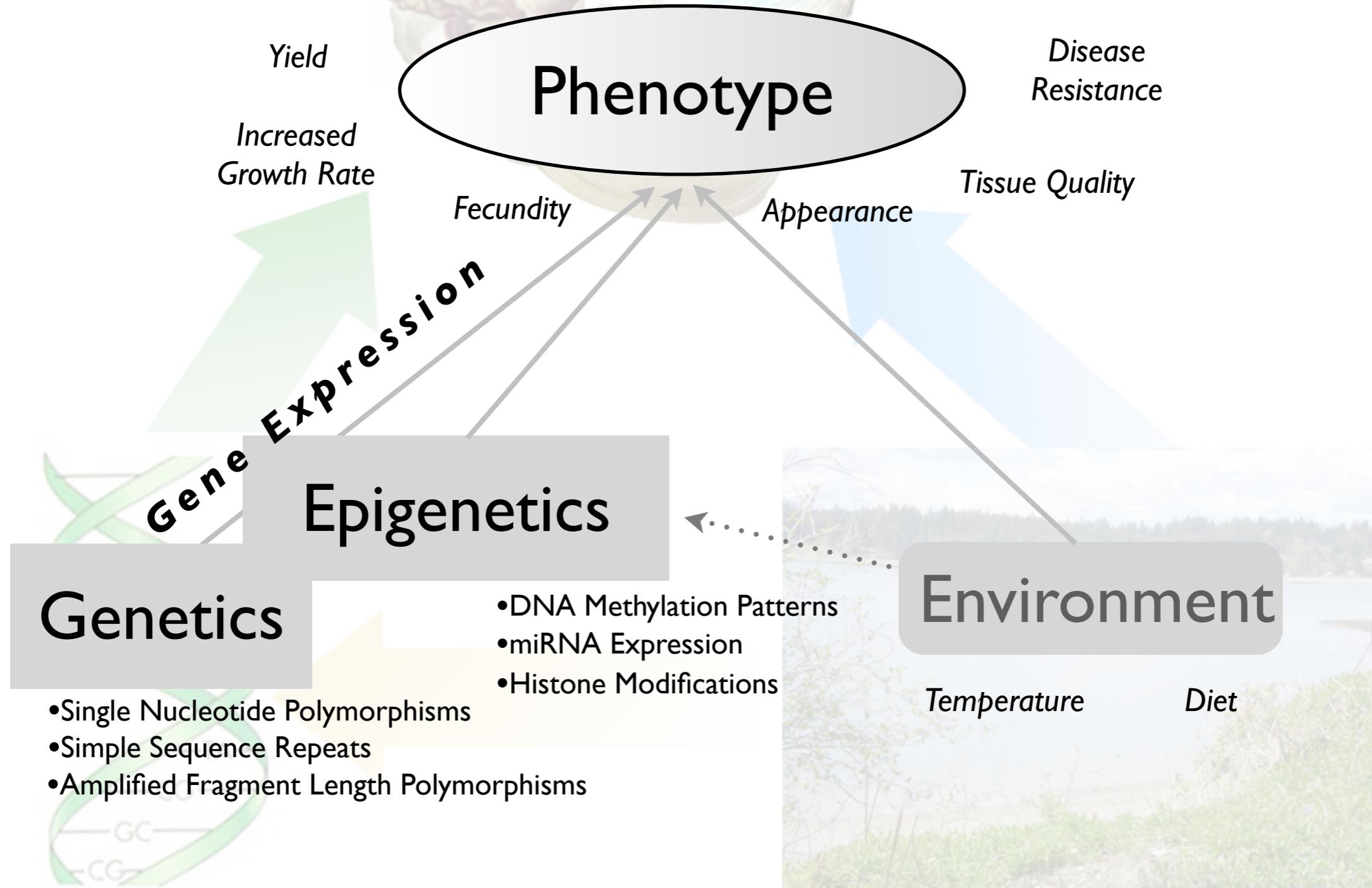
repo for escience seminar - January 6 2015 — Edit

 8 commits 1 branch 0 releases 1 contributor branch: master ▾**escience-talk-sqlshare-2015** / +

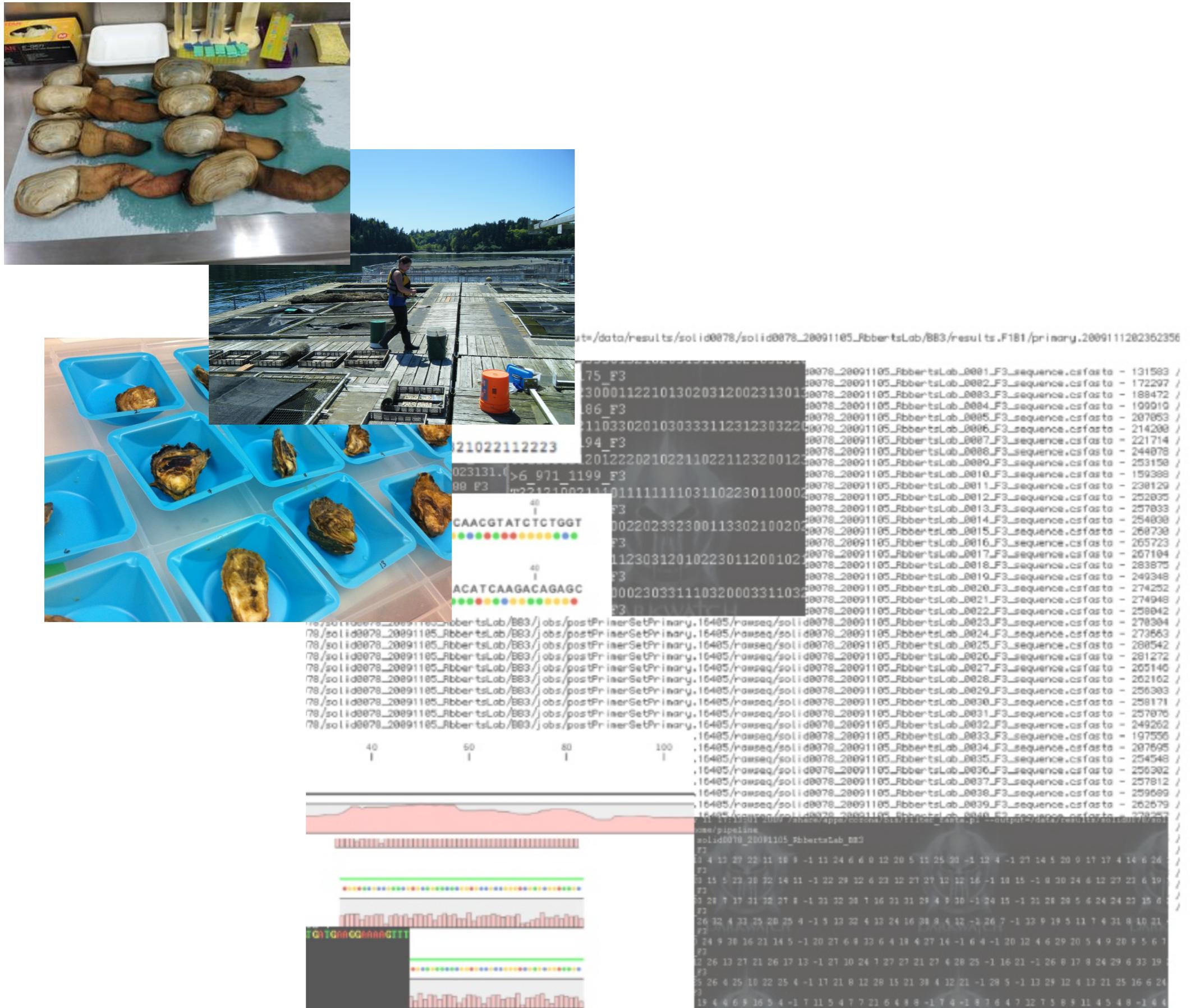
deleted html - added pdf

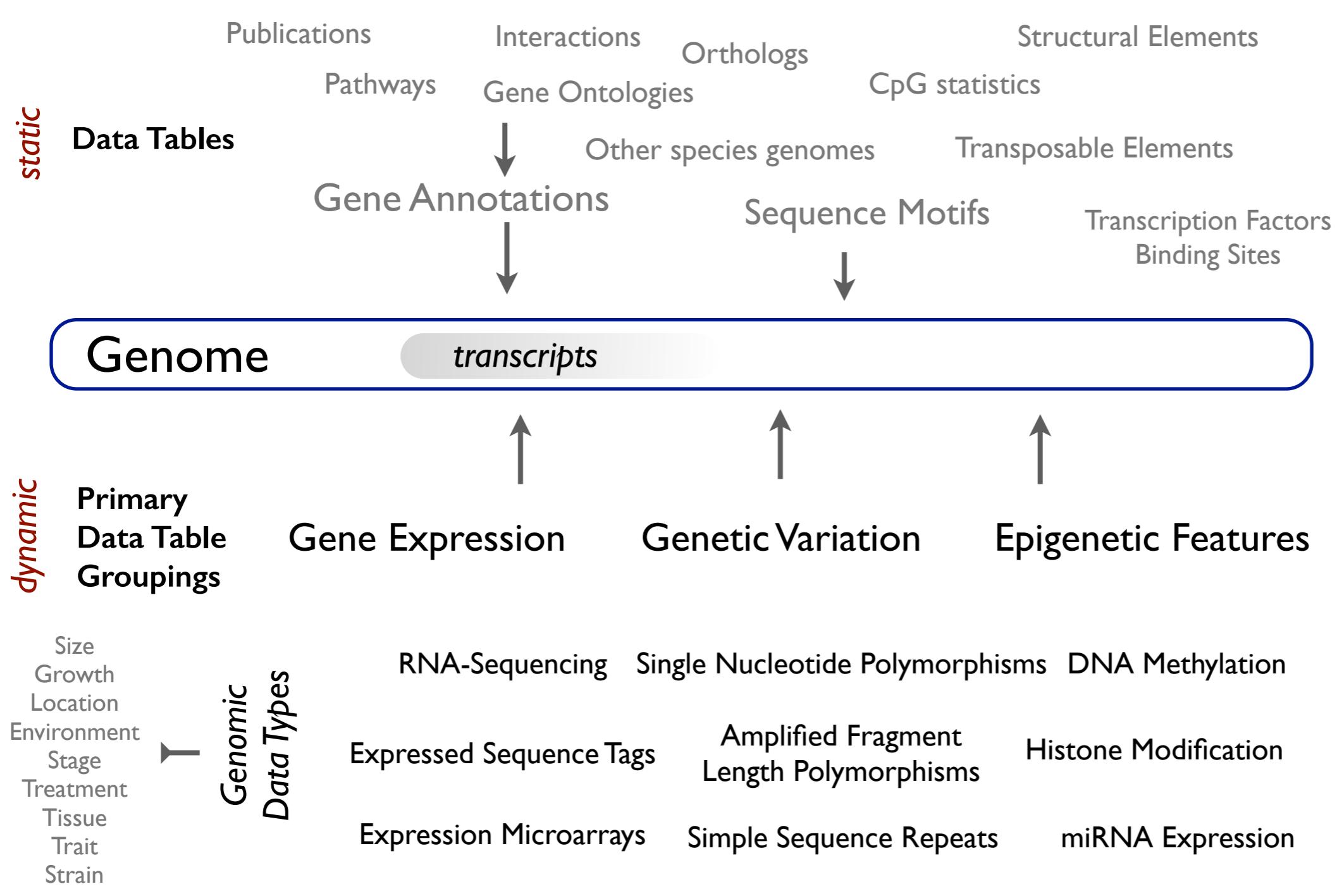
 Steven Roberts authored 32 seconds ago latest commit 48f8575540  example-data added some example data 17 hours ago README.md Create README.md 18 hours ago SQLShare-Intro.md init 18 hours ago fasta2slim.ipynb file ext an hour ago robertslab-qdod-snapshots.md init 18 hours ago robertslab-queries.md init 18 hours ago slidedeck.pdf deleted html - added pdf 32 seconds ago workflow-Analysis-of-BSMAP-data.md added more workflows an hour ago workflow-Annotating-Oyster-Genes.md init 18 hours ago workflow-Calculating-gene-specific-methylation... added more workflows an hour ago workflow-Generating-Quantitative-Genomic-Fea... added more workflows an hour ago MISC_Tools_Scripts_for_Seminar_Handout.pdf

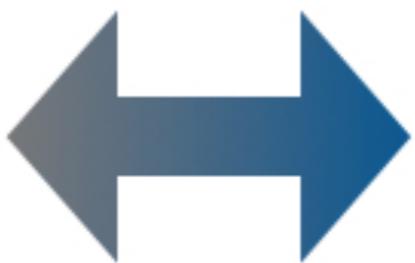
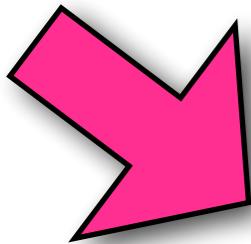
Domain Context



Environmental Physiology







blast

bedtools

file manipulation (unix shell)

R packages

trinity

stacks

tophat

bowtie

IGV

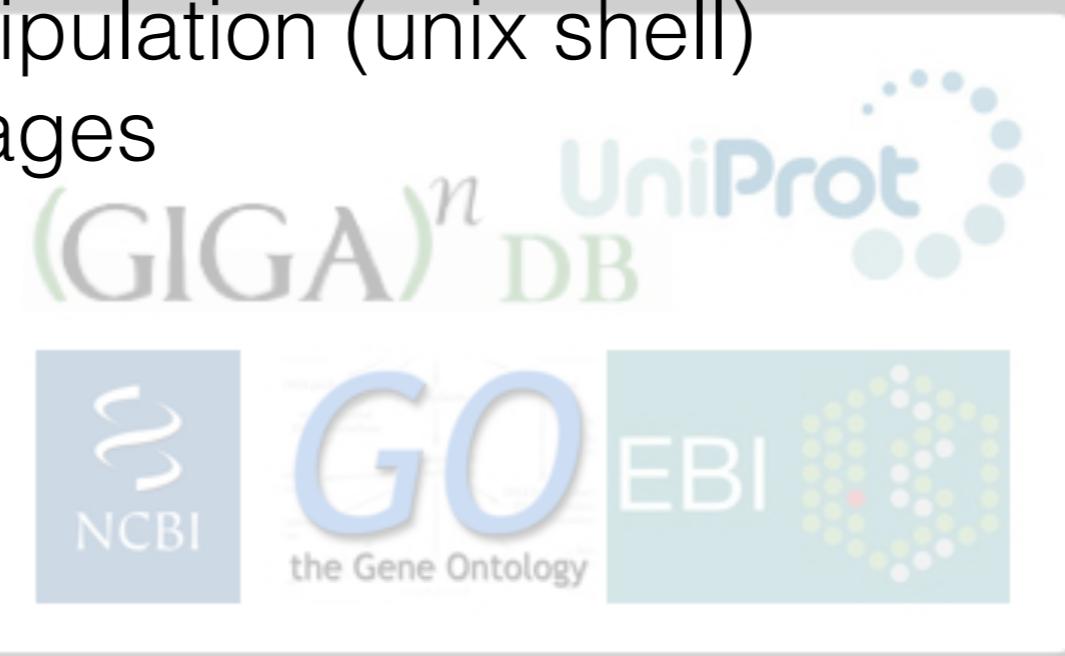
repeat masker

SQLSHARE

Galaxy

iPlant Collaborative™

Hyank



SQLShare for Annotation

case study - sea stars

Pycnopodia helianthoides
coelomocyte *de novo* transcriptome

Pycnopodia helianthoides coelomocyte de novo transcriptome



The top image is a healthy sunflower star. The middle image is a sick one. The last image is a Pisaster with the wasting sickness. (Hewson et. al)

Keystone species help maintain an ecosystem by eating quickly-reproducing prey species like urchins

The screenshot shows a news article from The Verge. At the top right are social sharing icons (Facebook, Twitter, Google+, YouTube, RSS) and a '78 NEW ARTICLES' badge. The main navigation bar includes LOGIN | SIGN UP, LONGFORM, VIDEO, REVIEWS, TECH, SCIENCE, CULTURE, DESIGN, BUSINESS, US & WORLD, FORUMS, and a search icon. Below the header, there's a 'TRENDING NOW' section with a thumbnail and the text 'This is the most insane wireless router in the history of mankind'. The main story title is 'In search of the starfish killer: the quest to save the original keystone species'. It features a sub-headline in italics: 'Starfish are dissolving up and down the West Coast, and scientists finally think they know why. They hope'. The author is Elizabeth Lopatto, dated November 18, 2014, 10:41 am. The byline includes links for Email and @mslopatto. The text discusses the impact of starfish absence on kelp forests and fish populations. The article has 27 comments.

THE VERGE

TRENDING NOW
This is the most insane wireless router in the history of mankind

LOG IN | SIGN UP LONGFORM ▾ VIDEO ▾ REVIEWS ▾ TECH ▾ SCIENCE ▾ CULTURE ▾ DESIGN ▾ BUSINESS ▾ US & WORLD ▾ FORUMS ▾

PREVIOUS STORY
Former BlackBerry CEO Thorsten Heins is now the top boss at Powermat

NEXT STORY
WhatsApp announces unprecedented encryption features

SCIENCE | US & WORLD | FEATURES | LONGFORM

In search of the starfish killer: the quest to save the original keystone species

Starfish are dissolving up and down the West Coast, and scientists finally think they know why. They hope

By Elizabeth Lopatto on November 18, 2014 10:41 am Email @mslopatto

University of California-Santa Cruz. "It completely changes the community if there's no kelp, especially with respect to fish. They use kelp as a structure for recruitment and refuge."

Pycnopodia's absence may also lead to more lingcod, which grow to about 5 feet and 80 pounds. They lay eggs in nests, which are essentially massive embryos attached to the ocean floor, Miner says. The nests are a favorite of *Pycnopodia*. Lingcod are good at cleaning water, so if their numbers increase, the water quality may change. Though the lingcod boom sounds



HK_CF2



V_CF26



HK_CF35



V_CF34



HK_CF70



V_CF71

Injected with
Heat Killed Stuff

Injected with
Stuff (filtered to 0.2 um)

Total length of sequence: 40747496 bp

Total number of sequences: 29476



Injected with
Heat Killed Stuff

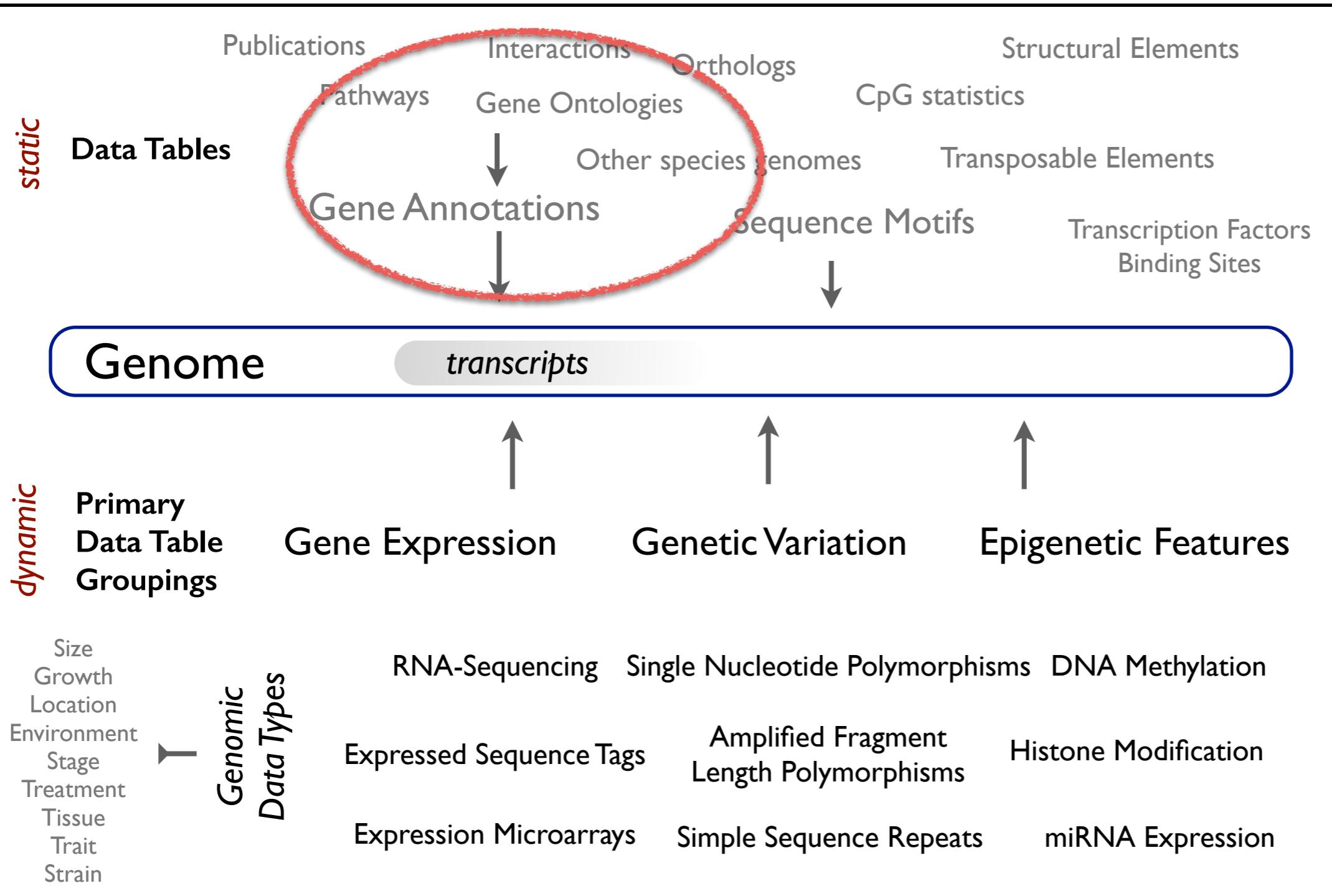


Injected with
Stuff (filtered to 0.2 um)

Total length of sequence: 40747496 bp
Total number of sequences: 29476

blastx

Phel_contig_100	sp Q16513 PKN2_HUMAN	81.33	332	61	1	7935	6940	653	983	5e-162	537
Phel_contig_1000	sp Q8R4U2 PDIA1_CRIGR	53.62	442	201	2	199	1512	31	472	5e-146	464
Phel_contig_10006	sp Q9Y2H9 MAST1_HUMAN	70.93	289	82	1	861	1	434	722	6e-132	415
Phel_contig_10021	sp Q96MU7 YTDC1_HUMAN	60.85	212	82	1	1115	1750	294	504	1e-73	258



[Your datasets](#)[All datasets](#)[Shared datasets](#)[Recent activity... 0](#)[Recently viewed »](#)[Upload dataset](#)[New query](#)**YOUR TOP VIEWED**

qDOD Cgigas ... 19

uniprot-review... 19

GPL11353_array 14

filt3_M3 14

UniprotProtNa... 12

POPULAR TAGS

proteomics 357

pnitzsch 189

oa 169

orbitrap 131

oyster 63

published 62

seaflow 52

protein 51

input 47

swissprot 38

techtrip 34

qdod2 30

bioinformatics 26

ssgcid 24

skyline 24

oceanography 23

blast 21

qdod 18

rna-seq 16

fluorimeter 16

optode 16

suna 16

tsg 16

qdod 16

eimd 15

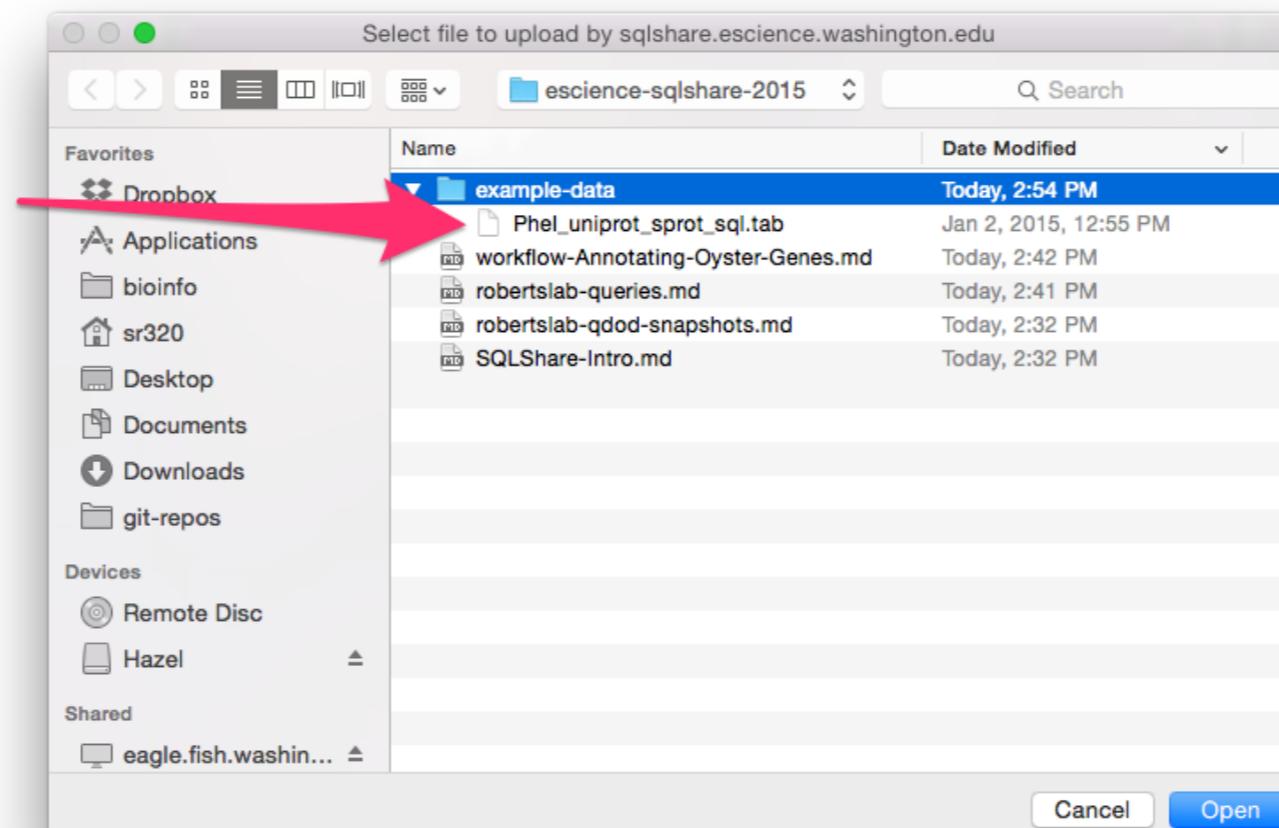
gill 15

Upload Dataset

1. Choose File 2. Import Settings 3. Save

File:
[Select file](#)

Accepted file types: Comma-separated values (.csv) and Plain text (.txt)

[Cancel](#)

Upload Dataset

1. Choose File

2. Import Settings

3. Save

Dataset was imported with the following settings:

You can change the parser options if your data was not properly imported.

 Contains column headerValues are separated by **Tab****DATASET PREVIEW** (Imported table with 14 columns)

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13
Phel_contig_100	sp	Q16513	PKN2_HUMAN	81.33	332	61	1	7935	6940	653	983	5e-162
Phel_contig_1000	sp	Q8R4U2	PDIA1_CRIGR	53.62	442	201	2	199	1512	31	472	5e-146
Phel_contig_10006	sp	Q9Y2H9	MAST1_HUMAN	70.93	289	82	1	861	1	434	722	6e-132
Phel_contig_10021	sp	Q96MU7	YTDC1_HUMAN	60.85	212	82	1	1115	1750	294	504	1e-73
Phel_contig_10025	sp	Q8BFQ8	PDDC1_MOUSE	55.56	216	92	2	60	704	8	220	2e-70

[Cancel](#)[Back](#) [Next](#)

Upload Dataset

1. Choose File

2. Import Settings

3. Save

Dataset was imported with the following settings:

You can change the parser options if your data was not properly imported.

 Contains column headerValues are separated by **Tab****DATASET PREVIEW** (Imported table with 14 columns)

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13
Phel_contig_100	sp	Q16513	PKN2_HUMAN	81.33	332	61	1	7935	6940	653	983	5e-162
Phel_contig_1000	sp	Q8R4U2	PDIA1_CRIGR	53.62	442	201	2	199	1512	31	472	5e-146
Phel_contig_10006	sp	Q9Y2H9	MAST1_HUMAN	70.93	289	82	1	861	1	434	722	6e-132
Phel_contig_10021	sp	Q96MU7	YTDC1_HUMAN	60.85	212	82	1	1115	1750	294	504	1e-73
Phel_contig_10025	sp	Q8BFQ8	PDDC1_MOUSE	55.56	216	92	2	60	704	8	220	2e-70

[Cancel](#)[Back](#) [Next](#)

SQLSHARE

Logged in: sr320@washington

Your datasets

All datasets

Shared datasets

Recent activity... 0

Recently viewed »

Upload dataset

New query

YOUR TOP VIEWED

qDOD Cgigas ... 19

uniprot-review... 19

GPL11353_array 14

filt3_M3 14

UniprotProtNa... 12

Upload Dataset

1. Choose File

2. Import Settings

3. Save

Title:

Phel_uniprot_sprot_sql.tab

Description:

Tags

 Share this dataset with everyone.[Cancel](#)[Back](#)[Save](#)

Your datasets
All datasets
Shared datasets
Recent activity... 0
Recently viewed »

Upload dataset
New query

YOUR TOP VIEWED

qDOD Cgigas ...	20
uniprot-review...	19
GPL11353_array	14
filt3_M3	14
UniprotProtNa...	12

POPULAR TAGS

proteomics	357
pnitzsch	189
oa	169
orbitrap	131
oyster	63
published	62
seaflow	52
protein	51
input	47
swissprot	38
techtrip	34
qdod2	30
bioinformatics	26
ssgcid	24
skyline	24

Phel_uniprot_sprot_sql.tab Last modified: Jan 5, 2015 2:29 PM  sr320@washington.edu*Click here to add a description**Click here to add a tag*

```
SELECT * FROM [sr320@washington.edu].[table_Phel_uniprot_sprot_sql.tab7D708]
```

[Edit dataset](#) [Derive dataset](#) [Create snapshot](#) [More actions ▾](#)

DATASET PREVIEW Rows 1 - 100 of 10513 | Columns 14 of 14

<< first < prev 1 [2](#) [3](#) [4](#) [5](#) [next >](#) [last >>](#)

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14
Phel_contig_100	sp	Q16513	PKN2_HUMAN	81.33	332	61	1	7935	6940	653	983	5E-162	537
Phel_contig_1000	sp	Q8R4U2	PDIA1_CRIGR	53.62	442	201	2	199	1512	31	472	5E-146	464
Phel_contig_10006	sp	Q9Y2H9	MAST1_HUMAN	70.93	289	82	1	861	1	434	722	6E-132	415
Phel_contig_10021	sp	Q96MU7	YTDC1_HUMAN	60.85	212	82	1	1115	1750	294	504	1E-73	258
Phel_contig_10025	sp	Q8BFQ8	PDDC1_MOUSE	55.56	216	92	2	60	704	8	220	2E-70	230
Phel_contig_10030	sp	Q5RA95	NDRG3_PONAB	44.7	264	139	1	2409	1618	69	325	3E-69	240
Phel_contig_10031	sp	Q8LEV3	Y2060_ARATH	23.11	251	142	9	254	916	344	573	5E-07	57
Phel_contig_10034	sp	B9EKI3	TMF1_MOUSE	45.89	717	354	7	3458	1344	392	1086	2E-161	521
Phel_contig_10039	sp	Q5FVB0	ATX10_XENTR	34.36	390	246	5	3	1166	93	474	1E-59	211
Phel_contig_10040	sp	Q6EAS5	GLGB_HORSE	67.75	555	177	1	2143	479	140	692	0	751
Phel_contig_10044	sp	Q13075	BIRC1_HUMAN	27.19	217	140	6	738	91	405	604	5E-12	73.9
Phel_contig_10047	sp	Q12767	K0195_HUMAN	46.72	122	61	2	6	368	1238	1356	1E-18	88.6
Phel_contig_10049	sp	Q54PZ6	Y5897_DICDI	25.82	275	177	7	1752	988	17	284	2E-20	99
Phel_contig_10050	sp	Q5R7A2	ALG1_PONAB	52.27	440	196	6	2351	1056	26	459	2E-150	457
Phel_contig_10053	sp	Q13075	BIRC1_HUMAN	24.71	255	165	5	738	1	405	641	2E-13	76.6
Phel_contig_10054	sp	P55175	Y601_SYNY3	40.94	254	143	6	839	93	13	264	7E-54	182
Phel_contig_10058	sp	P20749	BCL3_HUMAN	46.15	247	122	5	44	775	121	359	6E-58	204
Phel_contig_10068	sp	Q8BVE8	NSD2_MOUSE	62.76	333	118	2	3	995	999	1327	1E-127	406
Phel_contig_10069	sp	Q5XXB5	MSH2_CHLAE	65.11	599	202	2	3	1790	334	928	0	780
Phel_contig_10071	sp	A1A546	ISX_MOUSE	53.85	52	24	0	156	1	79	130	4E-11	65.5

Your datasets
All datasets
Shared datasets
Recent activity... 0
Recently viewed »

Upload dataset
New query

YOUR TOP VIEWED

qDOD Cgigas ... 20
uniprot-review... 19
GPL11353_array 14
filt3_M3 14
UniprotProtNa... 12

POPULAR TAGS

proteomics 357
pnitzsch 189
oa 169
orbitrap 131
oyster 63
published 62
seaflow 52
protein 51
input 47
swissprot 38
techtrip 34
qdod2 30
bioinformatics 26
ssgcid 24
skyline 24

Phel_uniprot_sprot_sql.tab Last modified: Jan 5, 2015 2:29 PM  sr320@washington.edu*Click here to add a description**Click here to add a tag*

```
SELECT * FROM [sr320@washington.edu].[table_Phel_uniprot_sprot_sql.tab7D708]
```

[Edit dataset](#) [Derive dataset](#) [Create snapshot](#) [More actions ▾](#)

DATASET PREVIEW Rows 1 - 100 of 10513 | Columns 14 of 14

<< first < prev 1 [2](#) [3](#) [4](#) [5](#) [next >](#) [last >>](#)

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14
Phel_contig_100	sp	Q16513	PKN2_HUMAN	81.33	332	61	1	7935	6940	653	983	5E-162	537
Phel_contig_1000	sp	Q8R4U2	PDIA1_CRIGR	53.62	442	201	2	199	1512	31	472	5E-146	464
Phel_contig_10006	sp	Q9Y2H9	MAST1_HUMAN	50.93	289	82	1	861	1	434	722	6E-132	415
Phel_contig_10021	sp	Q96MU7	YTDC1_HUMAN	60.5	212	82	1	1115	1750	294	504	1E-73	258
Phel_contig_10025	sp	Q8BFQ8	PDDC1_MOUSE	55.6	216	12	2	60	704	8	220	2E-70	230
Phel_contig_10030	sp	Q5RA95	NDRG3_PONAB	7	24	39	1	2409	1618	69	325	3E-69	240
Phel_contig_10031	sp	Q8LEV3	Y2060_ARATH	23.11	251	142	9	254	916	344	573	5E-07	57
Phel_contig_10034	sp	B9EKI3	TMF1_MOUSE	45.89	717	354	7	3458	1344	392	1086	2E-161	521
Phel_contig_10039	sp	Q5FVB0	ATX10_XENTR	34.36	390	246	5	3	1166	93	474	1E-59	211
Phel_contig_10040	sp	Q6EAS5	GLGB_HORSE	67.75	555	177	1	2143	479	140	692	0	751
Phel_contig_10044	sp	Q13075	BIRC1_HUMAN	27.19	217	140	6	738	91	405	604	5E-12	73.9
Phel_contig_10047	sp	Q12767	K0195_HUMAN	46.72	122	61	2	6	368	1238	1356	1E-18	88.6
Phel_contig_10049	sp	Q54PZ6	Y5897_DICDI	25.82	275	177	7	1752	988	17	284	2E-20	99
Phel_contig_10050	sp	Q5R7A2	ALG1_PONAB	52.27	440	196	6	2351	1056	26	459	2E-150	457
Phel_contig_10053	sp	Q13075	BIRC1_HUMAN	24.71	255	165	5	738	1	405	641	2E-13	76.6
Phel_contig_10054	sp	P55175	Y601_SYNY3	40.94	254	143	6	839	93	13	264	7E-54	182
Phel_contig_10058	sp	P20749	BCL3_HUMAN	46.15	247	122	5	44	775	121	359	6E-58	204
Phel_contig_10068	sp	Q8BVE8	NSD2_MOUSE	62.76	333	118	2	3	995	999	1327	1E-127	406
Phel_contig_10069	sp	Q5XXB5	MSH2_CHLAE	65.11	599	202	2	3	1790	334	928	0	780
Phel_contig_10071	sp	A1A546	ISX_MOUSE	53.85	52	24	0	156	1	79	130	4E-11	65.5

Query

Your datasets
All datasets
Shared datasets
Recent activity... 0
Recently viewed »

Upload dataset
New query

Phel_uniprot_sprot_sql.tab

Last modified: Jan 5, 2015 2:29 PM

sr320@washington.edu

[Click here to add a description](#)

[Click here to add a tag](#)

```
SELECT * FROM [sr320@washington.edu].[table_Phel_uniprot_sprot_sql.tab7D708]
```

[Edit dataset](#) [Derive dataset](#) [Create snapshot](#) [More actions ▾](#)

DATASET PREVIEW Rows 1 - 100 of 10513 | Columns 14 of 14

<< first < prev 1 [2](#) [3](#) [4](#) [5](#) [next >](#) [last >>](#)

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14
Phel_contig_100	sp	Q16513	PKN2_HUMAN	81.33	332	61	1	7935	6940	653	983	5E-162	537
Phel_contig_1000	sp	Q8R4U2	PDIA1_CRIGR	53.62	442	201	2	199	1512	31	472	5E-146	464
Phel_contig_10006	sp	Q9Y2H9	MAST1_HUMAN	70.93	289	82	1	861	1	434	722	6E-132	415
Phel_contig_10021	sp	Q96MU7	YTDC1_HUMAN	60.85	212	82	1	1115	1750	294	504	1E-73	258
Phel_contig_10025	sp	Q8BFQ8	PDDC1_MOUSE	55.56	216	92	2	60	704	8	220	2E-70	230
Phel_contig_10030	sp	Q5RA95	NDRG3_PONAB	44.7	264	139	1	2409	1618	69	325	3E-69	240
Phel_contig_10031	sp	Q8LEV3	Y2060_ARATH	23.11	251	142	9	254	916	344	573	5E-07	57
Phel_contig_10034	sp	B9EKI3	TMF1_MOUSE	45.89	717	354	7	3458	1344	392	1086	2E-161	521
Phel_contig_10039	sp	Q5FVB0	ATX10_XENTR	34.36	390	246	5	3	1166	93	474	1E-59	211
Phel_contig_10040	sp	Q6EAS5	GLGB_HORSE	67.75	555	177	1	2143	479	140	692	0	751
Phel_contig_10044	sp	Q13075	BIRC1_HUMAN	27.19	217	140	6	738	91	405	604	5E-12	73.9
Phel_contig_10047	sp	Q12767	K0195_HUMAN	46.72	122	61	2	6	368	1238	1356	1E-18	88.6
Phel_contig_10049	sp	Q54PZ6	Y5897_DICDI	25.82	275	177	7	1752	988	17	284	2E-20	99
Phel_contig_10050	sp	Q5R7A2	ALG1_PONAB	52.27	440	196	6	2351	1056	26	459	2E-150	457
Phel_contig_10053	sp	Q13075	BIRC1_HUMAN	24.71	255	165	5	738	1	405	641	2E-13	76.6
Phel_contig_10054	sp	P55175	Y601_SYN3	40.94	254	143	6	839	93	13	264	7E-54	182
Phel_contig_10058	sp	P20749	BCL3_HUMAN	46.15	247	122	5	44	775	121	359	6E-58	204
Phel_contig_10068	sp	Q8BVE8	NSD2_MOUSE	62.76	333	118	2	3	995	999	1327	1E-127	406
Phel_contig_10069	sp	Q5XXB5	MSH2_CHLAE	65.11	599	202	2	3	1790	334	928	0	780
Phel_contig_10071	sp	A1A546	ISX_MOUSE	53.85	52	24	0	156	1	79	130	4E-11	65.5

<< first < prev 1 [2](#) [3](#) [4](#) [5](#) [next >](#) [last >>](#)



Your datasets
All datasets
Shared datasets
Recent activity... 1
Recently viewed »

Upload dataset

New query

YOUR TOP VIEWED

qDOD Cgigas ...	19
uniprot-review...	19
GPL11353_array	14
filt3_M3	14
UniprotProtNa...	12

POPULAR TAGS

proteomics	357
pnitzsch	189
oa	169
orbitrap	131
oyster	63
published	62
seaflow	52
protein	51
input	47
swissprot	38
techtrip	34
qdod2	30
bioinformatics	26
secnid	24

Start a New Query

SQL Query *(E.g. select * from table)*

[Run query](#) [Save query](#) [Download results](#)

Your datasets
All datasets
Shared datasets
Recent activity... 1
Recently viewed »

Upload dataset
New query

YOUR TOP VIEWED

qDOD Cgigas ...	19
uniprot-review...	19
GPL11353_array	14
filt3_M3	14
UniprotProtNa...	12

POPULAR TAGS

proteomics	357
pnitzsch	189
oa	169
orbitrap	131
oyster	63
published	62
seaflow	52
protein	51
input	47
swissprot	38
techtrip	34
qdod2	30
bioinformatics	26
ssgcid	24
skyline	24
oceanography	23
blast	21
qdod	18
rna-seq	16

Phel_uniprot_sprot_sql.tab < Viewable by everyone

Last modified: Jan 5, 2015 2:29 PM sr320@washington.edu

Click here to add a description *Click here to add a tag*

```
SELECT * FROM [sr320@washington.edu].[table_Phel_uniprot_sprot_sql.tab7D708]
```

Actions: Edit dataset Derive dataset Create snapshot More actions ▾

DATASET PREVIEW Rows 1 - 100 of 10513 | Columns 14 of 14

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14
Phel_contig_100	sp	Q16513	PKN2_HUMAN	81.33	332	61	1	1	1	1	1	1	1
Phel_contig_1000	sp	Q8R4U2	PDIA1_CRIGR	53.62	442	201	2	2	2	2	2	2	2
Phel_contig_10006	sp	Q9Y2H9	MAST1_HUMAN	70.93	289	82	1	1	1	1	1	1	1
Phel_contig_10021	sp	Q96MU7	YTDC1_HUMAN	60.85	212	82	1	1	1	1	1	1	1
Phel_contig_10025	sp	Q8BFQ8	PDDC1_MOUSE	55.56	216	92	2	2	2	2	2	2	2
Phel_contig_10030	sp	Q5RA95	NDRG3_PONAB	44.7	264	139	1	1	1	1	1	1	1
Phel_contig_10031	sp	Q8LEV3	Y2060_ARATH	23.11	251	142	9	9	9	9	9	9	9
Phel_contig_10034	sp	B9EKI3	TMF1_MOUSE	45.89	717	354	7	7	7	7	7	7	7
Phel_contig_10039	sp	Q5FVB0	ATX10_XENTR	34.36	390	246	5	5	5	5	5	5	5
Phel_contig_10040	sp	Q6EAS5	GLGB_HORSE	67.75	555	177	1	1	1	1	1	1	1
Phel_contig_10044	sp	Q13075	BIRC1_HUMAN	27.19	217	140	6	6	6	6	6	6	6
Phel_contig_10047	sp	Q12767	K0195_HUMAN	46.72	122	61	2	2	2	2	2	2	2
Phel_contig_10049	sp	Q54PZ6	Y5897_DICDI	25.82	275	177	7	7	7	7	7	7	7
Phel_contig_10050	sp	Q5R7A2	ALG1_PONAB	52.27	440	196	6	6	6	6	6	6	6
Phel_contig_10053	sp	Q13075	BIRC1_HUMAN	24.71	255	165	5	5	5	5	5	5	5
Phel_contig_10054	sp	P55175	Y601_SYNY3	40.94	254	143	6	6	6	6	6	6	6
Phel_contig_10058	sp	P20749	BCL3_HUMAN	46.15	247	122	5	5	5	5	5	5	5
Phel_contig_10068	sp	Q8BVE8	NSD2_MOUSE	62.76	333	118	2	2	2	2	2	2	2
Phel_contig_10069	sp	Q5XXB5	MSH2_CHLAE	65.11	599	202	2	2	2	2	2	2	2
Phel_contig_10071	sp	A1A546	ISX_MOUSE	53.85	52	24	0	0	0	0	0	0	0

Actions: Edit dataset Derive dataset Create snapshot More actions ▾

DERIVE file0 close x

```
SELECT * FROM [sr320@washington.edu].[Phel_uniprot_sprot_sql]
```

Actions: Execute query

No table available!

- Your datasets
- All datasets
- Shared datasets
- Recent activity... 1
- Recently viewed »

- Upload dataset
- New query

YOUR TOP VIEWED

qDOD Cgigas ...	19
uniprot-review...	19
GPL11353_array	14
filt3_M3	14
UniprotProtNa...	12
<hr/>	
POPULAR TAGS	
proteomics	357
pnitzsch	189
oa	169
orbitrap	131
oyster	63
published	62
seaflow	52
protein	51
input	47
swissprot	38
techtrip	25
qdod2	30
bioinformatics	26
ssgcid	24
skyline	24
oceanography	23
blast	21
qdod	18
rna-seq	16
fluorimeter	16
optode	16

Phel_uniprot_sprot_sql.tab Viewable by everyone

Last modified: Jan 5, 2015 2:29 PM by sr320@washington.edu

[Click here to add a description](#)

[Click here to add a tag](#)

```
SELECT * FROM [sr320@washington.edu].[table_Phel_uniprot_sprot_sql.tab7D708]
```

[Edit dataset](#) [Derive dataset](#) [Create snapshot](#) [More actions ▾](#)

DATASET PREVIEW Rows 1 - 100 of 10513 | Columns 14 of 14

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

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14
Phel_contig_100	sp	Q16513	PKN2_HUMAN	81.33	332	61	1	1	1	1	1	1	1
Phel_contig_1000	sp	Q8R4U2	PDIA1_CRIGR	53.62	442	201	2	2	2	2	2	2	2
Phel_contig_10006	sp	Q9Y2H9	MAST1_HUMAN	70.93	289	82	1	1	1	1	1	1	1
Phel_contig_10021	sp	Q96MU7	YTDC1_HUMAN	60.85	212	82	1	1	1	1	1	1	1
Phel_contig_10025	sp	Q8BFQ8	PDDC1_MOUSE	55.56	216	92	2	2	2	2	2	2	2
Phel_contig_10030	sp	Q5RA95	NDRG3_PONAB	44.7	264	139	1	1	1	1	1	1	1
Phel_contig_10031	sp	Q8LEV3	Y2060_ARATH	23.11	251	142	9	9	9	9	9	9	9
Phel_contig_10034	sp	B9EKI3	TMF1_MOUSE	45.89	717	354	7	7	7	7	7	7	7
Phel_contig_10039	sp	Q5FVB0	ATX10_XENTR	34.36	390	246	5	5	5	5	5	5	5
Phel_contig_10040	sp	Q6EAS5	GLGB_HORSE	67.75	555	177	1	1	1	1	1	1	1
Phel_contig_10044	sp	Q13075	BIRC1_HUMAN	27.19	217	140	6	6	6	6	6	6	6
Phel_contig_10047	sp	Q12767	K0195_HUMAN	46.72	122	61	2	2	2	2	2	2	2
Phel_contig_10049	sp	Q54PZ6	Y5897_DICDI	25.82	275	177	7	7	7	7	7	7	7
Phel_contig_10050	sp	Q5R7A2	ALG1_PONAB	52.27	440	196	6	6	6	6	6	6	6
Phel_contig_10053	sp	Q13075	BIRC1_HUMAN	24.71	255	165	5	5	5	5	5	5	5
Phel_contig_10054	sp	P55175	Y601_SYNY3	40.94	254	143	6	6	6	6	6	6	6
Phel_contig_10058	sp	P20749	BCL3_HUMAN	46.15	247	122	5	5	5	5	5	5	5
Phel_contig_10068	sp	Q8BVE8	NSD2_MOUSE	62.76	333	118	2	2	2	2	2	2	2
Phel_contig_10069	sp	Q5XXB5	MSH2_CHLAE	65.11	599	202	2	2	2	2	2	2	2
Phel_contig_10071	sp	A1A546	ISX_MOUSE	53.85	52	24	0	0	0	0	0	0	0

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

DERIVE file0 close x

```
SELECT * FROM [sr320@washington.edu].[Phel_uniprot_sprot_sql]
```

[Execute query](#)

No table available!



Phel_uniprot_sprot_s... x

Log

Your datasets

All datasets

Shared datasets

Recent activity... 1

Recently viewed »

Upload dataset

New query

YOUR TOP VIEWED

qDOD Cgigas ... 19

uniprot-review... 19

GPL11353_array 14

filt3_M3 14

UniprotProtNa... 12

POPULAR TAGS

proteomics 357

pnitzsch 189

oa 169

orbitrap 131

oyster 63

published 62

seaflow 52

protein 51

input 47

swissprot 38

techtrip 34

qdod2 30

Error!

There was an error retrieving your list of queries

DERIVE file0

SELECT * FROM [sr320@washington.edu].|

Execute query

No table available!



Your datasets

All datasets

Shared datasets

Recent activity... 1

Recently viewed »

Upload dataset

New query

YOUR TOP VIEWED

qDOD Cgigas ... 19

uniprot-review... 19

GPL11353_array 14

filt3_M3 14

UniprotProtNa... 12

POPULAR TAGS

proteomics 357

pnitzsch 189

oa 169

orbitrap 131

oyster 63

published 62

seaflow 52

protein 51

input 47

swissprot 38

techtrip 34

qdod2 30

Error!

There was an error retrieving your list of queries

DERIVE file0

SELECT * FROM [sr320@washington.edu].|

Execute query

No table available!

tangent



branch: master

sqlshare / CONTRIBUTING.md



dhalperi on Aug 20, 2014 add CONTRIBUTING file

1 contributor

14 lines (11 sloc) 0.93 kb

[Raw](#) [Blame](#) [History](#) 

Common issues

- If you are receiving an error uploading data to SQLShare, please check your quota. You can do so by following the instructions [here](#).

When creating new issues

- Please search the existing issues to see if a related issue has been posted before. Please also check closed issues that have been solved.
- When creating an issue, please include information such as:
 - your SQLShare user name
 - when relevant, a URL link to the problematic dataset on [sqlshare.escience.washington.edu](#). (Or the dataset name, if it is private)
 - when relevant, the complete SQL code for a query.
 - when relevant, a URL link to the problematic file, if it is public on the web.

When issues are resolved

- please come back to the issue's page on GitHub and tell us that the issue is resolved. You can then also please [close](#) the issue.

[Issues](#)[Pull requests](#)[Labels](#)[Milestones](#)

Please review the [guidelines for contributing](#) to this repository.



Tag query produces error message

[Write](#)[Preview](#) [Markdown supported](#) [Edit in fullscreen](#)

When I click on a tag in the left side bar I receive an error.

see screenshot.

The screenshot shows the SQLShare web application. On the left, there's a sidebar with navigation links like 'Inbox (3) - ro... ', 'Upload File', 'Queries tagge...', 'fasta2slim/...', 'elmd_analysis/...', 'elmd_analysis/...', 'fasta2slim/...', 'nbviewer/pyth...', and 'escience-s...'. Below these are sections for 'Your datasets', 'Shared datasets', 'Recent activity...', 'Recently viewed', 'Upload dataset', and 'New query'. A 'POPULAR TAGS' section lists tags with their counts: proteomics (357), pntzach (189), os (189), orbitrap (131), oyster (63), published (62), seaflow (52), protein (51), input (47), swissprot (36), techtrip (36), and qdod2 (30). A red arrow points to the 'qdod2' tag in this list. The main area shows a query editor with a 'DERIVE file0' section containing the SQL query: `SELECT * FROM [arc320@washington.edu].|`. Below it is a button labeled 'Execute query'. To the right, a message says 'No table available!'. At the bottom right of the main area is a green 'Submit new issue' button.

DERIVE

close x

```
SELECT * FROM [sr320@washington.edu].[Phel_uniprot_sprot_sql].  
left join  
[sr320@washington.edu].[uniprot-reviewed_wGO_010714]des  
on phel.Column3=des.Entry
```

Execute query

DATASET PREVIEW Rows 1 - 100 of 100 or more | Columns 27 of 27

Column1	Column2	Column3	Column4	Column5
Phel_contig_11186	sp	Q5RCU0	TF2AA_PONAB	77.94
Phel_contig_11111	sp	Q9Y2B5	VP9D1_HUMAN	44.91
Phel_contig_1111	sp	P80146	SEPR_THESR	45.51

Column13	C
5E-162	5
5E-146	4
6E-132	4
1E-73	2
2E-70	2
3E-69	2
5E-07	5
2E-161	5
1E-59	2
0	7
5E-12	7
1E-18	8
2E-20	9
2E-150	4
2E-13	7
7E-54	1
6E-58	2
1E-127	4
0	7
4E-11	6

more actions ▾

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

 Upload dataset

New query

YOUR TOP VIEWED

uniprot-review... 19

gDOD Gaigas ... 18

GPL11353_array 14

filt3_M3

UniprotProtNa... 12

Start a New Query

SQL Query (*E.g. select * from table*)

```
SELECT * FROM [sr320@washington.edu].[Phel_uniprot_sprot_sql.tab]phel  
left join  
[sr320@washington.edu].[uniprot-reviewed_wGO_010714]des  
on phel.Column3=des.Entry
```

[Run query](#) [Save query](#) [Download results](#)

DATASET PREVIEW Rows 1 - 100 of 100 or more | Columns 27 of 27

Column12	Column13	Column14	Entry	Entry name	Gene ontology IDs	Interacts with	Cross-reference (GO)	Gene ontology (GO)	Status	InterPro	Pathway	Protein names	Gene names	Organism
5e-162	537	Q16513	PKN2_HUMAN		GO:0005524; GO:0043297; GO:0043296; GO:0006915; GO:0007155; GO:0007049; GO:0051301; GO:0032154; GO:0005737; GO:0005856; GO:0010631; GO:0042826; GO:0030027; GO:0016020; GO:0030496; GO:0005634; GO:0032467; GO:0045931; GO:0004697; GO:0004674; GO:2000145; GO:0006355; GO:0007165; GO:0006351	Q15118		ATP binding; apical junction assembly; apical junction complex; apoptotic process; cell adhesion; cell cycle; cell division; cleavage furrow; cytoplasm; cytoskeleton; epithelial cell migration; histone deacetylase binding; lamellipodium; membrane; midbody; nucleus; positive regulation of cytokinesis; positive regulation of mitotic cell cycle; protein kinase C activity; protein serine/threonine kinase activity; regulation of cell motility; regulation of transcription, DNA-dependent; signal transduction; transcription, DNA-dependent	reviewed	IPR000961; IPR000008; IPR011072; IPR011009; IPR017892; IPR000719; IPR017441; IPR002290; IPR008271;		Serine/threonine-protein kinase N2 (EC 2.7.11.13) (PKN gamma) (Protein kinase C-like 2) (Protein-kinase C-related kinase 2)	PKN2 PRK2 PRKCL2	Homo sapiens (Human)
5e-146	464	Q8R4U2	PDIA1_CRIGR		GO:0045454; GO:0005788; GO:0006662; GO:0042470; GO:0005886; GO:0003756; GO:0015035; GO:0006457			cell redox homeostasis; endoplasmic reticulum lumen; glycerol ether metabolic process; melanosome; plasma membrane; protein disulfide isomerase activity; protein disulfide oxidoreductase activity; protein folding	reviewed	IPR005788; IPR005792; IPR005746; IPR012336; IPR017937; IPR013766;		Protein disulfide-isomerase (PDI) (EC 5.3.4.1) (Prolyl 4-hydroxylase subunit beta) (p58)	P4HB PDIA1	Cricetulus griseus (Chinese hamster) (Cricetulus barabensis griseus)
6e-132	415	Q9Y2H9	MAST1_HUMAN		GO:0005524; GO:0005737; GO:0005856; GO:0007010; GO:0007243; GO:0000287; GO:0005886;	P00533		ATP binding; cytoplasm; cytoskeleton; cytoskeleton organization; intracellular protein kinase cascade; magnesium ion binding; plasma membrane; protein serine/threonine kinase	reviewed	IPR000961; IPR011009; IPR015022; IPR023142; IPR001478; IPR000719; IPR002290;		Microtubule-associated serine/threonine-protein kinase 1 (EC 2.7.11.1) (Syntrphin-associated serine/threonine-protein kinase)	MAST1 KIAA0973 SAST	Homo sapiens (Human)

- Your datasets
- All datasets
- Shared datasets
- Recent activity...
- Recently viewed

[Upload dataset](#)

New query

YOUR TOP VIEWED

uniprot-review... 19

qDOD Cqiqas ... 18

GPL11353_array 14

filt3_M3 14

Start a New Query

SQL Query (E.g. `select * from table`)

```
SELECT * FROM [sr320@washington.edu].[Phel_uniprot_sprot_sql.tab]phel
left join
[sr320@washington.edu].[uniprot-reviewed_wGO_010714]des
on phel.Column3=des.Entry
```

[Run query](#) [Save query](#) [Download results](#)

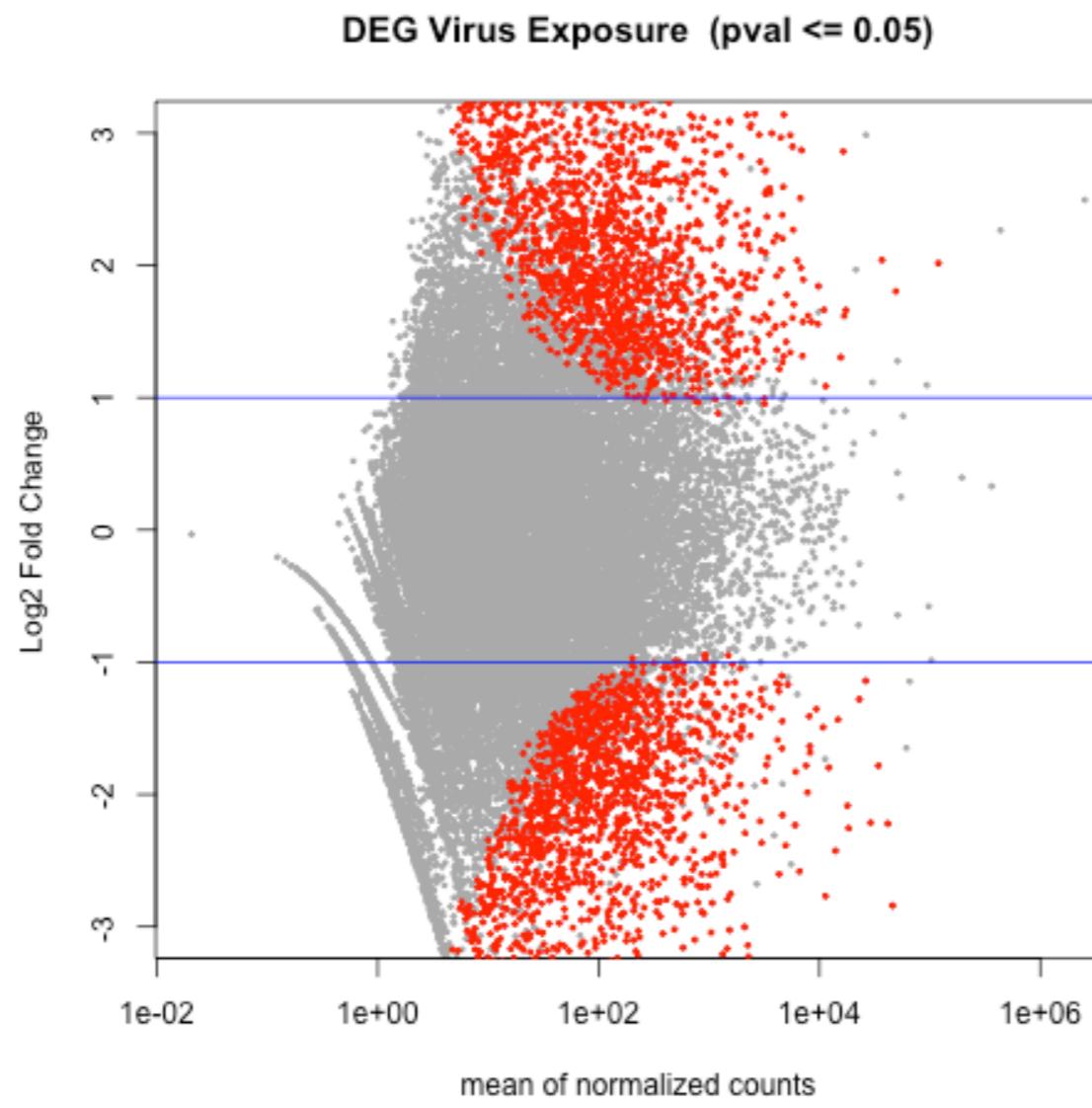
DATASET PREVIEW Rows 1 : 100 of 100 or more | Columns 27 of 27

Column12	Column13	Column14	Entry	Entry name	Gene ontology IDs	Interacts with	Cross-reference (GO)	Gene ontology (GO)	Status	InterPro	Pathway	Protein names	Gene names	Organism
5e-162	537	Q16513	PKN2_HUMAN		GO:0005524; GO:0043297; GO:0043296; GO:0006915; GO:0007155; GO:0007049; GO:0051301; GO:0032154; GO:0005737; GO:0005856; GO:0010631; GO:0042826; GO:0030027; GO:0016020; GO:0030496; GO:0005634; GO:0032467; GO:0045931; GO:0004697; GO:0004674; GO:2000145; GO:0006355; GO:0007165; GO:0006351	Q15118		ATP binding; apical junction assembly; apical junction complex; apoptotic process; cell adhesion; cell cycle; cell division; cleavage furrow; cytoplasm; cytoskeleton; epithelial cell migration; histone deacetylase binding; lamellipodium; membrane; midbody; nucleus; positive regulation of cytokinesis; positive regulation of mitotic cell cycle; protein kinase C activity; protein serine/threonine kinase activity; regulation of cell motility; regulation of transcription, DNA-dependent; signal transduction; transcription, DNA-dependent	reviewed	IPR000961; IPR000008; IPR011072; IPR011009; IPR017892; IPR000719; IPR017441; IPR002290; IPR008271;		Serine/threonine-protein kinase N2 (EC 2.7.11.13) (PKN gamma) (Protein kinase C-like 2) (Protein-kinase C-related kinase 2)	PKN2 PRK2 PRKCL2	Homo sapiens (Human)
5e-146	464	Q8R4U2	PDIA1_CRIGR		GO:0045454; GO:0005788; GO:0006662; GO:0042470; GO:0005886; GO:0003756; GO:0015035; GO:0006457			cell redox homeostasis; endoplasmic reticulum lumen; glycerol ether metabolic process; melanosome; plasma membrane; protein disulfide isomerase activity; protein disulfide oxidoreductase activity; protein folding	reviewed	IPR005788; IPR005792; IPR005746; IPR012336; IPR017937; IPR013766;		Protein disulfide-isomerase (PDI) (EC 5.3.4.1) (Prolyl 4-hydroxylase subunit beta) (p58)	P4HB PDIA1	Cricetulus griseus (Chinese hamster) (Cricetulus barabensis griseus)
6e-132	415	Q9Y2H9	MAST1_HUMAN		GO:0005524; GO:0005737; GO:0005856; GO:0007010; GO:0007243; GO:0000287; GO:0005886;	P00533		ATP binding; cytoplasm; cytoskeleton; cytoskeleton organization; intracellular protein kinase cascade; magnesium ion binding; plasma membrane; protein serine/threonine kinase	reviewed	IPR000961; IPR011009; IPR015022; IPR023142; IPR001478; IPR000719; IPR002290;		Microtubule-associated serine/threonine-protein kinase 1 (EC 2.7.11.1) (Synaptophelin-associated serine/threonine-protein kinase)	MAST1 KIAA0973 SAST	Homo sapiens (Human)

Outside of SQLShare

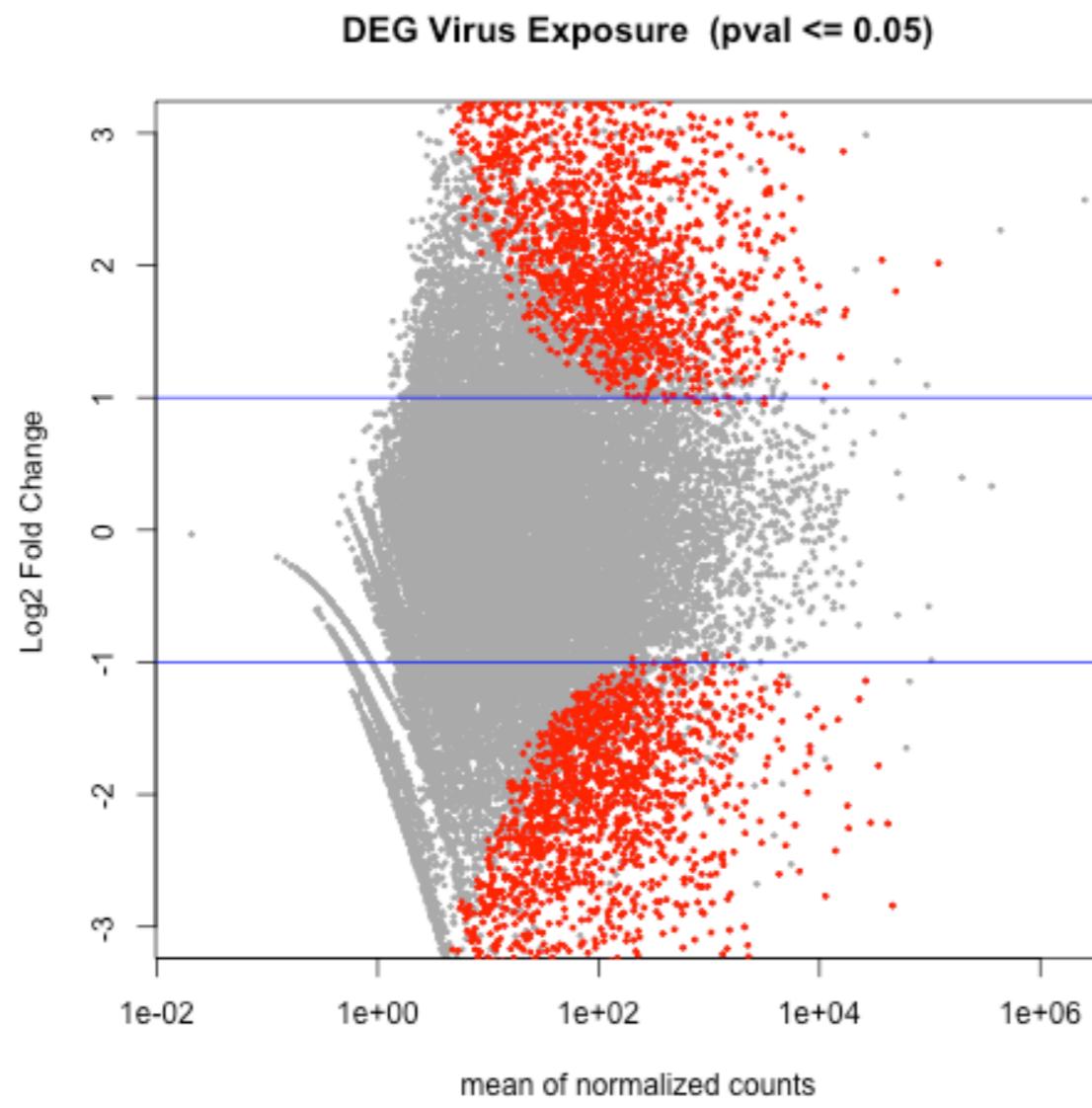
```
[26]: %R  
# Count number of hits with adjusted p-value less than 0.05  
dim(deseq2.res[!is.na(deseq2.res$padj) & deseq2.res$padj <= 0.05, ])  
[1] 3773    6
```

```
[27]: %R  
tmp <- deseq2.res  
# The main plot  
plot(tmp$baseMean, tmp$log2FoldChange, pch=20, cex=0.45, ylim=c(-3, 3), log="x", col="darkgray",  
     main="DEG Virus Exposure (pval <= 0.05)",  
     xlab="mean of normalized counts",  
     ylab="Log2 Fold Change")  
# Getting the significant points and plotting them again so they're a different color  
tmp.sig <- deseq2.res[!is.na(deseq2.res$padj) & deseq2.res$padj <= 0.05, ]  
points(tmp.sig$baseMean, tmp.sig$log2FoldChange, pch=20, cex=0.45, col="red")  
# 2 FC lines  
abline(h=c(-1,1), col="blue")
```



```
[26]: %R  
# Count number of hits with adjusted p-value less than 0.05  
dim(deseq2.res[!is.na(deseq2.res$padj) & deseq2.res$padj <= 0.05, ])  
[1] 3773    6
```

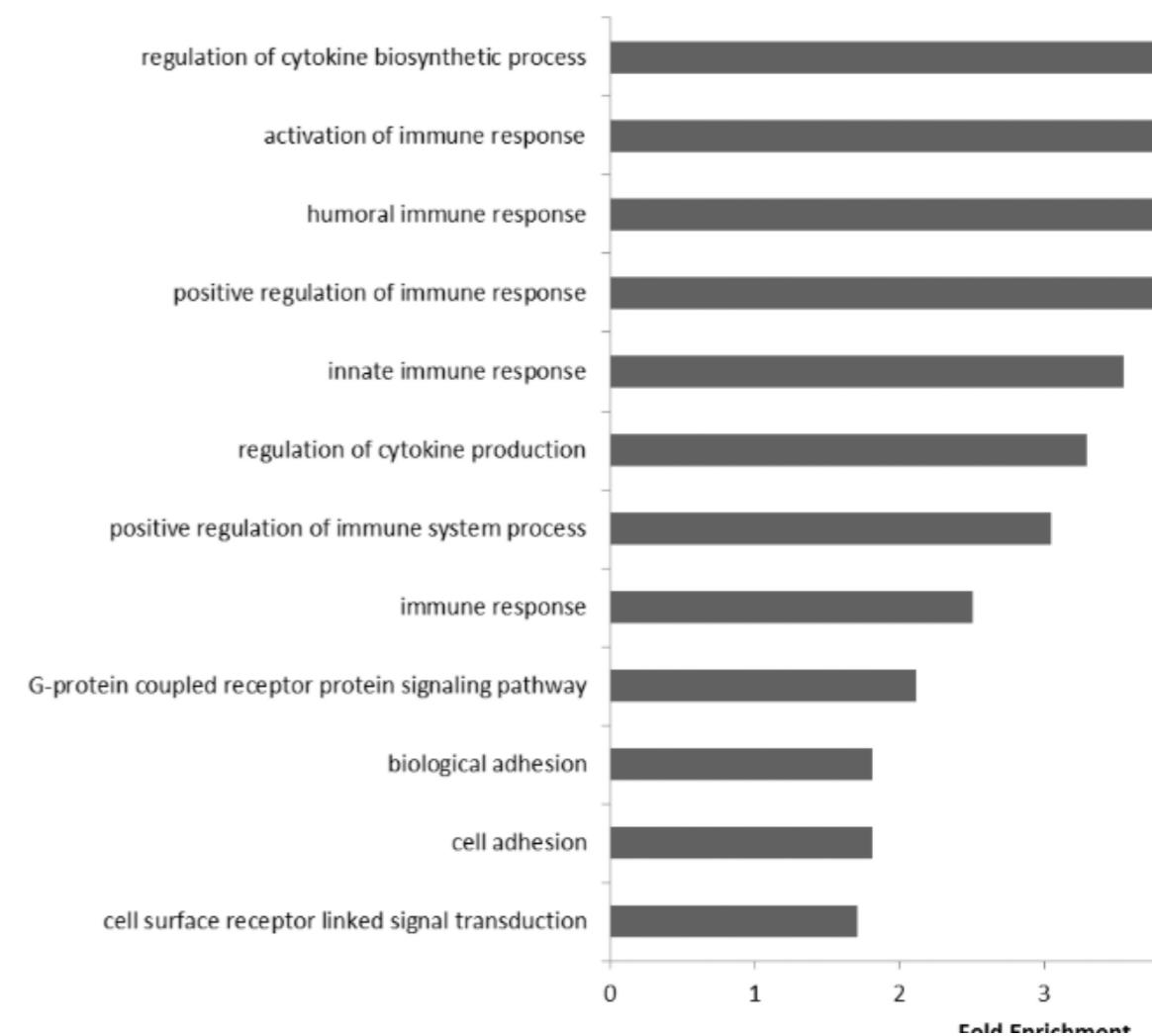
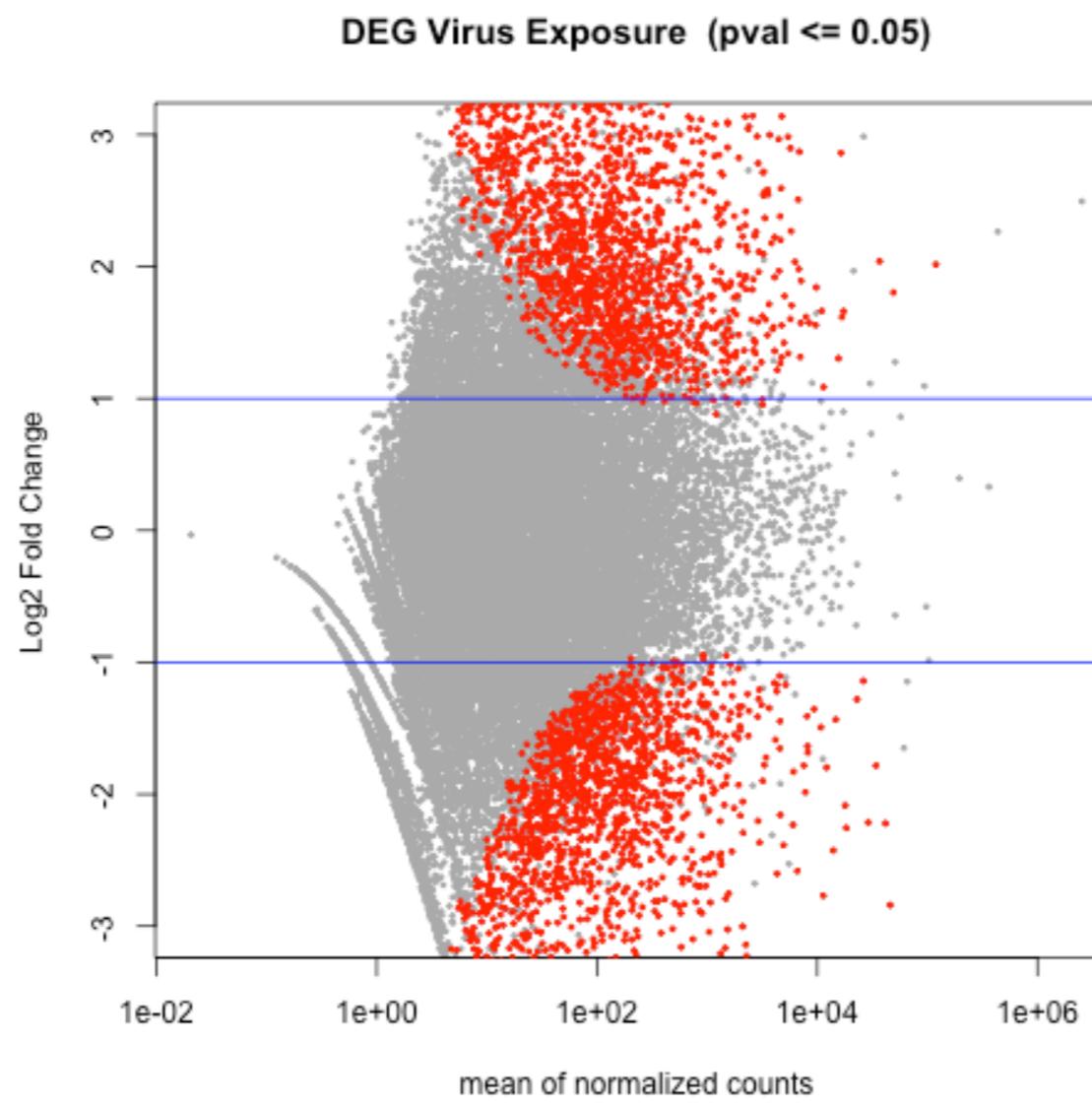
```
[27]: %R  
tmp <- deseq2.res  
# The main plot  
plot(tmp$baseMean, tmp$log2FoldChange, pch=20, cex=0.45, ylim=c(-3, 3), log="x", col="darkgray",  
     main="DEG Virus Exposure (pval <= 0.05)",  
     xlab="mean of normalized counts",  
     ylab="Log2 Fold Change")  
# Getting the significant points and plotting them again so they're a different color  
tmp.sig <- deseq2.res[!is.na(deseq2.res$padj) & deseq2.res$padj <= 0.05, ]  
points(tmp.sig$baseMean, tmp.sig$log2FoldChange, pch=20, cex=0.45, col="red")  
# 2 FC lines  
abline(h=c(-1,1), col="blue")
```



```
[26]: %R
# Count number of hits with adjusted p-value less than 0.05
dim(deseq2.res[!is.na(deseq2.res$padj) & deseq2.res$padj <= 0.05, ])

[1] 3773    6
```

```
[27]: %R
tmp <- deseq2.res
# The main plot
plot(tmp$baseMean, tmp$log2FoldChange, pch=20, cex=0.45, ylim=c(-3, 3), log="x", col="darkgray",
      main="DEG Virus Exposure (pval <= 0.05)",
      xlab="mean of normalized counts",
      ylab="Log2 Fold Change")
# Getting the significant points and plotting them again so they're a different color
tmp.sig <- deseq2.res[!is.na(deseq2.res$padj) & deseq2.res$padj <= 0.05, ]
points(tmp.sig$baseMean, tmp.sig$log2FoldChange, pch=20, cex=0.45, col="red")
# 2 FC lines
abline(h=c(-1,1), col="blue")
```



Value

- Very Large Delimited Files (frees disk space)
- Common Data Sources for Lab Group
- Ideal for Teaching (platform independant)

SQLShare -
pythonclient
generic annotation



branch: master ▾

[sqlshare-pythonclient / README.md](#) **billhowe** on May 28, 2014 Added link to nbviewer of the committed IPython notebook2 contributors  

177 lines (95 sloc) | 7.196 kb

[Raw](#)[Blame](#)[History](#)

☞ A Python client for the SQLShare REST API.



branch: master ▾

sqlshare-pythonclient / README.md

Quick Setup Instructions

These quick instructions are provided for experts comfortable with the command-line environments and open-source tools. For more detailed instructions, see [below](#).

1. Download the source and install the API

```
git clone git://github.com/uwescience/sqlshare-pythonclient.git
```

2. Install the Python libraries.

```
sudo python setup.py install
```

3. Make sure your own API key is configured in your home directory, in the file `.sqlshare/config`.

On Mac or Linux platforms, you can create the directory: `mkdir -p ~/.sqlshare`.

Then create a file called `config` in that directory using your favorite editor (e.g., `vim ~/.sqlshare/config`) with the following contents:

```
[sqlshare]
user=your-sql-share-account-name
password=your-sql-share-account-key
```

4. Use the command line tools in `tools/*` or write your own Python programs using these tools as examples.

[Blame](#) [History](#)

API.

Fasta2Slim

This IPython notebook is intended to serve as a structured means to annotate sequences using UniProt/SwissProt database. The notebook can be easily modified to personal preferences. As developed, the notebook requires the user has the following software installed ...

- IPython
 - NCBI Blast
 - SQLShare Python Client

Instructions for use.

In a working directory of your choosing place query fasta file, naming as `query.fa`. Edit the cell below, providing the path to said working directory.

Identify the location of the blast database you would like to use and indicate path in the cell below.

Identify the location of your `sqlshare-pythonclient/tools` and indicate path in the cell below.

Change the input to the `usr` variable to reflect your SQLShare user account.

```
In [2]: #Location Variables  
wd="~/Desktop/test/"  
  
db="/Volumes/Bay3/Software/ncbi-blast-2.2.29+/db/uniprot_sprot_r2013_12"  
  
sqls="~/sqlshare-pythonclient/tools/"  
  
usr="sr320@washington.edu"
```

```
In [254]: !head {wd}query.fa
```

```
>PiuraChilensis_v1_contig_1  
ATTTACAATAACGAAGTAAAATAGATAACGTAAAAATAATCTTGGTGCTGGATGATCGATC  
AAGTTCACCAATATTTATTGTAAAAAAATCATTCTAACAGCATGAAATCGTGTACAATG  
TATAAACAAAGCAAATATATAACACTAAAGCAAGAGGGCGTAAGTGGGGGGTGGGTGAGA  
GTAAAAAAATTCAAACATGTCAAATACCCCGGCCTTAGCCTTAAAAGCACCAGGACTTCTT  
CCGTTGAAATAGGATAAATTAAAGACGCTTATGACGATGAAATATGAGTAAAGAG
```



repo for escience seminar - January 6 2015 — Edit

6 commits

1 branch

0 releases

1 contributor



branch: master

escience-talk-sqlshare-2015 / +



file ext

Steven Roberts authored 30 seconds ago	latest commit c6aa0ff62f	
example-data	added some example data	16 hours ago
README.md	Create README.md	16 hours ago
SQLShare-Intro.md	init	16 hours ago
fasta2slim.ipynb	file ext	30 seconds ago
robertslab-qdod-snapshots.md	init	16 hours ago
robertslab-queries.md	init	16 hours ago
workflow-Analysis-of-BSMAP-data.md	added more workflows	7 minutes ago
workflow-Annotating-Oyster-Genes.md	init	16 hours ago
workflow-Calculating-gene-specific-methylation....	added more workflows	7 minutes ago
workflow-Generating-Quantitative-Genomic-Fea...	added more workflows	7 minutes ago
workflow-Transfer-Coordinates-from-Genome-to...	added more workflows	7 minutes ago



Fasta2Slim

This IPython notebook is intended to serve as a structured means to annotate sequences using UniProt/SwissProt database. The notebook can be easily modified to personal preferences. As developed, the notebook requires the user has the following software installed ...

- IPython
 - NCBI Blast
 - SQLShare Python Client

Instructions for use.

In a working directory of your choosing place query fasta file, naming as `query.fa`. Edit the cell below, providing the path to said working directory.

Identify the location of the blast database you would like to use and indicate path in the cell below.

Identify the location of your `sqlshare-pythonclient/tools` and indicate path in the cell below.

Change the input to the `usr` variable to reflect your SQLShare user account.

```
In [2]: #Location Variables  
wd="~/Desktop/test/"  
  
db="/Volumes/Bay3/Software/ncbi-blast-2.2.29+/db/uniprot_sprot_r2013_12"  
  
sqls="~/sqlshare-pythonclient/tools/"  
  
usr="sr320@washington.edu"
```

```
In [254]: !head {wd}query.fa
```

```
>PiuraChilensis_v1_contig_1  
ATTTACAATAACGAAGTAAAATAGATAACGTAAAAATAATCTTGGTGCTGGATGATCGATC  
AAGTTCACCAATATTTATTGTAAAAAAATCATTCTAACAGCATGAAATCGTGTACAATG  
TATAAACAAAGCAAATATATAACACTAAAGCAAGAGGGCGTAAGTGGGGGGTGGGTGAGA  
GTAAAAAAATTCAAACATGTCAAATACCCCGGCCTAGCCTTAAAAGCACCAGGACTTCTT  
CCGTTGAAATACGATTAATTAAGACGCTTATGACGATGAAATATGAGTAAAGAG
```

File Edit View Insert Cell Kernel Help



Run

Run and Select Below

Run and Insert Below

Run All

Run All Above

Run All Below

Cell Toolbar: None

Fasta2Slim

This IPython notebook is personal preferences. As

- IPython
- NCBI Blast
- SQLShare Python CLI

structured means to annotate sequences using UniProt/SwissProt requires the user has the following software installed ...

Cell Type

Current Output

All Output

Instructions for use.

In a working directory of your choosing place query fasta file, naming as `query.fa`. Edit the cell below, providing th

Identify the location of the blast database you would like to use and indicate path in the cell below.

Identify the location of your `sqlshare-pythonclient/tools` and indicate path in the cell below.

Change the input to the `usr` variable to reflect your SQLShare user account.

```
In [2]: #Location Variables  
wd="~/Desktop/test/"  
  
db="/Volumes/Bay3/Software/ncbi-blast-2.2.29+/db/uniprot_sprot_r2013_12"  
  
sqls="~/sqlshare-pythonclient/tools/"  
  
usr="sr320@washington.edu"
```

```
In [254]: !head {wd}query.fa
```

```
In [2]: !head {wd}query.fa
```

```
>PiuraChilensis_v1_contig_1
ATTACAATACGAAGTAAAATAGATAACGTAAAAATAATCTTGGTGCTGGATGATCGATC
AAGTTCACCAATATTTATTGTAAAAAATCATTCTAACAGCATGAAATCGTGTACAATG
TATAAACAAAGCAAATATATAACACTAAAGCAAGAGGGCGTAAGTGGGGGGTGGGTGAGA
GTAAAAAAATTCAAACATGTCAAATACCCCGGCCTAGCCTTAAAAGCACCATGGACTTCT
GCCTTCAATAAGCATAAAATTAAAACACCTAATACACAATGAATATACAGATAAAACAGA
TTTATGAATAGTTGGTGTACATCTTTACAGCCATAAGCCTTCATTGCTTCCAAACG
TATAAAATCTGACTTGGAACAAATATACAGCCATGAGATATGACACAGCGAGCACTACAAT
ATATATTATCTTGTACTATACAGCCTGTACAAGAAAATTCTGGAATTGTCTTCACAAGA
GACAGAAAAATAGTTGCAATGTGAATGCTAGTCTACTATTGATCACAATTGGATAGAAA
```

```
In [3]: #number of sequences
!fgrep -c ">" {wd}query.fa
```

282

Blast

```
In [4]: !blastx \
-query {wd}query.fa \
-db {db} \
-max_target_seqs 1 \
-max_hsps 1 \
-outfmt 6 \
-evalue 1E-05 \
-num_threads 2 \
-out {wd}blast_sprot.tab
```

Number of matched sequences:

```
In [5]: !wc -l {wd}blast_sprot.tab
```

```
211 /Users/sr320/Desktop/test/blast_sprot.tab
```

```
In [6]: !tr '||' "\t" <{wd}blast_sprot.tab> {wd}blast_sprot_sql.tab  
!head -1 {wd}blast_sprot.tab  
!echo SQLShare ready version has Pipes converted to Tabs ....  
!head -1 {wd}blast_sprot_sql.tab
```

PiuraChilensis_v1_contig_3	sp Q6P9A1 ZN530_HUMAN	33.33	105	61
3 825 1118 414	516 1e-07 57.4			
SQLShare ready version has Pipes converted to Tabs				
PiuraChilensis_v1_contig_3	sp Q6P9A1 ZN530_HUMAN	33.33	105	
61 3 825 1118	414 516 1e-07 57.4			

Joining in SQL Share

```
In [7]: !python {sqls}singleupload.py \  
-d _blast_sprot \  
{wd}blast_sprot_sql.tab
```

```
processing chunk line 0 to 211 (0.000264167785645 s elapsed)  
pushing /Users/sr320/Desktop/test/blast_sprot_sql.tab...  
parsing 983DD315...  
finished _blast_sprot
```

```
In [8]: !python {sqls}fetchdata.py \  
-s "SELECT Column1, term, GOSlim_bin, aspect, ProteinName FROM [{usr}].[_blast_sprot]md left join [samwhite@washington.edu].[UniprotProtNamesReviewed_yes20130610]sp on md.Column3=sp.SPID left join [sr320@washington.edu].[SPID and GO Numbers]go on md.Column3=go.SPID left join [sr320@washington.edu].[GO_to_GOslim]slim on go.GOID=slim.GO_id where aspect like 'P'" \  
-f tsv \  
-o {wd}GODEscriptions.txt
```

```
SELECT Column1, term, GOSlim_bin, aspect, ProteinName
FROM [{usr}].[_blast_sprot]md
left join
[samwhite@washington.edu].[UniprotProtNamesReviewed_yes20130610]sp
on md.Column3=sp.SPID
left join
[sr320@washington.edu].[SPID and GO Numbers]go
on md.Column3=go.SPID
left join
[sr320@washington.edu].[GO_to_Goslim]slim
on go.GOID=slim.GO_id
where aspect like 'P'
```

Plot GoSlim terms

```
In [10]: pylab inline
```

Populating the interactive namespace from numpy and matplotlib

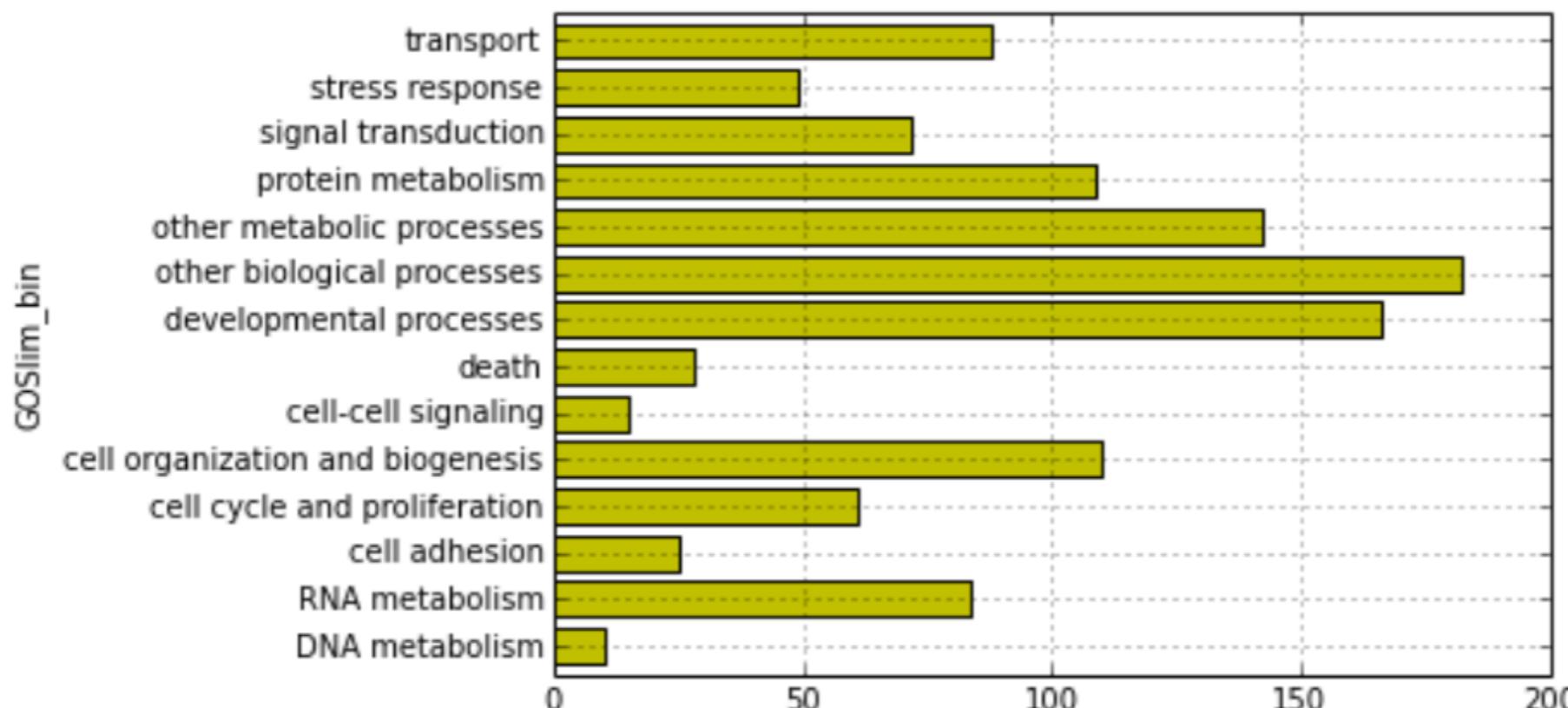
```
In [11]: cd {wd}
```

```
/Users/sr320/Desktop/test
```

```
In [12]: from pandas import *
```

```
gs = read_table('GOdescriptions.txt')
```

```
In [13]: gs.groupby('GOSlim_bin').Column1.count().plot(kind='barh', color=list('y'))
)
savefig('GOSlim.png', bbox_inches='tight')
```



Value

- Reproducible
- Great for teaching / troubleshooting

SQLShare

so much more...



repo for escience seminar - January 6 2015 — Edit

6 commits

1 branch

0 releases



branch: master ▾

escience-talk-sqlshare-2015 / +

file ext



Steven Roberts authored 22 minutes ago

1

example-data

added some example data

README.md

Create README.md

SQLShare-Intro.md

init

fasta2slim.ipynb

file ext

robertslab-qdod-snapshots.md

<https://github.com/sr320/escience-talk-sqlshare-2015/blob/master/robertslab-qdod-snapshots.md>

robertslab-queries.md

init

workflow-Analysis-of-BSMAP-data.md

added more workflows

workflow-Annotating-Oyster-Genes.md

init

workflow-Calculating-gene-specific-methylation.md

added more workflows

workflow-Generating-Quantitative-Genomic-Feat...

added more workflows

workflow-Transfer-Coordinates-from-Genome-to...

added more workflows

Crassostrea gigas male gonad transcriptome data comparison

Fastafile_Microarray	preview download
Rcode_OlsonandRoberts.R	preview download
SupplFig_OlsonandRoberts.pdf	preview download
Dheilly_SexSpecific_S1_GonadRNAseq.txt	preview download
Dheilly_DiffGametogenesis_S2_GonadR...txt	preview download

[Download all](#)

Share this:



0



1



0

Embed*

Cite this:

Olson, Claire; Roberts, Steven (2014): Crassostrea gigas male gonad transcriptome data comparison. figshare.

<http://dx.doi.org/10.6084/m9.figshare.1004464>

Retrieved 15:37, Jan 06, 2015 (GMT)

*The embed functionality can only be used for non commercial purposes... [more](#)

Description

This fileset represents a comparison of RNA-seq data from a Pacific oyster (*Crassostrea gigas*) male gamete library to microarray data from Dheilly et al. 2012 "Gametogenesis in the Pacific Oyster *Crassostrea gigas*: A Microarrays-Based Analysis Identifies Sex and Stage Specific Genes". Specifically sequence reads were mapped to full-length transcripts represented on the array platform used by Dheilly et al 2012. Data was joined using SQLShare.

126
views

1
shares

Published on 21 Apr 2014 - 19:42 (GMT)

Filesize in total is 24.93 MB

Categories

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

Authors

Claire Olson

Steven Roberts

Tags

- gene expression
- oyster
- dna methylation
- bivalves
- gametogenesis

License (what's this?)

CC-BY



repo for escience seminar - January 6 2015 — Edit

6 commits

1 branch

0 releases



branch: master ▾

escience-talk-sqlshare-2015 / +

file ext



Steven Roberts authored 22 minutes ago

1

example-data

added some example data

README.md

Create README.md

SQLShare-Intro.md

init

fasta2slim.ipynb

file ext

robertslab-qdod-snapshots.md

<https://github.com/sr320/escience-talk-sqlshare-2015/blob/master/robertslab-qdod-snapshots.md>

robertslab-queries.md

<https://github.com/sr320/escience-talk-sqlshare-2015/blob/master/robertslab-queries.md>

workflow-Analysis-of-BSMAP-data.md

added more workflows

workflow-Annotating-Oyster-Genes.md

init

workflow-Calculating-gene-specific-methylation.md

added more workflows

workflow-Generating-Quantitative-Genomic-Feat...

added more workflows

workflow-Transfer-Coordinates-from-Genome-to...

added more workflows

[Follow](#)  [Download as](#) [Introduction](#)[Methods](#)[Results](#)

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

Mackenzie R. Gavery, Steven B. Roberts 

DNA methylation landscape of genomic features

- In order to examine relationships between DNA methylation and genomic features, data from BSMP (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)

[Follow](#)  [Download as](#) [Introduction](#)[Methods](#)

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

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](#)[Download all](#)

Share this:

[Share](#)

0

[Tweet](#)

0

[g+1](#)

0

[Embed*](#)

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>

Retrieved 15:46, Jan 06, 2015 (GMT)

316
views

Published on 20 Sep 2013

Filesize in total is 92.1

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

Categories

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

Authors

Mackenzie Gavery

Steven Roberts

SQLShare

```
In [43]: #set location of sqlshare python  
pt="/Users/sr320/sqlshare-pythonclient/tools/"
```

```
In [258]: !python {pt}multiupload.py sqlr_*
```

```
uploading sT3D3  
uploading sT3D3 into []  
uploading sqlr_T3D3.txt  
uploading sqlr_T3D3.txt into ['sqlr_T3D3.txt']  
processing chunk line 0 to 1157463 (4.96696996689 s elapsed)  
pushing sqlr_T3D3.txt...  
parsing 567A0318...  
finished sqlr_T3D3.txt  
Successfully uploaded sqlr_T3D3.txt
```

```
In [426]: !python {pt}fetchdata.py -s "SELECT *, ([M1ratio]+[T1D3ratio]+[T1D5ratio])  
/3 as mean1 FROM [sr320@washington.edu].[sqlr_M1.txt]m1 inner join [sr320@  
washington.edu].[sqlr_T1D5.txt]t1d5 on m1.loci=t1d5.loci inner join [sr320  
@washington.edu].[sqlr_T1D3.txt]t1d3 on m1.loci=t1d3.loci where [M1ratio]-  
(([M1ratio]+[T1D3ratio]+[T1D5ratio])/3) < abs(.2) and [T1D3ratio]-(([M1rat  
io]+[T1D3ratio]+[T1D5ratio])/3) < abs(.2) and [T1D5ratio]-(([M1ratio]+[T1D  
3ratio]+[T1D5ratio])/3) < abs(.2)" -f tsv -o cglarv_lineage_1.txt
```

```
SELECT * ,
([M1ratio]+[T1D3ratio]+[T1D5ratio])/3 as mean1
FROM [sr320@washington.edu].[sqlr_M1.txt]m1
inner join
[sr320@washington.edu].[sqlr_T1D5.txt]t1d5
on m1.loci=t1d5.loci
inner join [sr320@washington.edu].[sqlr_T1D3.txt]t1d3
on m1.loci=t1d3.loci
where
[M1ratio]-(([M1ratio]+[T1D3ratio]+[T1D5ratio])/3) < abs(.2)
and
[T1D3ratio]-(([M1ratio]+[T1D3ratio]+[T1D5ratio])/3) < abs(.2)
and
[T1D5ratio]-(([M1ratio]+[T1D3ratio]+[T1D5ratio])/3) < abs(.2)
```

Shotgun proteomics reveals physiological response to ocean acidification in *Crassostrea gigas*

Emma Timmins-Schiffman, William D Coffey, Wilber Hua, Brook L Nunn, Gary H Dickinson and Steven B Roberts

BMC Genomics 2014, **15**:951 doi:10.1186/1471-2164-15-951

Published: 3 November 2014

Abstract (provisional)

Background

Ocean acidification as a result of increased anthropogenic CO₂ emissions is occurring in marine and estuarine environments worldwide. The coastal ocean experiences additional daily and seasonal fluctuations in pH that can be lower than projected end-of-century open ocean pH reductions. In order to assess the impact of ocean acidification on marine invertebrates, Pacific oysters (*Crassostrea gigas*) were exposed to one of four different pCO₂ levels for four weeks: 400 μatm (pH 8.0), 800 μatm (pH 7.7), 1000 μatm (pH 7.6), or 2800 μatm (pH 7.3).

SQLShare workflow for proteomics analysis

38
views

Supplementary Information S2

Proteomic Analysis Description

Workflow and data associated with expression values for oyster proteins based normalized spectral abundance factor (NSAF).

Published on 05 Jan 2

Filesize is 120.40 KB

Input files #1-48: Provided as worksheets in Supplemental Data S1. For each technical replicate the CGI ID, number of unique peptides, and total number of independent spectra are given. All input files can be found via this link

<https://sqlshare.escience.washington.edu/sqlshare#/s=tag/input>

Input file #49: [[all sequenced proteins all treatments.txt](#)] A non-redundant list of all the proteins sequenced across all 48 biological and technical replicates. This was compiled by combining the list of sequenced proteins for each of the 48 technical and biological replicates.

Query 1

To combine all technical replicates for each oyster, joined input file #49 (all proteins) with three technical replicate Input files for oysters to make 16 files (1 per oyster). Oysters 2-11 correspond to the 2800 μatm pCO₂ treatment, 26-35 are 2800 μatm + mechanical stress, 221-230 are 400 μatm pCO₂, and 242-251 are 400 μatm + mechanical stress.

Q1 outputs:

[[Oyster 2 spec counts](#)]

[[Oyster 5 spec counts](#)]

[[Oyster 8 spec counts](#)]

Categories

- Bioinformatics

Authors

Emma Timmins-Sch

Steven Roberts

Tags

NSAF based on avg spc

 Viewable by everyone

Las

No description provided by owner

oa orbitrap proteomics

```
SELECT [All Proteins], [CG2 unique peps sum], [CG5 unique peps sum],  
    spc.[CG2 SpC/L] / allspc.[SUM CG2 SpC/L] AS [NSAF CG2],  
    spc.[CG5 SpC/L] / allspc.[SUM CG5 SpC/L] AS [NSAF CG5],  
    spc.[CG8 SpC/L] / allspc.[SUM CG8 SpC/L] AS [NSAF CG8],  
    spc.[CG11 SpC/L] / allspc.[SUM CG11 SpC/L] AS [NSAF CG11],  
    spc.[CG26 SpC/L] / allspc.[SUM CG26 SpC/L] AS [NSAF CG26],  
    spc.[CG29 SpC/L] / allspc.[SUM CG29 SpC/L] AS [NSAF CG29],  
    spc.[CG32 SpC/L] / allspc.[SUM CG32 SpC/L] AS [NSAF CG32],  
    spc.[CG35 SpC/L] / allspc.[SUM CG35 SpC/L] AS [NSAF CG35],
```

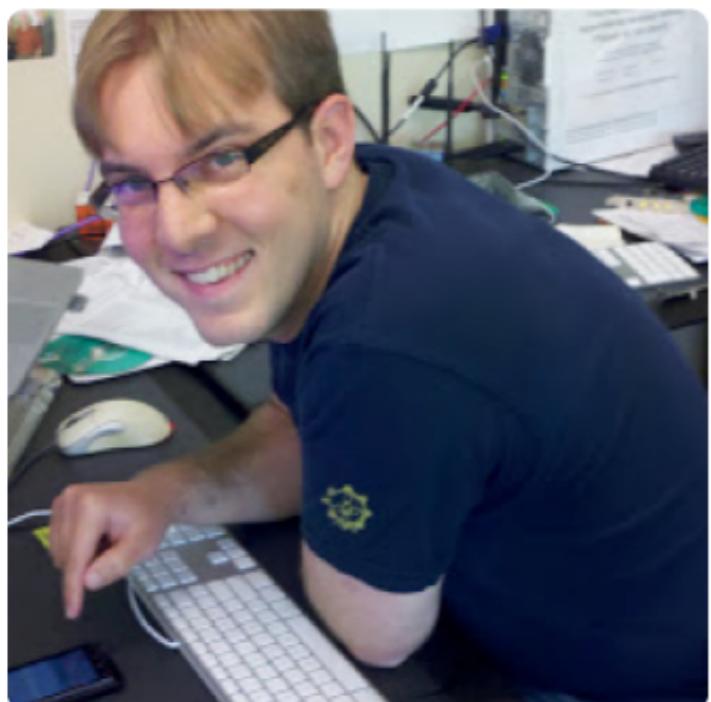
DATASET PREVIEW Rows 1 - 100 of 3826 | Columns 33 of 33

[<< first](#) [< prev](#) **1** [2](#) [3](#) [4](#) [5](#) [next >](#) [last >>](#)

All Proteins	CG2 unique peps sum	CG5 unique peps sum	CG8 unique peps sum	C
CGI_10000041	3	3	2	4
CGI_10000055	14	9	5	20

```
SELECT [All Proteins], [CG2 unique peps sum],
[CG5 unique peps sum], [CG8 unique peps sum],
[CG11 unique peps sum], [CG26 unique peps sum],
[CG29 unique peps sum], [CG32 unique peps sum],
[CG35 unique peps sum], [CG221 unique peps sum],
[CG224 unique peps sum], [CG227 unique peps sum],
[CG230 unique peps sum], [CG242 unique peps sum],
[CG245 unique peps sum], [CG248 unique peps sum],
[CG251 unique peps sum],
    spc.[CG2 SpC/L] / allspc.[SUM CG2 SpC/L] AS [NSAF CG2],
    spc.[CG5 SpC/L] / allspc.[SUM CG5 SpC/L] AS [NSAF CG5],
    spc.[CG8 SpC/L] / allspc.[SUM CG8 SpC/L] AS [NSAF CG8],
    spc.[CG11 SpC/L] / allspc.[SUM CG11 SpC/L] AS [NSAF CG11],
    spc.[CG26 SpC/L] / allspc.[SUM CG26 SpC/L] AS [NSAF CG26],
    spc.[CG29 SpC/L] / allspc.[SUM CG29 SpC/L] AS [NSAF CG29],
    spc.[CG32 SpC/L] / allspc.[SUM CG32 SpC/L] AS [NSAF CG32],
    spc.[CG35 SpC/L] / allspc.[SUM CG35 SpC/L] AS [NSAF CG35],
    spc.[CG221 SpC/L] / allspc.[SUM CG221 SpC/L] AS [NSAF CG221],
    spc.[CG224 SpC/L] / allspc.[SUM CG224 SpC/L] AS [NSAF CG224],
    spc.[CG227 SpC/L] / allspc.[SUM CG227 SpC/L] AS [NSAF CG227],
    spc.[CG230 SpC/L] / allspc.[SUM CG230 SpC/L] AS [NSAF CG230],
    spc.[CG242 SpC/L] / allspc.[SUM CG242 SpC/L] AS [NSAF CG242],
    spc.[CG245 SpC/L] / allspc.[SUM CG245 SpC/L] AS [NSAF CG245],
    spc.[CG248 SpC/L] / allspc.[SUM CG248 SpC/L] AS [NSAF CG248],
    spc.[CG251 SpC/L] / allspc.[SUM CG251 SpC/L] AS [NSAF CG251]
FROM [emmats@washington.edu].[SpC-L for all oysters (average spc)] spc,
[emmats@washington.edu].[Sum SpC-L for avg spc] allspc
```

Thanks



Daniel Halperin
dhalperi



billhowe



shrijain Shrainik

Questions