Genomics on the Half Shell

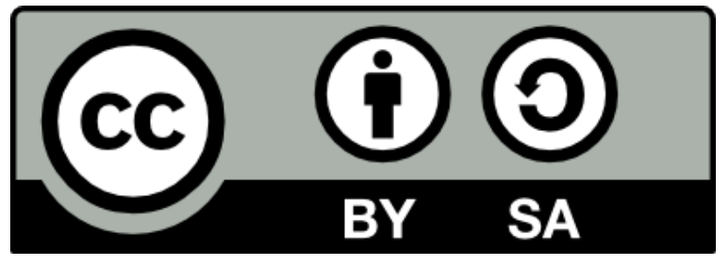
Environmental Epigenetics, Open Science, and the Oyster

Steven Roberts
Associate Professor
University of Washington
School of Aquatic and Fishery Sciences
robertslab.info
@sr320

OSU IB Seminar - May 4, 2015

Open Science

- You are free to Share!



- Our lab practices open notebook science

wikispaces
wikis for everyone

IP[y]: IPython
Interactive Computing

- Data, Preprints, Proposals, Lab Meetings, Web Cams, Slidedecks

These slides plus links @
github.com/sr320/talk-osu-2015

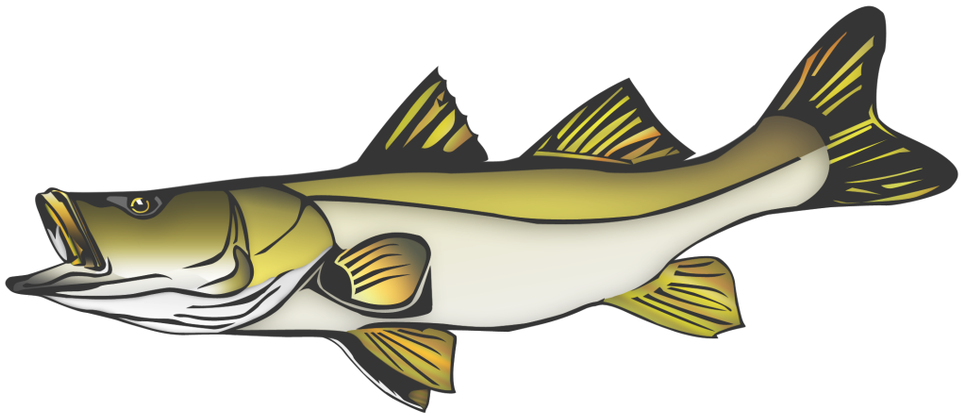
Background

Physiology

How fundamental processes work in aquatic species

Physiology

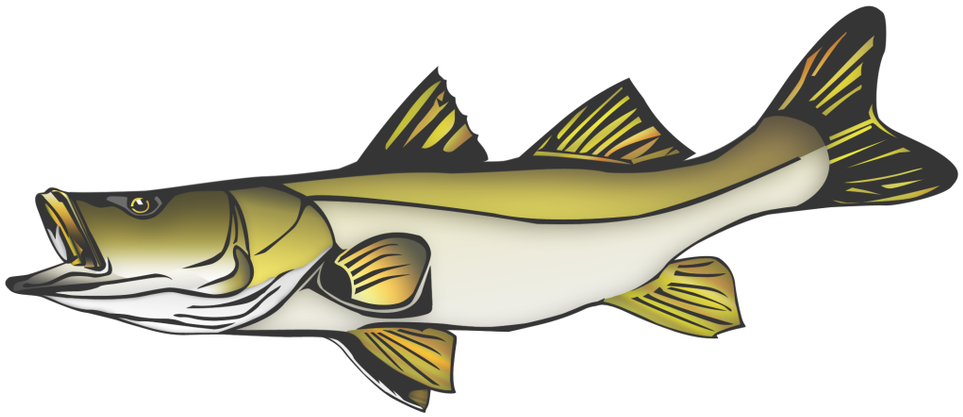
How fundamental processes work in aquatic species

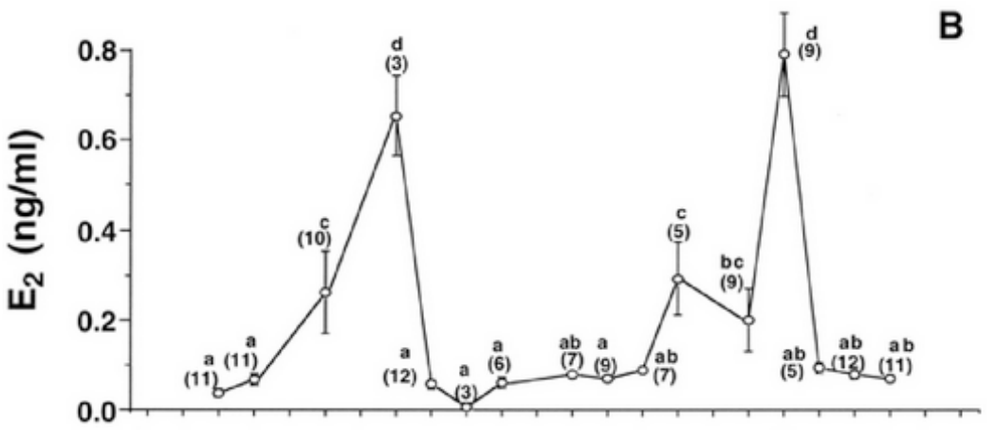


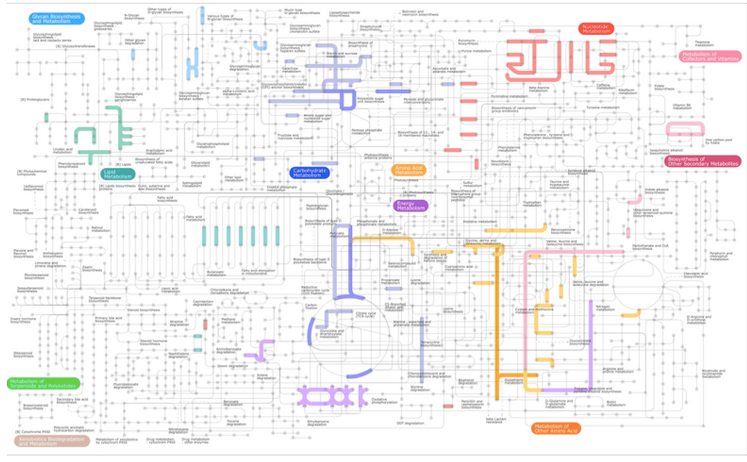
Background

Physiology

How fundamental processes work in aquatic species



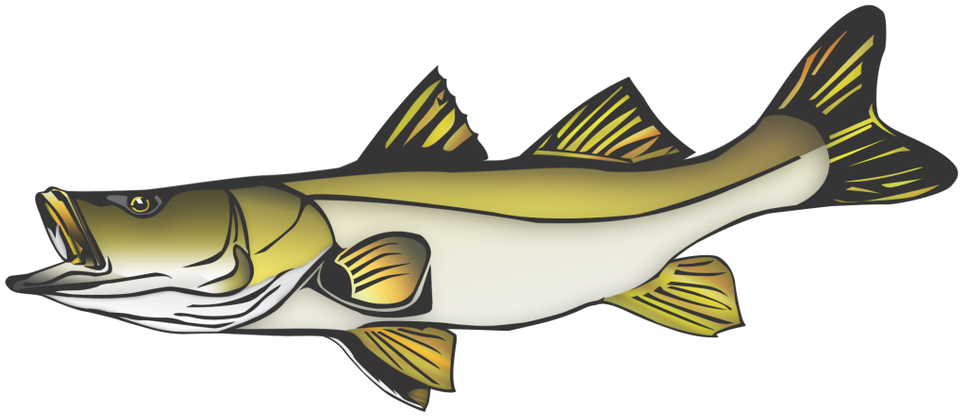




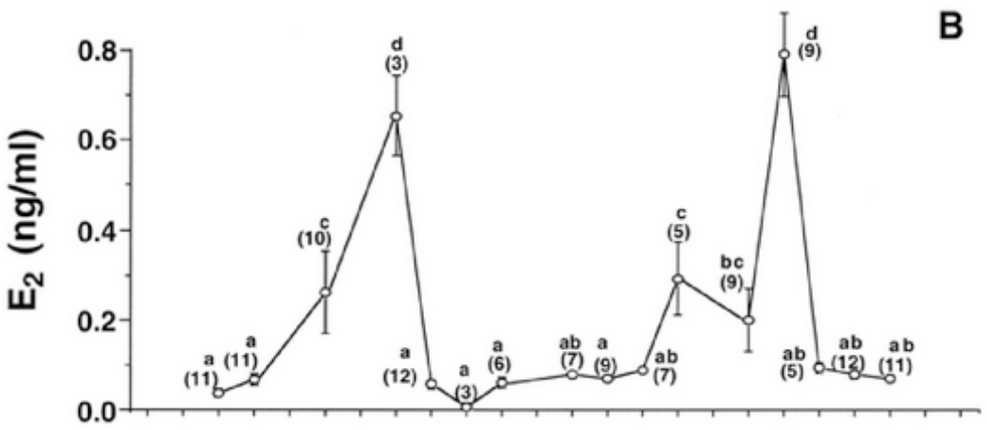
Background

Physiology

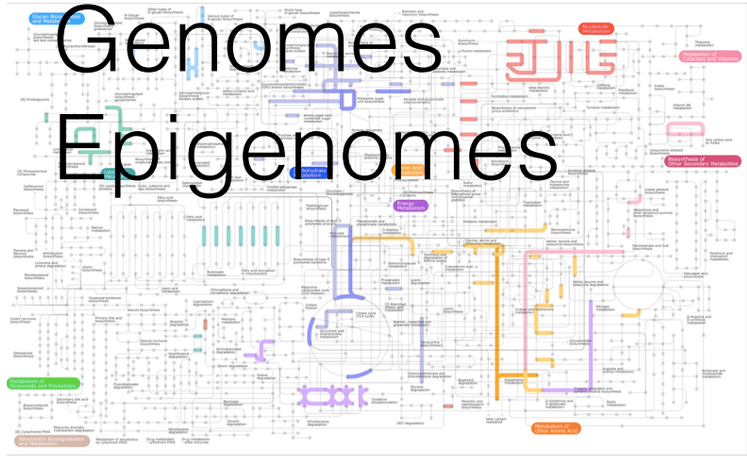
How fundamental processes work in aquatic species



Hormones
Proteins



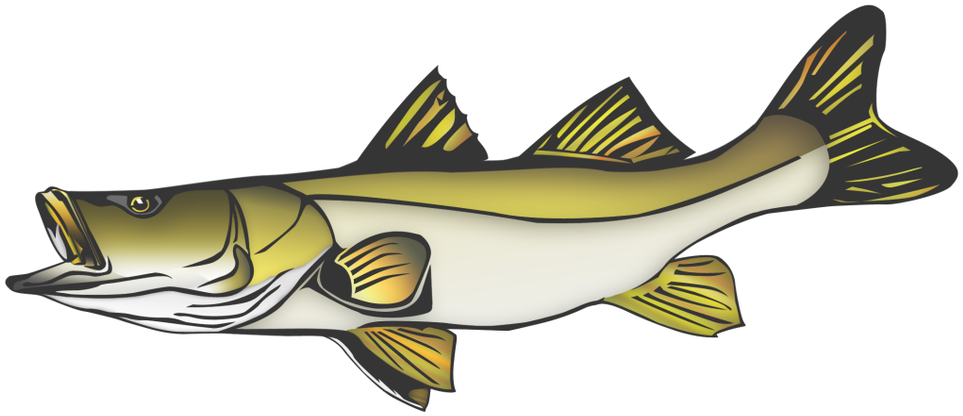
Transcriptomes
Proteomes
Genomes
Epigenomes

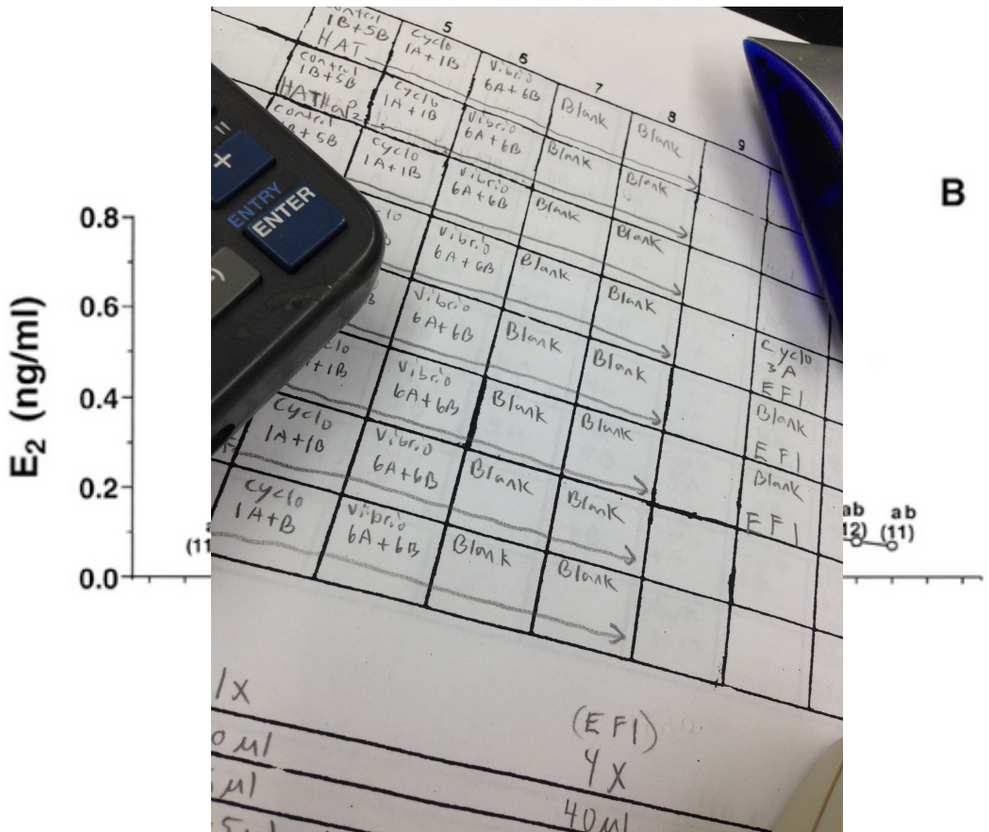


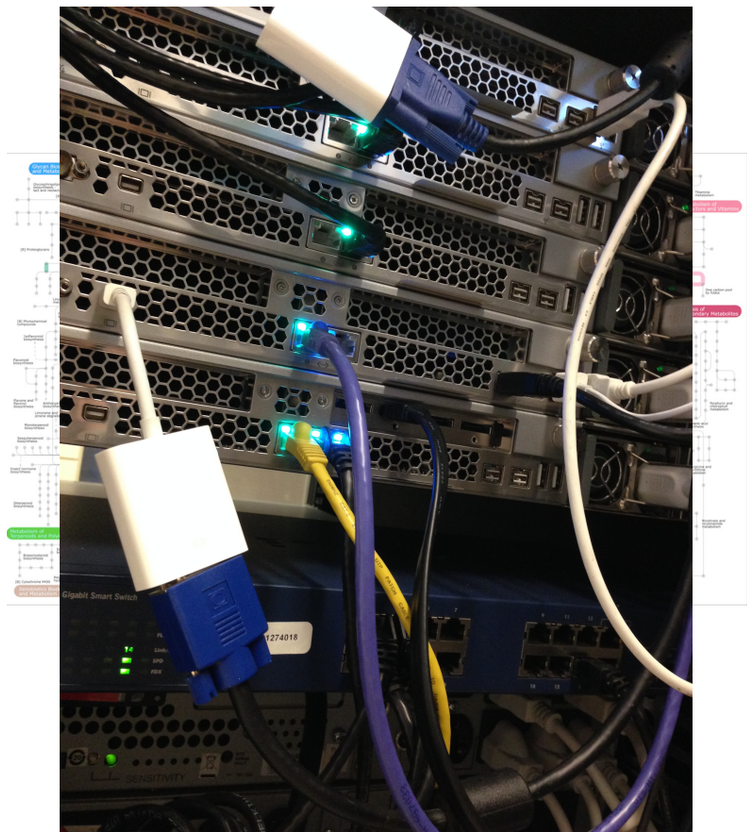
Background

Physiology

How fundamental processes work in aquatic species



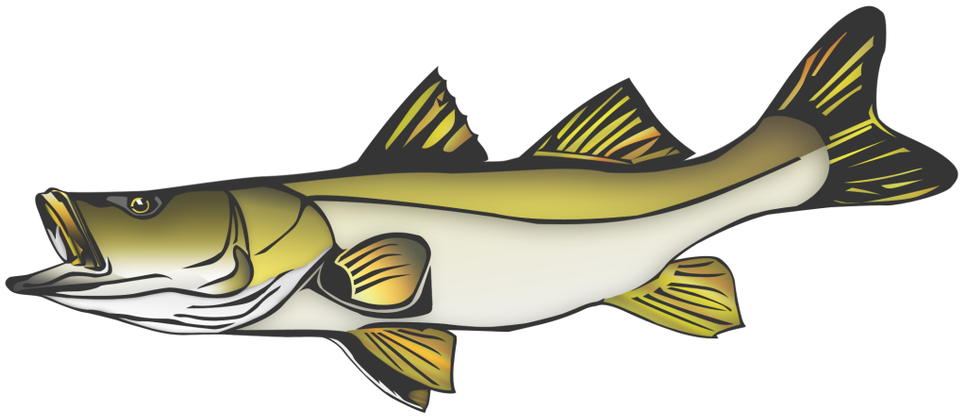


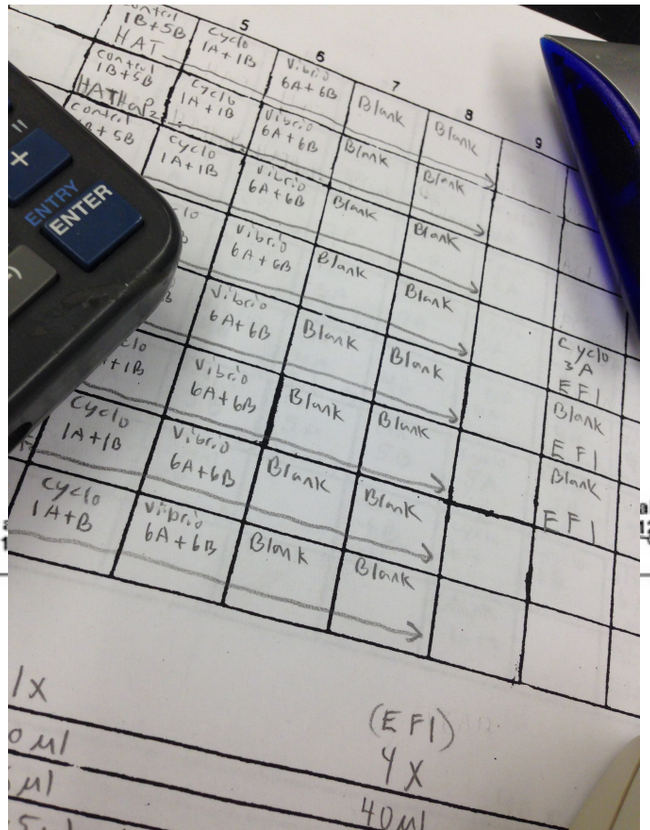


Background

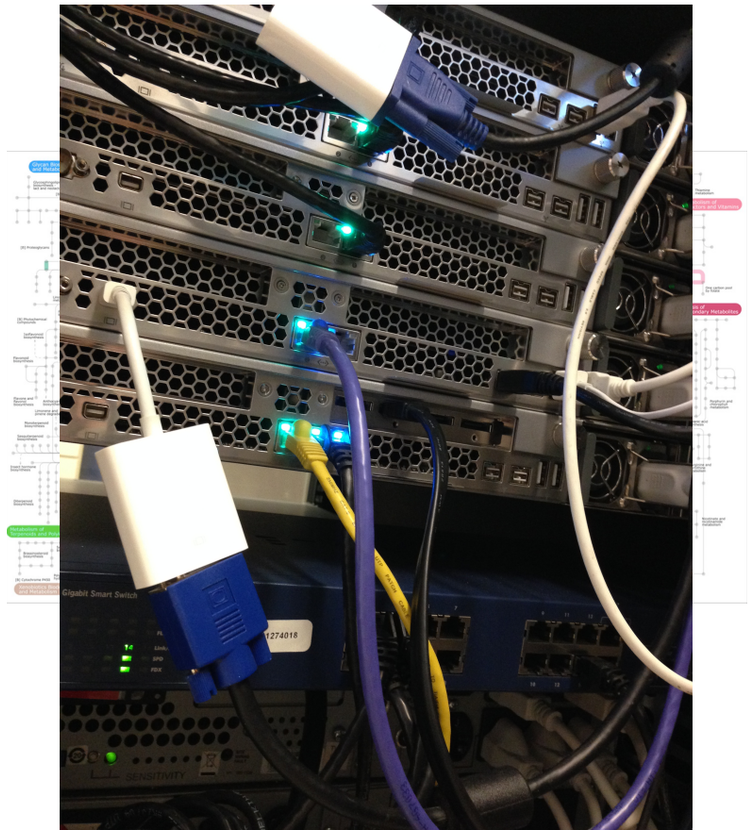
Physiology

How fundamental processes work in aquatic species

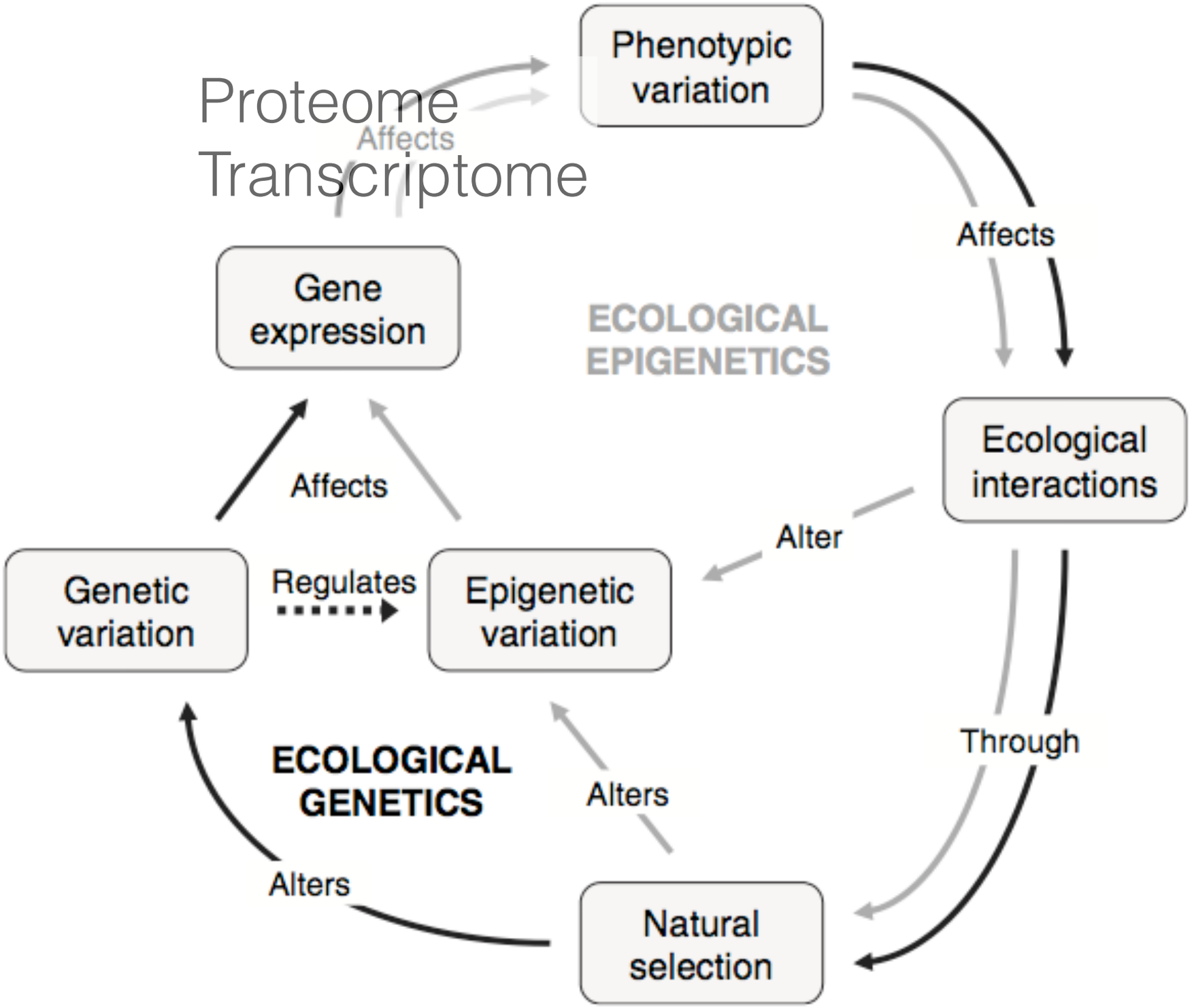




Changing Science



Proteome
Transcriptome



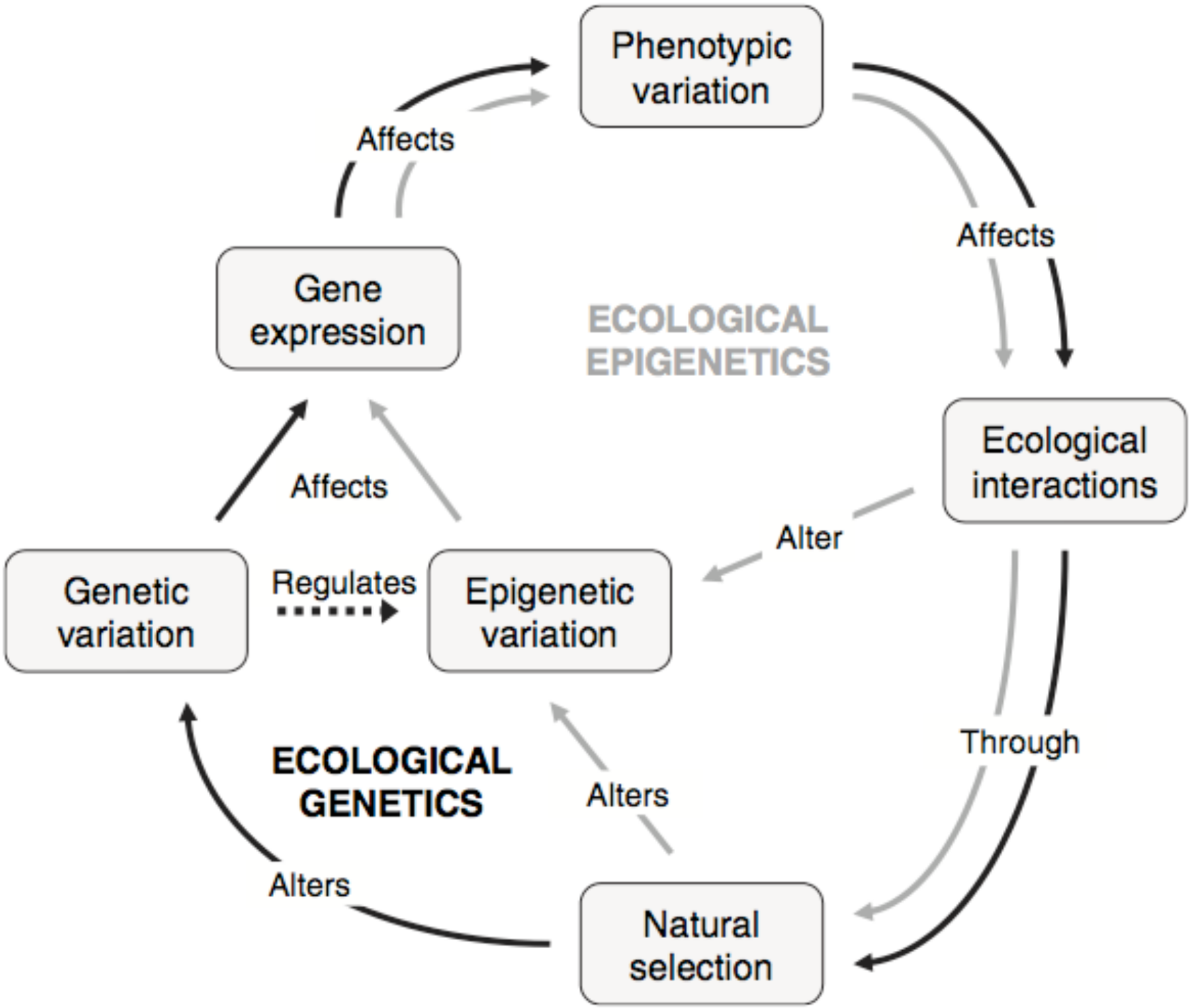
modified from

Ecology Letters, (2008) **11**: 106–115 doi: 10.1111/j.1461-0248.2007.01130.x

IDEA AND PERSPECTIVE

Epigenetics for ecologists

Oliver Bossdorf,[1]* Christina L. Richards[2] and Massimo Pigliucci[3]



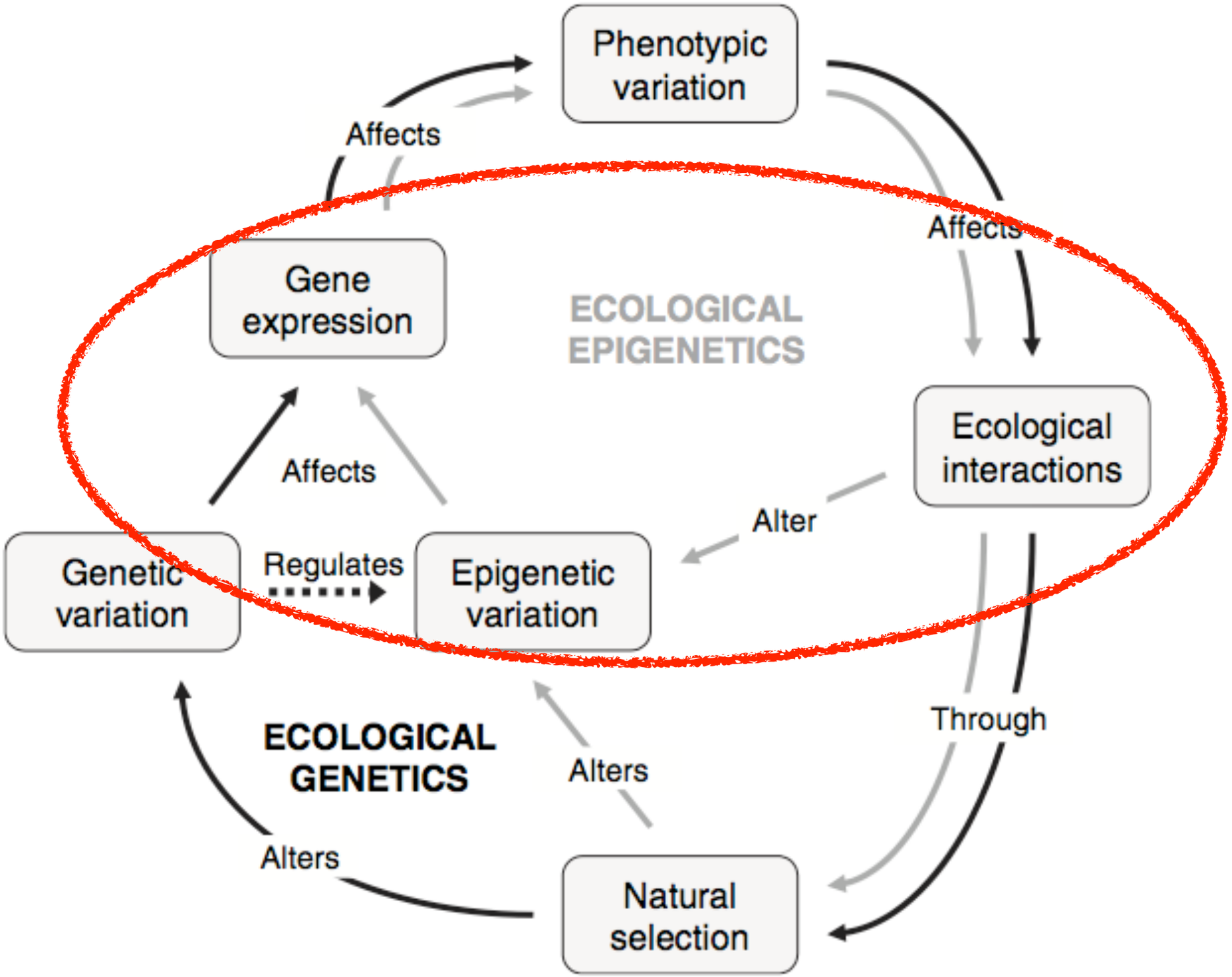
modified from

Ecology Letters, (2008) **11**: 106–115 doi: 10.1111/j.1461-0248.2007.01130.x

IDEA AND PERSPECTIVE

Epigenetics for ecologists

Oliver Bossdorf,[1]* Christina L. Richards[2] and Massimo Pigliucci[3]



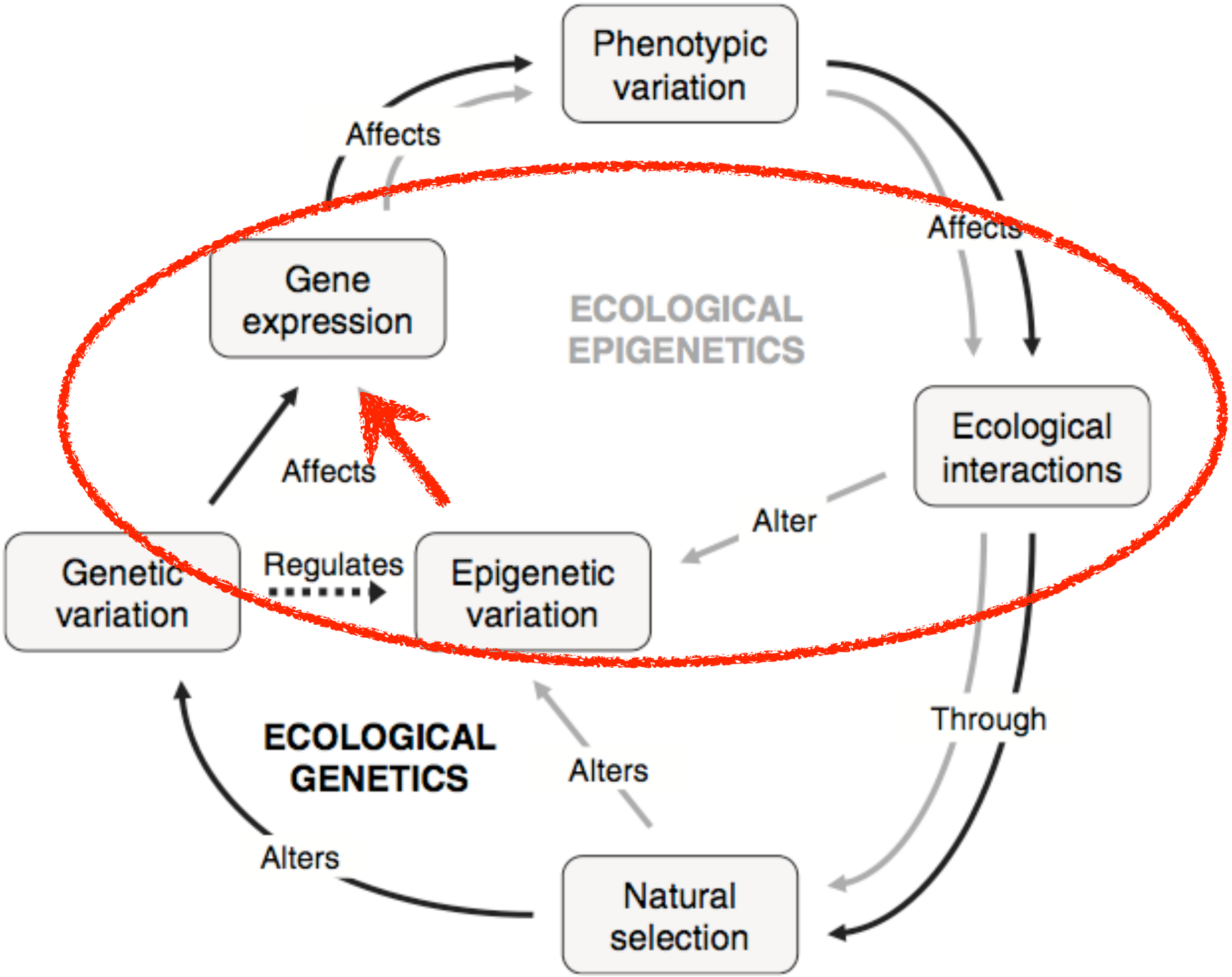
modified from

Ecology Letters, (2008) **11**: 106–115 doi: 10.1111/j.1461-0248.2007.01130.x

IDEA AND PERSPECTIVE

Epigenetics for ecologists

Oliver Bossdorf,[1]* Christina L. Richards[2] and Massimo Pigliucci[3]



modified from

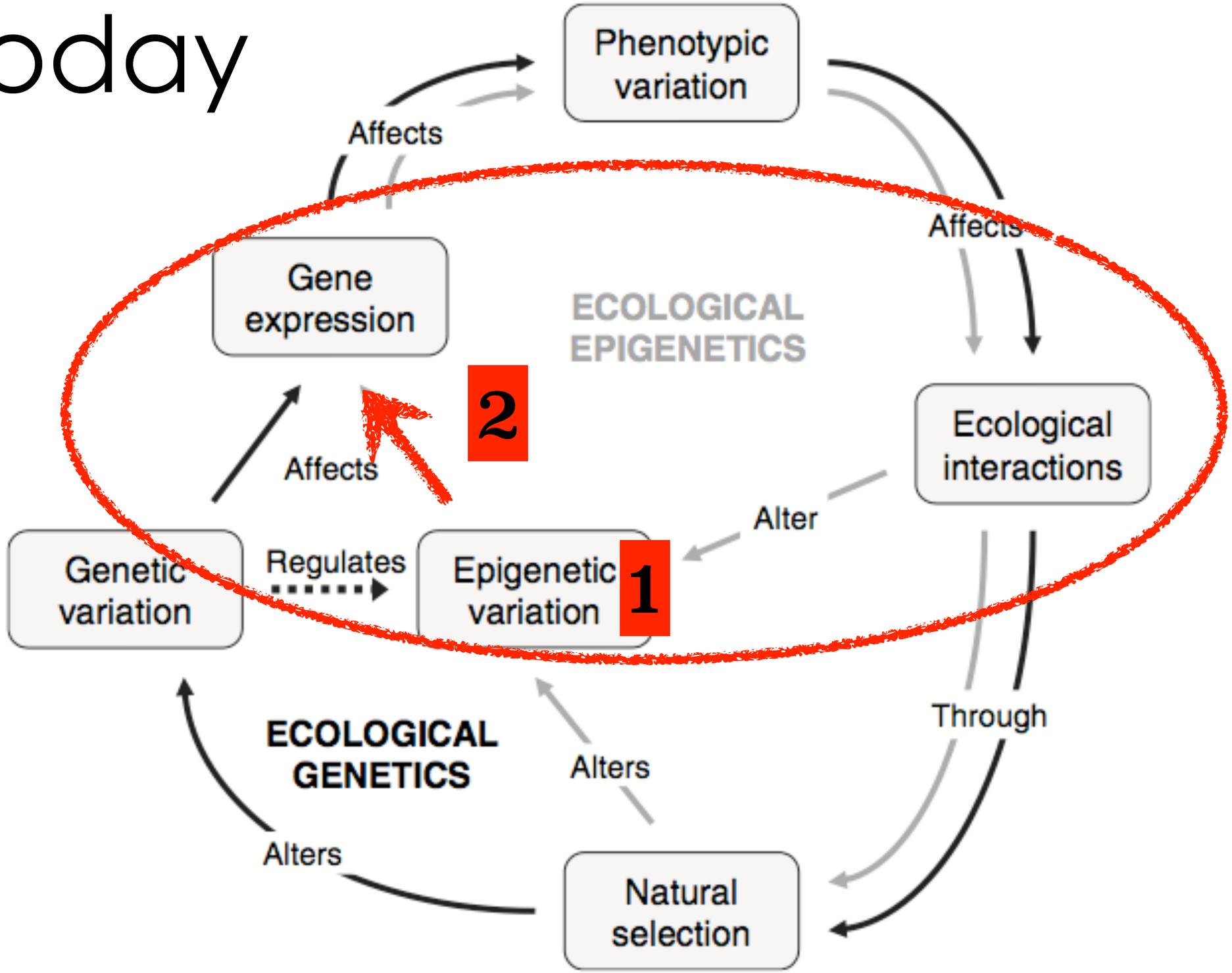
Ecology Letters, (2008) **11**: 106–115 doi: 10.1111/j.1461-0248.2007.01130.x

IDEA AND PERSPECTIVE

Epigenetics for ecologists

Oliver Bossdorf,[1]* Christina L. Richards[2] and Massimo Pigliucci[3]

Today



modified from

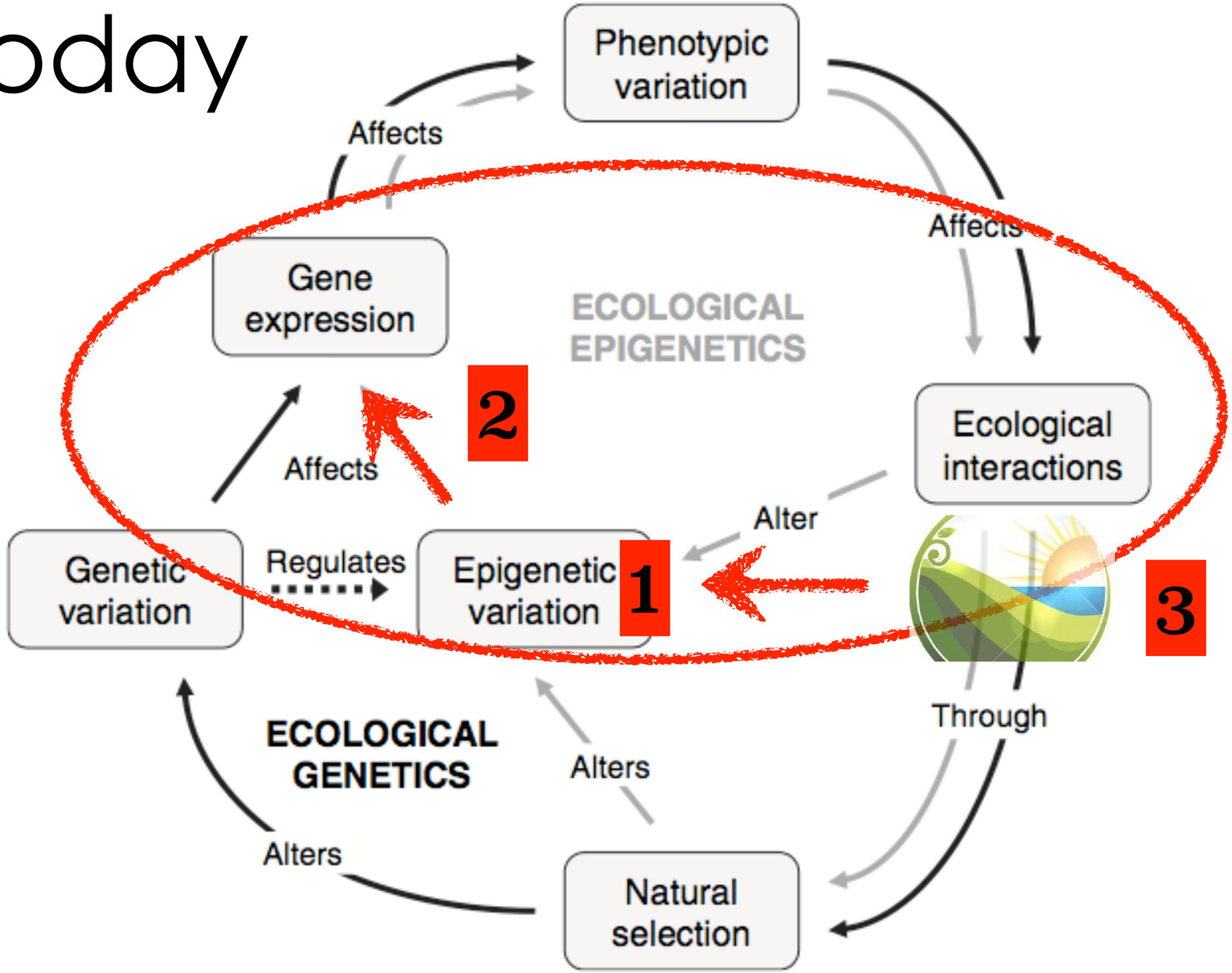
Ecology Letters, (2008) **11**: 106–115 doi: 10.1111/j.1461-0248.2007.01130.x

IDEA AND PERSPECTIVE

Epigenetics for ecologists

Oliver Bossdorf,[1]* Christina L. Richards[2] and Massimo Pigliucci[3]

Today



modified from

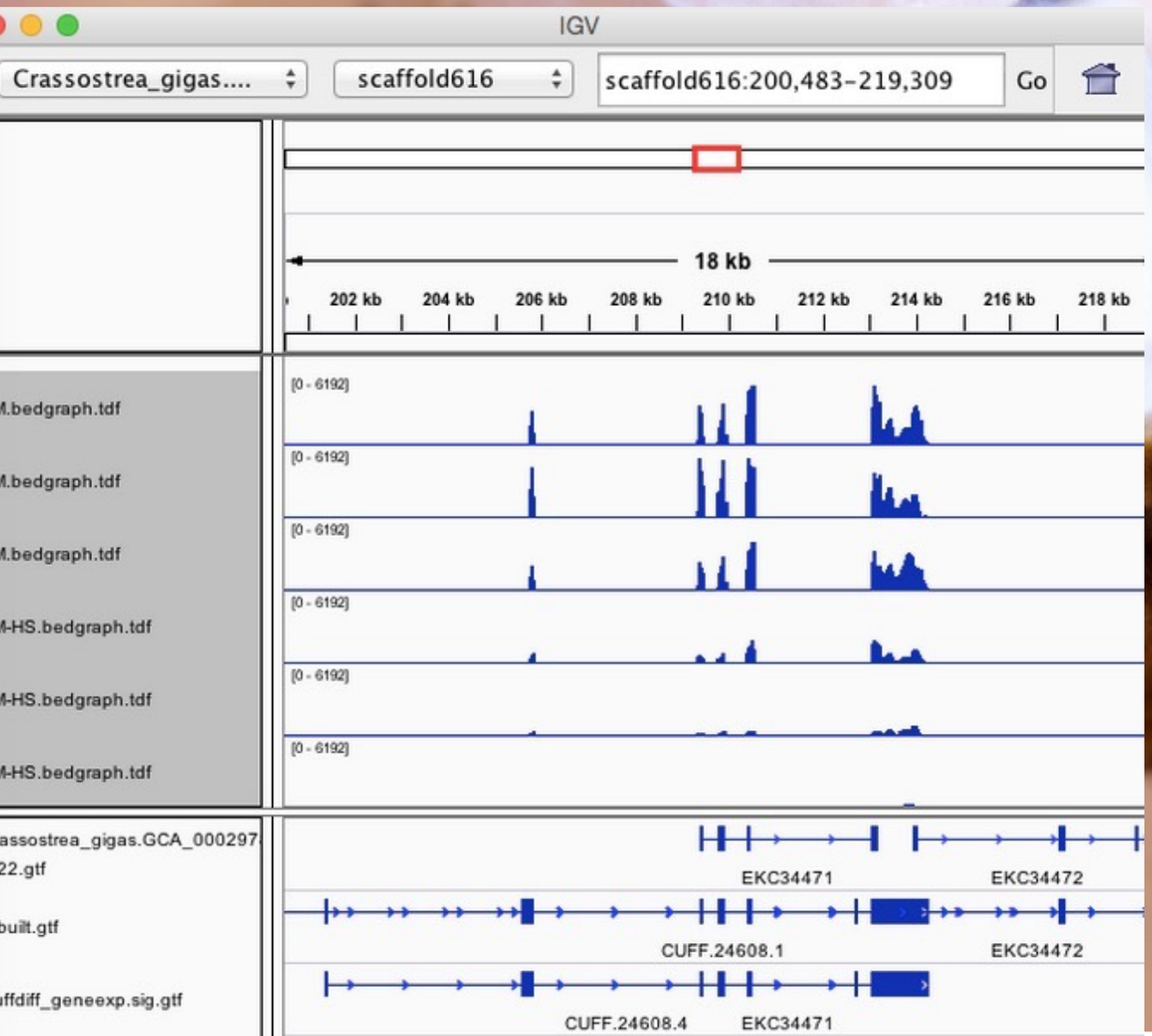
Ecology Letters, (2008) **11**: 106–115 doi: 10.1111/j.1461-0248.2007.01130.x

IDEA AND PERSPECTIVE

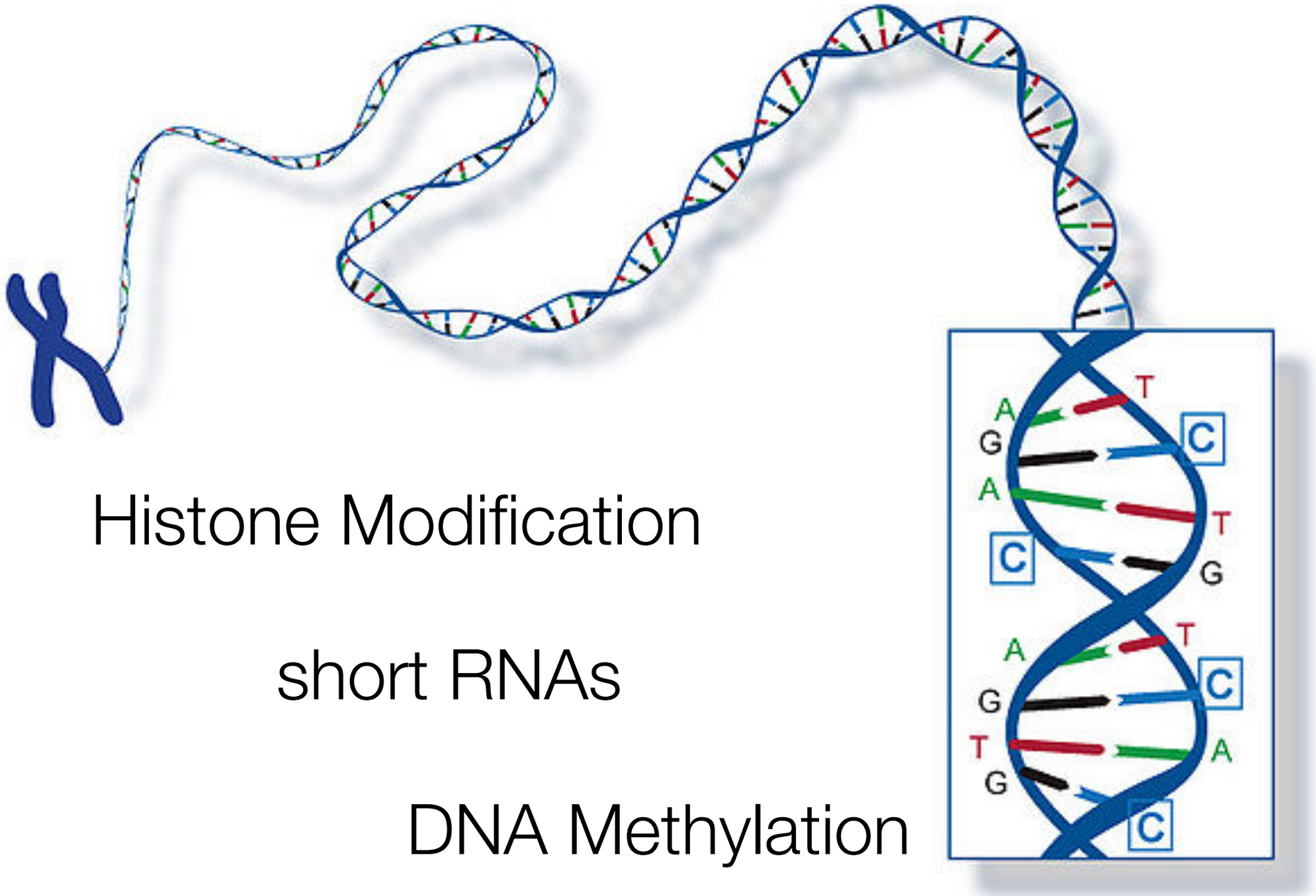
Epigenetics for ecologists

Oliver Bossdorf,[1]* Christina L. Richards[2] and Massimo Pigliucci[3]

Genome Resources



Epigenetics



Histone Modification

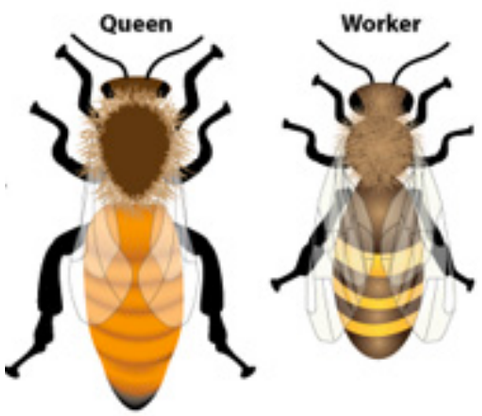
short RNAs

DNA Methylation

Epigenetics

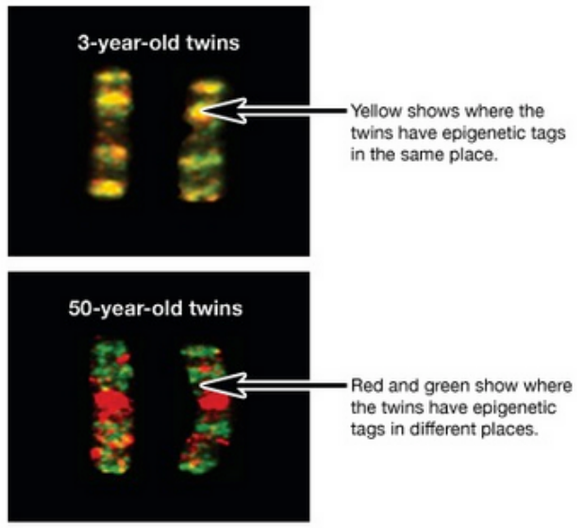
Learn.Genetics
GENETIC SCIENCE LEARNING CENTER

Queen Bee Larvae: Queens are raised in specially constructed cells called "queen cups," which are filled with royal jelly.

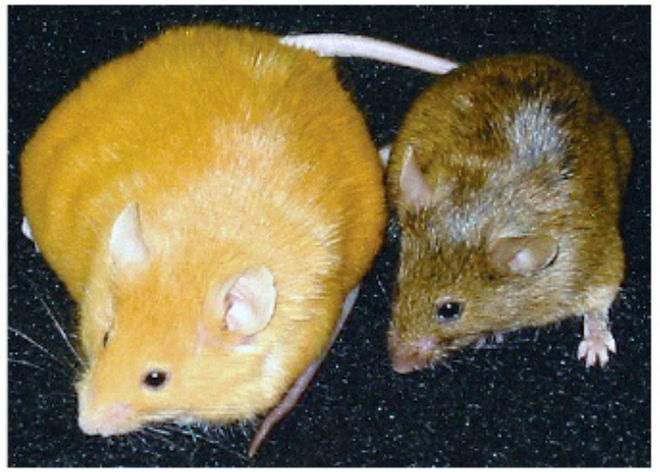


Nature AND Nurture

Chromosome 3 Pairs
3-year old twins vs. 50-year-old twins



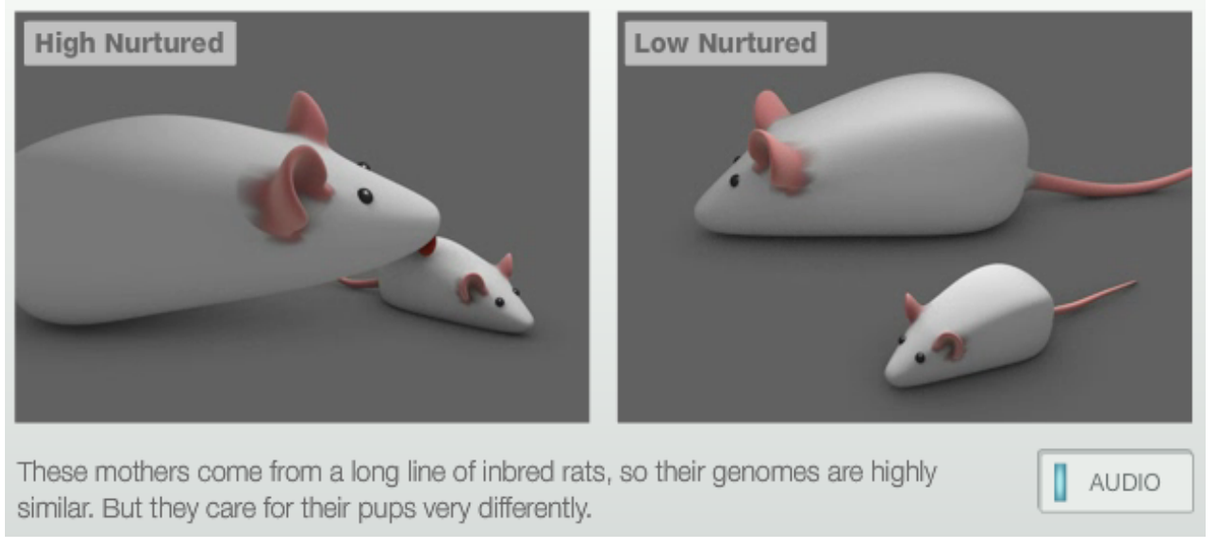
These Two Mice are Genetically Identical and the Same Age



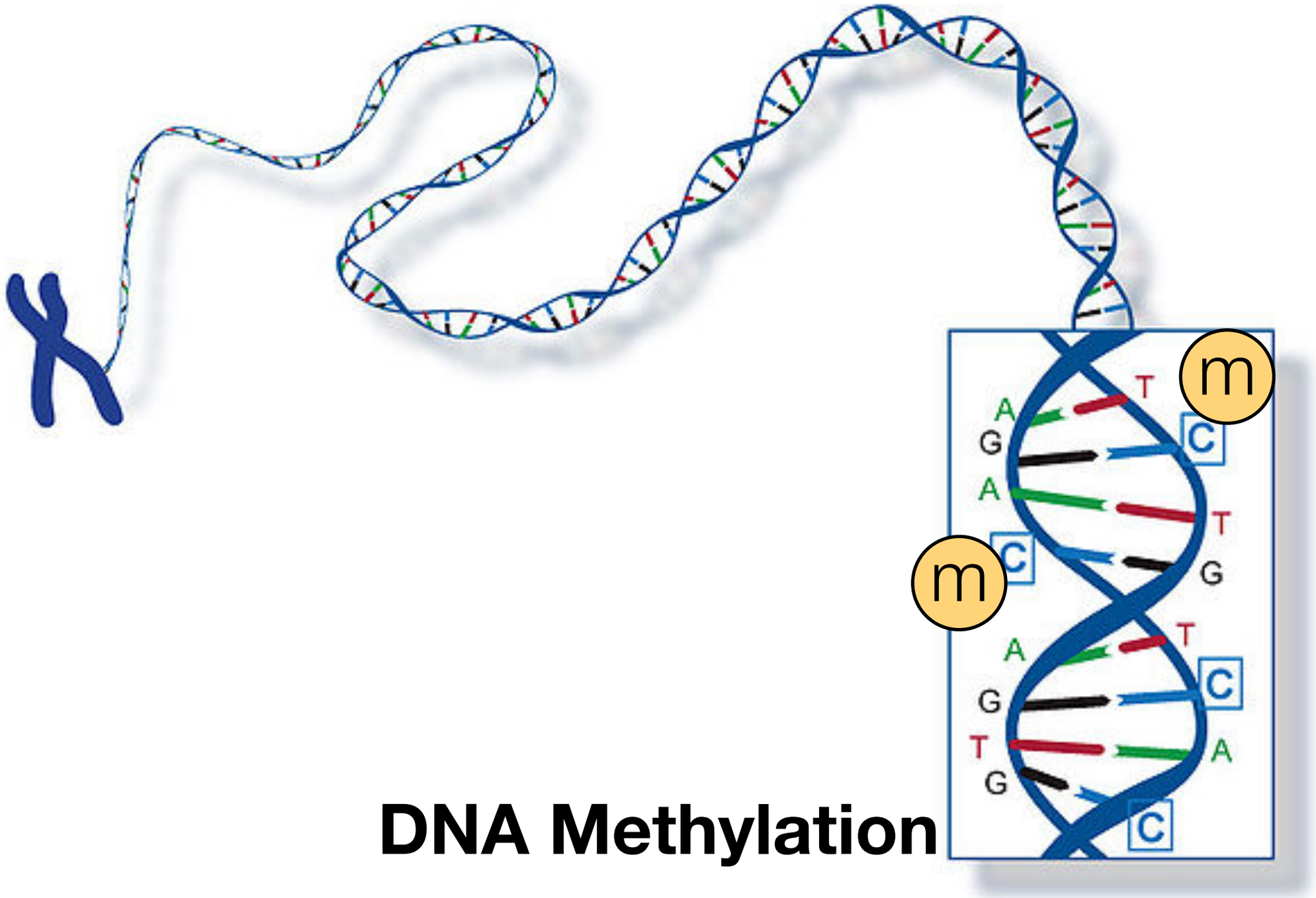
While pregnant, both of their mothers were fed Bisphenol A (BPA) but DIFFERENT DIETS:

The mother of this mouse received a **normal mouse diet**

The mother of this mouse received a diet **supplemented** with choline, folic acid, betaine and vitamin B12



http://learn.genetics.utah.edu



DNA Methylation

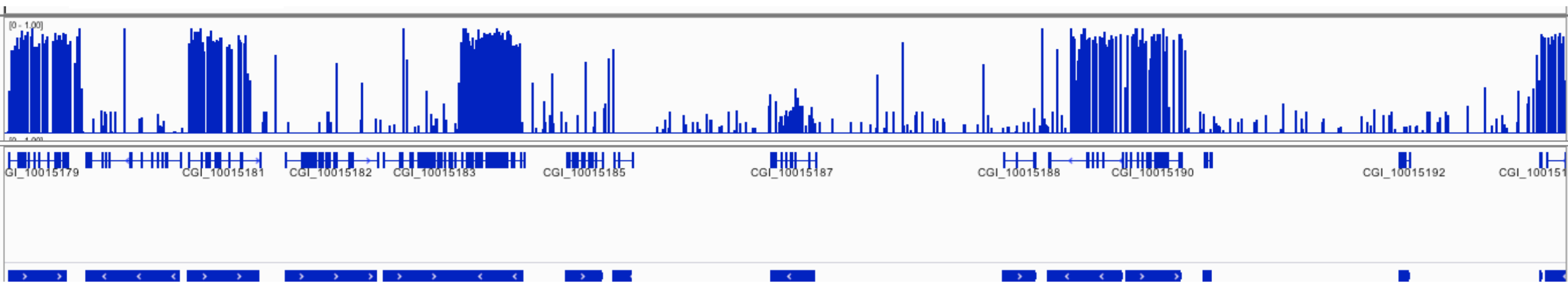
Non-Vertebrates?

landscape and function is very different than what is observed in vertebrates

Absent in several model organisms

Oysters?

Epigenetic variation 1

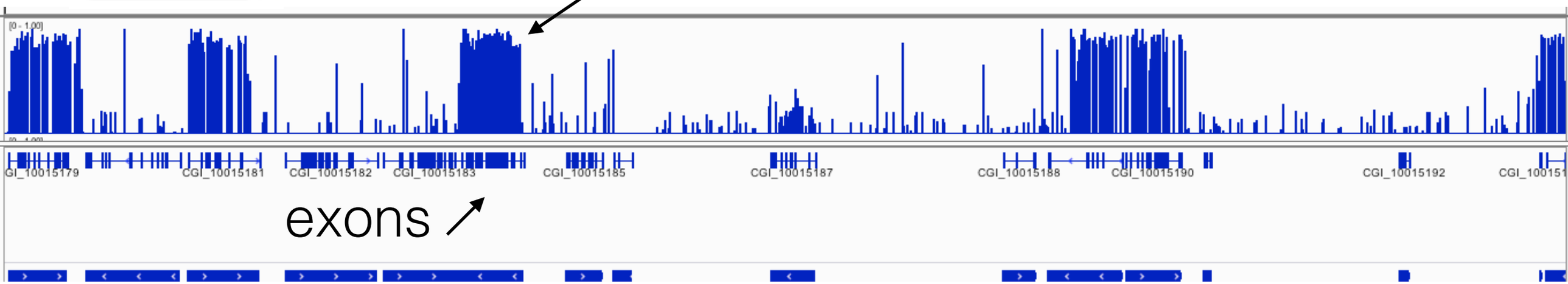


mosaic

associated with gene bodies

Epigenetic variation 1

DNA methylation level (0-100%) @ cytosines



exons

gene

mosaic

associated with gene bodies

Epigenetic variation 1

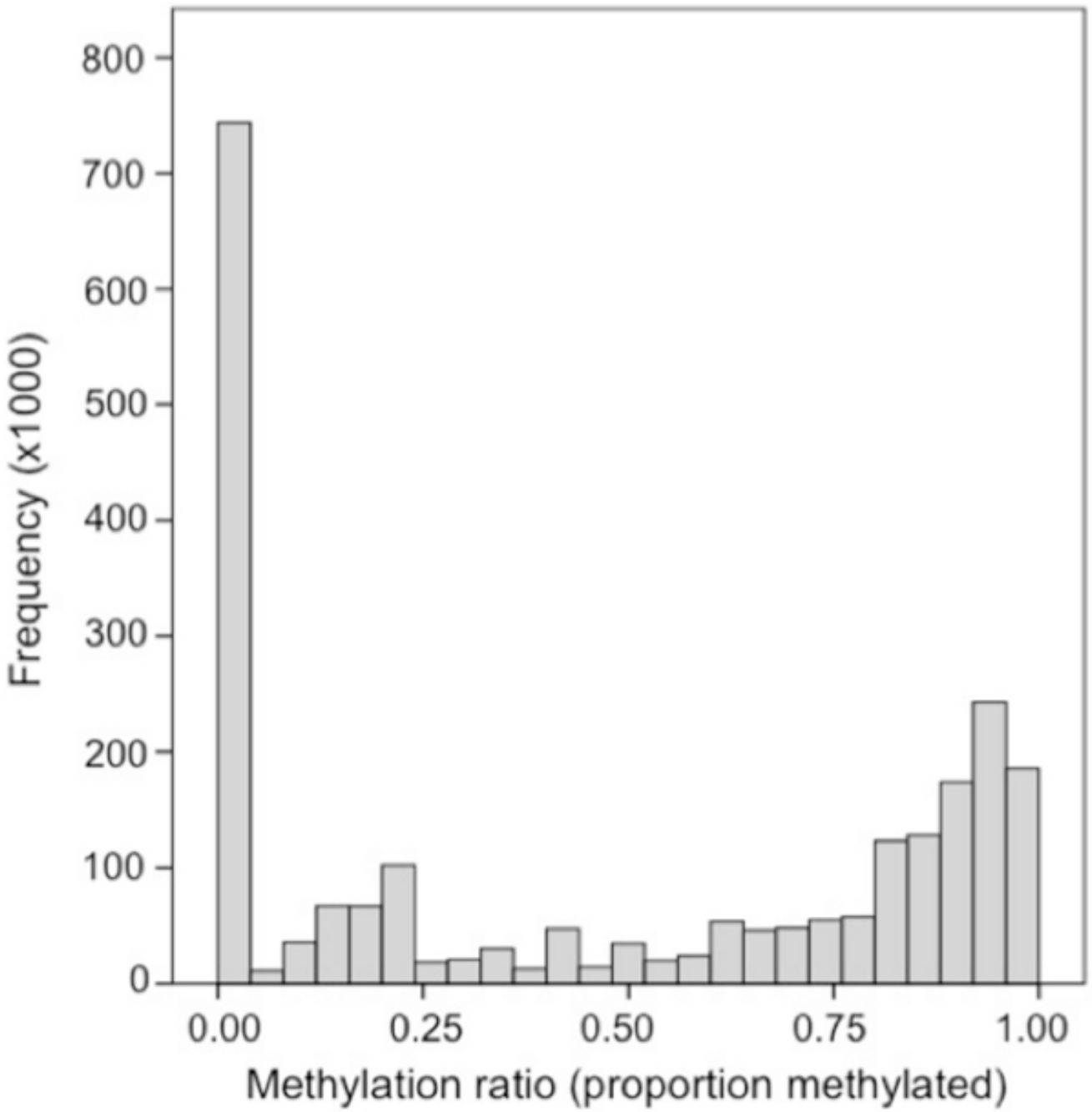
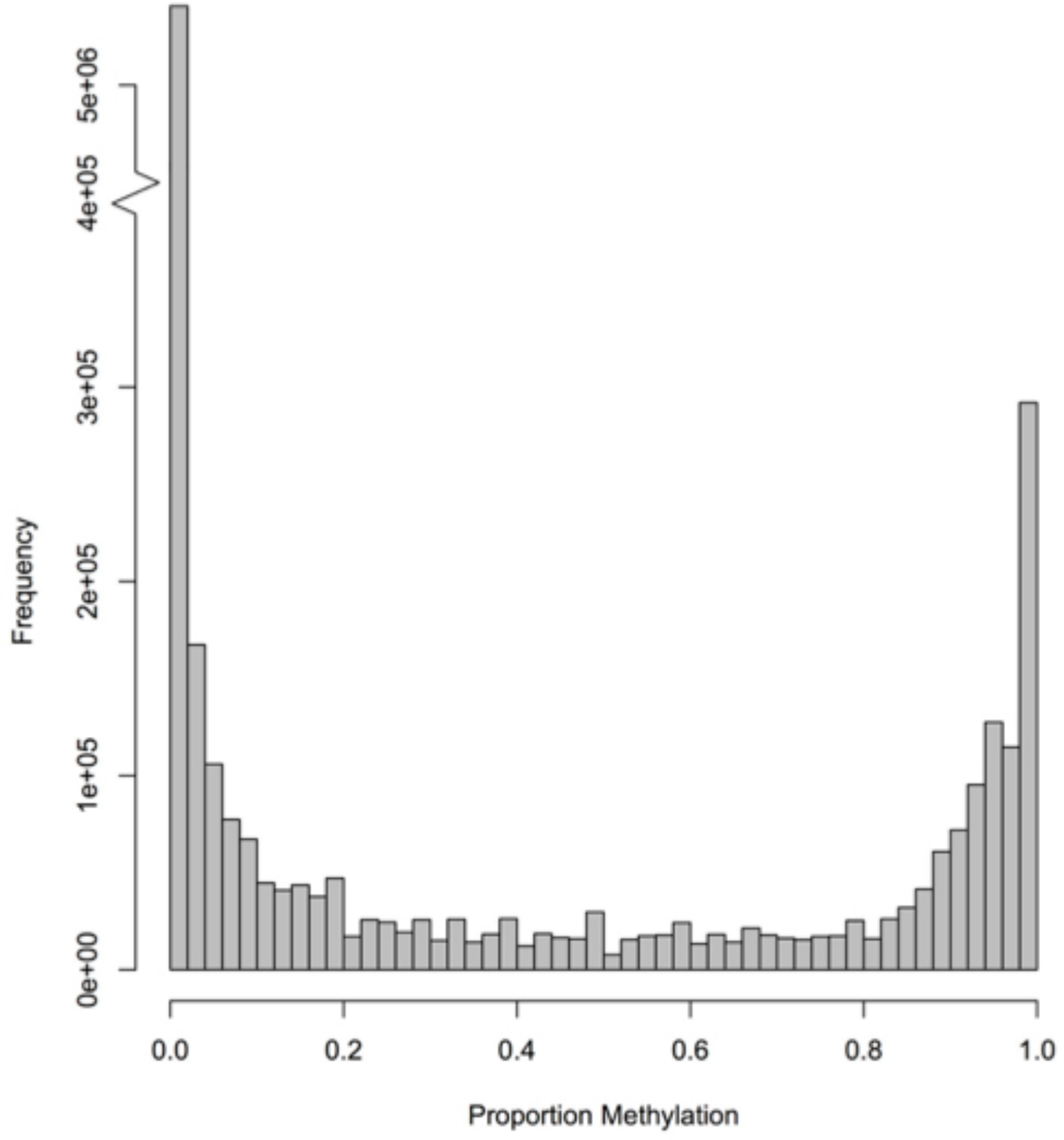


Figure 1 Frequency distribution of methylation ratios for CpG dinucleotides in oyster gill tissue. A total of 2,625,745 CpG dinucleotides with ≥5× coverage are represented.

Epigenetic variation 1



frontiers in PHYSIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 17 June 2014
doi: 10.3389/fphys.2014.00224

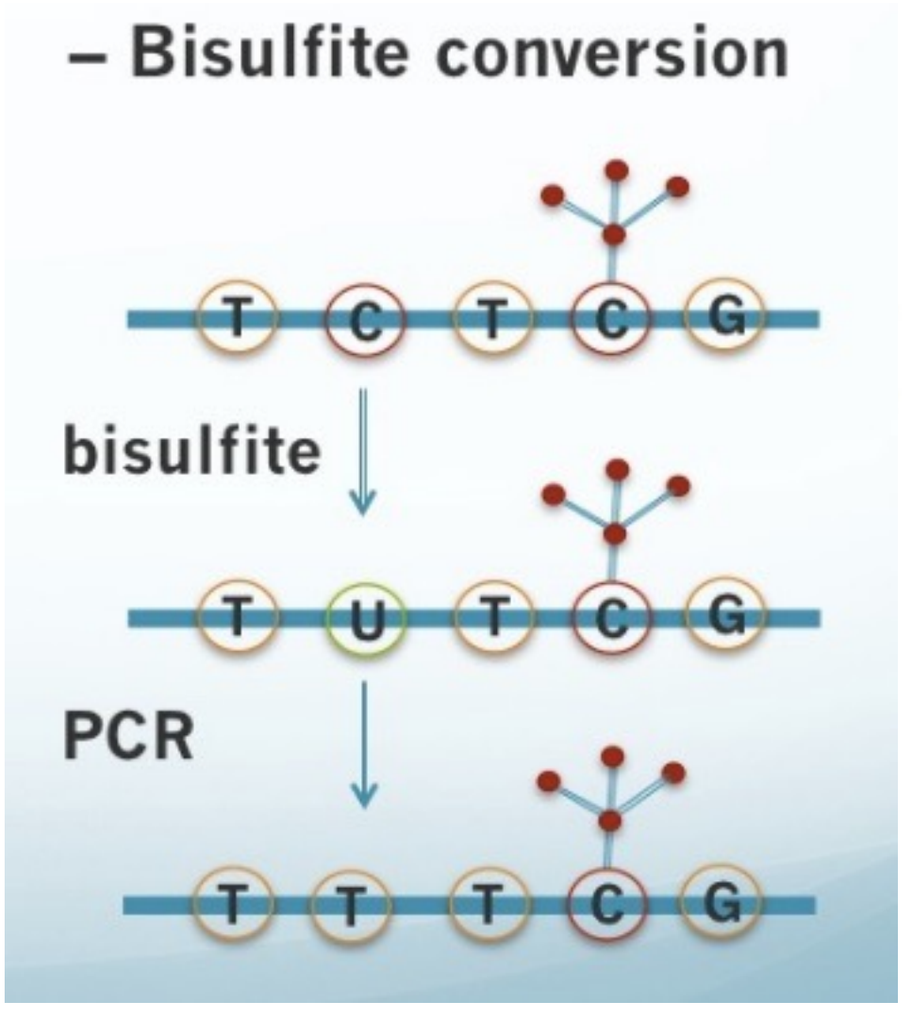
Genome-wide profiling of DNA methylation and gene expression in *Crassostrea gigas* male gametes

Claire E. Olson and Steven B. Roberts*

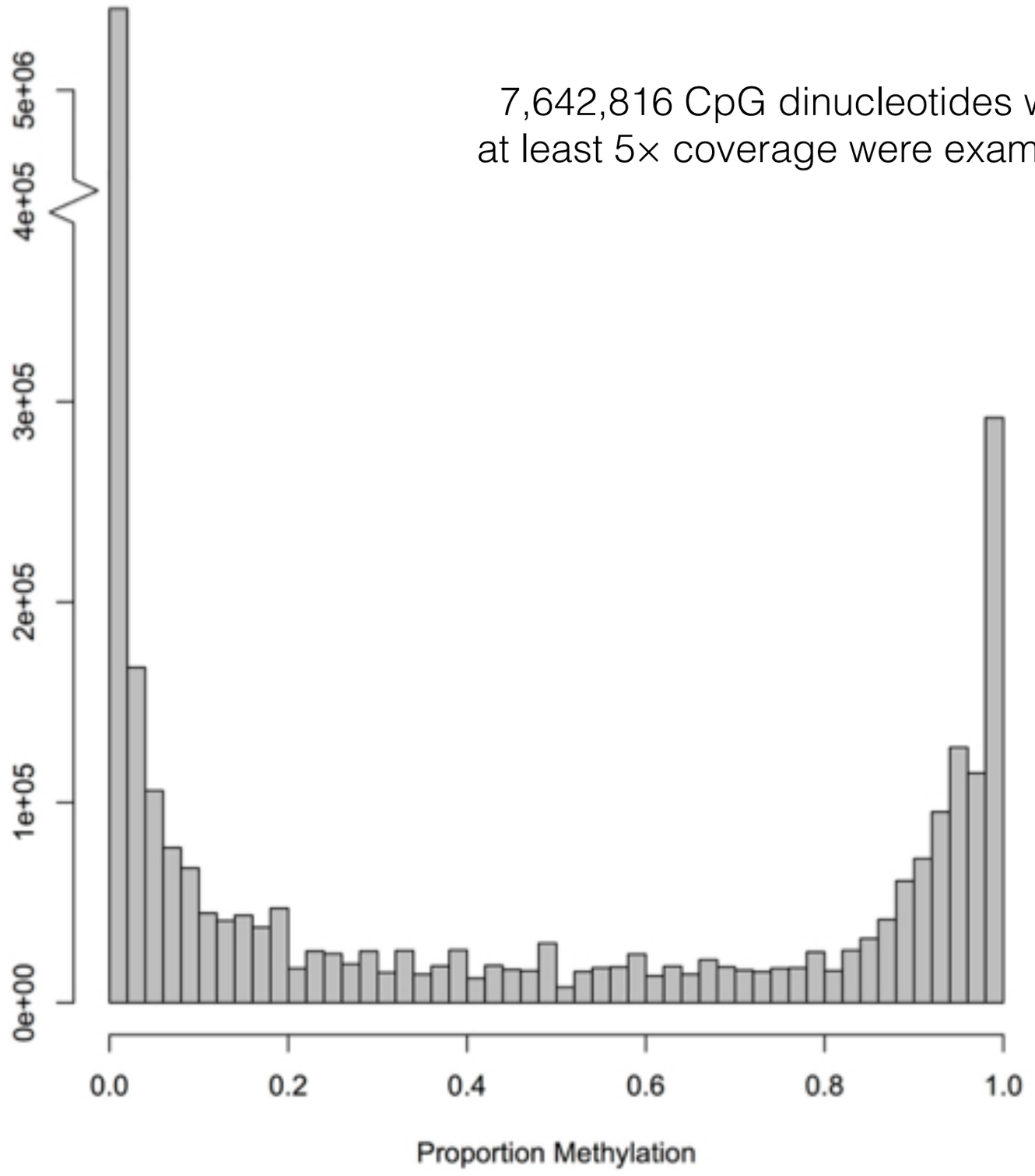
School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

Epigenetic variation **1**

– Bisulfite conversion



7,642,816 CpG dinucleotides with at least 5× coverage were examined



frontiers in PHYSIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 17 June 2014
doi: 10.3389/fphys.2014.00224

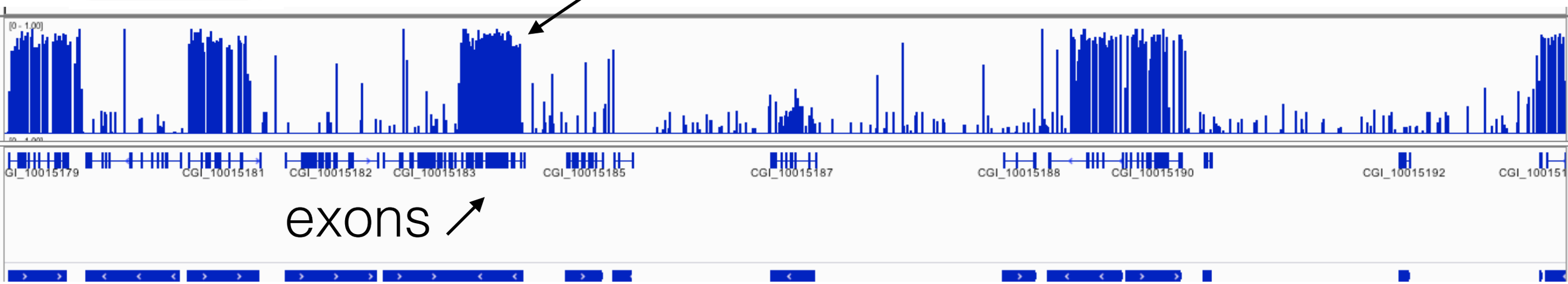
Genome-wide profiling of DNA methylation and gene expression in *Crassostrea gigas* male gametes

Claire E. Olson and Steven B. Roberts*

School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

Epigenetic variation 1

DNA methylation level (0-100%) @ cytosines

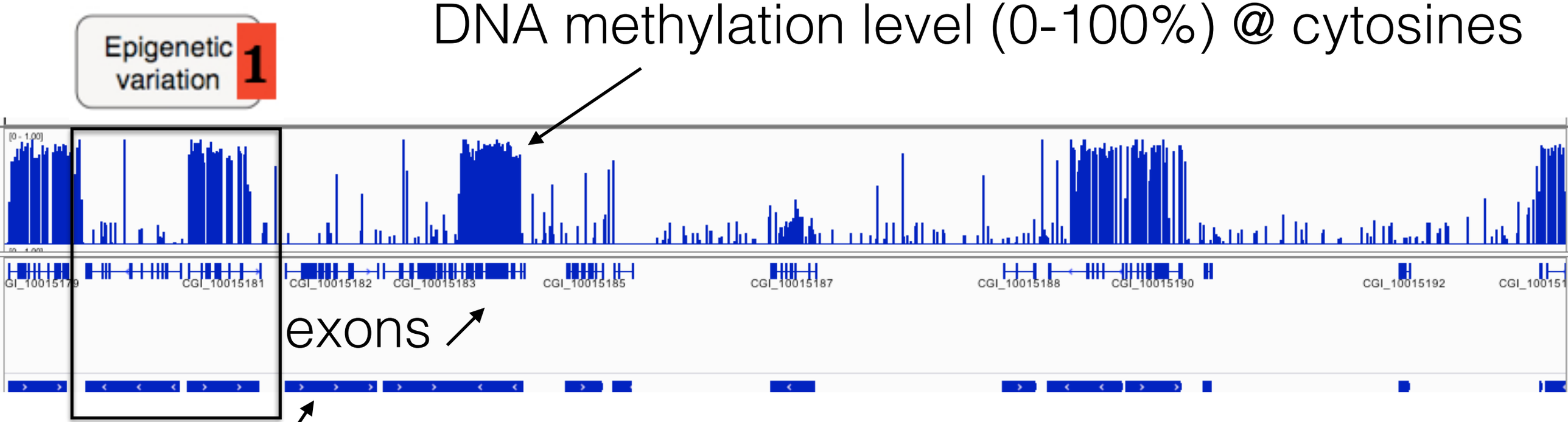


exons

gene

mosaic

associated with gene bodies



mosaic

Why are only a subset of genes methylated?

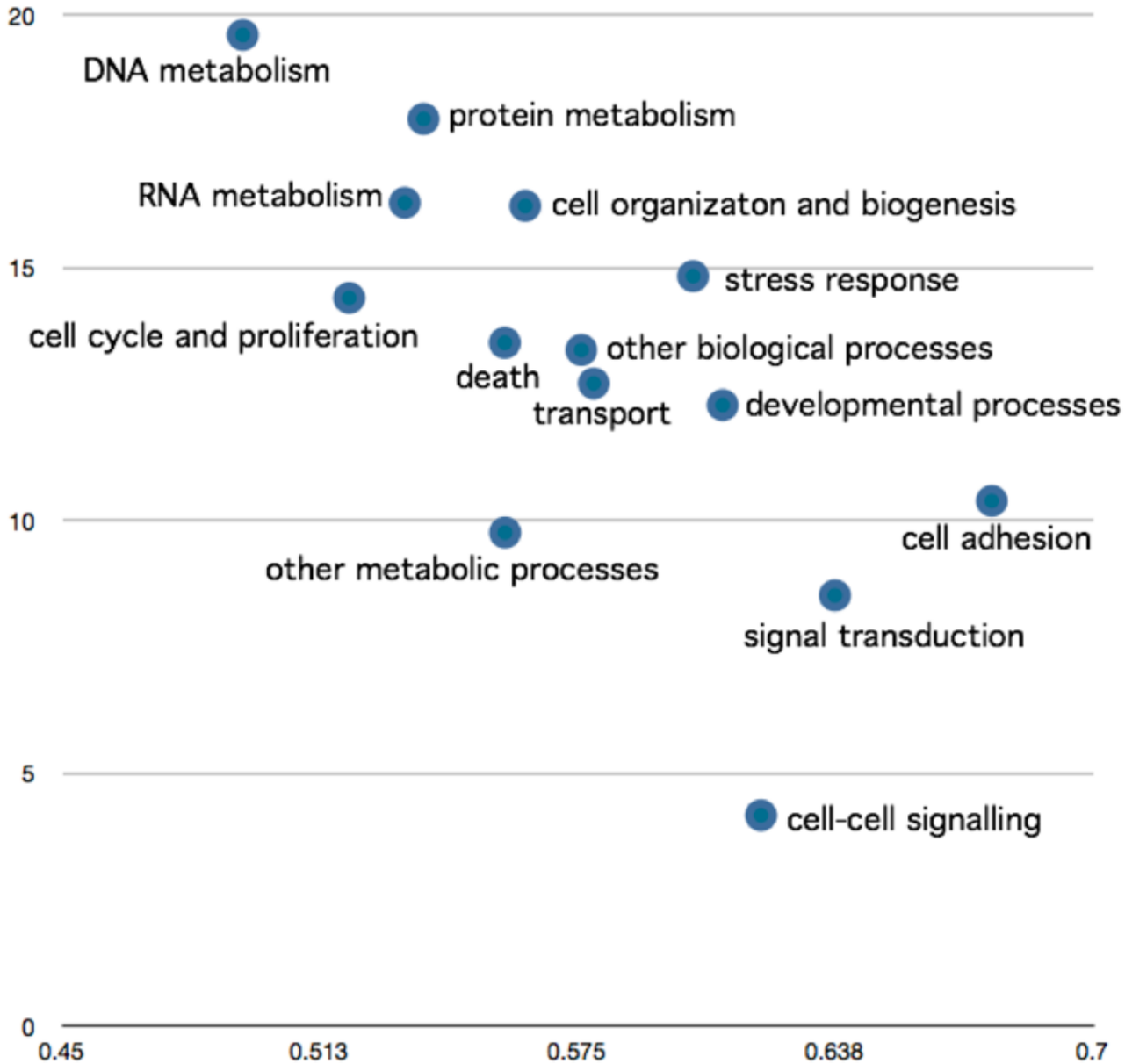
associated with gene bodies

Epigenetic variation 1

Measured degree of DNA methylation

MBD-Seq library

Enrichment level in



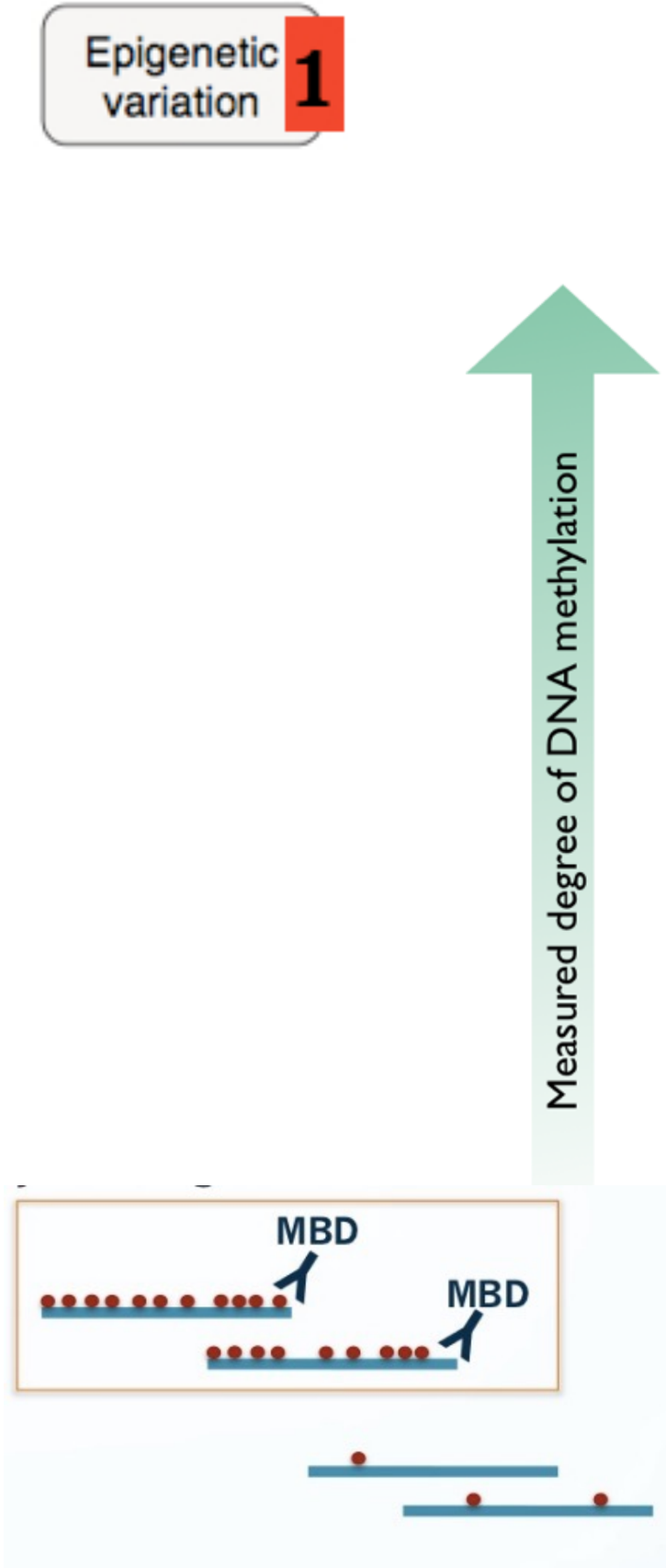
***in silico* analysis**

(modified from Gavery and Roberts 2010)

Predicted degree of DNA methylation

Roberts and Gavery 2012

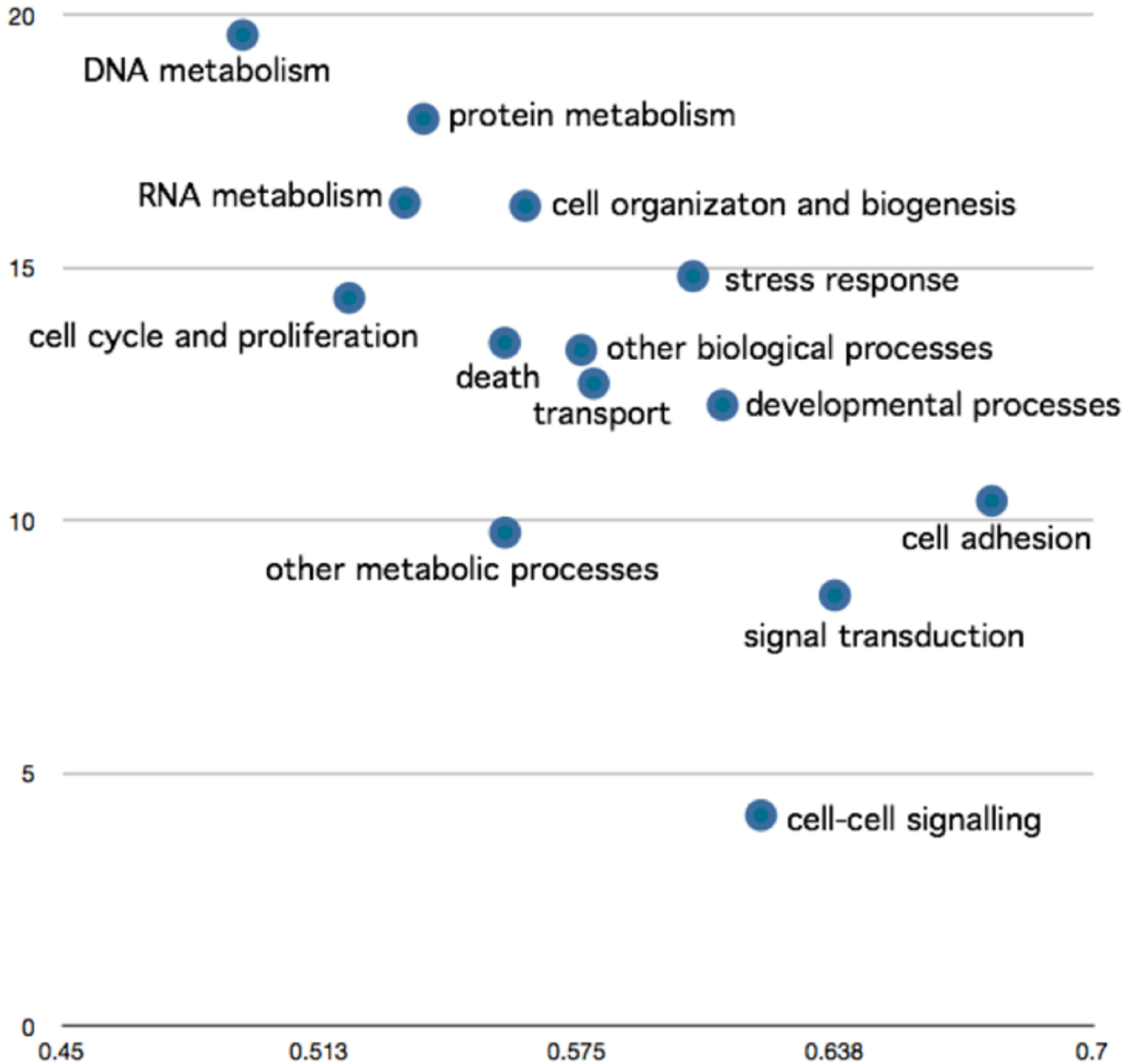
Epigenetic variation 1



Measured degree of DNA methylation

MBD-Seq library

Enrichment level in



***in silico* analysis**

(modified from Gavery and Roberts 2010)

Predicted degree of DNA methylation

Roberts and Gavery 2012

in silico approach

- Principle:
  - Methylated cytosines are highly mutable

$$\overset{m}{C} \rightarrow T$$

  - Methylated regions of DNA are depleted of CpG dinucleotides over evolutionary time (CpG to TpG)

$$\text{CpG O/E} = \frac{\text{CpG observed}}{\text{CpG expected}}$$

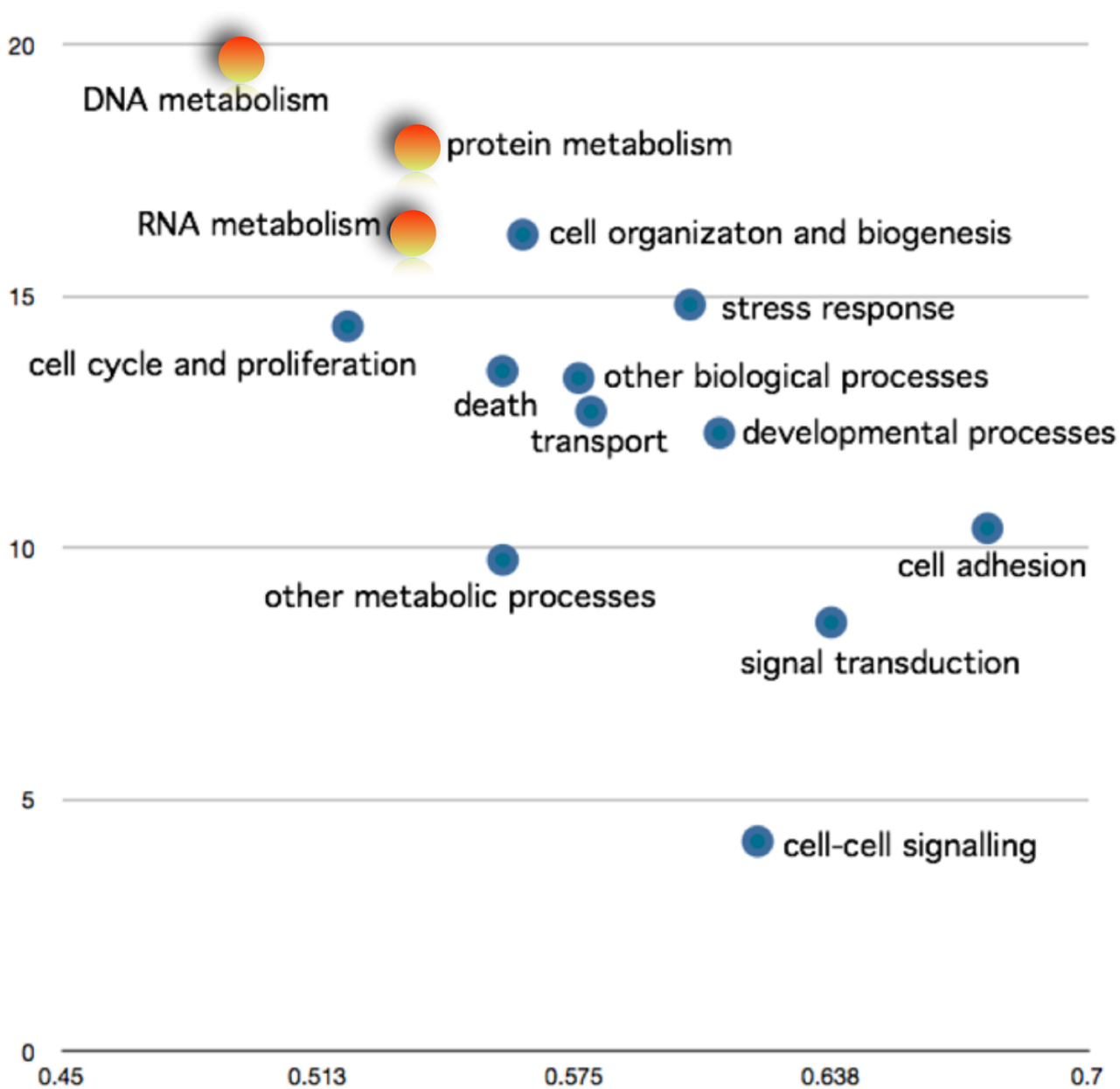
high = unmethylated

Epigenetic variation 1

Measured degree of DNA methylation

MBD-Seq library

Enrichment level in



***in silico* analysis**

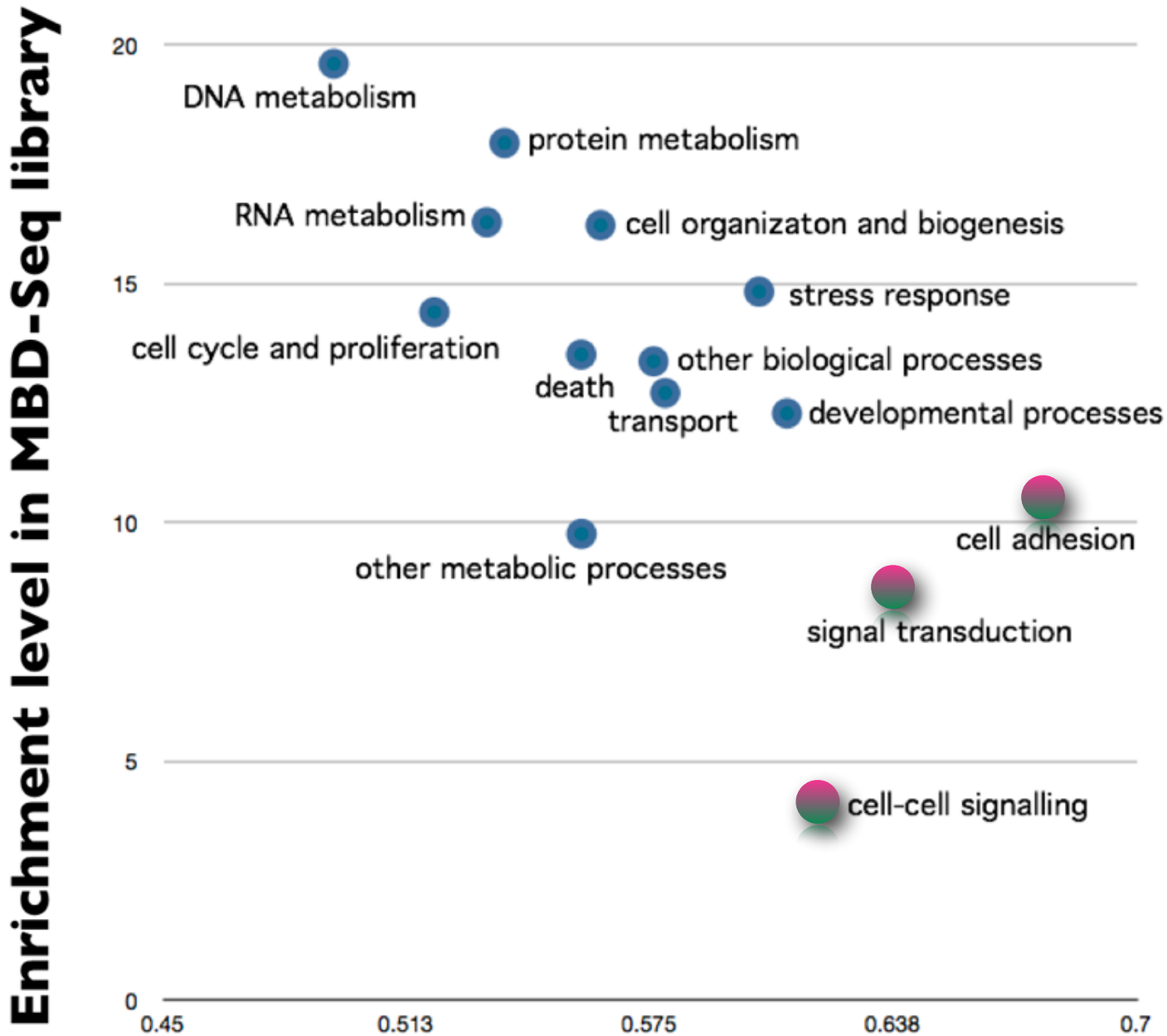
(modified from Gavery and Roberts 2010)

Predicted degree of DNA methylation

Roberts and Gavery 2012

Epigenetic variation 1

Measured degree of DNA methylation



***in silico* analysis**

(modified from Gavery and Roberts 2010)

Predicted degree of DNA methylation

Roberts and Gavery 2012

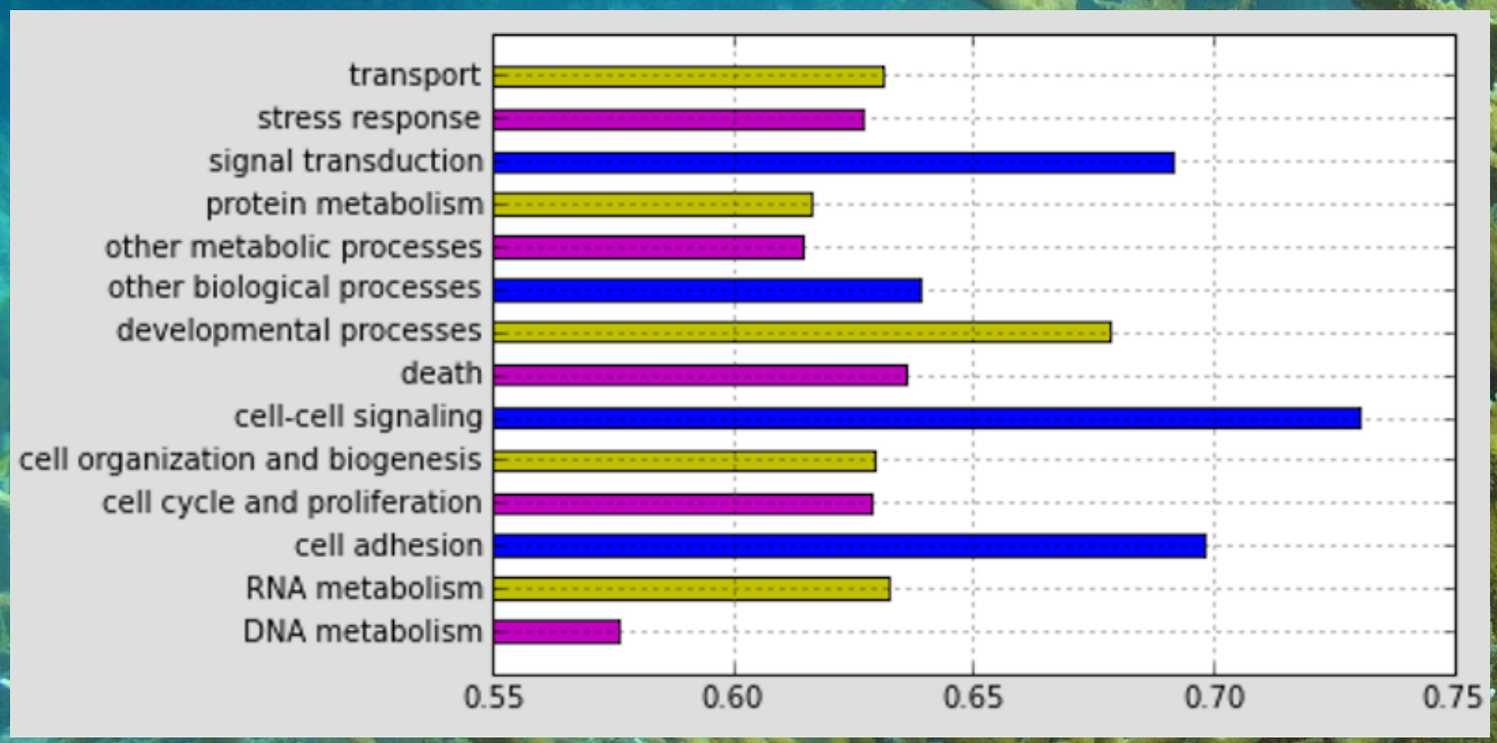
Jay Dimond

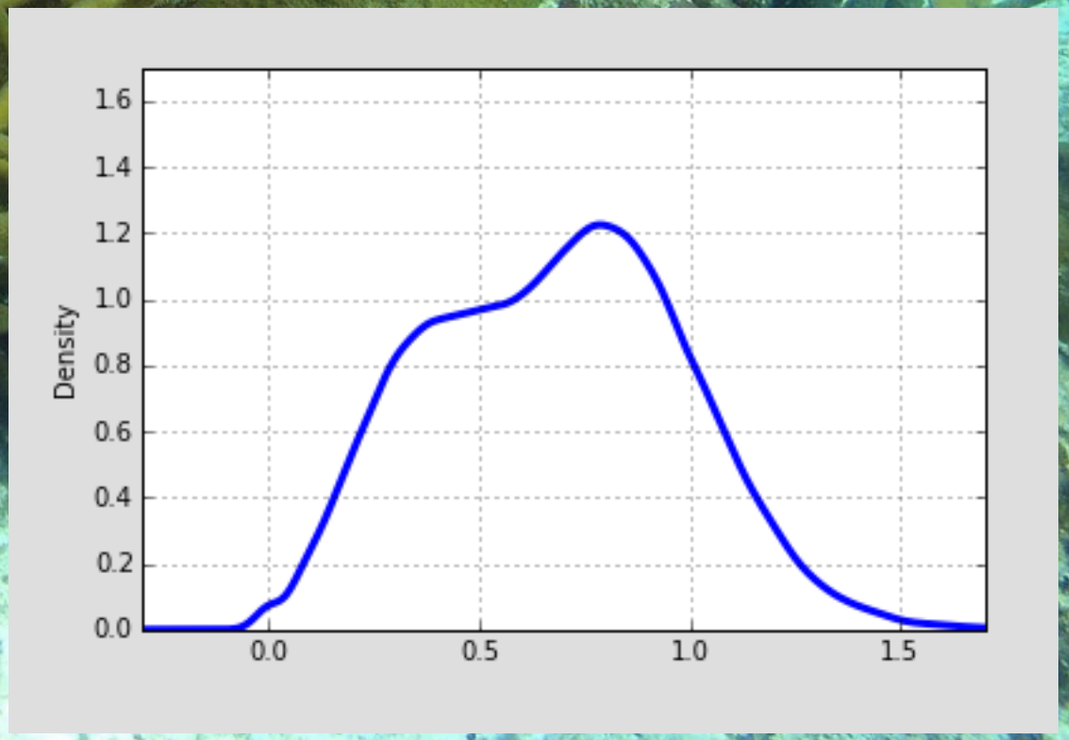
Acropora palmata

Acropora cervicornis

Jay Dimond

Acropora palmata



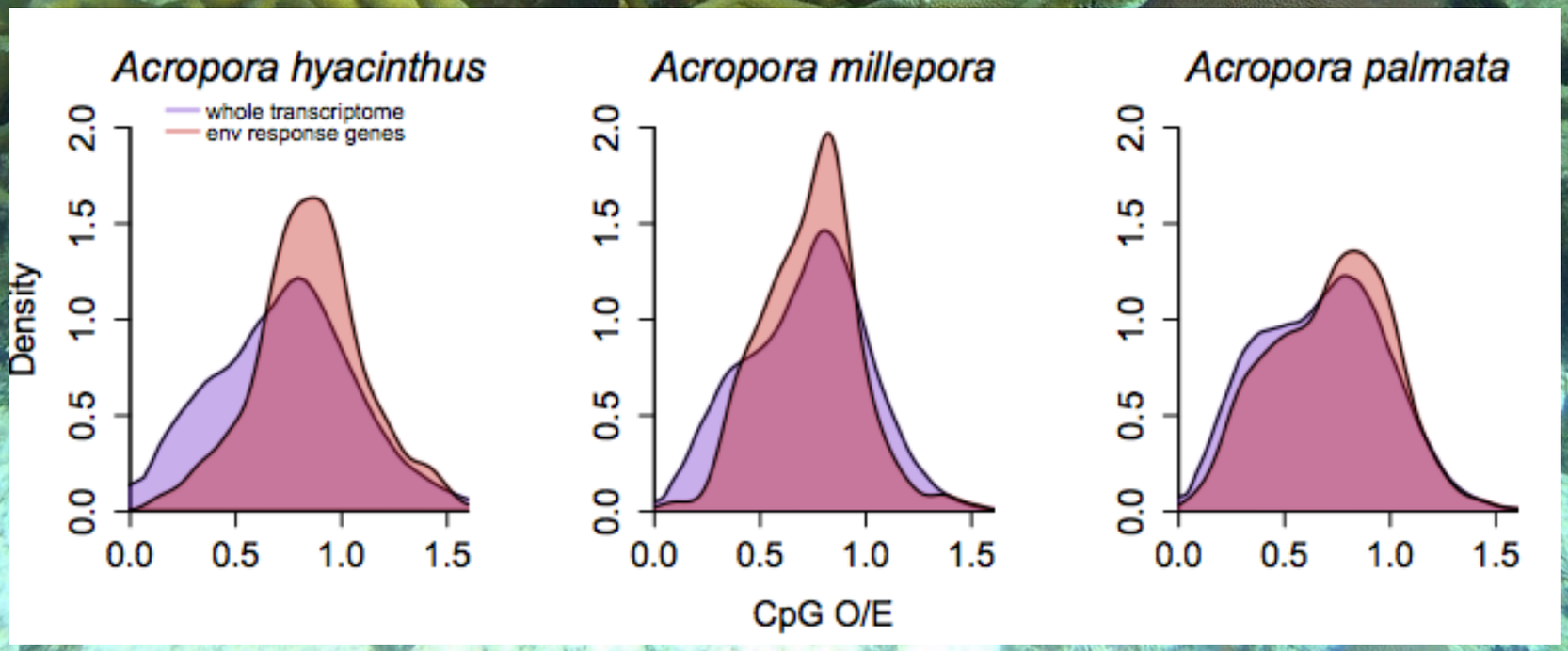


Jay Dimond

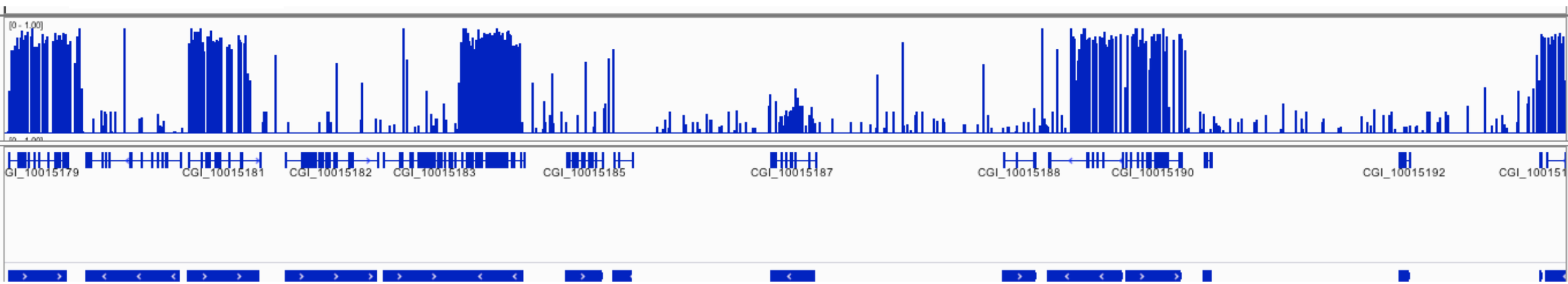
figshare

Germline DNA methylation in five coral transcriptomes

Acropora palmata



Epigenetic variation 1



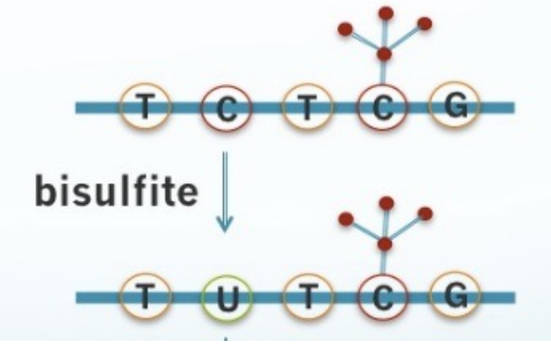
mosaic

associated with gene bodies
based on gene function

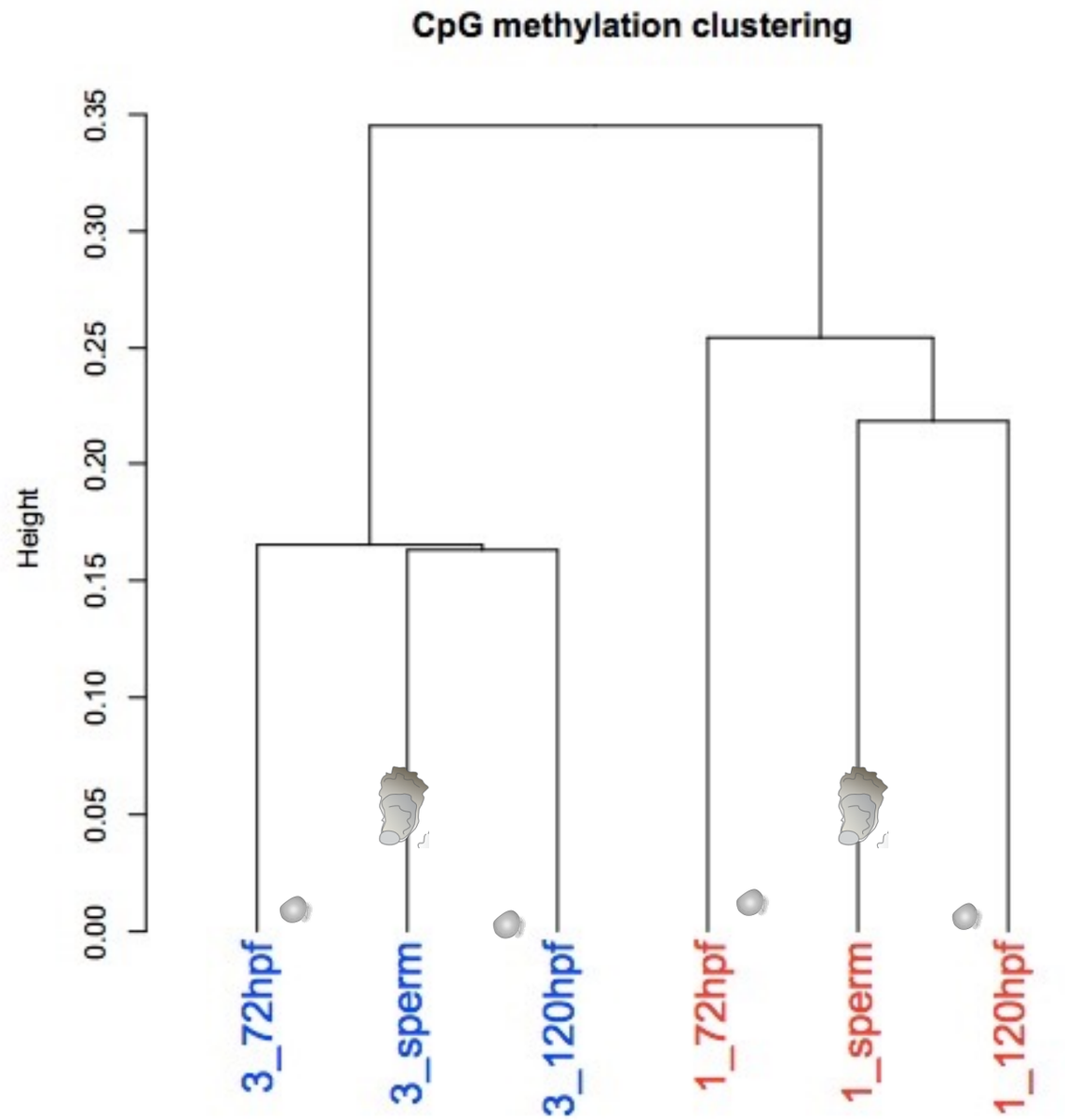
explanation?

Epigenetic variation 1

Family and Developmental Variation



Sperm &
Larvae
(72h & 120h)



Cold Spring Harbor Laboratory

bioRxiv beta
THE PREPRINT SERVER FOR BIOLOGY

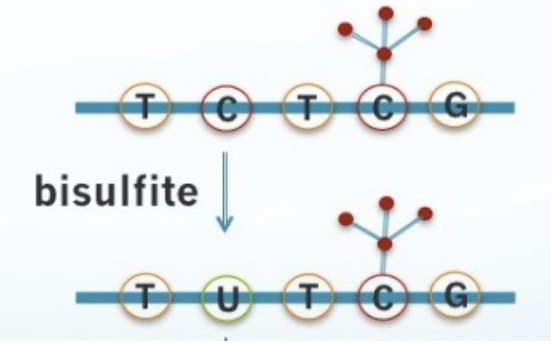
New Results

Indication of family-specific DNA methylation patterns in developing oysters

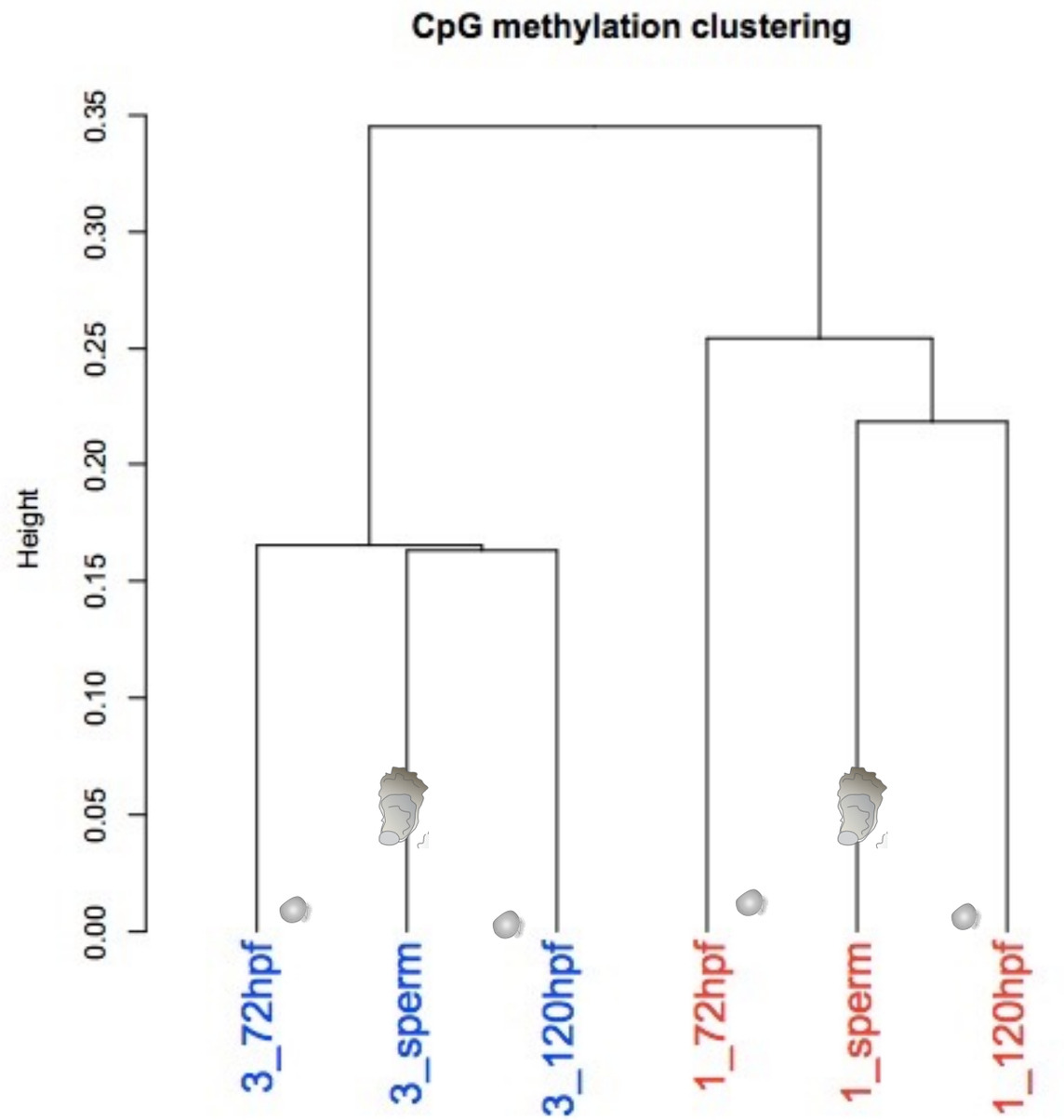
Claire E. Olson , Steven B. Roberts
doi: http://dx.doi.org/10.1101/012831

Epigenetic variation 1

Family and Developmental Variation



Inheritance



Cold Spring Harbor Laboratory

bioRxiv beta
THE PREPRINT SERVER FOR BIOLOGY

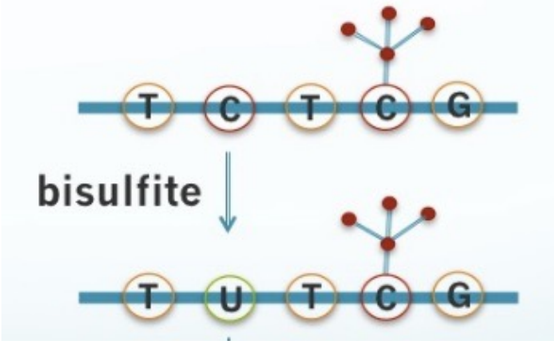
New Results

Indication of family-specific DNA methylation patterns in developing oysters

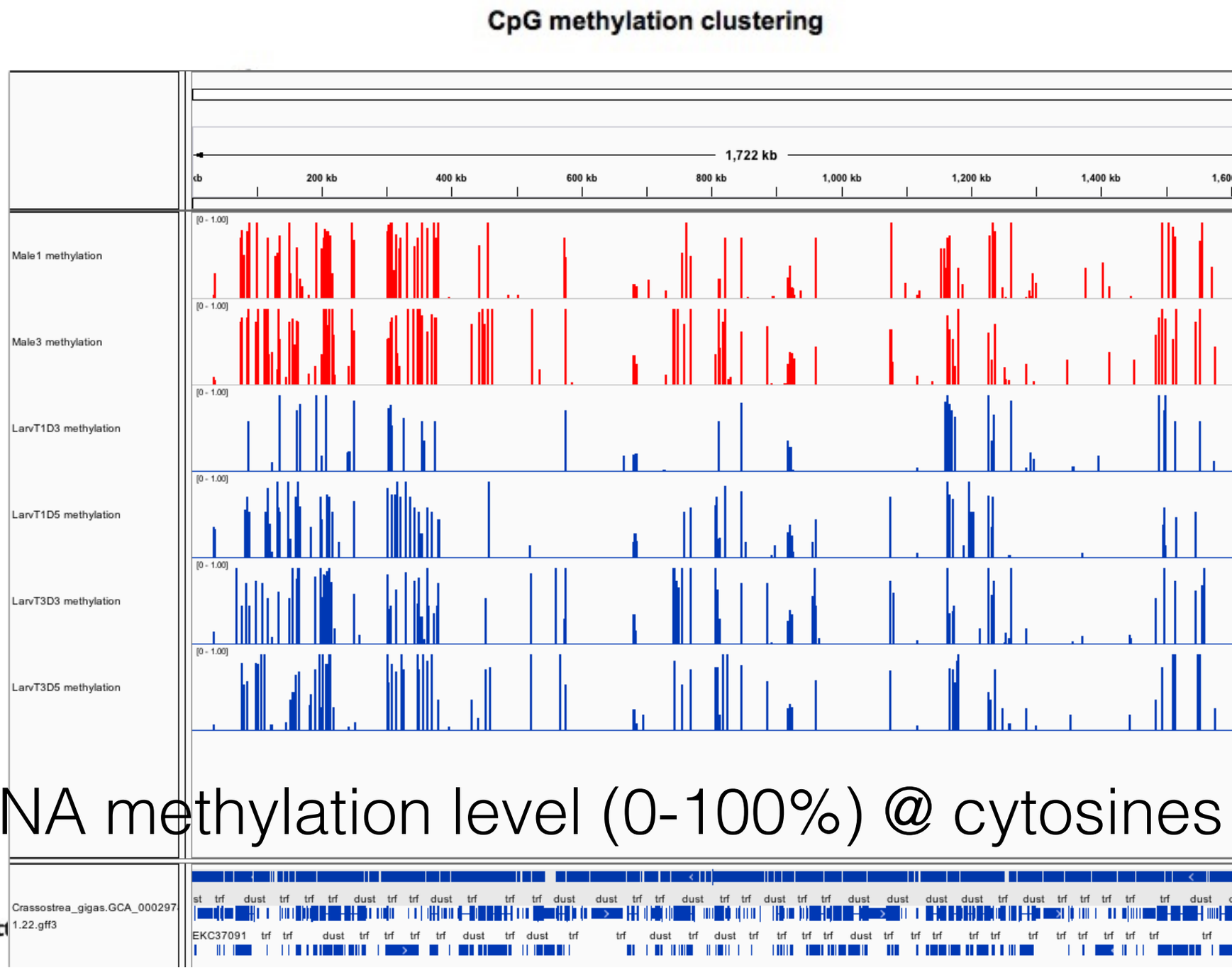
Claire E. Olson , Steven B. Roberts
doi: http://dx.doi.org/10.1101/012831

Epigenetic variation 1

Family and Developmental Variation



Sperm &
Larvae
(72h & 120h)



DNA methylation level (0-100%) @ cytosines

Cold Spring Harbor Laboratory

bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

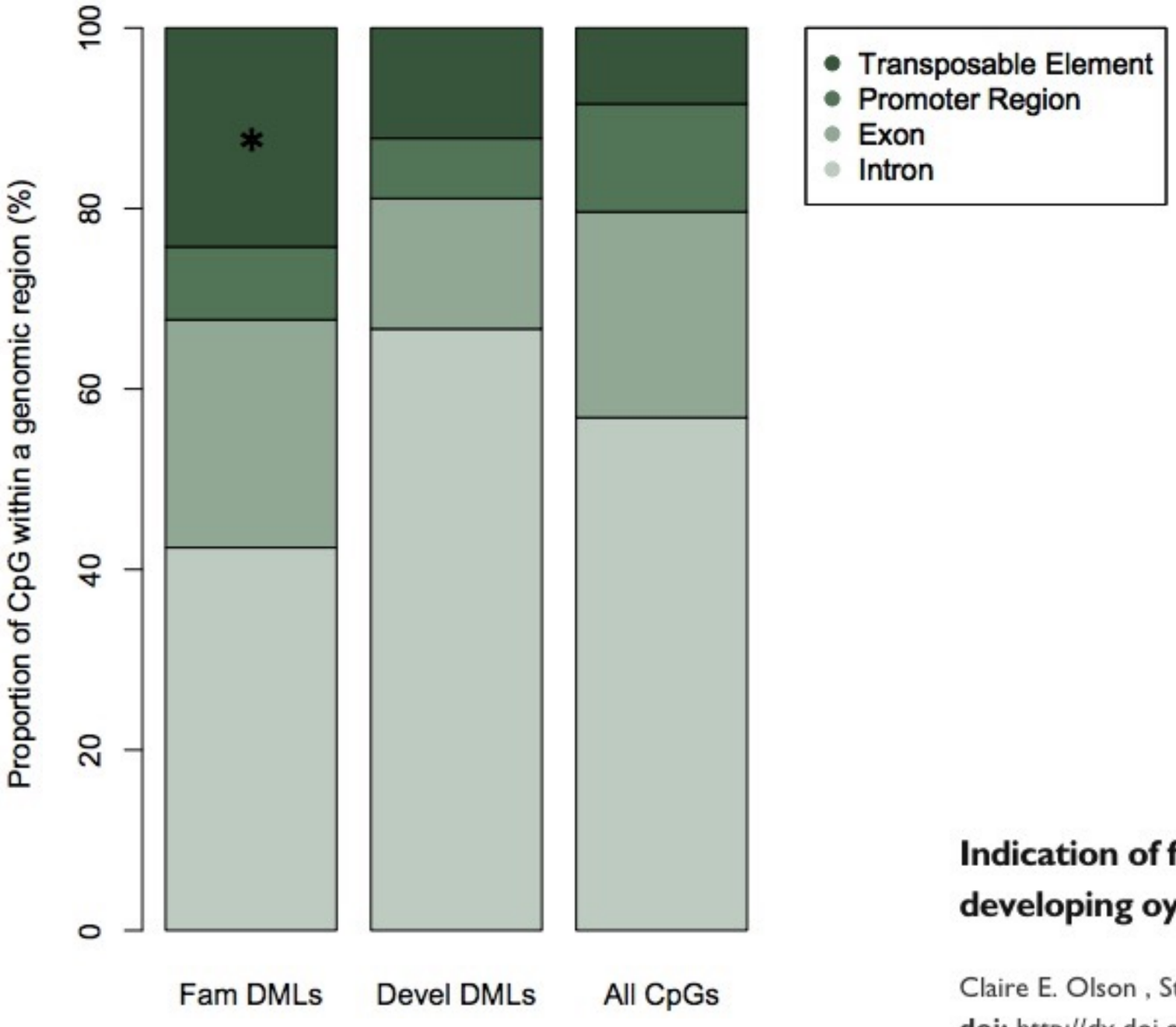
New Results

Indication of family-specific DNA methylation patt
developing oysters

Claire E. Olson , Steven B. Roberts

doi: http://dx.doi.org/10.1101/012831

Epigenetic variation 1

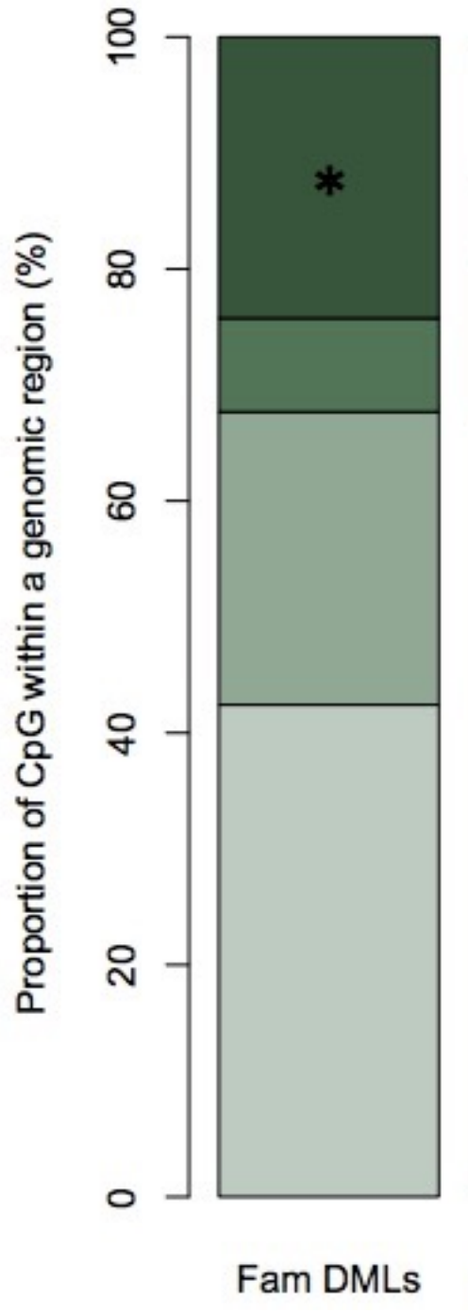


Indication of family-specific DNA methylation patterns in developing oysters

Claire E. Olson , Steven B. Roberts

doi: http://dx.doi.org/10.1101/012831

Epigenetic variation 1



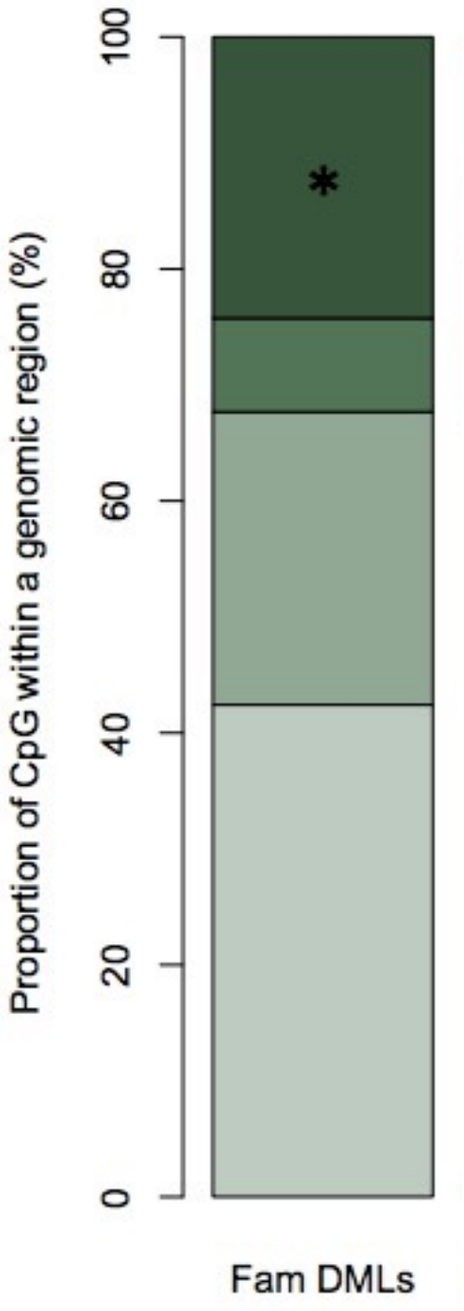
Differentially methylated loci in gene bodies could be lethal or deleterious, as they would alter gene expression.

Another possibility is that differentially methylated loci may provide advantageous phenotypic variation by increasing transposable element mobility.

Indication of family-specific DNA methylation patterns in developing oysters

Claire E. Olson , Steven B. Roberts
doi: http://dx.doi.org/10.1101/012831

Epigenetic variation 1

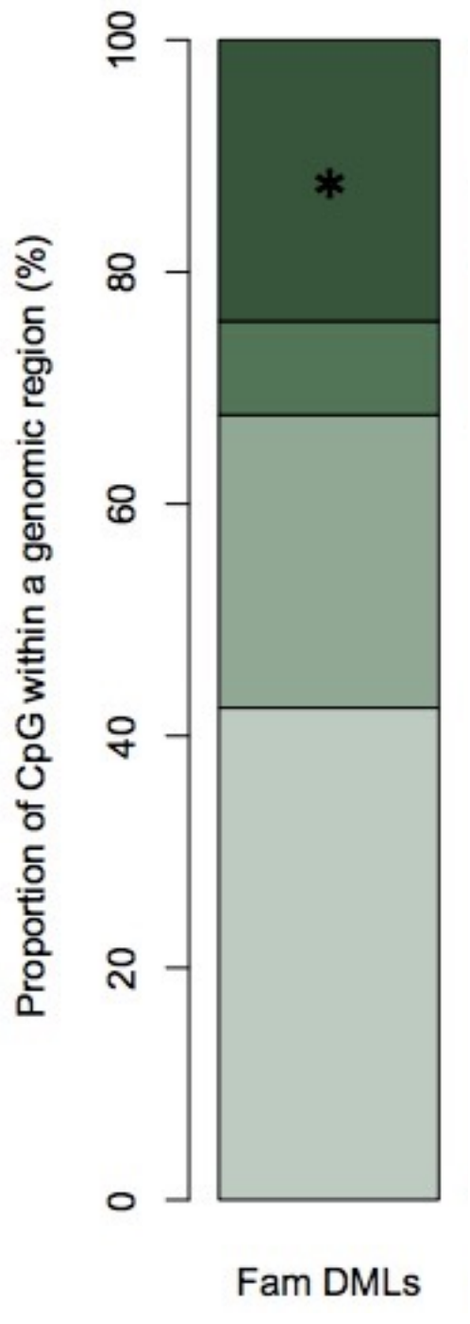


Assumes some form of random process occurring during gametogenesis?

Indication of family-specific DNA methylation patterns in developing oysters

Claire E. Olson , Steven B. Roberts
doi: http://dx.doi.org/10.1101/012831

Epigenetic variation 1



This repository Search

che625 / olson-ms-nb

github SOCIAL CODING

IPython Notebook for downloading and analyzing data the manusript: "Indication of family-specific DNA methylation patterns in developing oysters"

bioRix preprint - http://dx.doi.org/10.1101/012831

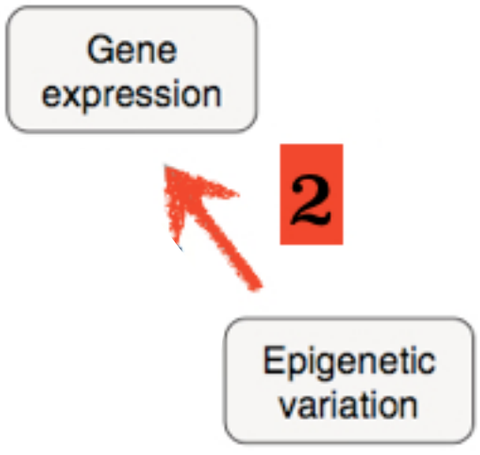
To execute the IPython Notebook in its entirety you will need:

- IPython - install instructions
- BSMAP - install instructions
- bedtools - install instructions
- R - install instructions
- rpy2 (interface to R from Python) - install instructions

Epigenetic variation 1

Summary

- Sparsely (~16 %), mosiac methylated genome
- Gene body methylation correlated with function
- DNA methylation patterns are inherited
- DMRs are predominant in transposable elements

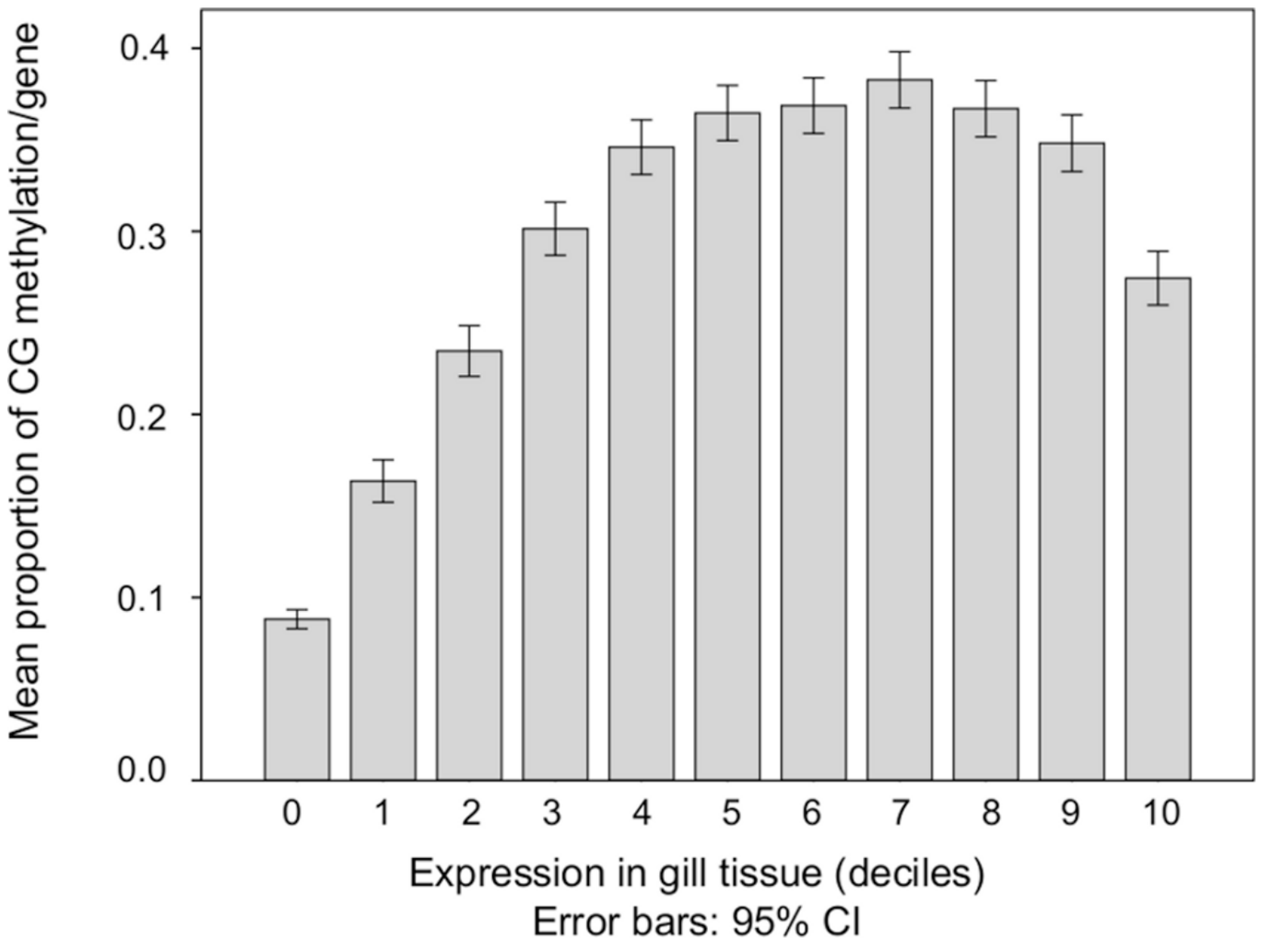


Function?

Gene expression

2

Epigenetic variation



PeerJ

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

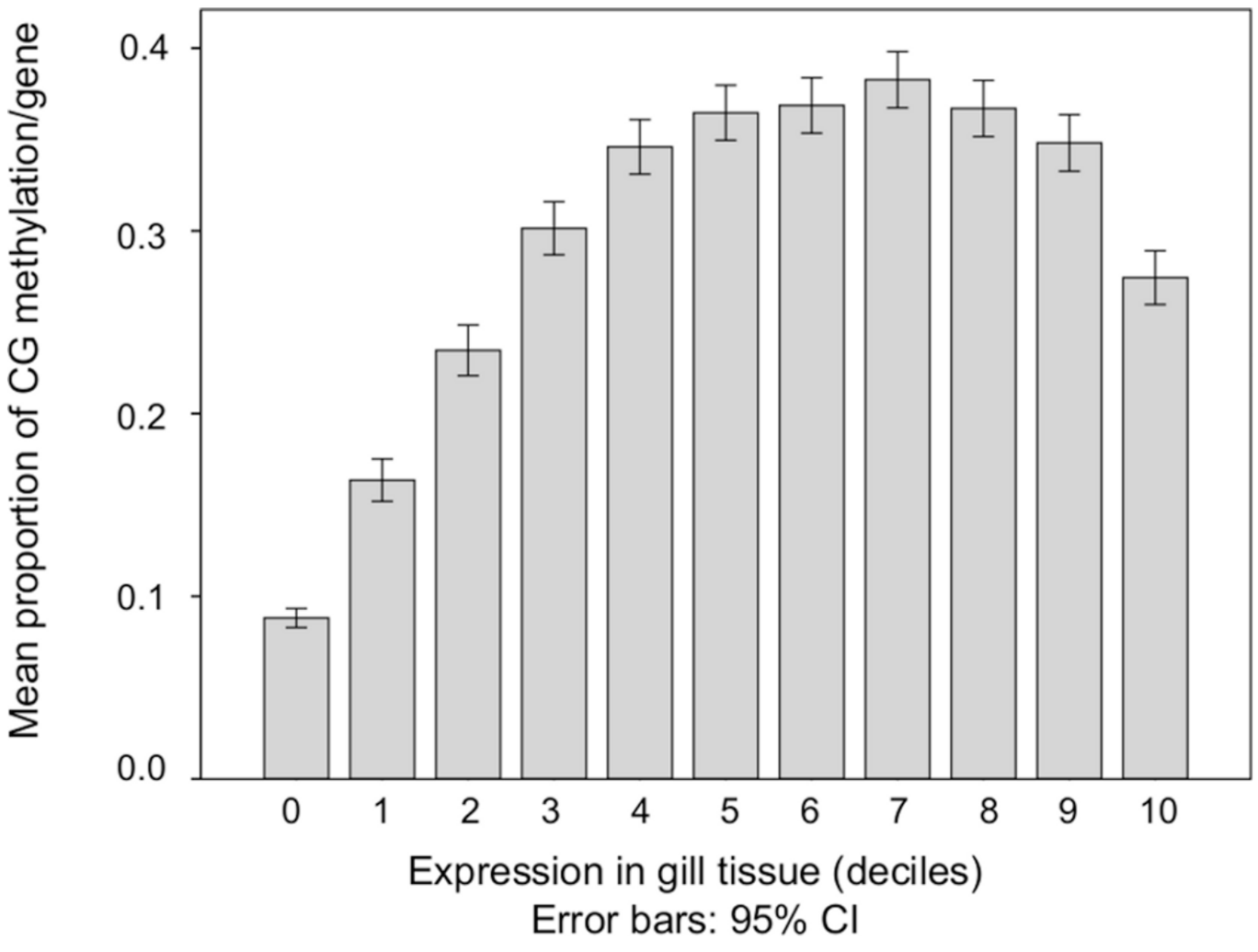
Mackenzie R. Gavery and Steven B. Roberts

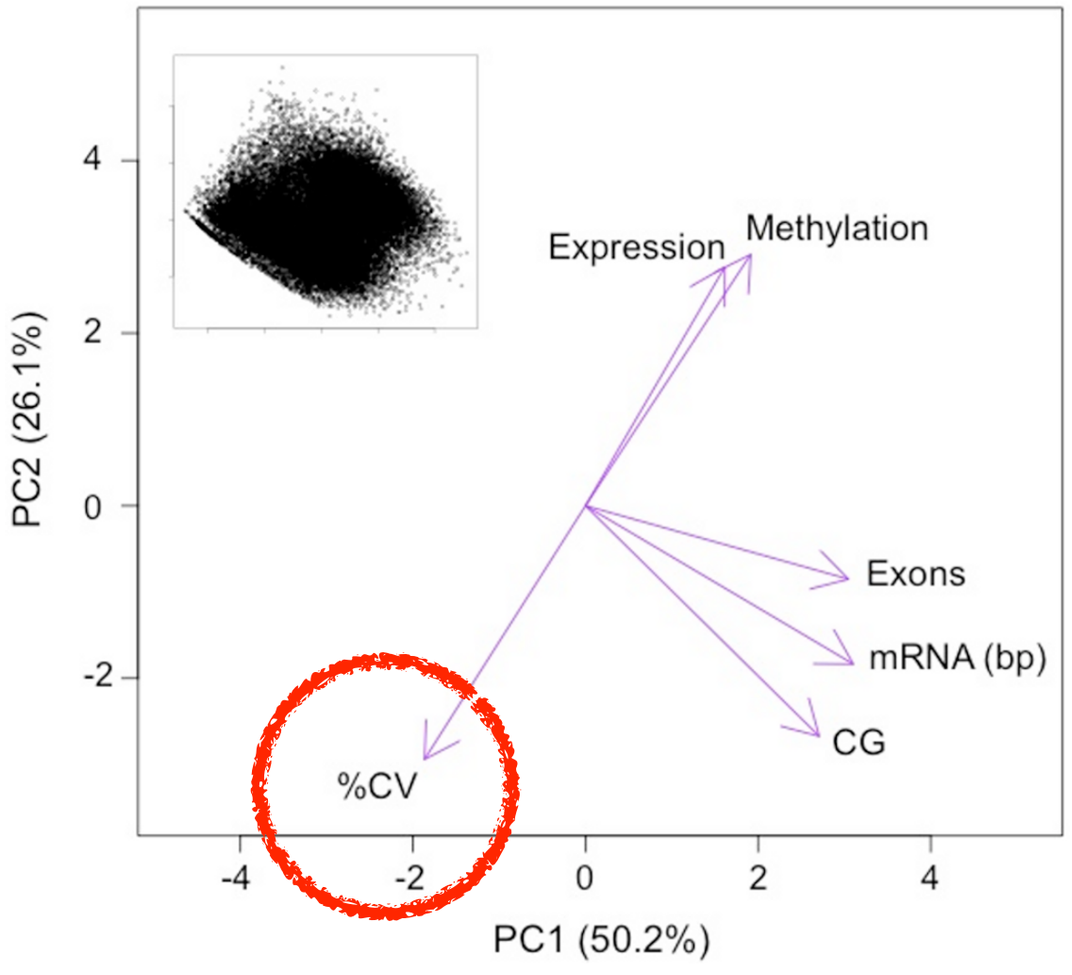
School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

Gene expression

2

Epigenetic variation





PeerJ

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

Mackenzie R. Gavery and Steven B. Roberts

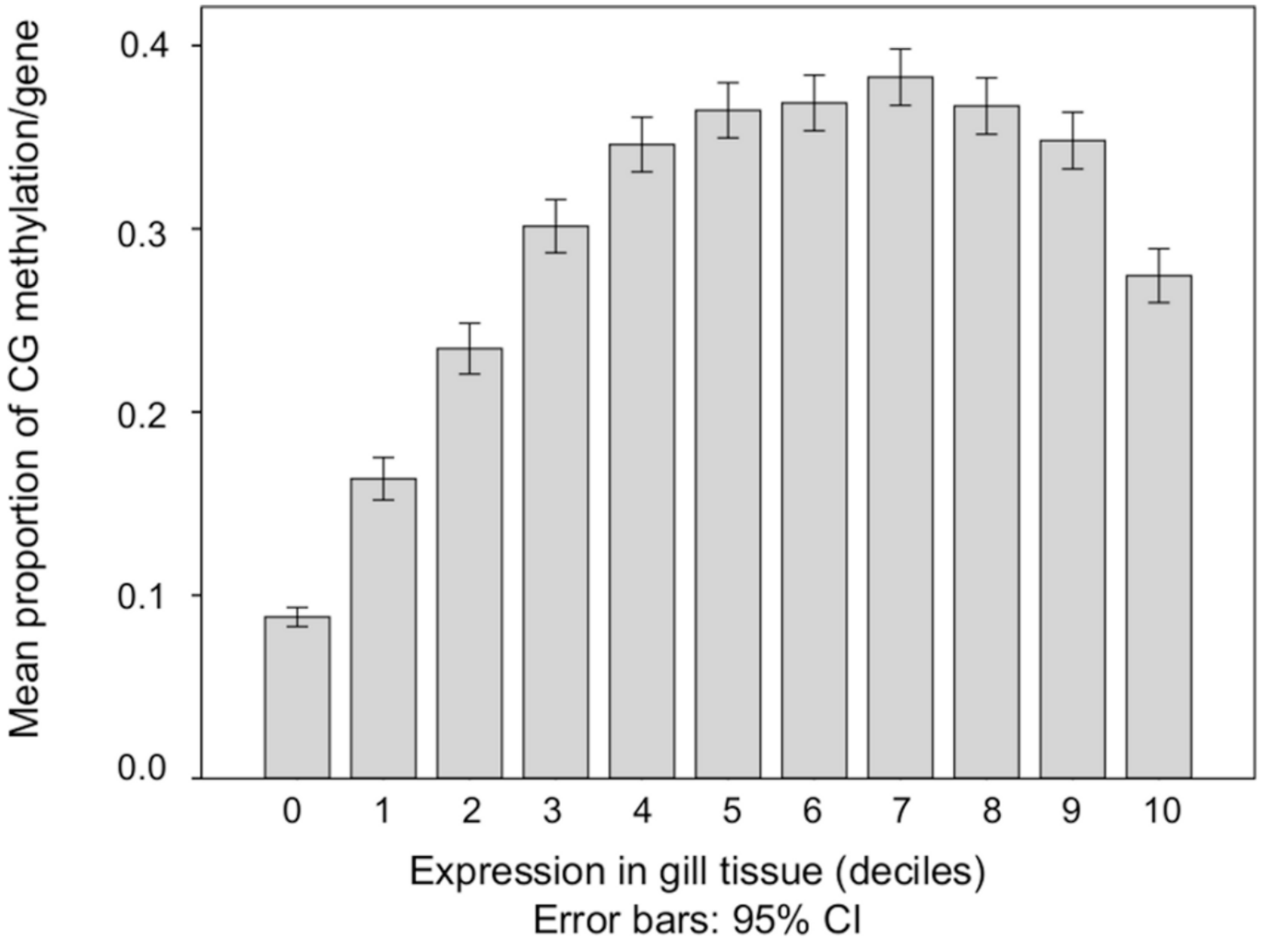
School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

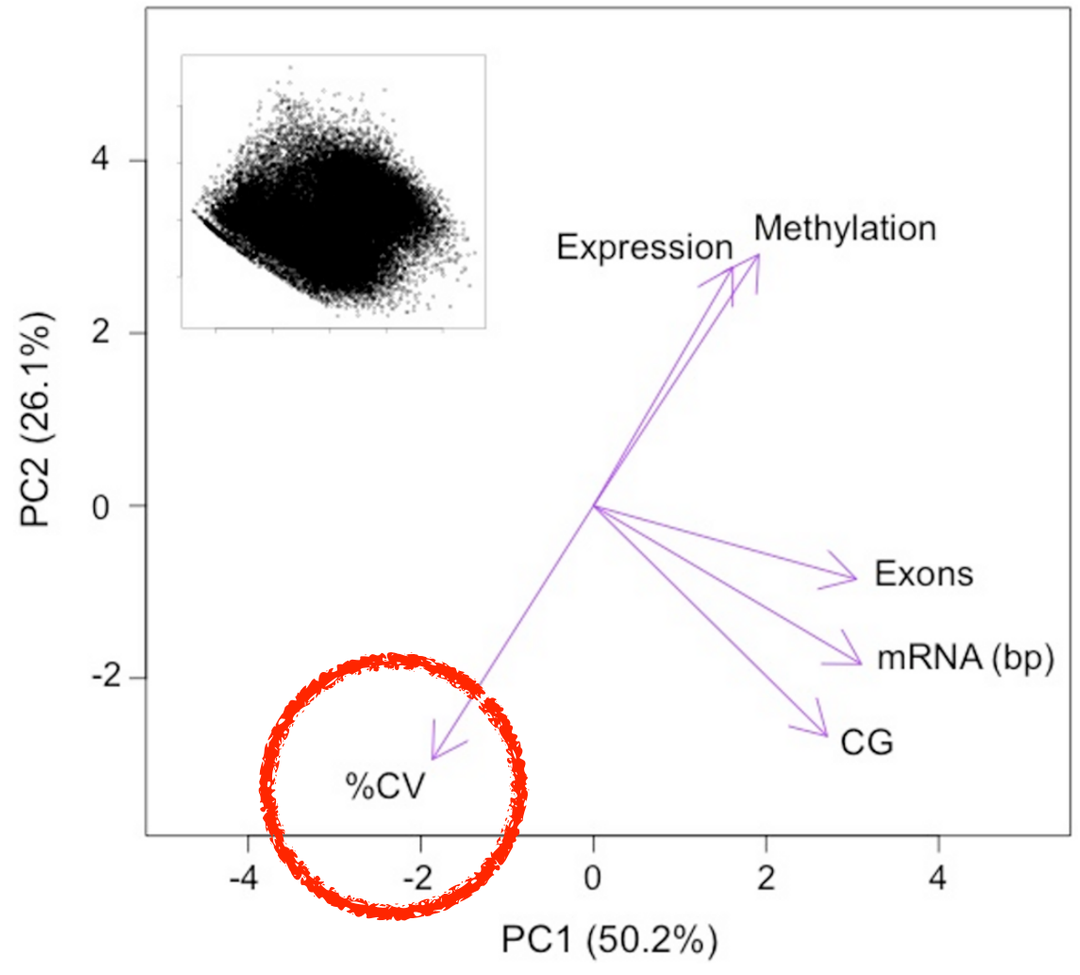
Gene expression

2

Epigenetic variation

Theory: **Does not** influence expression level but rather alternative splicing.





PeerJ

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

Mackenzie R. Gavery and Steven B. Roberts

School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

BRIEFINGS IN FUNCTIONAL GENOMICS. VOL 13. NO 3. 217–222

A context dependent role for DNA methylation in bivalves

Mackenzie R. Gavery and Steven B. Roberts

Advance Access publication date 7 January 2014

In species that experience a diverse range of environmental conditions, processes have evolved to increase the number of potential phenotypes in a population in order to improve the chances for an individual's survival.

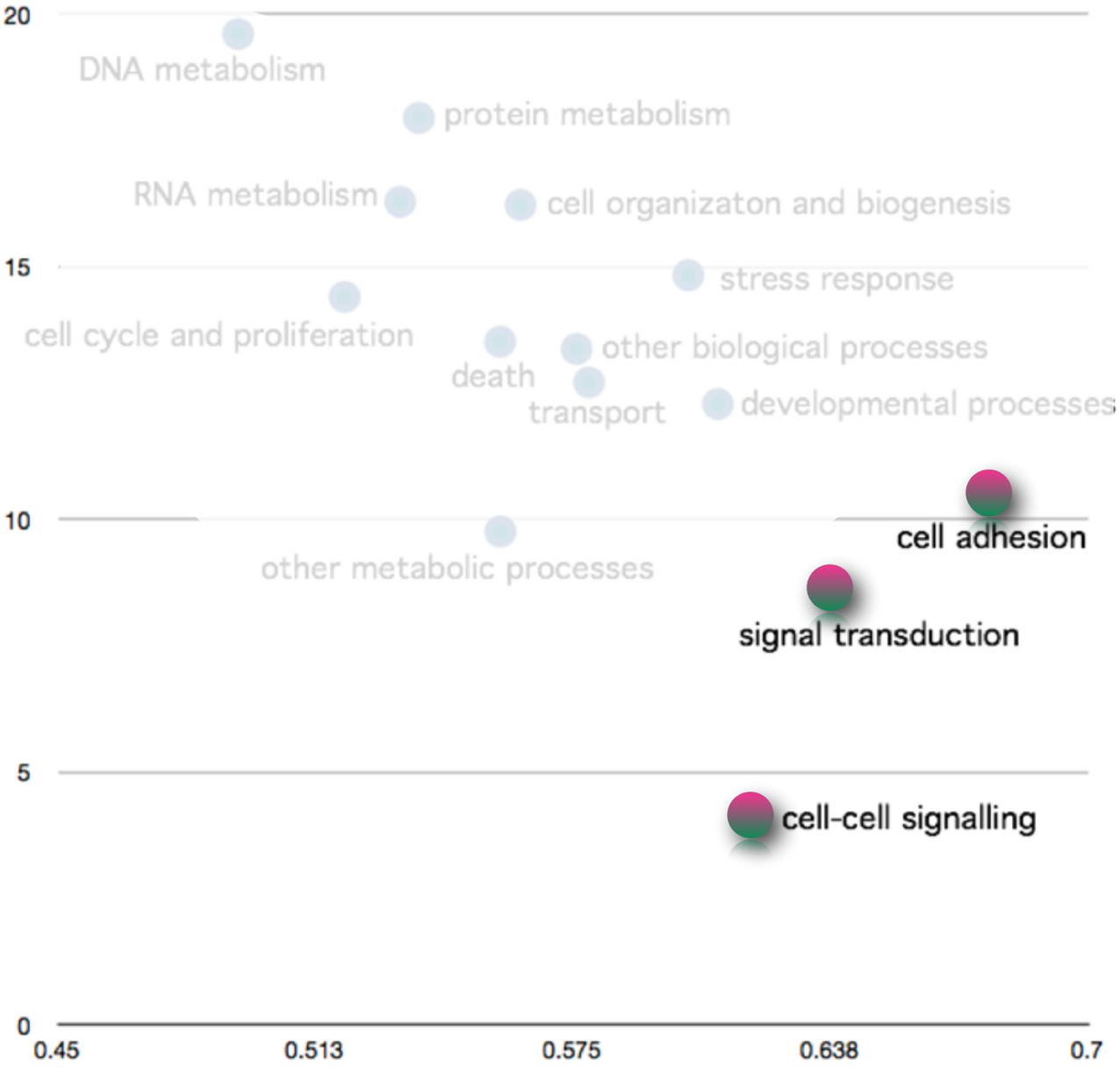
Gene expression

2

Epigenetic variation

Measured degree of DNA methylation

Enrichment level in MBD-Seq library

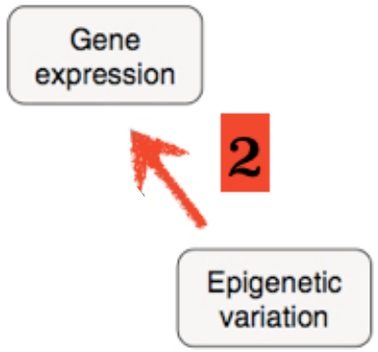


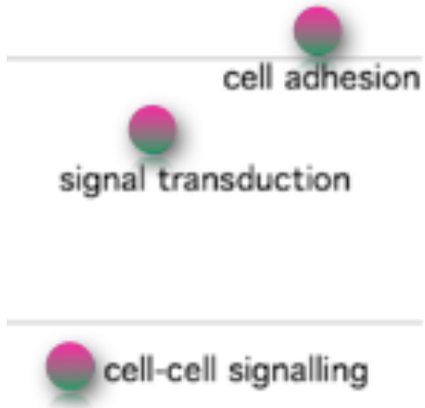
***in silico* analysis**

(modified from Gavery and Roberts 2010)

Predicted degree of DNA methylation

Roberts and Gavery 2012

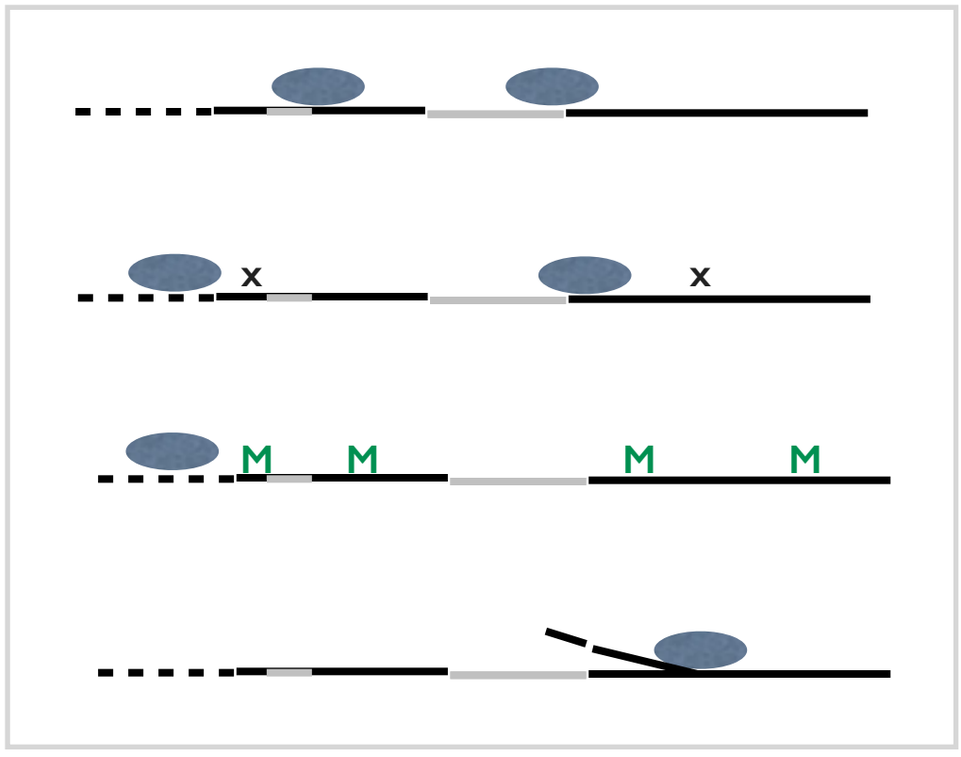




sparse methylation

tissue / temporal specific and inducible genes

Transcriptional opportunities



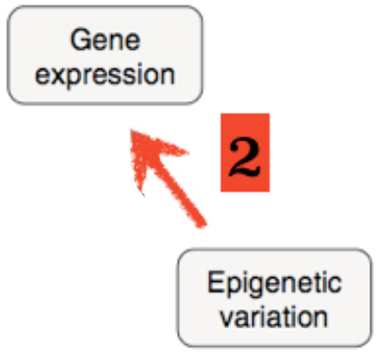
alternative start sites

sequence mutation
change AA, premature stop codon

conventional transcription
transient methylation

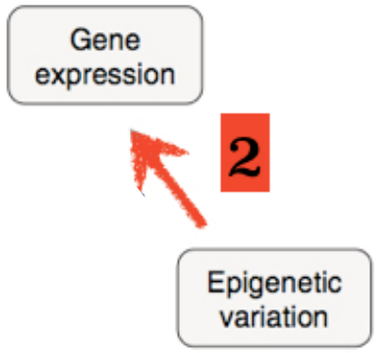
alternate transcript
exon skipping

Roberts and Gavery 2012



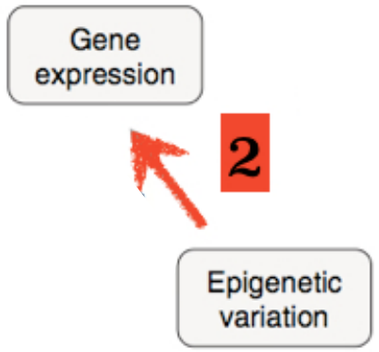
promoter exon intron exon

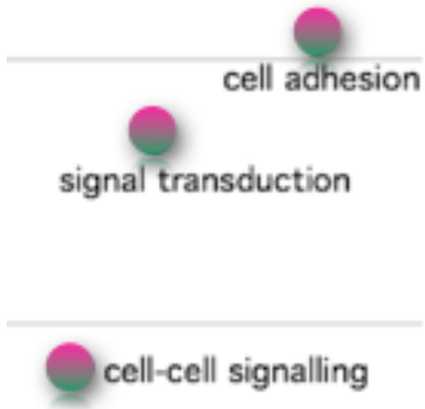
Roberts and Gavery 2012



transcript

Roberts and Gavery 2012

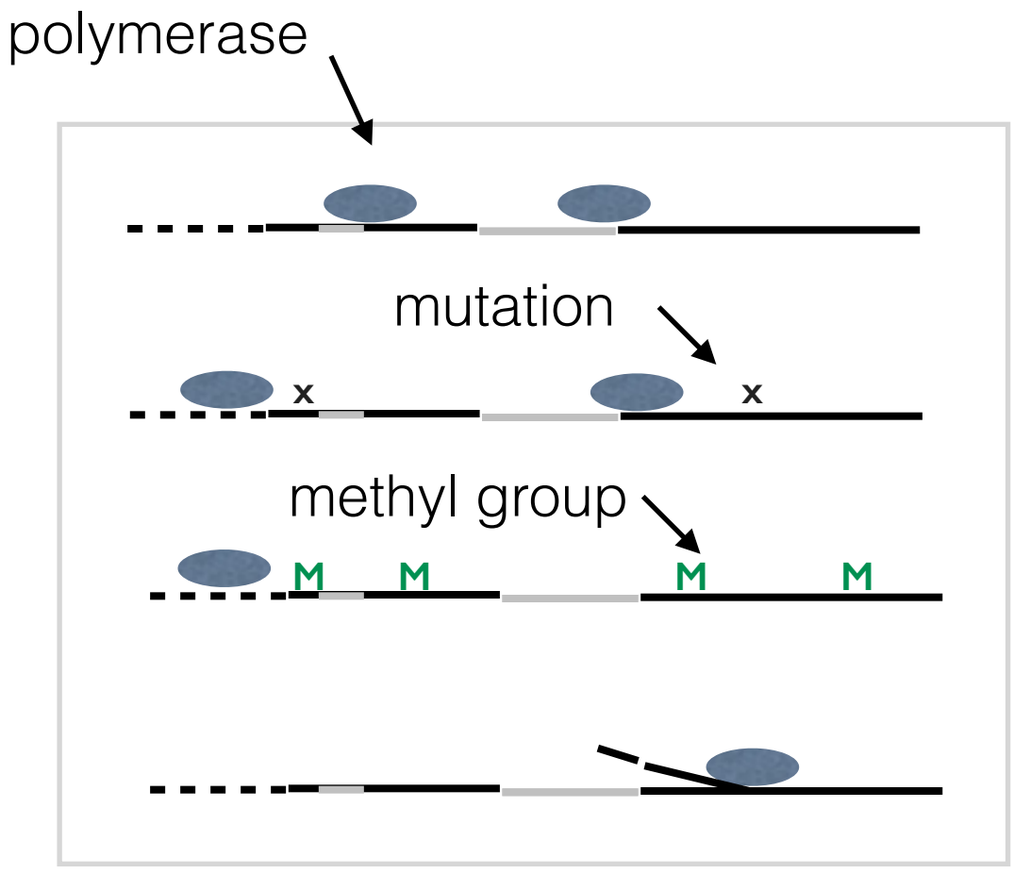




sparse methylation

tissue / temporal specific and inducible genes

Transcriptional opportunities



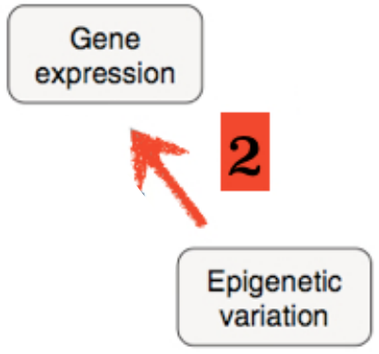
alternative start sites

sequence mutation
change AA, premature stop codon

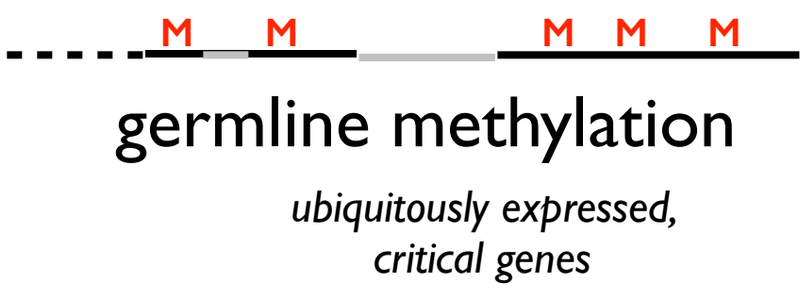
conventional transcription
transient methylation

alternate transcript
exon skipping

Roberts and Gavery 2012

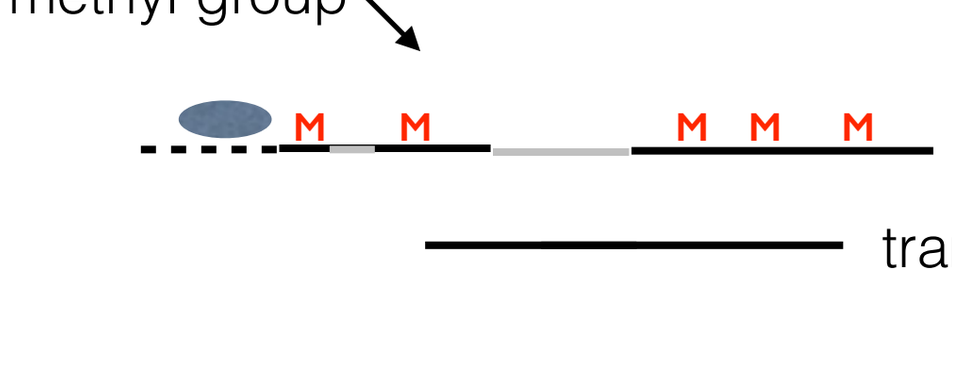


Transcriptional opportunities



germline methylation

ubiquitously expressed, critical genes



conventional transcription

5’ UTR promotor

Gene expression

2

Epigenetic variation

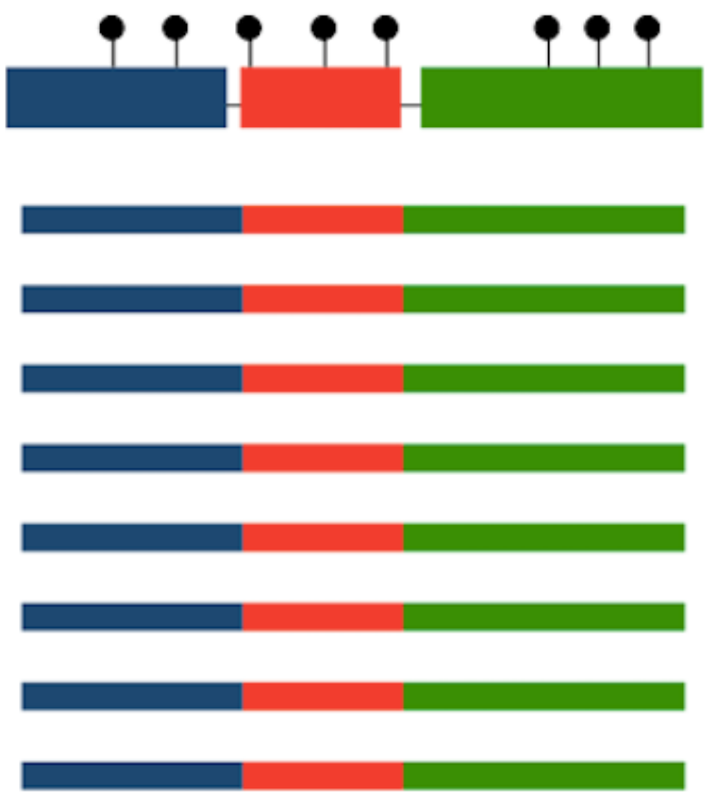
Stochastic Variation

BRIEFINGS IN FUNCTIONAL GENOMICS. VOL 13. NO 3. 217–222

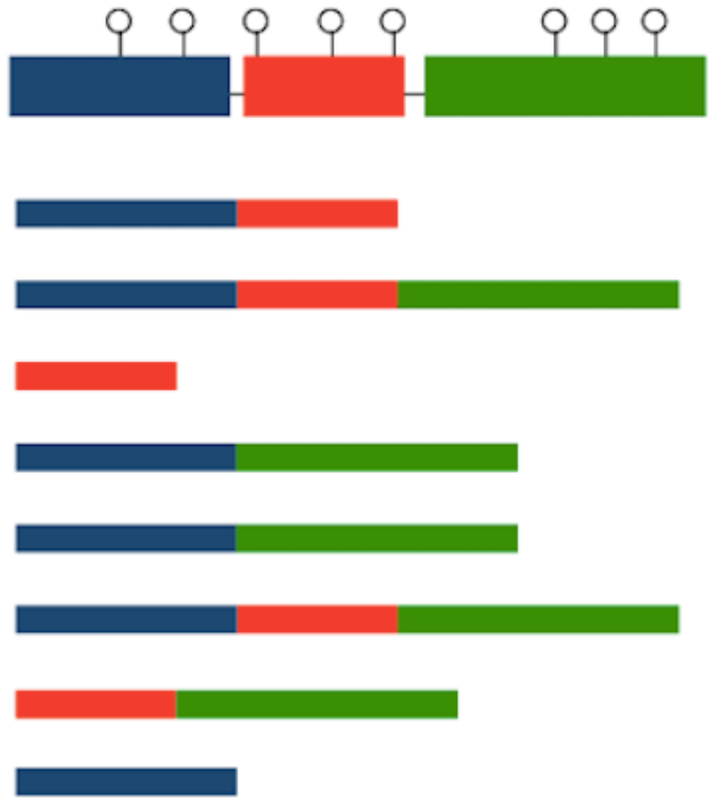
A context dependent role for DNA methylation in bivalves

Mackenzie R. Gavery and Steven B. Roberts

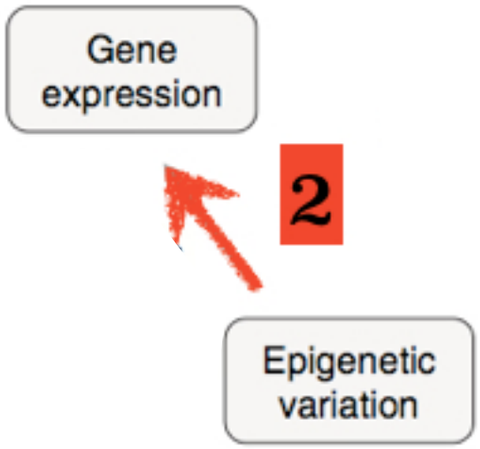
Advance Access publication date 7 January 2014



housekeeping



response to change



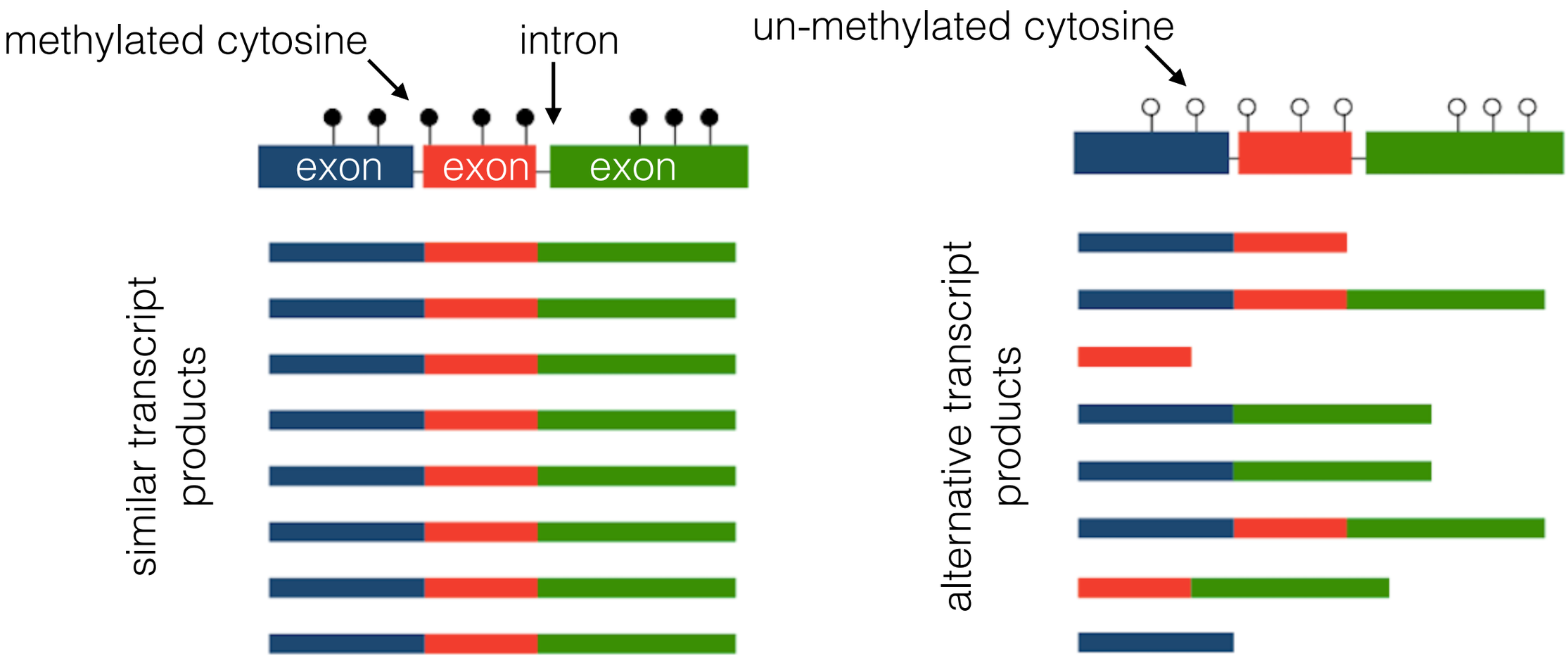
Stochastic Variation

BRIEFINGS IN FUNCTIONAL GENOMICS. VOL 13. NO 3. 217–222

A context dependent role for DNA methylation in bivalves

Mackenzie R. Gavery and Steven B. Roberts

Advance Access publication date 7 January 2014



housekeeping

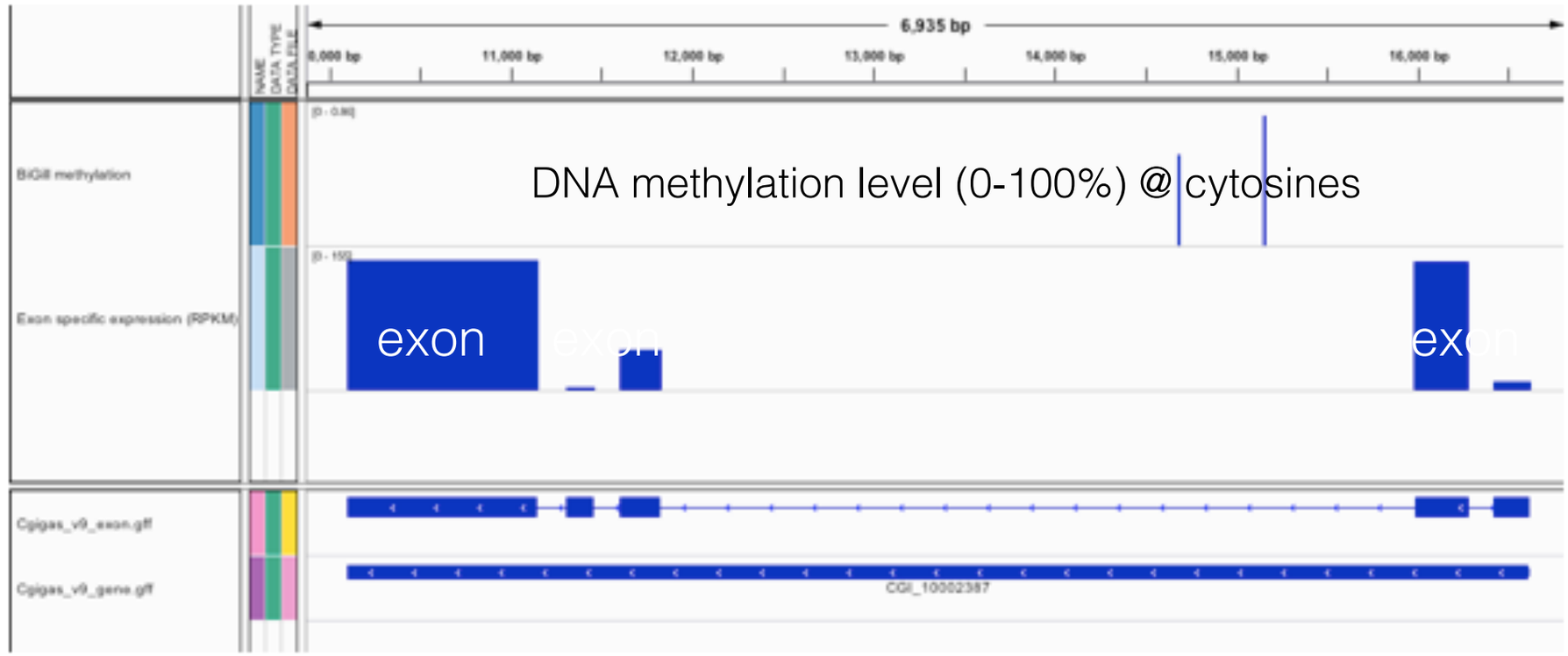
response to change

response to change

avg exp
199

%meth
2%

Heat shock 70 kDa protein 12A

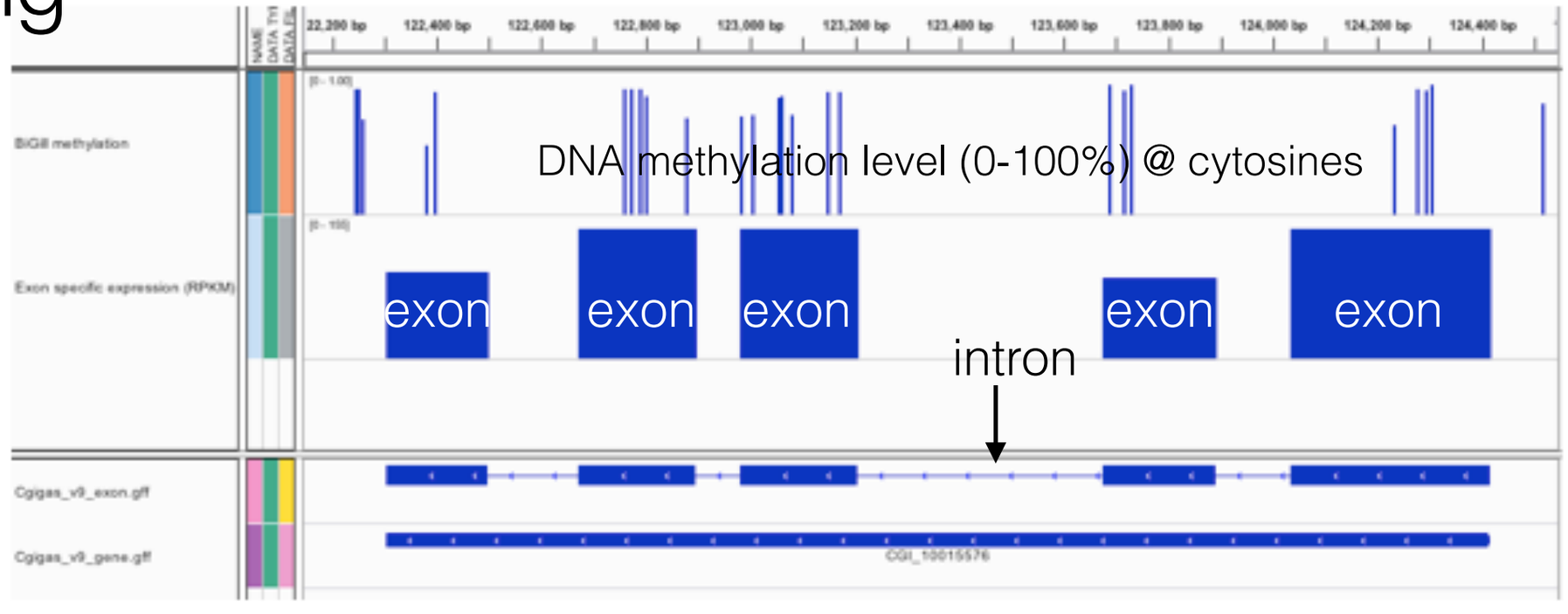


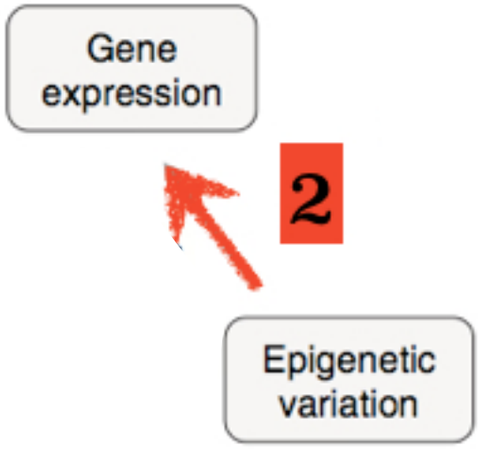
housekeeping

avg exp
197

%meth
91%

Tektin-2





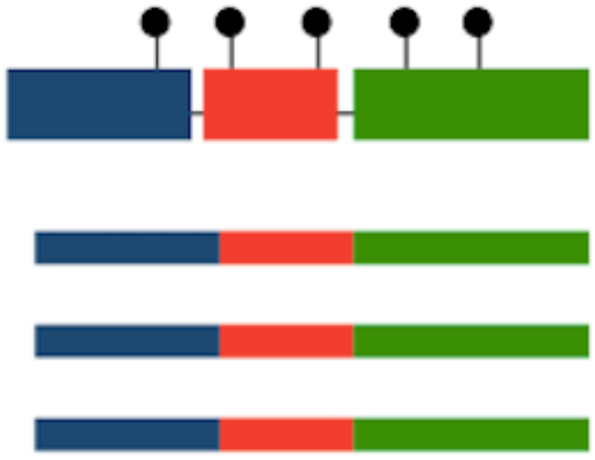
Targeted Regulation

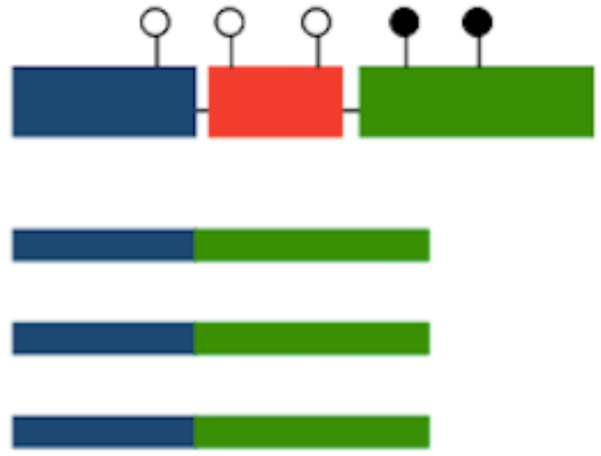
BRIEFINGS IN FUNCTIONAL GENOMICS. VOL 13. NO 3. 217–222

A context dependent role for DNA methylation in bivalves

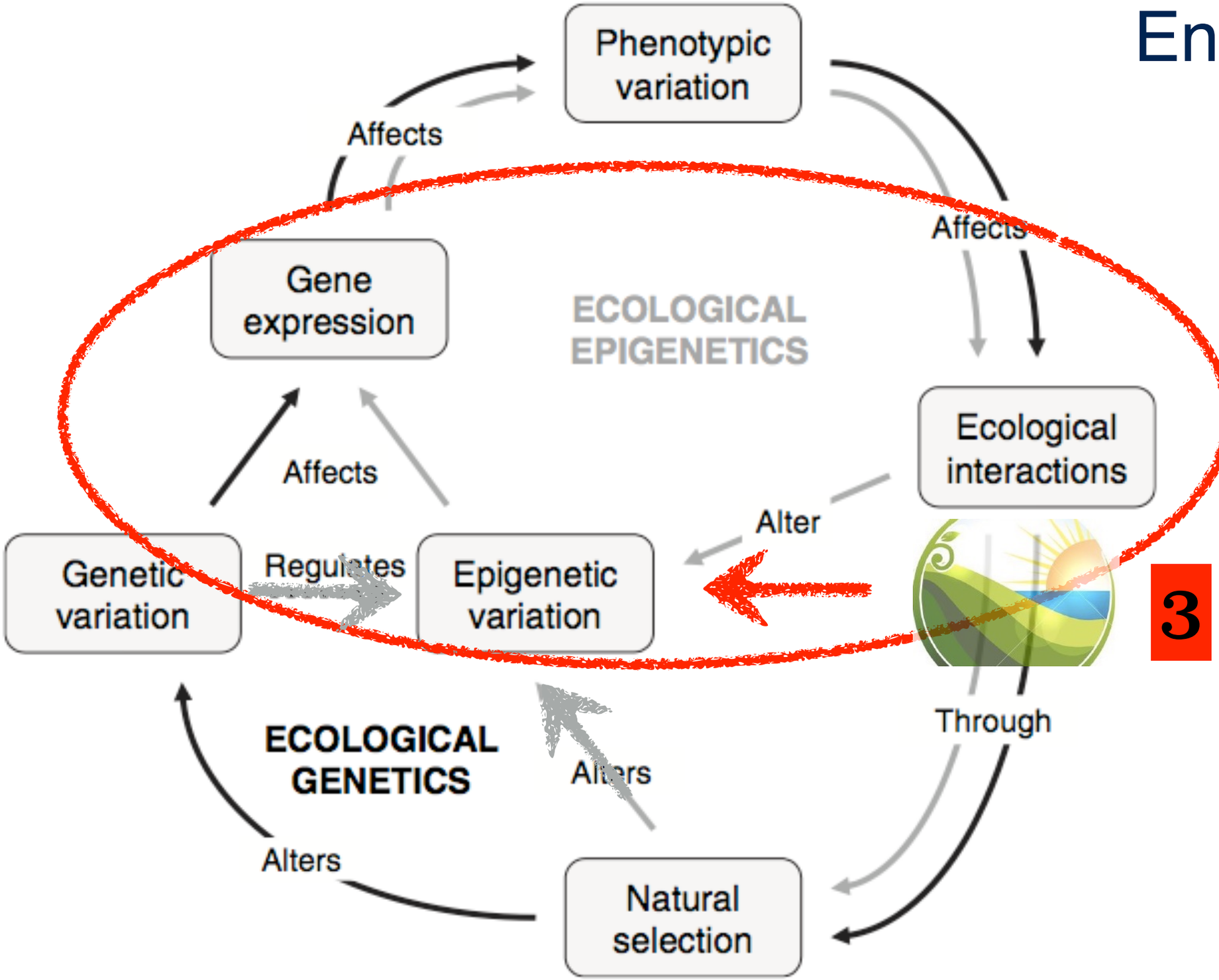
Mackenzie R. Gavery and Steven B. Roberts

Advance Access publication date 7 January 2014





Environmental Influence



3

modified from

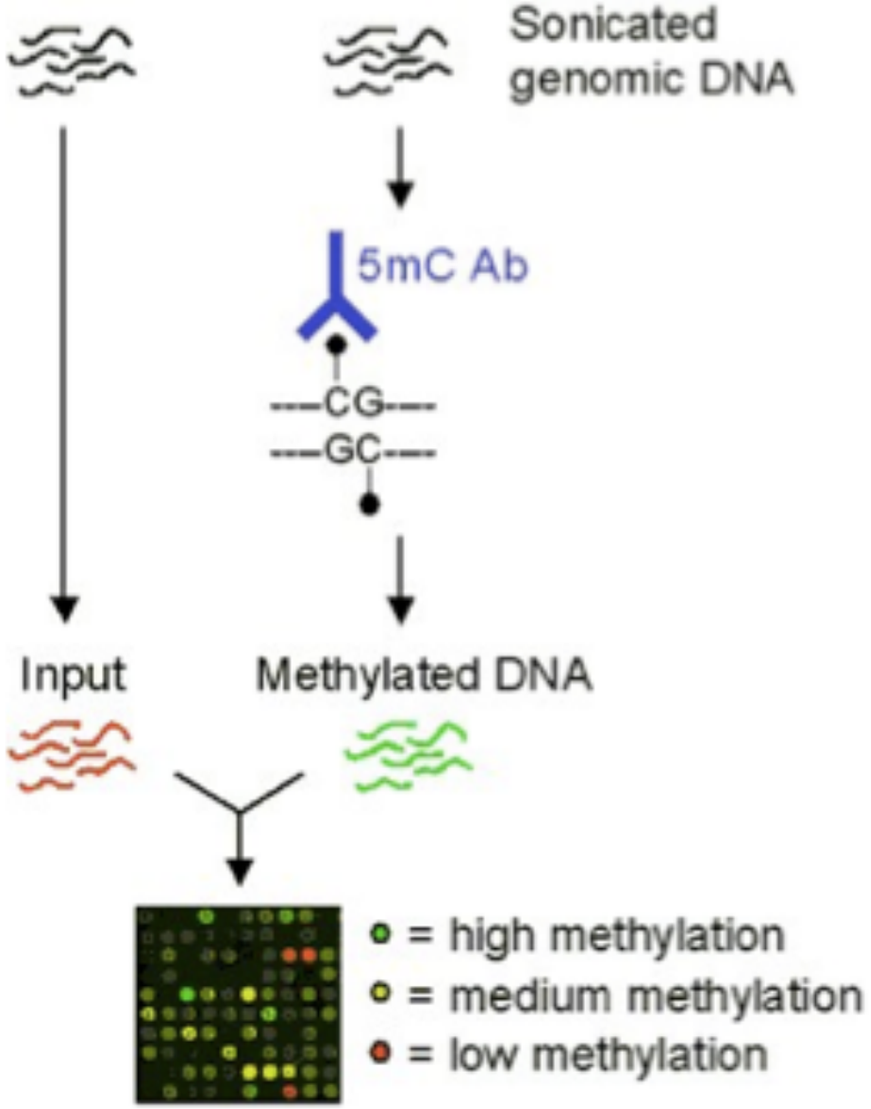
Ecology Letters, (2008) **11**: 106–115 doi: 10.1111/j.1461-0248.2007.01130.x

IDEA AND PERSPECTIVE

Epigenetics for ecologists

Oliver Bossdorf,[1]* Christina L. Richards[2] and Massimo Pigliucci[3]

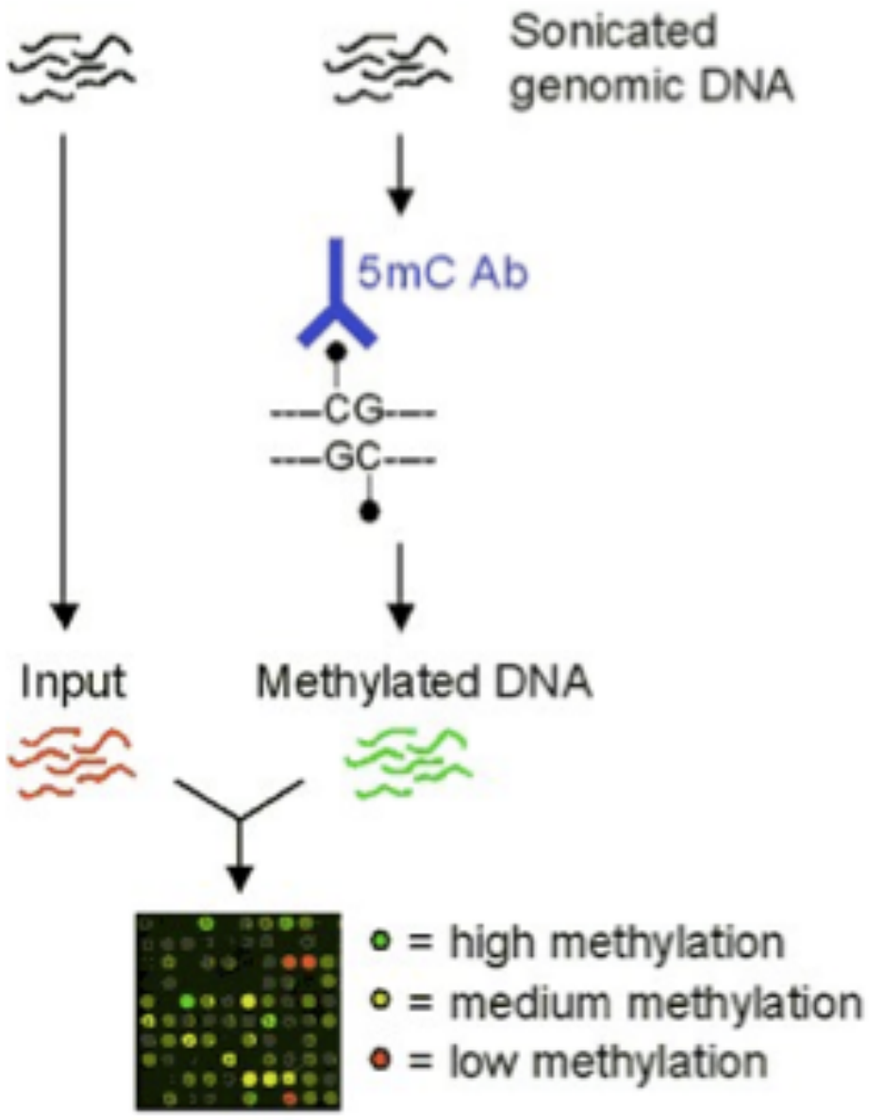
Very new data
Environmental impact (Estrogens)



- 96 regions that were hypermethylated in EE2
  - 90 are in genes
  - 52 of these cross exon/intron boundary another
  - 32 are in introns
  - 6 are just in exons
- 287 regions that were hypomethylated in EE2
  - 256 are in genes
  - 138 cross exon/intron boundary
  - 114 are in introns only and
  - 4 are just in exons

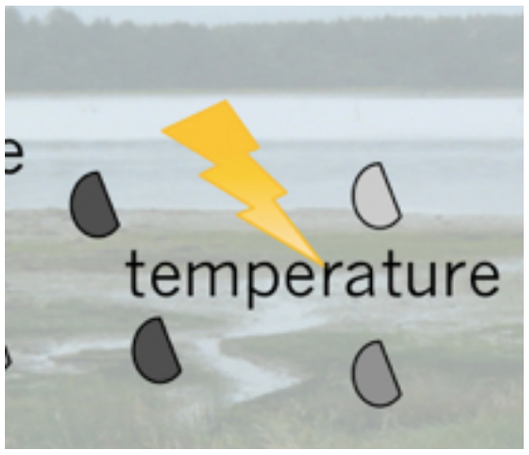
Very new data

Environmental impact (Estrogens)

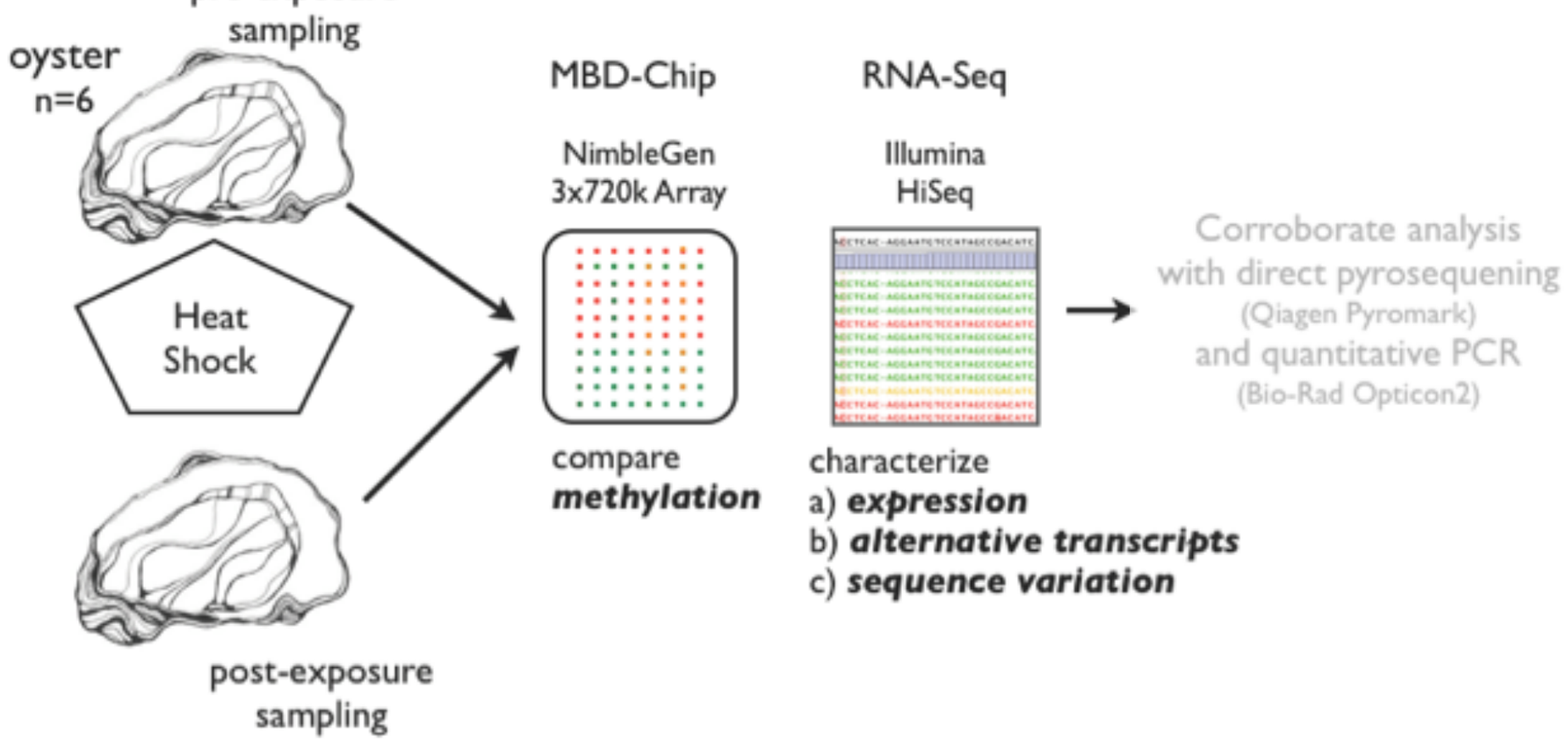


Exposure results in
significant hypomethylation

Very new data
Environment and gene expression

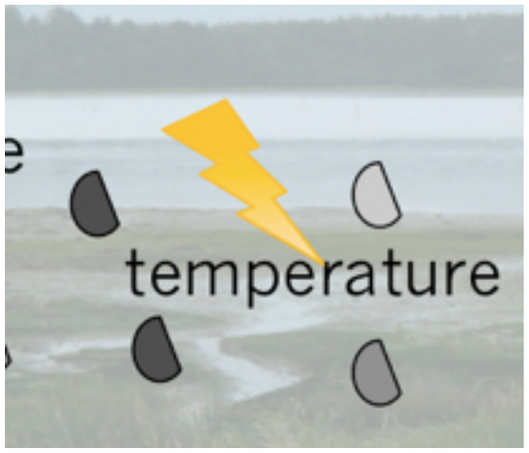


stochastic or targeted?



Very new data

Environment and gene expression

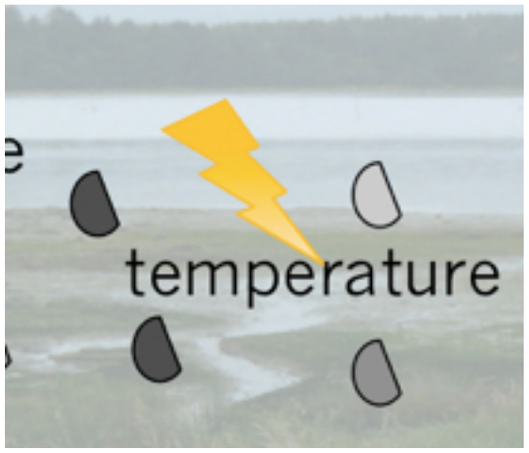


stochastic or targeted?

Oyster	Hypo-methylated	Hyper-methylated
2	7224	2803
4	6560	3587
6	7645	4044

Very new data

Environment and gene expression



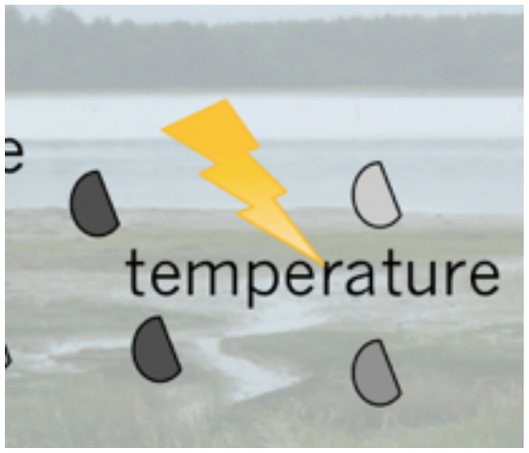
stochastic or targeted?

Oyster	Hypo-methylated	Hyper-methylated
2	7224	2803
4	6560	3587
6	7645	4044

No obvious association with genome feature including *differentially expressed genes*

Very new data

Environment and gene expression

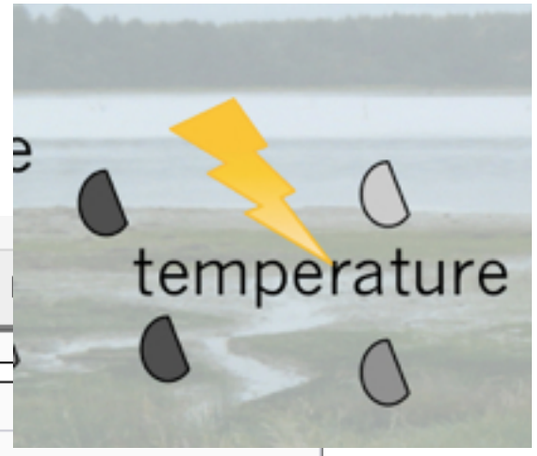


stochastic or targeted
or ..?

Oyster	Hypo-methylated	Hyper-methylated
2	7224	2803
4	6560	3587
6	7645	4044

Changes in methylation
(either direction)
are more prevalent
in introns, repeats,
and
transposable elements.

Very new data



IGV

Crassostrea_gigas.... | scaffold1009 | scaffold1009:677,703-719,650 | Go

41 kb

680 kb | 690 kb | 700 kb | 710 kb

RNA-seq 2 pre
RNA-seq 4 pre
RNA-seq 6 pre
RNA-seq 2 post
RNA-seq 4 post
RNA-seq 6 post

mCpG 2
mCpG 4
mCpG 6
mCpG 2
mCpG 4
mCpG 6

hypermethylated region (50bp)

hypomethylated region (50bp)

exon (x26)

Predicted methylation is limited.

Crassostrea_gigas.GCA_000297
1.22.gtf

Cuffdiff_geneexp.sig.gtf

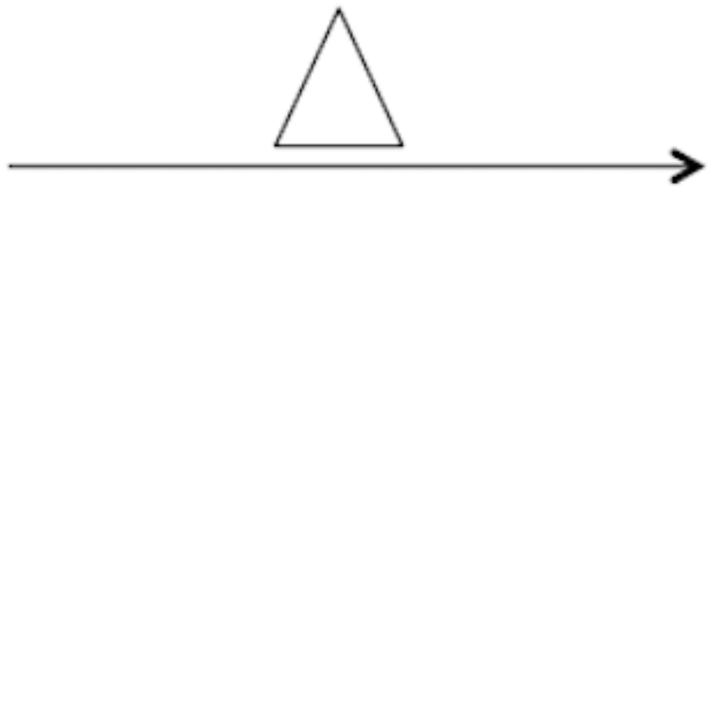
EKC31121

EKC31122

Gene expression

2

Epigenetic variation



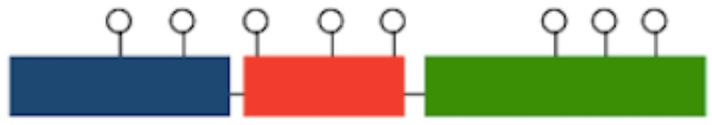
not in this experiment

not even consistent methylation changes at loci level

Gene expression

2

Epigenetic variation



more questions….
interesting but what is controlling?

Take Home

Oyster genome has a fantastic degree of diversity contributing to phenotypic plasticity & adaptation potential.

Take Home

Oyster genome has a fantastic degree of diversity contributing to phenotypic plasticity & adaptation potential.

- large gene families
- very high mutation rate (snp/50bp)
- numerous exons per gene (potential for alternatives)
- genome full of repeats region
- high number of transposable elements
- lack of methylation of transposable suggest mobility
- family variation of methylation
- limited methylation environmental response genes

is associated with spurious transcription

- inheritance of epigenetic marks as mechanism of improved adaptation

Next Steps

Dig into the heat-shock data

Transposable Elements

Consider other epigenetic process

Photo credit: Flickr, Creative Commons, csessums

Next Steps

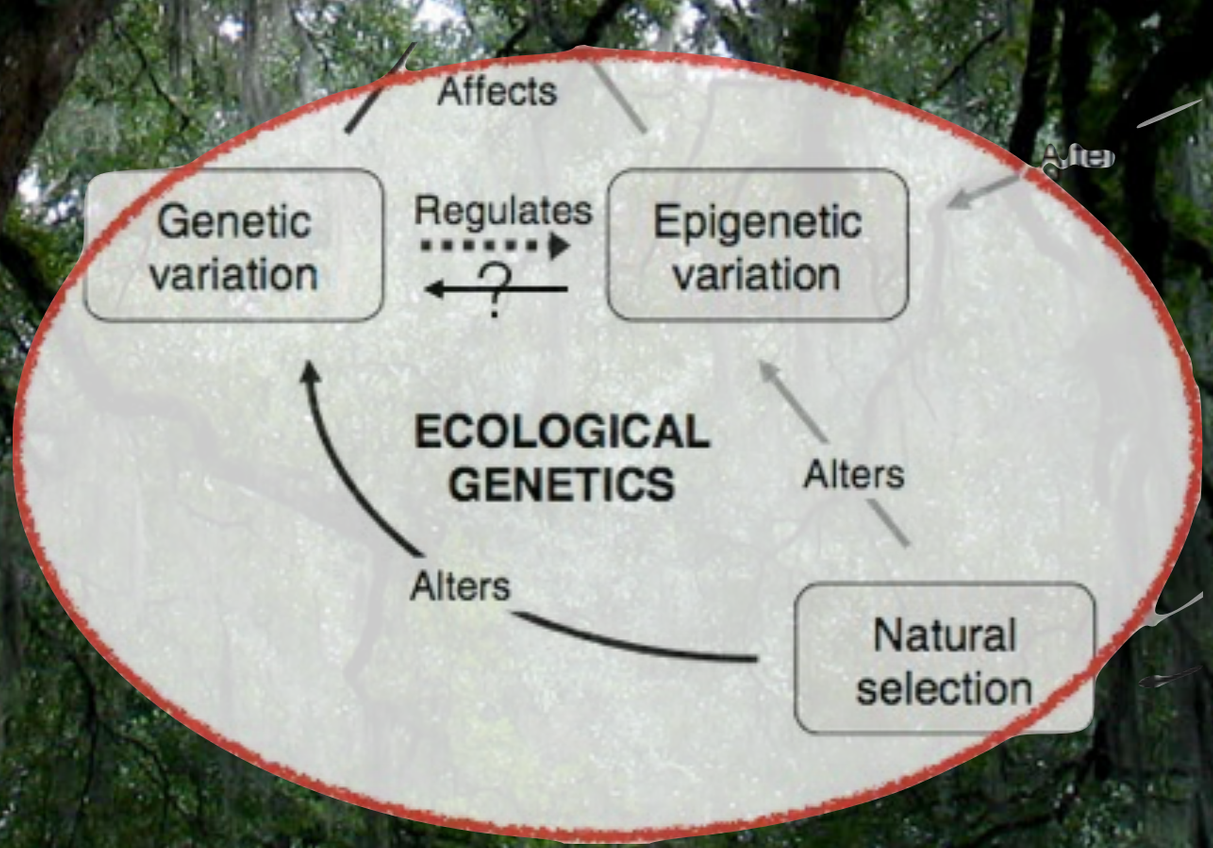


Photo credit: Flickr, Creative Commons, csessums

Very new ~~data~~
Heritability
Plasticity
Local Adaptation

Genetics
versus
Epigenetics

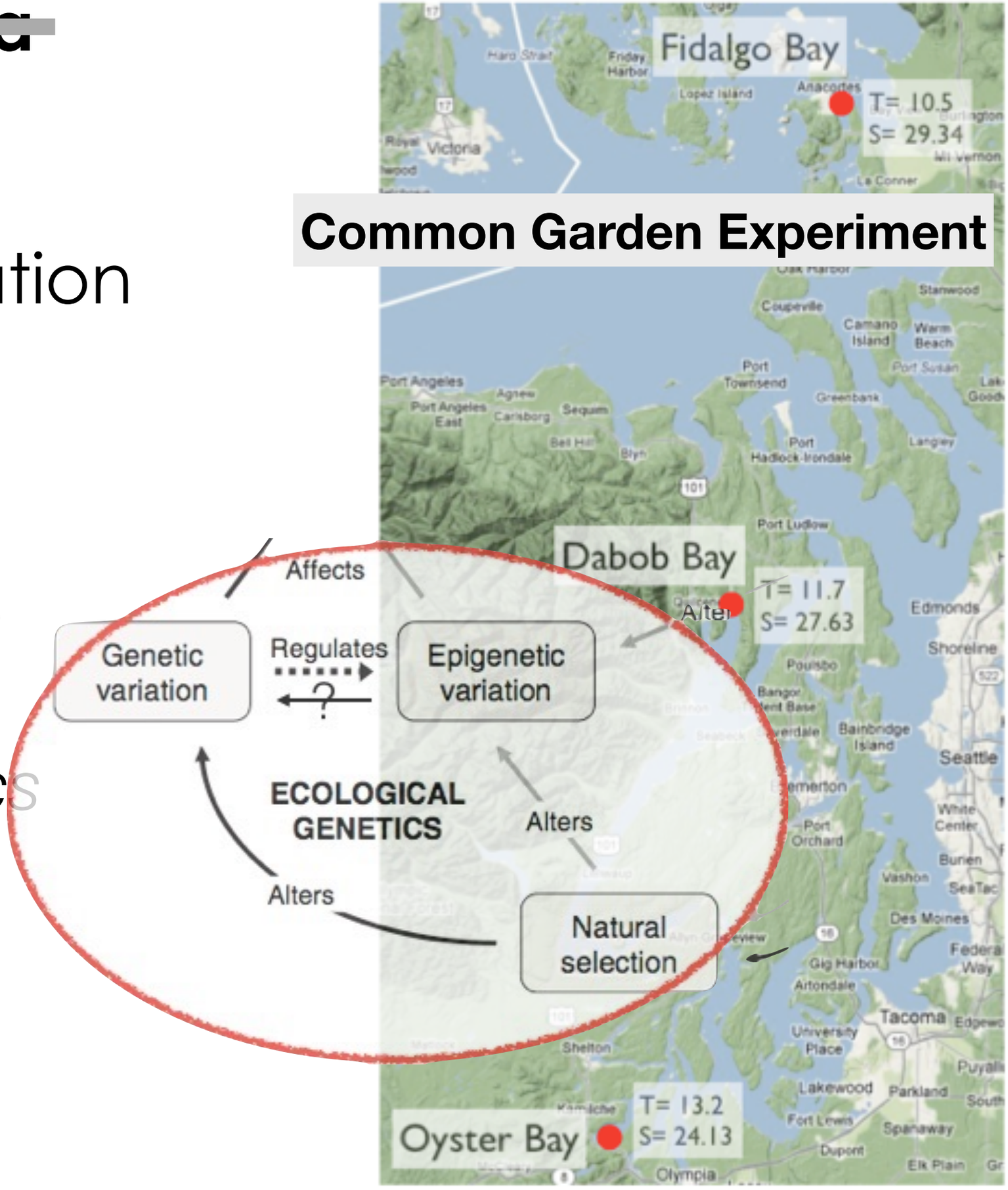
Very new ~~data~~

Heritability

Plasticity

Local Adaptation

Genetics versus Epigenetics



Open Science

web-native scholarship

Sharing

Photo credit: Flickr, Creative Commons, speechless

Open Data

Open Methods

Open Data

Sharing raw data

Example

s! Public Sigenae Contig Browser Oyster

Search e! Oyster: Anything Go

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

Based on Ensembl release 40 - Aug 2006

Use Ensembl to...

- Search Contig Browser
- Data mining [BioMart]
- Export data
- Download data
- Digital Differential Display
- Venn Diagrams

Docs and downloads

- About Contig Browser
- Contig Browser data

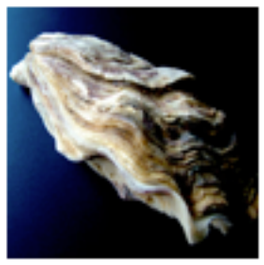
Other links

- Home
- Sigenae
- Agenae

wellcome trust sanger institute · EBI · INRA

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:

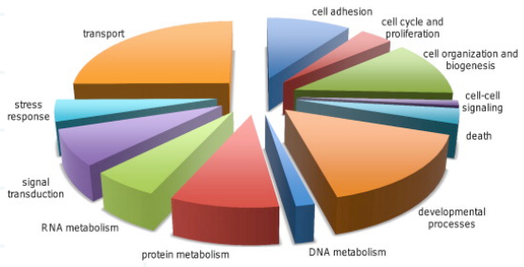
Contig: BQ426804.p.cg.8

SNP: snp_EW777925_683

Example

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



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

Mackenzie R. Gavery, Steven B. Roberts *

Raw Data

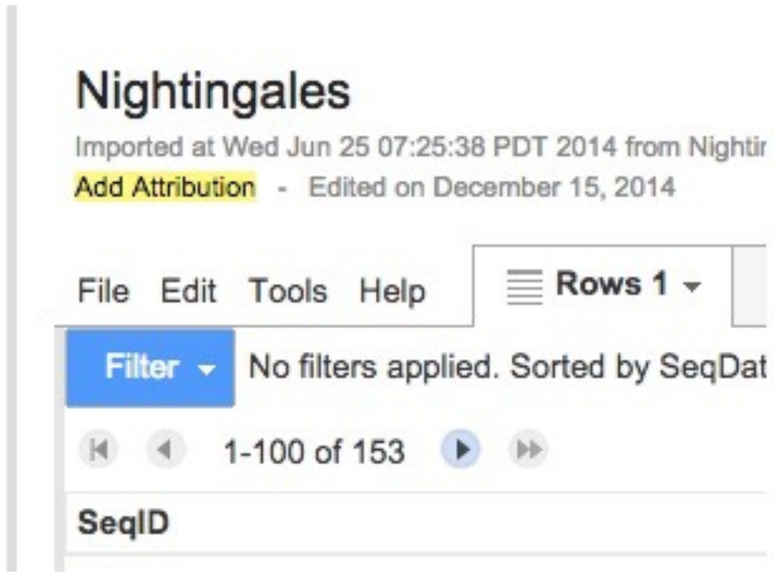
1) As sequencing facility provdes data, files are downloaded to our local NAS (owl), in the root `nightingales` directory. http://owl.fish.washington.edu/nightingales/

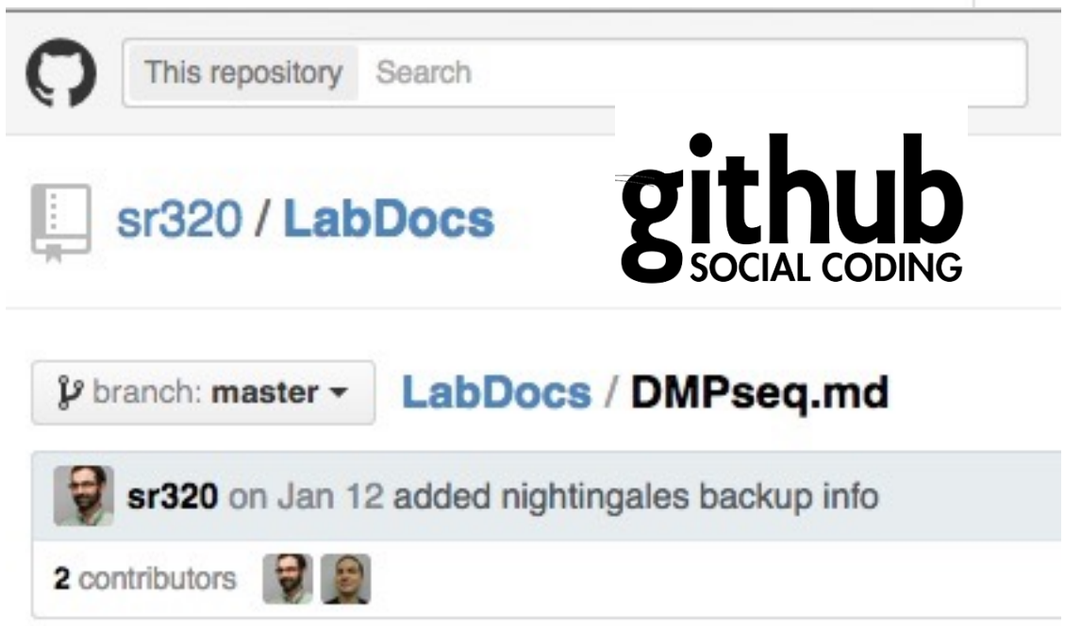
2) The Nightingales Google Spreadsheet is updated.

3) Update the Nightingales Google Fusion Table with new information from the Nightingales Google Spreadsheet. This is accomplished by:

- deleting all rows in the Nightingales Google Fusion Table (Edit > Delete all rows)
- Importing data from the Nightingales Google Spreadsheet (File > Import more rows...)

4) Once metadata is included in the Google Fusion Table Nightingales, respective data files are moved to subdirectory labelled with first letter of genus followed by species ie `C_gigas`. Check url in Nightingales Fusion table to ensure it is accurate.





Querying Disparate Datasets

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/oqCxl	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

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/oqCxl	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

SQLSHARE

Your datasets
All datasets
Shared datasets
Recent activity... 293
Recently viewed »

Upload dataset
New query

YOUR TOP VIEWED

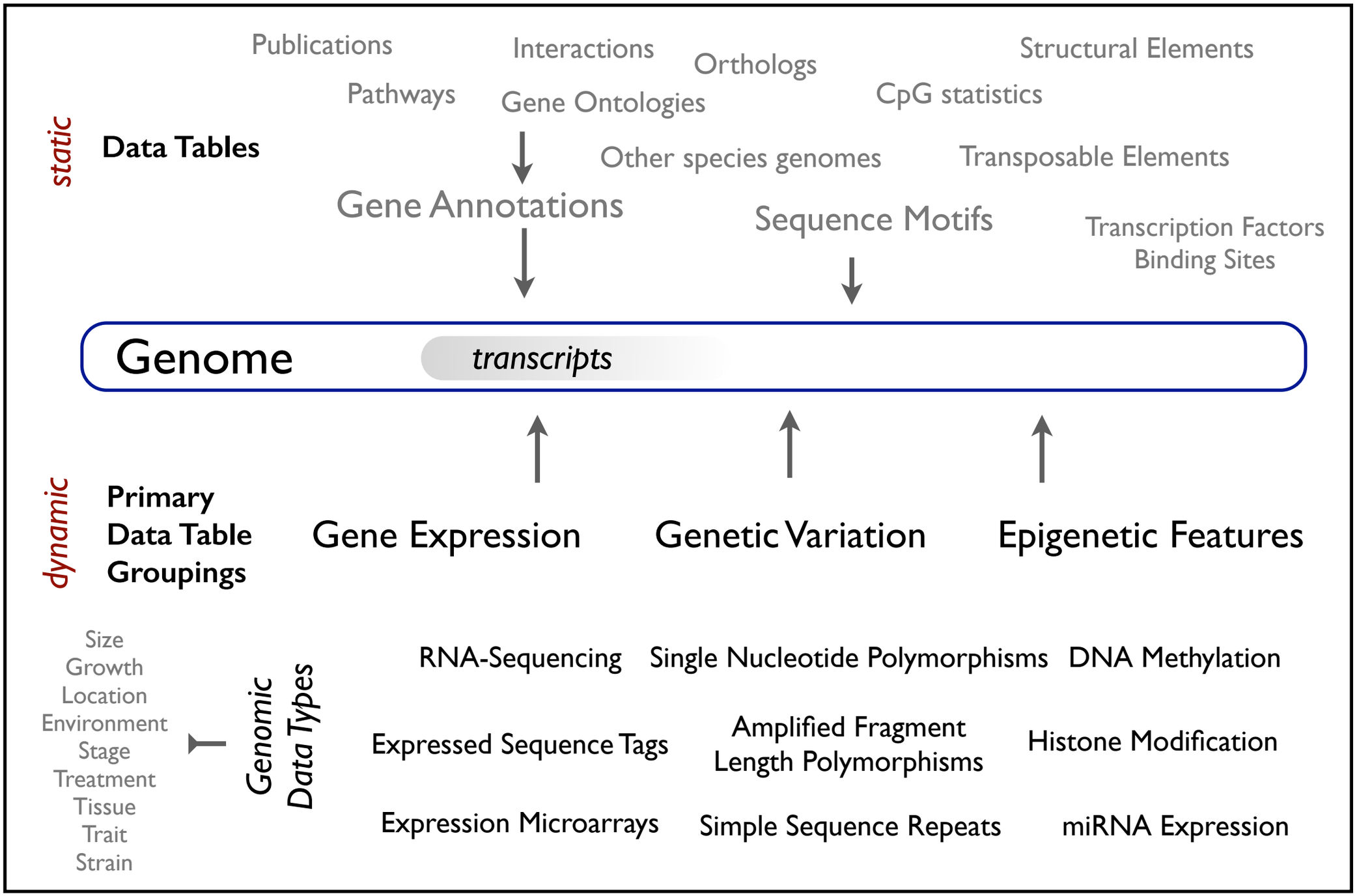
qDOD_scaffol... 35
oyster_v9_mR... 28
oyster_v9_CD... 23
BiGO_betty_pl... 16
qDOD_Cgigas... 16

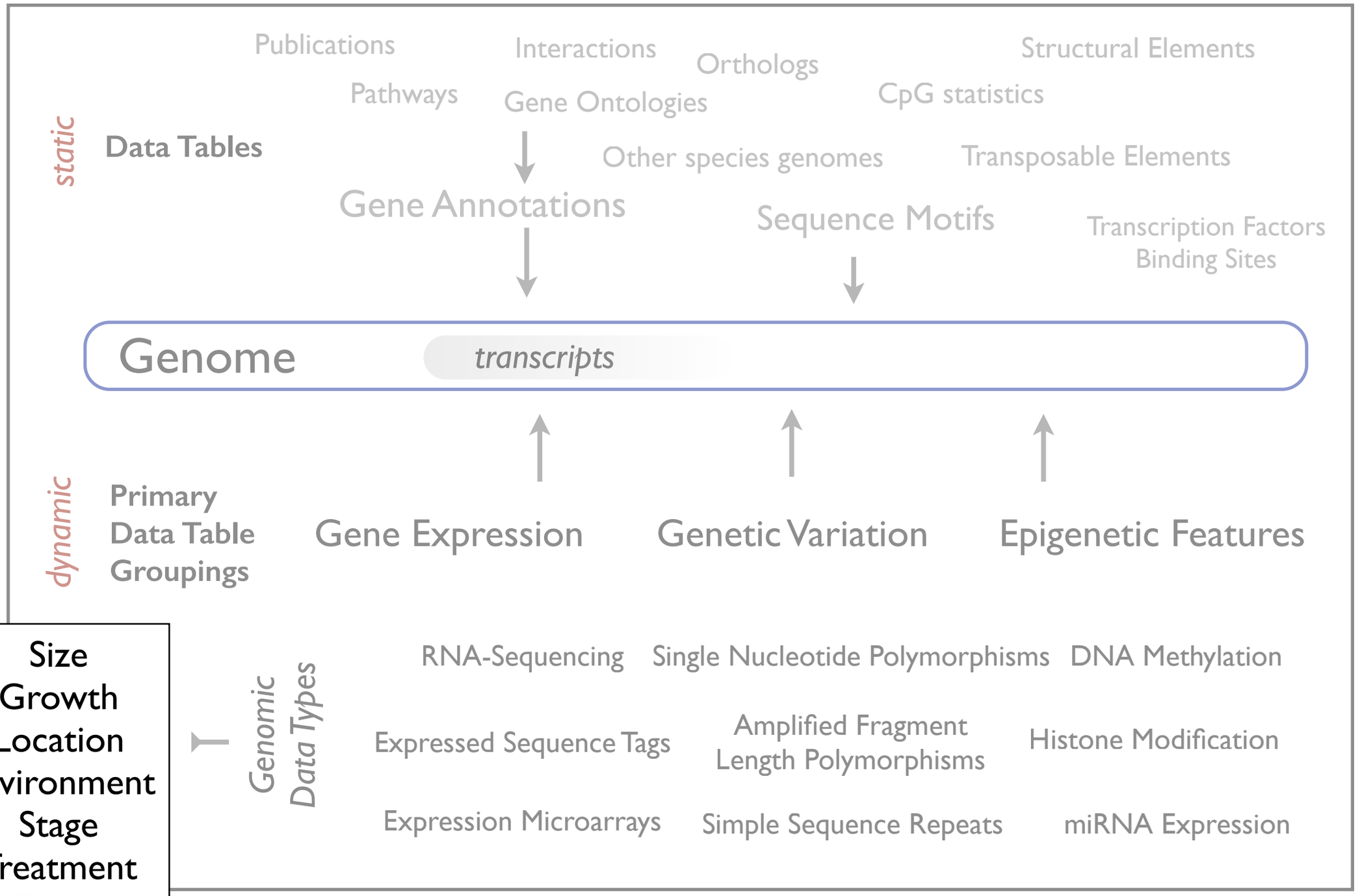
POPULAR TAGS

protein 47
oyster 45
proteomics 21
bioinformatics 21
oa 20

Tagged Datasets | qdod

Name	Sharing /
BiGO_Methylation_oysterv9_GFF GFF format file with percent methylation (score) information for bsmap gff gonad methylation oyster qdod sperm	sr320
qDOD_gene_length Sequence lengths of all genes in the oyster genome. This only includes trans fasta gene length oyster qdod	sr320
BiGill_methratio_v9_A.txt Methylation ratio information (MBD-Seq) from oyster gill tissue. Combir bsmap epigenetic gill mbd methylation oyster qdod	sr320
qDOD_Protein_Sequences Amino acid sequence for all proteins (v9) fasta oyster protein qdod	sr320
qDOD_oyster_gene_exon_number Number of exons for all genes in oyster genome (v9) exon oyster qdod	sr320
Zhang_etal_SuppTable14 S14. Transcriptomic representation of genes (RPKM) at different develc oyster qdod zhang	sr320
qDOD_Cgigas_gene_fasta Tabular format of Cgigas gene sequence fasta file Derived using Datas fasta oyster qdod	sr320
qDOD_Zhang_Gil_gene_RNA-seq Paired end Gill RNA-Seq data from Zhang et al 2012. Exported gill oyster qdod rna-seq zhang	sr320
Mgo_RNAseq_transcript Paired end Male Gonad RNA-Seq data from Zhang et al 2012 Exported f oyster qdod rna-seq zhang	sr320
qDOD_Zhang_Mgo_gene_RNA-seq Paired end Male Gonad RNA-Seq data from Zhang et al 2012 gonad oyster qdod rna-seq sperm zhang	sr320


Publications
Interactions
Orthologs
Structural Elements
Pathways
Gene Ontologies
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
Single Nucleotide Polymorphisms
DNA Methylation
Expressed Sequence Tags
Amplified Fragment
Length Polymorphisms
Histone Modification
Expression Microarrays
Simple Sequence Repeats
miRNA Expression


static
Data Tables
Publications
Interactions
Orthologs
Structural Elements
Pathways
Gene Ontologies
CpG statistics
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
Genomic
Data Types
RNA-Sequencing
Single Nucleotide Polymorphisms
DNA Methylation
Expressed Sequence Tags
Amplified Fragment
Length Polymorphisms
Histone Modification
Expression Microarrays
Simple Sequence Repeats
miRNA Expression
Size
Growth
Location
Environment
Stage
Treatment
Tissue
Trait
Strain

```
SELECT cgslim.CGI_ID,Description,evalue,SPID,GOID,term,GOSlim_bin,sequence
  FROM [sr320@washington.edu].[qDOD_Cgigas_GO_GOslim] cgslim

 LEFT JOIN [sr320@washington.edu].[qDOD_Cgigas_gene_fasta] cgf
  on cgslim.CGI_ID = cgf.CGI_ID

  Where term LIKE '%methyl%'
  OR
  term LIKE '%histone%'
```

What genes are associated with epigenetics?

What genes are associated with immune response?

Which genetic markers are associated with

fast growth?

resilience?

disease tolerance?

Publications

Interactions

Orthologs

Structural Elements

Pathways

Gene Ontologies

CpG statistics

static

Data Tables

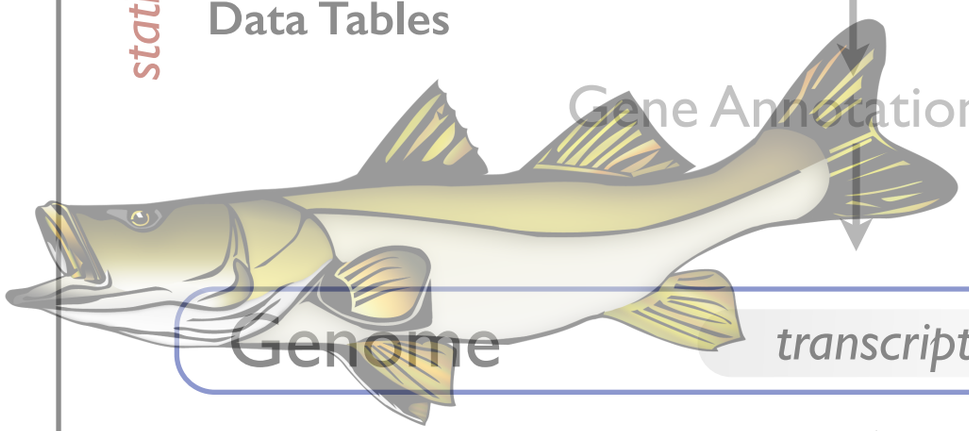
Other species genomes

Transposable Elements

Gene Annotations

Sequence Motifs

Transcription Factors
Binding Sites

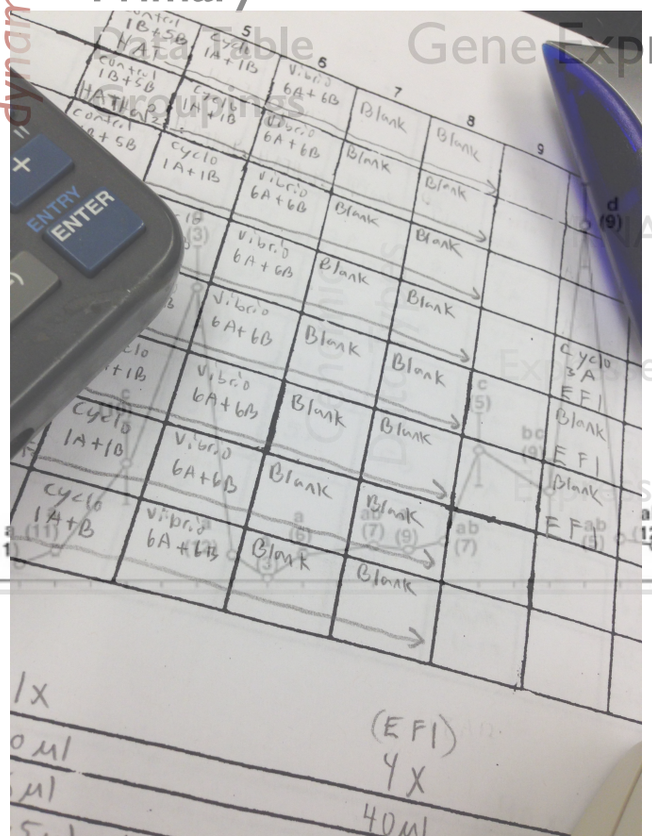


Genome

transcripts

dynamic

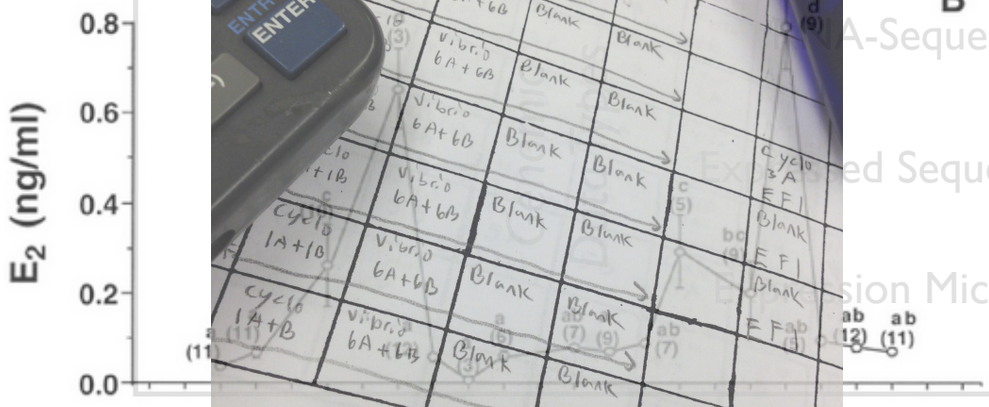
Primary



Gene Expression

Genetic Variation

Epigenetic Features



B

A-Sequencing

Single Nucleotide Polymorphisms

DNA Methylation

e Tags

Amplified Fragment
Length Polymorphisms

Histone Modification

rrays

Simple Sequence Repeats

miRNA Expression

Publications

Interactions

Orthologs

Structural Elements

Pathways

Gene Ontologies

CpG statistics

static

Data Tables

Other species genomes

Transposable Elements

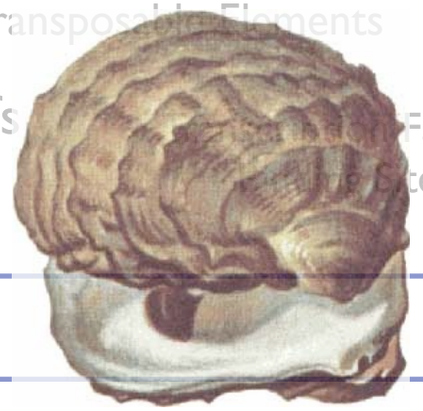
Gene Annotations

Sequence Motifs

Factors

tes

Web platform for collaboration, analyses, and sharing is essential



Genome

Statistics

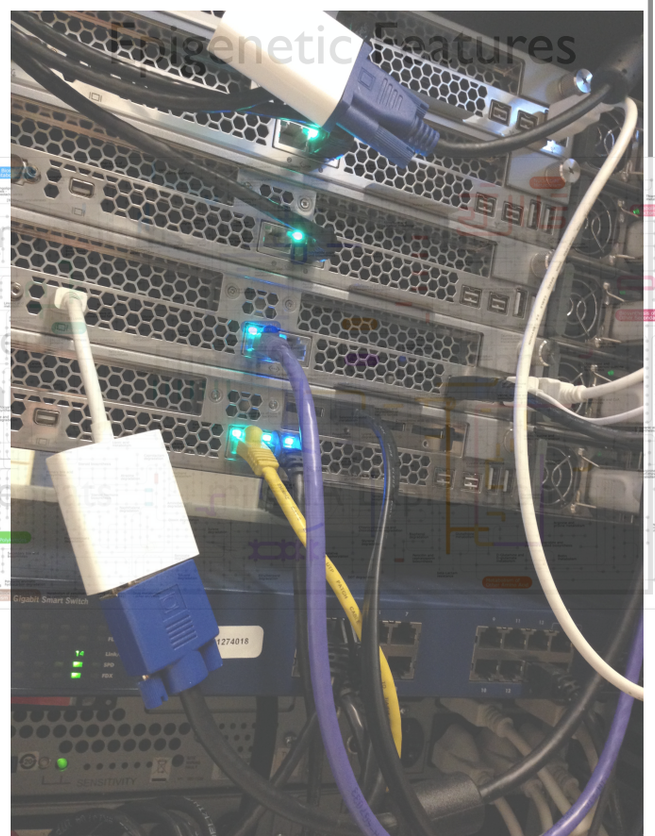
dynamic

Data Table Groupings

Gene Expression

Genetic Variation

Epigenetic Features



Genomic Data Types

RNA-Sequencing

Single Nucleotide Polym

Expressed Sequence Tags

Amplified Fragme
Length Polymorph

Expression Microarrays

Simple Sequence Re

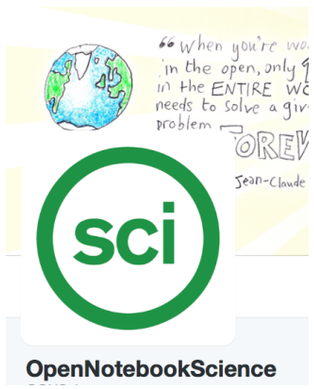
Open Methods

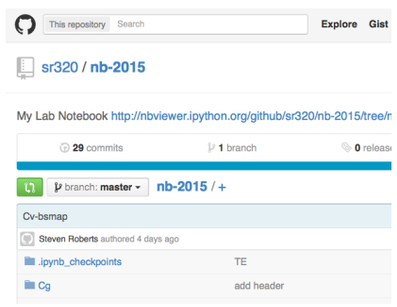
Open Methods

reproducibility

Data Acquisition and Analysis

Open Notebook Science





http://genefish.wikispaces.com/

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

How

open notebooks

-wiki

-evernote

-IPython

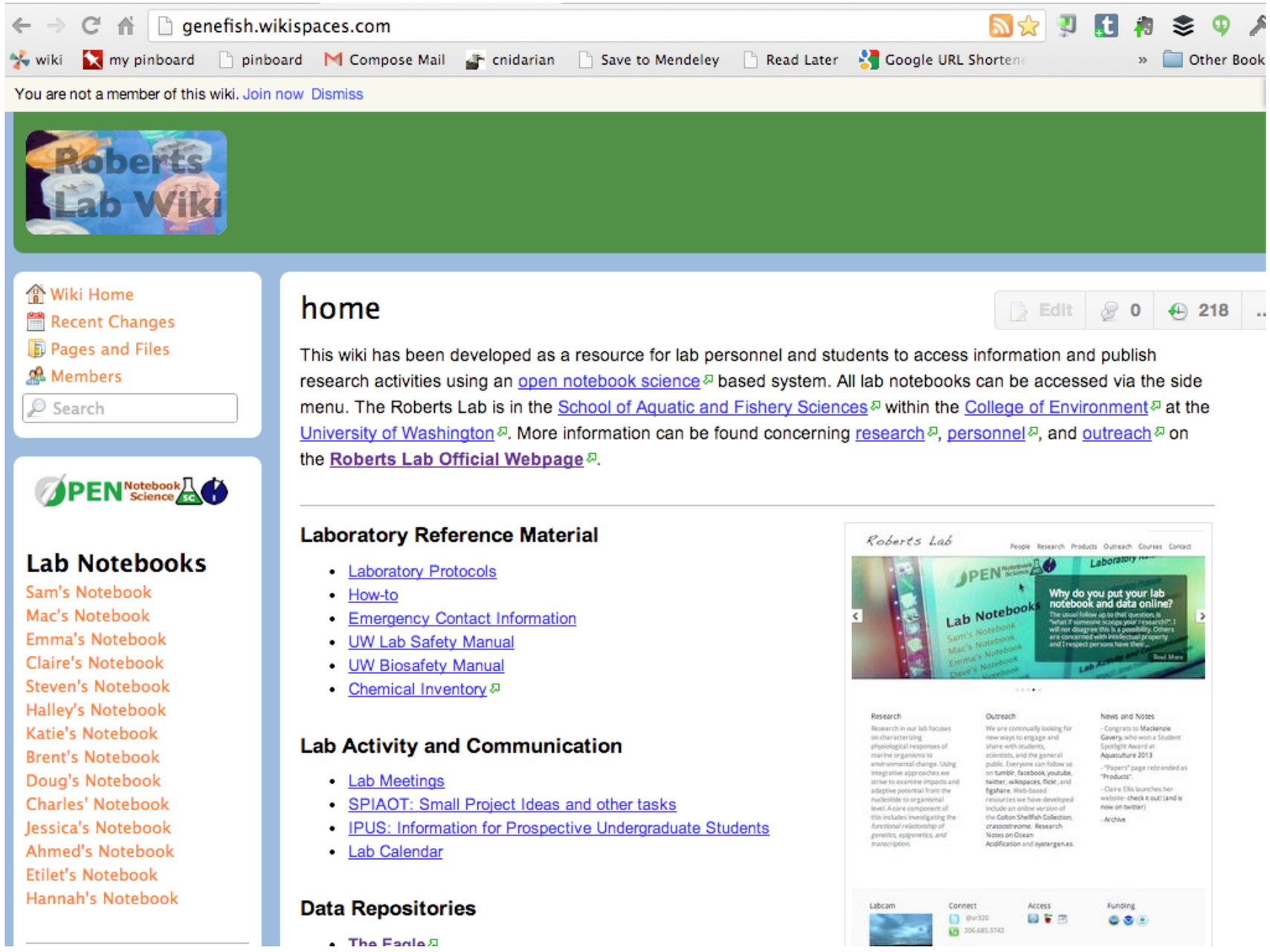
How

open notebooks

-wiki

-evernote

-IPython



How

open notebooks

-wiki

-evernote

-IPython

genefish.wikispaces.com

wiki | my pinboard | pinboard | Compose Mail | cnidarian | Save to Mendeley | Read Later | Google URL Shorten | Other Book

You are not a member of this wiki. Join now Dismiss

Roberts Lab Wiki

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

Search

OPEN Notebook Science

Lab Notebooks

Sam's Notebook
Mac's Notebook
Emma's Notebook
Claire's Notebook

home

Edit | 0 | 218

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

- Laboratory Protocols
- How-to
- Emergency Contact Information
- UW Lab Safety Manual
- UW Biosafety Manual
- Chemical Inventory

Lab Activity and Communication

- Lab Meetings
- SPIAOT: Small Project Ideas and other tasks
- IPUS: Information for Prospective Undergraduate Students
- Lab Calendar

Data Repositories

- The Eagle

Wiki Home
Recent Changes
Pages and Files
Members

Search

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
Ahmed's Notebook
Etilet's Notebook
Hannah's Notebook

Featured Pages
crassostreome

More Pages

Sam's Notebook

20130729

DNA Isolation - Claire's C.gigas Development
Isolated

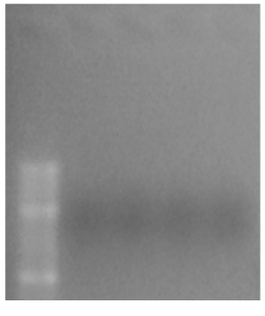
20130409

PCR - Hexokinase Partial CDS
Performed PCR using the primers CG_HK_CDS_2132-2158 (SRID: 1521) and Cg_Hk_CDS C.gigas cDNA (from DATE).

Master mix calcs and cycling params are here.

Samples were run in duplicate.

Results:



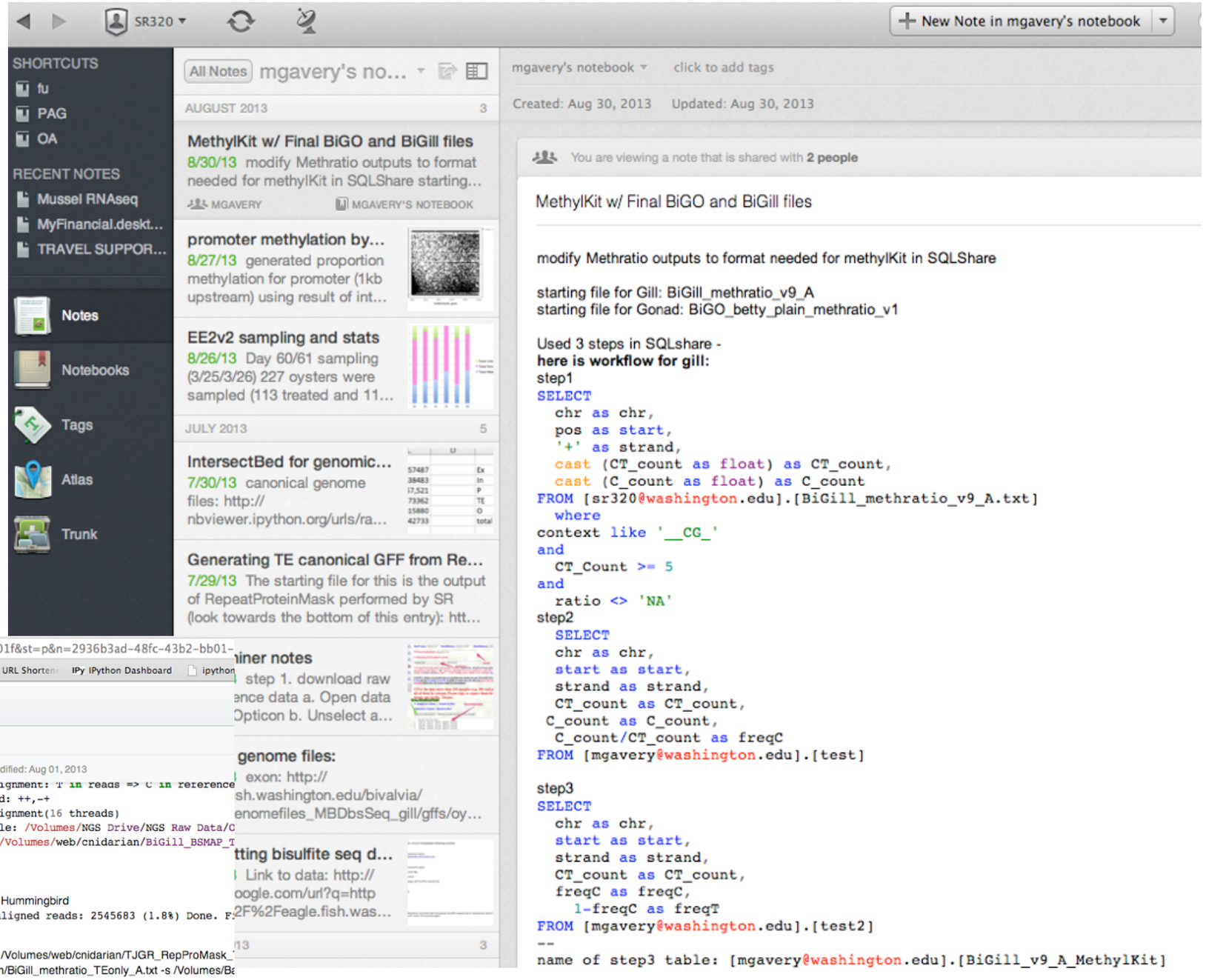
How

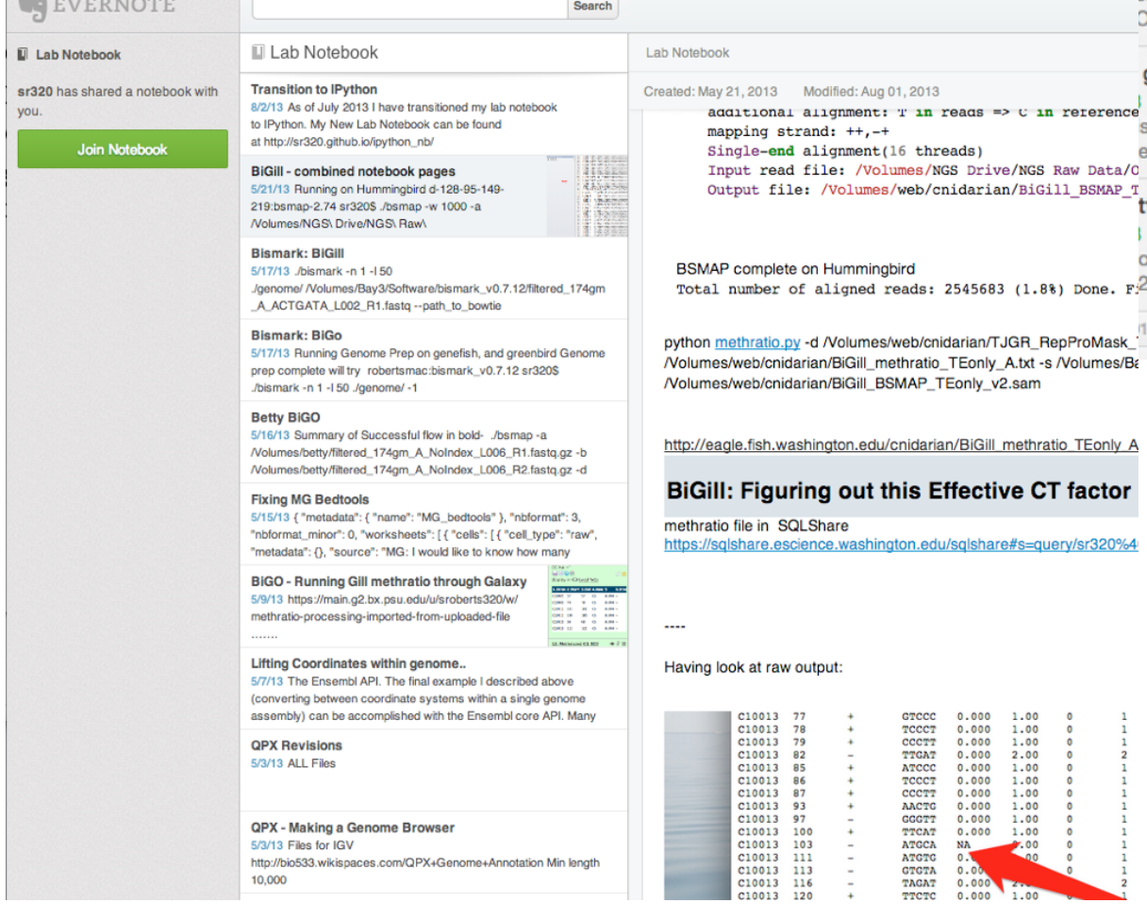
open notebooks

-wiki

-evernote

-IPython





How

open notebooks

-wiki

-evernote

-**IPython***

IP[y]: Notebook BiGo_methratio Last saved: Jul 30

File Edit View Insert Cell Kernel Help

Markdown

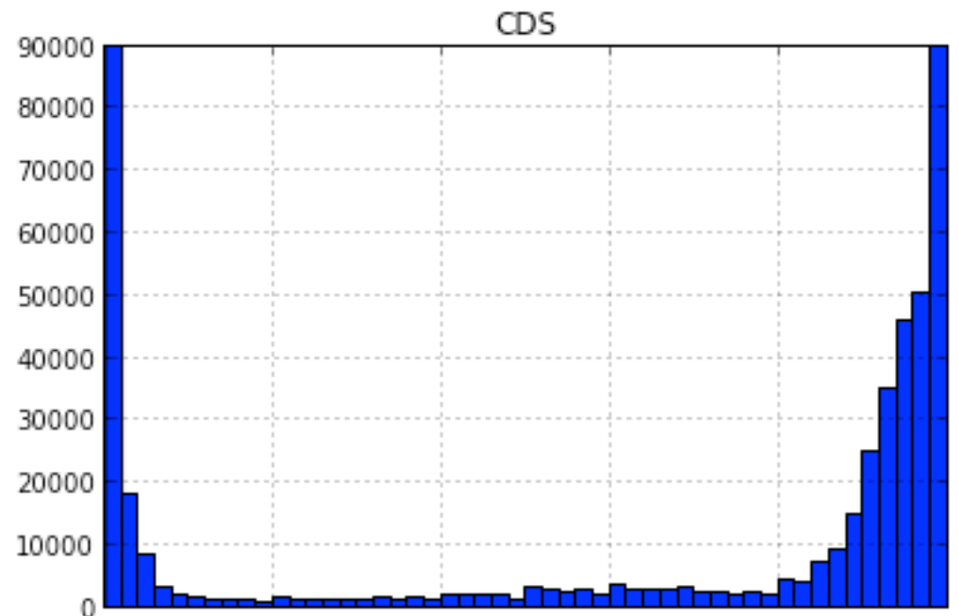
DNA methylation of Oyster Sperm based on

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

In [38]:
```
from pandas import *

# read data from data file into a pandas DataFrame
CDSmr = read_table("/Volumes/web/cnidarian/BiGo_methratio_boop_
                   #sep="\t", # what character separates each column?
                   #na_values=["", " "], # what values should be consi
                   header=None)
```

```
CDSmr[5].hist(bins=50);
#Axis limits are changed using the axis([xmin, xmax, ymin, ymax
plt.axis([0, 1, 0, 90000]);
plt.title('CDS');
```



IP[y]: Notebook TJGR_OysterGenome_IGV Last saved: Jul 31 5:34 AM

File Edit View Insert Cell Kernel Help

Markdown

In [44]:
```
!sed 's/mRNA/promoter/g' </Volumes/web/cnidarian/TJGR_Promoter_1k5p.gff> /Volumes/web
```

In [45]:
```
!head /Volumes/web/cnidarian/TJGR_Promoter_1k5p_b.gff
```

```
C16582  GLEAN   promoter        386     395     0.555898        -       .       ID=C
C17212  GLEAN   promoter        1       30      0.999572        +       .       ID=C
C17316  GLEAN   promoter        1       29      0.555898        +       .       ID=C
C17476  GLEAN   promoter        258     491     0.998947        -       .       ID=C
C17998  GLEAN   promoter        388     559     1       -       .       ID=CGI_10000
C18346  GLEAN   promoter        1       173     1       +       .       ID=CGI_10000
C18428  GLEAN   promoter        547     611     0.555898        -       .       ID=C
C18964  GLEAN   promoter        659     714     0.999572        -       .       ID=C
                                                                                ID=C
                                                                                ID=C
```

In [ ]:
```
#join in SQLShare
```

In [45]:
```
!python /Users/sr320/sqlshare-pythonclient/tools/singleupload.py -d OA_enrich2 /
```

```
processing chunk line 0 to 51 (0.00206899642944 s elapsed)
pushing /Volumes/web/cnidarian/mod_chart_B1049AF0BD891379525818063.txt...
parsing C8BE2D77...
finished OA_enrich2
```

In [46]:
```
!python /Users/sr320/sqlshare-pythonclient/tools/fetchdata.py -s "SELECT * FROM
```

In [47]:
```
!head /Volumes/web/cnidarian/OA_enrich2_join_SPID.csv
```

```
Category,Term,Count,%,PValue,Genes,List Total,Pop Hits,Pop Total,Fold Enrichment
,evalue,Description
GOTERM_BP_FAT,GO:0045449~regulation of transcription,8,8.98876404494382,0.092168
7272727,1,0.998517222392313,76.9925811614,CGI_10000058,Q2KJJ0,6E-16,Pre-mRNA-pro
GOTERM_BP_FAT,GO:0045216~cell-cell junction organization,4,4.49438202247191,0.00
```

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
ATTTACAATACGAAGTAAAATAGATAACGTGAAAATAATCTTGGTGCTGGATGATCGATC
AAGTTCACCAATATTTTATTGTAAAAAATCATTCTAAACAGCATGAAATCGTGTACAATG
TATAAACAAGCAAATATATAACACTAAAGCAAGAGGGCGTAAGTGGGGGGGTGGGTGAGA
GTAAAAAATTCAAACATGTCAAATACCCCGGCGTTAGCCTTAAAAGCACCATGGACTTCT
```

Fasta2Slim

This IPython notebook is ructured means to annotate sequences using UniProt/SwissPr
personal preferences. As equires the user has the following software installed ...

- IPython
- NCBI Blast
- SQLShare Python Cl

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
ATTTACAATACGAAGTAAAATAGATAACGTGAAAATAATCTTGGTGCTGGATGATCGATC
AAGTTCACCAATATTTTATTGTAAAAAATCATTCTAAACAGCATGAAATCGTGTACAATG
TATAAACAAGCAAATATATAACACTAAAGCAAGAGGGCGTAAGTGGGGGGGTGGGTGAGA
GTAAAAAATTCAAACATGTCAAATACCCCGGCGTTAGCCTTAAAAGCACCATGGACTTCT
GCCTTCAATAAGCATAAAATTAAAACACCTAATACACAATGAATATACAGATAAAACAGA
TTTATGAATAGTTGGTGTTACATCTTTTACAGCCATAAGCCTTCATTTTGCTTCCAAACG
TATAAAATCTGACTTGGAACAATATACAGCCATGAGATATGACACAGCGAGCACTACAAT
ATATATTTATCTTGTACTATACAGCCTGTACAAGAAAATTCTGGAATTGTCTTCACAAGA
GACAGAAAAATAGTTGCAATGTGAATGCTAGTCTACTATTTGATCACAATTGGATAGAAA
```

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}].[_b
last_sprot]md left join [samwhite@washington.edu].[UniprotProtNamesReviewe
d_yes20130610]sp on md.Column3=sp.SPID left join [sr320@washington.edu].[S
PID and GO Numbers]go on md.Column3=go.SPID left join [sr320@washington.ed
u].[GO_to_GOslim]slim on go.GOID=slim.GO_id where aspect like 'P'" \
-f tsv \
-o {wd}GOdescriptions.txt
```

Plot GoSlim terms

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

```
Populating the interactive namespace from numpy and matplotlib
```

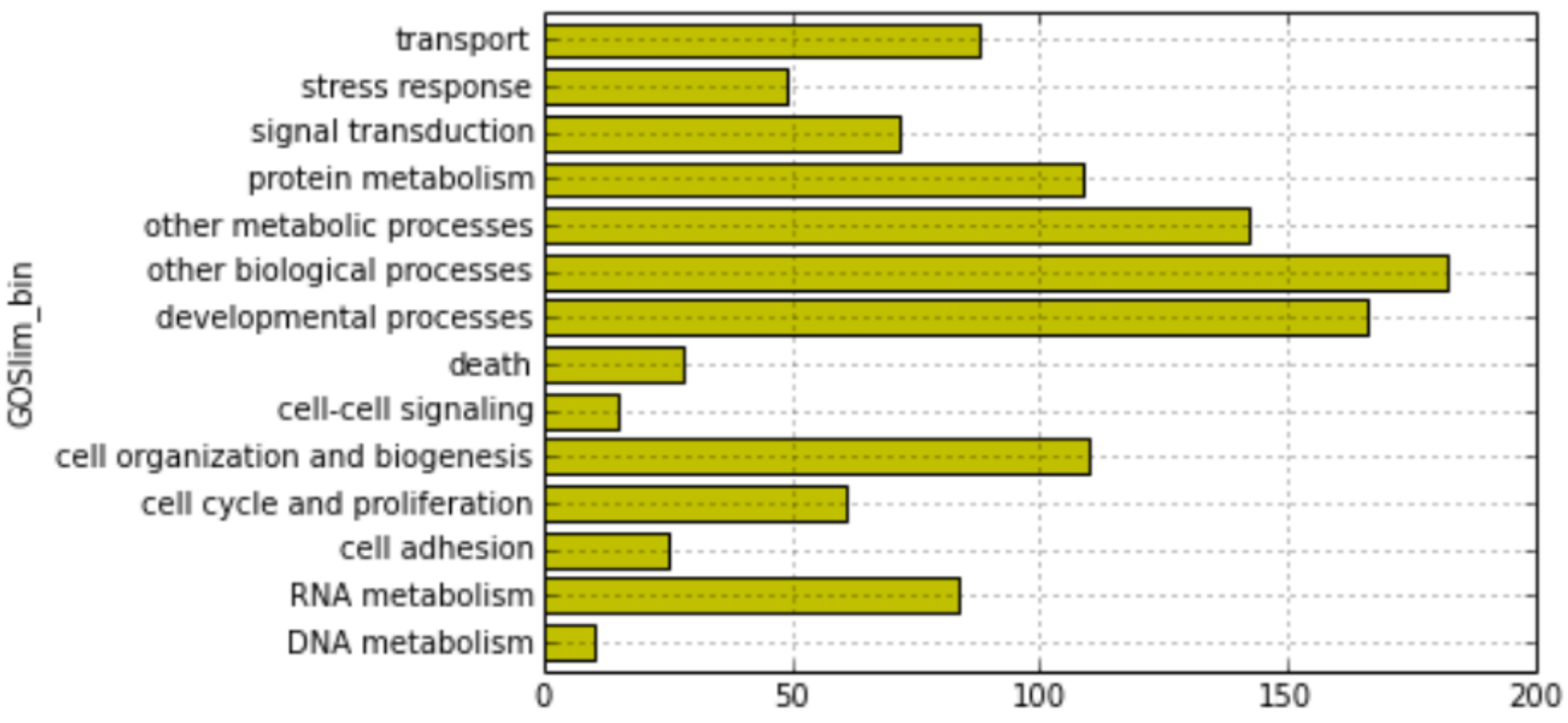
In [11]:
```
cd {wd}
```

```
/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')
```



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',
)
savefig('GOSlim.png', bbox_inches='tight')
```


This repository Search

che625 / olson-ms-nb

github SOCIAL CODING

IPython Notebook for downloading and analyzing data the manusript: "Indication of family-specific DNA methylation patterns in developing oysters"

bioRix preprint - http://dx.doi.org/10.1101/012831

To execute the IPython Notebook in its entirety you will need:

- IPython - install instructions
- BSMAP - install instructions
- bedtools - install instructions
- R - install instructions
- rpy2 (interface to R from Python) - install instructions

Plot GoSlim terms

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

```
Populating the interactive namespace from numpy and matplotlib
```

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

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

README.md

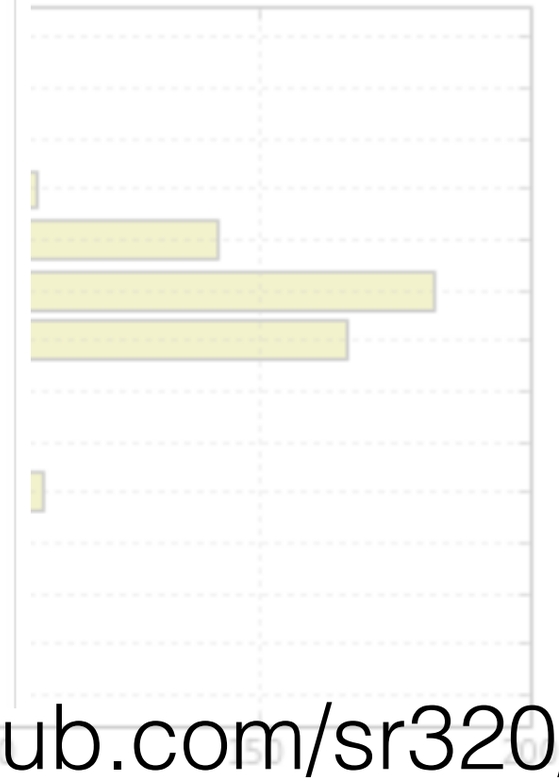
IPython Notebooks and data supplemental to the manuscript: "*Up in arms: Immune and nervous system response to sea star wasting disease*"

The repository includes IPython notebooks (.ipynb file) that can be downloaded locally and interactively executed. The code in the IPython notebook **eimd_analysis.ipynb** will process data such that figures in the manuscript are reproduced (in theory).

Description of Files

- `eimd_analysis.ipynb` - IPython notebok that can be interactively executed locally or viewed online designed so that user can **replicate all analysis**. Requires several dependancies (see below). nbviewer version
- `eimd_data-only.ipynb` - IPython notebok that can be interactively executed locally or viewed online designed so that user can **simply explore** data files. *Only requires IPython*. nbviewer version
- `data/Phel_transcriptome.fasta` - P hel coelocytes transcriptome. Contains xxxx contigs from de novo assembly.
- `data/Phel_countdata.txt` - Tab-delimited text file with read count data from 6 P hel RNA-seq libraries, 3 treated and 3 control libraries.

t(kind='barh',



github
SOCIAL CODING

github.com/sr320/eimd-sswd

Open Science Philosophy

Transparency with limited effort

will try just about anything

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

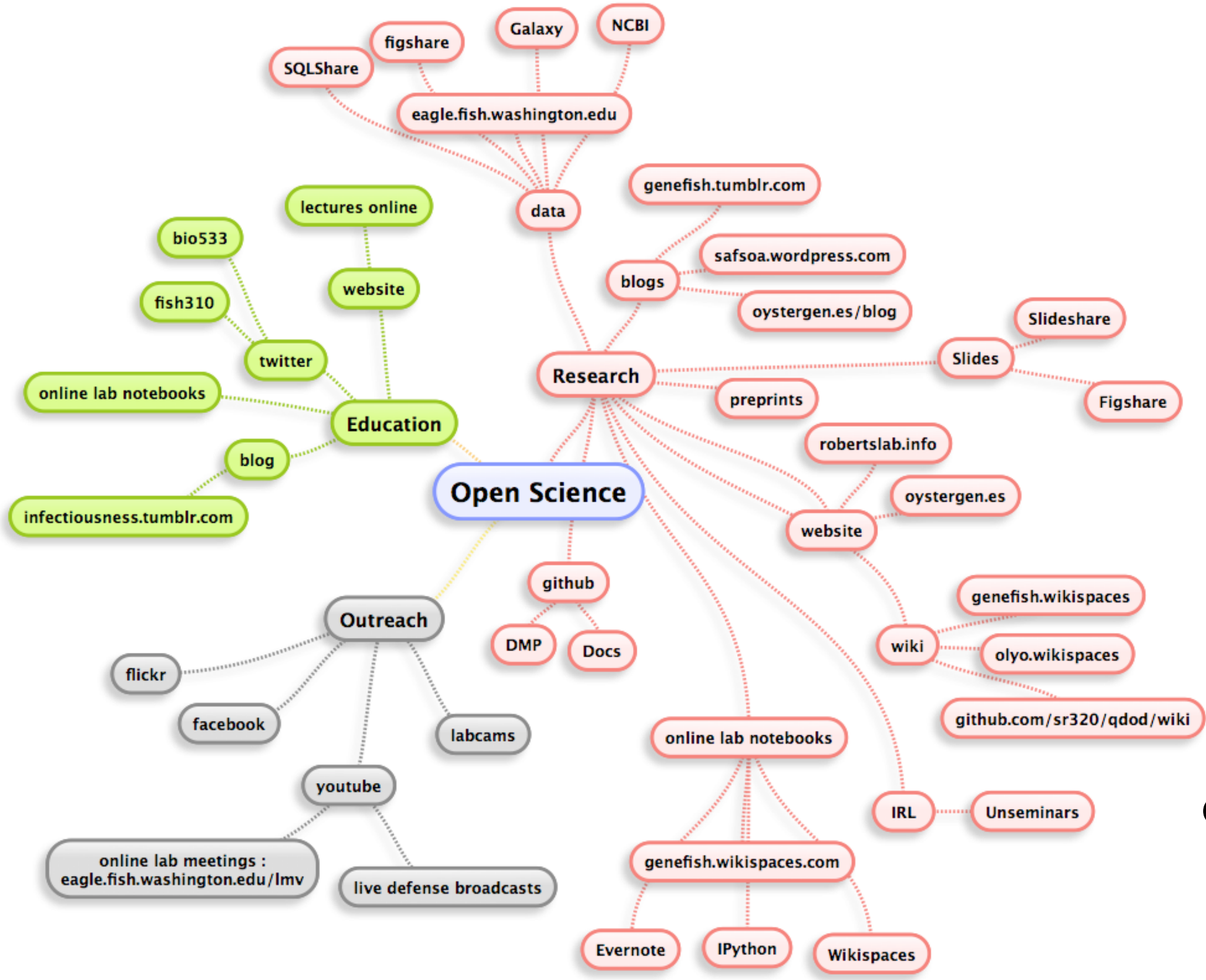
Rationale

Platforms

Open Science

Data

everything else...



Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science

Data

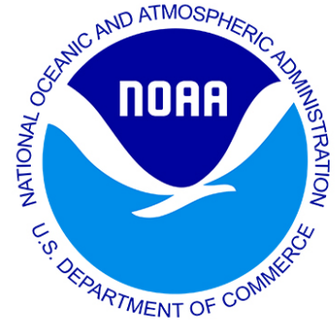
everything else...

start them young

Acknowledgements

Mackenzie Gavery
Claire Olson
Sam White
Brent Vadopalas
Jake Heare
Jay Dimond

Bill Howe
Dan Halperin



slides, data & more @
github.com/sr320/talk-osu-2015