Genomic analysis in non-model organisms

2013 Workshop on Genomics
Český Krumlov

Bill Cresko
Institute of Ecology and Evolution
Department of Biology
University of Oregon
Outline for today’s lecture

Genomic data and non-model organism research

RAD-seq for ecological & evolutionary genomics

Genomically enabling a non-model organism

Stacks software pipeline
Modified from Gilbert 1998
Model organism research has been very important

Vertebrate zygotes or embryos

28 day human  19h zebrafish

Video by Don Kane
Model organism research has been very important.
Studying brain cancer using somatic evolutionary genomics in a model organism

Hui Zong, Rui Galvao, Julian Catchen and Susie Bassham
Laser Capture Microdissection of cells
Transcriptomic and genomic analysis of cells

Sequence cells here...

... and here
Transcriptomic and genomic analysis of cells

Sequence cells here...

... and here
Multiple lines of genetic evidence for causative mutations

Gene Mutations

1. wild type
2. pre-cancerous
3. pre-cancerous
4. tumor
5. tumor

 Gene Mutations

- p16 mutation
- p53 mutation
- Gene?? mutation

Gene Expression

- WT
- p16
- Gene??
Genomic rearrangements in cancer cells
How can modern genomics improve studies of non-model organisms??
How do the major differences among lineages evolve?
How are organisms related to one another?

Phylogeography of pocket gophers using mtDNA

Avise, 1979
How do organisms adapt to novel environments?

Four fundamental processes in evolution

Origin of genetic variation;
- mutation
- migration

Sorting of variation;
- genetic drift
- natural selection
Genetic drift is a null model

Threespine stickleback fish (Gasterosteus aculeatus)
Population genomics

Simultaneous genotyping of **neutral** and **adaptive** loci

Genome-wide background provides more precise estimates:
- Demographic processes (e.g. $N_e$)
- Phylogeography

Outliers from background indicate:
- Selective sweeps
- Local adaptation
Population genomics of unordered markers

Population genomics of ordered markers

Genomic architecture:
- Distribution adaptive variation across the genome
- Correlations among genomic regions (linkage disequilibrium)
- Interactions among genomic regions (e.g. epistasis)
- Recombination rates and chromosomal inversions

How do we ‘genomically enable’ research on non-model organisms?

1. Genetic Markers & Maps
2. Physical Maps
3. Transcriptomes
4. Gene Expression Analyses
In the field and in the lab until a few years ago....
Next generation sequencing, high performance computing and new analytical approaches have fundamentally changed the scope of studies of non-model organisms.
Should we just sequence everything?
Why not sequence the entire genome??

• Still prohibitively expensive for many studies
  • Human height GWAS; over 15,000 individuals assayed
  • Identified many new regions contributing to the variation
  • Still only identified a fraction of the heritability

• For many studies a full sequence isn’t necessary
  • the genomes of many organisms are organized in linkage blocks
  • well spaced markers will provide the necessary coverage
  • the cost of genotyping will almost always be a fraction of full sequencing

• Genetic maps are very useful in genomic studies
  • a high density genetic map can facilitate genome assembly
  • genomes may be segregating a lot of structural variation
Alternative approach - Reduced representation NGS for genotyping

• Focus the sequencing on a homologous set of tags spread throughout the genome

• Can lead to the simultaneous identification and typing of single nucleotide polymorphisms (SNPs)

• The cost will always be a fraction of the cost of resequencing the genome
  • i.e. 1% genome coverage will be less than 1% the cost
  • often the coverage is more even than whole genome sequencing

• Can allow thousands of genomes to be assayed in just a few weeks

• WHY NOT - some cases complete genomic sequence is necessary
  • when linkage disequilibrium blocks (LD) are very short
  • Inferring patterns of LD may be easiest with full sequences
Different flavors of Reduced Representation Library (RRL) Sequencing for genotyping

• Common acronyms
  • **RRL** - Reduced Representation Library
  • **GBS** - Genotyping By Sequencing
  • **CRoPS** - Complexity Reduction of Polymorphic Sequences
  • **MSG** - Multiplex Shotgun Genotyping
  • **RAD** - Restriction site Associated DNA

• All rely on restriction enzyme digestion

• RRL, CRoPS, MSG and GBS use one or two restriction enzymes only

• RAD uses an extra shearing step to capture all restriction sites

• Incorporation of barcodes on adaptors for multiplexing

• Aligned against a reference genome or assembled *de novo*

• Statistical issues
  • new level of sampling variation (sequencing in addition to biological)
  • sequencing error and problems for aligning or clustering
What is RAD-seq?
(Restriction-site Associated DNA)

2007
Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers
Michael R. Miller,¹ Joseph P. Dunham,² Angel Amores,³ William A. Cresko,² and Eric A. Johnson¹,
¹Institute for Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA, ²Center for Ecology & Evolutionary Biology, University of Oregon, Eugene, Oregon 97403, USA, ³Institute of Neuroinformatics, University of Oregon, Eugene, Oregon 97403, USA.

2008
Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers
Nathan A. Baird¹,², Paul D. Etter¹,³, Tressa S. Atwood¹,², Mark C. Currey¹,², Anthony L. Shiver¹, Zachary A. Lewis¹, Eric U. Selsker¹, William A. Cresko³, Eric A. Johnson¹.
¹Institute for Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA, ²Center for Ecology & Evolutionary Biology, University of Oregon, Eugene, Oregon 97403, USA, ³Institute of Neuroinformatics, University of Oregon, Eugene, Oregon 97403, USA.
What is RAD-seq?
*(Restriction-site Associated DNA)*

Chr I

Chr II

Chr III

SbfI
What is RAD-seq?

(Restriction-site Associated DNA)
What is RAD-seq?

(Restriction-site Associated DNA)

What is RAD-seq?

(Restriction-site Associated DNA)
What is RAD-seq?
(Restriction-site Associated DNA)

22,830 SbfI sites in threespine stickleback

~ 45,000 RAD-Tags

HiSeq Illumina Lane:
160 million reads, 96 barcoded individuals
Restriction Enzyme (RE) digestion and first adaptor ligation

<table>
<thead>
<tr>
<th>0.0</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
<th>5.0kb</th>
<th>6.0</th>
<th>7.0</th>
<th>8.0</th>
<th>9.0</th>
<th>10.0</th>
</tr>
</thead>
</table>
Restriction Enzyme (RE) digestion and first adaptor ligation
Restriction Enzyme (RE) digestion and first adaptor ligation
Restriction Enzyme (RE) digestion and first adaptor ligation
Restriction Enzyme (RE) digestion and first adaptor ligation

A = Amplification primer
B = Sequencing primer
C = Barcode
Shearing and second adaptor ligation

A = Amplification primer
B = Sequencing primer
C = Barcode
Shearing and second adaptor ligation

A = Amplification primer
B = Sequencing primer
C = Barcode
Shearing and second adaptor ligation

A = Amplification primer
B = Sequencing primer
C = Barcode
Shearing makes consistent fragments for sequencing

A = Amplification primer
B = Sequencing primer
C = Barcode
Single (GBS) or Double Digest RAD (ddRAD)

A = Amplification primer
B = Sequencing primer
C = Barcode
Size selection is more problematic without shearing

A = Amplification primer
B = Sequencing primer
C = Barcode
A = Amplification primer
B = Sequencing primer
C = Barcode

2bRAD - type 2b restriction enzyme
2bRAD - can scale number of markers easily

A = Amplification primer
B = Sequencing primer
C = Barcode
2bRAD - size selection is difficult

A = Amplification primer
B = Sequencing primer
C = Barcode
## Summary of plusses and minuses of RAD family

<table>
<thead>
<tr>
<th></th>
<th>Sheared RAD</th>
<th>Single or ddRAD</th>
<th>2b-RAD</th>
</tr>
</thead>
</table>
| **plusses** | - Consistent reads  
- Local assemblies  
- Identify PCR duplicates | - Fewer steps  
- Easier marker scaling | - Fewest steps  
- Easiest marker scaling |
| **minuses** | - Shearing step  
- Scaling requires different enzymes | - Multiple enzymes  
- Poor consistency  
- PCR duplicates | - Very short reads  
- PCR duplicates |

- **Sheared RAD**: Consistent reads, Local assemblies, Identify PCR duplicates
- **Single or ddRAD**: Fewer steps, Easier marker scaling
- **2b-RAD**: Fewest steps, Easiest marker scaling
Benefits of random shearing in RAD

A) Restriction sites in genome

RAD tag sequence read

Sheared-end reads

B) Variable length RAD fragments isolated

200-1200bp in length

C) Contigs assembled from the sheared-end reads for each RAD tag
Acquire paired-end sequence

Match to marker catalog

Collate/Assemble PE reads

Associate markers / PE contigs with ESTs

Assign orthology to:
- markers
- PE contigs
- ESTs

Human Genome

Zebrafish Genome

EST Library
Random shearing benefits in RAD

Eliminating PCR duplicates:
Random shearing benefits in RAD

Eliminating PCR duplicates:
Considerations for RAD-seq studies
Experimental design considerations for RAD

Tradeoffs:
**Number** of sites versus **Depth** of sequencing per site versus **Number of samples**
Experimental design considerations for RAD

Tradeoffs:
- **Number** of sites versus **Depth** of sequencing per site versus **Number of samples**
Experimental design considerations for RAD

Tradeoffs:

**Number** of sites versus **Depth** of sequencing per site versus **Number of samples**

raw reads / samples / sites = coverage at each RAD locus

$$1,000,000 / 100 / 1,000 = 10x \text{ coverage}$$

25 to 50x average coverage per RAD locus is a good goal
Differentiating SNPs from error

Restriction enzyme recognition site

Reference genome sequence

Sequence reads
Differentiating SNPs from error
Differentiating SNPs from error

The reads are 14 T and 2 G:

GT heterozygote?
GG homozygote with error?
AA homozygote with lots of error?

Needed a rigorous method to call genotypes
Differentiating SNPs from error

Maximum likelihood genotyping based on multinomial distribution of nucleotide reads

$L(n_{1\text{ hom}}) = P(n_1,n_2,n_3,n_4) = \frac{n!}{n_1!n_2!n_3!n_4!} \left(1 - \frac{3\varepsilon}{4}\right)^{n_1} \left(\frac{\varepsilon}{4}\right)^{n_2} \left(\frac{\varepsilon}{4}\right)^{n_3} \left(\frac{\varepsilon}{4}\right)^{n_4}$

$L(n_{1\text{ hetero}}) = P(n_1,n_2,n_3,n_4) = \frac{n!}{n_1!n_2!n_3!n_4!} \left(\frac{0.5 - \varepsilon}{4}\right)^{n_1} \left(\frac{0.5 - \varepsilon}{4}\right)^{n_2} \left(\frac{\varepsilon}{4}\right)^{n_3} \left(\frac{\varepsilon}{4}\right)^{n_4}$
Making statistics continuous across the genome

Kernel-smoothing average of summary statistics along genome

Bootstrap re-sampling to estimate significance of moving average
Overall pipeline
Experimental design considerations for RAD

Tradeoffs:
**Number** of sites versus **Depth** of sequencing per site versus **Number of samples**

How many tags do I need?

Things to consider

**Choice of enzyme and genome size**  \((0.25)^n \times \text{genome size} = \text{expected \# sites}\)

Genomes are biased:

<table>
<thead>
<tr>
<th>Expected Sites</th>
<th>Actual Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>112,300 six-cutter sites in stickleback (460 Mb)</td>
<td>EcoRI sites = 90,000</td>
</tr>
<tr>
<td>7000 eight-cutter sites in stickleback</td>
<td>SbfI sites = 22,800</td>
</tr>
<tr>
<td>32,900 six-cutter sites in <em>C. remanei</em> (135 Mb)</td>
<td>EcoRI sites = 73,200</td>
</tr>
</tbody>
</table>
Tradeoffs:
Number of sites versus Depth of sequencing per site versus Number of samples

How many tags do I need?

Things to consider

Choice of enzyme and genome size
Polymorphism and read length

Nucleotide polymorphism rate = 0.01 to 0.001 for most vertebrates

Stickleback populations: 0.01 to 0.02. At least 1 SNP every 100 bp, on average
Experimental design considerations for RAD

**Tradeoffs:**

*Number* of sites versus *Depth* of sequencing per site versus *Number of samples*

How many samples should be multiplexed?

**Things to consider**

**Barcoded adapters**

- 5 to 8nt barcodes
- Variable length barcodes
- Combinatorial barcodes (PE)

Barcode distance - two mismatches
Molecular considerations in library building

How many samples should be multiplexed?

Things to consider

DNA Quality
  Multiplex only like samples to help equalize representation of poor quality samples
Molecular considerations in library building

How many samples should be multiplexed?

Things to consider

DNA Quality

Diversify barcodes

Illumina cluster calling is confused by repetition in first 4 bases - can offset barcodes

<table>
<thead>
<tr>
<th>CGATA</th>
<th>GTACA</th>
<th>TAGCC</th>
<th>ACTGC</th>
</tr>
</thead>
</table>

[Image of gel electrophoresis showing the barcodes: CGATA, GTACA, TAGCC, ACTGC]
Molecular considerations in library building

How can I get the best depth of coverage?

Things to consider

Fragment size
Smaller/tighter is better
Molecular considerations in library building

How can I get the best depth of coverage?

Things to consider

- Fragment size
- Library quality
- qPCR

qPCR control should be similar to measured sample:
Molecular considerations in library building

How can I get the best depth of coverage?

Things to consider

Fragment size

**Library quality**
- qPCR

Pilot Experiment:
- Spike or split a lane
A case study of using RAD for an organism with a reference genome: population genomics of threespine stickleback fish

Susie Bassham, Julian Catchen, Paul Hohenlohe, Emily Lescak and Frank von Hippel
Threespine stickleback, *Gasterosteus aculeatus*

- **Ancestral Oceanic Populations**
  - Marine and Anadromous
  - Old (> 10 million years)
  - Phenotypically similar

- **Derived Freshwater Populations**
  - Lake and stream
  - Young (<15,000 years)
  - Phenotypically diverse
Threespine stickleback, *Gasterosteus aculeatus*

- **Ancestral Oceanic Populations**
  - Marine and Anadromous
  - Old (> 10 million years)
  - Phenotypically similar

- **Derived Freshwater Populations**
  - Lake and stream
  - Young (<15,000 years)
  - Phenotypically diverse

Rundle and McKinnon 2002
Threespine stickleback, *Gasterosteus aculeatus*

- **Ancestral Oceanic Populations**
  - Marine and Anadromous
  - Old (> 10 million years)
  - Phenotypically similar

- **Derived Freshwater Populations**
  - Lake and stream
  - Young (<15,000 years)
  - Phenotypically diverse

Pelvic Structure  Lateral plates

Cresko et al. 2004. PNAS
Colosimo et al. 2005. Science
Albert et al. 2008. Evolution
Miller et al. 2007. Cell
Chan et al. 2010. Nature
Signatures of natural selection across the genome

20 individuals in each of 5 popn’s
2 Ocean & 3 Freshwater
45,000 SNPs in each individual

Hohenlohe, Bassham et al. 2010. PLoS Genetics
What genomic regions are subject to selection during parallel evolution?
Parallel signatures of selection across the genomes

- Bear Paw
  - (mean $F_{ST} = 0.121$)

- Boot
  - (mean $F_{ST} = 0.112$)

- Mud
  - (mean $F_{ST} = 0.117$)

Genomic location (mBases)
Previously identify quantitative trait loci (QTLs) are under selection

Natural populations

Lateral plate major locus on LGIV (4000 SNPs)

- Eda
- Enigma
- Foxi3b
- FGFR
Extensive LD across the genome

Freshwater

Ocean

Hohenlohe et al. 2012. Philosophical Transactions of the Royal Society of London
Extensive LD across the genome
More in oceanic than in freshwater populations

Hohenlohe et al. 2012. Philosophical Transactions of the Royal Society of London
Could genome rearrangements in the stickleback genome be affecting these patterns?

Julian Catchen, Susie Bassham and Kate Ituarte
## Genome Assembly

![Genome Assembly Diagram](image)

<table>
<thead>
<tr>
<th>N50</th>
<th>17,417 bp</th>
<th>18,982 bp</th>
<th>15,555 bp</th>
<th>15,534 bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>199,905 bp</td>
<td>192,283 bp</td>
<td>238,768 bp</td>
<td>254,734 bp</td>
</tr>
<tr>
<td>Total</td>
<td>488.8 Mb</td>
<td>472.5 Mb</td>
<td>456.4 Mb</td>
<td>473.4 Mb</td>
</tr>
<tr>
<td>Median Coverage</td>
<td>24.6x</td>
<td>26.5x</td>
<td>24.1x</td>
<td>25.8x</td>
</tr>
</tbody>
</table>
Illumina Paired-end Reads

Reference Genome

- Male: 161,305,595 pairs
- Female: 144,396,898 pairs
- Male: 131,471,548 pairs
- Female: 150,786,462 pairs
**Reference Genome**

**Illumina Paired-end Reads**

- 161,305,595 pairs
- 144,396,898 pairs
- 131,471,548 pairs
- 150,786,462 pairs
F1 Pseudo-testcross

Male Parent

A1 B1 C1
A2 B1 C2

Female Parent

A1 B1 C1
A1 B2 C3

Progeny

A1 B1 C2
A1 B1 C1
A1 B1 C2
A2 B1 C2

Infer

Male map

Male

Female map

Combined map

A B C
**F1 Pseudo-testcross**

- **Male Parent**
  - A\(_1\) B\(_1\) C\(_1\)
  - A\(_2\) B\(_1\) C\(_2\)

- **Female Parent**
  - A\(_1\) B\(_1\) C\(_1\)
  - A\(_1\) B\(_2\) C\(_3\)

- **Progeny**
  - A\(_1\) B\(_1\) C\(_2\)
  - A\(_1\) B\(_1\) C\(_1\)
  - A\(_2\) B\(_1\) C\(_2\)
  - A\(_1\) B\(_1\) C\(_3\)

**Female map**
- B CA

**Combined map**
- A B C

**Male map**
- A C

**Infer**
- A

---

**93 progeny**
- 66,071 loci
- 5,351 markers

---

**93 progeny**
- 45,301 loci
- 3,927 markers
Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences

Julian M. Catchen, Angel Amores, Paul Hohenlohe, William Cresko, and John H. Postlethwait

*Center for Ecology and Evolutionary Biology and Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403
Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences

Julian M. Catchen, * Angel Amorós, 1 Paul Hohenlohe, * William Cresko, * and John H. Postlethwait 1, 1

*Center for Ecology and Evolutionary Biology and 1Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403

G3: Genes, Genomes, Genetics
LG21

Fst

1Mb 2Mb 3Mb 4Mb 5Mb 6Mb 7Mb 8Mb 9Mb 10Mb 11Mb
Linkage Group XXI

RS (Marine)  Boot

Like
Bear Paw

Inverted
How quickly does stickleback evolution occur?
743 fish
27,878 RAD loci
110,000 SNPs

photo: e-Terra, LLC.
Marine x Marine $F_{ST}$ 0.001
Marine x Marine  $F_{ST} 0.001$
Fresh x Fresh  $F_{ST} 0.052$
NJ FST tree

photo: e-Terra, LLC.
Linkage Group XXI

Fresh vs Marine

Rabbit Sl x Rabbit Sl
Linkage Group XXI

Fresh vs Marine 7

Bear Paw Lk
Boot Lk vs Marine
Mud Lk (Hohenlohe, Bassham et al. 2010)
Linkage Group XXI

17 (Marine)  23 (Marine)  8 (Both)

Like
Bear Paw

Inverted
Other recent uses of RAD-seq

Quantitative Trait Loci (QTL) mapping
Population genomics & Genome Wide Association Studies (GWAS)
Phylogenetics and phylogeography
Genetic mapping, comparative genomics
*de novo* Genome assembly
Identifying signatures of selection in natural populations
Inferring parentage and pedigrees in the wild
Quantitative genetics in outbred populations
Allele specific transcriptional profiling using RNA-seq
What if you don’t have a genome sequence?

Genomically enabling very non-model organisms
Seahorses, sea dragons and pipefishes
Gasterosteidae and Syngnathidae are historically considered to be closely related.
Gulf Pipefish
Syngnathus scovelli

- 160 mm (6.3”)
- reversed sex roles
- sexual dimorphism
- specialized suction feeding
- no sequences in international databases

www.bio.tamu.edu/USERS/ajones/charlyn.html
Fishbase.org
We’re really interested in the head and body axis
Solution: ‘genomically enable’ pipefish

1) A high quality transcriptome

2) Very dense RAD genetic map

3) Deep coverage shotgun sequencing of genome

4) Order genomic and transcriptomic contigs against the RAD reference map
Pipefish Transcriptome
Building an EST database in pipefish

Pipefish embryonic mRNA

\[\text{Illumina sequencing:} \]

100 nt, paired-end

200 million reads (two lanes)

Assembly of transcripts
Transcriptome

- 30,000 solid contigs
- Mean depth of coverage = 24X
- Nearly all of the expected genes in the genome
Transcriptome

Very large genes were represented
We could use these genes right away
*Dlx2a* and *Dlx5a* expression in pipefish
Pipefish Genetic Map
Genetic map workflow

Generated an F1 family of 103 individuals
RAD sequenced the parents and offspring
Analyzed the data using Stacks
Paired end local assemblies
Output to JoinMap format
Created Linkage map
The pipefish genetic map is closed; 22 LGs
6000 segregating SNPs; 30,000 RAD sites
Pipefish Genome Project
Generated DNA from a single individual
Random Illumina shotgun sequencing
Removed highly repetitive kmers
Produced several different genome assemblies
Illumina genomic libraries for pipefish genome

**Paired end** 101bp

- 500–700bp

**Mate pair**

- 4500–7500bp

**Overlapping**

- 150–250bp

**Paired end RAD**

- ACTCTC
- 500–1200bp

- 15–25x of 3% of the genome
Pipefish genome assembly version 0.99
Nearly the whole genome is covered

<table>
<thead>
<tr>
<th>Coverage</th>
<th>Scaffolds</th>
<th>Contigs</th>
<th>Scaffold N50</th>
<th>Contig N50</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (66.6x)</td>
<td>33,911</td>
<td>307,317</td>
<td>26,109</td>
<td>1,840</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Max</th>
<th>Average Length</th>
<th>Total Length</th>
<th>Gap Length</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>198,155</td>
<td>9,916.35</td>
<td>336,273,415</td>
<td>38,303,839</td>
<td>(11.39%)</td>
</tr>
</tbody>
</table>
Bringing it all together; the spotted gar

Amores, Catchen et al. 2011. Genetics
94 Individuals
15,076 Markers
8,046 Mapped
974 In Genes
94 Individuals
15,076 Markers
8,046 Mapped
974 In Genes

<table>
<thead>
<tr>
<th>Organism</th>
<th>Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silver carp</td>
<td>483</td>
</tr>
<tr>
<td>Guppy</td>
<td>790</td>
</tr>
<tr>
<td>Barramundi</td>
<td>240</td>
</tr>
<tr>
<td>Catfish</td>
<td>331</td>
</tr>
<tr>
<td>Sea bass</td>
<td>368</td>
</tr>
<tr>
<td>Cichlid</td>
<td>204</td>
</tr>
<tr>
<td>Platyfish</td>
<td>290</td>
</tr>
<tr>
<td>Halibut</td>
<td>604</td>
</tr>
<tr>
<td>Sea bream</td>
<td>204</td>
</tr>
</tbody>
</table>
Physical Contig Size Distribution
- multiple RAD sites per segregating marker means that more of the genome can be tiled

- Mis-assemblies are easily identified
Using PE RAD for local assembly

Collate/Assemble PE reads

Match to marker catalog

Associate markers / PE contigs with ESTs & genomic contigs

Assign orthology to: markers PE contigs ESTs
Overall Conclusions

Genomics can be a tool for enabling new ecology and evolution research
- documenting patterns of genetic variation
- identifying the molecular genetic basis of important phenotypic variation
- assessing how ecological processes structure this genetic variation in genomes
- RAD-seq is a powerful tool for SNP identification and genotyping
- analytical and computational approaches are challenging but

Not your father’s genome assembly
- a mixture of data types can be efficiently combined
- a genetic map is extremely useful for pulling it all together
- having a tiled genome is good enough - it doesn’t have to be completely closed

Open Source Genomics provides a suite of breakthrough technologies
- the molecular approaches are not as daunting as they first appear
- analytical and computational approaches are challenging
- New software tools can help, but knowledge of Unix and Python is essential
TUTORIAL - USING STACKS

Stacks is a software pipeline for building loci out of a set of short-read sequenced samples. Stacks was developed for the purpose of building genetic maps from RAD-Tag Illumina sequence data, but can also be readily applied to population studies, and phylogeography.

G3: Genes, Genomes, Genetics

Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences

Julian M. Catchen,* Angélica Amores,† Paul Hohenlohe,* William Cresko,* and John H. Postlethwait†,†
*Center for Ecology and Evolutionary Biology and †Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403
Stacks workflow
Stacks workflow
Stacks workflow

**GENOTYPES**

**Locus 1**

Haplotypes: C A
Genotypes: A B

**Locus 2**

Haplotypes: AC CT
Genotypes: A B

**Locus N**

Haplotypes: C T
Genotypes: A B

---

**REFERENCE GENOME**

**PROSES RADTAGS**

**U STACKS**

**P STACKS**

**C STACKS**

**S STACKS**

**Build loci from REF**

**Build loci**

**Assemble Catalog**

**Match to Catalog**

**Genetic MAPPING**

**Population Genomics**
Stacks workflow

**PROCESS RADTAGS**

**REFERENCE GENOME**

**U STACKS**

**P STACKS**

**C STACKS**

**S STACKS**

**GENOTYPES**

Locus 1
Haplotypes: C A
Genotypes: A B

Population A Population B Population C

Progeny
C/AGCTCCCGACCAA TGCAGGACACGGACTG

Locus 2
Haplotypes: AC CT
Genotypes: A B

Population A Population B Population C

Progeny
A/C C/T AGCCCCTGCGGCACCA AAC TGCAGGACACACAGCTATTCCTGCGGCCCG

Locus N
Haplotypes: G T
Genotypes: A B

Population A Population B Population C

Progeny
CCAAACGTTT GG/T

**POPULATIONS**

Locus 1
Haplotypes: C A

Population A Population B Population C

Locus 2
Haplotypes: AC CT

Population A Population B Population C

Locus N
Haplotypes: G T

Population A Population B Population C
Stacks workflow

**PROCESS RADTAGS**

**REFERENCE GENOME**

**BUILD LOCI FROM REF ASSEMBLE CATALOG MATCH TO CATALOG**

**U STACKS**

**P STACKS**

**C STACKS**

**S STACKS**

**GENOTYPES**

**Locus 1**
Haplotypes: C A
Genotypes: A B

**Locus 2**
Haplotypes: AC CT
Genotypes: A B

**Locus N**
Haplotypes: G T

**POPULATIONS**

**JOIN MAP**
**R/QTL**
**ONE MAP**
**HAPLOTYPES**

**SUMMARY STATS**

**STRUCTURE**

**PHYLIP**
<table>
<thead>
<tr>
<th>Id</th>
<th>SNP</th>
<th>Consensus</th>
<th>Matching Parents</th>
<th>Progeny</th>
<th>Marker</th>
<th>Ratio</th>
<th>Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>N/A</td>
<td>2</td>
<td>92 / 91</td>
<td>ab/ac</td>
<td>91</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Alleles</th>
<th>Matching Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>G/A</td>
<td>a : GT</td>
<td>Progeny 1</td>
</tr>
<tr>
<td></td>
<td>b : GG</td>
<td>Progeny 2</td>
</tr>
<tr>
<td></td>
<td>c : AG</td>
<td>Progeny 3</td>
</tr>
<tr>
<td>T/G</td>
<td></td>
<td>Progeny 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progeny 49</td>
</tr>
</tbody>
</table>

Note: The table represents SNP annotation and matching samples with allele and genotype information.
<table>
<thead>
<tr>
<th>Id</th>
<th>SNP</th>
<th>Consensus</th>
<th>Matching Parents</th>
<th>Progeny</th>
<th>Marker</th>
<th>Ratio</th>
<th>Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>~103 annotate</td>
<td>Yes [2nuc]</td>
<td>TGCAAGGCCCTCCCTCCTGCTCGGTACCTGCTGCCACTCCTTCATGCTGGCAGGCCAGGACGGAACATACCTCAACTTCCTC</td>
<td>2</td>
<td>92 / 91</td>
<td>ab/ac</td>
<td>91</td>
<td></td>
</tr>
</tbody>
</table>

**SNPs**

- Column: 52; G/A
- Column: 70; T/G

**Alleles**

- a: GT
- b: GG
- c: AG

**Matching Samples**

<table>
<thead>
<tr>
<th>Progeny 1</th>
<th>Progeny 2</th>
<th>Progeny 3</th>
<th>Progeny 4</th>
<th>Progeny 5</th>
<th>Progeny 6</th>
<th>Progeny 7</th>
<th>Progeny 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / GG</td>
<td>GT / GT</td>
<td>AG / GT</td>
<td>GG / AG</td>
<td>GG / GT</td>
<td>GG / GT</td>
<td>GG / GT</td>
<td>GT / GG</td>
</tr>
<tr>
<td>34 / 13</td>
<td>12 / 14</td>
<td>8 / 16</td>
<td>26 / 14</td>
<td>15 / 11</td>
<td>14 / 8</td>
<td>29 / 11</td>
<td>12 / 5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 9</th>
<th>Progeny 10</th>
<th>Progeny 11</th>
<th>Progeny 12</th>
<th>Progeny 13</th>
<th>Progeny 14</th>
<th>Progeny 15</th>
<th>Progeny 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / GT</td>
<td>GT / GT</td>
<td>AG / GT</td>
<td>GG / AG</td>
<td>GG / GT</td>
<td>GG / GT</td>
<td>GG / GT</td>
<td>GT / GG</td>
</tr>
<tr>
<td>25 / 23</td>
<td>32 / 14</td>
<td>22 / 7</td>
<td>7 / 8</td>
<td>7 / 8</td>
<td>2 / 3</td>
<td>19 / 14</td>
<td>9 / 4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 17</th>
<th>Progeny 18</th>
<th>Progeny 19</th>
<th>Progeny 20</th>
<th>Progeny 21</th>
<th>Progeny 22</th>
<th>Progeny 23</th>
<th>Progeny 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>34</td>
<td>12</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 25</th>
<th>Progeny 26</th>
<th>Progeny 27</th>
<th>Progeny 28</th>
<th>Progeny 29</th>
<th>Progeny 30</th>
<th>Progeny 31</th>
<th>Progeny 32</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>9 / 4</td>
<td>14</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 33</th>
<th>Progeny 34</th>
<th>Progeny 35</th>
<th>Progeny 36</th>
<th>Progeny 37</th>
<th>Progeny 38</th>
<th>Progeny 39</th>
<th>Progeny 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>10 / 5</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 41</th>
<th>Progeny 42</th>
<th>Progeny 43</th>
<th>Progeny 44</th>
<th>Progeny 45</th>
<th>Progeny 46</th>
<th>Progeny 47</th>
<th>Progeny 48</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 49</th>
<th>Progeny 50</th>
<th>Progeny 51</th>
<th>Progeny 52</th>
<th>Progeny 53</th>
<th>Progeny 54</th>
<th>Progeny 55</th>
<th>Progeny 56</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>10 / 5</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 57</th>
<th>Progeny 58</th>
<th>Progeny 59</th>
<th>Progeny 60</th>
<th>Progeny 61</th>
<th>Progeny 62</th>
<th>Progeny 63</th>
<th>Progeny 64</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 65</th>
<th>Progeny 66</th>
<th>Progeny 67</th>
<th>Progeny 68</th>
<th>Progeny 69</th>
<th>Progeny 70</th>
<th>Progeny 71</th>
<th>Progeny 72</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 73</th>
<th>Progeny 74</th>
<th>Progeny 75</th>
<th>Progeny 76</th>
<th>Progeny 77</th>
<th>Progeny 78</th>
<th>Progeny 79</th>
<th>Progeny 80</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 81</th>
<th>Progeny 82</th>
<th>Progeny 83</th>
<th>Progeny 84</th>
<th>Progeny 85</th>
<th>Progeny 86</th>
<th>Progeny 87</th>
<th>Progeny 88</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progeny 89</th>
<th>Progeny 90</th>
<th>Progeny 91</th>
<th>Progeny 92</th>
<th>Progeny 93</th>
<th>Progeny 94</th>
<th>Progeny 95</th>
<th>Progeny 96</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
<td>GT / AG</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Id</td>
<td>SNP</td>
<td>Consensus</td>
<td>Matching Parents</td>
<td>Progeny</td>
<td>Marker</td>
<td>Ratio</td>
<td>Genotypes</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>-----------</td>
<td>------------------</td>
<td>---------</td>
<td>--------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>103</td>
<td>Yes</td>
<td>TCTCAAGAGAGCTCCCTACCTCGGCTAGATGGCAGCCAGGGGACATTCACTGAGGACCAGACCCAAAGGAGAATTTCCACTACACTTCC</td>
<td>2</td>
<td>92 / 91</td>
<td>ab/ac</td>
<td>91</td>
<td>25 (27.5%)</td>
</tr>
</tbody>
</table>

**SNPs**

- Column: 52; G/A
- Column: 70; T/G

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a: GT</td>
</tr>
<tr>
<td></td>
<td>b: GG</td>
</tr>
<tr>
<td></td>
<td>c: AG</td>
</tr>
</tbody>
</table>

**Matching Samples**

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Progeny 1</th>
<th>Progeny 2</th>
<th>Progeny 3</th>
<th>Progeny 4</th>
<th>Progeny 5</th>
<th>Progeny 6</th>
<th>Progeny 7</th>
<th>Progeny 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>13</td>
<td>13</td>
<td>13 / 15</td>
<td>13 / 20</td>
<td>13 / 20</td>
<td>13 / 20</td>
<td>13 / 20</td>
<td>13 / 20</td>
<td>13 / 20</td>
<td>13 / 20</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10 / 15</td>
<td>10 / 20</td>
<td>10 / 20</td>
<td>10 / 20</td>
<td>10 / 20</td>
<td>10 / 20</td>
<td>10 / 20</td>
<td>10 / 20</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1 / 15</td>
<td>1 / 20</td>
<td>1 / 20</td>
<td>1 / 20</td>
<td>1 / 20</td>
<td>1 / 20</td>
<td>1 / 20</td>
<td>1 / 20</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1 / 22</td>
<td>1 / 29</td>
<td>1 / 29</td>
<td>1 / 29</td>
<td>1 / 29</td>
<td>1 / 29</td>
<td>1 / 29</td>
<td>1 / 29</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0 / 22</td>
<td>0 / 29</td>
<td>0 / 29</td>
<td>0 / 29</td>
<td>0 / 29</td>
<td>0 / 29</td>
<td>0 / 29</td>
<td>0 / 29</td>
</tr>
</tbody>
</table>
### Batch #1 [2011-08-10: 80bp Lepisosteus oculatus F1 Genetic Map RAD-Tag Samples]

### RAD-Tag Sample #2 [female]

**Sequence #73**

<table>
<thead>
<tr>
<th>Catalog ID</th>
<th>Depth</th>
<th>SNPs</th>
<th>Alleles</th>
<th>Deleveraged?</th>
<th>Lumberjackstack?</th>
<th>Blacklisted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>#103</td>
<td>26x</td>
<td>Column: 52 G/A AG 46.15%</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Column: 70 T/G GT 53.85%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Seq ID</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>consensus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 primary</td>
<td>CAGTC_2_0018_768_1365_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>2 primary</td>
<td>CAGTC_2_0029_1628_1751_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>3 primary</td>
<td>CAGTC_2_0053_1692_1888_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>4 primary</td>
<td>CAGTC_2_0058_1588_1038_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>5 primary</td>
<td>CAGTC_2_0059_1524_1186_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>6 primary</td>
<td>CAGTC_2_0094_1356_1854_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>7 primary</td>
<td>CAGTC_2_0096_1791_1246_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>8 primary</td>
<td>CAGTC_2_0096_1791_1246_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>9 primary</td>
<td>CAGTC_2_0021_677_296_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>10 primary</td>
<td>CAGTC_2_0024_307_735_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>11 primary</td>
<td>CAGTC_2_0039_1352_1764_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>12 primary</td>
<td>CAGTC_2_0061_596_159_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>13 primary</td>
<td>CAGTC_2_0068_1310_997_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>14 primary</td>
<td>CAGTC_2_0070_644_2040_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>15 primary</td>
<td>CAGTC_2_0074_328_659_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>16 primary</td>
<td>CAGTC_2_0075_1666_1862_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>17 primary</td>
<td>CAGTC_2_0079_1481_305_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>18 primary</td>
<td>CAGTC_2_0084_805_1974_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>19 primary</td>
<td>CAGTC_2_0100_481_1043_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>20 secondary</td>
<td>CAGTC_2_0014_728_1008_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>21 secondary</td>
<td>CAGTC_2_0016_86_1022_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>22 secondary</td>
<td>CAGTC_2_0042_426_1001_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>23 secondary</td>
<td>CAGTC_2_0052_867_1387_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>24 secondary</td>
<td>CAGTC_2_0012_221_1043_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>25 secondary</td>
<td>CAGTC_2_0095_120_1067_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
<tr>
<td>26 secondary</td>
<td>CAGTC_2_0077_1003_356_1[35245]</td>
<td>TGCAGAGCCCTCCACCTGCTAGCCTGCTACCTACGCTACCTACATACCTC</td>
</tr>
</tbody>
</table>