Genetic basis and signatures of selection around loci underlying adaptation and speciation

Felicity Jones
Friedrich Miescher Laboratory of Max Planck Society
Tübingen, Germany

Male threespine stickleback
What is the molecular basis of adaptive divergence?

Two parts:

1. Methods used to identify loci underlying adaptive divergence
   - Forward genetic mapping
   - Population genomics

2. CASE STUDIES of molecular signatures of selection around known loci
   - EDA
   - Pitx1
   (what patterns might we expect to see at other adaptation loci?)
How to identify and test the loci and alleles underlying adaptation?

- Number and type of natural mutations?
- Functional effects on phenotype?
- Molecular mechanisms?
- Evolutionary processes & constraints?
- Affect on fitness and reproductive isolation?
Threespine sticklebacks have undergone a recent adaptive radiation

From Bell & Foster (1994)
Marine and freshwater sticklebacks differ in many traits.
Different types of sympatric “Species-Pairs” have evolved from Bell & Foster (1994).
Different species-pairs have evolved (and are at different ‘stages’ of speciation)

**Weak**

Lake-Stream
(100’s of independent events)

**Weak-Medium**

Marine - Freshwater
(10000’s of independent events)

Medium

**Medium**

Benthic - Limnetic
(5 independent events)

**Strong**

Japan Sea - Pacific Marine
(1 independent event)


McPhail, Schluter et al (many pubs)

(Kitano et al 2009)
Repeated & independent evolution provides biological replicates of the evolutionary process.
A broad suite of genetic and genomic tools in sticklebacks

Easy husbandry, viable inter-crosses

Transgenic techniques for reporter assays, knock-outs, knock-ins, & genome editing

Transparent embryos for developmental studies

Small (0.45GB), well-assembled Ref Genome >> hundreds of short-read sequenced genomes
What is the molecular basis of adaptive divergence and speciation?

Identifying and functionally testing adaptive loci

1. Forward genetics: from trait to basepairs

2. Population genomics
Finding the genes using forward genetics:
Quantitative Trait Locus (QTL) mapping in sticklebacks

Cross divergent forms

Establish F2 generation, Build Genetic Map

Identify 2LOD Interval

Fins
Vertebrae
Sex
Teeth
Jaw
Spines

QTL Candidate Interval
Finding the genes using forward genetics: Quantitative Trait Locus (QTL) mapping in sticklebacks

Cross divergent forms

Establish F2 generation, Build Genetic Map

Identify 2LOD Interval

Fins
Vertebrae
Sex
Teeth
Jaw
Spines

non-causal locus

causal locus

Phenotype

AA Aa aa

Phenotype

AA Aa aa
Finding the genes using forward genetics: QTL mapping in sticklebacks has been very successful
Forward genetics:
Narrowing QTL intervals to causal base pairs or haplotypes

Cross divergent forms

Establish F2 generation, Build Genetic Map

Identify 2LOD Interval

Colosimo et al 2005 Science
Forward genetics: Narrowing QTL intervals using association mapping in natural populations

Chromosome walking

QTL Candidate Interval
(0.68cM)

539kb Candidate Interval

Association mapping using a polymorphic natural population
(Friant River, California)

16kb candidate interval

Colosimo et al 2005 Science
Forward genetics:
Narrowing QTL intervals by demonstrating functional significance of genes

Transgenic Rescue using candidate gene (EDA) within interval

Rescuing plates on low plated fish

Colosimo et al 2005 Science
Forward genetics: Narrowing QTL intervals using Allele-Specific Expression tests for cis-regulatory element

Transgenic Rescue using candidate gene (EDA) within interval

Allele specific expression reveals a cis-acting regulatory mutation

Rescuing plates on low plated fish

Colosimo et al 2005 Science

O’Brien et al 2015 ELife
Forward genetics: Narrowing QTL intervals to causal base pairs or haplotypes using exceptional / unusual populations

"Naka - An unusual exception"
Low plated Japanese fish with high plated EDA haplotype

Colosimo et al 2005 Science
Forward genetics:
Narrowing QTL intervals to causal base pairs or haplotypes

Sequencing of an unusual exception (Low Plated Japanese Fish) identifies a single 3’ substitution

“NAKA - An unusual exception”
Low plated Japanese fish with high plated EDA haplotype

Colosimo et al 2005 Science

O’Brien et al 2015 ELife
Forward genetics:
Narrowing QTL intervals using reporter assays to demonstrate the functional significance of a single 1bp mutation

Regulatory enhancer assays functionally demonstrate loss of expression caused by 1bp change.

Wildtype marine enhancer expression pattern.

Single bp substitution reduces expression pattern.
# Forward genetics:
Narrowing QTL intervals to causal base pairs or haplotypes

1. **QTL mapping**
   - 0.63cM ~500kb (many genes)

2. **Association mapping in naturally polymorphic population**
   - 16kb (3 genes)

3. **Functional Assay:** Rescue of plates using transgenics
   - 1 gene (EDA)

4. **Allele specific expression assay** reveals cis-acting element
   - Cis-acting element underlies differences in EDA expression.

5. **Sequencing of a population** that is an unusual exception
   - 1bp substitution, 3’ of EDA

6. **Functional Assay:** Regulatory enhancer assays show sufficiency of substitution
   - 1bp substitution, 3’ of EDA
Forward genetics: Narrowing QTL intervals to causal base pairs or haplotypes

1. QTL mapping
   0.63cM ~500kb (many genes)

2. Association mapping in naturally polymorphic population
   16kb (3 genes)

3. **Functional Assay:** Rescue of plates using transgenics
   1 gene (EDA)

4. Allele specific expression assay reveals cis-acting element
   Cis-acting element underlies differences in EDA expression.

5. Sequencing of a population that is an unusual exception (low plated phenotype but high plated haplotype)
   1bp substitution, 3’ of EDA

6. **Functional Assay:** Regulatory enhancer assays show sufficiency of substitution
   1bp substitution, 3’ of EDA
Fine mapping, Allele-specific expression & Transgenic assays identify *cis*-regulatory elements controlling adaptive traits

**Plates (EDA)**

**Pelvis (Pitx1)**

**Single bp mutation**
- in a 3’ *cis*-regulatory element causes loss of EDA expression

**488bp deletion**
- of a 5’ *cis*-regulatory element causes loss of Pitx1 expression
Genetic basis and signatures of selection around loci underlying adaptation and speciation

Felicity Jones
Friedrich Miescher Laboratory of Max Planck Society
Tübingen, Germany

Male threespine stickleback
What is the molecular basis of adaptive divergence and speciation?

Identifying and functionally testing adaptive loci

1. Forward genetics: from trait to basepairs

2. Population genomics
Loss of lateral plates has evolved repeatedly via re-use of an ancient EDA haplotype.
Loss of lateral plates has evolved repeatedly via re-use of an ancient EDA haplotype.
Leveraging parallelism to identify adaptive loci
[Whole genome sequencing of replicate marine - freshwater species pairs]
(21 stickleback genomes, Jones et al 2012 Nature)
Parallel evolution via reuse of pre-existing genetic variation

Low Plated EDA alleles 'hide' at low frequency (0-6%) in the High Plated marine population

Selective sweep of “pre-screened” adaptive variation

Colosimo et al (2005) Science
Two different statistical approaches to identify parallel adaptive loci

Self-organized Map / Hidden Markov Model

Illumina short-read genomic sequencing

Genetic distance matrices

Two different statistical approaches to identify parallel adaptive loci

“Saguaro”
Self-organized Map / Hidden Markov Model

Illumina short-read genomic sequencing

Two different statistical approaches to identify parallel adaptive loci

Illumina short-read genomic sequencing

Genetic distance matrices

Two different statistical approaches to identify parallel adaptive loci

Illumina short-read genomic sequencing

Genetic distance matrices

P-values calculated by permutation tests

Shuffle group membership >350,000 times
Bayesian ‘Data-Driven’ Clustering
to identify other types of clustering patterns in the genome

Membership and phenotypic/ecological similarity of group members may inform us of possible genetic function and relevant selection pressures.

Bayesian ‘Data-Driven’ Clustering to identify other types of clustering patterns in the genome

Bayesian ‘Data-Driven’ Clustering to identify other types of clustering patterns in the genome

Fish falling in cluster K2

Fish falling in cluster K1
Two different statistical approaches to identify parallel adaptive loci

Self-organized Map / Hidden Markov Model

Illumina short-read genomic sequencing

Genetic distance matrices

Cluster Separation Score
[Parallel Divergence]

Clear signals of parallel marine-freshwater divergence
High resolution signals = good for functional follow-up

EDA

20kb

40kb

<5kb

Genes

RAW DATA

Marine

Fresh
Whole genome sequencing of marine – freshwater species pairs revealed parallel adaptive divergence at ~81 genomic loci

What is the molecular basis of adaptive divergence and speciation?

Identifying and functionally testing adaptive loci

1. Forward genetics: from trait to basepairs

2. Population genomics
Genetic basis and signatures of selection around loci underlying adaptation and speciation

Felicity Jones
Friedrich Miescher Laboratory of Max Planck Society
Tübingen, Germany

Male threespine stickleback
What is the molecular basis of adaptive divergence?

Two parts:

1. Methods used to identify loci underlying adaptive divergence
   - Forward genetic mapping
   - Population genomics

2. CASE STUDIES of molecular signatures of selection around known loci
   - EDA
   - Pitx1
   (what patterns might we expect to see at other adaptation loci?)
The EDA locus explains as much as 70% of variation in plate number.

Cross divergent forms

Establish F2 generation, Build Genetic Map

Identify 2LOD Interval

Colosimo et al 2005 Science
Loss of lateral plates has evolved repeatedly via re-use of an ancient EDA haplotype.
Parallel evolution via reuse of pre-existing genetic variation

Low Plated EDA alleles ‘hide’ at low frequency (0-6%) in the High Plated marine population

Colosimo et al (2005) Science
Molecular signatures of selection

Neutral allele

Adaptive allele
Molecular signatures of selection

Within population statistics:
- Reduction in genetic diversity
- Excess of derived alleles
- Aberrant site frequency spectrum
- Extended haplotype block size
  (Elevated rates of protein evolution)

Cross population statistics:
- Elevated genetic divergence
- Difference in haplotype size
Molecular signatures of selection around EDA using high density genotyping arrays

Felicity Jones & Frank Chan
Molecular signatures of selection around EDA using high density genotyping arrays

High Plated Individuals

Low Plated Individuals

SNPs 400kb window

Felicity Jones & Frank Chan
Molecular signatures of selection around \textit{EDA} using high density genotyping arrays

Colosimo et al (2005)

Minimal haplotype block 16kb

\textit{Colors} = different plate phenotypes (red = high plated; blue = low plated)
Molecular signatures of selection around *EDA* using high density genotyping arrays

Colors = different geographic regions (AK, CA, BC, AT)

Colosimo *et al* (2005)

Minimal haplotype block 16kb

**Reduction in Heterozygosity**

Colors = different plate phenotypes (red = high plated; blue = low plated)

**Excess of Derived Alleles**

Fay & Wu’s H (low plated populations)

Colors = different geographic regions (AK, CA, BC, AT)
Molecular signatures of selection around *EDA* using high density genotyping arrays

Colors = different geographic regions (AK, CA, BC, AT)

Colosimo et al (2005)

Minimal haplotype block 16kb

Reduction in Heterozygosity

Colors = different plate phenotypes (red = high plated; blue = low plated)

Excess of Derived Alleles

Fay & Wu’s H (low plated populations)

Colors = different geographic regions (AK, CA, BC, AT)

Increased Genetic Divergence

Average $F_{ST}$ between high and low plated populations

Chromosomal Position (Mb)
Comparisons across populations help narrow the EDA region.

How is high resolution achieved?

< 6.7kb

SNPs 400kb window
Parallel evolution via reuse of pre-existing variation

Via reuse of pre-existing variants

shared haplotype

An alternative mode of parallel evolution

Via independent mutations

unique haplotypes

An alternative mode of parallel evolution

Via reuse of pre-existing variants

Via independent mutations

Pelvic spine loss has evolved repeatedly
A major QTL explaining close to 80% variation on linkage group VII

Cross divergent forms

Establish F2 generation, Build Genetic Map

Identify 2LOD Interval

QTL Candidate Interval (including Pitx1 gene: a key hindlimb developmental regulator)

Shapiro et al 2004 Nature
Pitx1 allele-specific expression assay in F1 hybrids

- Higher expression of complete-pelvis Pitx1 allele than pelvic-reduced allele
- Cis-regulatory element, rather than transacting factor

Chan et al 2010 Science
Association mapping in two populations polymorphic for pelvic-reduction identifies a 23kb region 5’ of Pitx1.
A high density genotyping array reveals deletions in pelvic-reduced populations with a minimum shared interval of 488bp.
Transgenic assays reveal a regulatory element and phenotypic rescue

Fig. 3. (A) Juvenile pelvic-reduced BEPA stickleback expressing a Pitx1 transgene driven by the Pel-2.5-kb\textsuperscript{SALR} enhancer compared with (B) uninjected sibling. External spines form only in transgenic fish (arrowhead). (C and D) Alizarin red–stained pelvic structures of adult transgenic fish compared with BEPA parental phenotype. BEPA fish normally develop only a small ovoid vestige (OV) of the anterior pelvic process (AP). Transgenic fish show clear development of the AP, ascending branch (AB), and posterior process (PP) of the pelvis, and a prominent serrated pelvic spine. Pectoral fin (PF) rays develop in both fish.
Molecular signatures of selection at the Pel regulatory element in pelvic reduced populations

Excess of Derived Alleles

Reduction in Diversity ($\theta - \pi$)


Chan et al 2010 Science
Pelvic spine loss has evolved repeatedly (including in two benthic populations)
Benthic and limnetic ecotypes evolved following a “Double Invasion” event caused by isostatic rebound.
What is the genetic basis of the divergence?
At neutral loci, fish cluster together by lake

Genome wide SNP genotyping array
(~1000 SNPs)

Jones et al (2012) Current Biology
What is the genetic basis of the divergence? At neutral loci, fish cluster together by lake

Genome wide SNP genotyping array
(~1000 SNPs)

Projection of global populations onto PCA using loadings from Benthic Limnetic eigenvectors

Global marine & freshwater fish populations are not predictive of the source of neutral variation

Jones et al (2012) Current Biology
But at selected loci, fish cluster by ecotype

Genome wide SNP genotyping array
(~1000 SNPs)

F_{ST} Outlier SNPs:

Jones et al (2012) Current Biology
But at selected loci, fish cluster by ecotype

Genome wide SNP genotyping array
(~1000 SNPs)

F_{ST} Outlier SNPs:
At selected loci, benthic fish carry freshwater alleles, limnetic carry marine alleles

Genome wide SNP genotyping array (~1000 SNPs)

F<sub>ST</sub> Outlier SNPs:

At divergent loci, benthic fish carry alleles similar to those carried by nearby freshwater populations.

Limnetic fish carry marine alleles

Jones et al (2012) Current Biology
Whole genome sequencing of 48 benthic & limnetic ecotypes from 4 lakes

6 x Limnetic
6 x Benthic
6 x Limnetic
6 x Benthic
6 x Limnetic
6 x Benthic
6 x Limnetic
6 x Benthic

PAXTON LAKE
PRIEST LAKE
QUARRY LAKE
ENOS LAKE ("collapsed")

[Historical samples sequenced]
Parallel benthic-limnetic divergence across large parts of the genome

Muhua Wang
Genetic basis and signatures of selection around loci underlying adaptation and speciation

Felicity Jones
Friedrich Miescher Laboratory of Max Planck Society
Tübingen, Germany

Male threespine stickleback
TODAY’s EXERCISE: Molecular signatures of selection around Pitx1 in Benthic fish?
Lessons from case studies of known adaptive loci: EDA and PITX1

Association mapping in natural populations is extremely helpful at narrowing the relevant interval (EDA ~20kb; PITX1 ~23kb)

The signature of phenotype-genotype association is so small that reduced representation sequencing methods (eg GBS, RAD) may miss it entirely.

Whole genome sequencing (including low pass “skim” sequencing) can give excellent high resolution signal

In the case of EDA, whole genome sequencing of individuals from different populations (as opposed to many individuals from a single population) helped narrow the boundaries of the region.

Signatures of selection can be narrow: 20-50kb wide are not unusual.
Molecular signatures of a selective sweep
Molecular signatures of a selective sweep
Molecular signatures of a selective sweep
Molecular signatures of a selective sweep
Molecular signatures of a selective sweep
Summarising Polymorphism Data with a Site Frequency Spectrum
Summarising Polymorphism Data with a Site Frequency Spectrum