## Genomic studies of speciation and gene flow



## Why study speciation genomics?

Long-standing questions (role of geography/gene flow)

How do genomes diverge?

Find speciation genes

## Genomic divergence during speciation



1. Speciation as a bi-product of physical isolation

2. Speciation due to selection - without isolation

## Genomic divergence during speciation



1. Speciation as a bi-product of physical isolation


Cline theory - e.g. Barton and Gale 1993
2. Speciation due to selection - without isolation

## Genomic divergence during speciation



1．Speciation as a bi－product of physical isolation


Original species


Daughter specics with hybrids


[^0]2．Speciation due to selection－without isolation

# Stage 1 - one or few loci under disruptive selection 

Gene<br>under<br>selection

Genome


Feder, Egan and Nosil TiG

## Stage 2 - Divergence hitchhiking

Genome


Feder, Egan and Nosil TiG

## Stage 2b - Inversion

Inversion links co-adapted alleles


III
Genome


Feder, Egan and Nosil TiG

## Stage 3 - Genome hitchhiking



Genome


Feder, Egan and Nosil TiG

## Stage 4 - Genome wide isolation

## Genome

## $F_{\mathrm{ST}}$



Feder, Egan and Nosil TiG

# Some clear examples of 'speciation islands' 

Wing pattern "races" of Heliconius melpomene

Heliconius melpomene

H. melpomene amaryllis

H. melpomene aglaope


# Some clear examples of 'speciation islands' 

Wing pattern "races" of Heliconius melpomene

S. H. Martin et al. Genome Res. 23, 1817-1828 (2013).
O. Seehausen et al. Nat. Rev. Genet. 15, 176-92 (2014).

# Some clear examples of 'speciation islands' 

## Carrion and hooded Crows





Poelstra, J. W. et al. Science 344, 1410-4 (2014).


Hench, K., Vargas, M., Höppner, M. P., McMillan, W. O. \& Puebla, O. Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. Nat Ecol Evol 3, 657-667 (2019).


Malinsky et al., Science 350, 1493 (2015).

Aa Parapatric races: H. m. amaryllis (Per) versus H. m. aglaope (Per)


Ab Allopatric races: H. m. rosina (Pan) versus H. m. melpomene (FG)


Ac Sympatric species: H. cydno (Pan) versus H.m. rosina (Pan)


Ad Allopatric species: H. cydno (Pan) versus H. m. melpomene (FG)


Seehausen et al., Nature Reviews Genetics, 2014

## Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species

MAF Noor and SM Bennett
Biology Department, Duke Universily, Durham, NC, USA


INVITED REVIEWS AND SYNTHESES
Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow

TAMI E. CRUICKSHANK* and MATTHEW W. HAHN* $\dagger$
*Depariment of Biology, Indiana University, Bloomington. IN 47405, USA, +School of Informatics and Computing, Indiana University, Blcomington, IN 47405, USA


## What do patterns of $F_{\text {st }}$ really mean?

- Fst $_{\text {st }}$ measures relative divergence $\quad F_{\mathrm{ST}}=\frac{H_{\mathrm{T}}-H_{\mathrm{S}}}{H_{\mathrm{T}}}$,
- Peaks indicate regions of higher than expected between population divergence, given the within population divergence
- Peaks can therefore result from reduced diversity within species
- This could be due to lower Ne within species (selective sweeps, background selection)
- So peaks NOT NECESSARILY due to reduced gene flow


## No evidence for higher Dxy in wing pattern loci

Wing pat Heliconia

One further issue with interpreting the data from these two races is whether this comparison relates to speciation at all. There is strong geographic structure involving the wing colour patterns that define these morphs as races, largely due to selection determined by colour morphs in the Müllerian mimic, H. erato (Mallet et al. 1990). But the races are not separate species: they do not show evidence of hybrid sterility or inviability and appear to be randomly mating in the narrow zone where the colour morphs overlap (Mallet et al. 1990). This raises the possibility that the colour-
 patterning loci contain locally adapted alleles within a largely panmictic (or at least continuously distributed) population and that gene flow outside of these regions represents nothing more than the normal movement of alleles within a species. In this case, there should be


## Suggestion that we use absolute measures of divergence?




Ravinet, M. et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. 30, 1450-1477 (2017).
' Selective Sweeps


Note that sometimes sweeps within species = speciation genes


Sweeps across the species barrier can also lead to Fst peaks

Nicolas Bierne, Daniel Berner and others

## Sweeps within and between species

Anopheles gambiae and A. coluzzi
Formerly $M$ and $S$ forms of $A$. gambiae


M-wt vs. S


Clarkson et al. 2014 Nature Communications

## Sweeps within and between species




M-kdr vs. S


21
M-wt vs. M-kdr
kdr
3R

' Background selection


Sequenced 20 individuals per population at 20x coverage

## Collared and Pied Flycatchers



Fixed differences
Ellegren, et al. Nature 491, 756- (2012).

c
Cryomosome 4A

$\pi^{+}+0^{3}$

$\pi_{i} * 30$


Supplemental Table S4. ABBA-BABA tests for gene flow. Populations/species among which the test indicates gene flow are highlighted in bold.

| 1. Inner | 2. Inner | 1. Outgroup | Mean(D) | SE(D) | p-value |
| :--- | :--- | :--- | :--- | :--- | :--- |
| collared Italy | collared CZ | pied CZ | 0.0010 | 0.0010 | 0.3344 |
| pied Spain | pied CZ | collared CZ | 0.0004 | 0.0005 | 0.4186 |
| pied Spain | Atlas | collared Italy | -0.1648 | 0.0027 | $<10^{-4}$ |
| pied Spain | Atlas | semicollared | -0.0108 | 0.0016 | $<10^{-4}$ |
| pied Spain | collared Italy | semicollared | 0.1162 | 0.0018 | $<10^{-4}$ |
| Atlas | collared Italy | semicollared | 0.1242 | 0.0016 | $<10^{-4}$ |

## Background selection - theory

## MOLECULAR ECOLOGY

ORIGINAL ARTICLE 自 full Acress
Background selection and $F_{5 \mathrm{~T}}$ : Consequences for detecting local adaptation

Remi Mathey-Dorets Michael C. Whitlock
First published: $\mathbf{2 6}$ July 2019 | https'//dol.org/10.1111/mec. 15197 | Clations: 33
Simulate diverging populations and calculate Fst
Demonstrate little effect of BGS on between-locus variation in Fst even with no recombination

Selfing, asexuals etc not simulated. Mainly focused on local adaptation - early stages of divergence

- Genetic architecture


## Lake Nicaragua cichlids



Lip size - single locus
Body size and pharyngeal jaws polygenic

Contrasting signatures of genomic divergence during sympatric speciation


Ravinet, M. et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. 30, 1450-1477 (2017).

## Better to test specific hypotheses: e.g is there gene flow?

Need to design sampling so the expectations in the absence of gene flow are clear and testable

Not always possible, but 'control' populations that are not influenced by admixture

## Explicit tests for gene flow: Neanderthal genome

- Isolated DNA from bones 38,000 yrs old in Croatia
-We diverged from Neanderthals around $270-440,000 \mathrm{yrs}$ ago
-Evidence for gene exchange with humans (1-4\% of genome?)


## Explicit tests for gene flow: ABBABABA test



African European Neanderthal Chimp human human


African human

European Neanderthal Chimp human

$$
\begin{equation*}
D\left(P_{1}, P_{2}, P_{3}, 0\right)=\frac{\sum C_{A B B A}(i)-C_{B A B A}(i)}{\sum C_{A B B A}(i)+C_{B A B A}(i)} \tag{1}
\end{equation*}
$$

## Explicit tests for gene flow: ABBABABA test

Denisovans
Modern humans
Neanderthals


## Explicit tests for gene flow: Heliconius butterflies



Martin et al., Genome Research 2013

## Explicit tests for gene flow: Heliconius butterflies



# Whole-genome phylogeny supports grouping by species 

Many sources of reproductive isolation:

Female hybrids are sterile
Different host plant use
Different habitat preference
Strong assortative mating

## Explicit tests for gene flow: Heliconius butterflies





- 100 kb trees
- Only 53\% group by species
- 42\% group by geography!



## Explicit tests for gene flow: Heliconius butterflies

- Much larger proportion of genome is flowing as compared to Neanderthals
- Similarly strong effect on sex chromosome





## Recombination rate strongly correlated with admixture proportion





Recombination Rate (cM/MB)

Short chromosomes have more admixture:


And chromosome ends have more admixture:


All of these patterns are consistent with a highly polygenic architecture.

## Experimental design for speciation genomics

- Use allopatric 'control' populations
- Use replicate comparisons - e.g hybridising species in different localities
- Use systems with well understood natural history
- Think about readily accessible traits if you want to go beyond genomic patterns
- Choose a system that allows you to address specific hypotheses


## Conclusions so far

- Genomics has great power to detect patterns of divergence and identify genes underlying speciation
- But patterns of differentiation are complex and influenced by many factors
- These include intrinsic properties of the genome of little direct relevance to speciation
- Need clear hypotheses
- Combine multiple signals
- Model background selection, recombination, positive selection?


## Adaptive introgression











- G-test: $G=7.25$, d.f. $=1, p=0.007$

Merrill et al., Proc. Roy. Soc 2012



Peaks of divergence correspond to wing pattern genes

Colour pattern differences between species are controlled by large effect loci

Ac - band shape


D - red patterns


Yb - yellow patterns


K - white/yellow colour


## Optix - the red locus



D - red patterns


Linlin Zhang, Anyi Mazo-Vargas, and Robert D. Reed PNAS 2017


Transcription factor, paints red in pupal wing (Richard Wallbank)


Richard Wallbank

## The first evidence for combinatorial evolution in

 Heliconius

NNbb

NNBB


Heliconius heurippa

The first evidence for combinatorial evolution in Heliconius


Salazar et al., PLoS Genetics 2010






Generate dated trees using this node as a reference point

## Diverse history of adjacent enhancers






b


## Supergenes and mimicry



## Supergenes in mimicry

- Batesian mimics suffer if they become too common - so negative frequency dependent selection favours rare forms
- Muellerian mimics NOT expected to be polymorphic; where they are, this is probably due to spatial or temporal heterogeneity in selection

A spatial mosaic of mimicry maintains polymorphism in Heliconius numata


All inherited at a single locus


There is a dominance hierarchy of alleles controlling the polymorphism


## The supergene locus is associated with two alternative inversion rearrangements



## One of the inversion haplotypes has arisen through introgression from a related species



## Signatures of selection at wing pattern genes



## Selective sweeps



Hard selective sweep leaves a signature in the genome

Is there evidence for recent selection on pattern loci?

Can we distinguish forms of selection?

## Comparing signatures of sweeps under different forms of selection - theory



## Signatures of sweeps - empirical patterns



Pi


Tajima's D

Sweepfinder
Likelihood Ratio

## Signatures of selection at all wing pattern genes




WntA


Cortex


Optix

Invasion of a mega-pest - Helicoverpa


Admixture has declined through time in H . armigera






## ALX1 associated with beak shape




## Pupfish radiation sorts ancestral variation



Richards et al., PNAS 118(20)e2011811118

## Implications for tree-thinking



Mallet, Hahn and Besansky BioEssays 2015
Hahn and Nakhleh Evolution 2015

The true phylogeny showing bifurcation events as well


# Implications for tree-thinking 



The tree of life is reticulated

## Okay, so what have we learnt and where do we go from here?

## An alternative is to take an explicit modelling approach



IM and IMa Jody Hey


Martin et al., Biorxiv 2015



So far models have mostly just estimated genomewide parameters...assuming the genome is homogenous

Where we need to go next is to incorporate genome heterogeneity in selection and recombination

## Identifying Loci Under Selection Against Gene Flow in Isolation-with-Migration Models

Vitor C. Sousa,**1.2 Miguel Carneiro, ${ }^{+}$Nuno Ferrand, ${ }^{+}$and Jody Hey ${ }^{* .1}$
*Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, New lersey 08854, and ${ }^{1}$ CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, 4099-002 Porto, Portugal

## High density linkage maps to map the recombination landscape



## The effect of background selection on introgression in humans




Admixture is less in gene rich regions supporting this model.....


[^0]:    」1 ل｜いいいいい

