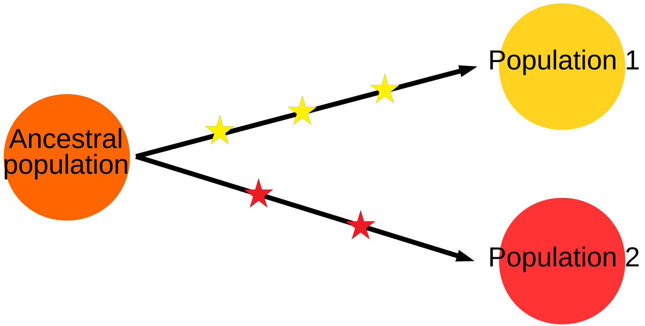
Inferring the relationship between molecular divergence and genetic isolation

September 27, 2018

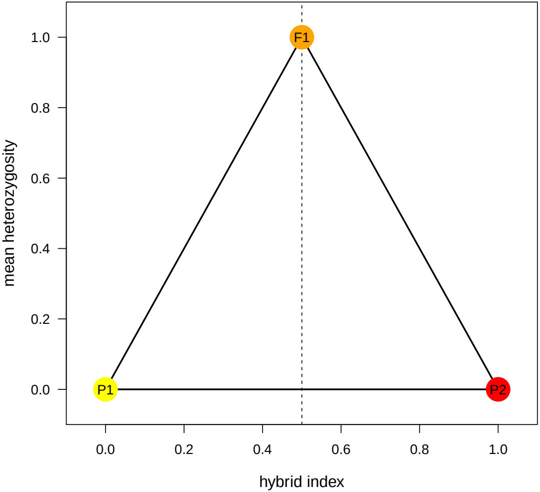
SPECIATION



- Species barriers have no effect within P1 and P2, but reduce gene flow between P1 and P2
- P1 and P2 are different at d genomic positions

REPRODUCTIVE ISOLATION

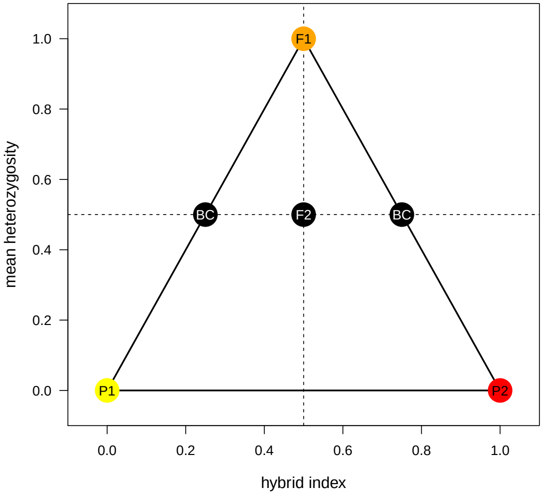
(Alexis Simon thesis; Montpellier + Cambridge)



- heterozygosity = proportion of $P1/P2$ positions over d
- hybrid index = proportion of $P2$ alleles over d genotypes
- $ln(w) \propto -S(\mathbf{d})^{\frac{\beta}{2}}$

REPRODUCTIVE ISOLATION

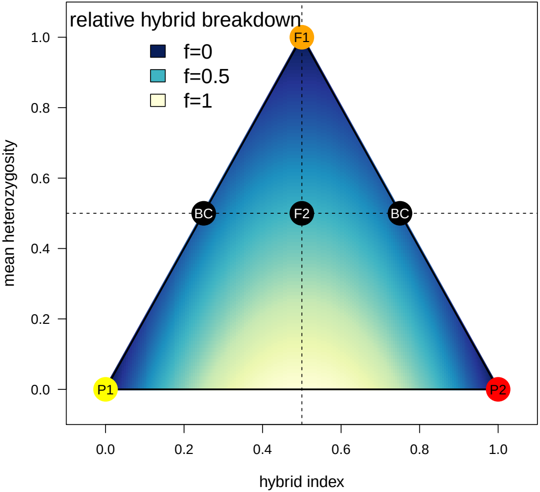
(Alexis Simon thesis; Montpellier + Cambridge)



- heterozygosity = proportion of $P1/P2$ positions over d
- hybrid index = proportion of $P2$ alleles over d genotypes
- $ln(w) \propto -S(\mathbf{d})^{\frac{\beta}{2}}$

REPRODUCTIVE ISOLATION

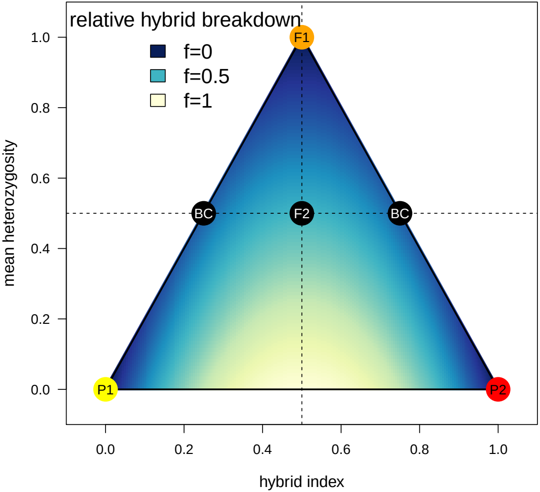
(Alexis Simon thesis; Montpellier + Cambridge)



- heterozygosity = proportion of $P1/P2$ positions over d
- hybrid index = proportion of $P2$ alleles over d genotypes
- $ln(w) \propto -S(\mathbf{d})^{\frac{\beta}{2}}$

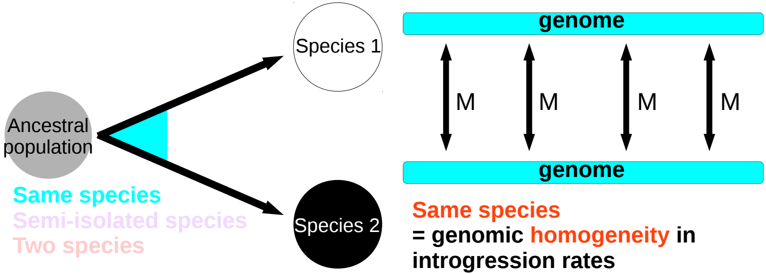
REPRODUCTIVE ISOLATION

(Alexis Simon thesis; Montpellier + Cambridge)



- heterozygosity = proportion of $P1/P2$ positions over d
- hybrid index = proportion of $P2$ alleles over d genotypes
- $ln(w) \propto -S(\boldsymbol{d})^{\frac{\beta}{2}}$

FROM ONE TO TWO SPECIES

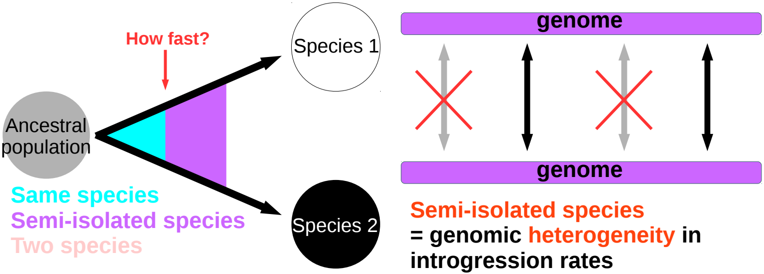


Same species
Semi-isolated species
Two species

Same species
= genomic homogeneity in introgression rates

- When can we detect the effects of the first barriers on gene flow?
- Is there a threshold of divergence above which gene flow is impossible?

FROM ONE TO TWO SPECIES

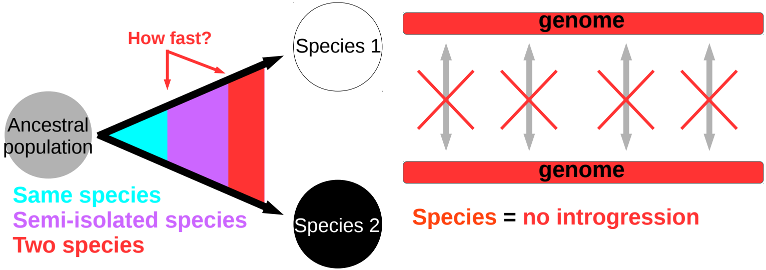


Same species
Semi-isolated species
Two species

Semi-isolated species = genomic heterogeneity in introgression rates

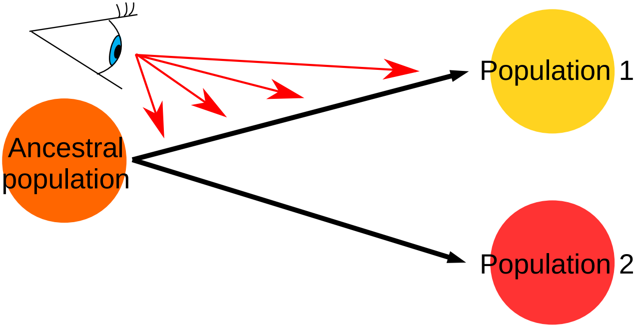
- When can we detect the effects of the first barriers on gene flow?
- Is there a threshold of divergence above which gene flow is impossible?

FROM ONE TO TWO SPECIES



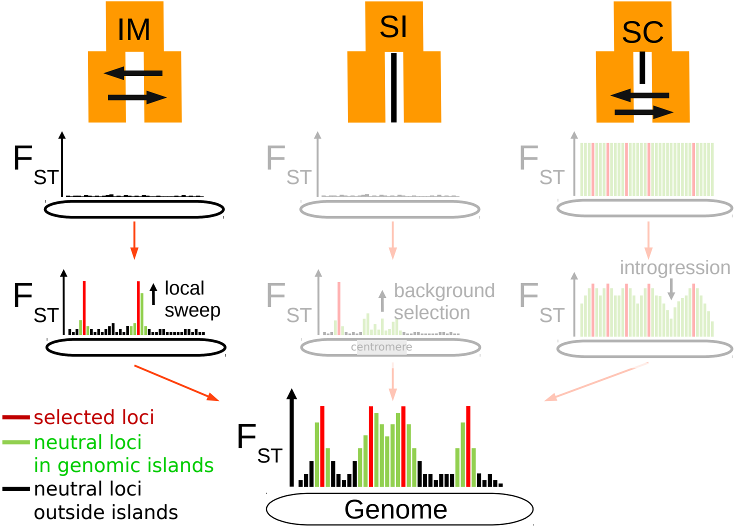
- When can we detect the effects of the first barriers on gene flow?
- Is there a threshold of divergence above which gene flow is impossible?

QUESTIONS:



- Comparing alternative scenarios for 61 pairs of species along a continuum of d (ABC)

MULTIPLE VERBAL MODELS TO EXPLAIN OBSERVED PATTERNS



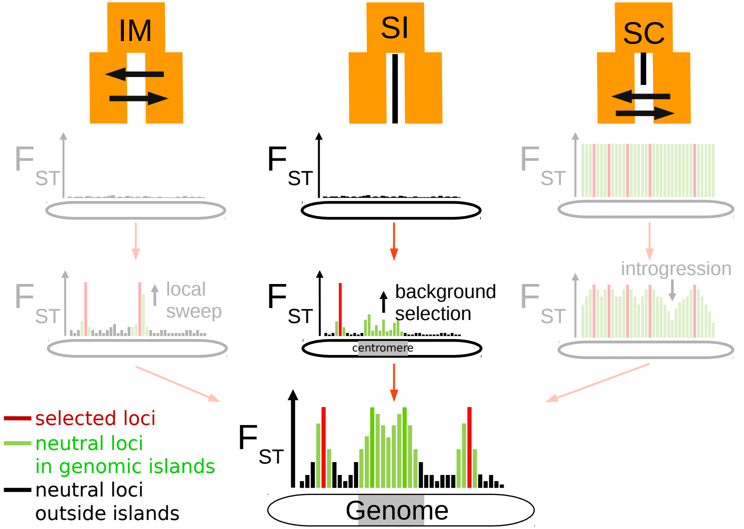
🟥 selected loci

🟩 neutral loci in genomic islands

⬛ neutral loci outside islands

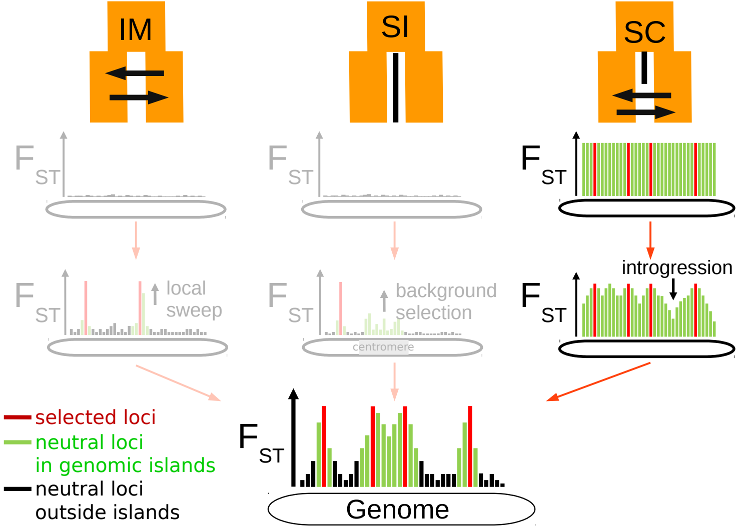
Endler (1977) Barton & Hewitt (1985) Noor & Bennett (2009) Bierne et al (2013) Cruickshank & Hahn (2014)

Multiple verbal models to explain observed patterns



Endler (1977) Barton & Hewitt (1985) Noor & Bennett (2009) Bierne et al (2013) Cruickshank & Hahn (2014)

Multiple verbal models to explain observed patterns



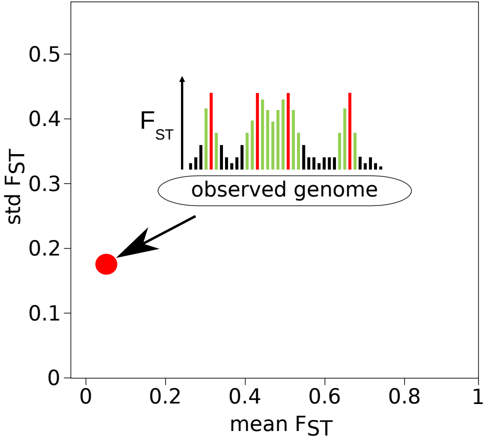
🟥 selected loci

🟩 neutral loci in genomic islands

■ neutral loci outside islands

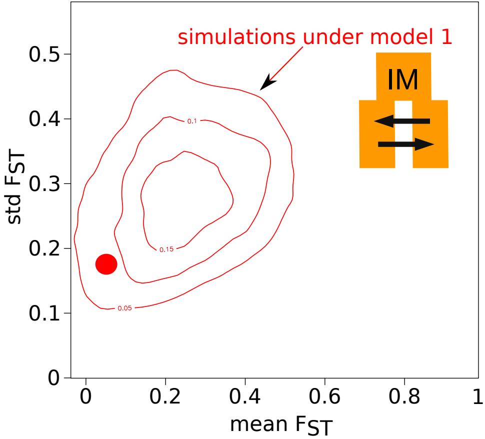
Endler (1977) Barton & Hewitt (1985) Noor & Bennett (2009) Bierne et al (2013) Cruickshank & Hahn (2014)

TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC



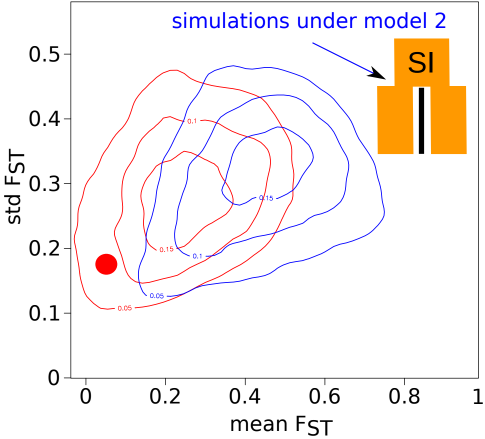
- Summaryzing the observed data by N statistics (here $N = 2$)
- Random simulations under the model *IM*
- Random simulations under the model *SI*
- Comparisons between observation and simulations

Testing for introgression between species using ABC



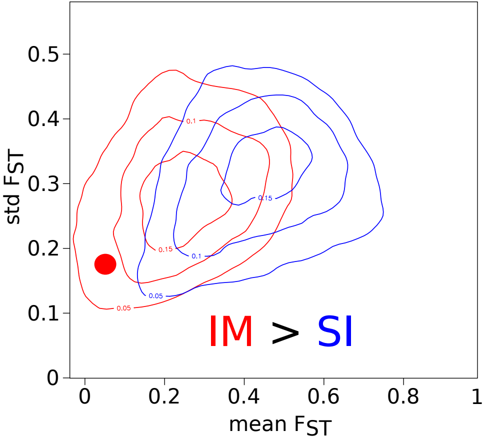
- Summaryzing the observed data by N statistics (here $N = 2$)
- Random simulations under the model *IM*
- Random simulations under the model *SI*
- Comparisons between observation and simulations

TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC



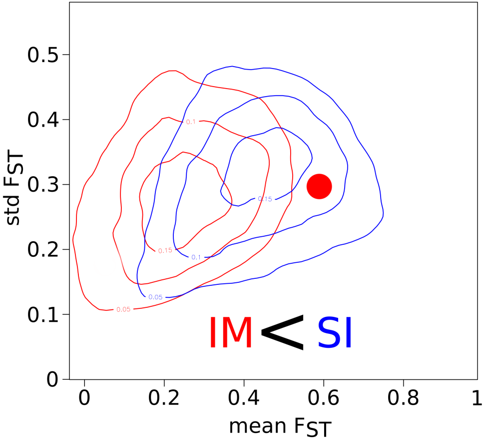
- Summaryzing the observed data by N statistics (here $N = 2$)
- Random simulations under the model *IM*
- Random simulations under the model *SI*
- Comparisons between observation and simulations

TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC



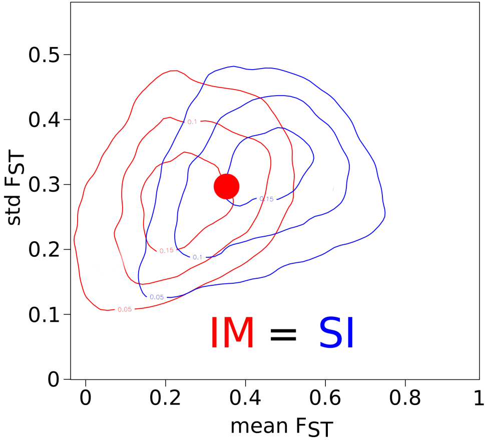
- Summaryzing the observed data by N statistics (here $N = 2$)
- Random simulations under the model *IM*
- Random simulations under the model *SI*
- Comparisons between observation and simulations

Testing for introgression between species using ABC



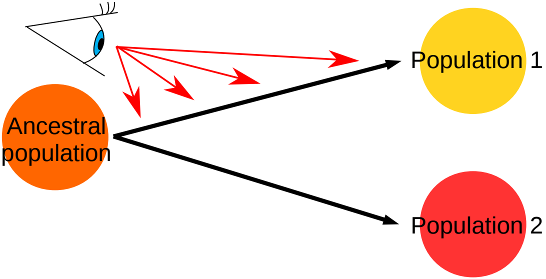
- Summaryzing the observed data by N statistics (here $N = 2$)
- Random simulations under the model *IM*
- Random simulations under the model *SI*
- Comparisons between observation and simulations

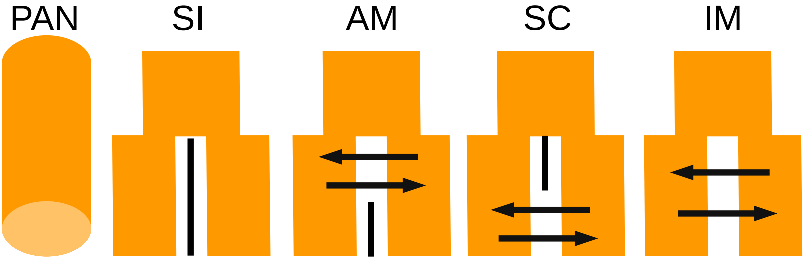
TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC



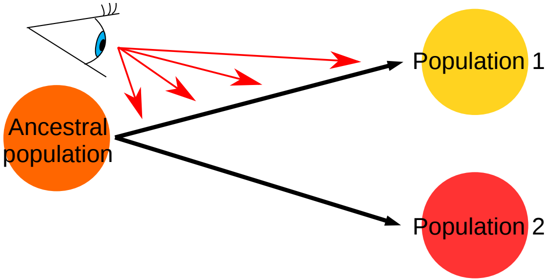
- Summaryzing the observed data by N statistics (here $N = 2$)
- Random simulations under the model *IM*
- Random simulations under the model *SI*
- Comparisons between observation and simulations

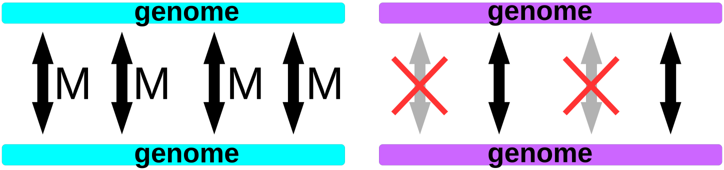
QUESTIONS:


Ancestral population
Population 1
Population 2


PAN
SI
AM
SC
IM

QUESTIONS:


Ancestral population
Population 1
Population 2


genome
M
M
M
M
genome
genome
genome

STUDIED DATASETS

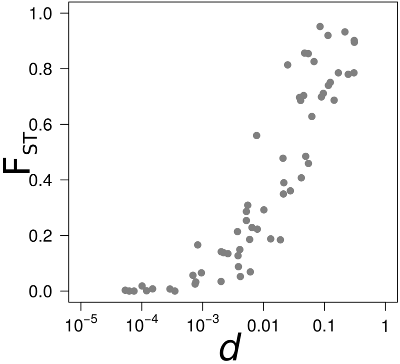
Published datasets

10 pairs of species

Obtained datasets

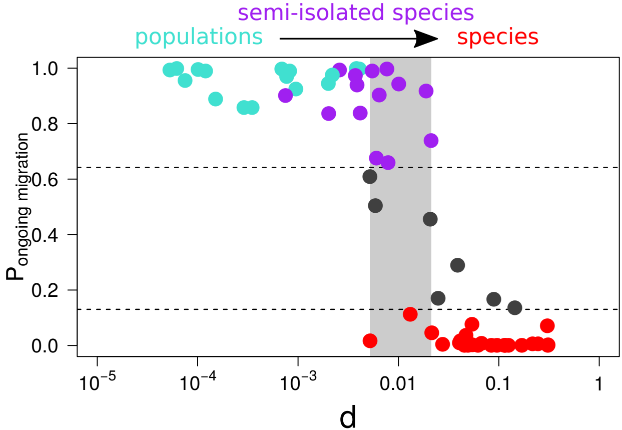
51 pairs of species

Explored range of divergence



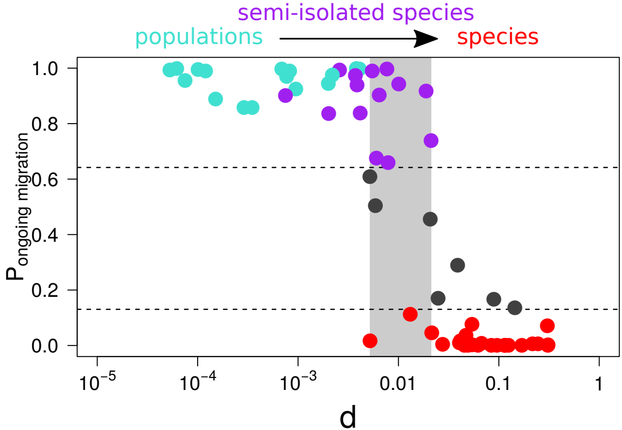
- Range of d: $5.10^{-5} - 0.31$
- Range of F_{ST}: $0 - 0.95$

GENERAL RELATION BETWEEN d AND $P_{ongoing\ migration}$



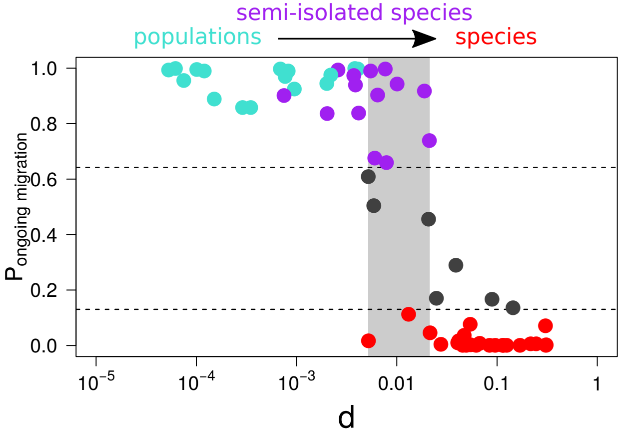
- First barriers detected for $d = 0.075\%$
- Important occurence of semi-isolated species
- Co-occurence of isolated and semi-isolated pairs for d in $[0.5\% - 2\%]$

GENERAL RELATION BETWEEN d AND $P_{ongoing\ migration}$



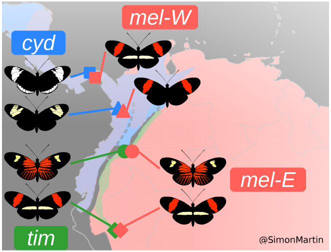
- First barriers detected for $d = 0.075\%$
- Important occurence of semi-isolated species
- Co-occurence of isolated and semi-isolated pairs for d in $[0.5\% - 2\%]$

GENERAL RELATION BETWEEN d AND $P_{ongoing\ migration}$



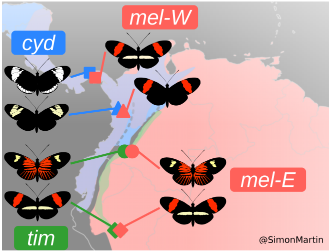
- First barriers detected for $d = 0.075\%$
- Important occurence of semi-isolated species
- Co-occurence of isolated and semi-isolated pairs for d *in* $[0.5\% - 2\%]$

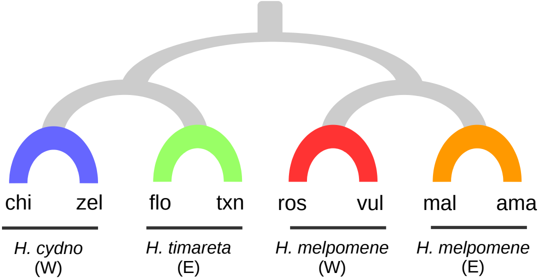
WITHIN THE SAME GENUS: *Heliconius*



- Sympatric and allopatric hybridising pairs
- Continuum of divergence
- Which pairs among 28 experiment ongoing introgression?
- How species barriers are distributed along genomes?

WITHIN THE SAME GENUS: *Heliconius*

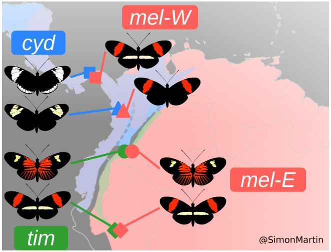


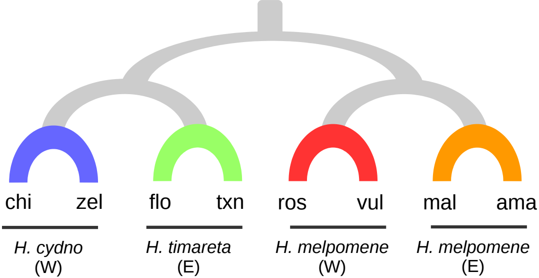


10 whole-genome sequenced individuals per population
(~270Mb in 795 contigs)

- Sympatric and allopatric hybridising pairs
- Continuum of divergence
- Which pairs among 28 experiment ongoing introgression?
- How species barriers are distributed along genomes?

WITHIN THE SAME GENUS: *Heliconius*

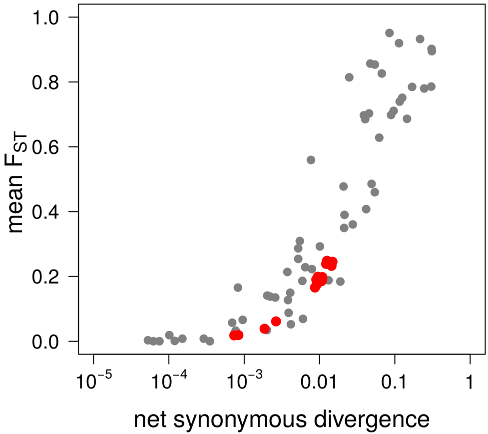




10 whole-genome sequenced individuals per population
(~270Mb in 795 contigs)

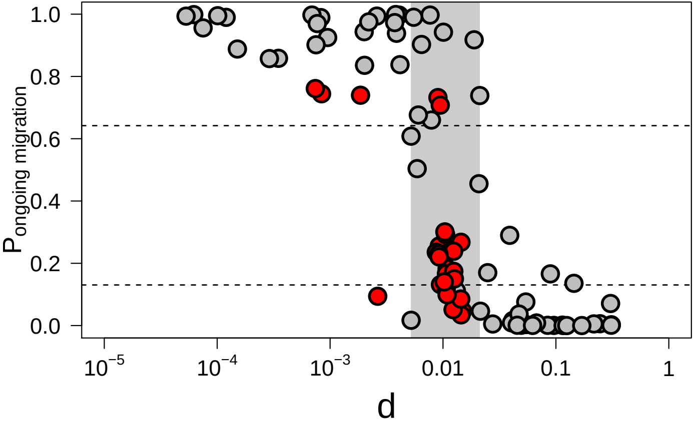
- Sympatric and allopatric hybridising pairs
- Continuum of divergence
- Which pairs among 28 experiment ongoing introgression?
- How species barriers are distributed along genomes?

EXPLORED RANGE OF DIVERGENCE IN *Heliconius*



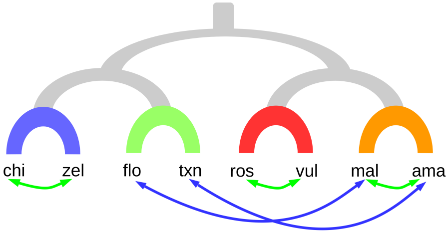
- Range of *d*: 0.07% – 1.5%
- Range of F_{ST}: 1.8% – 25%

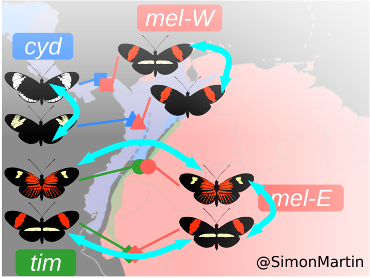
Tests for ongoing migration (ABC)



- 5 pairs supported by models with ongoing migration
- 7 pairs supported by models with current isolation
- 16 pairs not supported by any models

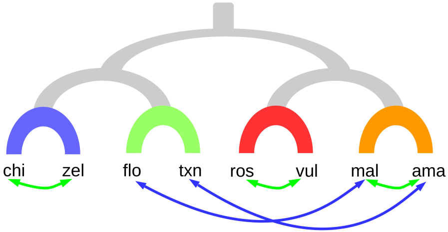
RESULTS OF MODEL COMPARISONS OVER 28 PAIRS

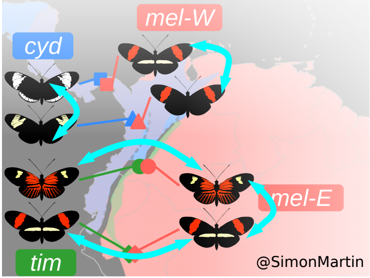




- Support for ongoing introgression for:
  - 3 pairs of populations over 4
  - 2 pairs of sympatric species over 4
  - 0 pair of allopatric species over 20
- Are there any species barriers?

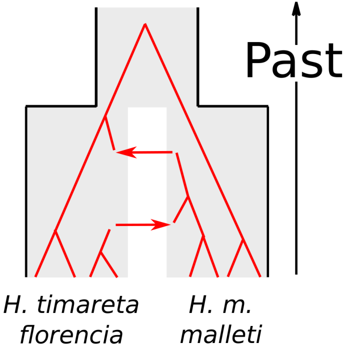
RESULTS OF MODEL COMPARISONS OVER 28 PAIRS





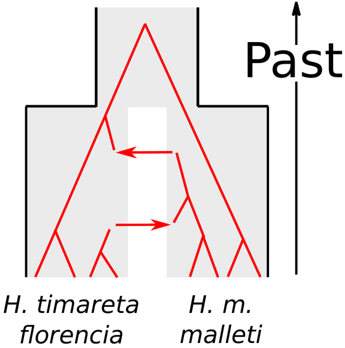
- Support for ongoing introgression for:
  - 3 pairs of populations over 4
  - 2 pairs of sympatric species over 4
  - 0 pair of allopatric species over 20
- Are there any species barriers?

LOCUS SPECIFIC MODEL COMPARISON



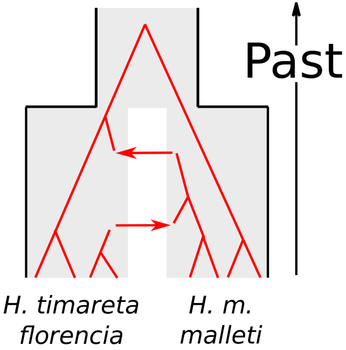
- **Estimating parameters shared by all loci**
  - Time of split
  - *Ne* (β distributed across the genome)
- **Simulating individual loci using the estimated model**
  - with $N.m = 0$ model M0
  - with $N.m > 0$ model M1
- **ABC model comparisons:**
  - 16,626 10kb windows (autosomes)
  - 7,727 coding genes
  - 1,032 10 kb windows (Z-chromosome)
  - 397 coding genes

LOCUS SPECIFIC MODEL COMPARISON



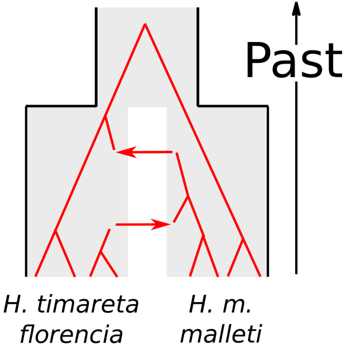
- **Estimating parameters shared by all loci**
  - Time of split
  - *Ne* (β distributed across the genome)
- **Simulating individual loci using the estimated model**
  - with $N.m = 0$ model M0
  - with $N.m > 0$ model M1
- **ABC model comparisons:**
  - 16,626 10kb windows (autosomes)
  - 7,727 coding genes
  - 1,032 10 kb windows (Z-chromosome)
  - 397 coding genes

LOCUS SPECIFIC MODEL COMPARISON



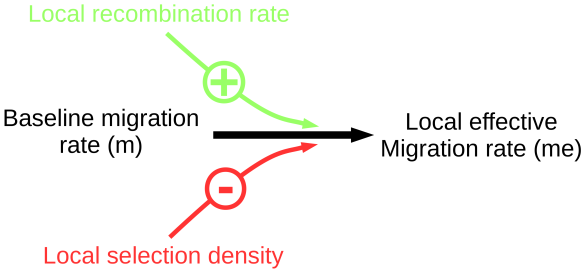
- **Estimating parameters shared by all loci**
  - Time of split
  - *Ne* (β distributed across the genome)
- **Simulating individual loci using the estimated model**
  - with $N.m = 0$ model M0
  - with $N.m > 0$ model M1
- **ABC model comparisons:**
  - 16,626 10kb windows (autosomes)
  - 7,727 coding genes
  - 1,032 10 kb windows (Z-chromosome)
  - 397 coding genes

LOCUS SPECIFIC MODEL COMPARISON



- **Estimating parameters shared by all loci**
  - Time of split
  - *Ne* (β distributed across the genome)
- **Simulating individual loci using the estimated model**
  - with $N.m = 0$ model M0
  - with $N.m > 0$ model M1
- **ABC model comparisons:**
  - 16,626 10kb windows (autosomes)
  - 7,727 coding genes
  - 1,032 10 kb windows (Z-chromosome)
  - 397 coding genes

HOW THE LOCI WITH A REDUCED *me* ARE DISTRIBUTED?



Local recombination rate

+

Baseline migration rate (m)

Local effective Migration rate (me)

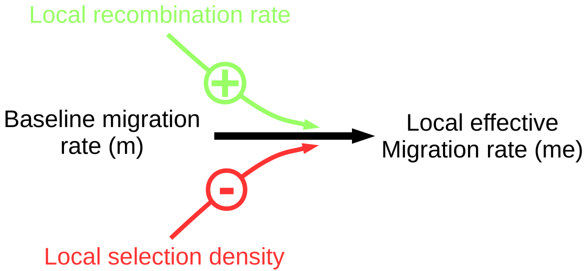
-

Local selection density

Aeschbacher et *al.* 2017

- Increased recombination rates maintain *me* close to *m*
- Increased selection density reduces *me*
- Subdivision of the genome in 9 bins of equal size:
  - low/intermediate/high recombination rates
  - low/intermediate/high selection density

HOW THE LOCI WITH A REDUCED *me* ARE DISTRIBUTED?



Local recombination rate

⊕

Baseline migration rate (m)

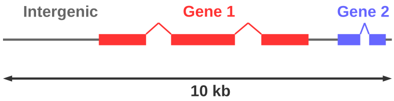
Local effective Migration rate (me)

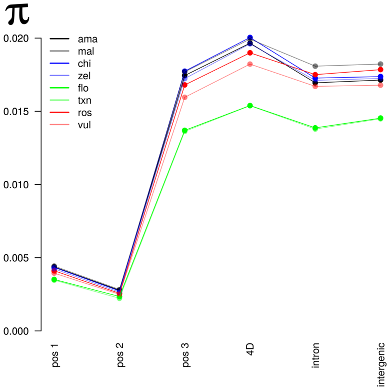
-

Local selection density

Aeschbacher et *al.* 2017

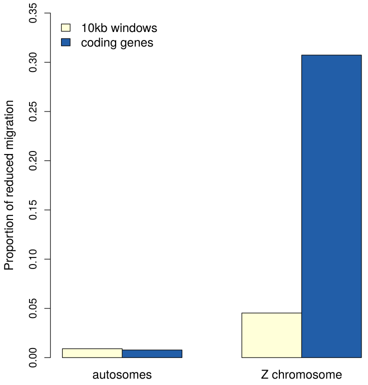
- Increased recombination rates maintain *me* close to *m*
- Increased selection density reduces *me*
- Subdivision of the genome in 9 bins of equal size:
  - low/intermediate/high recombination rates
  - low/intermediate/high selection density





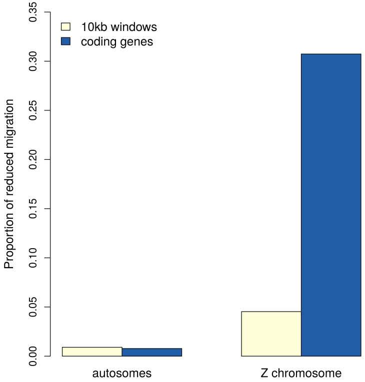
- Coding density: percentage of 1*st* and 2*nd* coding positions in a genomic window around the midpoint (10kb)
- Recombination rates: four-gamete test (FGT) on unphased genomic data

GENOMIC DISTRIBUTION OF INTROGRESSION EVENTS



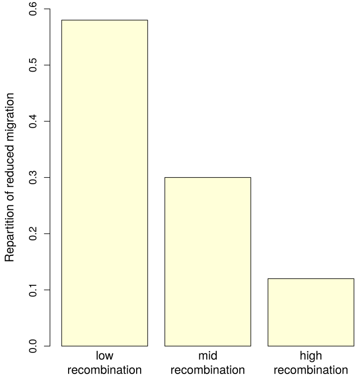
- **Species barriers represent a low fraction of the genome**
  - ≈ 0.86% of the autosomal genome
- **Species barriers are concentrated in:**
  - coding genes of the Z chromosome
  - lower recombining regions

GENOMIC DISTRIBUTION OF INTROGRESSION EVENTS



- **Species barriers represent a low fraction of the genome**
  - $\approx 0.86\%$ of the autosomal genome
- **Species barriers are concentrated in:**
  - coding genes of the Z chromosome
  - lower recombining regions

GENOMIC DISTRIBUTION OF INTROGRESSION EVENTS



- **Species barriers represent a low fraction of the genome**
  - $\approx 0.86\%$ of the autosomal genome
- **Species barriers are concentrated in:**
  - coding genes of the Z chromosome
  - lower recombining regions

TAKE HOME MESSAGES

- First barriers are detected for $d \approx 0.075\%$
- No gene flow detected for $d > 2\%$
- In *Heliconius* : ≈ 6 times more loci inferred as being isolated in Low Recombination regions than in High Recombination
- In *Heliconius* : Z-chromosome is more associated to species barriers than autosomes

TAKE HOME MESSAGES

- First barriers are detected for $d \approx 0.075\%$
- No gene flow detected for $d > 2\%$
- In *Heliconius* : $\approx$ 6 times more loci inferred as being isolated in Low Recombination regions than in High Recombination
- In *Heliconius* : Z-chromosome is more associated to species barriers than autosomes

TAKE HOME MESSAGES

- First barriers are detected for $d \approx 0.075\%$
- No gene flow detected for $d > 2\%$
- In *Heliconius* : ≈ 6 times more loci inferred as being isolated in Low Recombination regions than in High Recombination
- In *Heliconius* : Z-chromosome is more associated to species barriers than autosomes


Jonathan Romiguier


Chris Jiggins


Nicolas Bierne


Nicolas Galtier


Simon Martin


Christelle Fraïsse