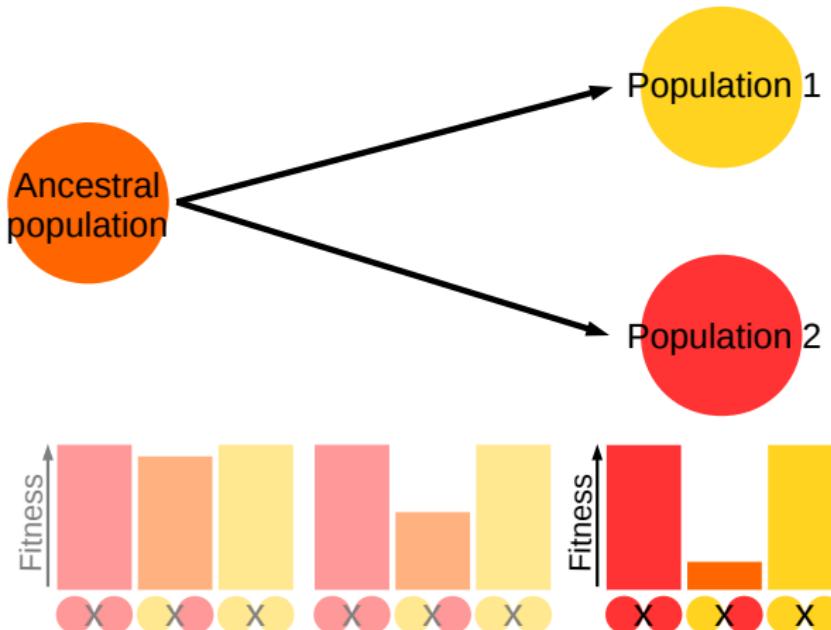


Inferring the relationship between molecular divergence and genetic isolation

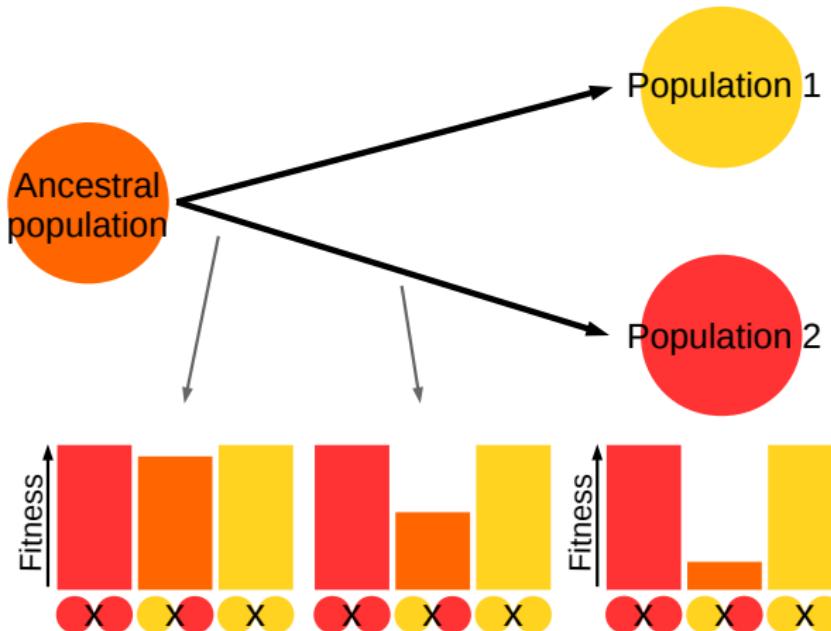
GDR «Génomique Environnementale» et «Approche Interdisciplinaire de l'Evolution Moléculaire»
16th May 2018

SPECIATION: EVOLUTION OF REPRODUCTIVE ISOLATION



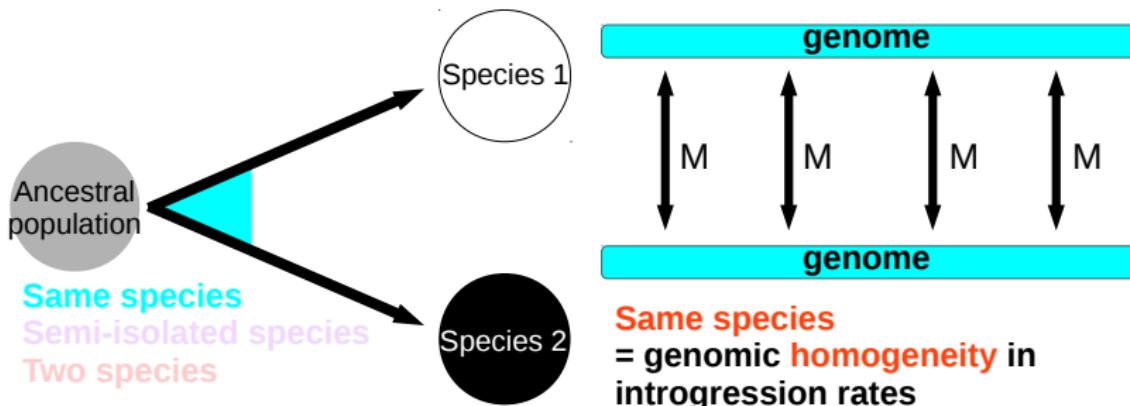
- ▶ Parental fitness > Hybrid fitness
- ▶ Number and effects of barriers increase over time

SPECIATION: EVOLUTION OF REPRODUCTIVE ISOLATION



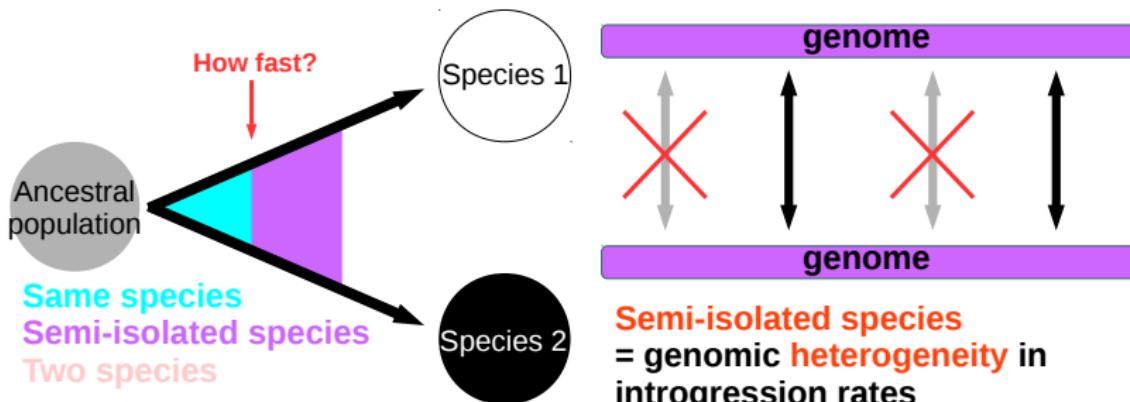
- ▶ Parental fitness > Hybrid fitness
- ▶ Number and effects of barriers increase over time

FROM ONE TO TWO SPECIES



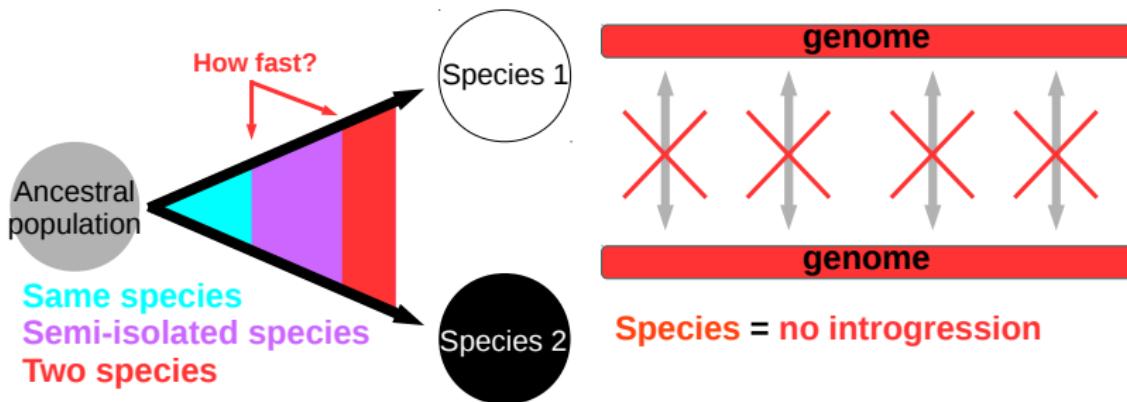
- When can we detect the effects of the first barriers on gene flow?
- Is there a threshold of divergence above which gene flow is impossible?
- No integrative theory of speciation connecting 'divergence' to 'reproductive isolation'

FROM ONE TO TWO SPECIES



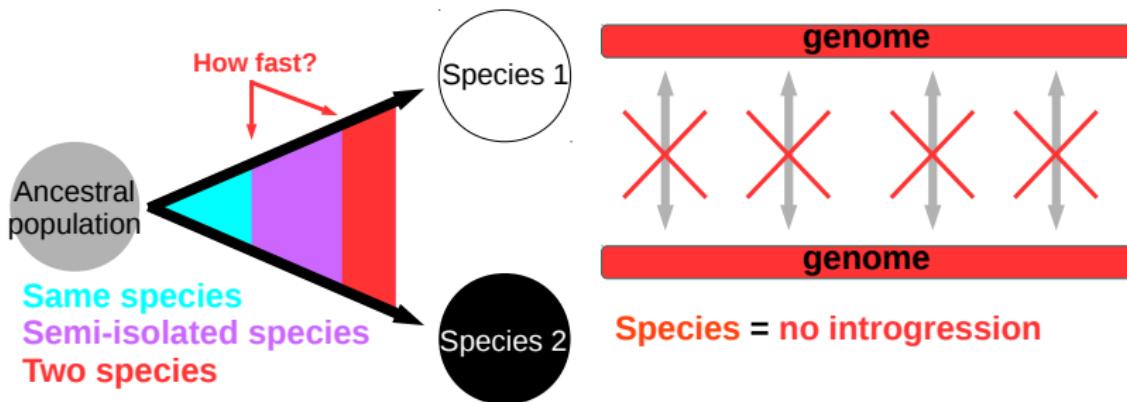
- ▶ When can we detect the effects of the first barriers on gene flow?
- ▶ Is there a threshold of divergence above which gene flow is impossible?
- ▶ No integrative theory of speciation connecting 'divergence' to 'reproductive isolation'

FROM ONE TO TWO SPECIES



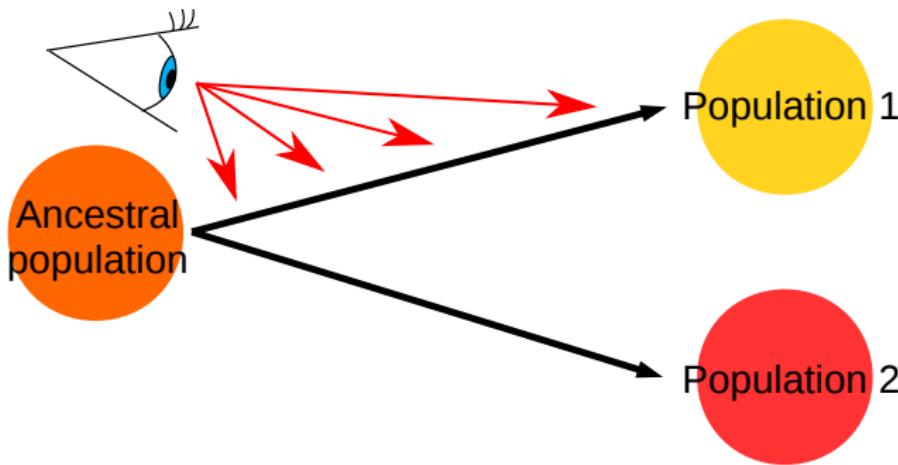
- When can we detect the effects of the first barriers on gene flow?
- Is there a threshold of divergence above which gene flow is impossible?
- No integrative theory of speciation connecting 'divergence' to 'reproductive isolation'

FROM ONE TO TWO SPECIES



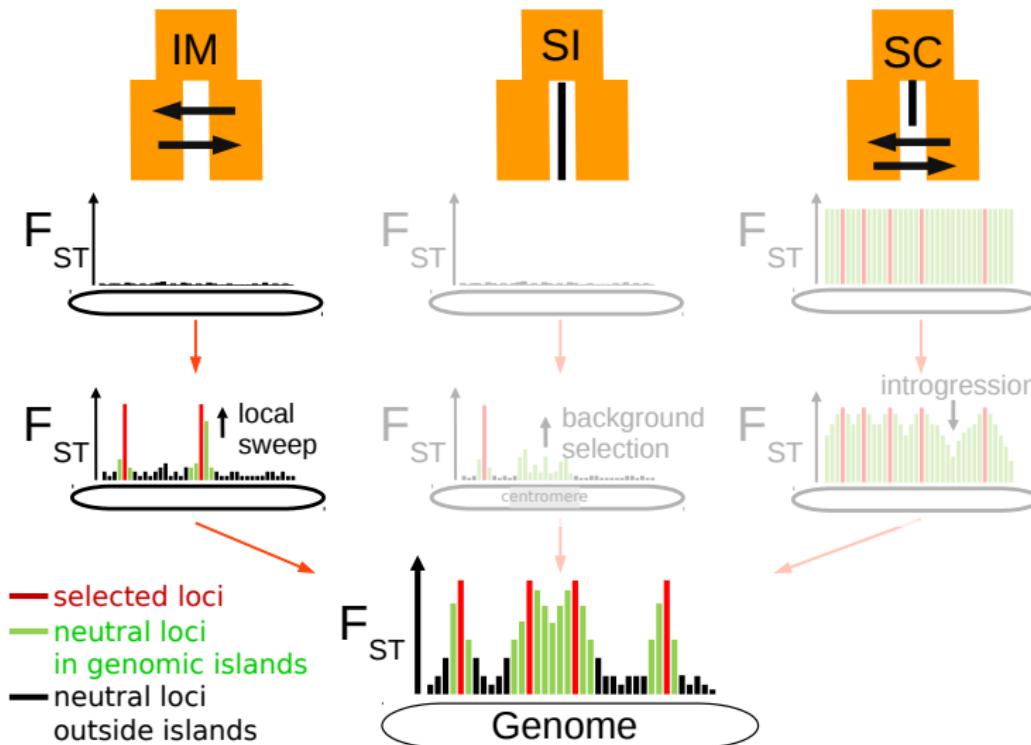
- When can we detect the effects of the first barriers on gene flow?
- Is there a threshold of divergence above which gene flow is impossible?
- No integrative theory of speciation connecting 'divergence' to 'reproductive isolation'

QUESTIONS:

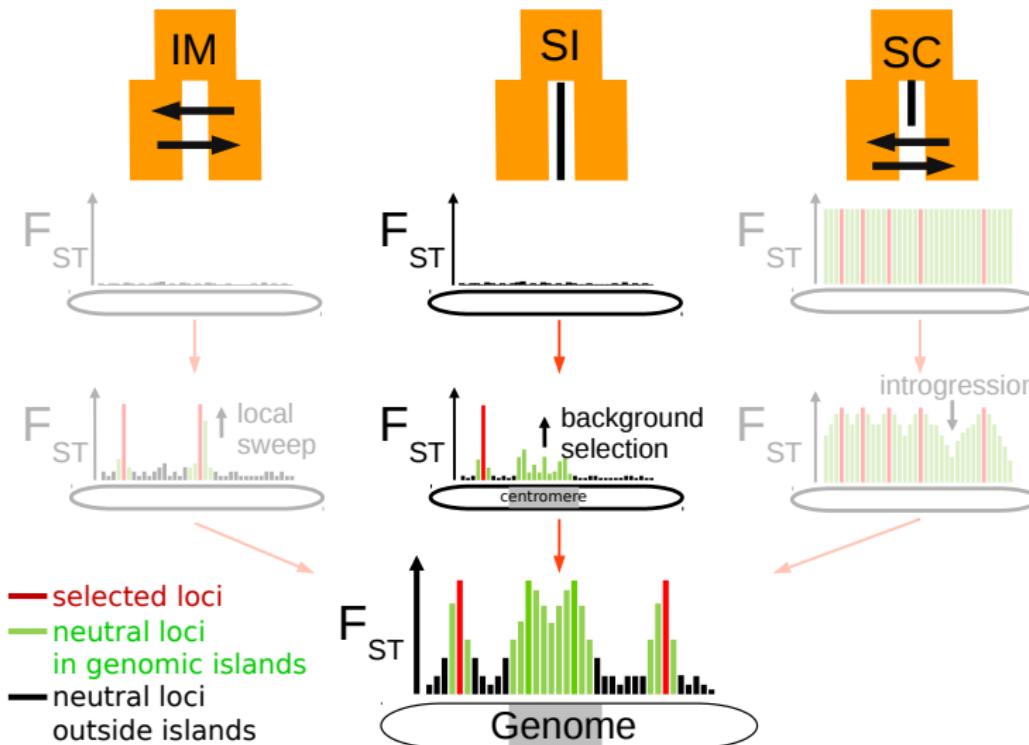


- ▶ Comparing alternative scenarios for 61 pairs of species along a continuum of d (ABC)

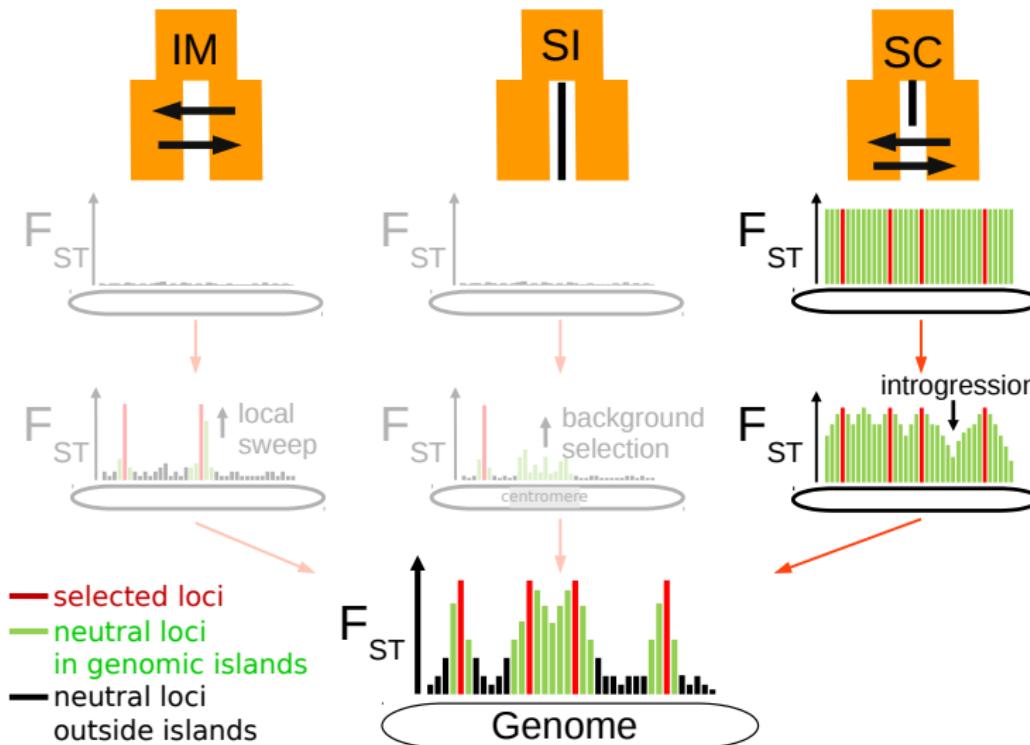
MULTIPLE VERBAL MODELS TO EXPLAIN OBSERVED PATTERNS



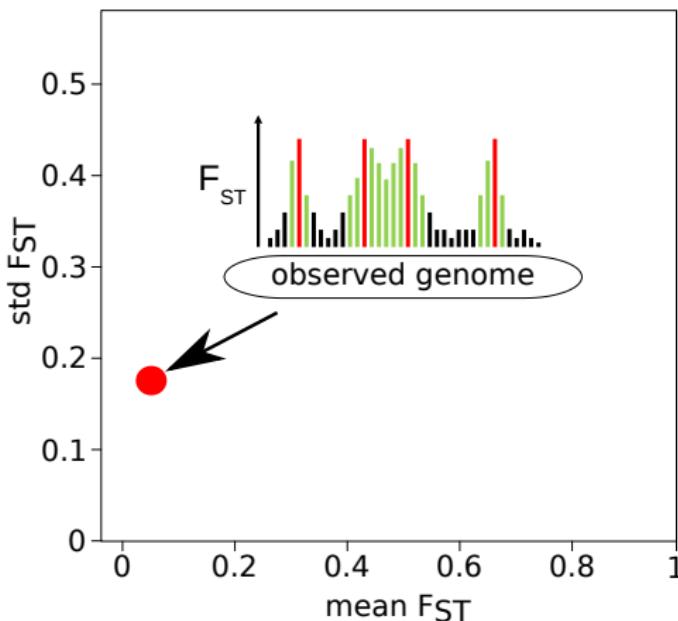
MULTIPLE VERBAL MODELS TO EXPLAIN OBSERVED PATTERNS



MULTIPLE VERBAL MODELS TO EXPLAIN OBSERVED PATTERNS

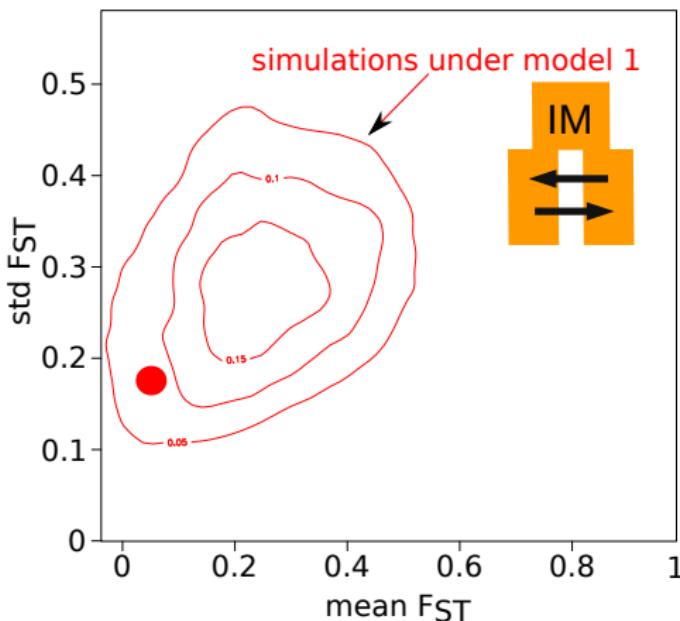


TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC



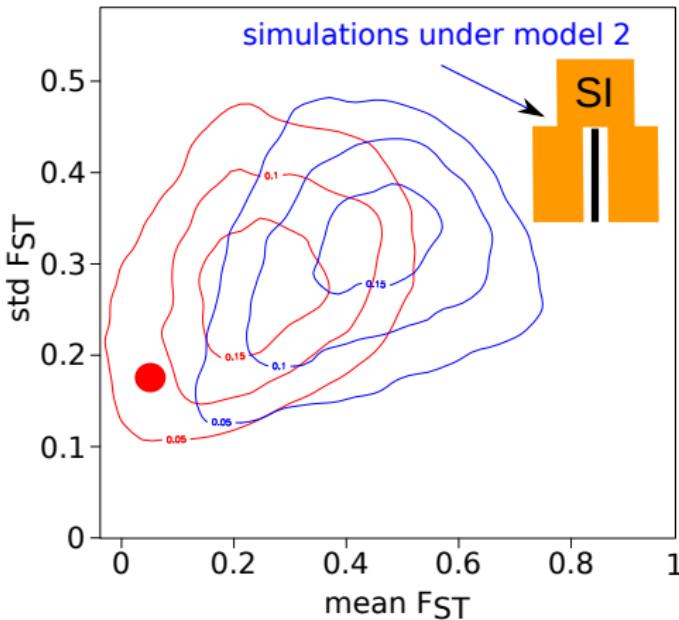
- ▶ Summarying the observed data by N statistics (here $N = 2$)
- ▶ Random simulations under the model IM
- ▶ Random simulations under the model SI
- ▶ Comparisons between observation and simulations

TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC



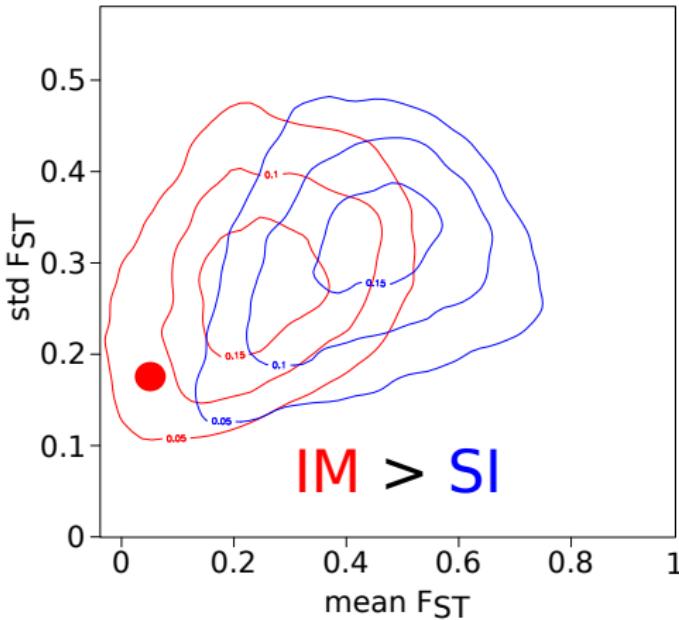
- ▶ Summaryzing the observed data by N statistics (here $N = 2$)
- ▶ Random simulations under the model IM
- ▶ Random simulations under the model SI
- ▶ Comparisons between observation and simulations

TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC



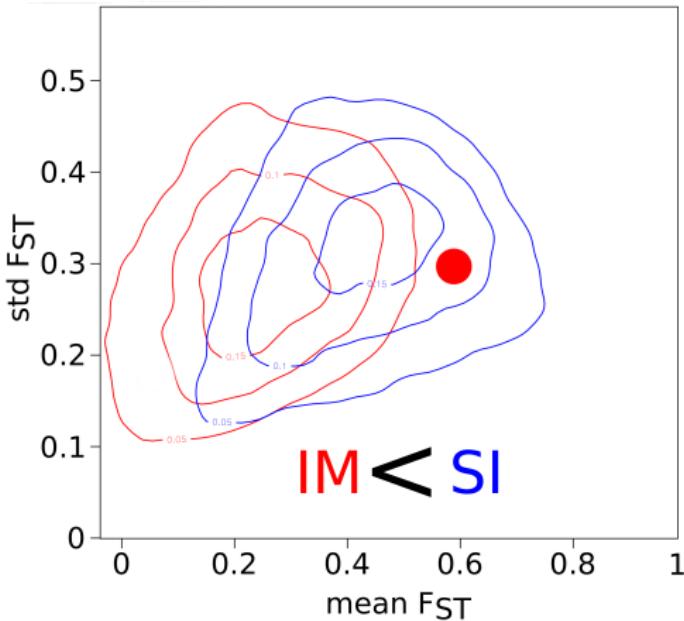
- ▶ Summarying the observed data by N statistics (here $N = 2$)
- ▶ Random simulations under the model IM
- ▶ Random simulations under the model SI
- ▶ Comparisons between observation and simulations

TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC



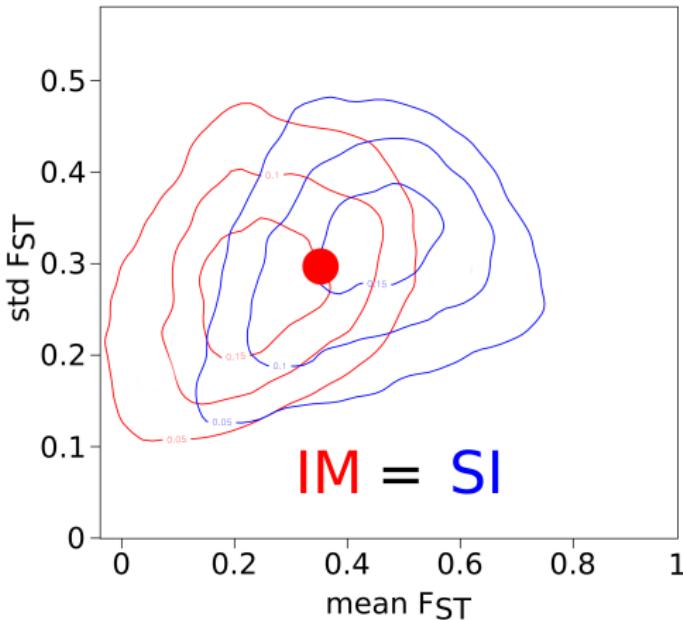
- ▶ Summarying the observed data by N statistics (here $N = 2$)
- ▶ Random simulations under the model IM
- ▶ Random simulations under the model SI
- ▶ Comparisons between observation and simulations

TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC



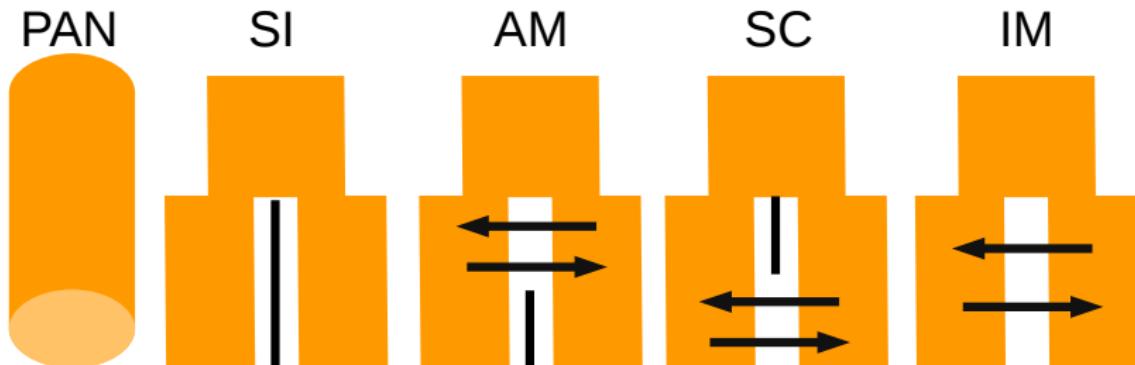
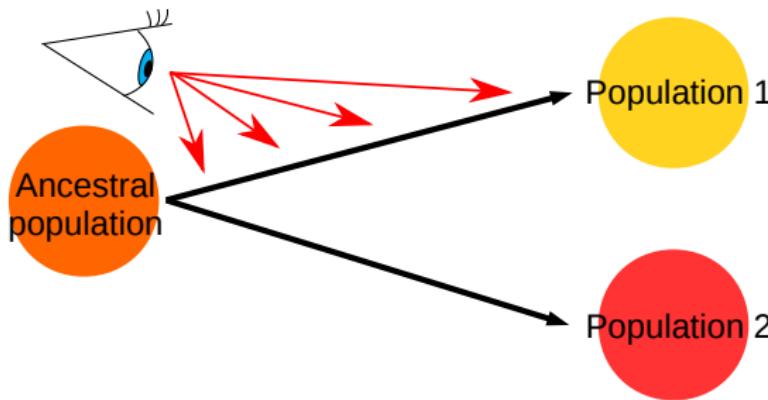
- ▶ Summarying the observed data by N statistics (here $N = 2$)
- ▶ Random simulations under the model *IM*
- ▶ Random simulations under the model *SI*
- ▶ Comparisons between observation and simulations

TESTING FOR INTROGRESSION BETWEEN SPECIES USING ABC

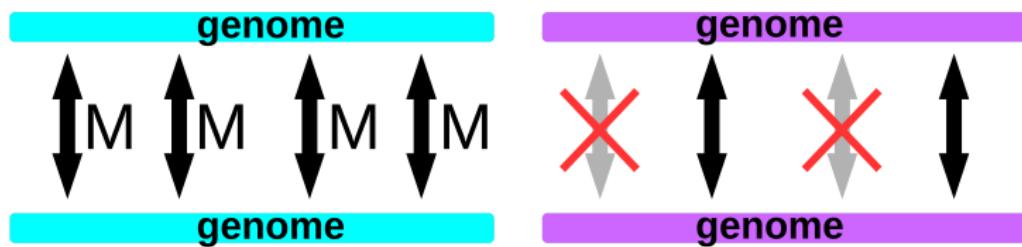
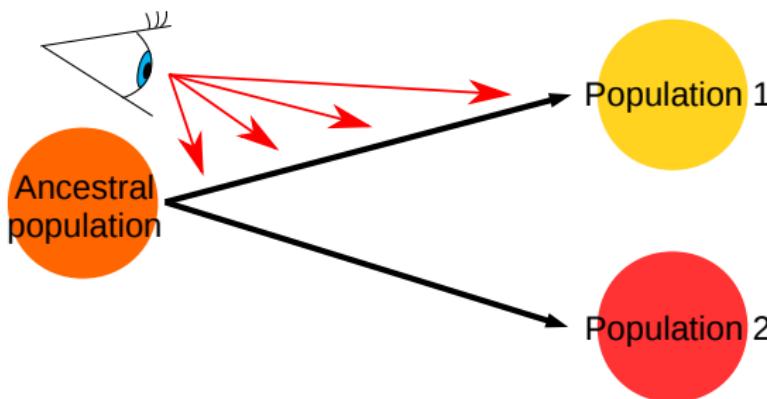


- ▶ Summarying the observed data by N statistics (here $N = 2$)
- ▶ Random simulations under the model *IM*
- ▶ Random simulations under the model *SI*
- ▶ Comparisons between observation and simulations

QUESTIONS:



QUESTIONS:



STUDIED DATASETS

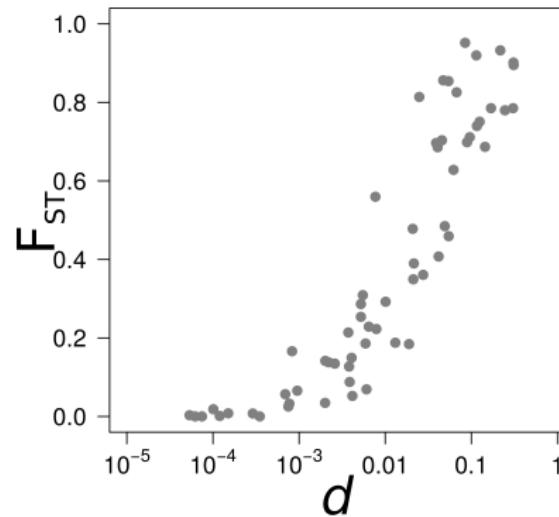
Published datasets
10 pairs of species



Obtained datasets
51 pairs of species

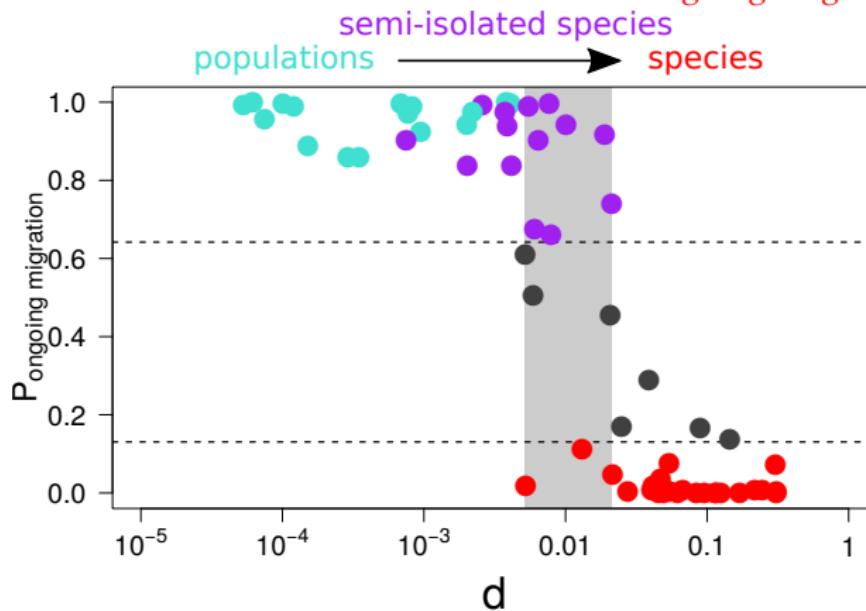


EXPLORED RANGE OF DIVERGENCE



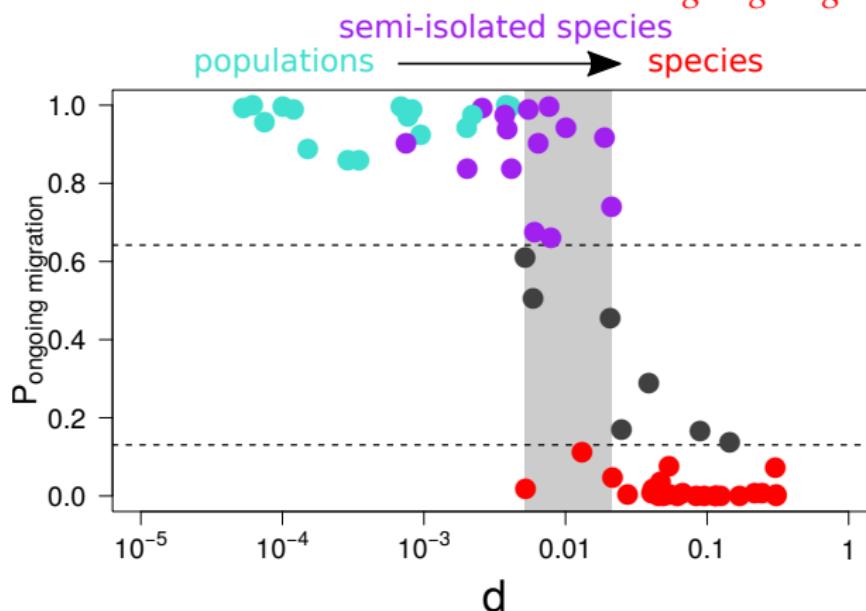
- Range of d : $5.10^{-5} – 0.31$
- Range of F_{ST} : $0 – 0.95$

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



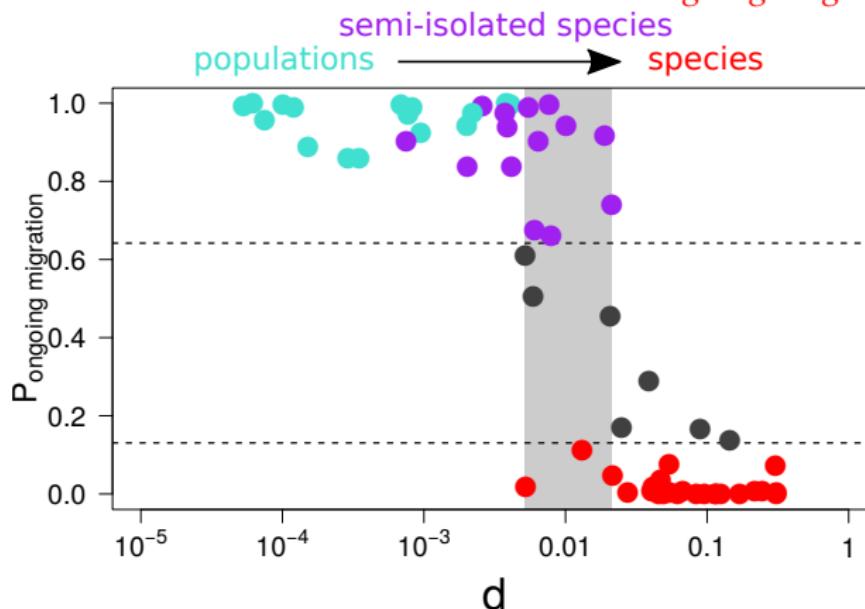
- ▶ First barriers detected for $d = 0.075\%$
- ▶ Important occurrence of semi-isolated species
- ▶ Co-occurrence of isolated and semi-isolated pairs for d in $[0.5\% - 2\%]$

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



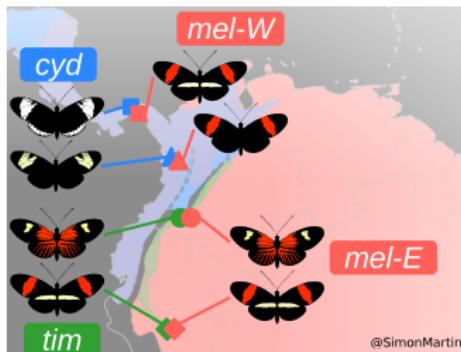
- ▶ First barriers detected for $d = 0.075\%$
- ▶ Important occurrence of semi-isolated species
- ▶ Co-occurrence of isolated and semi-isolated pairs for d in $[0.5\% - 2\%]$

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



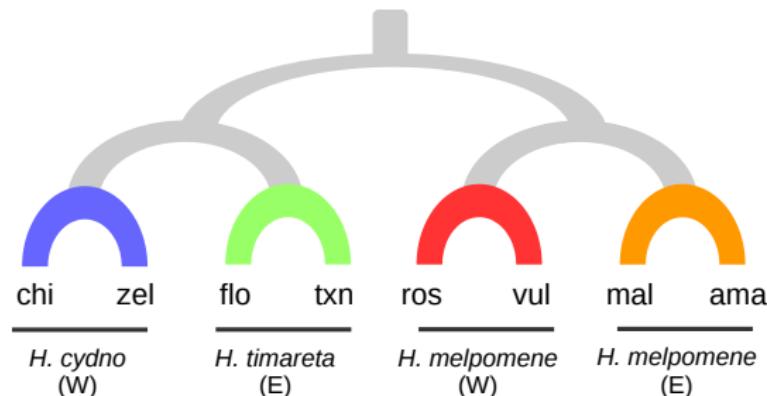
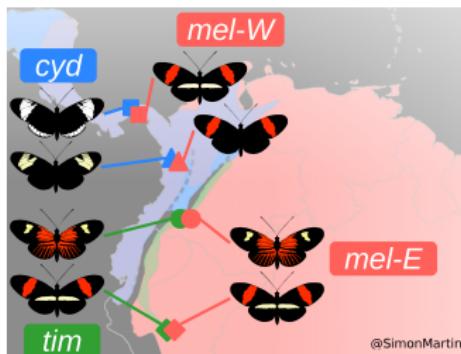
- ▶ First barriers detected for $d = 0.075\%$
- ▶ Important occurrence of semi-isolated species
- ▶ Co-occurrence of isolated and semi-isolated pairs for d in $[0.5\% - 2\%]$

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?

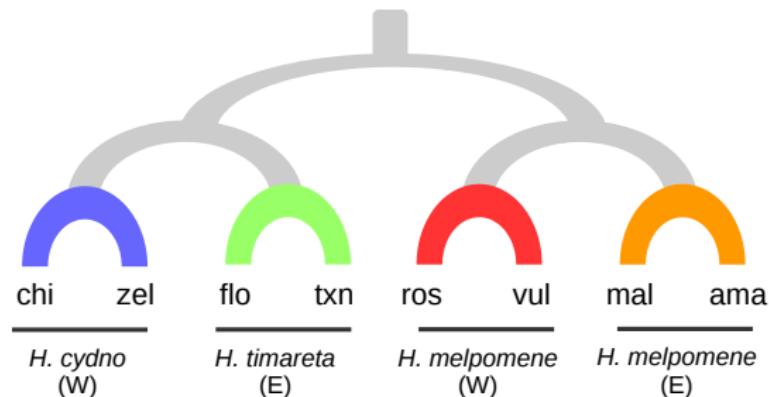
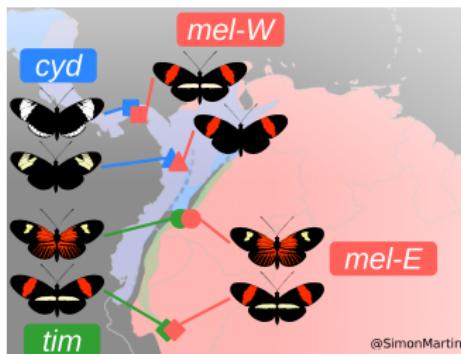
WITHIN THE SAME GENUS: *Heliconius*



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

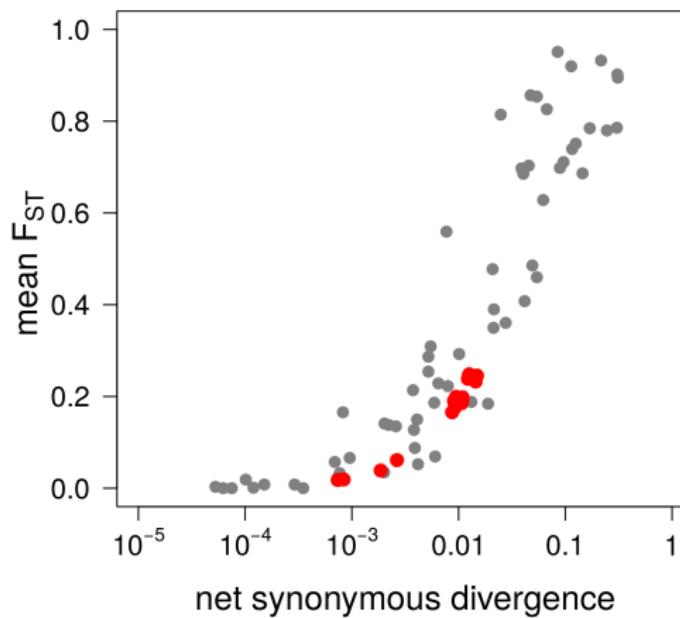
- ▶ Sympatric and allopatric hybridising pairs
- ▶ Continuum of divergence
- ▶ Which pairs among 28 experiment ongoing introgression?
- ▶ How species barriers are distributed along genomes?

WITHIN THE SAME GENUS: *Heliconius*



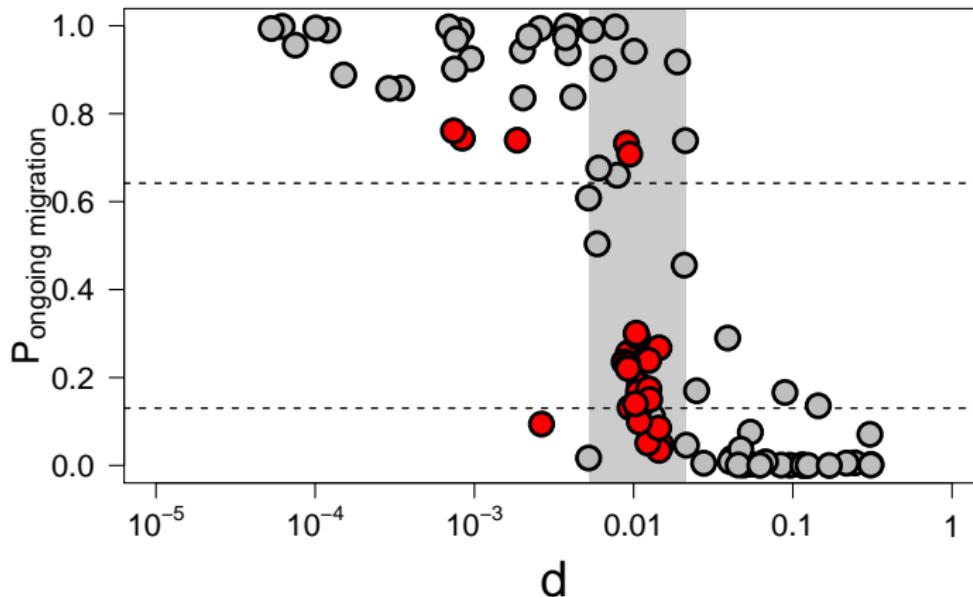
10 whole-genome sequenced individuals per population
(~270Mb in 795 contigs)

- ▶ 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*

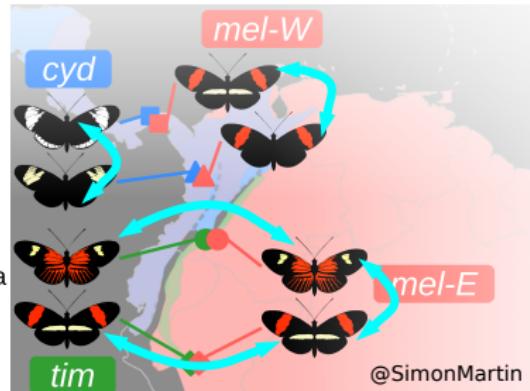
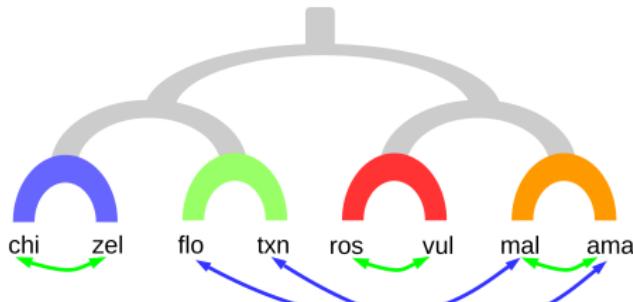
- Range of d : 0.07% – 1.5%
- Range of F_{ST} : 1.8% – 25%

TESTS FOR ONGOING MIGRATION (ABC)



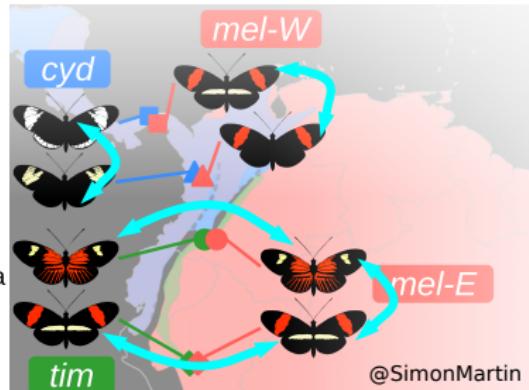
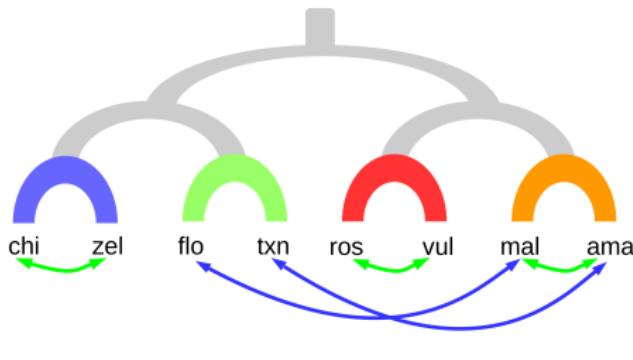
- ▶ 5 pairs supported by models with ongoing migration
- ▶ 7 pairs supported by models with current isolation
- ▶ 16 pairs not supported by any models

RESULTS OF MODEL COMPARISONS OVER 28 PAIRS



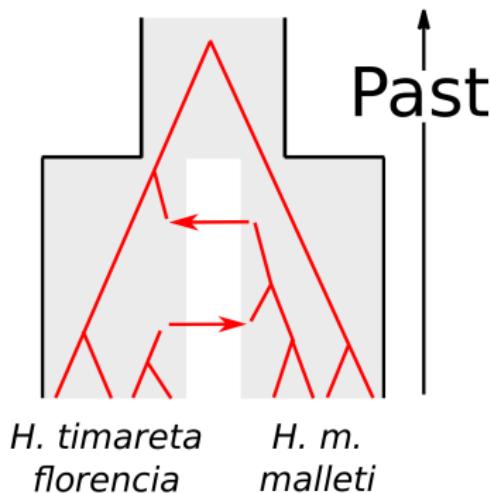
- ▶ Support for ongoing introgression for:
 - ▶ 3 pairs of **populations** over 4
 - ▶ 2 pairs of **sympatric species** over 4
 - ▶ 0 pair of **allopatric species** over 20
- ▶ Are there any species barriers?

RESULTS OF MODEL COMPARISONS OVER 28 PAIRS



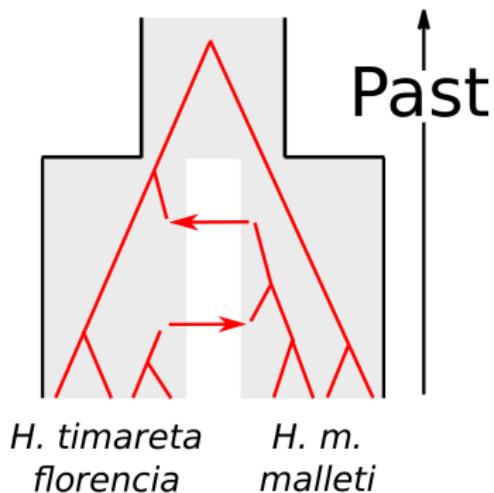
- ▶ Support for ongoing introgression for:
 - ▶ 3 pairs of **populations** over 4
 - ▶ 2 pairs of **sympatric species** over 4
 - ▶ 0 pair of **allopatric species** over 20
 - ▶ Are there any species barriers?

LOCUS SPECIFIC MODEL COMPARISON



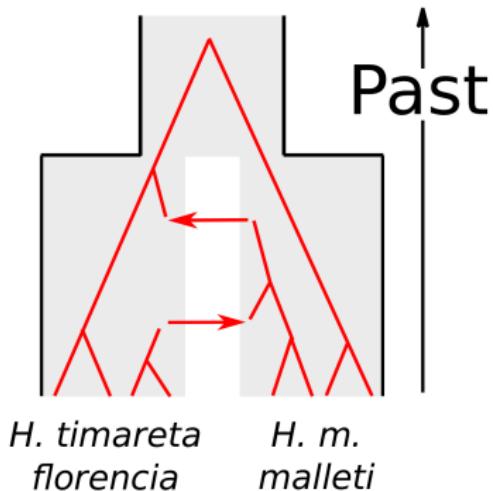
- ▶ Estimating parameters shared by all loci
 - ▶ Time of split
 - ▶ N_e (β distributed across the genome)
- ▶ Simulating individual loci using the estimated model
 - ▶ with $N.m = 0$ model M0
 - ▶ with $N.m > 0$ model M1
- ▶ ABC model comparison for 10,557 loci
 - ▶ 1,723 ambiguous
 - ▶ 8,531 supported by M1
 - ▶ 303 supported by M0

LOCUS SPECIFIC MODEL COMPARISON



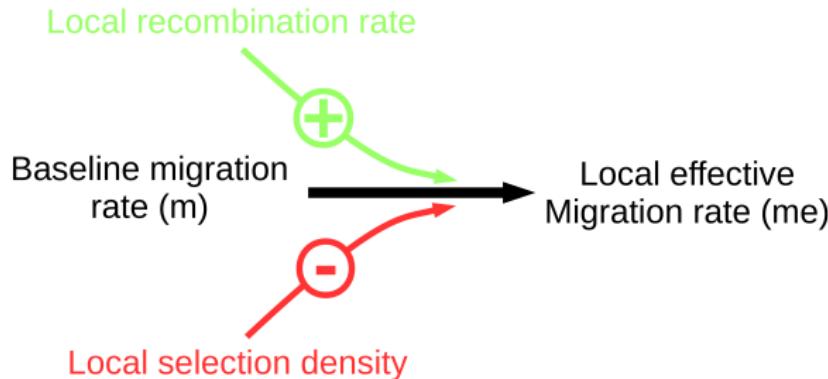
- ▶ Estimating parameters shared by all loci
 - ▶ Time of split
 - ▶ N_e (β distributed across the genome)
- ▶ Simulating individual loci using the estimated model
 - ▶ with $N.m = 0$ model M0
 - ▶ with $N.m > 0$ model M1
- ▶ ABC model comparison for 10,557 loci
 - ▶ 1,723 ambiguous
 - ▶ 8,531 supported by M1
 - ▶ 303 supported by M0

LOCUS SPECIFIC MODEL COMPARISON



- ▶ Estimating parameters shared by all loci
 - ▶ Time of split
 - ▶ N_e (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 comparison for 10,557 loci
 - ▶ 1,723 ambiguous
 - ▶ 8,531 supported by M1
 - ▶ 303 supported by M0

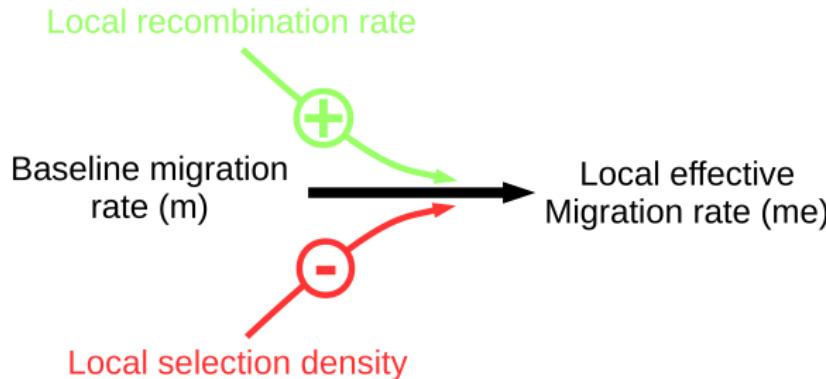
HOW THE 303 LOCI WITH A REDUCED me ARE DISTRIBUTED?



Aeschbacher et al. 2017

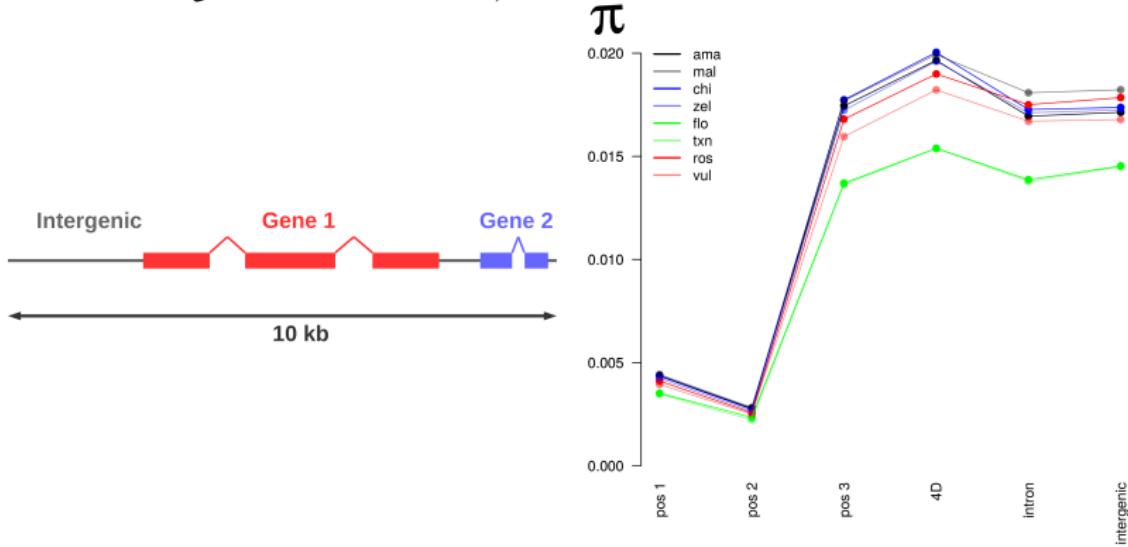
- ▶ Increased **recombination rates** maintain me close to m
- ▶ Increased **selection density** reduces me
- ▶ Subdivision of the genome in 9 bins of $\approx 1,170$ genes:
 - ▶ low/intermediate/high recombination rates
 - ▶ low/intermediate/high selection density

HOW THE 303 LOCI WITH A REDUCED me ARE DISTRIBUTED?



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 $\approx 1,170$ genes:
 - ▶ low/intermediate/high recombination rates
 - ▶ low/intermediate/high selection density

PRODUCING 9 BINS OF $\approx 1,170$ GENES

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

GENOMIC DISTRIBUTION OF THE 303 LOCI WITH A REDUCED-*me*

		Recombination		
		Low	Intermediate	High
Density in coding positions	High	21.1 %	11.9 %	3.3 %
	Intermediate	20.1 %	9.2 %	4.6 %
	Low	17.8 %	8.6 %	3.3 %

- ▶ 6 times more loci inferred as being isolated in **LowRecomb+HighDensity** regions than in **HighRecomb+LowDensity**
 - ▶ Does background selection increases the rate of false positive?
 - ▶ 6 times more DMIs in **LowRecomb+HighDensity** than in **HighRecomb+LowDensity**?

GENOMIC DISTRIBUTION OF THE 303 LOCI WITH A REDUCED-*me*

		Recombination		
		Low	Intermediate	High
Density in coding positions	High	21.1 %	11.9 %	3.3 %
	Intermediate	20.1 %	9.2 %	4.6 %
	Low	17.8 %	8.6 %	3.3 %

- ▶ 6 times more loci inferred as being isolated in **LowRecomb+HighDensity** regions than in **HighRecomb+LowDensity**
 - ▶ Does background selection increases the rate of false positive?
 - ▶ 6 times more DMIs in **LowRecomb+HighDensity** than in **HighRecomb+LowDensity**?

GENOMIC DISTRIBUTION OF THE 303 LOCI WITH A REDUCED-*me*

		Recombination		
		Low	Intermediate	High
Density in coding positions	High	21.1 %	11.9 %	3.3 %
	Intermediate	20.1 %	9.2 %	4.6 %
	Low	17.8 %	8.6 %	3.3 %

- ▶ 6 times more loci inferred as being isolated in **LowRecomb+HighDensity** regions than in **HighRecomb+LowDensity**
 - ▶ Does background selection increases the rate of false positive?
 - ▶ 6 times more DMIs in **LowRecomb+HighDensity** than in **HighRecomb+LowDensity**?

TAKE HOME MESSAGES

- ▶ First barriers are detected for $d \approx 0.075\%$
- ▶ No gene flow detected for $d > 2\%$
- ▶ 6 times more loci inferred as being isolated in **LowRecomb+HighDensity** regions than in **HighRecomb+LowDensity**

TAKE HOME MESSAGES

- ▶ First barriers are detected for $d \approx 0.075\%$
- ▶ No gene flow detected for $d > 2\%$
- ▶ 6 times more loci inferred as being isolated in **LowRecomb+HighDensity** regions than in **HighRecomb+LowDensity**

TAKE HOME MESSAGES

- ▶ First barriers are detected for $d \approx 0.075\%$
- ▶ No gene flow detected for $d > 2\%$
- ▶ 6 times more loci inferred as being isolated in **LowRecomb+HighDensity** regions than in **HighRecomb+LowDensity**



Jonathan Romiguier



Chris Jiggins



Nicolas Bierne



Nicolas Galtier



Simon Martin



Christelle Fraïsse

erc