

Comparative divergence genomics enlightens a large grey zone of speciation

Camille Roux¹, Christelle Fraïsse², Jonathan Romiguier³,
Yoann Anciaux¹, Nicolas Galtier¹ and Nicolas Bierne¹

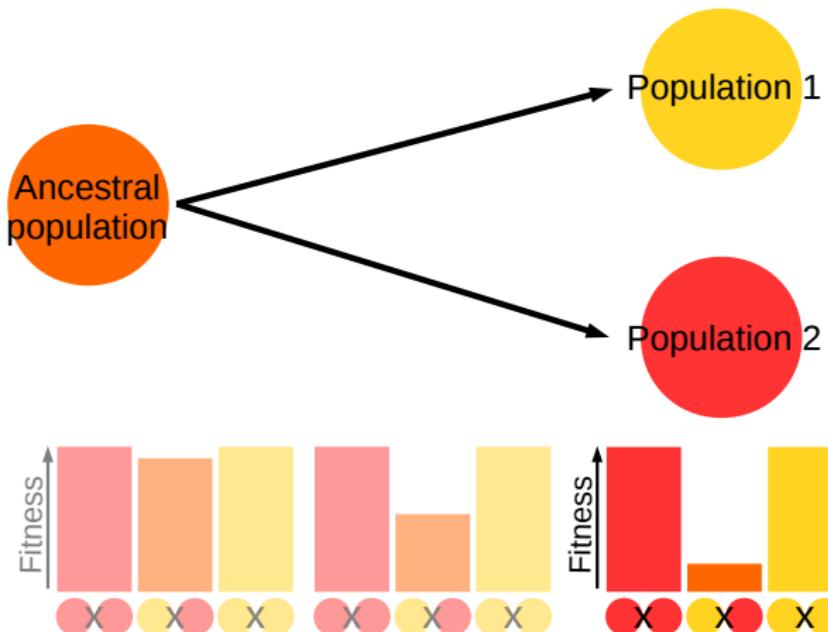
¹Institut des Sciences de l'Évolution (Montpellier, France)

²Institute of Science and Technology (Vienna, Austria)

³Department of Ecology and Evolution (Lausanne, Switzerland)

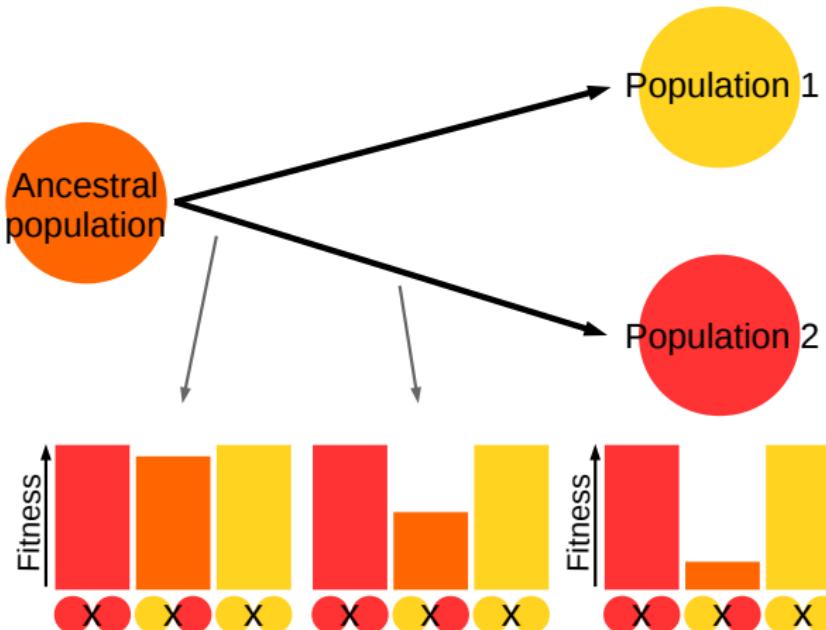
May 25, 2016

SPECIATION: EVOLUTION OF REPRODUCTIVE ISOLATION



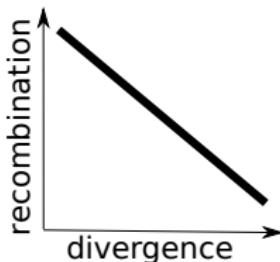
- ▶ Parental fitness > Hybrid fitness
- ▶ Number and effects of barriers increase over time

SPECIATION: EVOLUTION OF REPRODUCTIVE ISOLATION



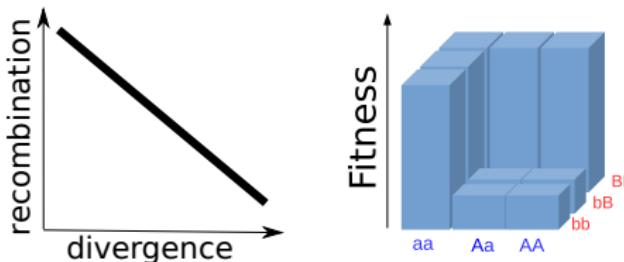
- ▶ Parental fitness > Hybrid fitness
- ▶ Number and effects of barriers increase over time

CAUSES OF POSTZYGOTIC ISOLATION



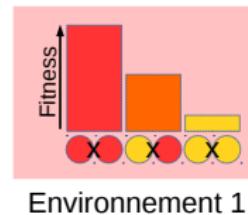
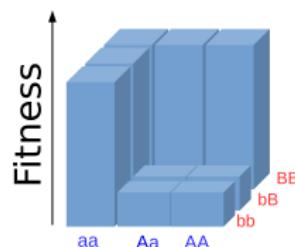
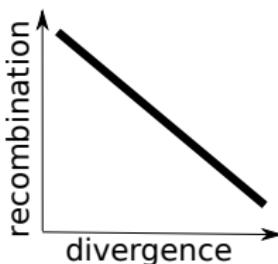
- ▶ Decreased recombination with divergence d
- ▶ Accumulation of Dobzhansky-Muller incompatibilities with d
- ▶ Hybrids can be unadapted to parental environments

CAUSES OF POSTZYGOTIC ISOLATION



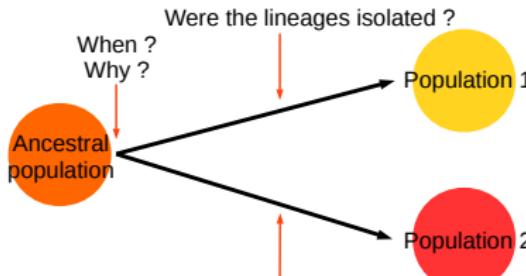
- ▶ Decreased recombination with divergence d
- ▶ Accumulation of Dobzhansky-Muller incompatibilities with d
- ▶ Hybrids can be unadapted to parental environments

CAUSES OF POSTZYGOTIC ISOLATION



- Decreased recombination with divergence d
- Accumulation of Dobzhansky-Muller incompatibilities with d
- Hybrids can be unadapted to parental environments

STUDYING SPECIATION



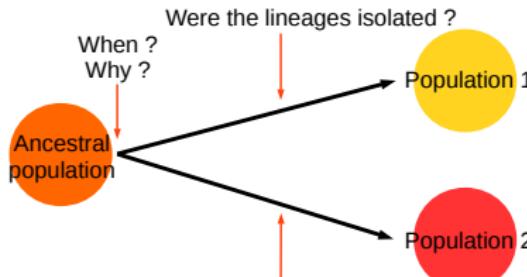
What was the rate of accumulation of barriers ?

What was the order of appearance of barriers ?

What forces drove the fixation of barriers ?

- ▶ Retracing history of populations and mutations
- ▶ 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 direct observation → main source of information available are current genomes

STUDYING SPECIATION



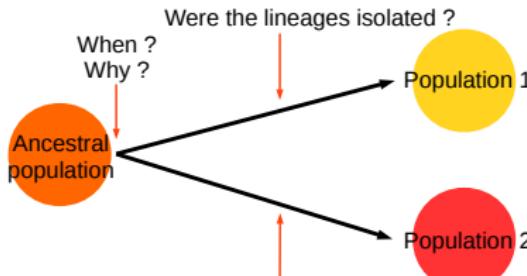
What was the rate of accumulation of barriers ?

What was the order of appearance of barriers ?

What forces drove the fixation of barriers ?

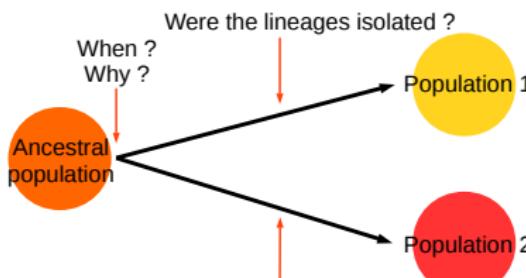
- ▶ Retracing history of populations and mutations
- ▶ When can we detect the effects of the first barriers on gene flow?
- ▶ Is there a threshold of divergence above which gene flow is impossible?
- ▶ No direct observation → main source of information available are current genomes

STUDYING SPECIATION

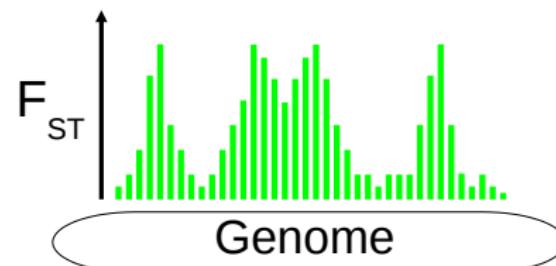


- ▶ Retracing history of populations and mutations
- ▶ When can we detect the effects of the first barriers on gene flow?
- ▶ Is there a threshold of divergence above which gene flow is impossible?
- ▶ No direct observation → main source of information available are current genomes

STUDYING SPECIATION

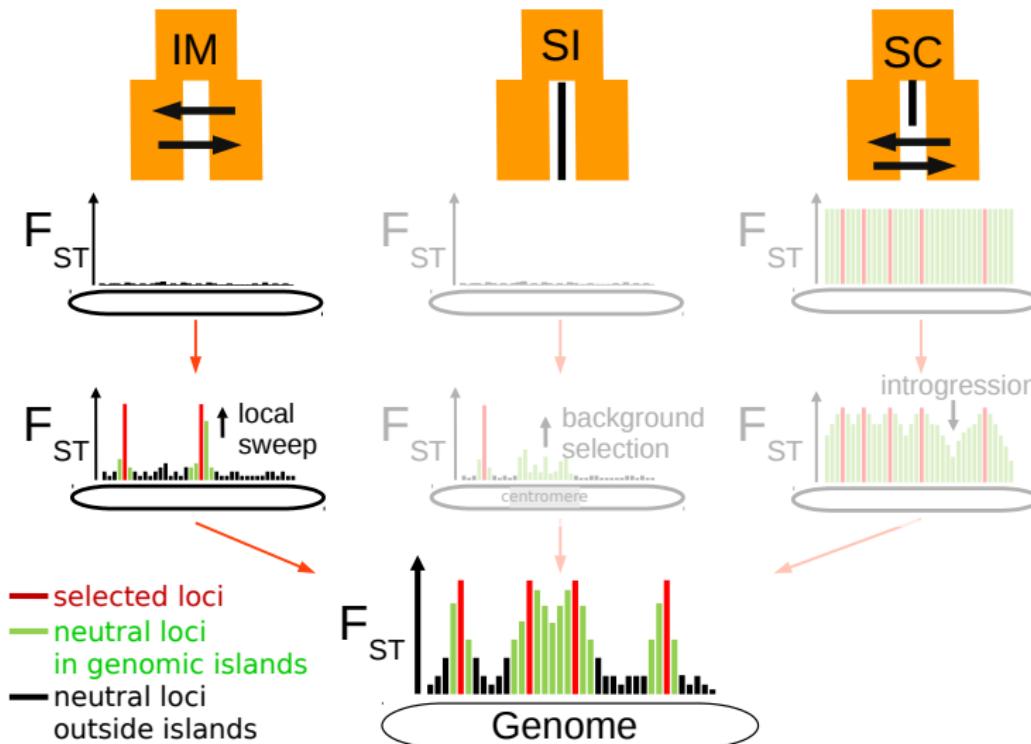


What was the rate of accumulation of barriers?
What was the order of appearance of barriers?
What forces drove the fixation of barriers?

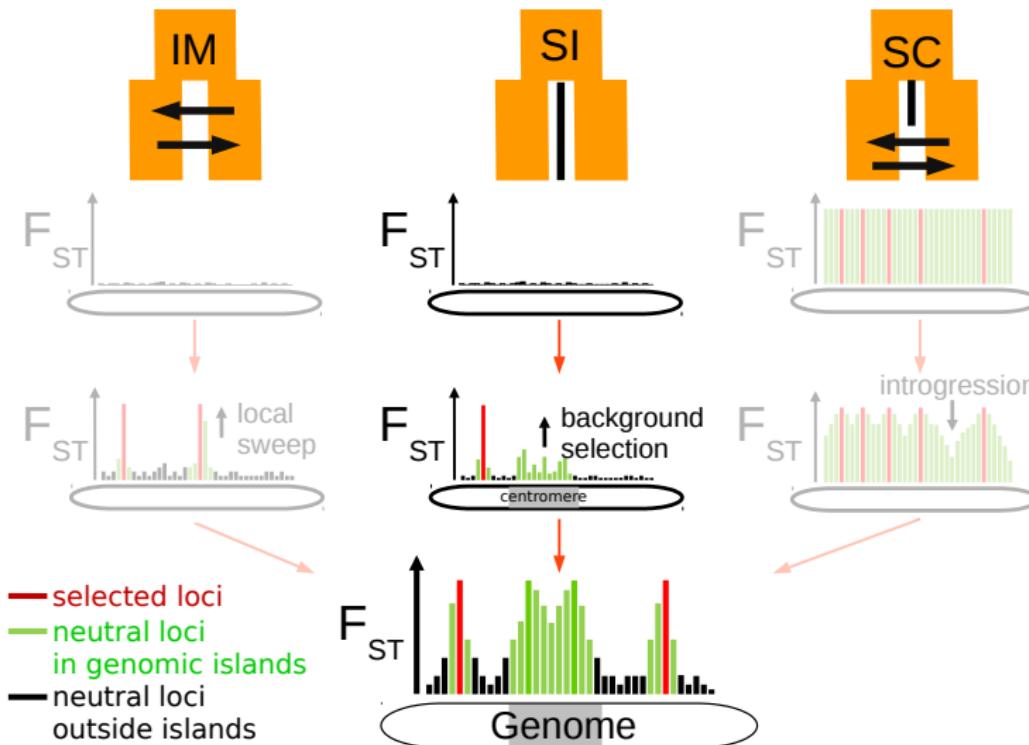


- ▶ Retracing history of populations and mutations
- ▶ When can we detect the effects of the first barriers on gene flow?
- ▶ Is there a threshold of divergence above which gene flow is impossible?
- ▶ No direct observation → main source of information available are current genomes

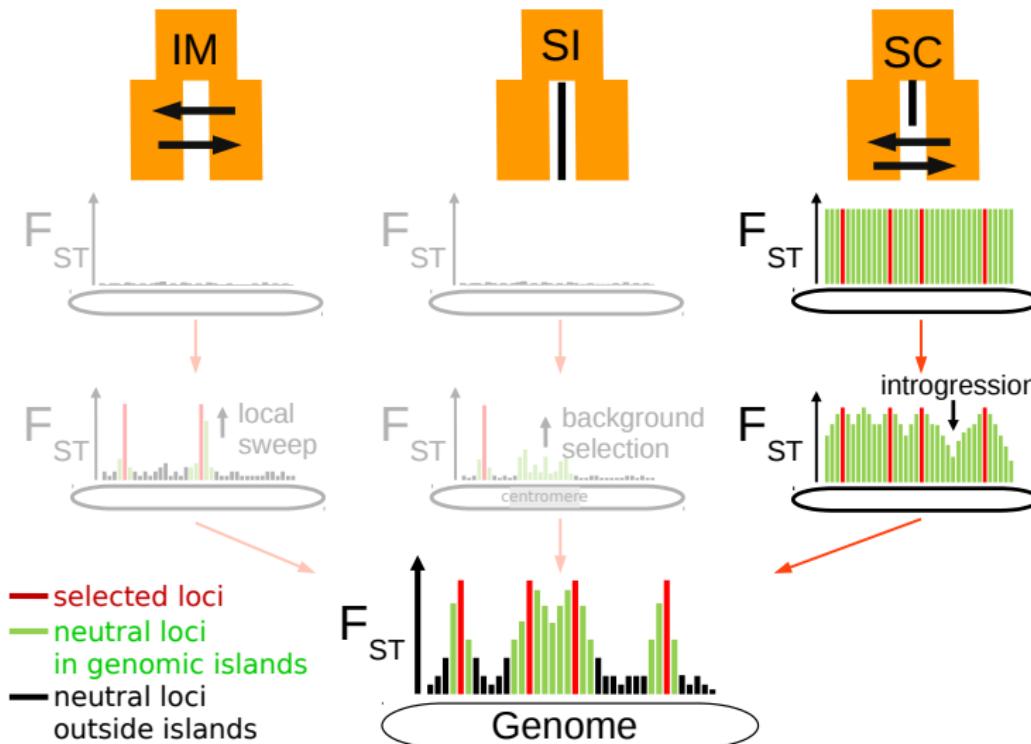
MULTIPLE VERBAL MODELS TO EXPLAIN OBSERVED PATTERNS



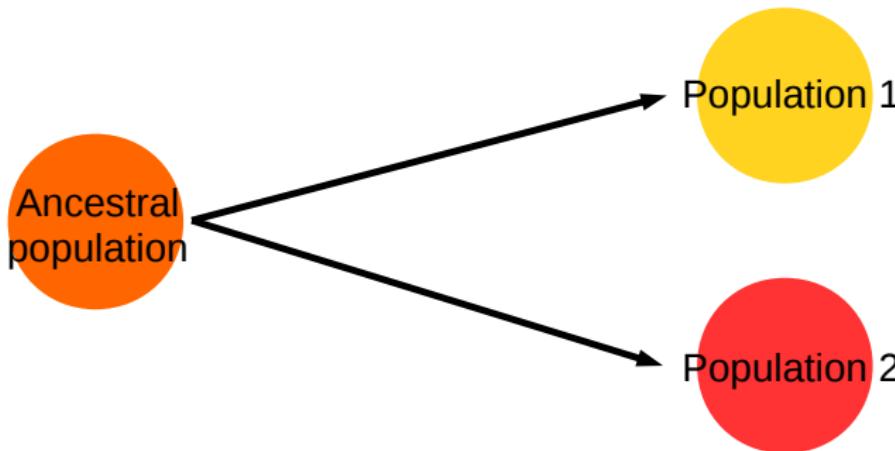
MULTIPLE VERBAL MODELS TO EXPLAIN OBSERVED PATTERNS



MULTIPLE VERBAL MODELS TO EXPLAIN OBSERVED PATTERNS

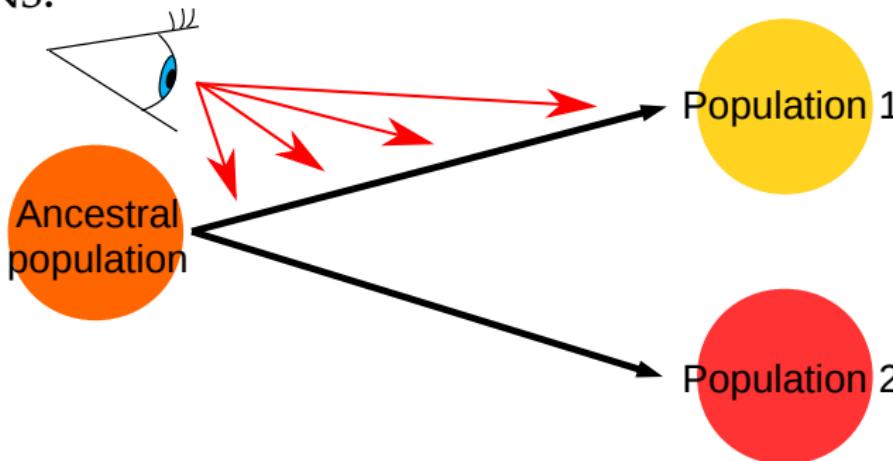


QUESTIONS:



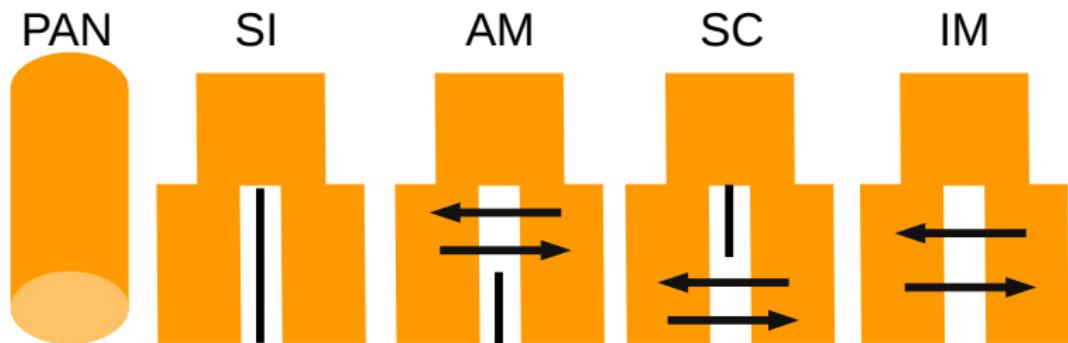
- ▶ When can we detect the effects of the first barriers?
- ▶ Is there a threshold of divergence above which gene flow is impossible?
- ▶ Testing gene flow for 61 pairs of species along a continuum of d

QUESTIONS:



- ▶ When can we detect the effects of the first barriers?
- ▶ Is there a threshold of divergence above which gene flow is impossible?
- ▶ Testing gene flow for 61 pairs of species along a continuum of d

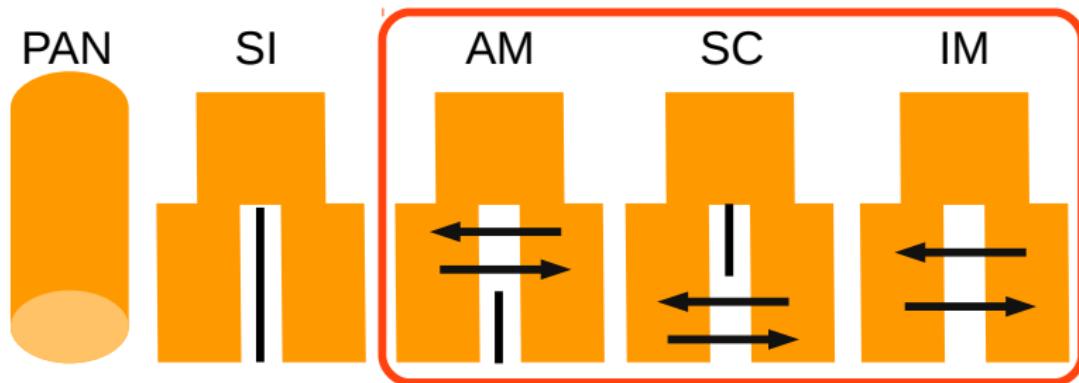
COMPARED SCENARIOS OF SPECIATION: ABC



COMPARED SCENARIOS OF SPECIATION: ABC

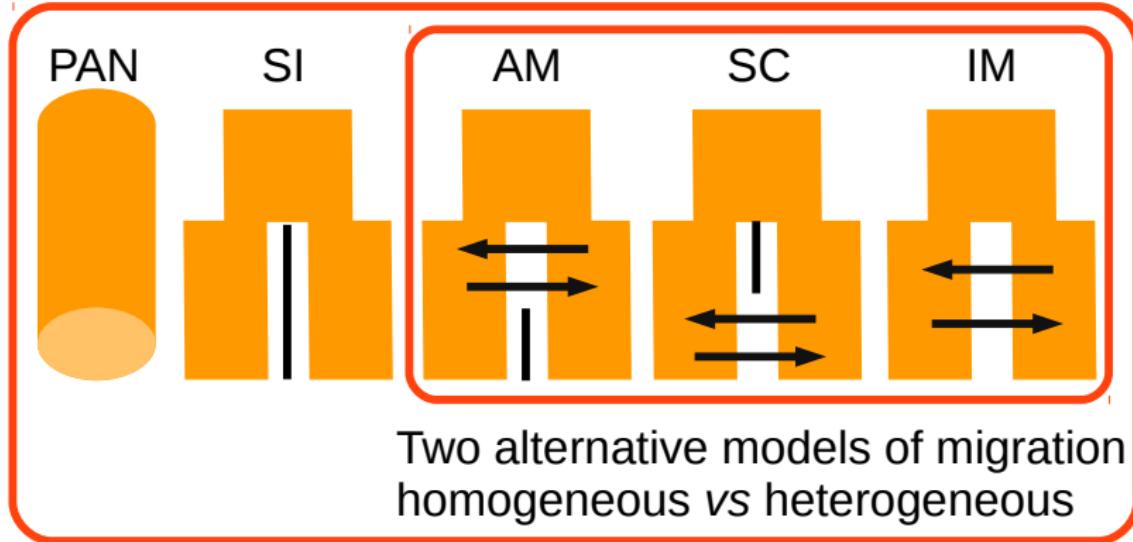
GENOMIC VARIATION IN F_{ST} CAN
BE THE RESULT OF GENOMIC
HETEROGENEITIES IN Ne AND/OR M

COMPARED SCENARIOS OF SPECIATION: ABC



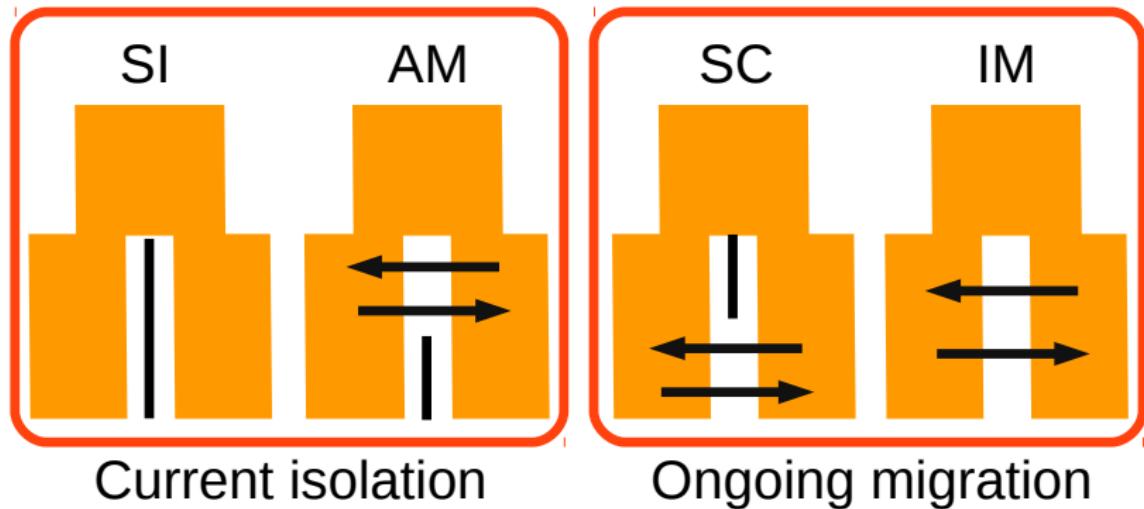
Two alternative models of migration
homogeneous vs heterogeneous

COMPARED SCENARIOS OF SPECIATION: ABC

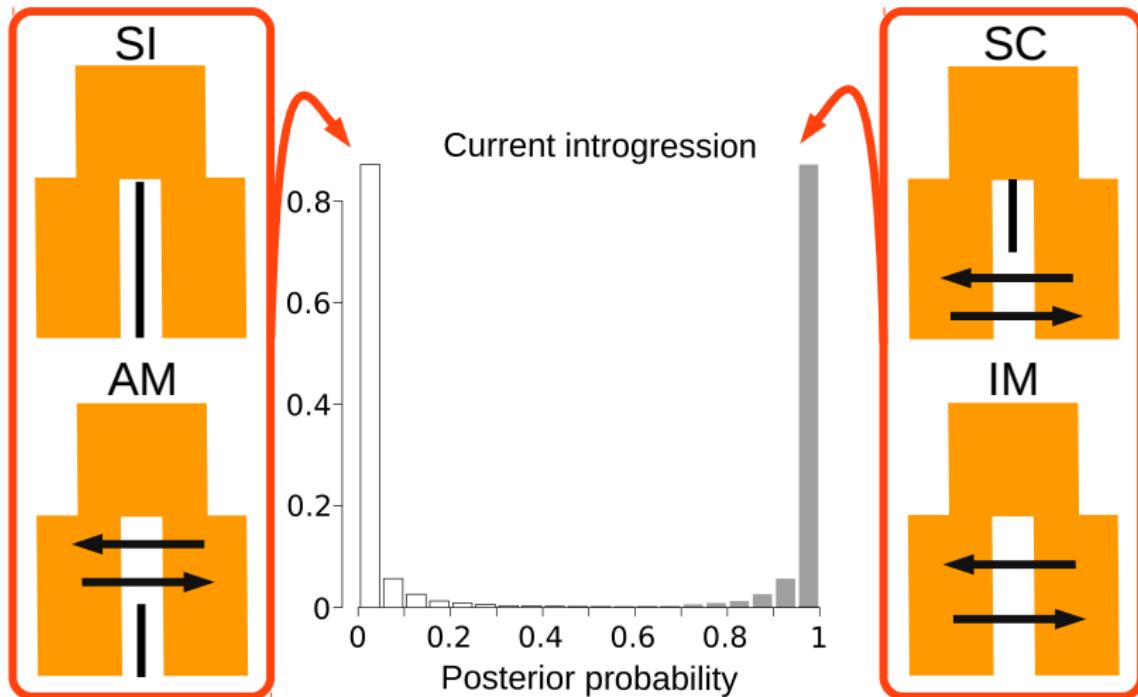


Two alternative models of drift
homogeneous vs heterogeneous

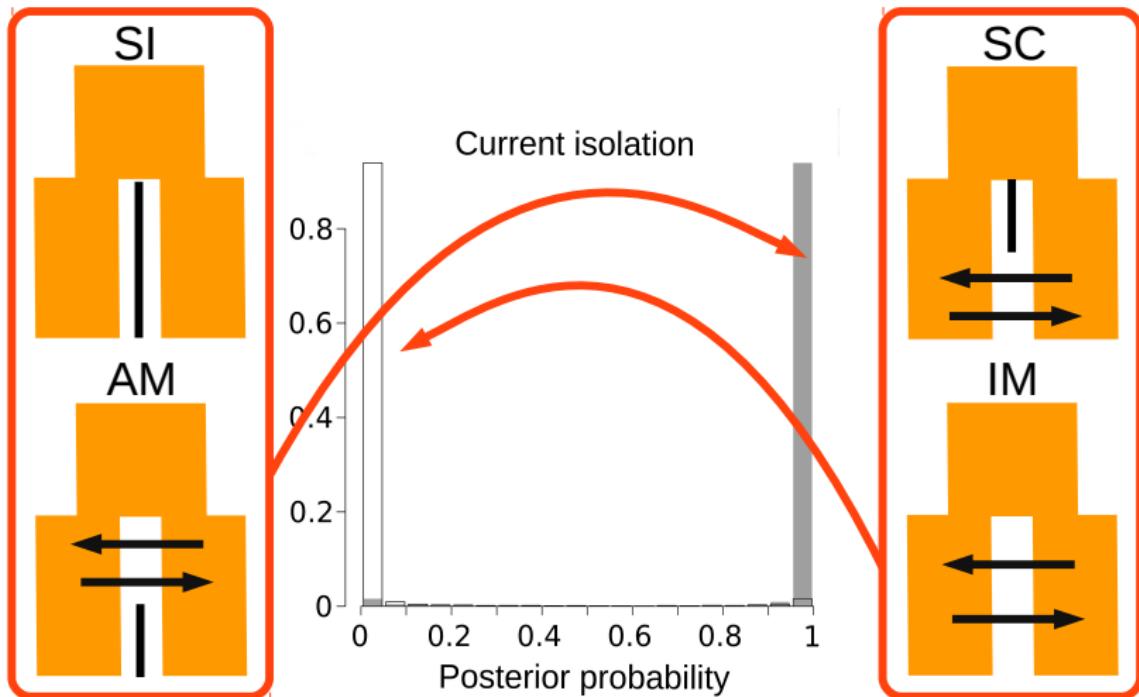
CAN WE DISTINGUISH ISOLATION FROM INTROGRESSION?



CAN WE DISTINGUISH ISOLATION FROM INTROGRESSION?

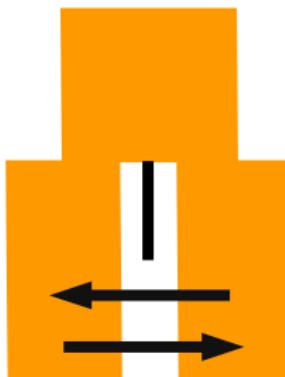


CAN WE DISTINGUISH ISOLATION FROM INTROGRESSION?

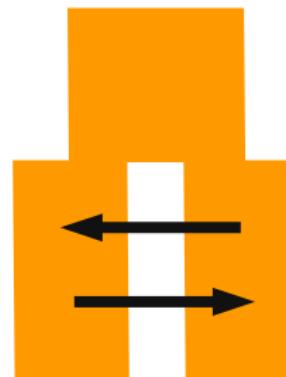


CAN WE DISTINGUISH IM FROM SC?

SC

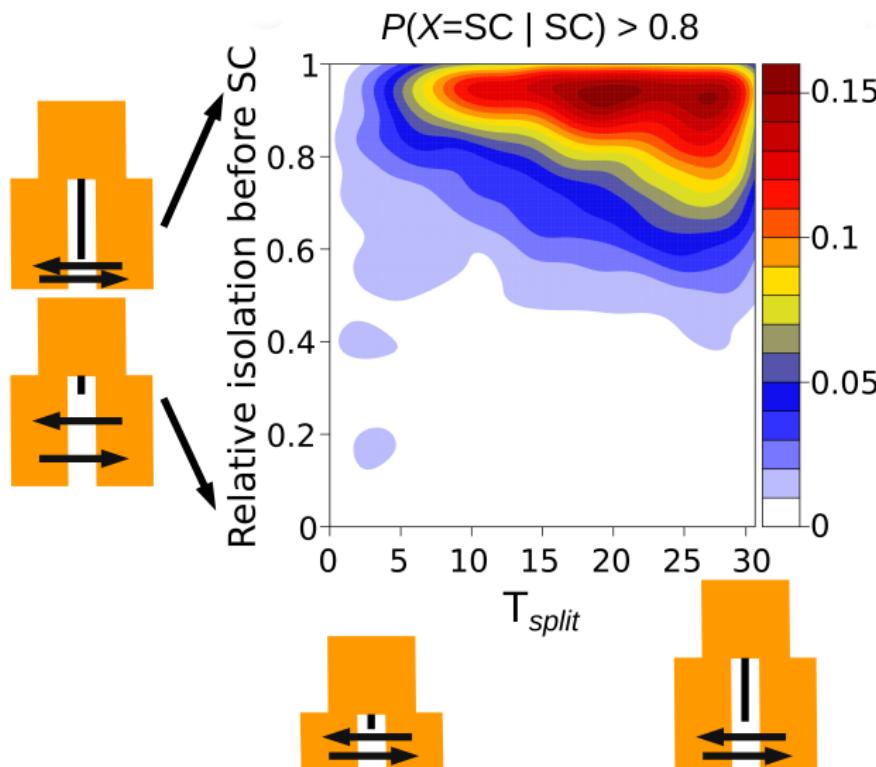


IM

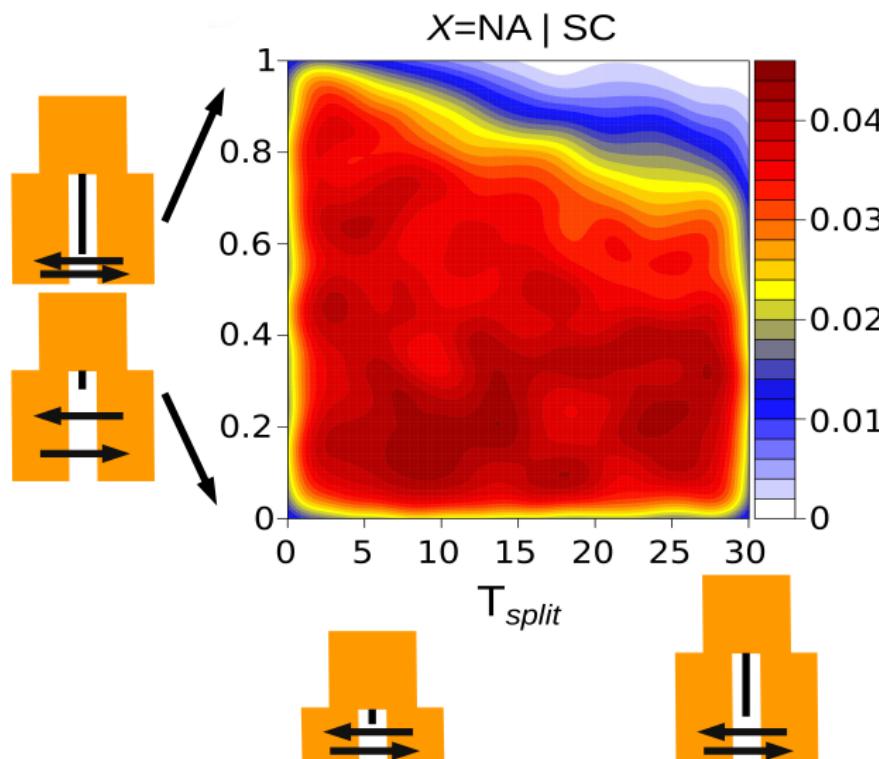


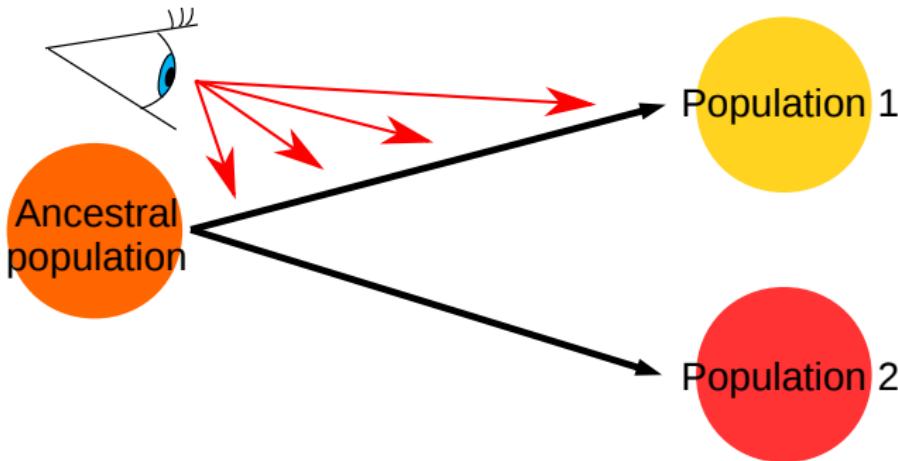
VS

CAN WE DISTINGUISH IM FROM SC?

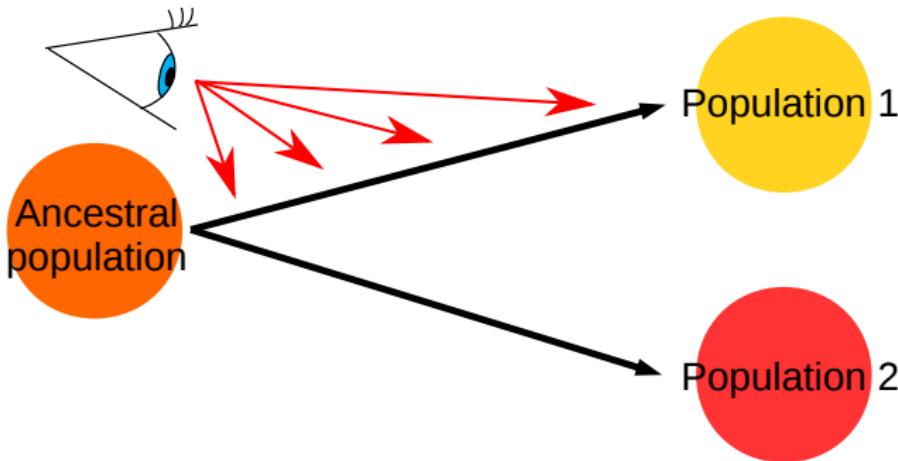


CAN WE DISTINGUISH IM FROM SC?





- ▶ We can reject/support ongoing gene flow
- ▶ Going further is more hazardous
- ▶ Testing ongoing gene flow along a continuum of divergence



- ▶ We can reject/support ongoing gene flow
- ▶ Going further is more hazardous
- ▶ Testing ongoing gene flow along a continuum of divergence

STUDIED DATASETS

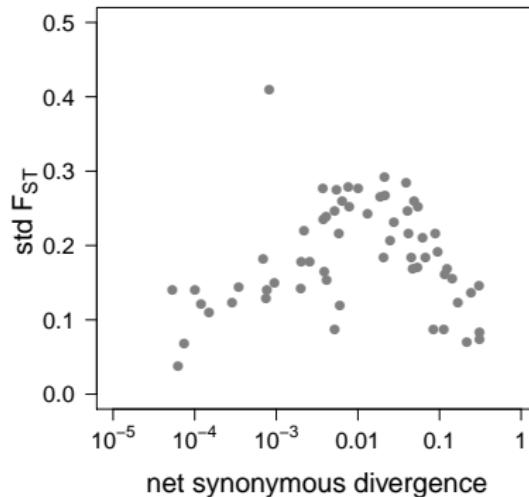
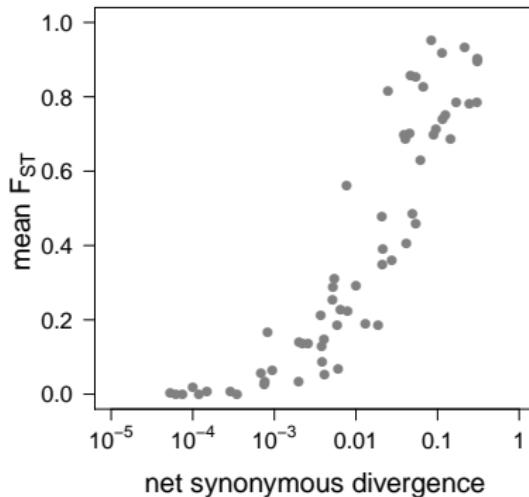
Published datasets
10 pairs of species



Obtained datasets
51 pairs of species

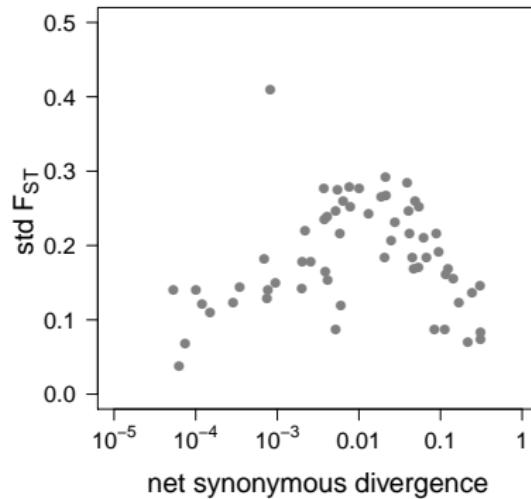
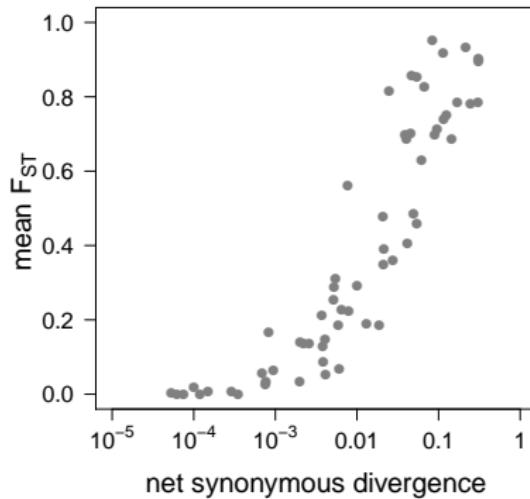


EXPLORED RANGE OF DIVERGENCE



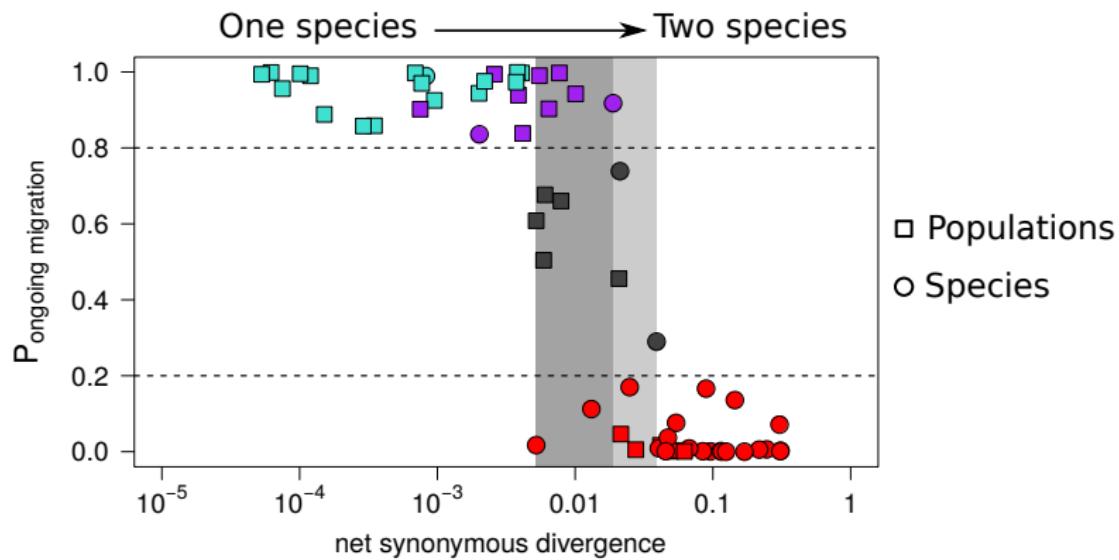
- Range of d : $5.10^{-5} - 0.31$
Range of F_{ST} : $0 - 0.95$
- Across-loci variance in F_{ST} was minimal for low and high values of d

EXPLORED RANGE OF DIVERGENCE



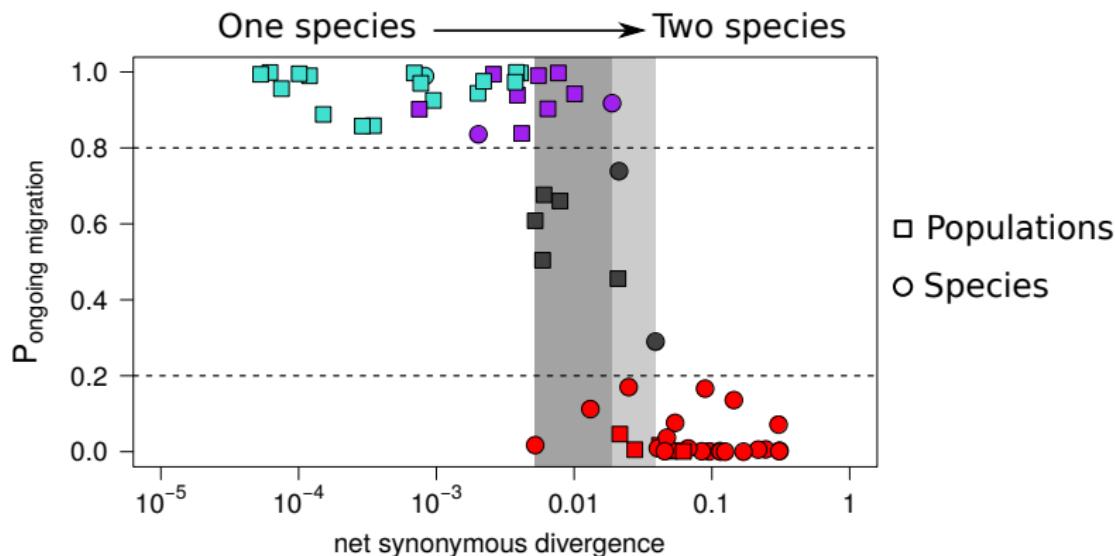
- Range of d : $5.10^{-5} - 0.31$
Range of F_{ST} : $0 - 0.95$
- Across-loci variance in F_{ST} was minimal for low and high values of d

RELATION BETWEEN d AND $P_{\text{introgression}}$



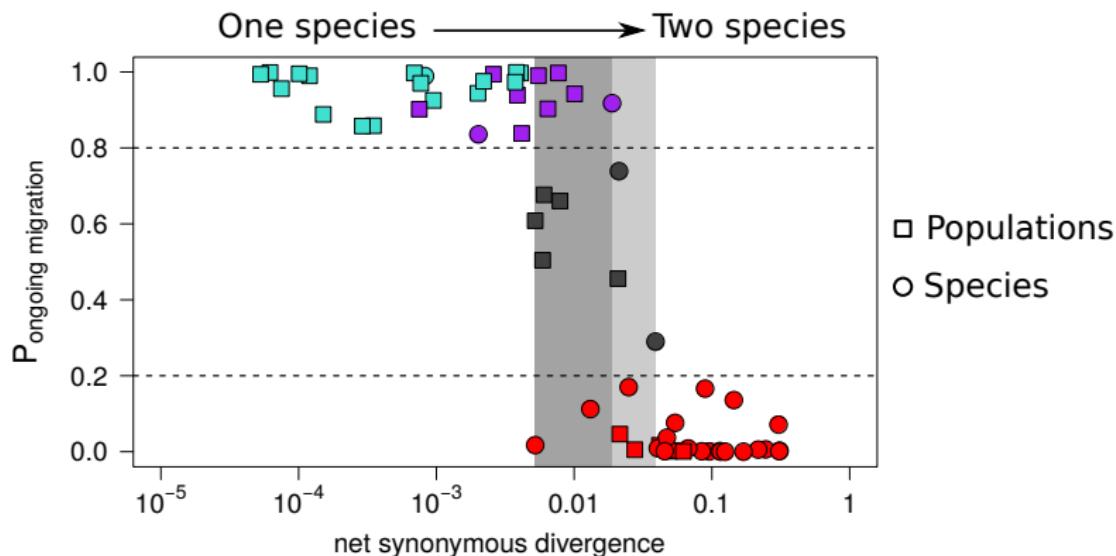
- ▶ Statistical grey zone of speciation for d in $[0.5\% - 4\%]$
- ▶ Conceptual grey zone of speciation for d in $[0.5\% - 2\%]$

RELATION BETWEEN d AND $P_{\text{introgression}}$

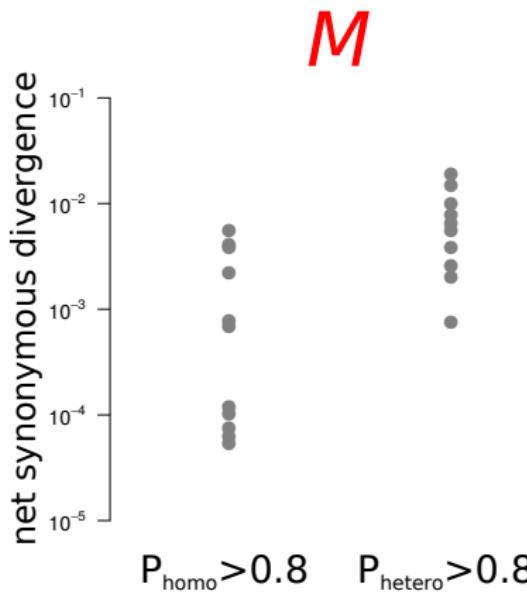


- ▶ Statistical grey zone of speciation for d in $[0.5\% - 4\%]$
- ▶ Conceptual grey zone of speciation for d in $[0.5\% - 2\%]$

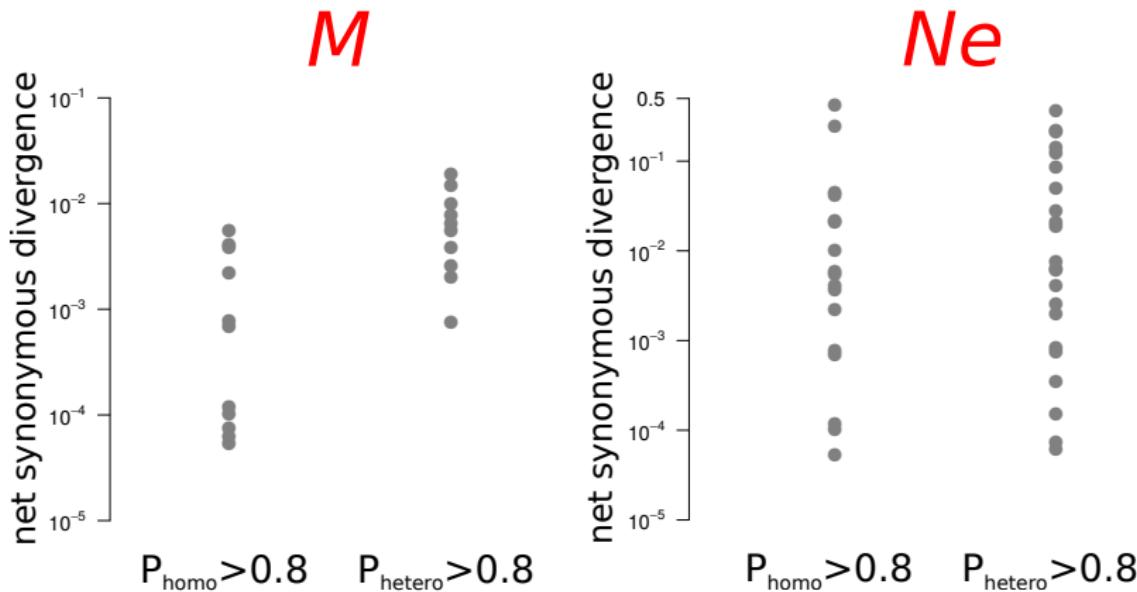
RELATION BETWEEN d AND $P_{\text{introgression}}$



- ▶ Statistical grey zone of speciation for d in $[0.5\% - 4\%]$
- ▶ Conceptual grey zone of speciation for d in $[0.5\% - 2\%]$

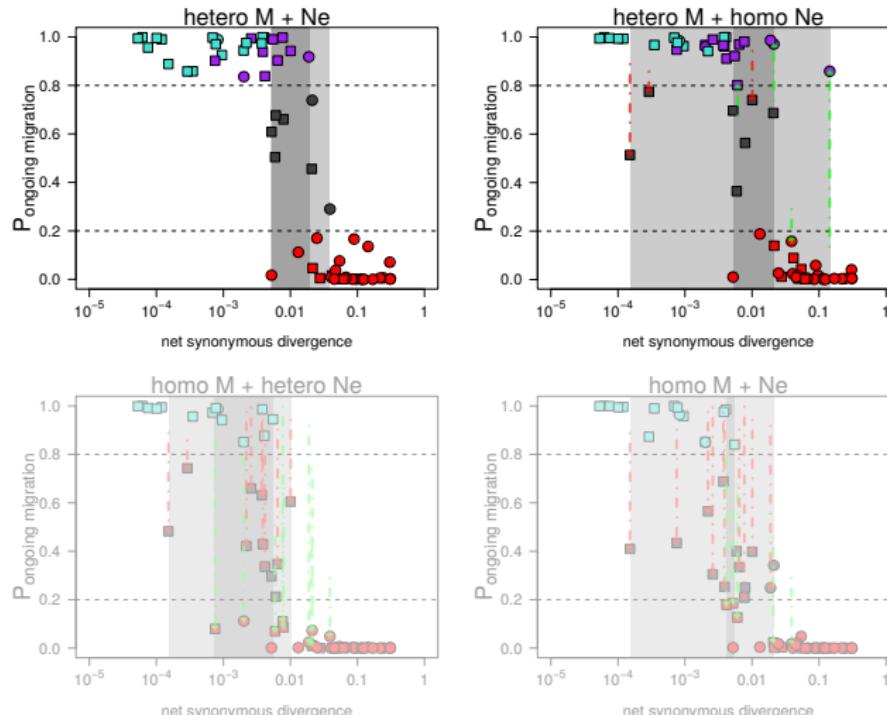
RELATION BETWEEN d AND GENOMIC HETEROGENEITIES

- ▶ Heterogeneity in M increases with d
 - ▶ First effects of barriers detected for $d = 0.00075$
- ▶ No effect of d on heterogeneity in Ne

RELATION BETWEEN d AND GENOMIC HETEROGENEITIES

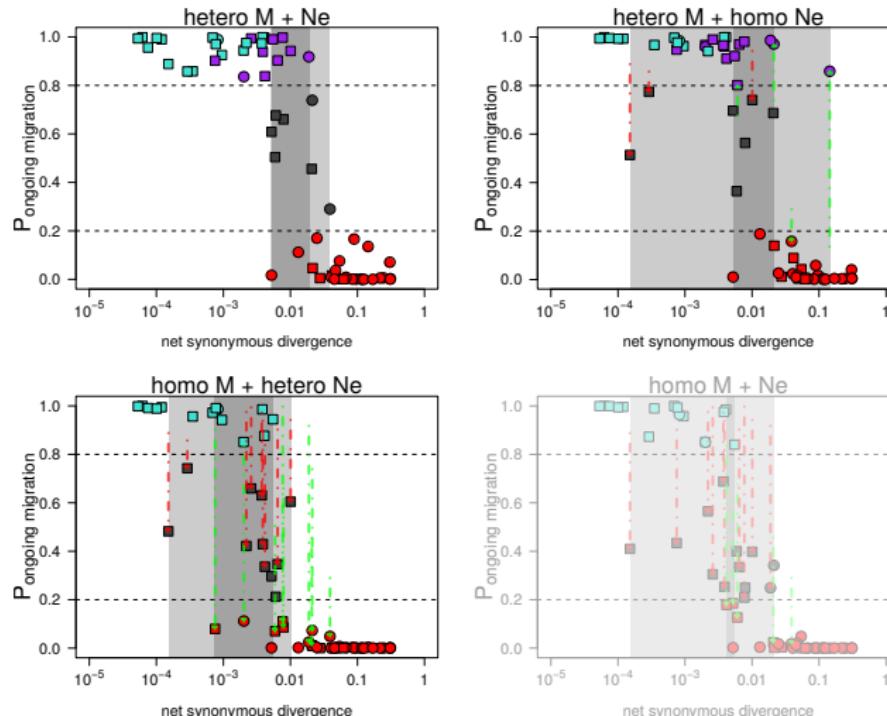
- ▶ Heterogeneity in M increases with d
 - ▶ First effects of barriers detected for $d = 0.00075$
- ▶ No effect of d on heterogeneity in Ne

EFFECTS OF MODEL HYPOTHESIS ON INFERENCES



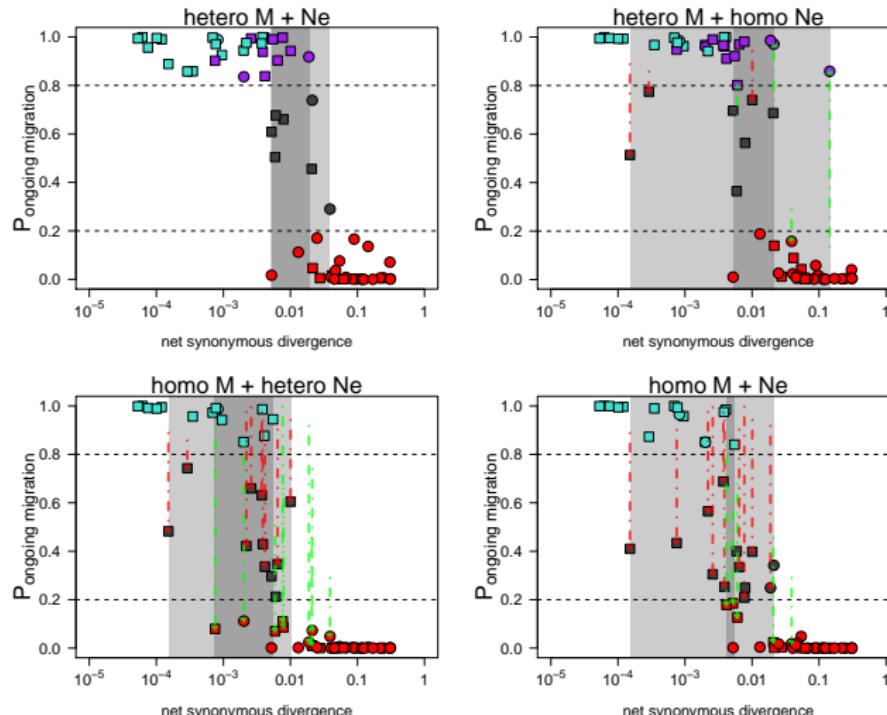
- ▶ Allowing M_{hetero} alone: false positives for introgression
- ▶ Allowing N_{hetero} alone: semi-isolated species become fully isolated

EFFECTS OF MODEL HYPOTHESIS ON INFERENCES



- ▶ Allowing M_{hetero} alone: false positives for introgression
- ▶ Allowing N_{hetero} alone: semi-isolated species become fully isolated

EFFECTS OF MODEL HYPOTHESIS ON INFERENCES



- ▶ Allowing M_{hetero} alone: false positives for introgression
- ▶ Allowing N_{hetero} alone: semi-isolated species become fully isolated

TAKE HOME MESSAGES

- ▶ Allowing for both N_{hetero} and M_{hetero} is important when reconstructing the demographic history
 - ▶ Neglecting N_{hetero} → over-estimating gene flow
 - ▶ Neglecting M_{hetero} → under-estimating gene flow
- ▶ Large statistical grey zone makes inferences complicated
- ▶ First barriers are detected for $d \approx 0.075\%$
- ▶ No gene flow detected for $d > 2\%$

TAKE HOME MESSAGES

- ▶ Allowing for both N_{hetero} and M_{hetero} is important when reconstructing the demographic history
 - ▶ Neglecting N_{hetero} → over-estimating gene flow
 - ▶ Neglecting M_{hetero} → under-estimating gene flow
- ▶ Large statistical grey zone makes inferences complicated
- ▶ First barriers are detected for $d \approx 0.075\%$
- ▶ No gene flow detected for $d > 2\%$

TAKE HOME MESSAGES

- ▶ Allowing for both N_{hetero} and M_{hetero} is important when reconstructing the demographic history
 - ▶ Neglecting N_{hetero} → over-estimating gene flow
 - ▶ Neglecting M_{hetero} → under-estimating gene flow
- ▶ Large statistical grey zone makes inferences complicated
- ▶ First barriers are detected for $d \approx 0.075\%$
- ▶ No gene flow detected for $d > 2\%$

TAKE HOME MESSAGES

- ▶ Allowing for both N_{hetero} and M_{hetero} is important when reconstructing the demographic history
 - ▶ Neglecting N_{hetero} → over-estimating gene flow
 - ▶ Neglecting M_{hetero} → under-estimating gene flow
- ▶ Large statistical grey zone makes inferences complicated
- ▶ First barriers are detected for $d \approx 0.075\%$
- ▶ No gene flow detected for $d > 2\%$

TAKE HOME MESSAGES

- ▶ Allowing for both N_{hetero} and M_{hetero} is important when reconstructing the demographic history
 - ▶ Neglecting N_{hetero} → over-estimating gene flow
 - ▶ Neglecting M_{hetero} → under-estimating gene flow
- ▶ Large statistical grey zone makes inferences complicated
- ▶ First barriers are detected for $d \approx 0.075\%$
- ▶ No gene flow detected for $d > 2\%$