

Polygenic barriers to gene flow

The role of dominance, haploid selection and heterogeneous genetic architectures

Arthur Zwaenepoel

19 October 2023

Local adaptation, gene flow & speciation

1. How can a population remain adapted to local environmental conditions in the face of maladaptive gene flow (**local adaptation**)?



Local adaptation, gene flow & speciation

1. How can a population remain adapted to local environmental conditions in the face of maladaptive gene flow (**local adaptation**)?

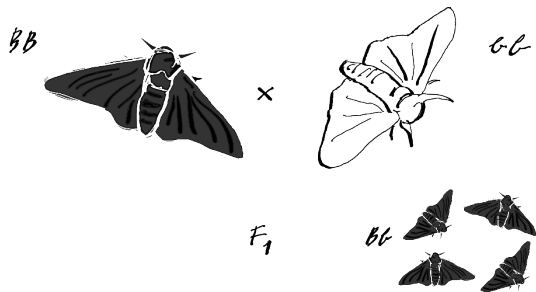


2. How does local adaptation promote **speciation**? (To what extent can populations which are under divergent selection pressures become **reproductively isolated**?)

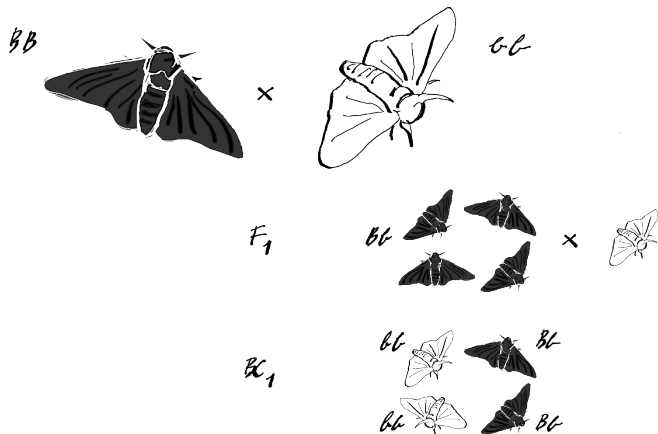
Barriers to gene flow



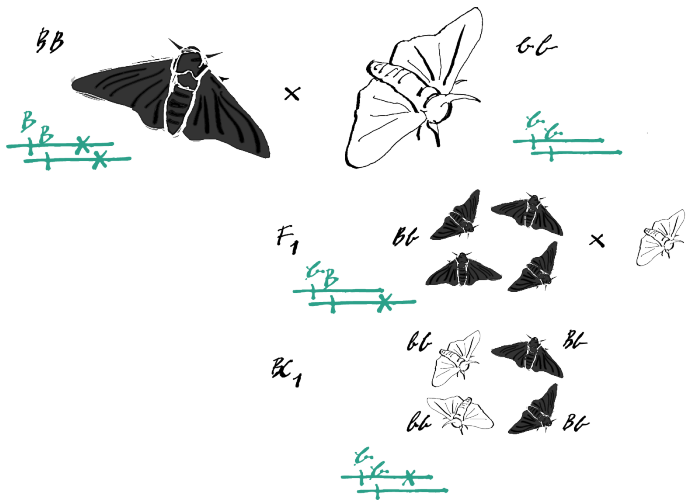
Barriers to gene flow



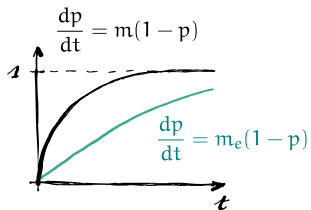
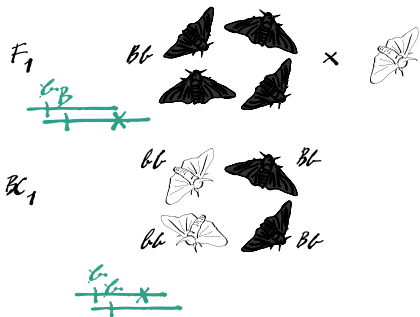
Barriers to gene flow



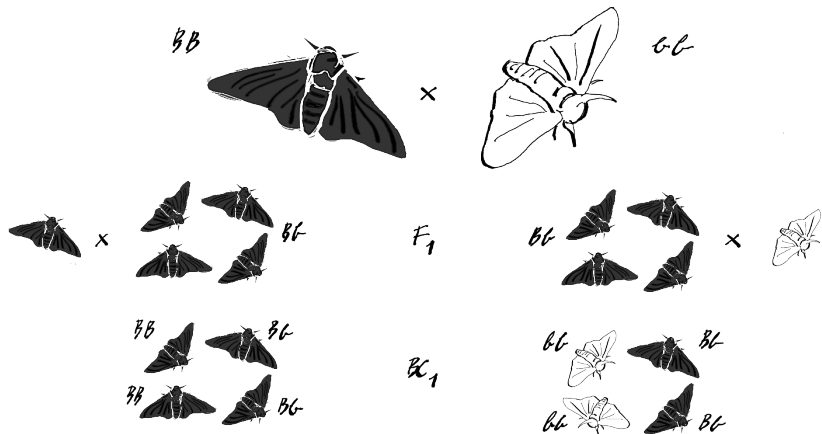
Barriers to gene flow



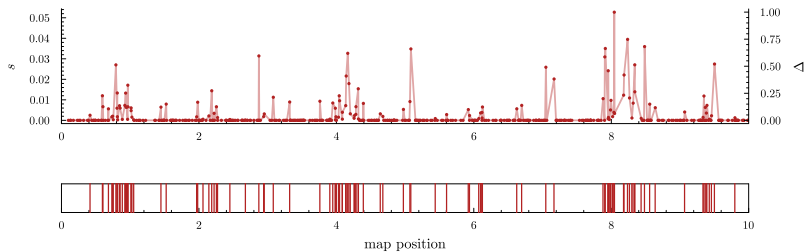
Barriers to gene flow



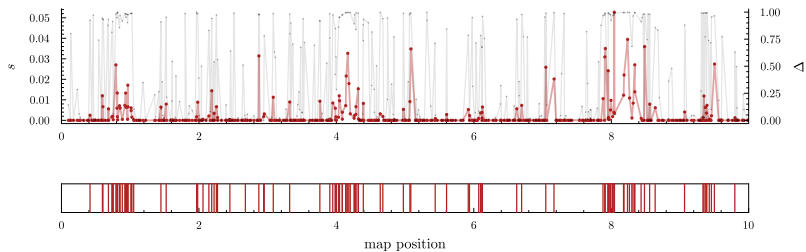
Barriers to gene flow



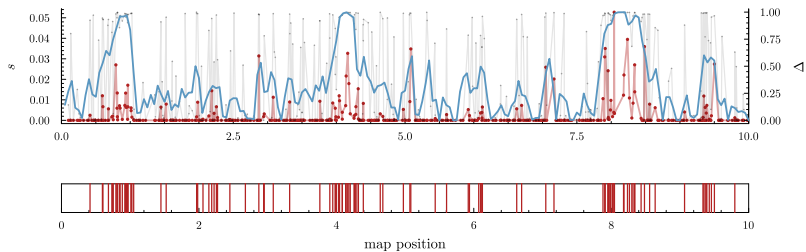
Polygenic barriers to gene flow



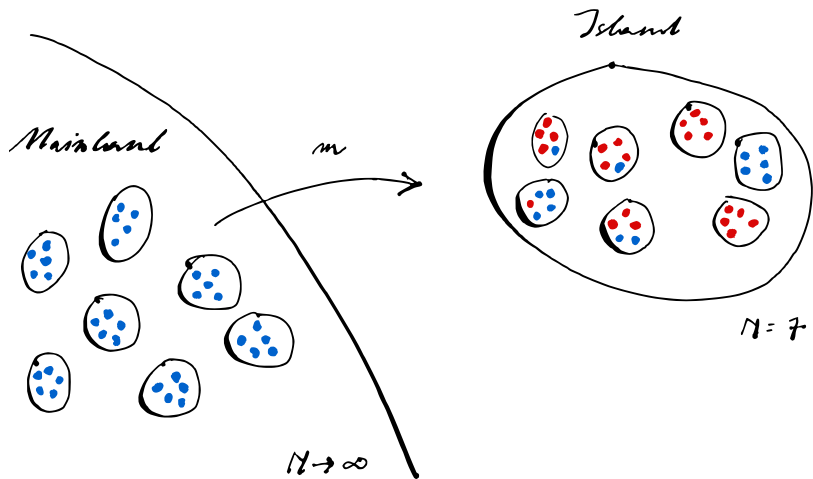
Polygenic barriers to gene flow



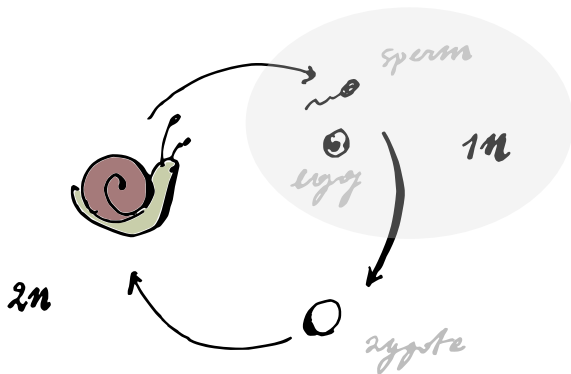
Polygenic barriers to gene flow



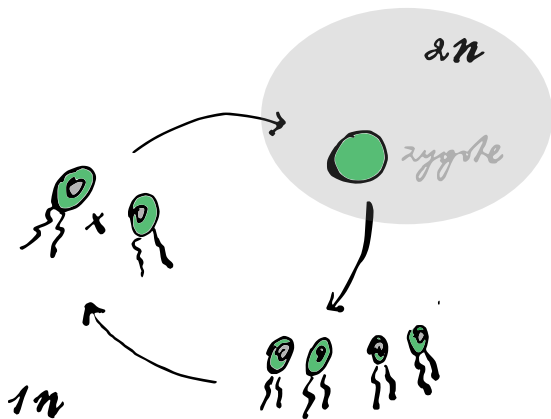
The model



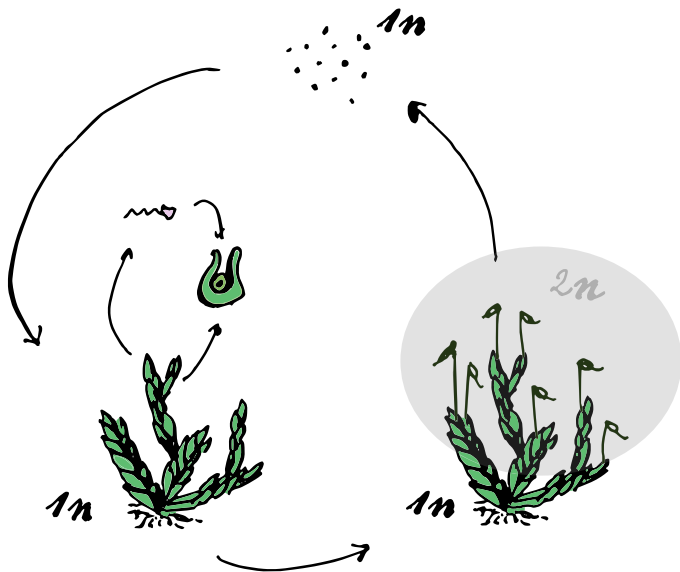
The life cycle – diplontic



The life cycle – haplontic



The life cycle – haplodiplontic



- ▶ **Island population** with N haploid individuals, generating Nk diploid individuals (non-overlapping alternation of generations)

- ▶ **Island population** with N haploid individuals, generating Nk diploid individuals (non-overlapping alternation of generations)
- ▶ **Mainland population** with a constant genetic composition

- ▶ **Island population** with N haploid individuals, generating Nk diploid individuals (non-overlapping alternation of generations)
- ▶ **Mainland population** with a constant genetic composition
- ▶ Each (haploid) generation, $M \sim \text{Poisson}(mN)$ island individuals are replaced by individuals from the mainland (**migration**)

- ▶ **Island population** with N haploid individuals, generating Nk diploid individuals (non-overlapping alternation of generations)
- ▶ **Mainland population** with a constant genetic composition
- ▶ Each (haploid) generation, $M \sim \text{Poisson}(mN)$ island individuals are replaced by individuals from the mainland (**migration**)
- ▶ Genetic architecture of L **unlinked** loci that affect **local adaptation**. Haploid relative fitness on the island

$$W_h = e^{\sum_{i=1}^L s_{i1} X_i}$$

Diploid relative fitness on the island

$$W_d = e^{\sum_{i=1}^L s_{i01} X_{i01} + s_{i11} X_{i11}}$$

- ▶ **Island population** with N haploid individuals, generating Nk diploid individuals (non-overlapping alternation of generations)
- ▶ **Mainland population** with a constant genetic composition
- ▶ Each (haploid) generation, $M \sim \text{Poisson}(mN)$ island individuals are replaced by individuals from the mainland (**migration**)
- ▶ Genetic architecture of L **unlinked** loci that affect **local adaptation**. Haploid relative fitness on the island

$$W_h = e^{\sum_{i=1}^L s_{i1} X_i}$$

Diploid relative fitness on the island

$$W_d = e^{\sum_{i=1}^L s_{i01} X_{i01} + s_{i11} X_{i11}}$$

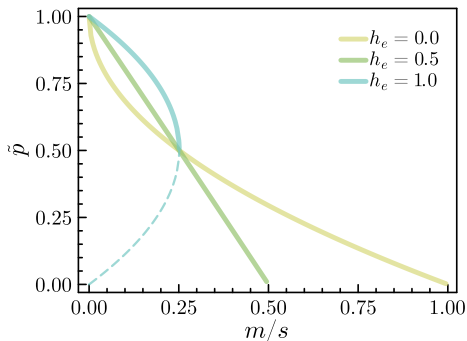
Goal 1: characterize island population at migration-selection equilibrium

Goal 2: how does local adaptation and reproductive isolation depend on the genetic architecture underlying divergent selection

Deterministic model at a single locus

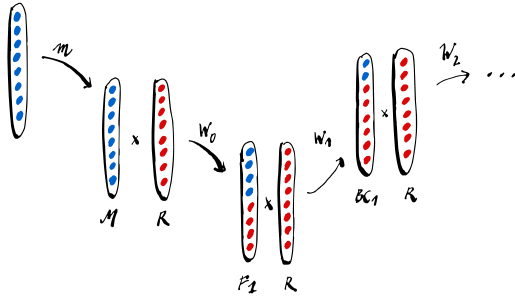
- ▶ p : the frequency of the locally beneficial allele on the island, mainland is fixed for the other allele
- ▶ $s_a = s_1 + s_{01} = -s_e h_e$ and $s_b = s_{11} - 2s_{01} = -s_e(1 - 2h_e)$

$$\frac{dp}{dt} = -mp - q(s_a + s_b p q)$$



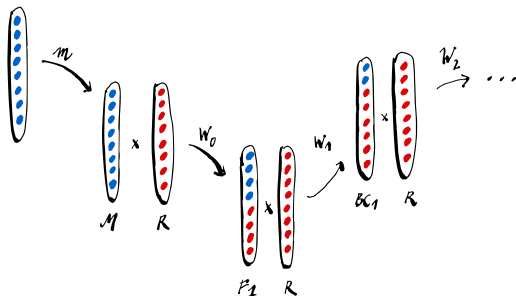
Polygenic migration & selection: effective migration rate

$m_e = gm$, where g = *gene flow factor*



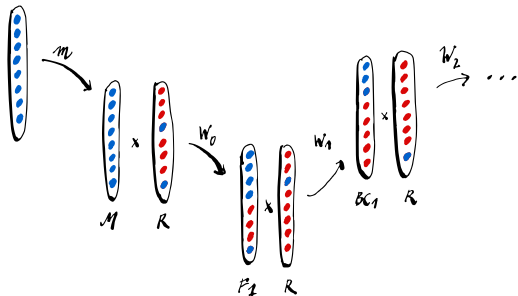
Polygenic migration & selection: effective migration rate

$m_e = gm$, where $g = \text{gene flow factor}$



Approximate g

1. L **unlinked** or **weakly-linked** loci under weak selection
2. Sufficiently **weak migration** ($O(m^2)$ is negligible)
3. Resident and migrant gene pools are in HWLE
4. Segregation variance and selection thereon is negligible



1. L **unlinked** or **weakly-linked** loci under weak selection
2. Sufficiently **weak migration** ($O(m^2)$ is negligible)
3. Resident and migrant gene pools are in HWLE
4. Segregation variance and selection thereon is negligible

$$\begin{aligned}
 g &= \frac{m_e}{m} = \mathbb{E} \left[W_h^{(0)} \prod_{k=0}^{\infty} W_d^{(k)} W_h^{(k)} \right] \\
 &\approx e^{L(s_a p + s_b p q)} e^{\frac{1}{2}(s_a p + s_b p q)} \dots \\
 &= e^{2L(s_a p + s_b p q)}
 \end{aligned}$$

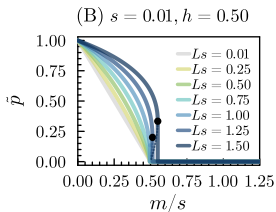
Deterministic multilocus model

System of L nonlinear ODEs, coupled through g

$$\begin{aligned}\frac{dp_i}{dt} &= -p_i q_i (s_{i,a} + s_{i,b} q_i) - m g_i [p_{-i}] p_i \\ &= -p_i q_i (s_{i,a} + s_{i,b} q_i) - m \exp \left(\sum_{j \neq i} s_{j,a} p_j + s_{j,b} p_j q_j \right) p_i\end{aligned}$$

Deterministic multilocus model

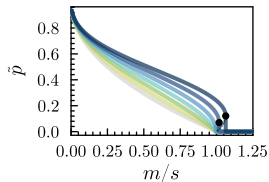
$$\begin{aligned}\frac{dp}{dt} &= -pq(s_a + s_b q) - mg[p]p \\ &= -pq(s_a + s_b q) - m \exp(2L(s_a p + s_b p q))p\end{aligned}$$



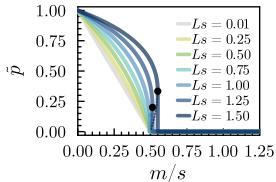
Deterministic multilocus model

$$\begin{aligned}\frac{dp}{dt} &= -pq(s_a + s_b q) - mg[p]p \\ &= -pq(s_a + s_b q) - m \exp(2L(s_a p + s_b pq))p\end{aligned}$$

(A) $s = 0.01, h = 0.00$

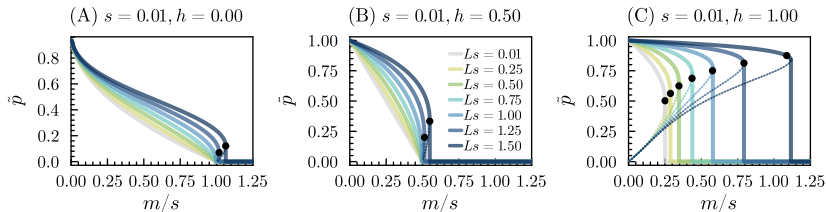


(B) $s = 0.01, h = 0.50$



Deterministic multilocus model

$$\begin{aligned}\frac{dp}{dt} &= -pq(s_a + s_b q) - mg[p]p \\ &= -pq(s_a + s_b q) - m \exp(2L(s_a p + s_b p q))p\end{aligned}$$



Accounting for genetic drift

Single locus: **diffusion approximation** for the equilibrium distribution of p on the island (Wright 1937)

$$\phi(p) \propto \underbrace{p^{2N\mu-1}}_{\text{mutation}} \times \underbrace{q^{2N\mu+2Nm-1}}_{\text{mutation + migration}} \times \underbrace{e^{N(2s_a q + s_b q^2)}}_{\text{selection}}$$

Substitute $N_e = \left(\frac{1}{N} + \frac{1}{2Nk}\right)^{-1}$ for N for biphasic life cycle

Accounting for genetic drift

Single locus: **diffusion approximation** for the equilibrium distribution of p on the island (Wright 1937)

$$\phi(p) \propto \underbrace{p^{2N\mu-1}}_{\text{mutation}} \times \underbrace{q^{2N\mu+2Nm-1}}_{\text{mutation + migration}} \times \underbrace{e^{N(2s_a q + s_b q^2)}}_{\text{selection}}$$

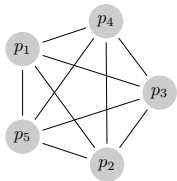
Substitute $N_e = \left(\frac{1}{N} + \frac{1}{2Nk}\right)^{-1}$ for N for biphasic life cycle

Multilocus heuristic: substitute $m_e = mg(p)$ for m

$$\phi_i(p_i|p_{-i}) \propto p_i^{2N_e\mu-1} q_i^{2N_e\mu+2Nm g_i[p_{-i}]} e^{N_e(2s_{i,a} q_i + s_{i,b} q_i^2)}$$

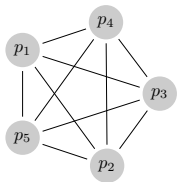
where:

$$g_i[p_{-i}] = \exp\left(2 \sum_{j \neq i}^L s_{a,j} p_j + s_{b,j} p_j q_j\right)$$



$$\phi(\mathbf{p}) \propto \prod_{i=1}^L \phi_i(\mathbf{p}_i | \mathbf{p}_{-i})$$

Defines a **Markov random field**, which yields a Gibbs distribution as joint distribution over $[0, 1]^L$



$$\phi(\mathbf{p}) \propto \prod_{i=1}^L \phi_i(\mathbf{p}_i | \mathbf{p}_{-i})$$

Defines a **Markov random field**, which yields a Gibbs distribution as joint distribution over $[0, 1]^L$

Or view it as a kind of **mean field**: i.e. solve self-consistently for the $\mathbb{E}[p_i]$ and $\mathbb{E}[p_i q_i]$ in:

$$\mathbb{E}[p_i] = \int p_i \phi \left(p_i, g_i \left[\mathbb{E}[p_{-i}], \mathbb{E}[p q_{-i}] \right] \right) dp_i$$

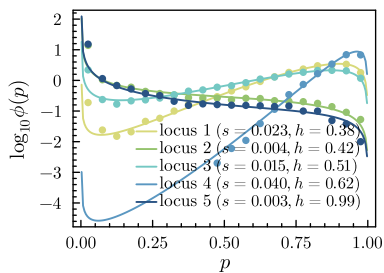
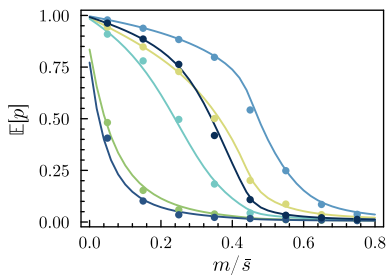
$$\mathbb{E}[p_i q_i] = \int p_i q_i \phi \left(p_i, g_i \left[\mathbb{E}[p_{-i}], \mathbb{E}[p q_{-i}] \right] \right) dp_i$$

where:

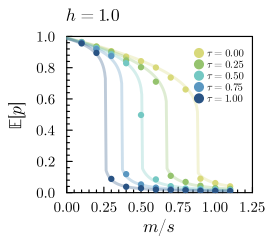
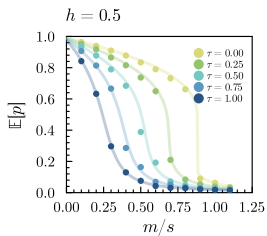
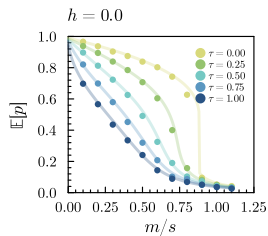
$$g_i \left[\mathbb{E}[p_{-i}], \mathbb{E}[p q_{-i}] \right] = \exp \left(2 \sum_{j \neq i}^L s_{a,j} \mathbb{E}[p_j] + s_{b,j} \mathbb{E}[p_j q_j] \right)$$

Evaluating the approximation

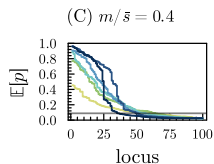
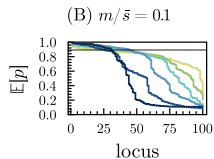
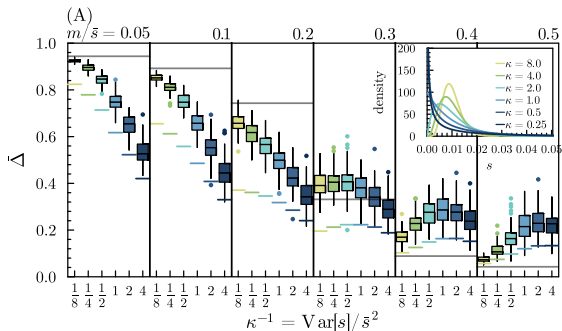
$L = 50, L\bar{s} = 1.0, N_e\bar{s} = 10.0$



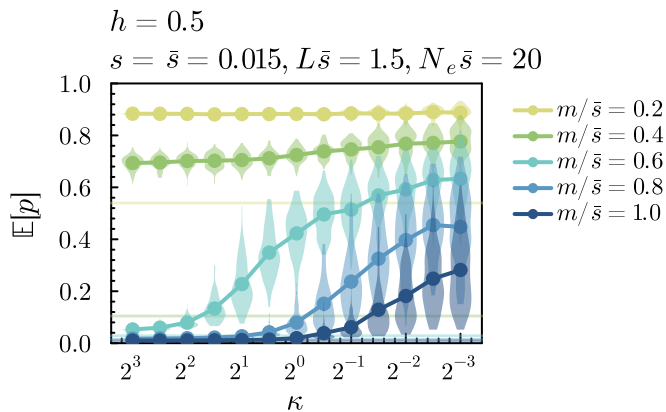
Evaluating the approximation: life cycle



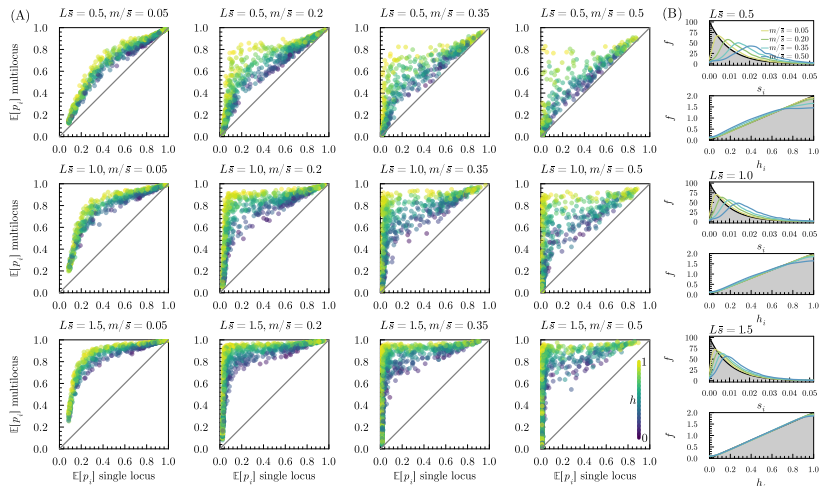
Heterogeneous architectures and barrier strength



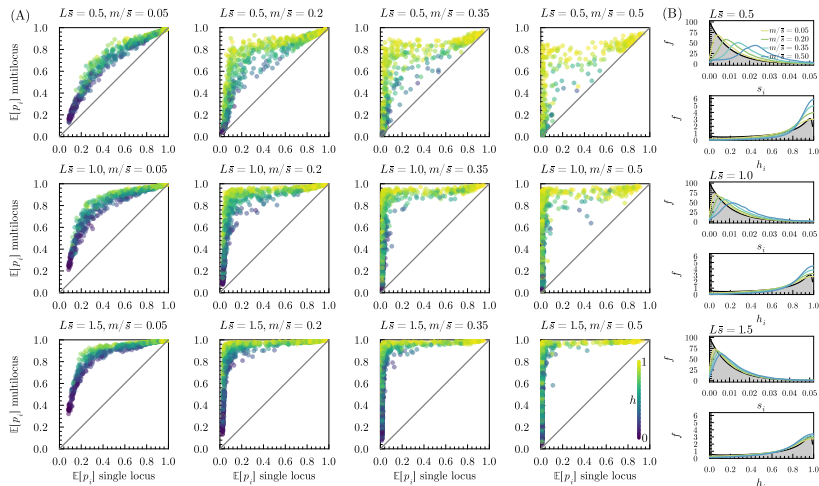
Effect on a focal locus



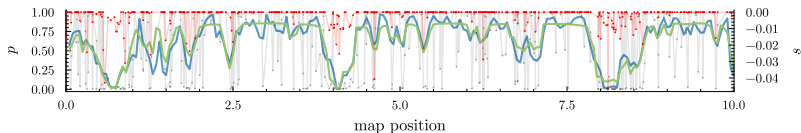
Variation in the barrier effect



The importance of the distribution of fitness effects



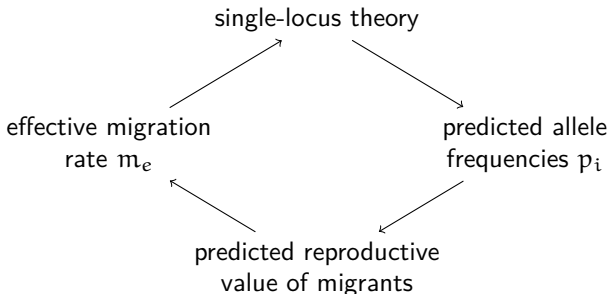
Accounting for linkage



$$g_i \left[\mathbb{E}[p_{-i}], \mathbb{E}[pq_{-i}] \right] = \exp \left(\sum_{j \neq i}^L \frac{s_{a,j} \mathbb{E}[p_j] + s_{b,j} \mathbb{E}[p_j q_j]}{r_{ij}} \right)$$

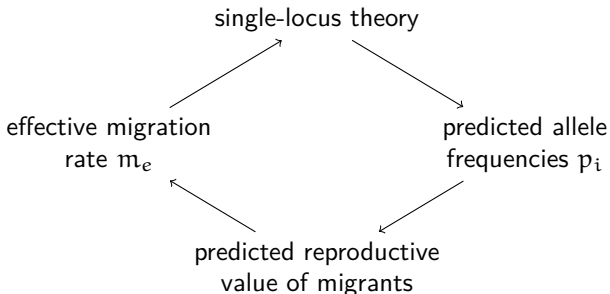
Conclusion

- ▶ Single-locus theory + m_e, s_e, h_e, N_e yields **accurate multilocus predictions for *equilibrium* frequencies** when there is no/weak linkage, for **heterogeneous genetic architectures** and a general **haplodiplontic life cycle**.



Conclusion

- ▶ Single-locus theory + m_e, s_e, h_e, N_e yields **accurate multilocus predictions for *equilibrium* frequencies** when there is no/weak linkage, for **heterogeneous genetic architectures** and a general **haplodiplontic life cycle**.



- ▶ Efficient numerics to study how the **genetic architecture** underlying locally adaptive traits determines the extent of **observable adaptive differentiation at migration-selection equilibrium**

Preprint

Zwaenepoel, A., Sachdeva, H., & Fraïsse, C. (2023).

Polygenic barriers to gene flow: the role of dominance, haploid selection and heterogeneous genetic architectures.

bioRxiv, 2023-09. <https://doi.org/10.1101/2023.09.24.559235>

Thanks to

Christelle Fraïsse, Himani Sachdeva (University of Vienna)



European Research Council
Established by the European Commission

