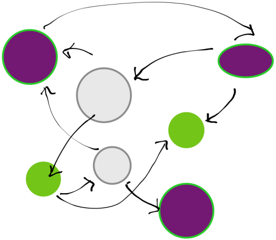
Polygenic local adaptation in subdivided populations: effects of LD and drift

Himani Sachdeva

Department of Mathematics, University of Vienna



Model

- infinite island model: subdivision but no space.
- fixed population size (**N** haploids) on each island: "soft selection"
- fraction **m** replaced by migrants drawn *uniformly* from metapopulation.
- Two habitats: fractions $\boldsymbol{\rho}$, $\mathbf{1}-\boldsymbol{\rho}$ of islands in rare/common habitat.
  $\rho \to 0$: mainland-island model.
- Habitat-dependent fitness; influenced by **L** *unlinked* bi-allelic loci.
- Alternative alleles favoured at each locus in the two habitats:
  000000000000000 111111111111111
  Habitat 1 Habitat 2
  $W(X) = \exp[-\mathbf{s}X]$ where $X \to \#$ locally deleterious alleles.
- Neglect mutation except for mainland-island model.

$\boldsymbol{\rho}$ → fraction of islands with rare habitat ($\rho < 1/2$)

L → number of divergently selected loci

s, **m**, $\mathbf{1/N}$: per generation change due to selection, migration, drift.

r: rate of recombination between selected loci; $r = 1/2$ for unlinked loci

$$\Delta D = -rD$$

$\boldsymbol{\rho} \rightarrow$ fraction of islands with rare habitat ($\rho < 1/2$)

L $\rightarrow$ number of divergently selected loci

s, **m**, $\mathbf{1/N}$: per generation change due to selection, migration, drift.

r: rate of recombination between selected loci; $r = 1/2$ for unlinked loci

$$\Delta D = -rD$$

Theory challenging because:

- multiple loci under divergent selection $\implies$ **LD**
- individual sub-populations often small $\implies$ genetic **drift**

Typically neglect LD (“linkage equilibrium”) or drift (deterministic analyses).

LD and drift in subdivided populations

- Reproductive isolation and speciation require "coupling" (associations) between different genetic incompatibilities.

  Kulmuni et al, 2020; Butlin & Smadja, 2018; Barton & De Cara 2009.
- LD between large numbers of incompatibilities $\implies$ "Tipping points"
  Nosil et al, 2017

LD and drift in subdivided populations

- Reproductive isolation and speciation require "coupling" (associations) between different genetic incompatibilities.

  Kulmuni et al, 2020; Butlin & Smadja, 2018; Barton & De Cara 2009.
- LD between large numbers of incompatibilities $\Longrightarrow$ "Tipping points"
  Nosil et al, 2017
- Many small local populations + polygenic architectures $\Longrightarrow$
  - Direct selection less effective ($Ns \sim 1$).
  - Negative LD due to Hill-Robertson interference? Collective elimination of sets of alleles less effective?

Multi-locus local adaptation in subdivided "patchy" populations?

Outline

1. Approximations
2. Mainland-island model
3. Infinite island model with two habitats.

Ref: H Sachdeva, bioRxiv 2021.11.05.467433, 2021.

Linkage equilibrium analysis (neglecting LD)

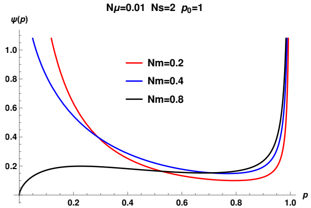
- recombination faster than all processes: $1/N \sim m \sim s \ll Ls \ll r \sim 1$

  $\implies$ independently evolving loci ("LE").

- *Single locus* allele frequency distribution under migration-mutation-selection-drift balance:

$$\psi_{island}[p] \propto p^{2Nm+2N\mu-1}(1-p)^{2N\mu-1}e^{-2\,N\,s\,p}$$

(Wright, 1937)



Deterministic analysis (neglecting drift)

- drift weaker than all other processes: $1/N \ll m \sim s \ll Ls \sim r \sim 1$ $\implies$ deterministic coupled equations for genotype frequencies $\{P_1, P_2, \ldots P_{L-1}, P_L\}$ under migration-selection-recombination balance.

- deleterious alleles rare $\implies$ *linear* equations

  avg. allele freq $$p_{det} = m \frac{e^{-sL}}{(2-e^{-s})^{L-1}} \sum_{k=0}^{\infty} e^{-s\,k} \left(1 + \frac{e^{-sk}}{2^k} - \frac{e^{-s(k+1)}}{2^k}\right)^{L-1}$$ $$\approx (m/s)\ e^{-2Ls} \qquad \text{for large } L, \text{ small } s$$

- LD between deleterious alleles $\implies$ reduced 'effective' rate of migration $m_e \approx m\,e^{-2Ls}$ for individual alleles.

- Analogous definition of m_e for neutral alleles (Barton & Bengtsson 1986).

Approximating combined effects of LD and drift.

What if $\frac{1}{N} \sim m \sim s \ll Ls \sim r \sim 1$?

Approximating combined effects of LD and drift.

What if $\frac{1}{N} \sim m \sim s \ll Ls \sim r \sim 1$?

Drift affects *single* deleterious alleles but not genotypes with *many* alleles.

Approximating combined effects of LD and drift.

What if $\frac{1}{N} \sim m \sim s \ll Ls \sim r \sim 1$?

Drift affects *single* deleterious alleles but not genotypes with *many* alleles.

Allele frequency distribution under mainland-island model:

$$\psi_{island}[p] \;\approx\; p^{2N m_e + 2N\mu - 1}(1-p)^{2N\mu-1} e^{-2Ns\,p}$$

Diffusion approximation with effective migration rates

Approximating combined effects of LD and drift.

What if $\frac{1}{N} \sim m \sim s \ll Ls \sim r \sim 1$?

Drift affects *single* deleterious alleles but not genotypes with *many* alleles.

Allele frequency distribution under mainland-island model:

$$\psi_{island}[p] \;\approx\; p^{2Nm_e + 2N\mu - 1}(1-p)^{2N\mu-1} e^{-2Ns\,p}$$

Diffusion approximation with effective migration rates

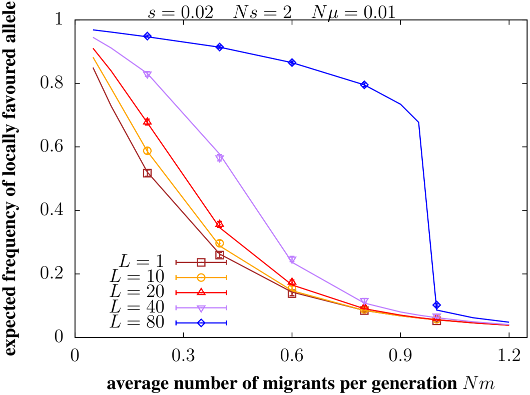
$$m_e \;\approx\; m\, e^{-2sX}$$

$X \to$ number of nearly fixed selective differences between populations.

$\approx \; L(1 - \mathbf{E}[p])$

Need to find $\mathbf{E}[p]$ numerically.

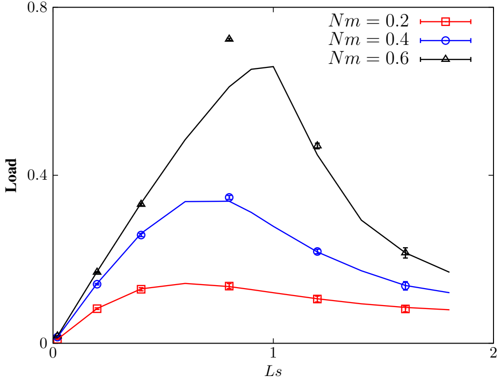
Frequency of locally favoured alleles increases with *Ls*



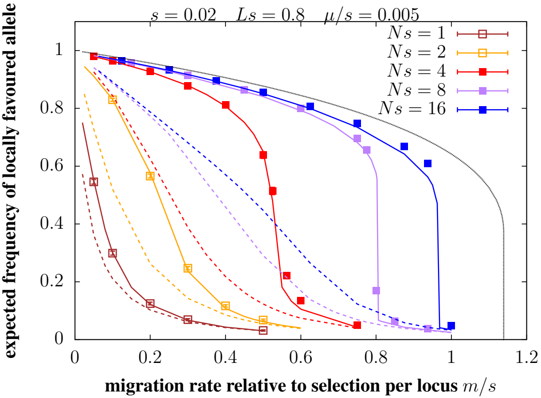
symbols → simulations
lines → theory

higher *Ls* $\Longrightarrow$ higher frequencies of locally adaptive alleles AND sharper thresholds (“tipping points”) for loss of adaptation.

But maladaptation load maximum for intermediate *Ls*.



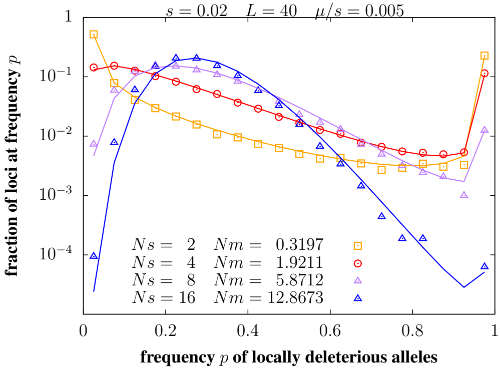
LD, drift and migration thresholds.



symbols → simulations
solid lines → theory (LD+drift)
dashed lines → theory (no LD)
dotted line → theory (no drift)

- Drift can strongly reduce migration thresholds for local adaptation.
- LD more significant in larger populations.

Predicting allele frequency distributions



- Predictions fairly accurate even when heterozygosity high.

1. Approximations

2. Mainland-island model

3. Infinite island model with two habitats.

When is the rare habitat locally adapted?

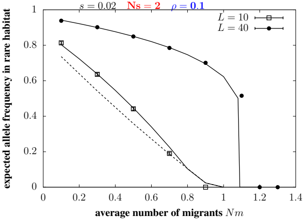
$m_e \approx m\, e^{-2sL(\mathbb{E}[p_1]-\mathbb{E}[p_2])}$ for migration between different habitats

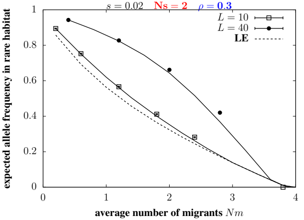
$m_e \approx m$ for migration within same habitat

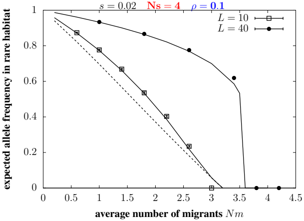
When is the rare habitat locally adapted?

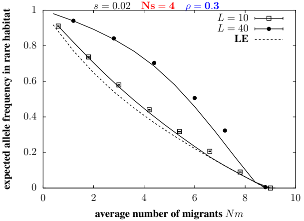
$m_e \approx m\, e^{-2sL(\mathbb{E}[p_1]-\mathbb{E}[p_2])}$ for migration between different habitats

$m_e \approx m$ for migration within same habitat

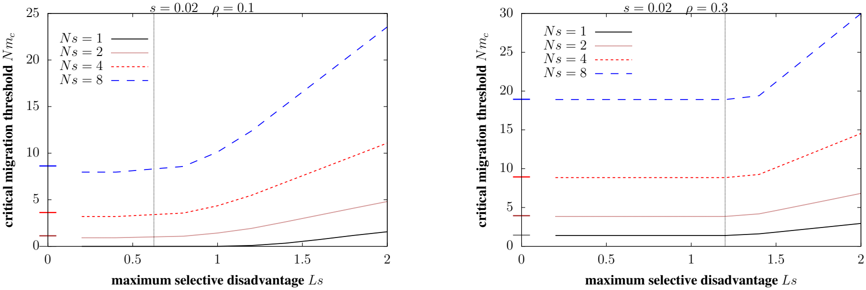








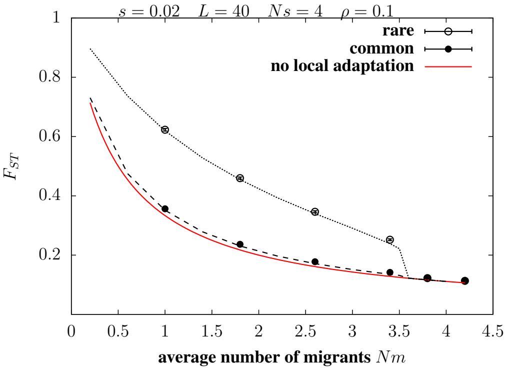
Sharper vs. shifted thresholds for loss of adaptation



- m_c only increases above a threshold $(Ls)_*$
- More significant effect of LD on m_c when one habitat much rarer.

Barrier to neutral gene flow: F_{ST} in rare vs. common habitat

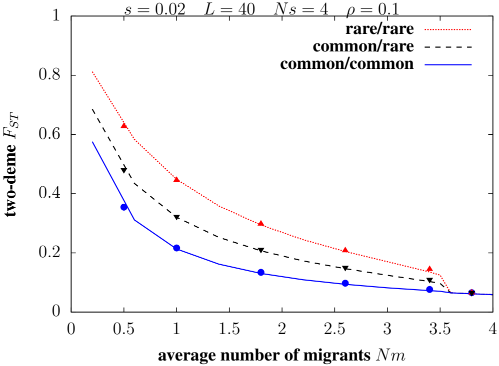
$$F_{ST}^{(rare)} = 1 - \frac{\mathbb{E}_{rare}[pq]}{\overline{p}_T \overline{q}_T} \qquad F_{ST}^{(com)} = 1 - \frac{\mathbb{E}_{com}[pq]}{\overline{p}_T \overline{q}_T}$$



No local adaptation:
$F_{ST} = \frac{1}{1+2Nm}$ (red line)

most immigrants into rare habitat originate from dissimilar habitat $\Longrightarrow$ lower net effective immigration into rare habitat $\Longrightarrow$ lower heterozygosity $\Longrightarrow$ higher F_{ST}.

F_{ST} between pairs of islands



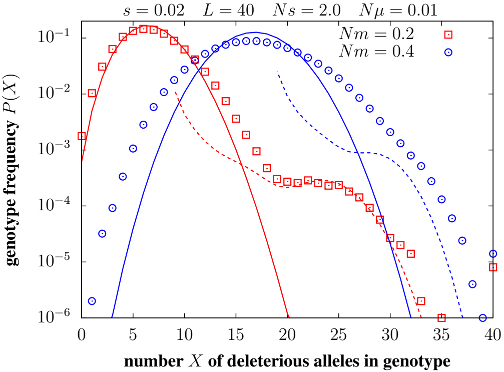
Summary

- Polygenic local adaptation $\implies$ **separation of timescales** (drift comparable to selection on alleles but not individuals) $\implies$ 'splice' **deterministic** description of **LD** with **stochastic** description of **allele frequencies**.
- Allele frequencies predicted very accurately by introducing **effective migration rates**: $m_e \approx m e^{-2sL(\mathbb{E}[p_1]-\mathbb{E}[p_2])}$ into the single-locus diffusion approximation.
- **LD** between divergently selected alleles
  $\implies$ **sharp thresholds** for loss of adaptation for $Ns \gtrsim 4$ and $\rho \ll 1$.
  $\implies$ **shifted thresholds** for $Ls \gtrsim \frac{1}{2(1-\rho)}$
- Neutral F_{ST} **higher** in **rare** habitat due to lower effective immigration. Pairwise F_{ST} not correlated with extent of adaptive divergence.
- **Temporal** dynamics: buildup of divergence?

Outlook

- Does separation of timescales extend to **linked** loci, **assortative** mating?
- Effective migration rates when loci have a **distribution of fitness** effects? When fitness has both **'local'** and **'global'** components?
  1. Exchange of weakly beneficial alleles between habitats?
  2. Interplay between local adaptation and inbreeding depression / heterosis?
  3. Associative overdominance
- **Hard selection**: effect of LD on extinction thresholds?
- How does this connect to **effective size** of metapopulation?

Predicting genotype frequencies



symbols → simulations
solid lines → LE
dashed lines → deterministic

Barriers to genetic exchange at neutral loci: expected F_{ST}

	simulations	coalescent	based on migration rates
F_{ST} (no LA)	$1-\frac{\mathbb{E}[pq]}{\overline{p}_T\overline{q}_T}$	$1-\frac{\mathbb{E}[T_w]}{\mathbb{E}[T_{tot}]}$	$\frac{1}{1+2Nm}$
$F_{ST}^{(rare)}$	$1-\frac{\mathbb{E}_{rare}[pq]}{\overline{p}_T\overline{q}_T}$	$1-\frac{\mathbb{E}[T_{w(rare)}]}{\mathbb{E}[T_{tot}]}$	$\frac{1}{1+2N[\boldsymbol{m_{r\,c}^{(e)}+m_{r\,r}^{(e)}}]}$
$F_{ST}^{(com)}$	$1-\frac{\mathbb{E}_{com}[pq]}{\overline{p}_T\overline{q}_T}$	$1-\frac{\mathbb{E}[T_{w(com)}]}{\mathbb{E}[T_{tot}]}$	$\frac{1}{1+2N[\boldsymbol{m_{c\,c}^{(e)}+m_{c\,r}^{(e)}}]}$

$$m_{c\,c}^{(e)} \approx (1-\rho)\,m \qquad m_{r\,r}^{(e)} \approx \rho\,m$$

$$m_{r\,c}^{(e)} \approx (1-\rho)\,m\,e^{-2sL(\mathbb{E}[p_1]-\mathbb{E}[p_2])} \qquad m_{c\,r}^{(e)} \approx \rho\,m\,e^{-2sL(\mathbb{E}[p_1]-\mathbb{E}[p_2])}$$

Different (effective) rates of immigration for rare vs. common habitats

$$\frac{1/F_{\mathrm{ST}}^{(rare)}-1}{1/F_{\mathrm{ST}}^{(com)}-1}=\frac{\text{eff. number of migrants into deme within rare habitat}}{\text{eff. number of migrants into deme within common habitat}}$$