MULTILOCUS MODELS, SELECTION, AND GENETIC DRIFT

Denis Roze
Institute of Evolutionary Biology
University of Edinburgh
MULTILOCUS POPULATION GENETICS

When do multilocus processes matter?
MULTILOCUS POPULATION GENETICS

When do multilocus processes matter?

• Interference between loci $\rightarrow$ rates of adaptation
MULTILOCUS POPULATION GENETICS

When do multilocus processes matter?

- Interference between loci $\rightarrow$ rates of adaptation
- Evolution of genetic systems and life-cycles
MULTILOCUS POPULATION GENETICS

When do multilocus processes matter?

- Interference between loci → rates of adaptation
- Evolution of genetic systems and life-cycles
  - mutation
  - recombination
  - sex
  - mating systems (e.g., selfing vs outcrossing)
  - mate choice
  - ploidy levels
  ...
MULTILOCUS POPULATION GENETICS

When do multilocus processes matter?

• Interference between loci $\rightarrow$ rates of adaptation

• Evolution of genetic systems and life-cycles

• Speciation
MULTILOCUS POPULATION GENETICS

When do multilocus processes matter?

Genetic drift
MULTILOCUS POPULATION GENETICS

When do multilocus processes matter?

Genetic drift

Finite population size
MULTILOCUS POPULATION GENETICS

When do multilocus processes matter?

Genetic drift

Finite population size

Population structure, « local drift »
MULTILOCUS POPULATION GENETICS

When do multilocus processes matter?

Genetic drift

A general formalism?
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002

• Description of populations in terms of allele frequencies and genetic associations
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002

• Description of populations in terms of allele frequencies and genetic associations

Ex: 3 biallelic loci, haploid

\[ p_i, p_j, p_k, D_{ij}, D_{ik}, D_{jk}, D_{ijk} \]
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002

• Description of populations in terms of allele frequencies and genetic associations

**Ex:** 3 biallelic loci, haploid

\[
\begin{array}{ccc}
& i & j & k \\
p_i, p_j, p_k, D_{ij}, D_{ik}, D_{jk}, D_{ijk}
\end{array}
\]

• Recursions
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002

• Description of populations in terms of allele frequencies and genetic associations

Ex: 3 biallelic loci, haploid

\[ p_i, p_j, p_k, D_{ij}, D_{ik}, D_{jk}, D_{ijk} \]

• Recursions

Selection:

\[ p_i' = p_i + \sum_U a_U D_{Ui} \]

\[ D_X' = D_X + \sum_U a_U \left( D_{UX} - D_U D_X \right) \]
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002

• Description of populations in terms of allele frequencies and genetic associations

**Ex:** 3 biallelic loci, haploid

\[ \begin{array}{c}
\begin{array}{ccc}
\text{i} & \text{j} & \text{k}
\end{array}
\end{array} \]

\[ p_i, p_j, p_k, D_{ij}, D_{ik}, D_{jk}, D_{ijk} \]

• Recursions

**Selection:**

\[ p_i' = p_i + \sum_U a_U D_{Ui} \]

\[ D'_X = D_X + \sum_U a_U (D_{UX} - D_U D_X) \]

**Recombination:**

\[ D''_X = \sum_{S+T=X} r_{S,T} D'_S D'_T \]
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002

• Description of populations in terms of allele frequencies and genetic associations

• Recursions
  → exact recursions
  → general results in terms of $a_U$ and $r_{S,T}$ coefficients
  → can be extended to diploidy, non-random mating, structured populations (infinite demes)....
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002

• Description of populations in terms of allele frequencies and genetic associations

• Recursions

• Quasi-linkage equilibrium approximation
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002

• Description of populations in terms of allele frequencies and genetic associations

• Recursions

• Quasi-linkage equilibrium approximation
  
  Separation of timescales, when selection is weak relative to recombination
A MULTILOCUS FORMALISM

Barton & Turelli 1991, Kirkpatrick et al 2002

• Description of populations in terms of allele frequencies and genetic associations

• Recursions

• *Quasi-linkage equilibrium* approximation
  
  Separation of timescales, when selection is weak relative to recombination

  Recursions in terms of allele frequencies only
A MULTILOCUS FORMALISM

Example: two biallelic loci $i$ and $j$, haploidy
A MULTILOCUS FORMALISM

Example: two biallelic loci $i$ and $j$, haploidy

Fitnesses:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>$1+s$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$1+s$</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>$(1+s)^2 + e$</td>
<td></td>
</tr>
</tbody>
</table>
A MULTILOCUS FORMALISM

Example: two biallelic loci \( i \) and \( j \), haploidy

Fitnesses:

\[
\begin{align*}
00 & \quad 1 \\
01 & \quad 1 + s \\
10 & \quad 1 + s \\
11 & \quad (1 + s)^2 + e
\end{align*}
\]

\[
\begin{align*}
a_i & \approx s + e p_j \\
a_j & \approx s + e p_i \\
a_{ij} & \approx e
\end{align*}
\]
A MULTILOCUS FORMALISM

Example: two biallelic loci $i$ and $j$, haploidy

Fitnesses:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>1</td>
<td>$a_i \approx s + e p_j$</td>
</tr>
<tr>
<td>01</td>
<td>1 + $s$</td>
<td>$a_j \approx s + e p_i$</td>
</tr>
<tr>
<td>10</td>
<td>1 + $s$</td>
<td>$a_{ij} \approx e$</td>
</tr>
<tr>
<td>11</td>
<td>$(1 + s)^2 + e$</td>
<td></td>
</tr>
</tbody>
</table>

Quasi-linkage equilibrium:

$$\Delta p_i \approx a_i p_i q_i$$

$$\Delta D_{ij} \approx -r_{ij} D_{ij} + \left(1 - r_{ij}\right) a_{ij} p_i q_i p_j q_j$$
A MULTILOCUS FORMALISM

Example: two biallelic loci $i$ and $j$, haploidy

Fitnesses:

| 00 | 1     | $a_i \approx s + e p_j$ |
| 01 | $1+s$ | $a_j \approx s + e p_i$ |
| 10 | $1+s$ | $a_{ij} \approx e$      |
| 11 | $(1+s)^2 + e$ |

Quasi-linkage equilibrium:

$$\Delta p_i \approx a_i p_i q_i$$

$$\Delta D_{ij} \approx -r_{ij} D_{ij} + \left(1-r_{ij}\right) a_{ij} p_i q_i p_j q_j$$

$$\longrightarrow D_{ij} \approx a_{ij} \left(1/r_{ij} - 1\right) p_i q_i p_j q_j$$
A MULTILocus FORMALISM

Example: two biallelic loci \(i\) and \(j\), haploidy

\[
\begin{align*}
\text{s = e = 0.01, } r = 0.1 \\
\text{s = e = 0.05, } r = 0.1 \\
\text{s = e = 0.1, } r = 0.1 \\
\text{s = e = 0.01, } r = 0.05 \\
\text{s = e = 0.05, } r = 0.05 \\
\text{s = e = 0.1, } r = 0.05
\end{align*}
\]
INCLUDING GENETIC DRIFT
INCLUDING GENETIC DRIFT

Find appropriate variables
INCLUDING GENETIC DRIFT

Find appropriate variables

Complete system of recursions?
INCLUDING GENETIC DRIFT

Find appropriate variables

Complete system of recursions?

Which approximations can be used?
INCLUDING GENETIC DRIFT

Find appropriate variables

Complete system of recursions?

Which approximations can be used?

General framework?
POPULATION STRUCTURE AND LOCAL DRIFT
POPULATION STRUCTURE AND LOCAL DRIFT

Drift generated by finite deme size and local competition
Drift generated by finite deme size and local competition

Infinite number of demes: deterministic model
POPULATION STRUCTURE AND LOCAL DRIFT

Drift generated by finite deme size and local competition

Infinite number of demes: deterministic model

Possible variables: moments of allele frequencies and genetic associations (averages and variances across demes, higher-order moments)
POPULATION STRUCTURE AND LOCAL DRIFT

Drift generated by finite deme size and local competition

Infinite number of demes: deterministic model

**Possible variables:** moments of allele frequencies and genetic associations (averages and variances across demes, higher-order moments)

These moments can also be represented by associations between genes present in different individuals from the same deme (link with concept of IBD)
POPULATION STRUCTURE AND LOCAL DRIFT

Example: infinite island model
Example: infinite island model

Association between two genes sampled with replacement at locus $i$:

$$D_{i/i} = p^2_{i(d)} - p^2_i$$
POPULATION STRUCTURE AND LOCAL DRIFT

Example: infinite island model

Association between two genes sampled with replacement at locus $i$:

$$D_{i/i} = p_{i(d)}^2 - p_i^2$$

Equilibrium value in the neutral case:

$$p_{i(d)}^2 = F p_i + (1-F) p_i^2$$

$$= p_i^2 + F p_i q_i$$

$$D_{i/i} = F p_i q_i$$
POPULATION STRUCTURE AND LOCAL DRIFT

Example: infinite island model

Association between four genes sampled in two individuals:

\[ D_{ij/ij} \]
POPULATION STRUCTURE AND LOCAL DRIFT

Example: infinite island model

Association between four genes sampled in two individuals:

\[ D_{ij/ij} \]

Neutral equilibrium:

\[ D_{ij/ij} = \phi p_i q_i p_j q_j \]

probability that the two pairs of genes are IBD
POPULATION STRUCTURE AND LOCAL DRIFT

Example: infinite island model

Association between four genes sampled in two individuals:

\[ D_{ij/ij} \]

Neutral equilibrium: \[ D_{ij/ij} = \phi p_i q_i p_j q_j \]

Selection: not possible to obtain complete sets of recursions
Example: infinite island model

Association between four genes sampled in two individuals:

\[ D_{ij/ij} \]

Neutral equilibrium: \[ D_{ij/ij} = \phi p_i q_i p_j q_j \]

Selection: not possible to obtain complete sets of recursions

**Quasi-equilibrium approximation:** under weak selection (relative to migration and recombination), all associations are close to their neutral equilibrium value, and equilibrate fast.
POPULATION STRUCTURE AND LOCAL DRIFT

Single-locus example, soft selection
POPULATION STRUCTURE AND LOCAL DRIFT

Single-locus example, soft selection

\[ \Delta p_i = \frac{(1+s)p_{i(d)}}{1 + sp_{i(d)}} - p_i \]
POPULATION STRUCTURE AND LOCAL DRIFT

Single-locus example, soft selection

\[
\Delta p_i = \frac{(1+s)p_{i(d)}}{1 + sp_{i(d)}} - p_i
\]

Weak selection:

\[
\Delta p_i = s\left(p_i - p_{i(d)}^2\right) + o(s)
\]

\[
= s\left(p_iq_i - D_{i/i}\right) + o(s)
\]
Single-locus example, soft selection

\[ \Delta p_i = \frac{(1+s)p_{i(d)}}{1 + sp_{i(d)}} - p_i \]

Weak selection:

\[ \Delta p_i = s\left(p_i - p_{i(d)}^2\right) + o\left(s\right) \]
\[ = s\left(p_i q_i - D_{i/i}\right) + o\left(s\right) \]

Quasi-equilibrium:

\[ \Delta p_i = s\left(1 - F\right)p_i q_i + o\left(s\right) \]
Single-locus example, soft selection

Quasi-equilibrium:

\[ \Delta p_i = s \left(1 - F \right) p_i q_i + o(s) \]

Possible to express associations at quasi-equilibrium, to leading order in selection coefficients
POPULATION STRUCTURE AND LOCAL DRIFT

Single-locus example, soft selection

\[ \Delta p_i = \frac{(1+s)p_{i(d)}}{1+sp_{i(d)}} - p_i \]
Single-locus example, soft selection

\[
\Delta p_i = \frac{(1 + s) p_{i(d)}}{1 + s p_{i(d)}} - p_i
\]

Large deme size:

\[
p_{i(d)} = p_i + \zeta_{i(d)} - p_i
\]  
\[
\zeta_{i(d)} \quad \text{(small)}
\]
Single-locus example, soft selection

\[ \Delta p_i = \frac{(1+s)p_{i(d)}}{1+sp_{i(d)}} - p_i \]

Large deme size:

\[ p_{i(d)} = p_i + p_{i(d)} - p_i \underbrace{\zeta_{i(d)}}_{(small)} \]

\[ \Delta p_i = \frac{sp_iq_i}{1+sp_i} - \frac{s(1+s)}{(1+sp_i)^3} \zeta_{i(d)}^2 + o\left(\zeta_i^2\right) \]
Single-locus example, soft selection

\[ \Delta p_i = \frac{(1 + s) p_{i(d)}}{1 + s p_{i(d)}} - p_i \]

Large deme size:
\[ p_{i(d)} = p_i + p_{i(d)} - p_i \]
\[ \zeta_{i(d)} \] (small)

\[ \Delta p_i = \frac{s p_i q_i}{1 + s p_i} - \frac{s (1 + s)}{(1 + s p_i)^3} D_{i/i} + o\left(\zeta_i^2\right) \]
Single-locus example, soft selection

\[ \Delta p_i = \frac{(1+s) p_{i(d)}}{1+s p_{i(d)}} - p_i \]

\[ \Delta p_i = \frac{sp_i q_i}{1+s p_i} - \frac{s(1+s)}{(1+s p_i)^3} D_{i/i} + o\left(\zeta_i^2\right) \]
Single-locus example, soft selection

\[ \Delta p_i = \frac{(1+s)p_{i(d)}}{1+sp_{i(d)}} - p_i \]

\[ \Delta p_i = \frac{sp_i q_i}{1+sp_i} - \frac{s(1+s)}{(1+sp_i)^3} D_{i/i} + o(\zeta_i^2) \]

→ need hypothesis on higher-order moments to close the system of recursions
Single-locus example, soft selection

\[ \Delta p_i = \frac{(1+s)p_{i(d)}}{1+sp_{i(d)}} - p_i \]

\[ \Delta p_i = \frac{sp_i q_i}{1+sp_i} - \frac{s(1+s)}{(1+sp_i)^3} D_{i/i} + o\left(\zeta_i^2\right) \]

→ need hypothesis on higher-order moments to close the system of recursions

→ method used by Barton & Otto (2005), Martin et al (2005)
INBREEDING DEPRESSION AND THE EVOLUTION OF DISPERSAL RATES

Evolution of dispersal:
INBREEDING DEPRESSION AND THE EVOLUTION OF DISPERSAL RATES

Evolution of dispersal:

• Individual selection: cost of dispersal
INBREEDING DEPRESSION AND THE EVOLUTION OF DISPERSAL RATES

Evolution of dispersal:

• Individual selection: cost of dispersal

• Kin selection: dispersal reduces competition among kin
INBREEDING DEPRESSION AND THE EVOLUTION OF DISPERSAL RATES

Evolution of dispersal:

- Individual selection
- Kin selection

\[
\text{ESS dispersal rate} \approx \frac{1}{2Nc}
\]

(infinite island model)
INBREEDING DEPRESSION AND THE EVOLUTION OF DISPERSAL RATES

Evolution of dispersal:

• Individual selection

• Kin selection

• Indirect selection: selection at other loci than loci determining dispersal rates
INBREEDING DEPRESSION AND THE EVOLUTION OF DISPERSAL RATES

Evolution of dispersal:

- Individual selection
- Kin selection
- Indirect selection: selection at other loci than loci determining dispersal rates

**Inbreeding depression** (due to recessive deleterious alleles):

→ generates heterosis (gives an advantage to migrants)
→ decreases strength of kin selection for dispersal
TWO-LOCUS MODEL

- Infinite island model, random mating within demes
- Finite deme size $N$
- Migration cost $c$

![Diagram](image)

**locus 1: migration modifier**
- $aa$: $m$
- $Aa$: $m + \varepsilon / 2$
- $AA$: $m + \varepsilon$

**locus 2: selected locus**
- $bb$: 1
- $Bb$: $1 - hs$ (fecundities)
- $BB$: $1 - s$
VARIABLES

Allele frequencies (in the whole metapopulation):

\[ p_1 = \text{freq}(A), \quad p_2 = \text{freq}(B) \quad q_1 = 1 - p_1, \quad q_2 = 1 - p_2 \]
VARIABLES

Allele frequencies (in the whole metapopulation):

\[ p_1 = \text{freq}(A), \quad p_2 = \text{freq}(B), \quad q_1 = 1 - p_1, \quad q_2 = 1 - p_2 \]

Genetic associations:
VARIABLES

Allele frequencies (in the whole metapopulation):

\[ p_1 = \text{freq}(A) , \ p_2 = \text{freq}(B) \quad q_1 = 1 - p_1, \ q_2 = 1 - p_2 \]

Genetic associations:
VARIABLES

Allele frequencies (in the whole metapopulation):

\[ p_1 = \text{freq}(A), \quad p_2 = \text{freq}(B) \quad q_1 = 1 - p_1, \quad q_2 = 1 - p_2 \]

Genetic associations:

\[ D_{1/1}, \quad D_{1,1}, \quad D_{12} : \text{linkage disequilibrium} \]
VARIABLES

Allele frequencies (in the whole metapopulation):

\[ p_1 = \text{freq}(A), \quad p_2 = \text{freq}(B) \]

\[ q_1 = 1 - p_1, \quad q_2 = 1 - p_2 \]

Genetic associations:

Neutrality:

\[ D_{1,1} = F_{IT} p_1 q_1 \]
\[ D_{1/1} = F_{ST} p_1 q_1 \]

Random mating within demes:

\[ F_{IT} = F_{ST} = F \approx \frac{1}{1 + 4Nm(1-c)} \]

\[ D_{12}: \text{linkage disequilibrium} \]
LIFE-CYCLE

• **Reproduction** within demes: differences in fecundity according to genotype at locus 2

• **Dispersal**: differences in dispersal rates according to genotype at locus 1
CHANGE IN FREQUENCY OF THE MODIFIER

During reproduction:

\[ \Delta_s p_1 = -s \left(1 - 2h\right) \left(1 - \frac{1}{N}\right) \left(D_{12,2} - D_{1/2,2}\right) + o(s) \]
CHANGE IN FREQUENCY OF THE MODIFIER

During reproduction:

\[ \Delta_S p_1 = -s \left(1 - 2h \right) \left(1 - \frac{1}{N} \right) \left( D_{12,2} - D_{1/2,2} \right) + o(s) \]

\[ \varepsilon > 0 : \]

\[ D_{12,2} < 0: \]

individuals carrying allele A are less homozygous at locus 2
CHANGE IN FREQUENCY OF THE MODIFIER

During reproduction:

\[ \Delta_S p_1 = -s \left( 1 - 2h \right) \left( 1 - \frac{1}{N} \right) \left( D_{12,2} - D_{1/2,2} \right) + o(s) \]

\( \varepsilon > 0 \):

- \( D_{12,2} < 0 \): individuals carrying allele A are less homozygous at locus 2

- \( D_{1/2,2} < 0 \): individuals in deme with higher frequency of allele A are less homozygous at locus 2
CHANGE IN FREQUENCY OF THE MODIFIER

During dispersal:

\[ \Delta_M p_1 = \frac{\varepsilon}{2(1-cm)} \left[ -c p_1 q_1 + 2 \left( \frac{1-m}{1-cm} - c \right) D_{1/1} \right] + o(\varepsilon) \]
CHANGE IN FREQUENCY OF THE MODIFIER

During dispersal:

$$\Delta_M p_1 = \frac{\varepsilon}{2(1-cm)} \left[-c p_1 q_1 + 2 \left( \frac{1-m}{1-cm} - c \right) D_{1/1} \right] + o(\varepsilon)$$
CHANG IN FREQUENCY OF THE MODIFIER

During dispersal:

\[
\Delta_M p_1 = \frac{\epsilon}{2(1-cm)} \left[ -c p_1 q_1 + 2 \left( \frac{1-m}{1-cm} - c \right) D_{1/1} \right] + o(\epsilon)
\]

\[
s = 0 \rightarrow D_{1/1} = F p_1 q_1 + O(\epsilon)
\]
CHANGE IN FREQUENCY OF THE MODIFIER

During dispersal:

\[
\Delta_{M} p_1 = \frac{\epsilon}{2(1-cm)} \left[ -c p_1 q_1 + 2 \left( \frac{1-m}{1-cm} - c \right) D_{1/1} \right] + o(\epsilon)
\]

\[
s = 0 \quad \rightarrow \quad D_{1/1} = F p_1 q_1 + O(\epsilon)
\]

\[
s \neq 0 \quad \rightarrow \quad D_{1/1} \text{ is affected by } s \quad \rightarrow \quad \text{term in } \epsilon s
\]
 APPROXIMATE SOLUTIONS

$m$ small, $N$ large

Change in frequency of the modifier during reproduction:

$$\Delta_S p_1 \approx \varepsilon s (1 - 2h) \frac{1-c}{r} \frac{1 + 2(1-c)Nm}{[1 + 4(1-c)Nm]^2} p_1 q_1 p_2 q_2$$

Change in frequency of the modifier during dispersal:

$$\Delta_M p_1 \approx (\Delta_M p_1)_{s=0} - \varepsilon s (1 - 2h) \frac{2(1-c)(2-c)Nm(1+r)}{r[1 + 4(1-c)Nm]^3} p_1 q_1 p_2 q_2$$
APPROXIMATE SOLUTIONS

$m$ small, $N$ large

Change in frequency of the modifier during reproduction:

$$\Delta_s p_1 \approx \varepsilon s (1-2h) \frac{1-c}{r} \frac{1+2(1-c)Nm}{\left[1+4(1-c)Nm\right]^2} p_1 q_1 p_2 q_2$$

Change in frequency of the modifier during dispersal:

$$\Delta_M p_1 \approx (\Delta_M p_1)_{s=0} - \varepsilon s (1-2h) \frac{2(1-c)(2-c)Nm(1+r)}{r\left[1+4(1-c)Nm\right]^3} p_1 q_1 p_2 q_2$$

→ two different effects of inbreeding depression, of opposite sign
ESS DISPERSAL RATE

Multilocus simulations:

\[ U \]

\[ c = 0.4, \ N = 30 \]
\[ h = 0.1, \ s = 0.01 \]
\[ R = 10 \]
ESS DISPERSAL RATE

Multilocus simulations:

\[ r, \quad s, \quad h \text{ constant} \]

\[ c = 0.8, \quad N = 30 \]
\[ h = 0.1, \quad s = 0.001 \]
\[ R = 10 \]
CONCLUSION

Multilocus models, selection and drift
CONCLUSION

Multilocus models, selection and drift

Not possible to derive complete systems of recursions
CONCLUSION

Multilocus models, selection and drift

Not possible to derive complete systems of recursions

Approximations:

– weak selection / quasi-equilibrium
– weak effects of drift / moment closure
CONCLUSION

Multilocus models, selection and drift

Not possible to derive complete systems of recursions

Approximations:

– weak selection / quasi-equilibrium
– weak effects of drift / moment closure

Becomes tedious when associations between more than a few loci do matter...