Introgression & Deriving Patterson's D

BIOL 1435

March 7, 2023

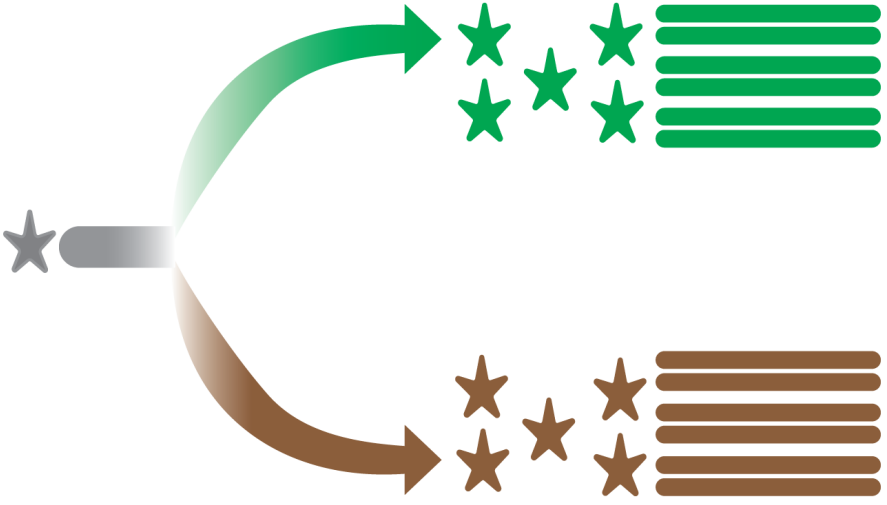
Overview

1. Motivation
2. Preliminaries
3. Derivations

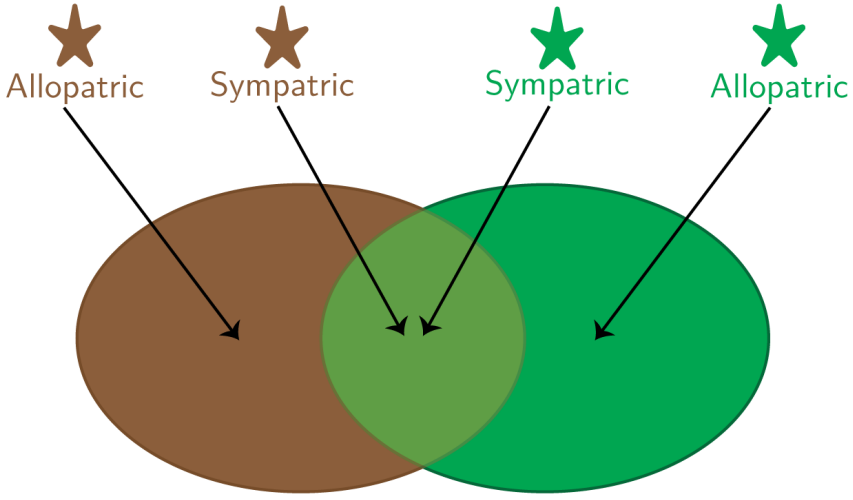
Overview

1. **Motivation**
2. **Preliminaries**
3. **Derivations**

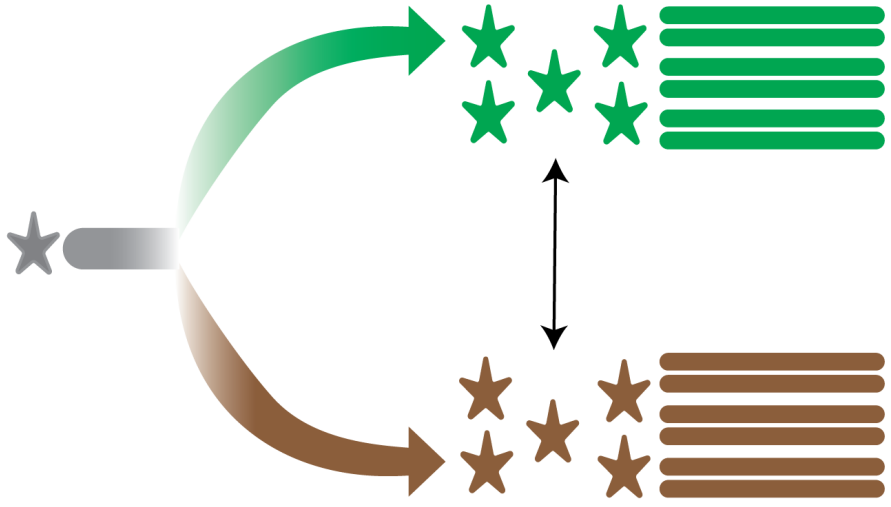
Evolution as a bifurcating process



Sympatry is necessary for gene flow

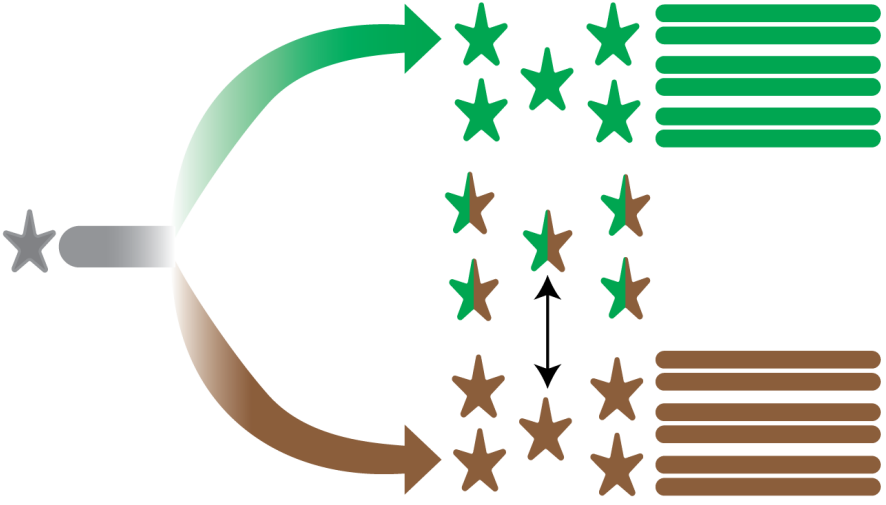


Incomplete RI leads to hybridization



F1 hybrids = uniform mixture of parental ancestry

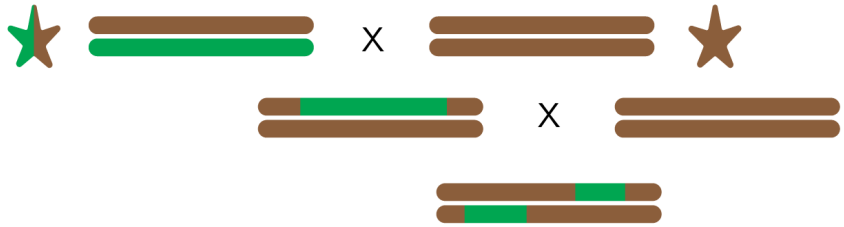
Hybridization is NOT sufficient for introgression



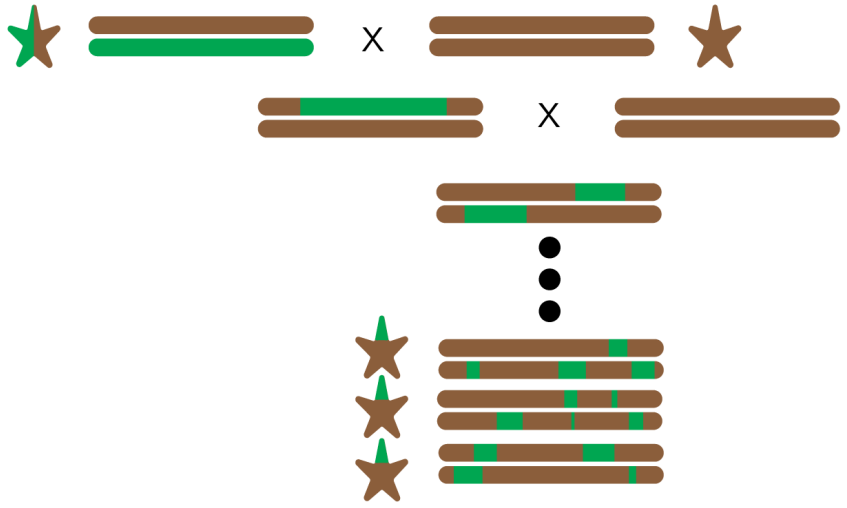
Incorporation of heterospecific loci via backcrossing

Incorporation of heterospecific loci via backcrossing

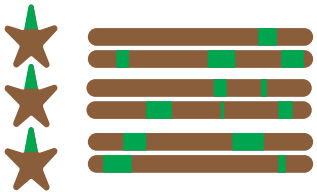
Incorporation of heterospecific loci via backcrossing



Incorporation of heterospecific loci via backcrossing

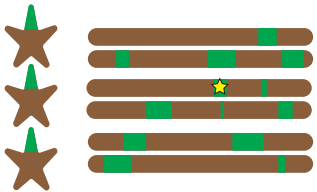


Adaptive introgression



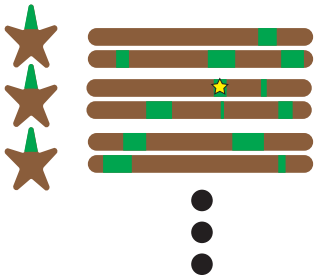
Before selection

Introgression may introduce beneficial mutations



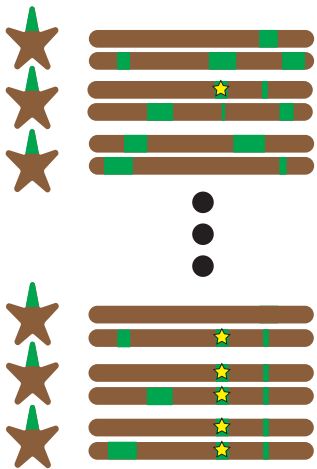
Before selection

The beneficial mutation rises in frequency



Before selection

The beneficial mutation is eventually fixed or nearly fixed

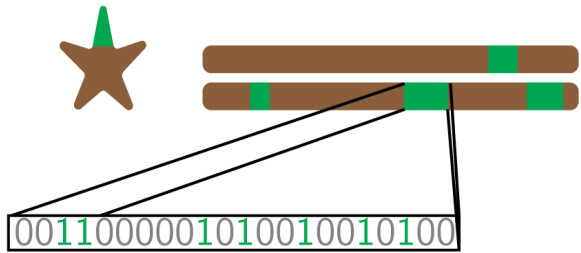


Before selection

After selection

Genetic composition of an introgressed segment

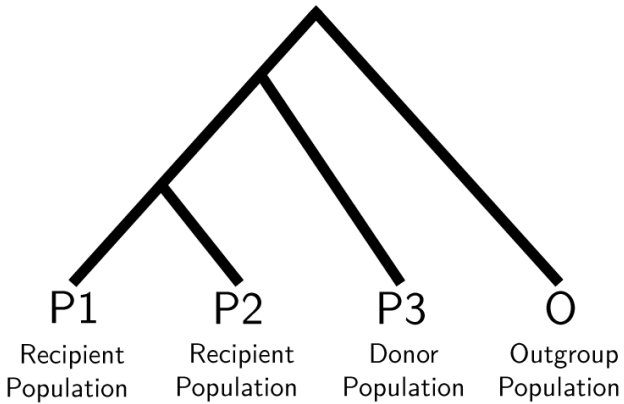
Introgression segments leaves a genomic footprint



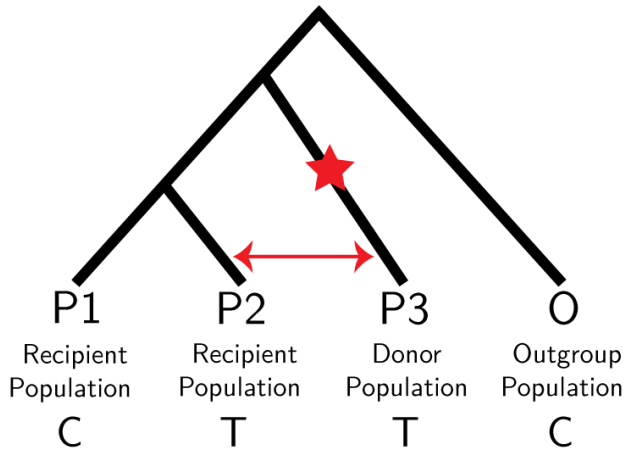
0 = reintroduced ancestral allele

1 = newly introduced derived allele

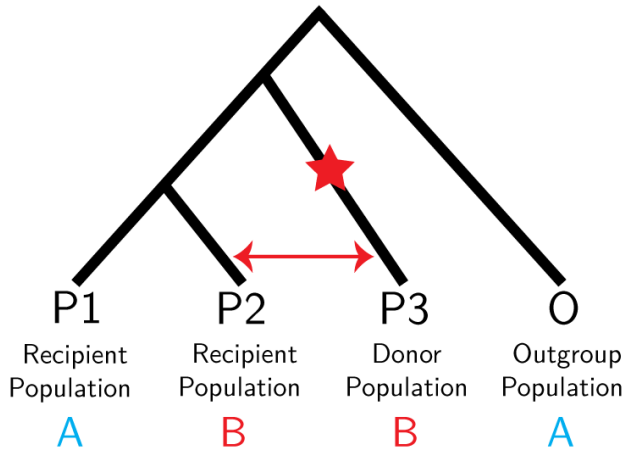
Site pattern tests of introgression



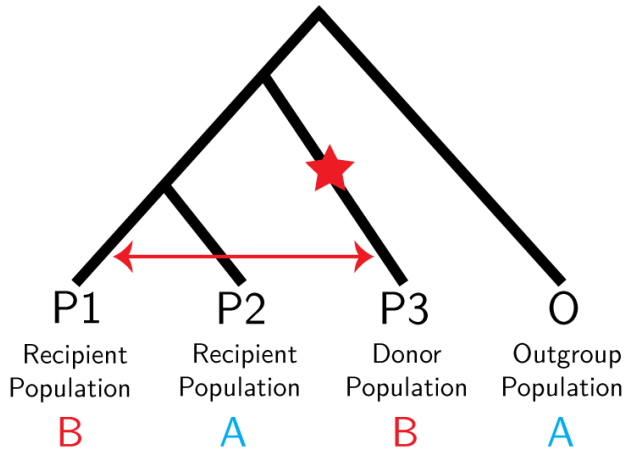
Derived allele sharing (P2 $\longleftrightarrow$ P3)



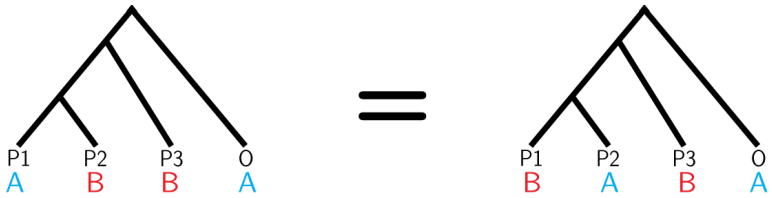
Derived allele sharing (P2 $\longleftrightarrow$ P3)



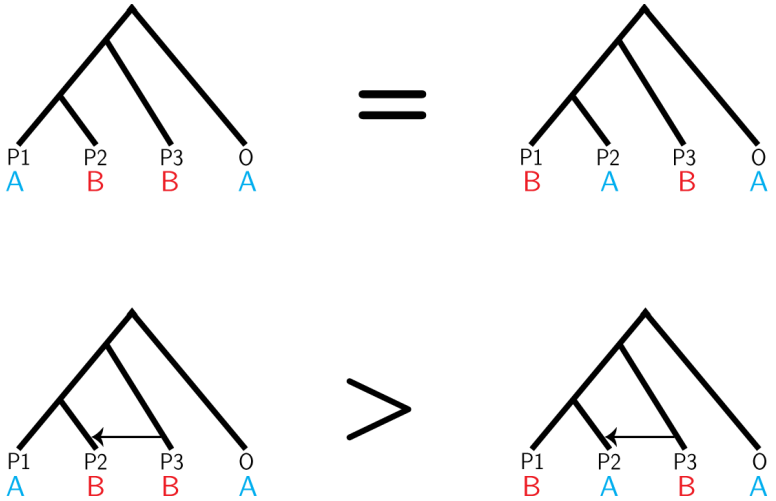
Derived allele sharing (P1 $\longleftrightarrow$ P3)



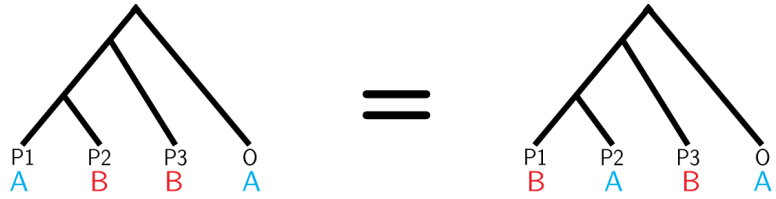
Patterson's D



Patterson's D

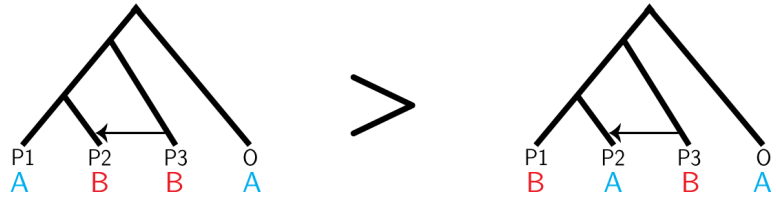


Patterson's D

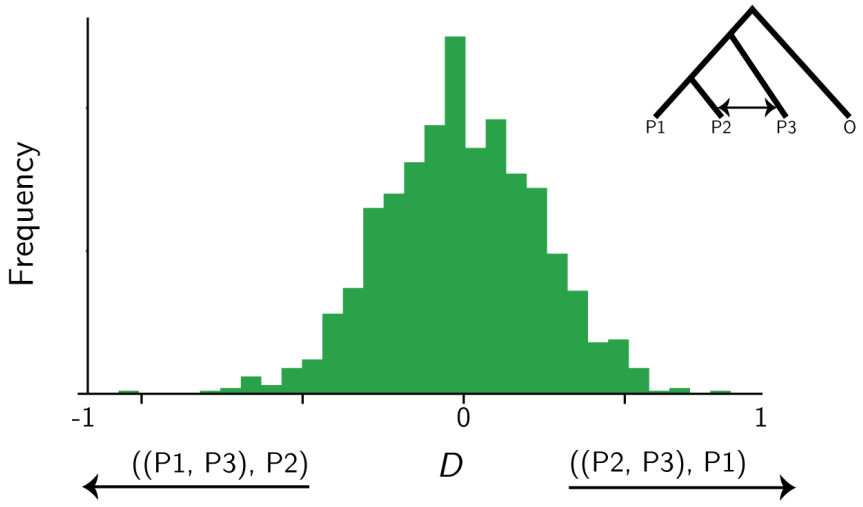


$$Patterson's\ D = \frac{\sum_{i=1}^{n} ABBA_i - \sum_{i=1}^{n} BABA_i}{\sum_{i=1}^{n} ABBA_i + \sum_{i=1}^{n} BABA_i}$$

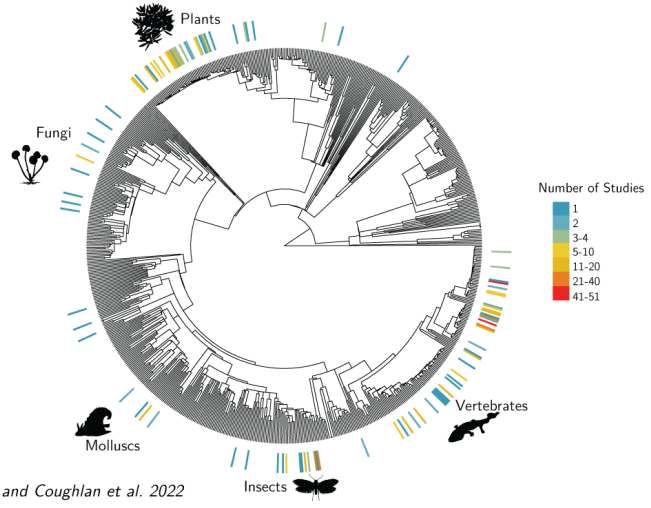
Green et al. 2010 & Durand et al. 2011



Patterson's D

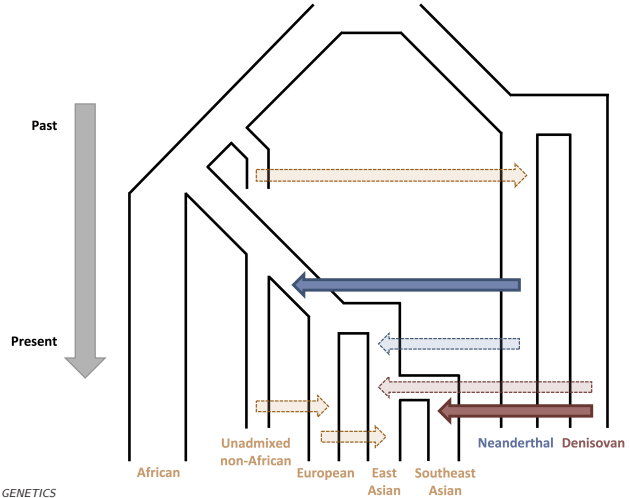


How common is introgression in eukaryotes?



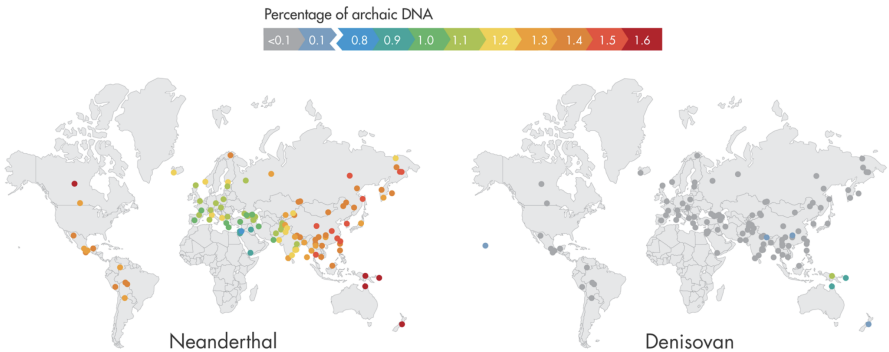
Dagilis, **Peede**, *and Coughlan et al. 2022*

An oversimplified view of human evolution



Wolf & Akey 2018, *PLOS GENETICS*

Global distribution of archaic DNA



From Quanta Magazine

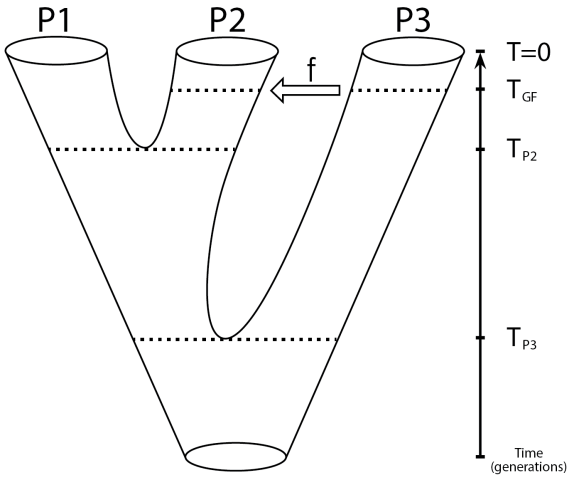
Overview

1. Motivation

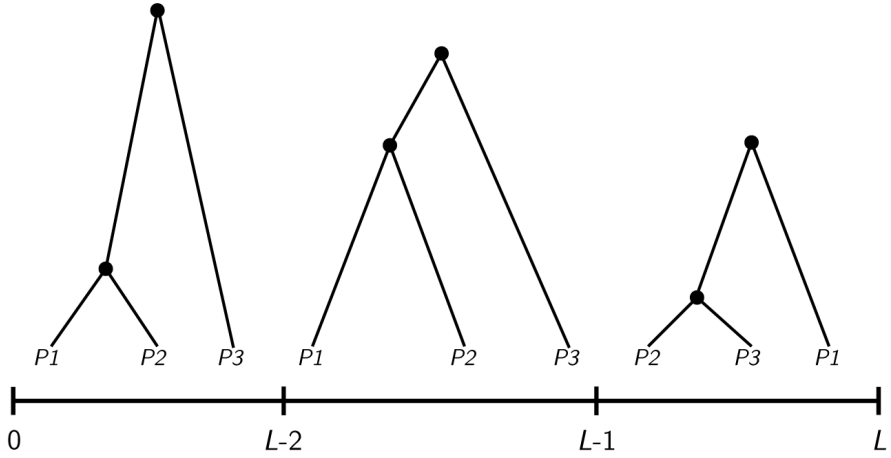
2. **Preliminaries**

3. Derivations

Instantaneous unidirectional admixture model



Each locus has its own coalescent history



Probability of gene flow

Equation

$$Pr\,(\textit{gene flow}) = f \tag{1}$$

$$Pr\,(\textit{no gene flow}) = (1 - f) \tag{2}$$

Where f represents the admixture proportion—the probability that any lineage from *P3* migrates to *P2*.

Probability of no coalescence during time interval t

Probability of no coalescence during time interval t

Equation

$$Pr\left(no\ coalescences\right) = \left(1 - \frac{1}{2N}\right)^{t} \tag{3}$$

Where t denotes the time interval where coalescence can occur.

Probability of coalescence during time interval t

Probability of coalescence during time interval t

Equation

$$Pr\left(coalescences\right) = 1 - \left(1 - \frac{1}{2N}\right)^{t} \tag{4}$$

Where t denotes the time interval where coalescence can occur.

Expected time of coalescence T_2

Expected time of coalescence T_2

Equation

$$T_2 = 2N \tag{5}$$

Expected time of coalescence | coalescence during t

Expected time of coalescence | coalescence during t

Equation

$$i \sim Tgeo\left(t|p, c\right) \tag{6}$$

Expected time of coalescence | coalescence during t

Equation

$$i \sim Tgeo\left(t|p,c\right) \tag{6}$$

$$f\left(i\right) = \frac{\frac{1}{2N}\left(1-\frac{1}{2N}\right)^{i-1}}{1-\left(1-\frac{1}{2N}\right)^{c}} \tag{7}$$

Expected time of coalescence | coalescence during t

Equation

$$i \sim Tgeo\left(t|p,c\right) \tag{6}$$

$$f\left(i\right) = \frac{\frac{1}{2N}\left(1-\frac{1}{2N}\right)^{i-1}}{1-\left(1-\frac{1}{2N}\right)^{c}} \tag{7}$$

$$\bar{t} = \mathbb{E}\left(i\right) = \sum_{i=1}^{c} i\frac{\frac{1}{2N}\left(1-\frac{1}{2N}\right)^{i-1}}{1-\left(1-\frac{1}{2N}\right)^{c}} = \frac{2N-\left(\left(1-\frac{1}{2N}\right)^{c}\left(c+2N\right)\right)}{1-\left(1-\frac{1}{2N}\right)^{c}} \tag{8}$$

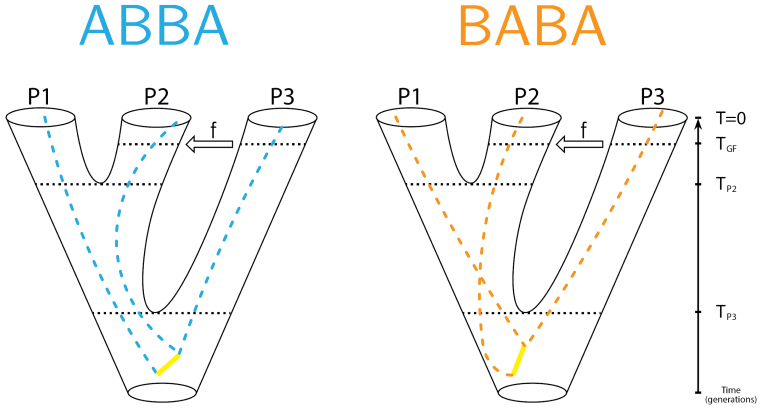
Where c denotes the time interval where coalescence must occur.

Overview

1. Motivation
2. Preliminaries
3. **Derivations**

Coalescent history 1

No gene flow from $P3 \rightarrow P2$, *P1* & *P2* don't coalesce between T_{P2} & T_{P3}, and *P1*/*P2* & *P3* coalesce first.



Coalescent history 1

No gene flow from $P3 \rightarrow P2$, *P1* & *P2* don't coalesce between T_{P2} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Coalescent history 1

No gene flow from $P3 \rightarrow P2$, *P1* & *P2* don't coalesce between T_{P2} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Derivation

$$Pr\left(\textit{no gene flow}\right) = (1 - f) \tag{9}$$

Coalescent history 1

No gene flow from $P3 \rightarrow P2$, *P1* & *P2* don't coalesce between T_{P2} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Derivation

$$Pr\left(\textit{no gene flow}\right) = (1 - f) \tag{9}$$

$$Pr\left(\textit{P1 \& P2 don't coalesce between } T_{P2} \textit{ \& } T_{P3}\right) = \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{P2}} \tag{10}$$

Coalescent history 1

No gene flow from $P3 \rightarrow P2$, *P1* & *P2* don't coalesce between T_{P2} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Derivation

$$Pr\left(\textit{no gene flow}\right) = (1 - f) \tag{9}$$

$$Pr\left(\textit{P1 \& P2 don't coalesce between } T_{P2} \textit{ \& } T_{P3}\right) = \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{P2}} \tag{10}$$

$$Pr\left(\textit{P1/P2 \& P3 coalesce first}\right) = \frac{1}{3} \tag{11}$$

Coalescent history 1

No gene flow from $P3 \rightarrow P2$, *P1* & *P2* don't coalesce between T_{P2} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Derivation

$$Pr\left(\textit{no gene flow}\right) = (1 - f) \tag{9}$$

$$Pr\left(\textit{P1 \& P2 don't coalesce between } T_{P2} \textit{ \& } T_{P3}\right) = \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{P2}} \tag{10}$$

$$Pr\left(\textit{P1/P2 \& P3 coalesce first}\right) = \frac{1}{3} \tag{11}$$

$$\mathbb{E}\left(\textit{Branch length between the } 1^{st} \textit{ \& } 2^{nd} \textit{coalescent event}\right) = 2N \tag{12}$$

Coalescent history 1

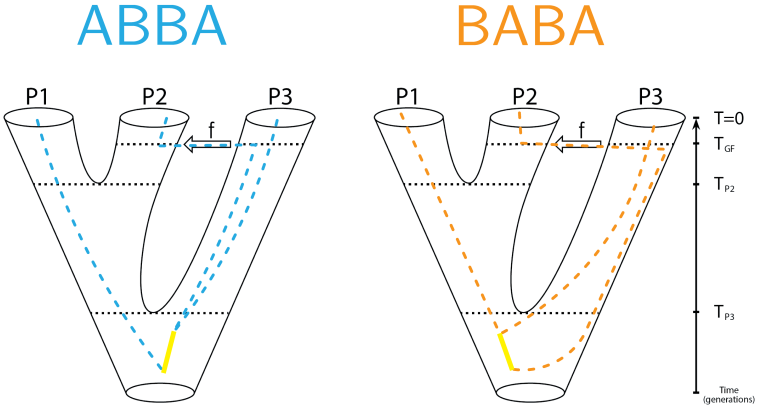
No gene flow from $P3 \rightarrow P2$, *P1* & *P2* don't coalesce between T_{P2} & T_{P3}, and *P1*/*P2* & *P3* coalesce first.

Derivation

$$\mathbb{E}\left(C_{ABBA_1}\right) = \mathbb{E}\left(C_{BABA_1}\right) = (1-f)\left(1-\frac{1}{2N}\right)^{T_{P3}-T_{P2}}\frac{2N}{3} \tag{13}$$

Coalescent history 2

Gene flow from $P3 \rightarrow P2$, *P2* & *P3* don't coalesce between T_{GF} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.



Coalescent history 2

Gene flow from $P3 \rightarrow P2$, *P2* & *P3* don't coalesce between T_{GF} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Coalescent history 2

Gene flow from $P3 \rightarrow P2$, *P2* & *P3* don't coalesce between T_{GF} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Derivation

$$Pr\,(gene\ flow) = f \tag{14}$$

Coalescent history 2

Gene flow from $P3 \rightarrow P2$, *P2* & *P3* don't coalesce between T_{GF} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Derivation

$$Pr\left(\textit{gene flow}\right) = f \tag{14}$$

$$Pr\left(\textit{P2 \& P3 don't coalesce between } T_{GF} \textit{ \& } T_{P3}\right) = \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{GF}} \tag{15}$$

Coalescent history 2

Gene flow from $P3 \rightarrow P2$, $P2$ & $P3$ don't coalesce between T_{GF} & T_{P3}, and $P1$ /$P2$ & $P3$ coalesce first.

Derivation

$$Pr\left(gene\ flow\right) = f \tag{14}$$

$$Pr\left(P2\ \&\ P3\ don't\ coalesce\ between\ T_{GF}\ \&\ T_{P3}\right) = \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{GF}} \tag{15}$$

$$Pr\left(P1/P2\ \&\ P3\ coalesce\ first\right) = \frac{1}{3} \tag{16}$$

Coalescent history 2

Gene flow from $P3 \rightarrow P2$, *P2* & *P3* don't coalesce between T_{GF} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Derivation

$$Pr\left(\textit{gene flow}\right) = f \tag{14}$$

$$Pr\left(\textit{P2 \& P3 don't coalesce between } T_{GF} \textit{ \& } T_{P3}\right) = \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{GF}} \tag{15}$$

$$Pr\left(\textit{P1/P2 \& P3 coalesce first}\right) = \frac{1}{3} \tag{16}$$

$$\mathbb{E}\left(\textit{Branch length between the } 1^{st} \textit{ \& } 2^{nd} \textit{coalescent event}\right) = 2N \tag{17}$$

Coalescent history 2

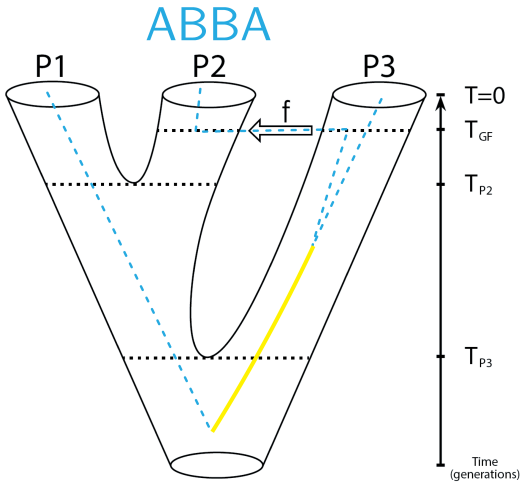
Gene flow from $P3 \rightarrow P2$, *P2* & *P3* don't coalesce between T_{GF} & T_{P3}, and *P1* /*P2* & *P3* coalesce first.

Derivation

$$\mathbb{E}\left(C_{ABBA_2}\right) = \mathbb{E}\left(C_{BABA_2}\right) = f\left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{GF}} \frac{2N}{3} \tag{18}$$

Coalescent history 3

Gene flow from $P3 \rightarrow P2$ and $P2$ & $P3$ coalesce between T_{GF} & T_{P3}.



Coalescent history 3

Gene flow from $P3 \rightarrow P2$ and $P2$ & $P3$ coalesce between T_{GF} & T_{P3}.

Coalescent history 3

Gene flow from $P3 \rightarrow P2$ and $P2$ & $P3$ coalesce between T_{GF} & T_{P3}.

Derivation

$$Pr\left(gene\ flow\right) = f \tag{19}$$

Coalescent history 3

Gene flow from $P3 \rightarrow P2$ and $P2$ & $P3$ coalesce between T_{GF} & T_{P3}.

Derivation

$$Pr\left(\textit{gene flow}\right) = f \tag{19}$$

$$Pr\left(\textit{P2 \& P3 coalesce between } T_{GF} \textit{ \& } T_{P3}\right) = 1 - \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{GF}} \tag{20}$$

Coalescent history 3

Gene flow from $P3 \rightarrow P2$ and $P2$ & $P3$ coalesce between T_{GF} & T_{P3}.

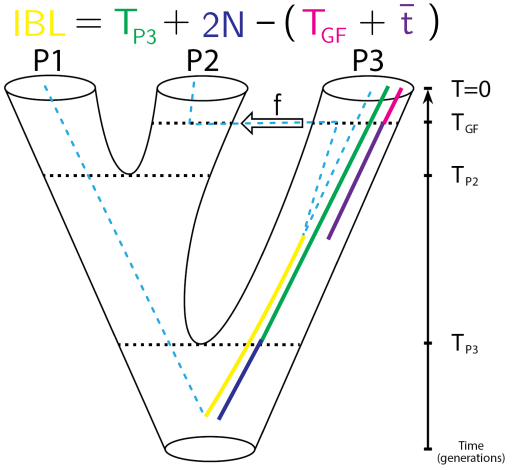
Derivation

$$Pr\left(\textit{gene flow}\right) = f \tag{19}$$

$$Pr\left(\textit{P2 \& P3 coalesce between } T_{GF} \textit{ \& } T_{P3}\right) = 1 - \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{GF}} \tag{20}$$

$$\mathbb{E}\left(\textit{Branch length between the } 1^{st} \textit{ \& } 2^{nd} \textit{coalescent event}\right) = \left(T_{P3} + 2N\right) - \left(T_{GF} + \overline{t}\right) \tag{21}$$

Coalescent history 3 ($\bar{t}$)



Coalescent history 3

Gene flow from $P3 \rightarrow P2$ and $P2$ & $P3$ coalesce between T_{GF} & T_{P3}.

Derivation

$$\mathbb{E}\left(C_{ABBA_3}\right) = f\left(1 - \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{GF}}\right)\left(\left(T_{P3} + 2N\right) - \left(T_{GF} + \overline{t}\right)\right) \tag{22}$$

Coalescent history 3

Gene flow from $P3 \rightarrow P2$ and $P2$ & $P3$ coalesce between T_{GF} & T_{P3}.

Derivation

$$\mathbb{E}\left(C_{ABBA_3}\right) = f\left(1 - \left(1 - \frac{1}{2N}\right)^{T_{P3} - T_{GF}}\right)\left(\left(T_{P3} + 2N\right) - \left(T_{GF} + \overline{t}\right)\right) \tag{22}$$

$$\mathbb{E}\left(C_{BABA_3}\right) = 0 \tag{23}$$

$\mathbb{E}(ABBA)$

Derivation

$$\mathbb{E}(\tau_{ABBA}) = C_{ABBA_1} + C_{ABBA_2} + C_{ABBA_3} \tag{24}$$

$\mathbb{E}(ABBA)$

Derivation

$$\mathbb{E}(\tau_{ABBA}) = C_{ABBA_1} + C_{ABBA_2} + C_{ABBA_3} \tag{24}$$

$$\begin{aligned}\mathbb{E}(\tau_{ABBA}) = (1-f)\left(2N/3\,(1-1/2N)^{T_{P3}-T_{P2}}\right) \\ + (f)\left(\left(2N/3\,(1-1/2N)^{T_{P3}-T_{GF}}\right) + (T_{P3}-T_{GF})\right)\end{aligned} \tag{25}$$

$\mathbb{E}(ABBA)$

Derivation

$$\mathbb{E}(\tau_{ABBA}) = C_{ABBA_1} + C_{ABBA_2} + C_{ABBA_3} \tag{24}$$

$$\begin{aligned}\mathbb{E}(\tau_{ABBA}) = (1-f)\left(2N/3\,(1-1/2N)^{T_{P3}-T_{P2}}\right) \\ + (f)\left(\left(2N/3\,(1-1/2N)^{T_{P3}-T_{GF}}\right) + (T_{P3}-T_{GF})\right)\end{aligned} \tag{25}$$

$$\mathbb{E}(ABBA_{sites}) = \mathbb{E}(\tau_{ABBA}) \times \mu \times L \tag{26}$$

Where μ represents the mutation rate and L represents the sequence length.

$\mathbb{E}(BABA)$

Derivation

$$\mathbb{E}(\tau_{BABA}) = C_{BABA_1} + C_{BABA_2} \tag{27}$$

$\mathbb{E}(BABA)$

Derivation

$$\mathbb{E}(\tau_{BABA}) = C_{BABA_1} + C_{BABA_2} \tag{27}$$

$$\begin{aligned}\mathbb{E}(\tau_{BABA}) = (1-f)\left(2N/3\,(1-1/2N)^{T_{P3}-T_{P2}}\right) \\ + (f)\left(2N/3\,(1-1/2N)^{T_{P3}-T_{GF}}\right)\end{aligned} \tag{28}$$

$\mathbb{E}(BABA)$

Derivation

$$\mathbb{E}(\tau_{BABA}) = C_{BABA_1} + C_{BABA_2} \tag{27}$$

$$\begin{aligned}\mathbb{E}(\tau_{BABA}) = (1-f)\left(2N/3\,(1-1/2N)^{T_{P3}-T_{P2}}\right) \\ + (f)\left(2N/3\,(1-1/2N)^{T_{P3}-T_{GF}}\right)\end{aligned} \tag{28}$$

$$\mathbb{E}(BABA_{sites}) = \mathbb{E}(\tau_{BABA}) \times \mu \times L \tag{29}$$

Where μ represents the mutation rate and L represents the sequence length.

Patterson's D

Derivation

$$\mathbb{E}(D) = \frac{\mathbb{E}(\tau_{ABBA}) - \mathbb{E}(\tau_{BABA})}{\mathbb{E}(\tau_{ABBA}) + \mathbb{E}(\tau_{BABA})} \tag{30}$$

Patterson's D

Derivation

$$\mathbb{E}(D) = \frac{\mathbb{E}(\tau_{ABBA}) - \mathbb{E}(\tau_{BABA})}{\mathbb{E}(\tau_{ABBA}) + \mathbb{E}(\tau_{BABA})} \tag{30}$$

$$\mathbb{E}(D) = \frac{(f)(T_{P3} - T_{GF})}{(1-f)\left[{}^{4N}/_{3}(1 - {}^{1}/_{2N})^{T_{P3}-T_{P2}}\right] + (f)\left[\left({}^{4N}/_{3}(1 - {}^{1}/_{2N})^{T_{P3}-T_{GF}}\right) + (T_{P3} - T_{GF})\right]} \tag{31}$$

Patterson's D

Derivation

$$\mathbb{E}(D) = \frac{\mathbb{E}(\tau_{ABBA}) - \mathbb{E}(\tau_{BABA})}{\mathbb{E}(\tau_{ABBA}) + \mathbb{E}(\tau_{BABA})} \tag{30}$$

$$\mathbb{E}(D) = \frac{(f)(T_{P3} - T_{GF})}{(1-f)\left[{}^{4N}/_{3}(1 - {}^{1}/_{2N})^{T_{P3}-T_{P2}}\right] + (f)\left[\left({}^{4N}/_{3}(1 - {}^{1}/_{2N})^{T_{P3}-T_{GF}}\right) + (T_{P3} - T_{GF})\right]} \tag{31}$$

$$\mathbb{E}(D) = \frac{\sum_{i=1}^{L}(1-p_{i1})p_{i2}p_{i3}(1-p_{iO}) - p_{i1}(1-p_{i2})p_{i3}(1-p_{iO})}{\sum_{i=1}^{L}(1-p_{i1})p_{i2}p_{i3}(1-p_{iO}) + p_{i1}(1-p_{i2})p_{i3}(1-p_{iO})} \tag{32}$$

Where $p_{i\#}$ represents the derived allele frequency at site i and L represents the sequence length.