#### Allele frequencies in polyploids

Botany 201!

Tests for introgression

f estimators

Ponetomon

attenuatus

Population genetics models

# Estimating allele frequencies in non-model polyploids using high throughput sequencing data

Paul Blischak<sup>1</sup>, Laura Kubatko<sup>1,2</sup>, Andrea Wolfe<sup>1</sup>

<sup>1</sup>Dept. of EEOB <sup>2</sup>Dept. of Statistics The Ohio State University

June 10, 2015

#### Outline

Allele frequencies in polyploids

Botany 2019

introgression
Patterson's
D-statistic

Penstemon attenuatus

attenuatus

Population genetics models

- 1 Tests for introgression
  - Patterson's D-statistic
  - $\blacksquare f$  estimators

2 Penstemon attenuatus

3 Population genetics models

#### ABBA-BABA statistics

Allele frequencies in polyploids

Botany 201!

Tests for introgression

Patterson's D-statistic

f estimator

Penstemon attenuatus

genetics models

$$\mathbf{D} = \frac{\sum \mathcal{C}_{ABBA} - \mathcal{C}_{BABA}}{\sum \mathcal{C}_{ABBA} + \mathcal{C}_{BABA}}$$

Tests for introgression

Patterson's

f estimators

Penstemon attenuatus

genetics models

$$f = \frac{\sum \mathcal{C}_{ABBA} - \mathcal{C}_{BABA}}{\sum \mathcal{C}_{ABBA} + \mathcal{C}_{BABA}}$$

#### Penstemon attenuatus

Allele frequencies in polyploids

Botany 2019

Tests for introgression Patterson's D-statistic f estimator.

Penstemon attenuatus

Population

The Penstemon attenuatus Doug. ex Lindl. species complex . . .

## Posterior distribution of allele frequencies

Allele frequencies in polyploids

Botany 2015

Tests for introgression Patterson's D-statistic f estimators

attenuatus

Population genetics models

$$P(p_l, g_{li}|R_{li}^b, \epsilon) \propto \prod_{l} \prod_{i} P(R_{li}^b|g_{li}, \epsilon) P(g_{li}|p_l) P(p_l). \tag{1}$$

### Notation

Allele frequencies in polyploids

Botany 201!

ntrogression
Patterson's
D-statistic
f estimators

attenuatus

Population genetics models

Symbol	Description
L	The number of loci.
l	Index for loci $(l \in \{1, \dots, L\})$ .
$N_k$	The number of individuals sampled from population k.
k	Index for populations $(k \in \{P_1, P_{poly}, P_2, O\}).$
i	Index for individuals in a population $k$ ( $i \in \{1, \dots, N_k\}$ ).
$N_{lk}$	The number of individuals sampled at locus $\boldsymbol{l}$ in population $\boldsymbol{k}.$
$N_{lk}^a$	The number of individuals homozygous for A at locus $\boldsymbol{l}$ in population $\boldsymbol{k}.$
$N_{lk}^b$	The number of individuals homozygous for B at locus $\boldsymbol{l}$ in population $k.$
$N_{lk}^{ab}$	The number of heterozygous individuals at locus $\boldsymbol{l}$ in population $\boldsymbol{k}.$
$\hat{p}_{lk}$	Frequency of the derived allele (B) at locus $l$ in population $k$ .
$P_k$	The ploidy of individuals in population $k$ .
$R_{li}$	The number of reads for individual $i$ at locus $l$ .
$R_{li}^a$	The number of reads with allele A for individual $i$ at locus $l$ .
$R_{li}^b$	The number of reads with allele B for individual $i$ at locus $l$ .
$r_{li}^a$	Proportion of reads with allele A $(R_{li}^a/R_{li})$ .
$r_{li}^b$	Proportion of reads with allele A $(R_{li}^b/R_{li})$ .