Exploring conservation genomic data in R

Josh Jahner
jpjahner@gmail.com

UNIVERSITY OF WYOMING

modelscape

NEW MEXICO TECH
SCIENCE • ENGINEERING • RESEARCH UNIVERSITY

Outline of today's topics

1) Big questions in conservation genetics
2) Types of conservation genetic data
3) Simulating data with the beta distribution
4) Exploring conservation genetic analyses

Genetic diversity and inbreeding

FROM THE COVER

MOLECULAR ECOLOGY RESOURCES WILEY

Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations

Samuel A. Speak[1,2,3] | Thomas Birley[1] | Chiara Bortoluzzi[4,5] | Matthew D. Clark[2,6] | Lawrence Percival-Alwyn[7] | Hernán E. Morales[8] | Cock van Oosterhout[1]

Identifying adaptive genetic variation

Received: 30 January 2024 | Revised: 23 April 2024 | Accepted: 30 April 2024

DOI: 10.1111/mec.17383

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Natural selection on feralization genes contributed to the invasive spread of wild pigs throughout the United States

Niek W. G. Barmentlo[1,2] | Patrick G. Meirmans[2] | William H. Stiver[3] | Joseph G. Yarkovich[3] | Blake E. McCann[4] | Antoinette J. Piaggio[5] | Dominic Wright[6] | Timothy J. Smyser[5] | Mirte Bosse[1,7]

Population structure and connectivity

Molecular Ecology

WILEY

MOLECULAR ECOLOGY

ORIGINAL ARTICLE OPEN ACCESS

Shared Dispersal Patterns but Contrasting Levels of Gene Flow in Two Anadromous Salmonids Along a Broad Subarctic Coastal Gradient

Raphaël Bouchard[1,2] | Charles Babin[1,2] | Eric Normandeau[3] | Amanda Xuereb[1,2] | Félix Boulanger[4] | Angela Coxon[4] | Sanford Diamond[5] | Robert Fireman[5] | John Lameboy[5] | Natasha Louttit[4,5] | George Natawapineskum[5] | Derek Okimaw[5] | Dante Torio[6] | Stephanie Varty[5] | Jean-Sébastien Moore[1,2] | Dylan Fraser[7] | Louis Bernatchez[1,2]

Population size estimation

Evolutionary Applications

WILEY

Evolutionary Applications

ORIGINAL ARTICLE OPEN ACCESS

Estimating the Effective Size of European Wolf Populations

Joachim Mergeay[1,2] | Sander Smet[2] | Sebastian Collet[3] | Sabina Nowak[4] | Ilka Reinhardt[5] | Gesa Kluth[5] | Maciej Szewczyk[6] | Raquel Godinho[7] | Carsten Nowak[3] | Robert W. Mysłajek[4] | Gregor Rolshausen[3]

Demographic history reconstruction

Whole Genomes Inform Genetic Rescue Strategy for Montane Red Foxes in North America

Cate B. Quinn [1,2,3,*] Sophie Preckler-Quisquater,[1] Michael R. Buchalski [2] Benjamin N. Sacks [1,4]

Species delimitation

CellPress
OPEN ACCESS

Current Biology

Report

Comparative species delimitation of a biological conservation icon

Ava Ghezelayagh,[1,2,13,14,*] Jeffrey W. Simmons,[3] Julia E. Wood,[1] Tsunemi Yamashita,[4] Matthew R. Thomas,[5] Rebecca E. Blanton,[6] Oliver D. Orr,[1] Daniel J. MacGuigan,[7] Daemin Kim,[1] Edgar Benavides,[8,9] Benjamin P. Keck,[10] Richard C. Harrington,[1,11] and Thomas J. Near[1,12]

Hybridization and admixture

OXFORD

Evolution Letters, 2025, **9**(1), 1–12

https://doi.org/10.1093/evlett/qrae057
Advance access publication 19 October 2024
Letter

Genetic structure and common ancestry expose the dingo-dog hybrid myth

Andrew R. Weeks[1,2], Peter Kriesner[1], Nenad Bartonicek[1,†], Anthony van Rooyen[1], Kylie M. Cairns[3], Collin W. Ahrens[1]

Assisted gene flow, genetic rescue, translocations

Evolutionary Applications

WILEY

Evolutionary Applications

ORIGINAL ARTICLE OPEN ACCESS

Genetic Rescue of the Dinaric Lynx Population: Insights for Conservation From Genetic Monitoring and Individual-Based Modelling

Elena Pazhenkova[1,2] | Matej Bartol[3] | Barbara Boljte[1,2] | Urša Fležar[1] | Andrea Gazzola[4] | Tomislav Gomerčić[5] | Marjeta Konec[1,2] | Ivan Kos[1] | Miha Krofel[1] | Jakub Kubala[6] | Ladislav Paule[6] | Mihai Pop[4] | Hubert Potočnik[1] | Barbara Promberger[7] | Robin Rigg[1,8] | Teodora Sin[4] | Magda Sindičić[5] | Vedran Slijepčević[9] | Astrid Vik Stronen[1,2,10] | Ira Topličanec[5] | Tomaž Skrbinšek[1,2]

Genetic monitoring

Received: 26 September 2023 | Revised: 1 December 2023 | Accepted: 19 January 2024

DOI: 10.1002/ece3.10934

Ecology and Evolution Open Access WILEY

RESEARCH ARTICLE

Evaluating genotyping-in-thousands by sequencing as a genetic monitoring tool for a climate sentinel mammal using non-invasive and archival samples

Kate E. Arpin[1] | Danielle A. Schmidt[1] | Bryson M. F. Sjodin[1] | Anthony L. Einfeldt[2] | Kurt Galbreath[3] | Michael A. Russello[1]

Outline of today's topics

1) Big questions in conservation genetics
2) Types of conservation genetic data
3) Simulating data with the beta distribution
4) Exploring conservation genetic analyses

Mitochondrial DNA (Sanger sequencing)

Individual 1

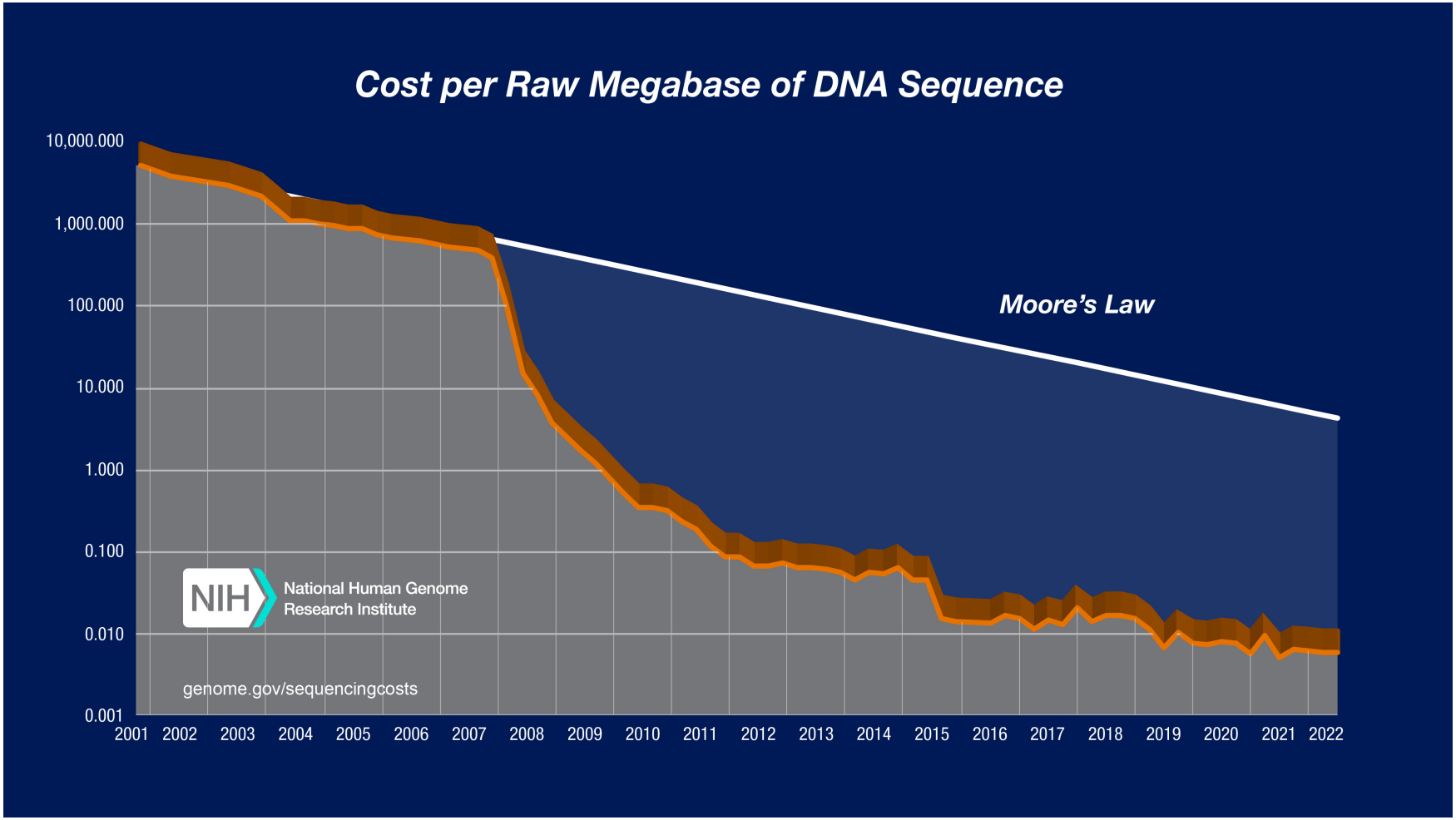
ACTTTGAGCTTTCTGA ...
ACTCTGAGCTTTCTGA ...

Individual 2

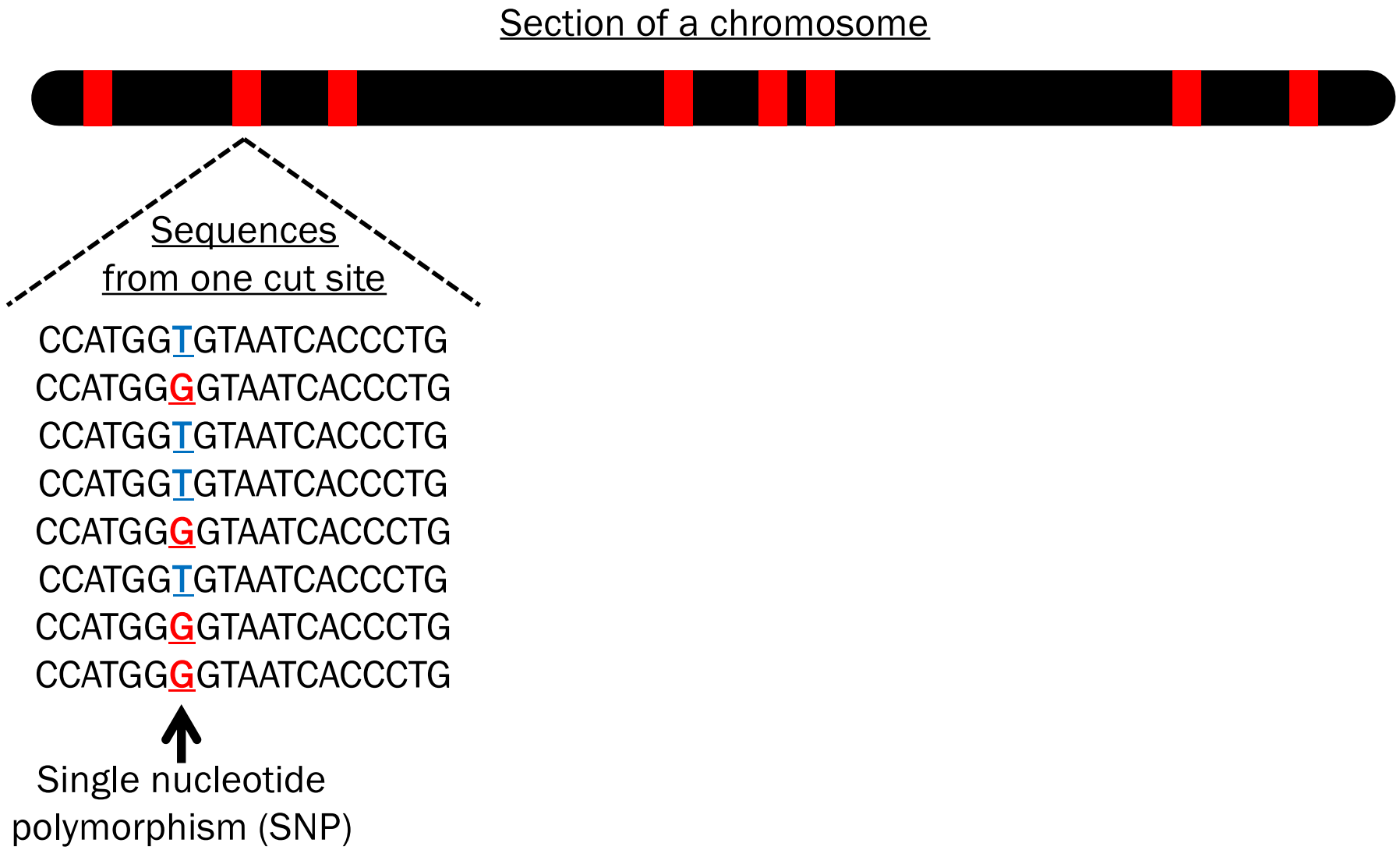
ACTCTGGGCTATCTGA ...
ACTCTGGGCTATCTGA ...

Microsatellites
(aka: simple sequence repeats [SSRs])

Individual 1	AGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	10 repeats
	AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	12 repeats
Individual 2	AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	14 repeats
	AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	14 repeats


Cost per Raw Megabase of DNA Sequence
10,000.000
1,000.000
100.000
10.000
1.000
0.100
0.010
0.001
Moore's Law
NIH
National Human Genome
Research Institute
genome.gov/sequencingcosts
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

Reduced representation sequencing



Reduced representation sequencing

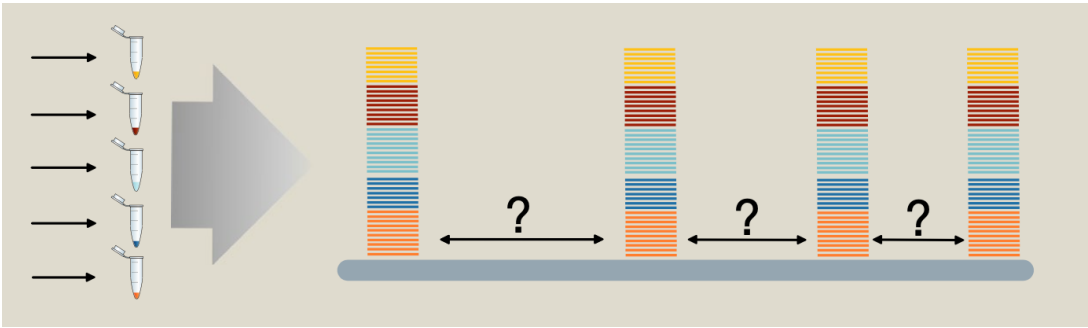
Section of a chromosome

Sequences
from one cut site

CCATGGTGTAATCACCCTG
CCATGGGGTAATCACCCTG
CCATGGTGTAATCACCCTG
CCATGGTGTAATCACCCTG
CCATGGGGTAATCACCCTG
CCATGGTGTAATCACCCTG
CCATGGGGTAATCACCCTG
CCATGGGGTAATCACCCTG

Single nucleotide
polymorphism (SNP)

RADseq approaches



Lou et al. 2021 *Mol Ecol*

Reduced representation sequencing

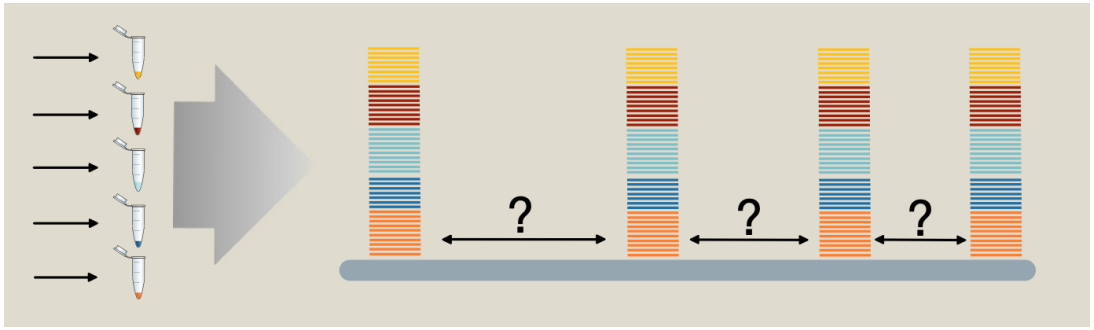
Section of a chromosome

Sequences
from one cut site

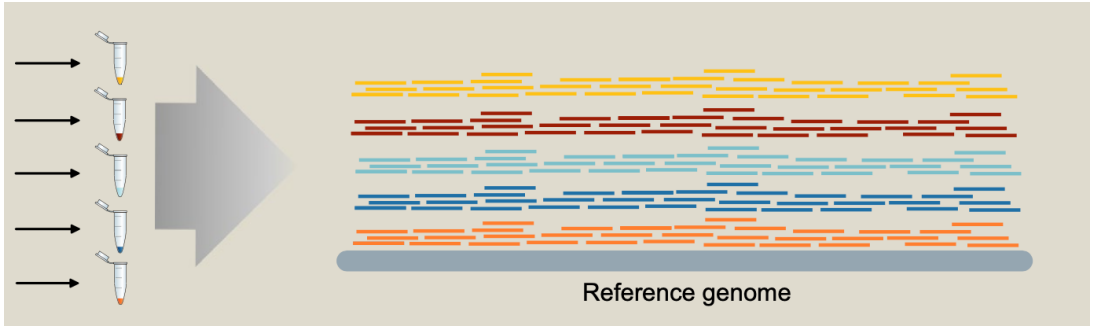
CCATGGTGTAATCACCCTG
CCATGGGGTAATCACCCTG
CCATGGTGTAATCACCCTG
CCATGGTGTAATCACCCTG
CCATGGGGTAATCACCCTG
CCATGGTGTAATCACCCTG
CCATGGGGTAATCACCCTG
CCATGGGGTAATCACCCTG

Single nucleotide
polymorphism (SNP)

RADseq approaches



Whole genome resequencing



Lou et al. 2021 *Mol Ecol*

Outline of today's topics

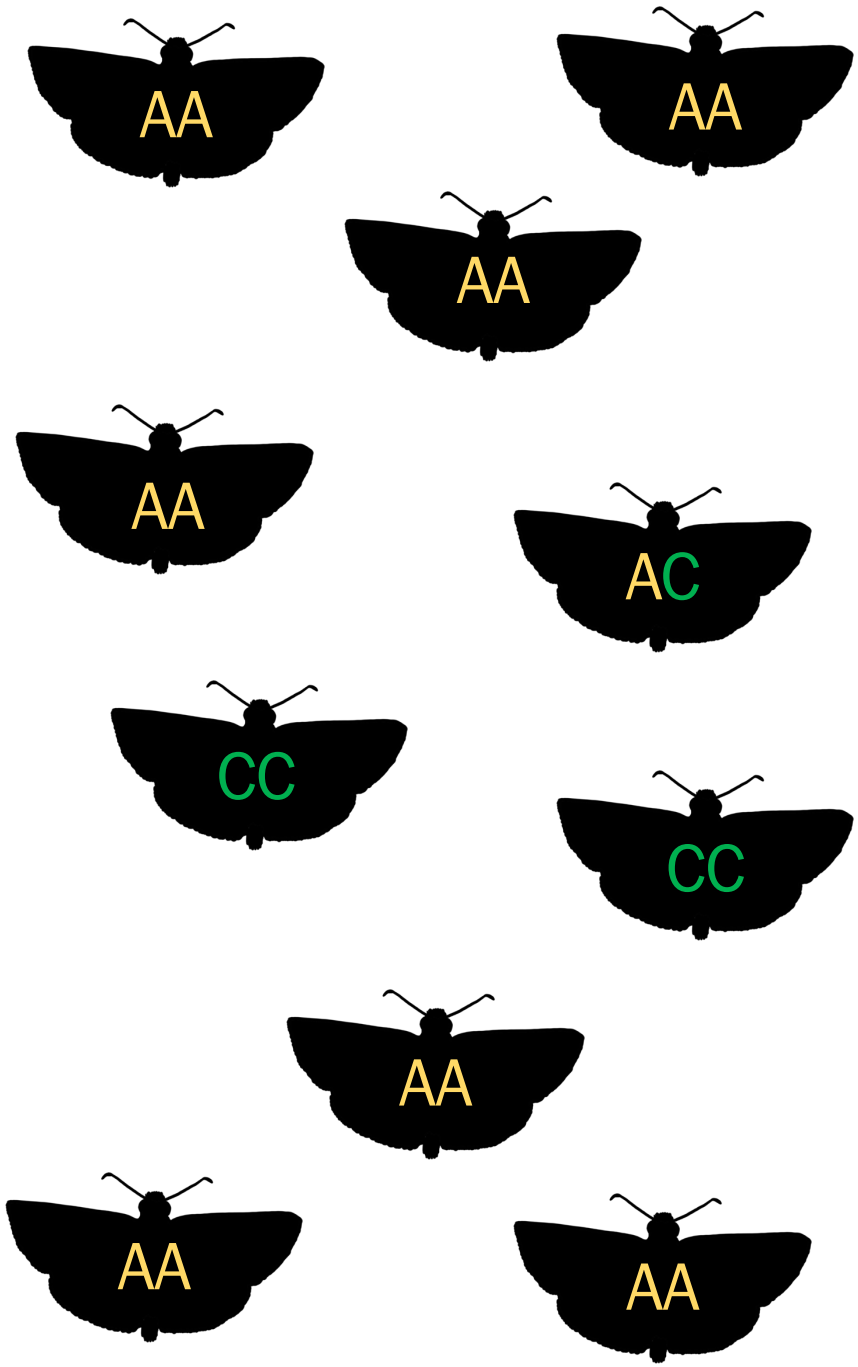
1) Big questions in conservation genetics
2) Types of conservation genetic data
3) Simulating data with the beta distribution
4) Exploring conservation genetic analyses

Find the R scripts at:

https://github.com/jpjahner/congen_workshop

01_beta_simulate.R
02_explore_analyses.R

Allele frequency of A:
15 / 20 = 0.75

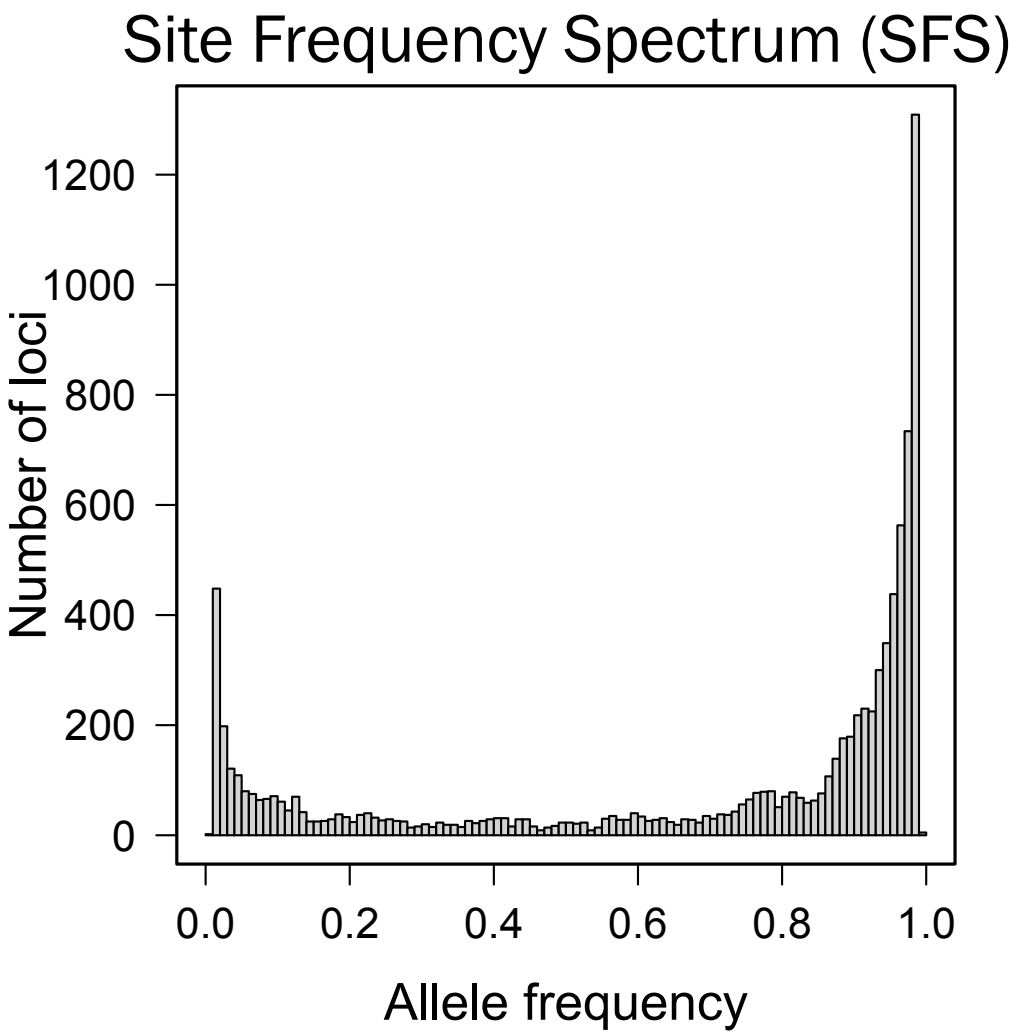


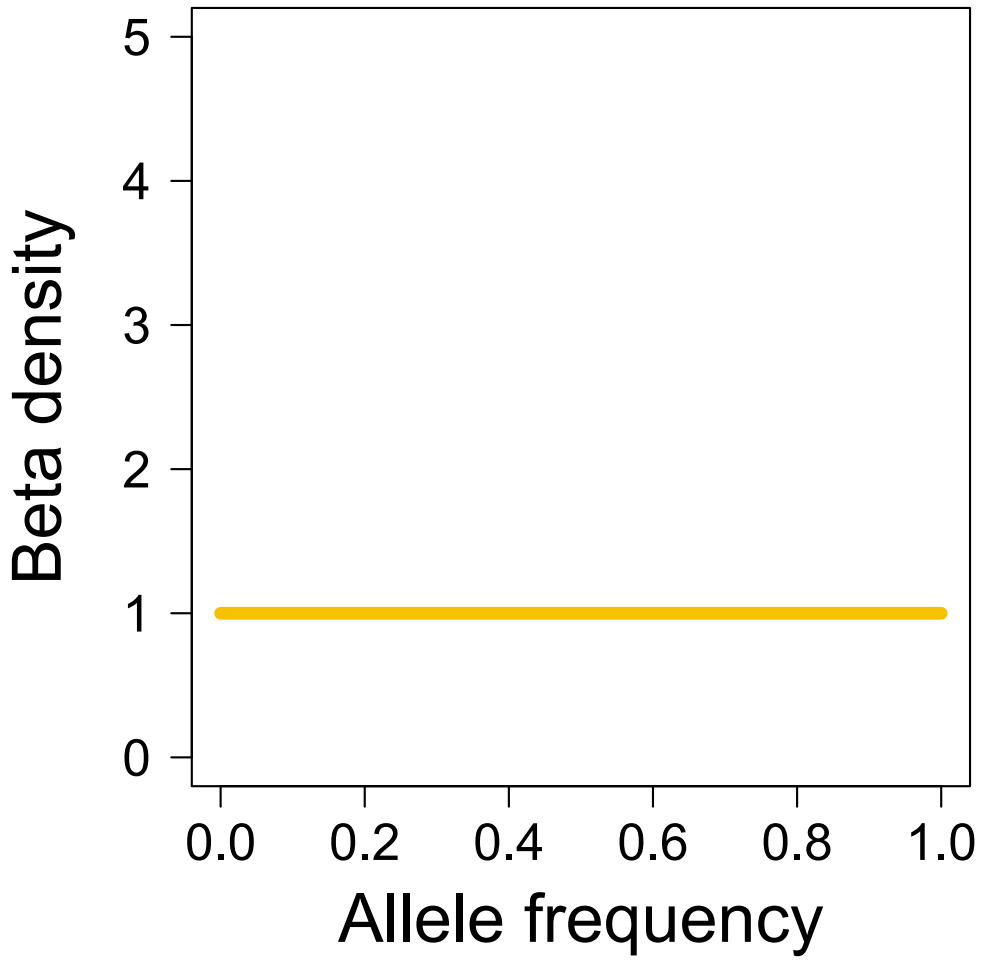
RESEARCH ARTICLE

Ecology and Evolution Open Access WILEY

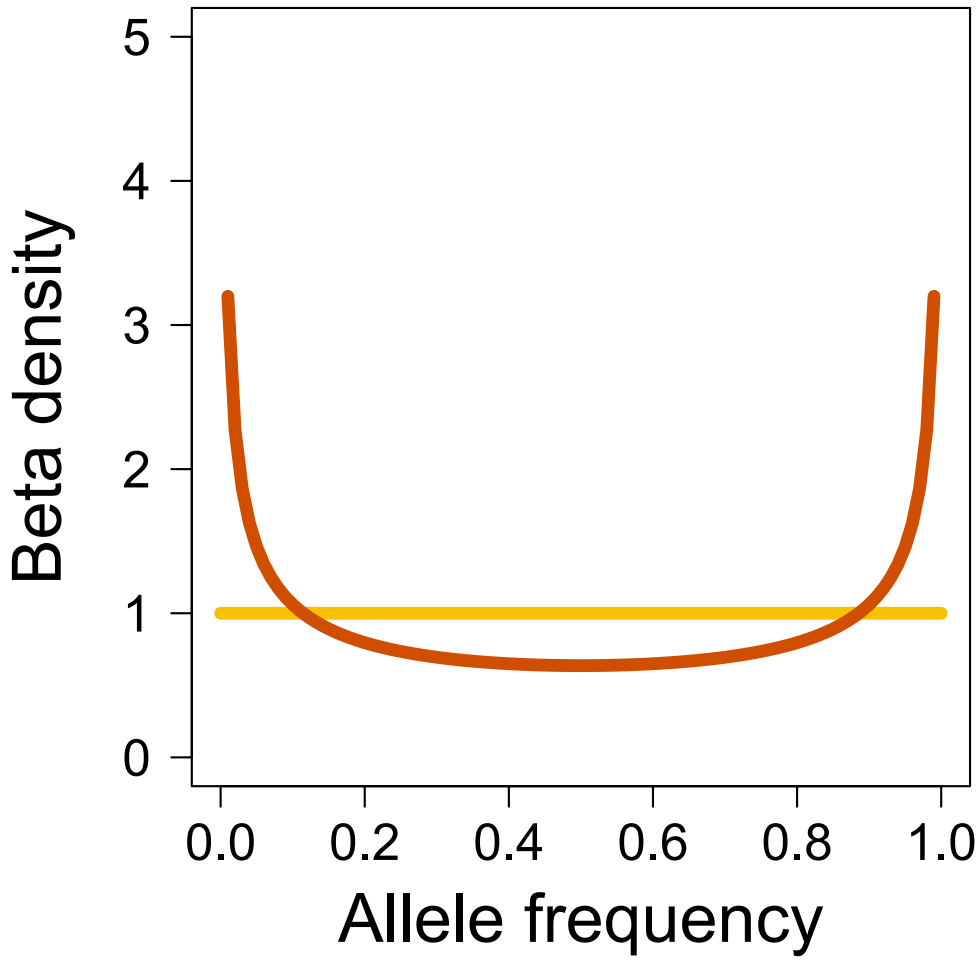
Influence of dams on sauger population structure and hybridization with introduced walleye

William C. Rosenthal[1,2] | Elizabeth G. Mandeville[1,3] | Ashleigh M. Pilkerton[2,4,5] | Paul C. Gerrity[6] | Joseph A. Skorupski[6] | Annika W. Walters[2,4,5,7] | Catherine E. Wagner[1,2]



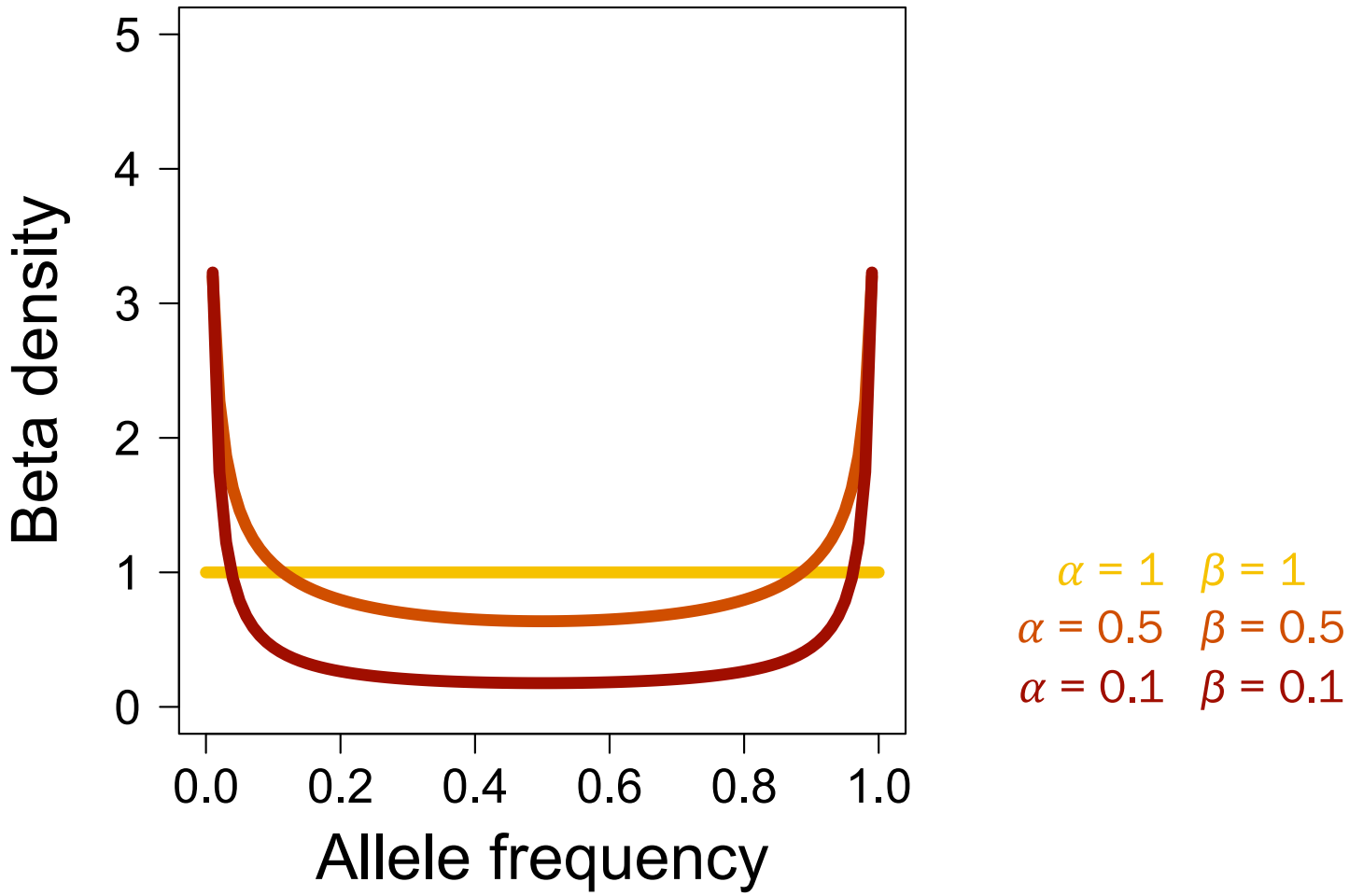


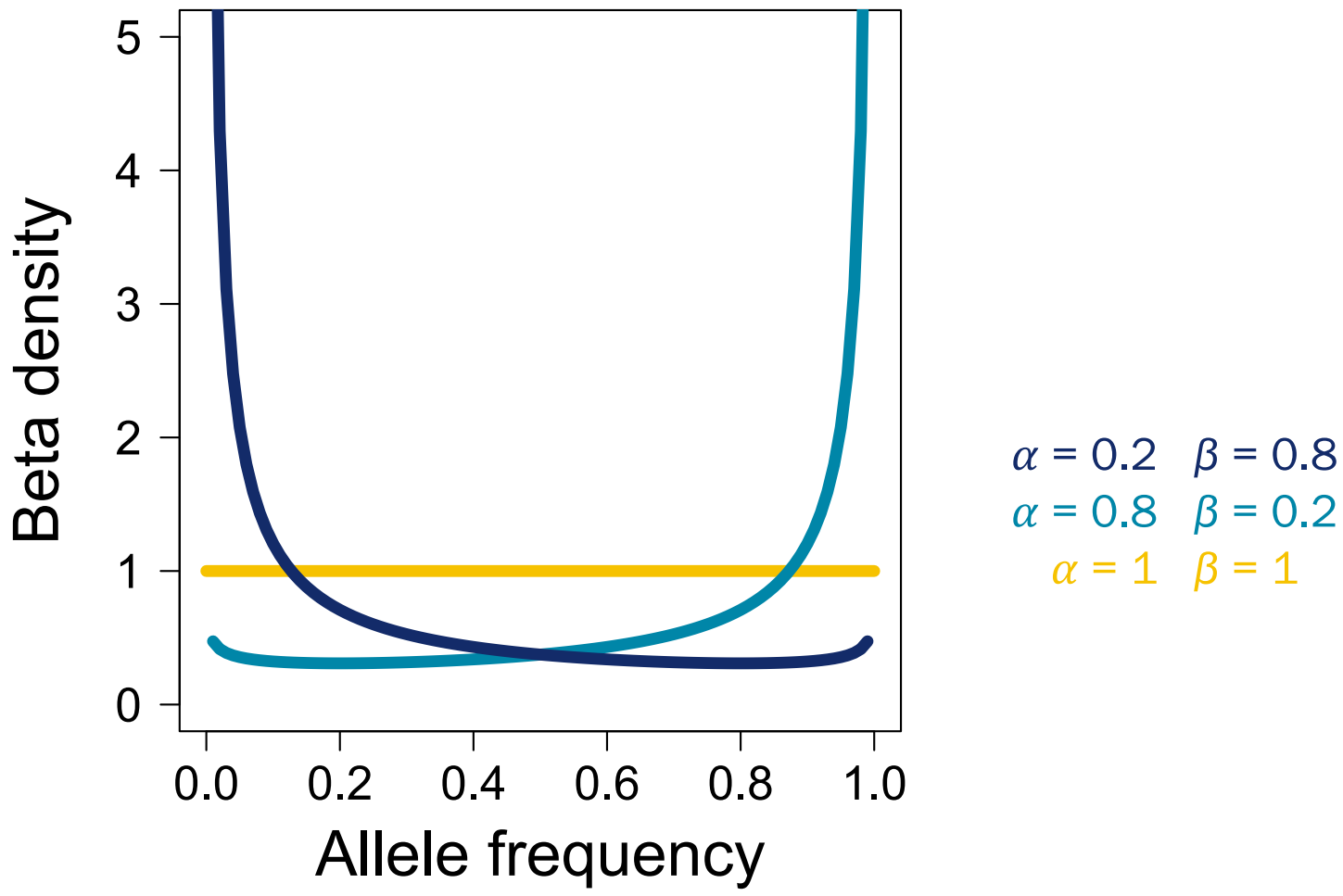
$\alpha = 1 \quad \beta = 1$

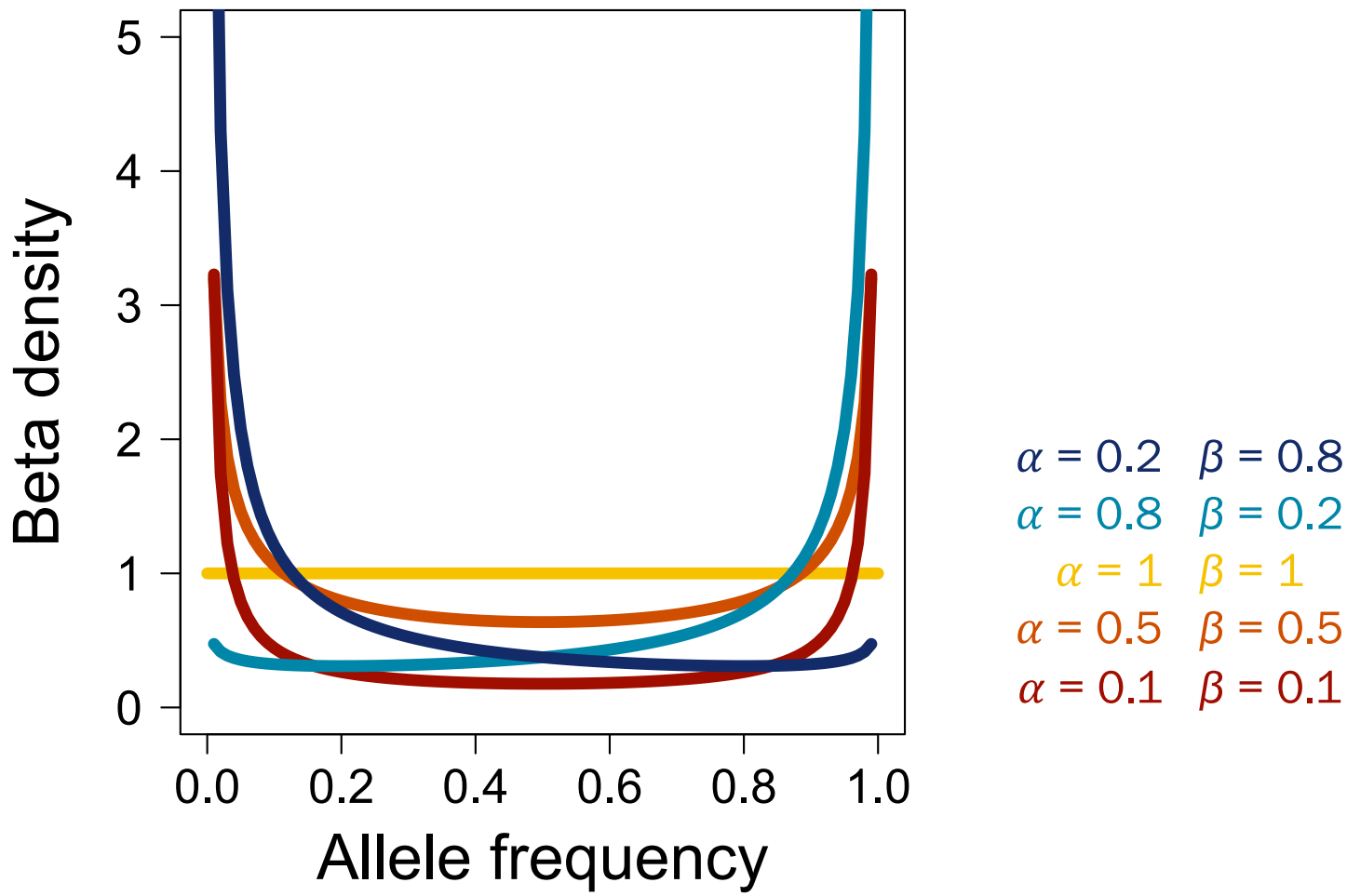


$\alpha = 1 \quad \beta = 1$

$\alpha = 0.5 \quad \beta = 0.5$


Beta density
5
4
3
2
1
0
0.0
0.2
0.4
0.6
0.8
1.0
Allele frequency
α = 1 β = 1
α = 0.5 β = 0.5
α = 0.1 β = 0.1


Beta density
5
4
3
2
1
0
0.0
0.2
0.4
0.6
0.8
1.0
Allele frequency
α = 0.2 β = 0.8
α = 0.8 β = 0.2
α = 1 β = 1

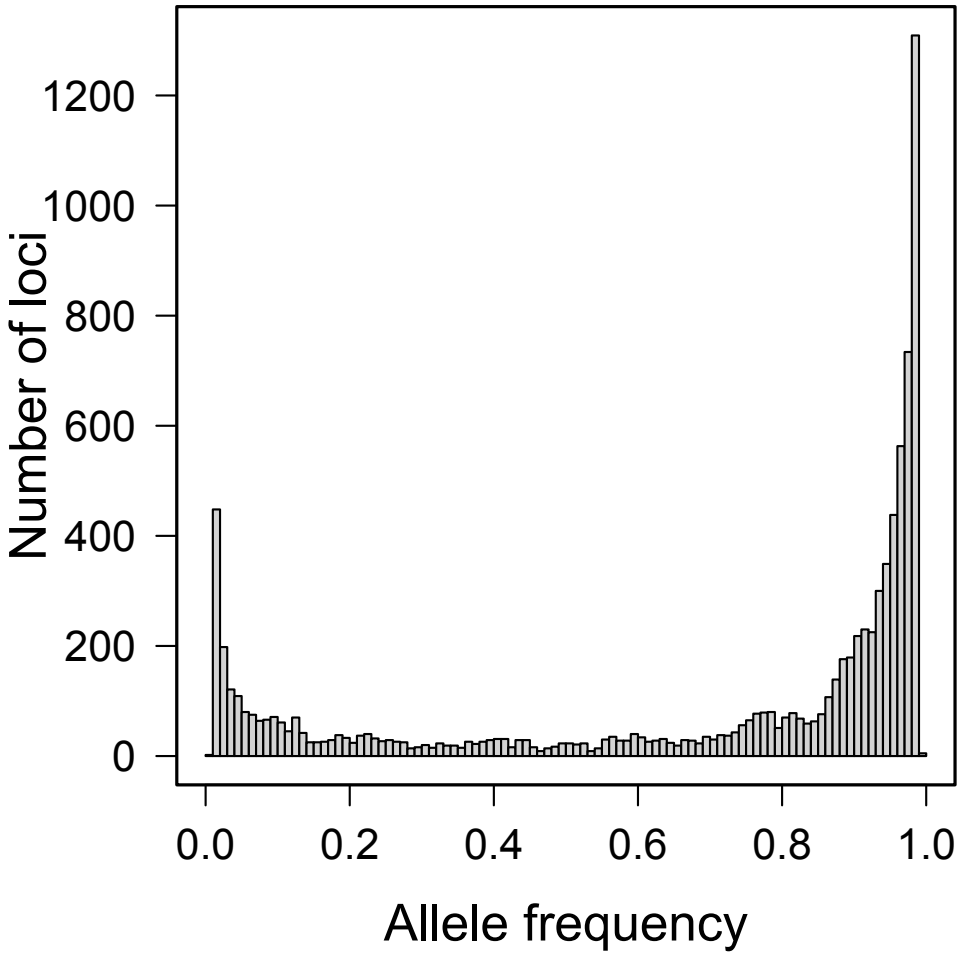

Beta density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Allele frequency
α = 0.2 β = 0.8
α = 0.8 β = 0.2
α = 1 β = 1
α = 0.5 β = 0.5
α = 0.1 β = 0.1

RESEARCH ARTICLE

Ecology and Evolution Open Access WILEY

Influence of dams on sauger population structure and hybridization with introduced walleye

William C. Rosenthal[1,2] | Elizabeth G. Mandeville[1,3] | Ashleigh M. Pilkerton[2,4,5] | Paul C. Gerrity[6] | Joseph A. Skorupski[6] | Annika W. Walters[2,4,5,7] | Catherine E. Wagner[1,2]



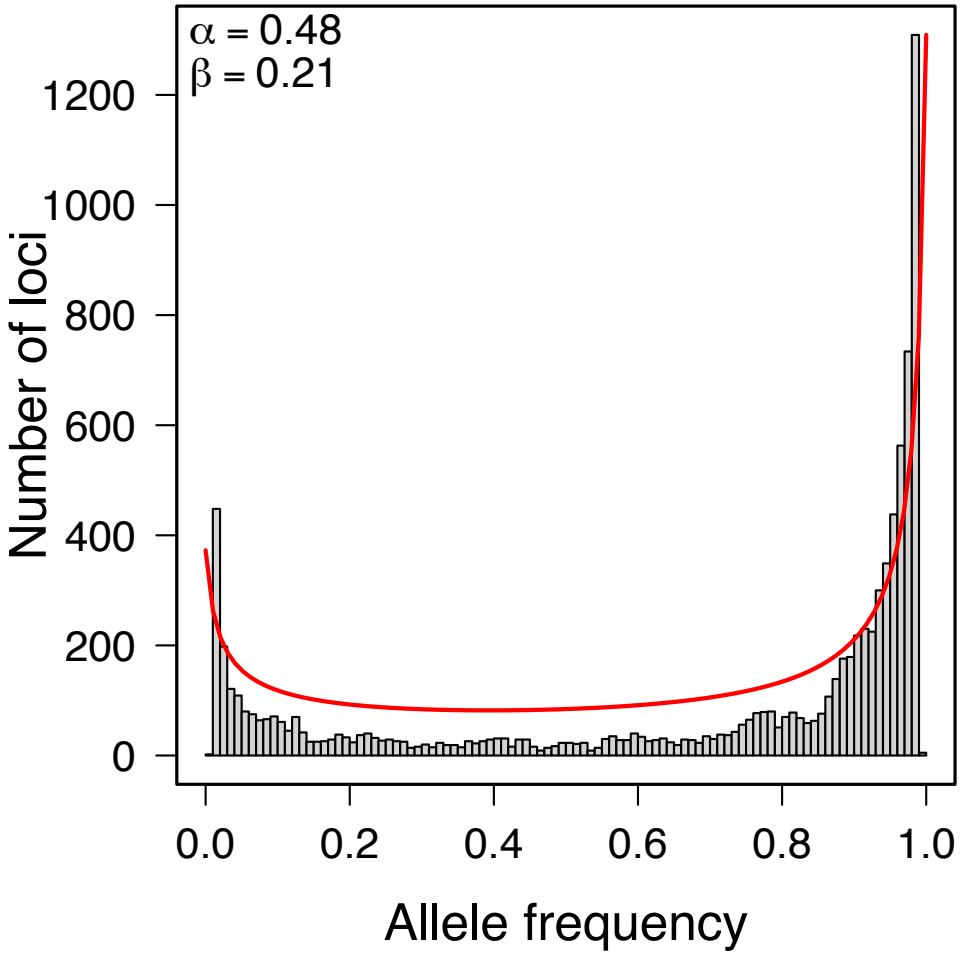

```
dbeta(xvals, shape1=NA, shape2=NA)
```

RESEARCH ARTICLE

Ecology and Evolution Open Access WILEY

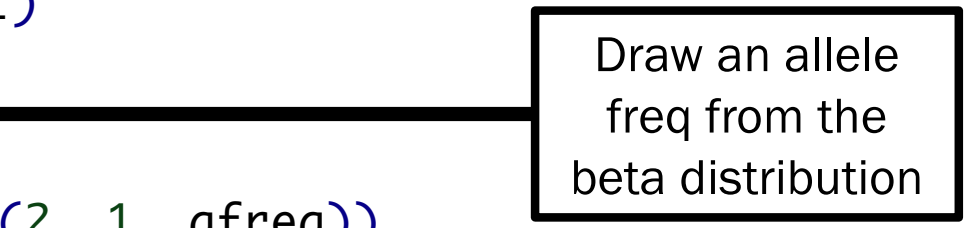
Influence of dams on sauger population structure and hybridization with introduced walleye

William C. Rosenthal[1,2] | Elizabeth G. Mandeville[1,3] | Ashleigh M. Pilkerton[2,4,5] | Paul C. Gerrity[6] | Joseph A. Skorupski[6] | Annika W. Walters[2,4,5,7] | Catherine E. Wagner[1,2]

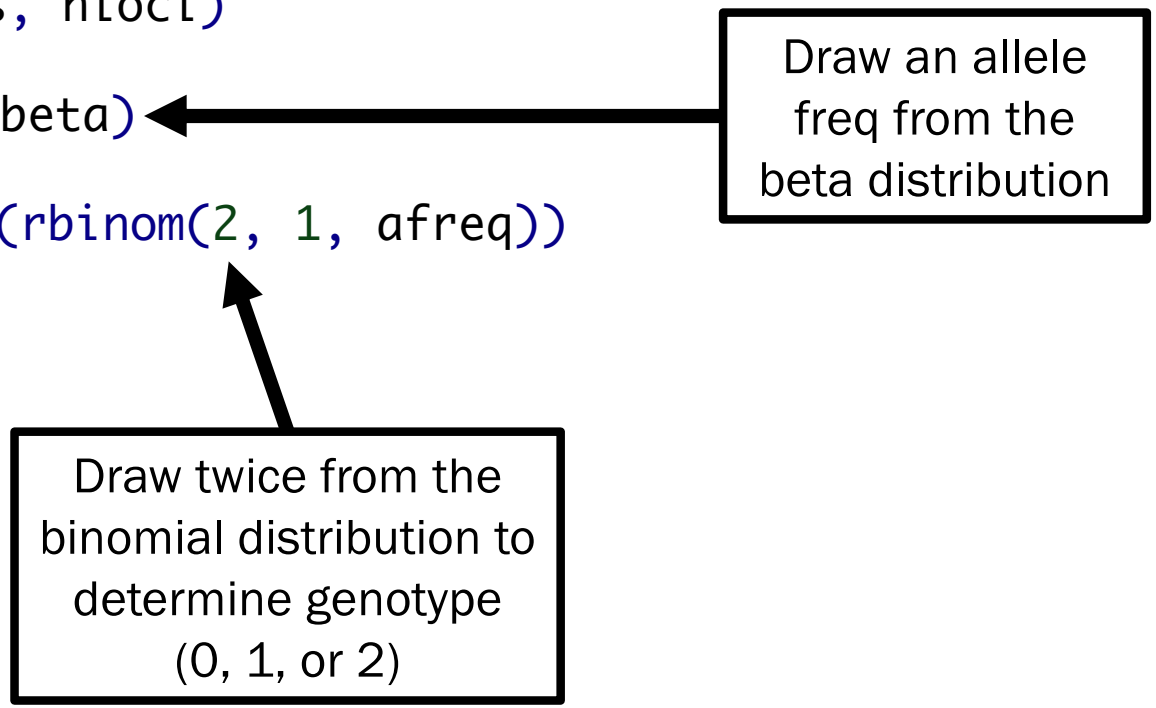


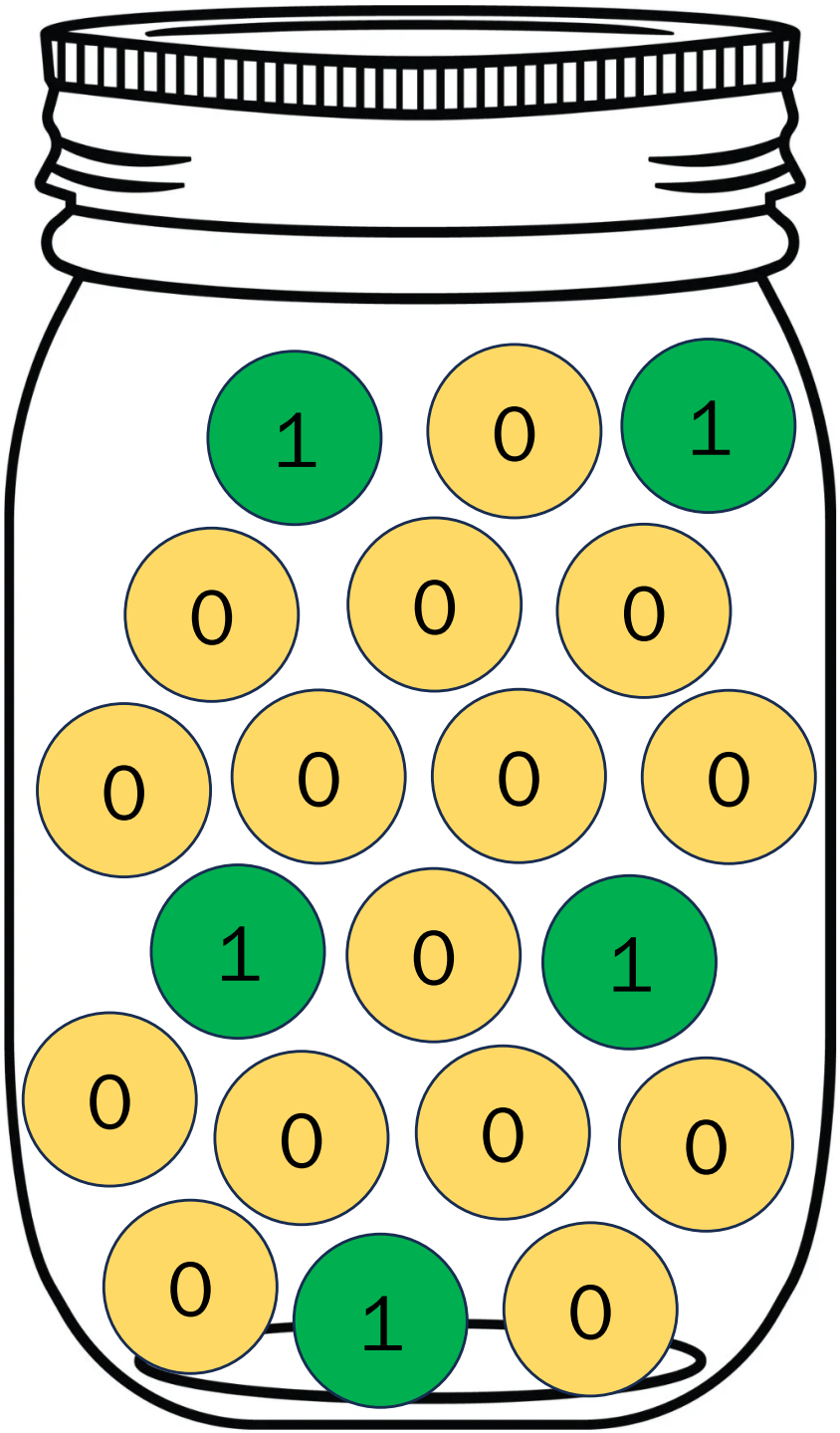
```r
## Function for simulating genetic data from the beta distribution


single_pop_sim <- function(nloci=NA, ninds=NA, alpha=NA, beta=NA){
genotypes <- matrix(NA, ninds, nloci)
for (i in 1:nloci){
    afreq <- rbeta(1, alpha, beta)
        for (j in 1:ninds){
        genotypes[j,i] <- sum(rbinom(2, 1, afreq))
        }
    }
    return(genotypes)
}
```

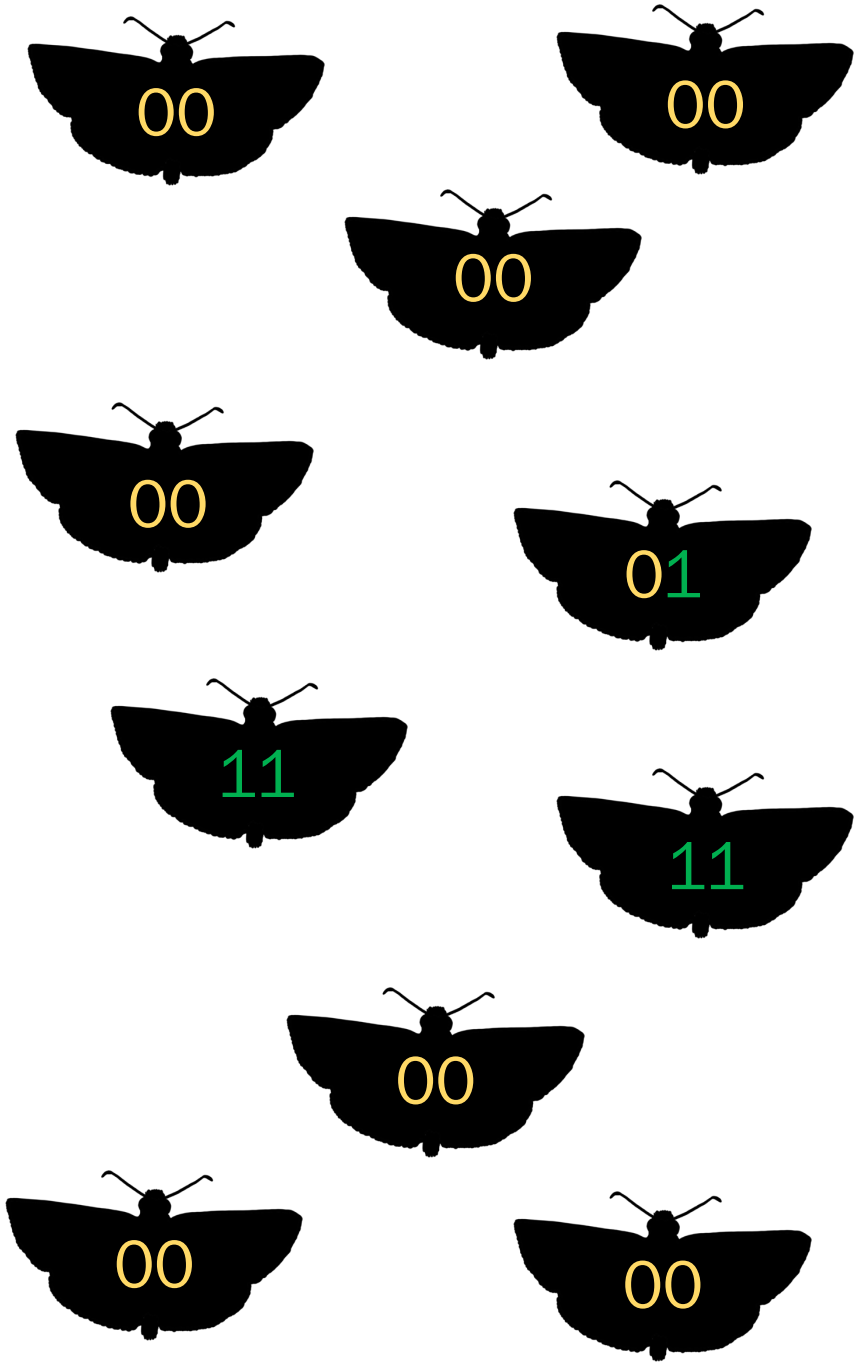
```
## Function for simulating genetic data from the beta distribution


single_pop_sim <- function(nloci=NA, ninds=NA, alpha=NA, beta=NA){
genotypes <- matrix(NA, ninds, nloci)
for (i in 1:nloci){
    afreq <- rbeta(1, alpha, beta)
        for (j in 1:ninds){
        genotypes[j,i] <- sum(rbinom(2, 1, afreq))
        }
    }
    return(genotypes)
}
```



```
## Function for simulating genetic data from the beta distribution


single_pop_sim <- function(nloci=NA, ninds=NA, alpha=NA, beta=NA){
genotypes <- matrix(NA, ninds, nloci)
for (i in 1:nloci){
    afreq <- rbeta(1, alpha, beta)
        for (j in 1:ninds){
        genotypes[j,i] <- sum(rbinom(2, 1, afreq))
        }
    }
    return(genotypes)
}
```




1 0 1
0 0 0
0 0 0 0
1 0 1
0 0 0 0
0 1 0


00
00
00
00
01
11
11
00
00
00

Each column is a genetic marker

Each row is an individual

2	2	2	1	1	2	2	2	2	2
2	2	2	2	2	2	1	2	1	2
1	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
0	2	2	1	2	2	1	2	1	2
0	2	2	1	2	2	2	2	0	2
1	2	2	2	2	2	1	2	1	2
1	2	2	1	2	2	1	2	1	2
0	2	2	2	1	2	2	2	2	2
0	2	2	2	2	2	2	2	2	2

Outline of today's topics

1) Big questions in conservation genetics
2) Types of conservation genetic data
3) Simulating data with the beta distribution
4) Exploring conservation genetic analyses

Goals for the rest of our time:

1. Simulate genetic data using the beta distribution
2. Explore the data with three different genetic analyses

Goals for the rest of our time:

1. Simulate genetic data using the beta distribution
2. Explore the data with three different genetic analyses

How are genetic analyses
affected by the shape of the
site frequency spectrum?

How many genetic markers
do I need to detect a signal?