Developing computational tools to enable genomic analyses in polyploids

Paul Blischak
29 October 2020

My background

I am trained in mathematics, statistics, and biology.

My background

I am trained in mathematics, statistics, and biology.

I use theory and computation to answer biological questions.

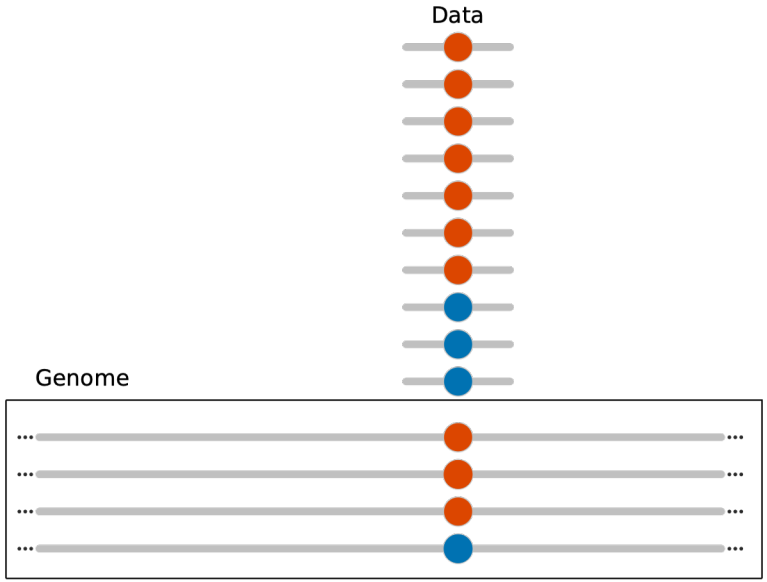
My background

I am trained in mathematics, statistics, and biology.

I use theory and computation to answer biological questions.

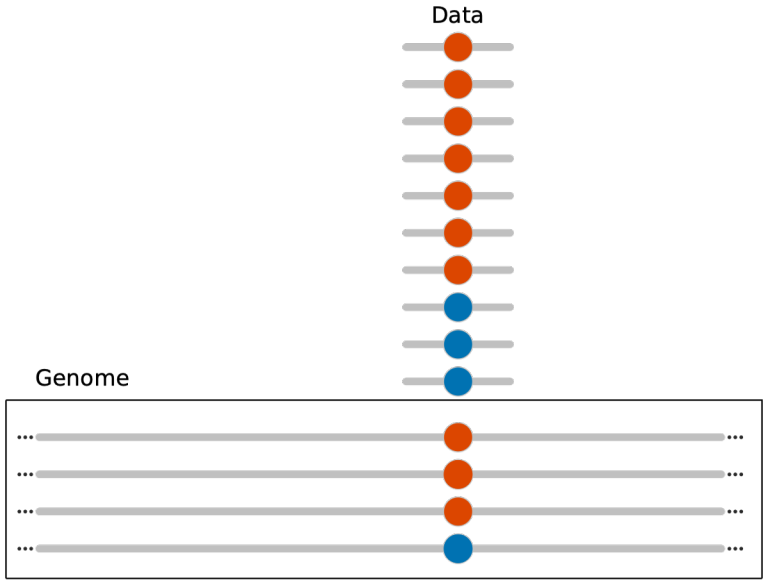
I develop and distribute software for members of the research community.

Road map



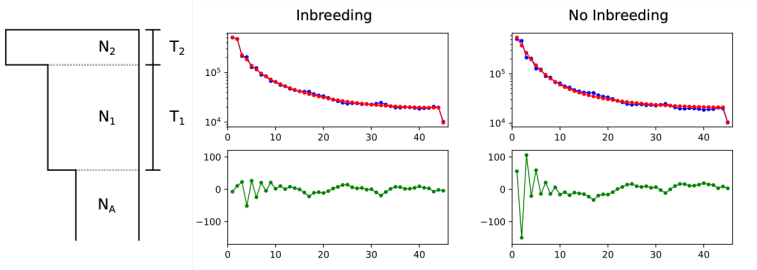
Genotyping and parameter estimation in polyploids

Road map

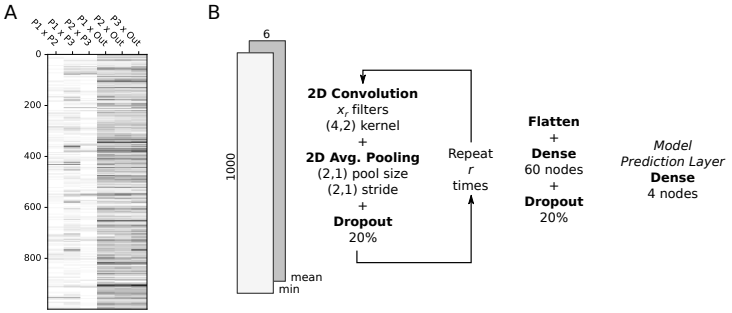


Genotyping and parameter estimation in polyploids

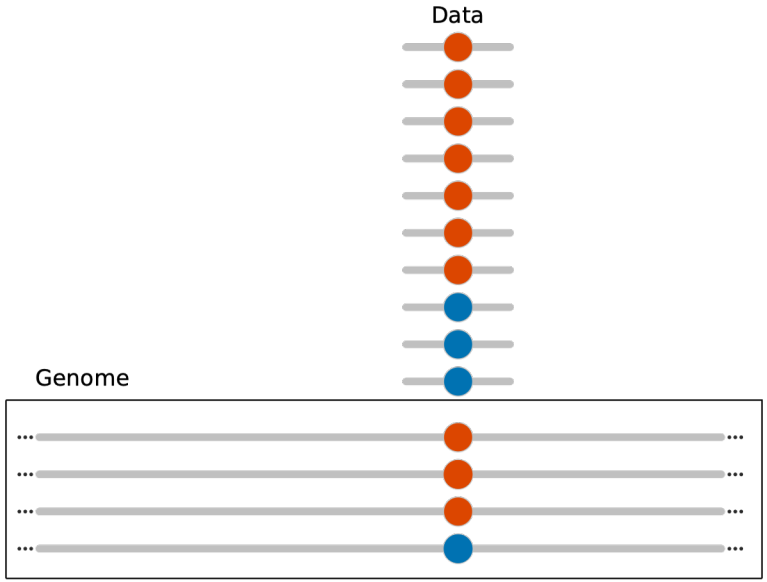
Demographic inference in inbred species



Hybridization detection with convolutional neural networks



Road map



Genotyping and parameter estimation in polyploids

Demographic inference in inbred species

Hybridization detection with convolutional neural networks

What is polyploidy?

What is polyploidy?

Having more than two sets of chromosomes

What is polyploidy?

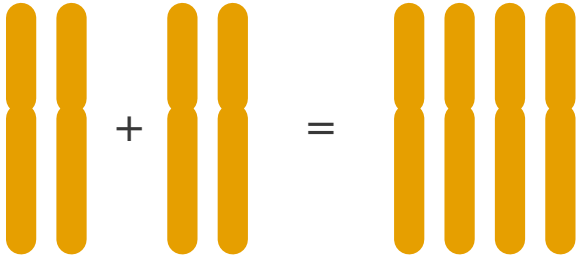
Having more than two sets of chromosomes

Whole genome duplication

Types of polyploids

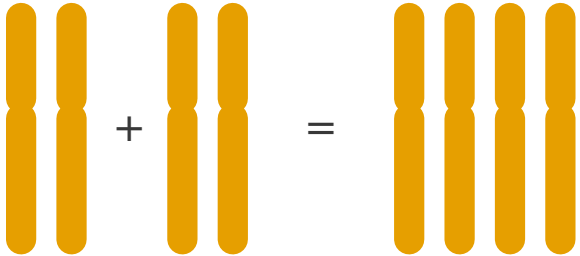
Types of polyploids

Autopolyploid:

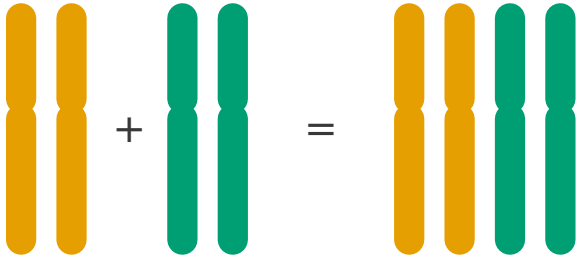


Types of polyploids

Autopolyploid:



Allopolyploid:



Why polyploidy?

Penstemon

Photo credit: Andi Wolfe

Penstemon

Photo credit: Andi Wolfe

Why polyploidy?

nature plants

LETTERS

PUBLISHED: 1 AUGUST 2016 | ARTICLE NUMBER: 16115 | DOI: 10.1038/NPLANTS.2016.115

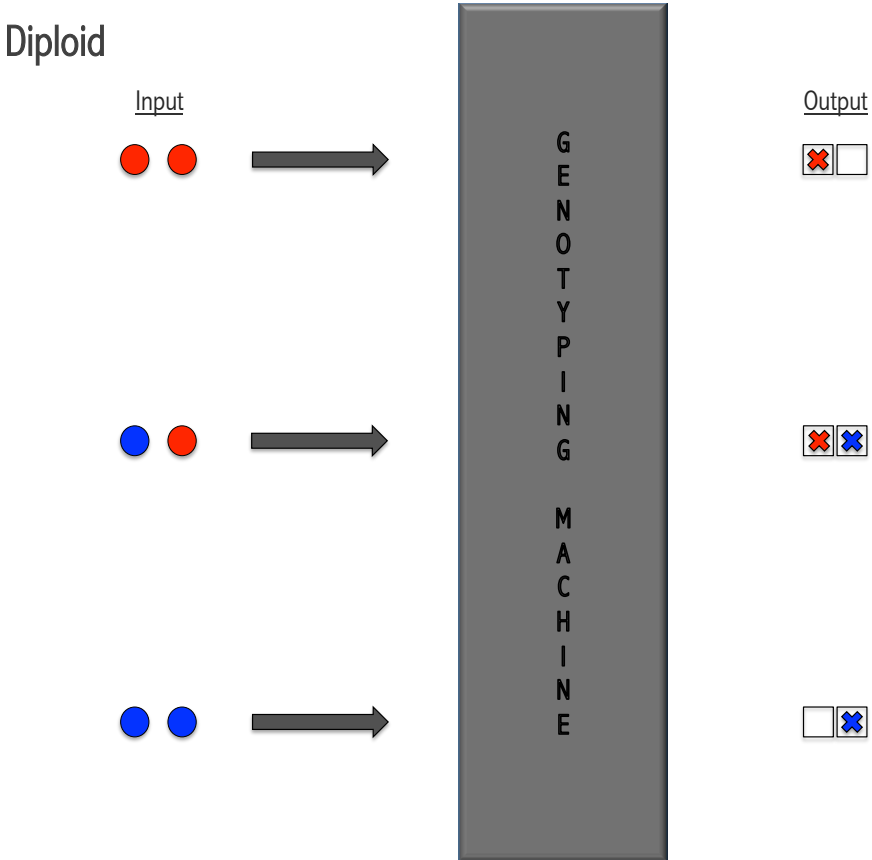
Whole-genome duplication as a key factor in crop domestication

Ayelet Salman-Minkov†, Niv Sabath† and Itay Mayrose*

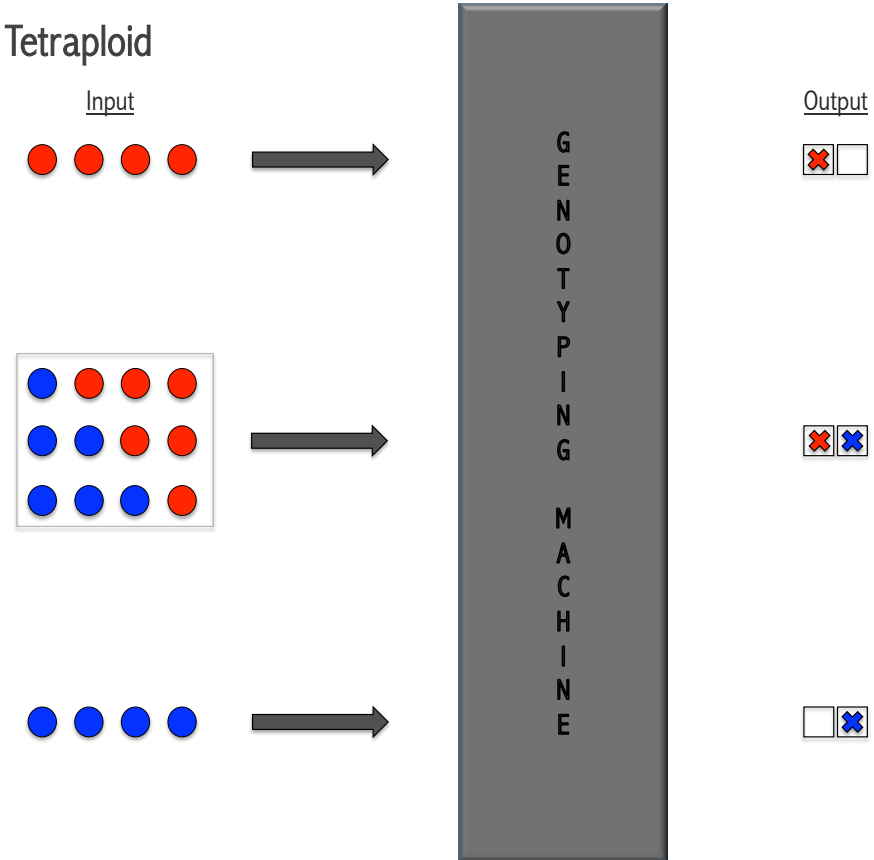
Photo credits: Wikipedia; Wikimedia Commons

The challenge with polyploid genomics

The challenge with polyploid genomics



The challenge with polyploid genomics



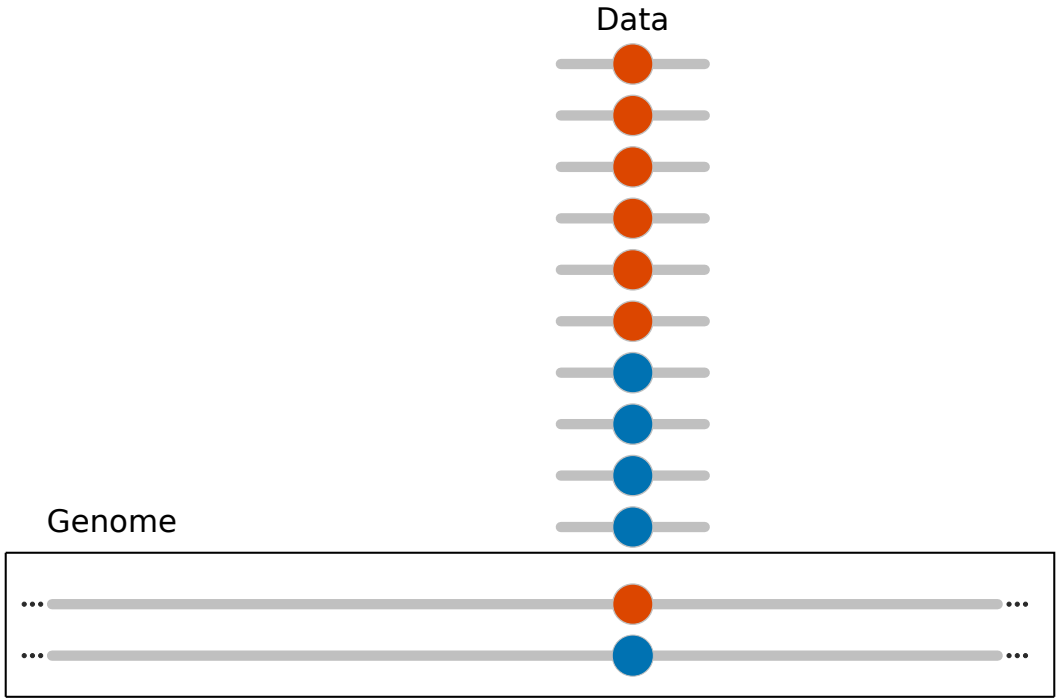
Estimating genotypes

Estimating genotypes

We never directly observe genotypes.

Estimating genotypes

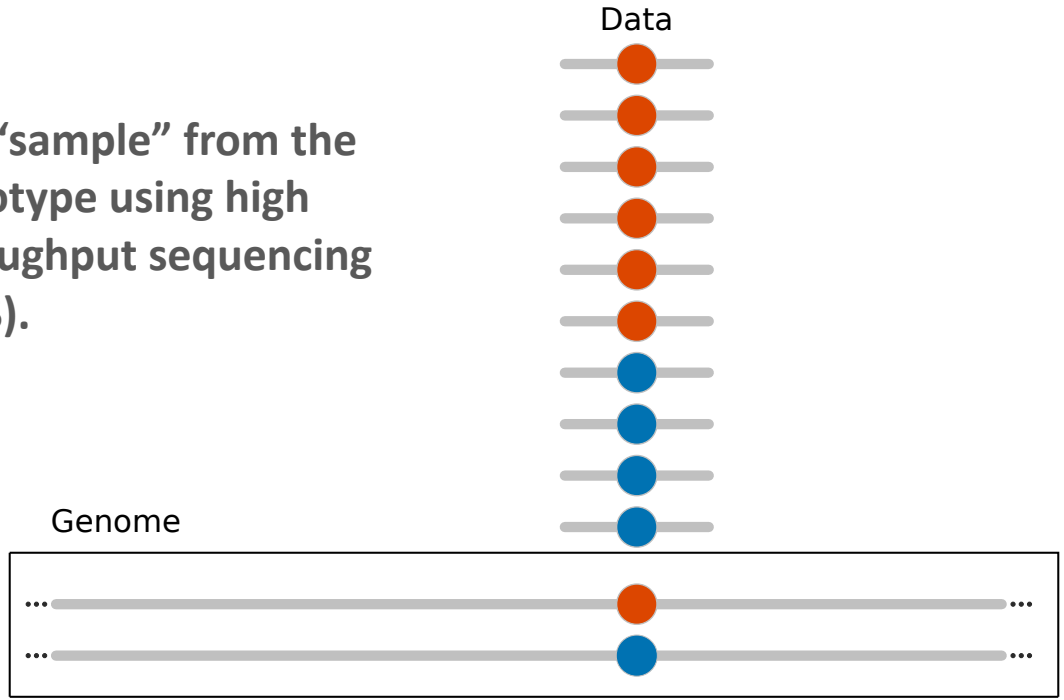
We never directly observe genotypes.



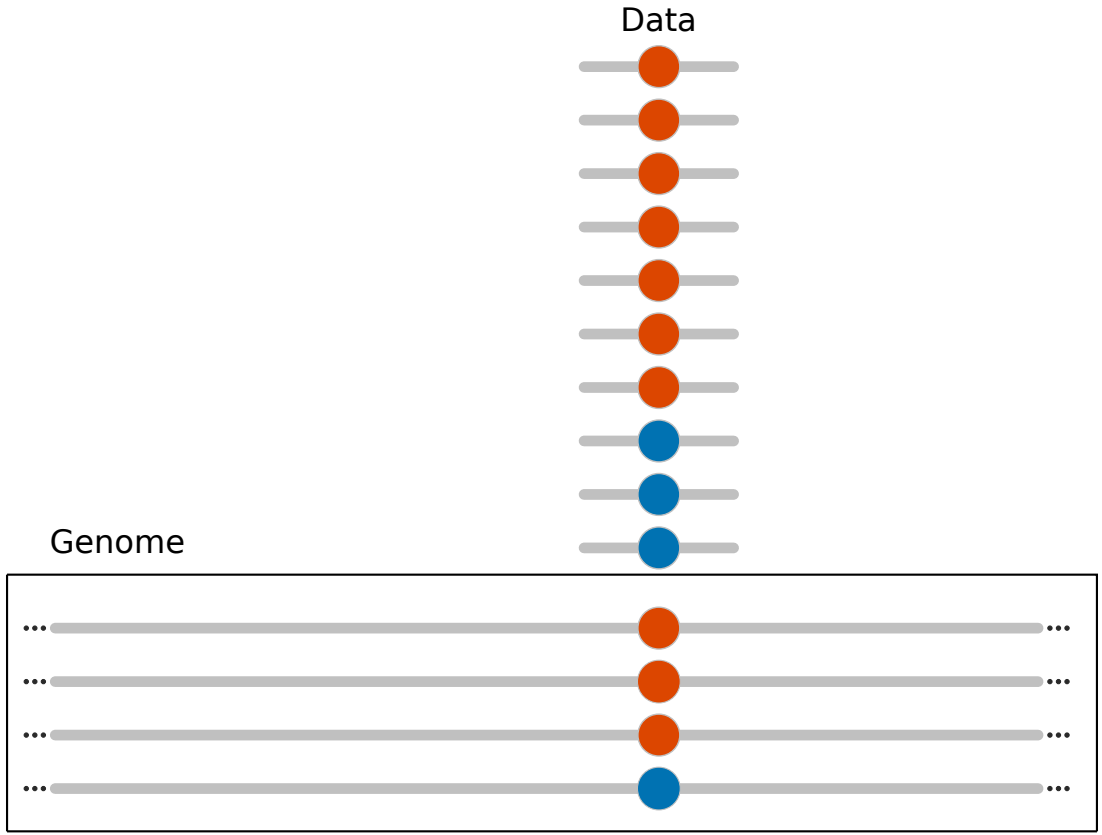
Estimating genotypes

We never directly observe genotypes.

We “sample” from the genotype using high throughput sequencing (HTS).



Estimating genotypes



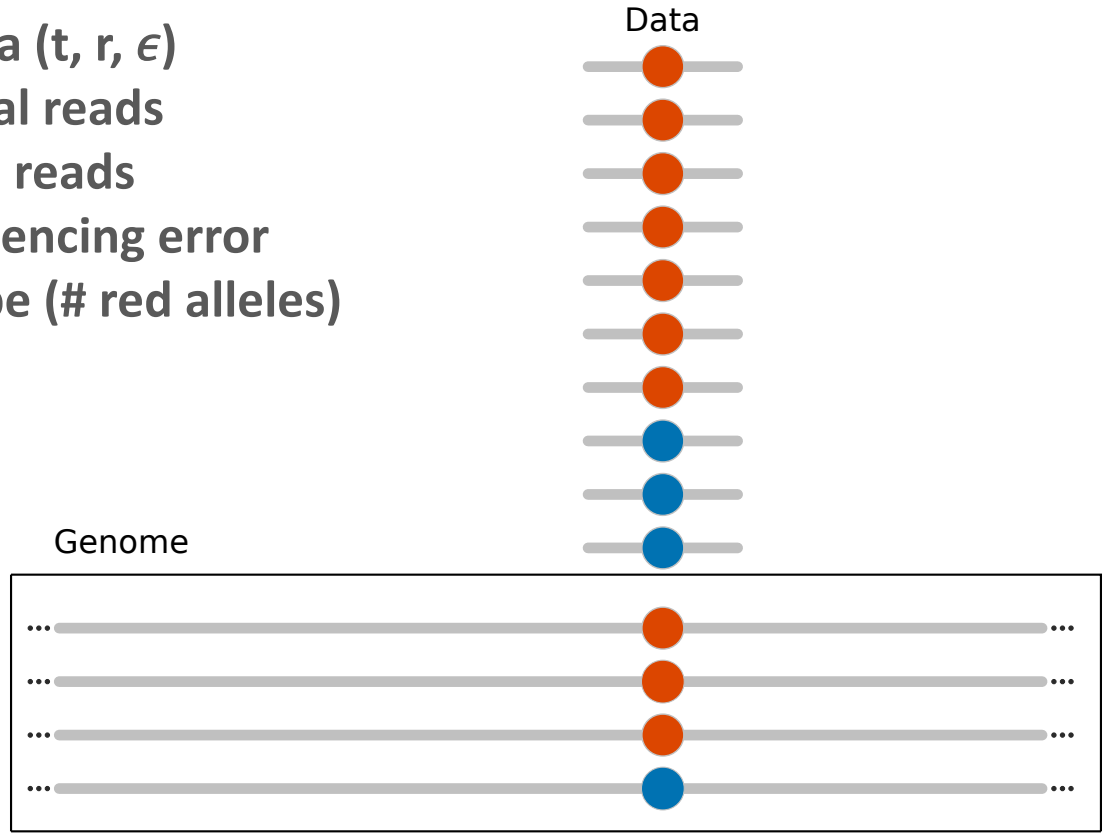
Estimating genotypes

There is a need for a method to apply the same principles in diploids to polyploids.

Blischak *et al*. 2016. *Mol. Ecol. Resour.*
Blischak *et al*. 2018. *Bioinformatics*.

Estimating genotypes – genotype likelihoods

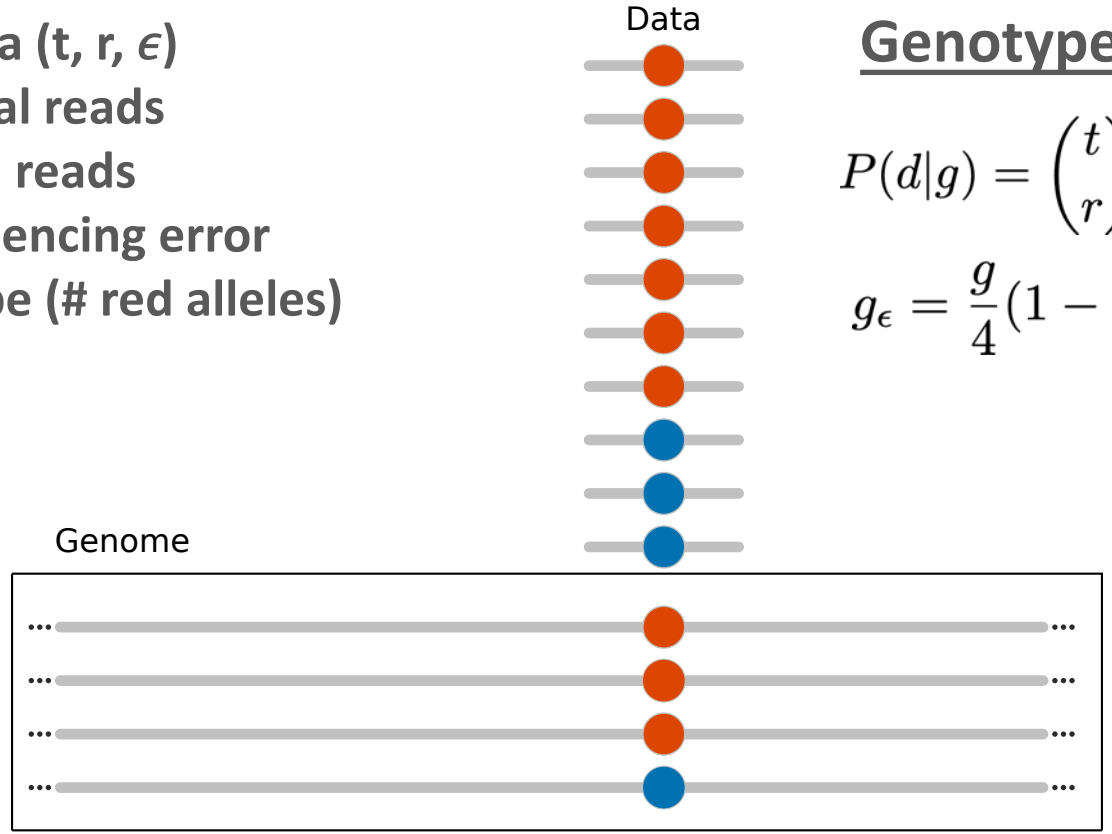
- d = HTS data (t, r, ϵ)
  - t = # total reads
  - r = # red reads
  - ϵ = sequencing error
- g = genotype (# red alleles)



Blischak *et al*. 2016. *Mol. Ecol. Resour.*
Blischak *et al*. 2018. *Bioinformatics*.

Estimating genotypes – genotype likelihoods

- d = HTS data (t, r, ϵ)
  - t = # total reads
  - r = # red reads
  - ϵ = sequencing error
- g = genotype (# red alleles)



Genotype likelihood

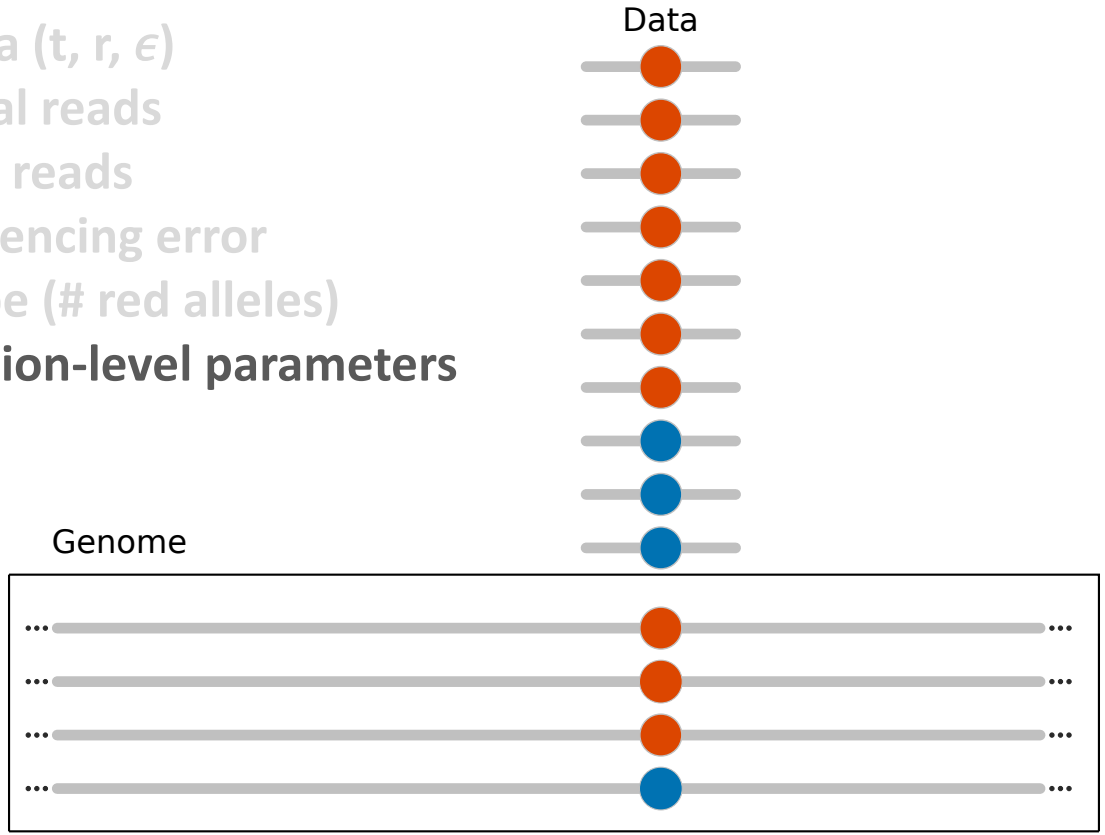
$$P(d|g) = \binom{t}{r} g_\epsilon^r (1 - g_\epsilon)^{t-r}$$

$$g_\epsilon = \frac{g}{4}(1 - \epsilon) + (1 - \frac{g}{4})\epsilon$$

Blischak *et al*. 2016. *Mol. Ecol. Resour.*
Blischak *et al*. 2018. *Bioinformatics*.

Estimating genotypes – genotype priors

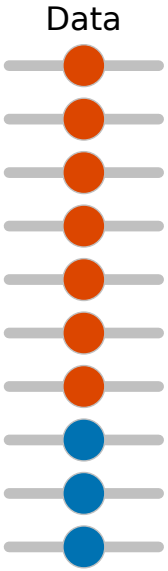
- d = HTS data (t, r, ϵ)
  - t = # total reads
  - r = # red reads
  - ϵ = sequencing error
- g = genotype (# red alleles)
- **θ = population-level parameters**



Blischak *et al*. 2016. *Mol. Ecol. Resour.*
Blischak *et al*. 2018. *Bioinformatics*.

Estimating genotypes – genotype priors

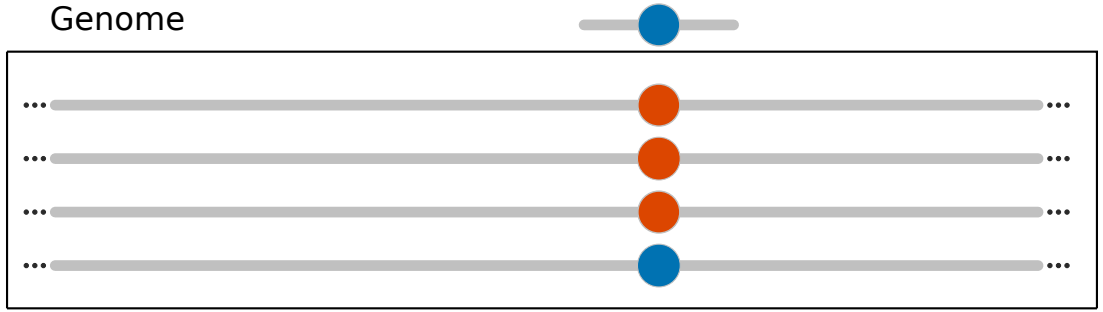
- d = HTS data (t, r, ϵ)
  - t = # total reads
  - r = # red reads
  - ϵ = sequencing error
- g = genotype (# red alleles)
- **θ = population-level parameters**



HWE genotype prior

Eg: θ could be the red allele frequency, *p*.

$$P(g|p) = \binom{g}{4} p^g (1-p)^{4-g}$$



Building a Bayesian model

$$P(D|G) \sim \text{Binomial}(D, G, \epsilon) \quad \text{Genotype likelihood}$$

$$P(G|p) \sim \text{Binomial}(G, p) \quad \text{Genotype prior}$$

$$P(p) \sim \text{Beta}(1, 1) \quad \text{Allele frequency prior}$$

Building a Bayesian model

$$P(D|G) \sim \text{Binomial}(D, G, \epsilon) \quad \text{Genotype likelihood}$$

$$P(G|p) \sim \text{Binomial}(G, p) \quad \text{Genotype prior}$$

$$P(p) \sim \text{Beta}(1, 1) \quad \text{Allele frequency prior}$$

$$P(p, G|D) = \frac{P(D|G)P(G|p)P(p)}{P(D)}$$

Blischak *et al*. 2016. *Mol*. *Ecol*. *Resour*.

Building a Bayesian model

$$P(D|G) \sim \text{Binomial}(D, G, \epsilon) \quad \text{Genotype likelihood}$$

$$P(G|p) \sim \text{Binomial}(G, p) \quad \text{Genotype prior}$$

$$P(p) \sim \text{Beta}(1, 1) \quad \text{Allele frequency prior}$$

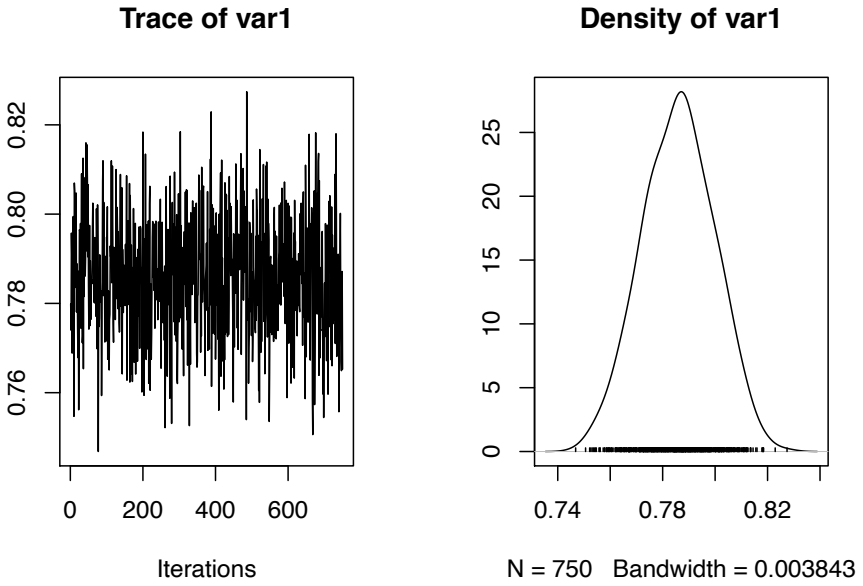
$$P(p, G|D) = \frac{P(D|G)P(G|p)P(p)}{P(D)}$$

We generally have no idea what this is…

Blischak *et al*. 2016. *Mol*. *Ecol*. *Resour*.
GitHub: pblischak/polyfreqs

Building a Bayesian model

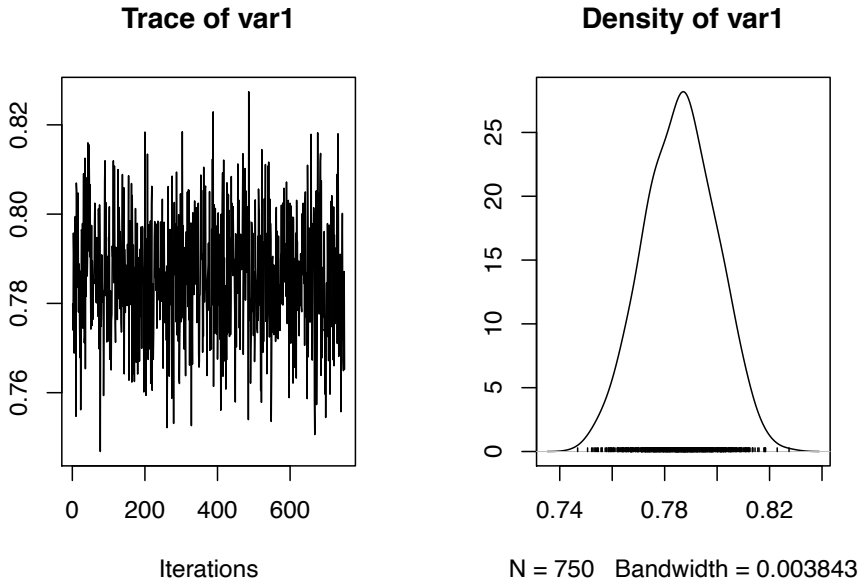
- To bypass the *P(D)* problem and sample from the posterior, I used Markov chain Monte Carlo (MCMC).



Blischak *et al*. 2016. *Mol*. *Ecol*. *Resour*.
GitHub: pblischak/polyfreqs

Building a Bayesian model

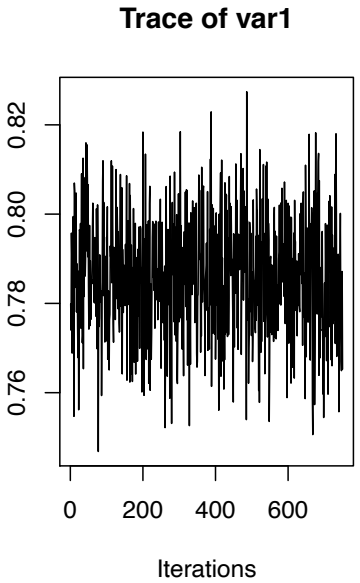
- To bypass the *P(D)* problem and sample from the posterior, I used Markov chain Monte Carlo (MCMC).
- The Binomial and Beta distributions work well together, allowing me to use a special form of MCMC called Gibbs sampling.

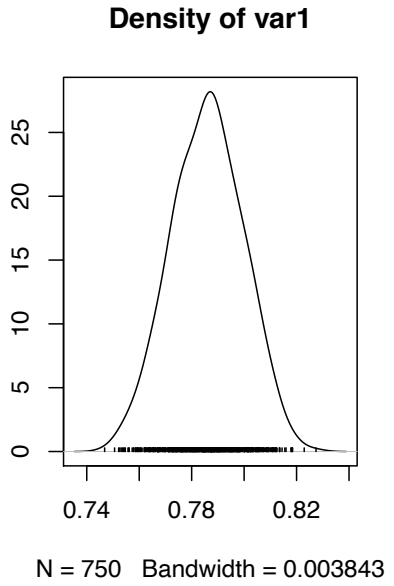


Blischak *et al*. 2016. *Mol*. *Ecol*. *Resour*.
GitHub: pblischak/polyfreqs

Building a Bayesian model

- To bypass the *P(D)* problem and sample from the posterior, I used Markov chain Monte Carlo (MCMC).
- The Binomial and Beta distributions work well together, allowing me to use a special form of MCMC called Gibbs sampling.
- I wrote an R package, POLYFREQS, to implement the MCMC algorithm.





Blischak *et al*. 2016. *Mol*. *Ecol*. *Resour*.
GitHub: pblischak/polyfreqs-ms-data

Application to autotetraploid potato

- Reanalyzed data for 224 individuals genotyped at 384 SNP markers from Voorrips *et al*. (2011).

Photo credit: https://harvesttotable.com

Blischak *et al*. 2016. *Mol*. *Ecol*. *Resour*.
GitHub: pblischak/polyfreqs-ms-data

Application to autotetraploid potato

- Reanalyzed data for 224 individuals genotyped at 384 SNP markers from Voorrips *et al*. (2011).
- Obtained estimates of allele frequencies and genotypes using POLYFREQS R package.

Photo credit: https://harvesttotable.com

Blischak *et al*. 2016. *Mol*. *Ecol*. *Resour*.
GitHub: pblischak/polyfreqs-ms-data

Application to autotetraploid potato

- Reanalyzed data for 224 individuals genotyped at 384 SNP markers from Voorrips *et al*. (2011).
- Obtained estimates of allele frequencies and genotypes using POLYFREQS R package.
- Used these estimates to calculate observed and expected heterozygosity.

Photo credit: https://harvesttotable.com

Blischak *et al*. 2016. *Mol*. *Ecol*. *Resour*.
GitHub: pblischak/polyfreqs-ms-data

Application to autotetraploid potato

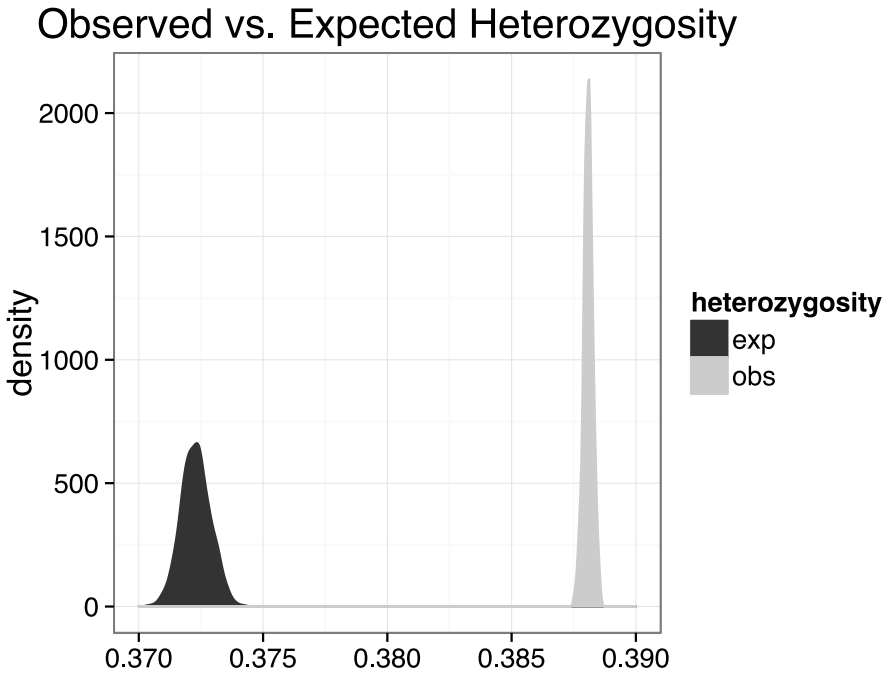
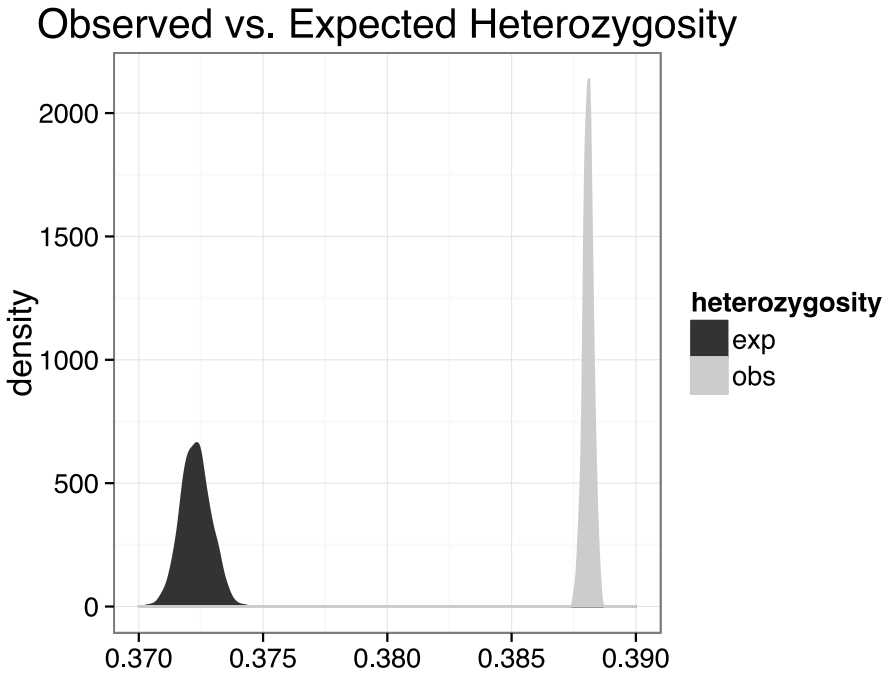


Photo credit: Wikipedia

Blischak *et al*. 2016. *Mol*. *Ecol*. *Resour*.
GitHub: pblischak/polyfreqs-ms-data

Application to autotetraploid potato



Potato is known to be outcrossed, so $H_o > H_e$ is exactly what we would expect.

Photo credit: Wikipedia

POLYFREQS

- R package implementing Gibbs sampling for allele frequency, genotype, and heterozygosity estimation.

POLYFREQS

- R package implementing Gibbs sampling for allele frequency, genotype, and heterozygosity estimation.
- Has functions to generate HTS read count data and to perform posterior predictive simulations as well.

POLYFREQS

- R package implementing Gibbs sampling for allele frequency, genotype, and heterozygosity estimation.
- Has functions to generate HTS read count data and to perform posterior predictive simulations as well.
- Computationally intensive code is written in C++ using the RCPP package.

POLYFREQS

- R package implementing Gibbs sampling for allele frequency, genotype, and heterozygosity estimation.
- Has functions to generate HTS read count data and to perform posterior predictive simulations as well.
- Computationally intensive code is written in C++ using the RCPP package.
- Available on CRAN and GitHub:
  - https://cran.r-project.org/package=polyfreqs
  - https://github.com/pblischak/polyfreqs.git

Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Adding complexity – inbreeding

- Polyploidy is often associated with changes in mating system.

Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Adding complexity – inbreeding

- Polyploidy is often associated with changes in mating system.
- Crops are often highly inbred as well, creating challenges for genotyping.

Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Adding complexity – inbreeding

- Polyploidy is often associated with changes in mating system.
- Crops are often highly inbred as well, creating challenges for genotyping.
- Assuming HWE is likely not appropriate in many instances – need a genotype prior that incorporates disequilibrium scenarios.

Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Beta-Binomial distribution

- Genotype prior that includes both the allele frequency (*p*) and the inbreeding coefficient (*F*).

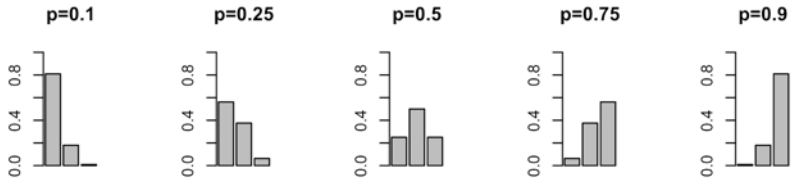
$$P(G|p,F) \sim \text{Beta-Binomial}\left(G, \alpha = p\frac{1-F}{F}, \beta = (1-p)\frac{1-F}{F}\right)$$

Balding & Nichols. 1995. *Genetica*.
Balding & Nichols. 1997. *Heredity*.

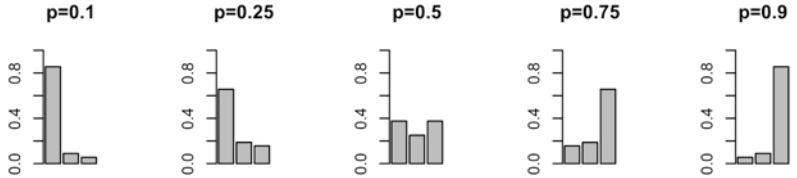
Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Beta-Binomial distribution

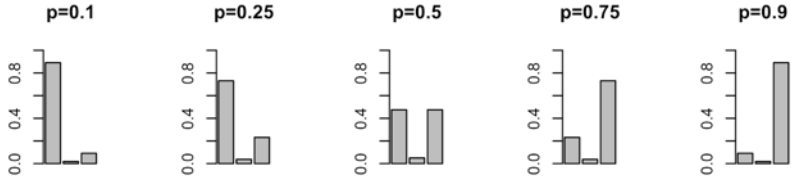
F=0



F=0.5



F=0.9



Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Integrating over genotypes

- We don’t need to estimate genotypes while also inferring p and F.

Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Integrating over genotypes

- We don't need to estimate genotypes while also inferring *p* and *F*.
- How? Build a sum over genotypes into the model.
  - This is known as 'empirical Bayes'.
  - Maximum likelihood inference is conducted using an expectation-maximization algorithm.

Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Integrating over genotypes

- We don't need to estimate genotypes while also inferring *p* and *F*.
- How? Build a sum over genotypes into the model.
  - This is known as 'empirical Bayes'.
  - Maximum likelihood inference is conducted using an expectation-maximization algorithm.
- This is the approach taken by SAMtools (Li 2009), ANGSD (Korneliussen *et al*. 2014), and others.

Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Inbreeding in big bluestem

- Big bluestem (*Andropogon gerardii*) is an important tall grass native to the Great Plains.

Photo credit: Wikipedia

Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Inbreeding in big bluestem

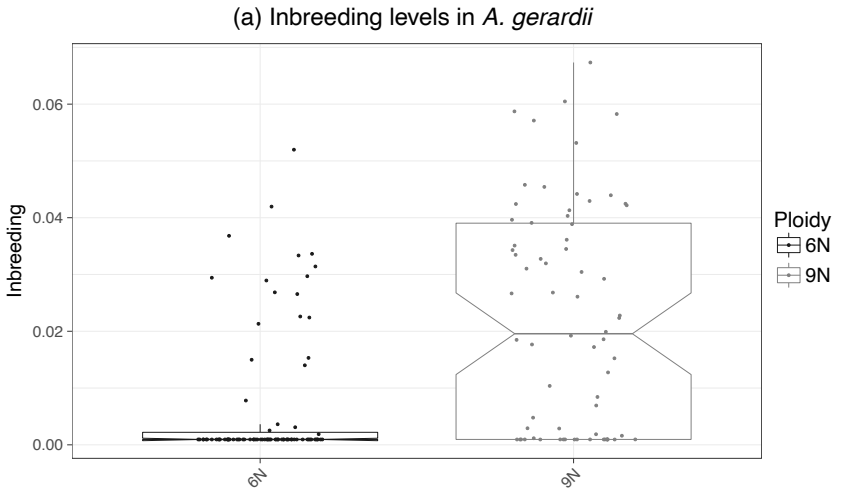
- Big bluestem (*Andropogon gerardii*) is an important tall grass native to the Great Plains.
- Has two main ploidal levels: 6N and 9N.

Photo credit: Wikipedia

Blischak *et al*. 2018. *Bioinformatics*.
GitHub: pblischak/polyploid-genotyping

Inbreeding in big bluestem

- Big bluestem (*Andropogon gerardii*) is an important tall grass native to the Great Plains.
- Has two main ploidal levels: 6N and 9N.
- I estimated inbreeding in the different ploidal levels using SNP data from McAllister & Miller (2016).



EBG (Empirical Bayes Genotyping)

- C++ program implementing several algorithms to estimate allele frequencies, genotypes, and inbreeding coefficients in autopolyploids.

EBG (Empirical Bayes Genotyping)

- C++ program implementing several algorithms to estimate allele frequencies, genotypes, and inbreeding coefficients in autopolyploids.
- Also has genotyping algorithms for allopolyploids based on parental allele frequency information.

EBG (Empirical Bayes Genotyping)

- C++ program implementing several algorithms to estimate allele frequencies, genotypes, and inbreeding coefficients in autopolyploids.
- Also has genotyping algorithms for allopolyploids based on parental allele frequency information.
- Available on GitHub:
  - https://github.com/pblischak/polyploid-genotyping.git

Summary – genotyping polyploids

- Developed flexible Bayesian and empirical Bayes methods to estimate genotypes and to incorporate genotype uncertainty in other parameter estimates.

$$P(p, G|D) = \frac{P(D|G)P(G|p)P(p)}{P(D)}$$

Summary – genotyping polyploids

- Developed flexible Bayesian and empirical Bayes methods to estimate genotypes and to incorporate genotype uncertainty in other parameter estimates.

$$P(p, G|D) = \frac{P(D|G)P(G|p)P(p)}{P(D)}$$

- Demonstrated the effectiveness of these methods for different applications in two empirical examples: Potato and big bluestem.

Photo credit: Wikipedia

Summary – genotyping polyploids

- Developed flexible Bayesian and empirical Bayes methods to estimate genotypes and to incorporate genotype uncertainty in other parameter estimates.

$$P(p, G|D) = \frac{P(D|G)P(G|p)P(p)}{P(D)}$$

- Demonstrated the effectiveness of these methods for different applications in two empirical examples: Potato and big bluestem.

- Implemented the approaches in open-source software to allow other researchers to use these models in their work.

Community applications

Chinook salmon – McKinney *et al*. 2018

Miscanthus (biofuel crop) – Clark *et al*. 2019

ploidyverse

Community for developing tools in polyploid genomics

Photo credit: Wikipedia

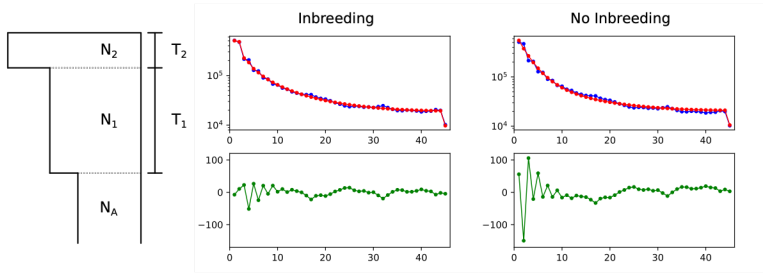
Road map

Data

Genome

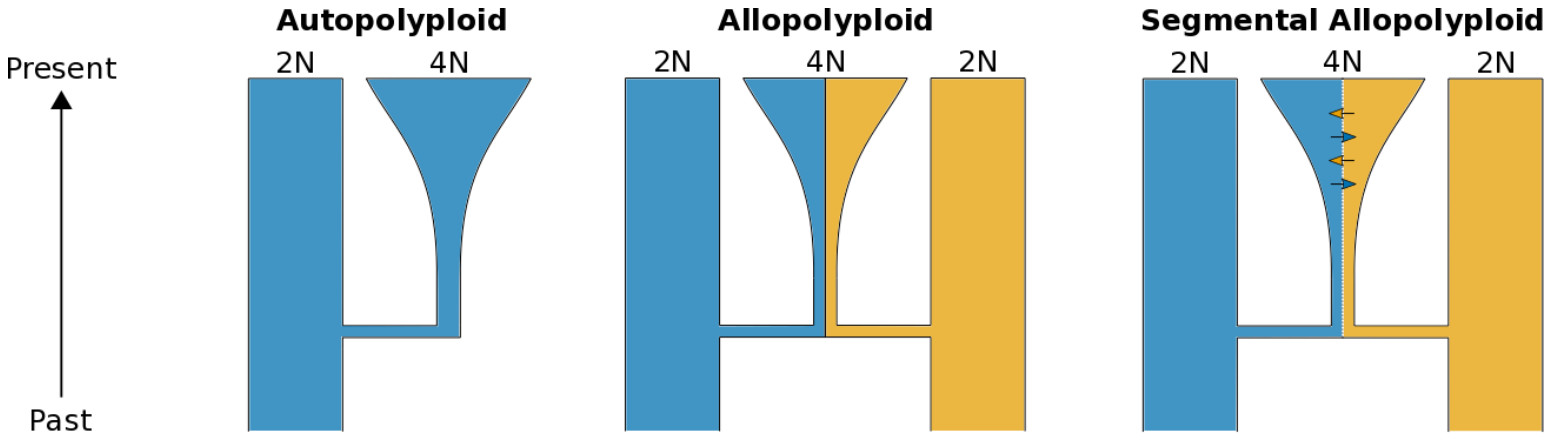
Genotyping and parameter estimation in polyploids

Demographic inference in inbred species



Hybridization detection with convolutional neural networks

Demographic inference in polyploids



Demographic inference in polyploids

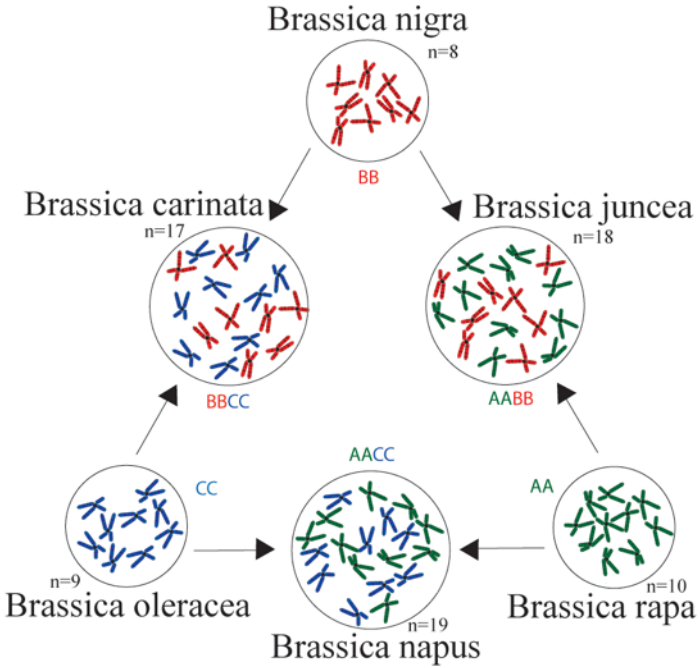


Photo credits: Wikipedia; Wikimedia Commons

Demographic inference in polyploids

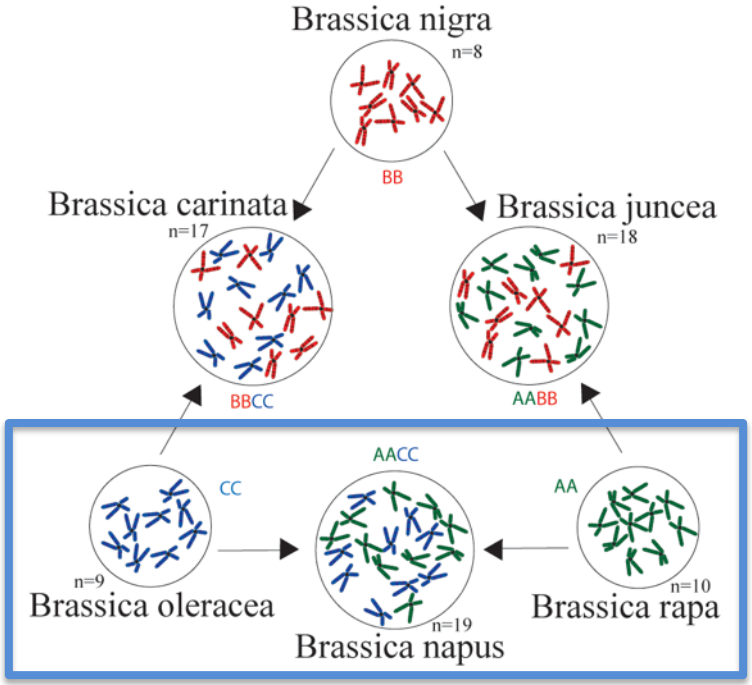


Photo credits: Wikipedia; Wikimedia Commons

Demographic inference in polyploids

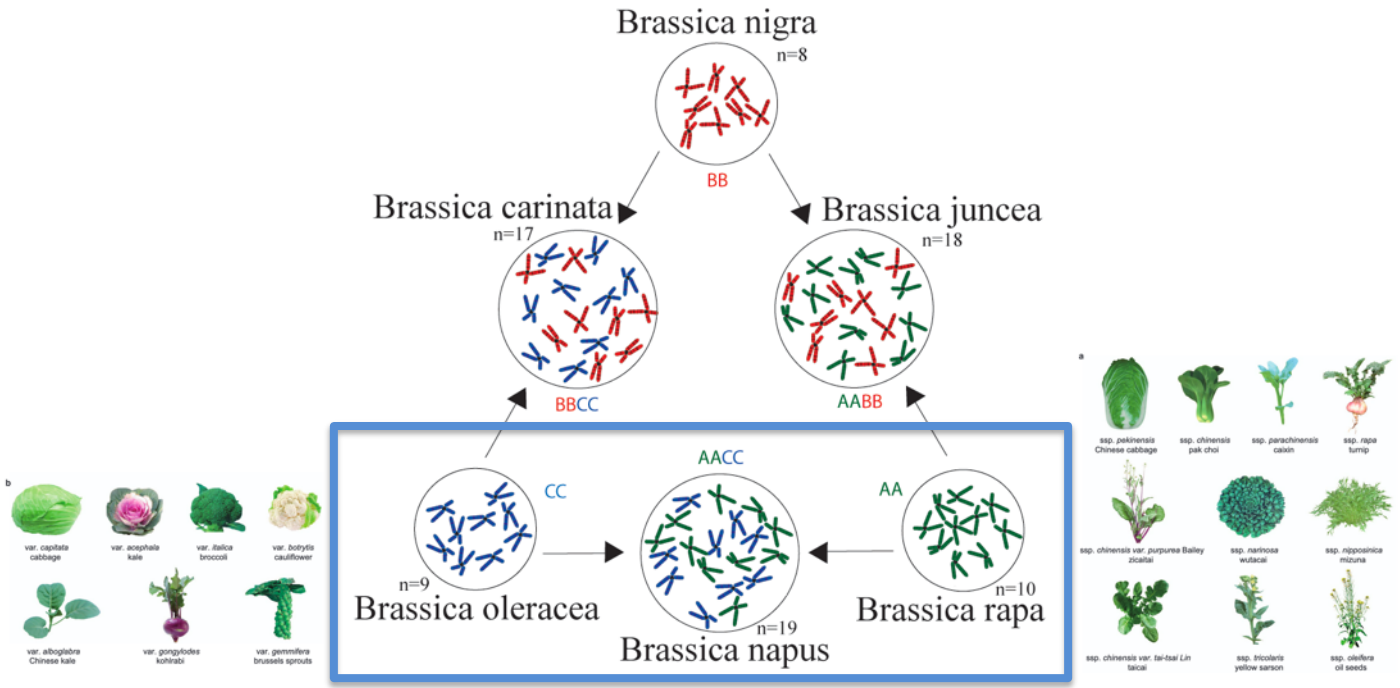


Photo credits: Wikipedia; Wikimedia Commons

Demographic inference in polyploids

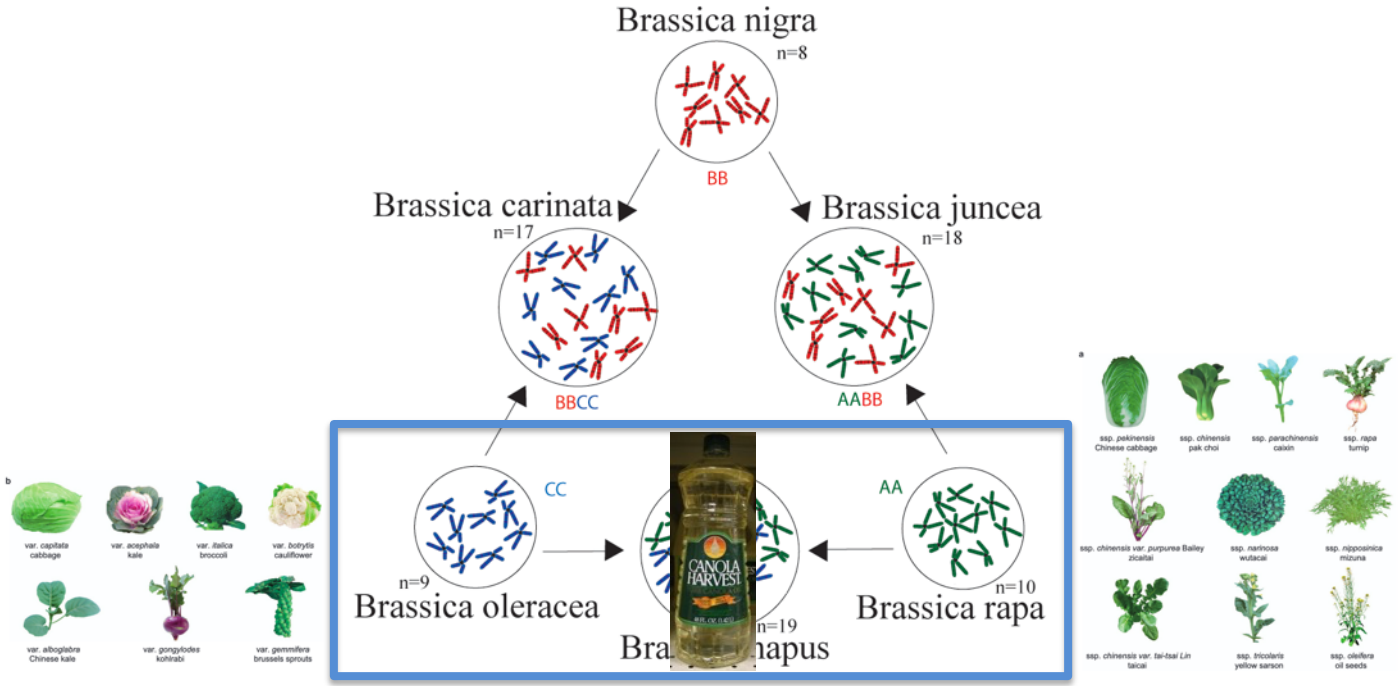


Photo credits: Wikipedia; Wikimedia Commons

Starting simple – demography of cabbage

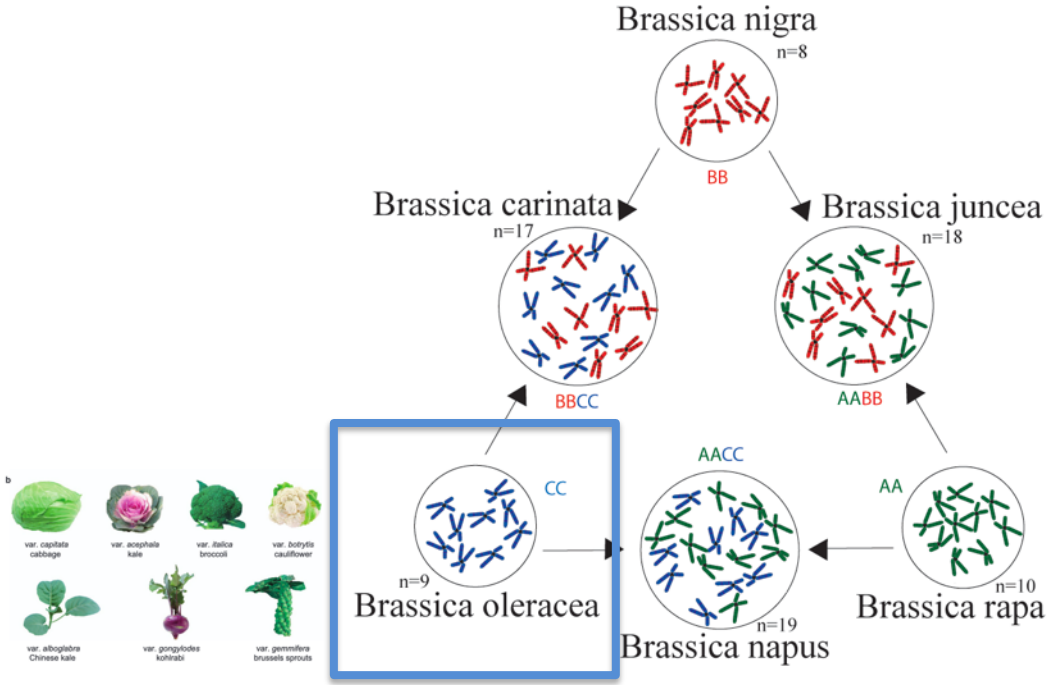


Photo credits: Wikipedia; Wikimedia Commons

The site frequency spectrum (SFS)

- The SFS records how often mutations with different frequencies occur in a population.

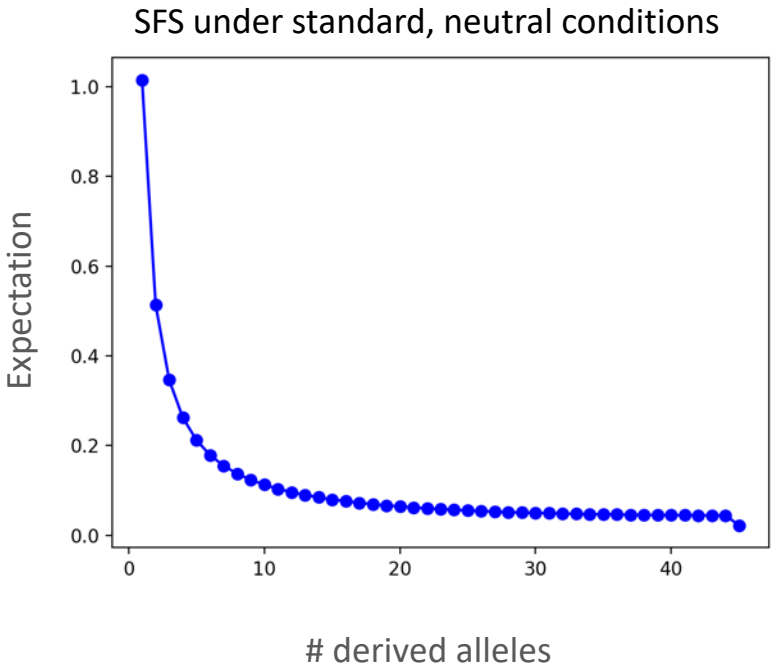
The site frequency spectrum (SFS)

- The SFS records how often mutations with different frequencies occur in a population.
- Demographic events (e.g., changes in population size, divergence, gene flow, etc.) change the shape of the SFS in predictable ways.

The site frequency spectrum (SFS)

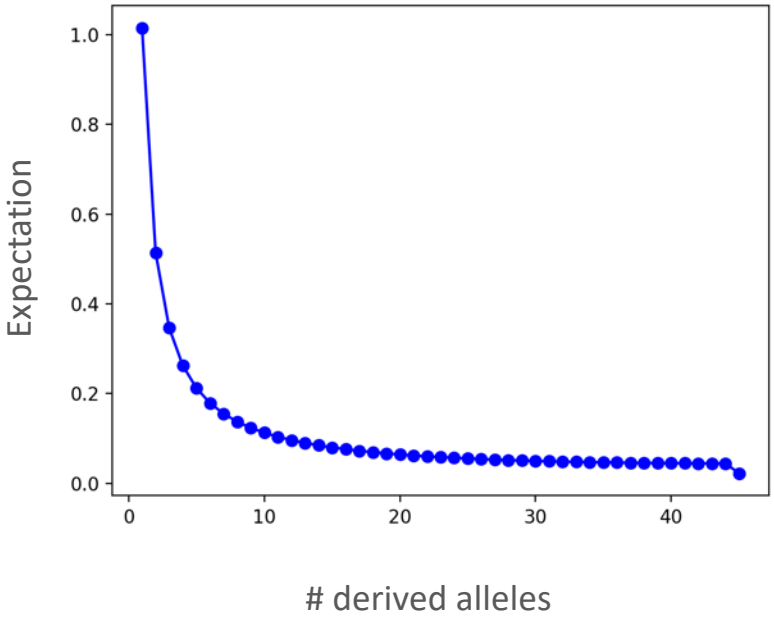
- The SFS records how often mutations with different frequencies occur in a population.
- Demographic events (e.g., changes in population size, divergence, gene flow, etc.) change the shape of the SFS in predictable ways.
- Matching the observed SFS to the predictions of a model allows us to make inferences about demographic events.

The site frequency spectrum (SFS)

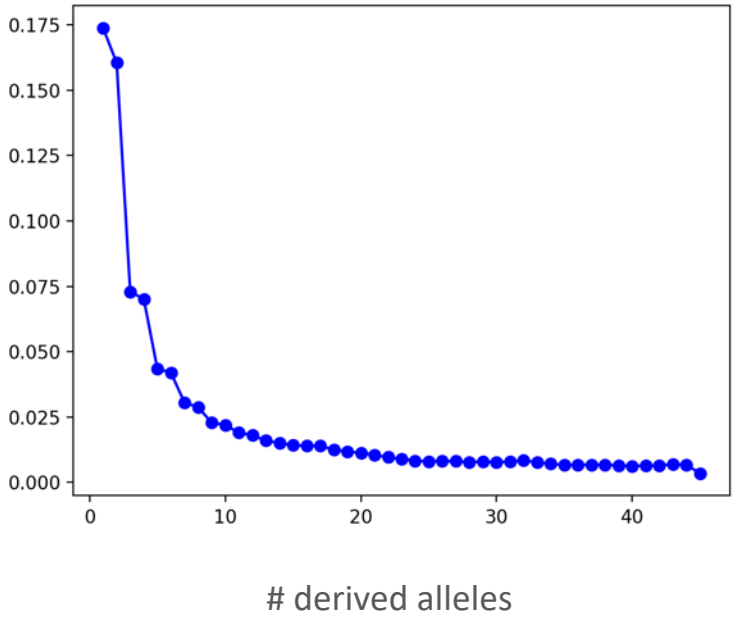


The site frequency spectrum (SFS)

SFS under standard, neutral conditions

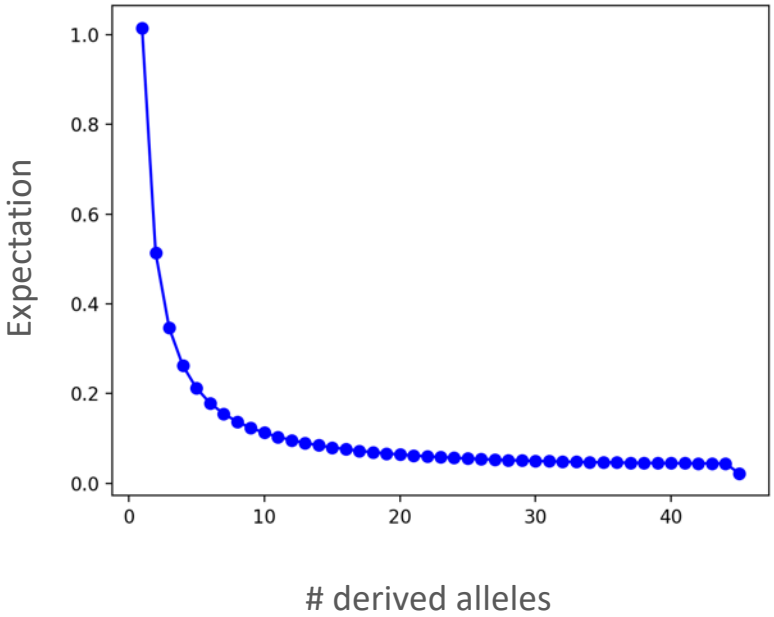


SFS for cabbage

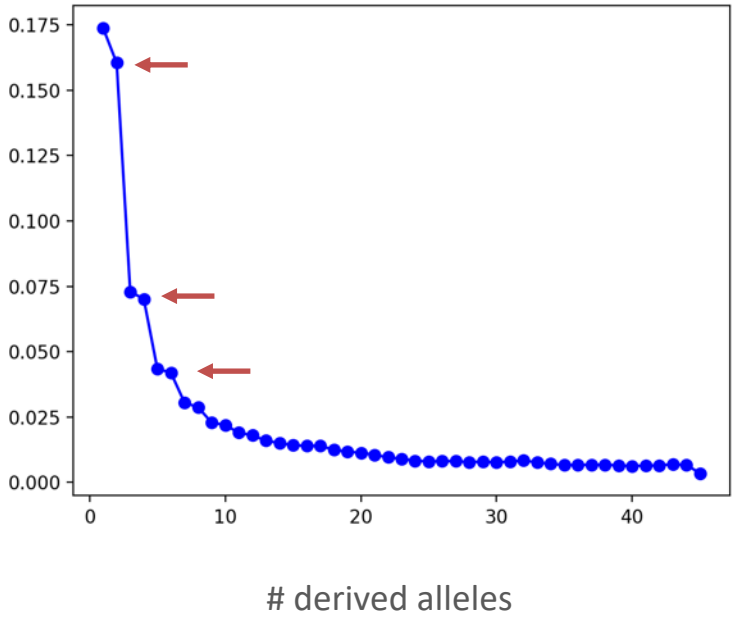


The site frequency spectrum (SFS)

SFS under standard, neutral conditions

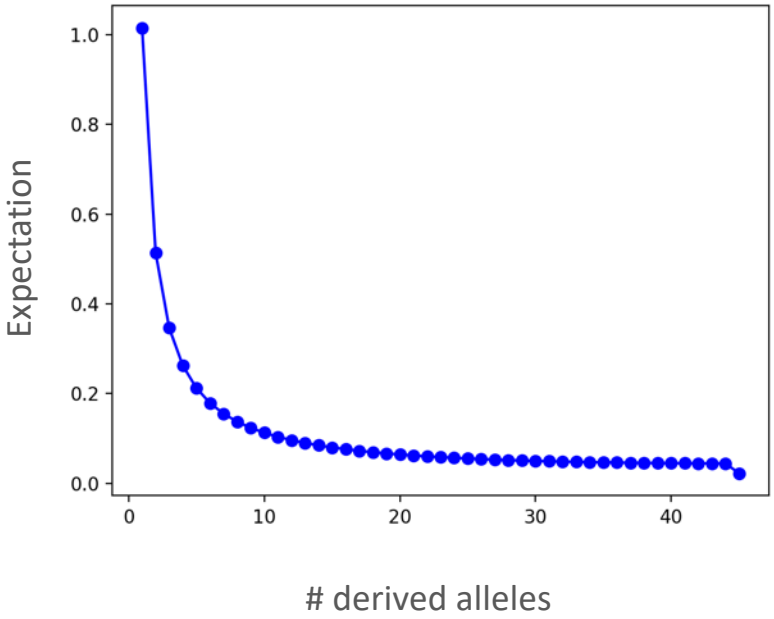


SFS for cabbage

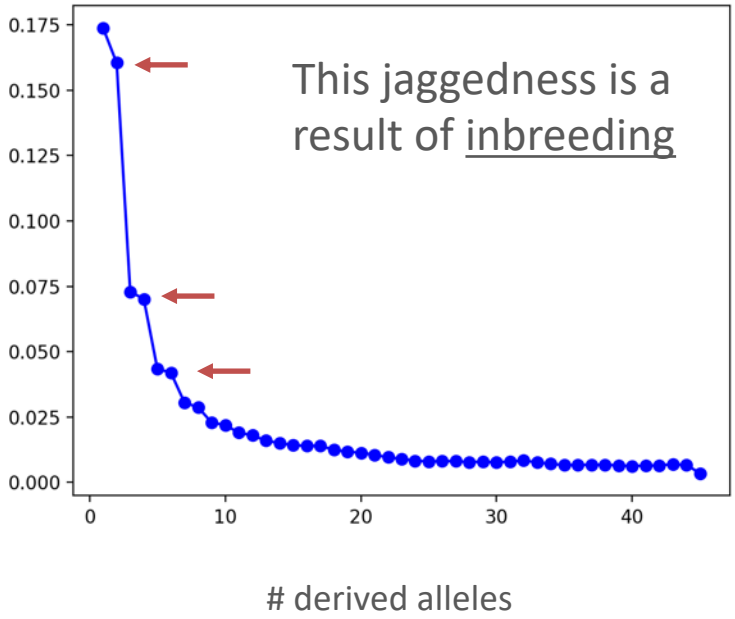


The site frequency spectrum (SFS)

SFS under standard, neutral conditions



SFS for cabbage



Blischak *et al*. 2020. *Mol*. *Biol*. *Evol*.
GitHub: pblischak/inbreeding-sfs

Inbreeding and the SFS

- Developed an approach using convolutions of Beta-Binomial distributions to incorporate inbreeding into the expected SFS.

Blischak *et al*. 2020. *Mol*. *Biol*. *Evol*.
GitHub: pblischak/inbreeding-sfs

Inbreeding and the SFS

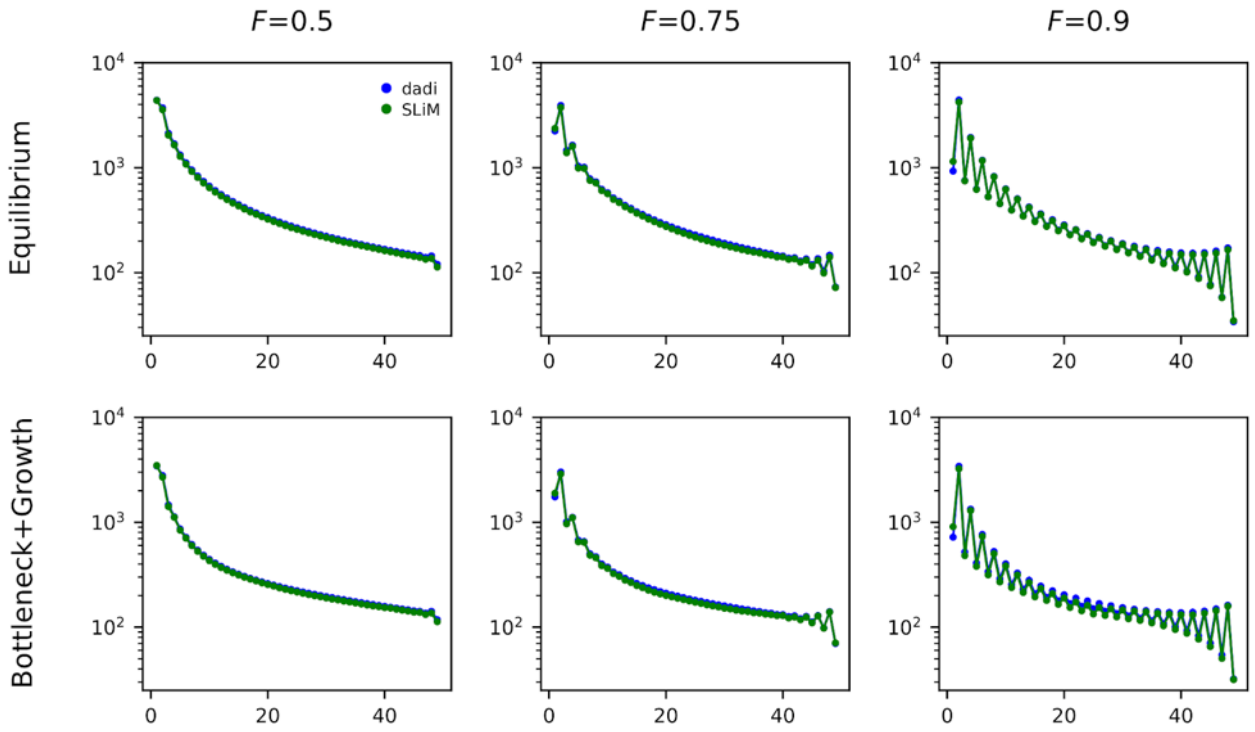
- Developed an approach using convolutions of Beta-Binomial distributions to incorporate inbreeding into the expected SFS.
- Implemented joint inference of demography and inbreeding in the Python package *dadi* (Gutenkunst *et al.* 2009).

Blischak *et al*. 2020. *Mol*. *Biol*. *Evol*.
GitHub: pblischak/inbreeding-sfs

Inbreeding and the SFS

- Developed an approach using convolutions of Beta-Binomial distributions to incorporate inbreeding into the expected SFS.
- Implemented joint inference of demography and inbreeding in the Python package *dadi* (Gutenkunst *et al.* 2009).
- Validated the approach using forward genetic simulations in SLiM 3 (Haller & Messer 2019).

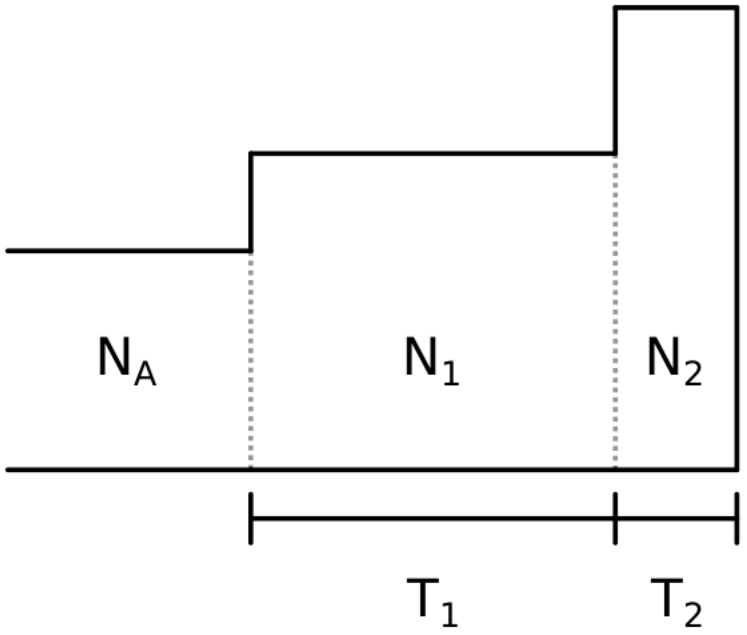
Inbreeding and the SFS



Blischak *et al*. 2020. *Mol*. *Biol*. *Evol*.
GitHub: pblischak/inbreeding-sfs

Demography and inbreeding in cabbage

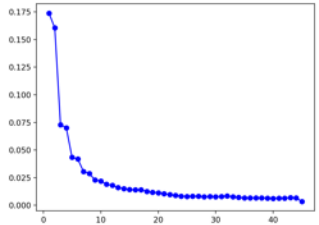
Parameter	Inbreeding	No Inbreeding
N_A	17,500 ind.	19,100 ind.
N_1	31,600 ind.	123,000 ind.
N_2	215,000 ind.	592 ind.
T_1	16,600 yrs.	5,870 yrs.
T_2	322 yrs.	38.8 yrs.
F	0.578	--
Log-likelihood	-4,821	-24,330



Blischak *et al*. 2020. *Mol*. *Biol*. *Evol*.
GitHub: pblischak/inbreeding-sfs

Summary – demography and inbreeding

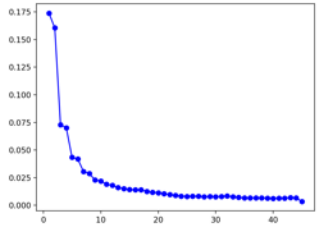
- Inbreeding has a strong effect on the SFS and can mislead demographic inference if left unmodeled.
  - This is likely an issue for many crops and could negatively influence inferences of domestication history.



Blischak *et al*. 2020. *Mol*. *Biol*. *Evol*.
GitHub: pblischak/inbreeding-sfs

Summary – demography and inbreeding

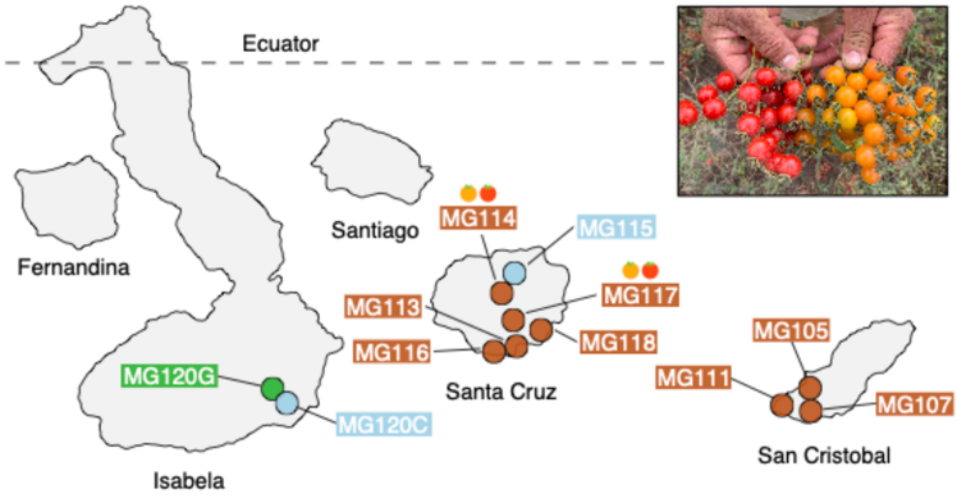
- Inbreeding has a strong effect on the SFS and can mislead demographic inference if left unmodeled.
  - This is likely an issue for many crops and could negatively influence inferences of domestication history.



- Joint inference of demography and inbreeding in *dadi* makes this approach accessible to any researchers studying inbred species (especially crops).

Photo credits: Britannica.com

Community applications



Wild tomato on the Galápagos – Figure 1 from Gibson *et al*. 2020

Road map

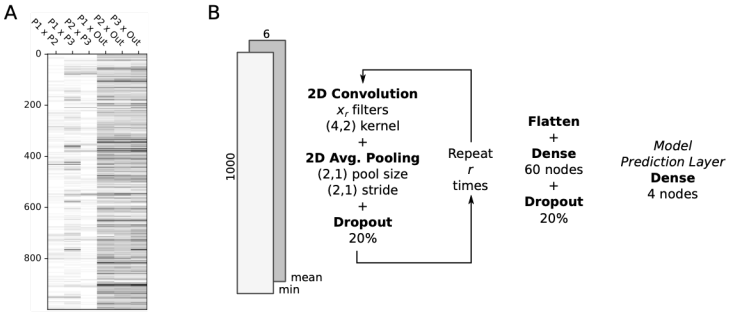
Data

Genome

Genotyping and parameter estimation in polyploids

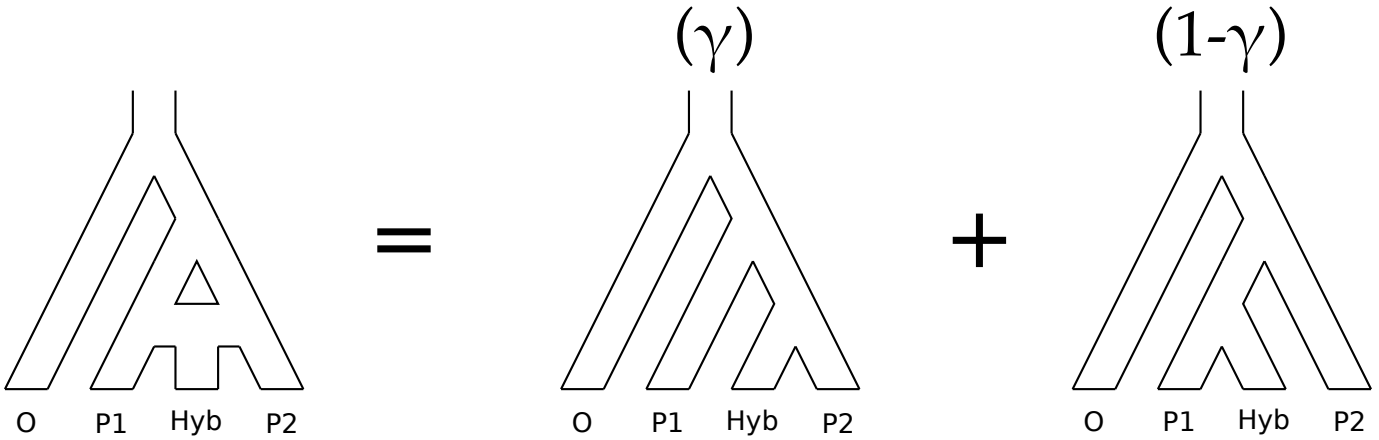
Demographic inference in inbred species

Hybridization detection with convolutional neural networks



Blischak *et al*. 2018. *Syst. Biol.*
GitHub: pblischak/HyDe

Hybridization detection



Blischak *et al*. 2018. *Syst. Biol.*
GitHub: pblischak/HyDe

Hybridization detection

- Previous approaches rely on counting SNP patterns (e.g., HyDe and “ABBA-BABA” tests).

Blischak *et al*. 2018. *Syst. Biol.*
GitHub: pblischak/HyDe

Hybridization detection

- Previous approaches rely on counting SNP patterns (e.g., HyDe and “ABBA-BABA” tests).
- Assume all SNPs are independent.

Blischak *et al*. 2018. *Syst. Biol.*
GitHub: pblischak/HyDe

Hybridization detection

- Previous approaches rely on counting SNP patterns (e.g., HyDe and “ABBA-BABA” tests).
- Assume all SNPs are independent.
- Binary decision:
  - We detect hybridization.
  - We fail to detect hybridization.

Blischak *et al*. 2018. *Syst. Biol.*
GitHub: pblischak/HyDe

Chromosome-scale hybridization detection

- Things that are ignored:
  - SNPs are linked.
  - The size of haplotype blocks exchanged between species can help to determine the timing of hybridization.
  - The mode of hybridization can be complex (e.g., hybrid speciation vs. admixture vs. gene flow).

Blischak *et al*. 2018. *Syst. Biol.*
GitHub: pblischak/HyDe

Chromosome-scale hybridization detection

- Things that are ignored:
  - SNPs are linked.
  - The size of haplotype blocks exchanged between species can help to determine the timing of hybridization.
  - The mode of hybridization can be complex (e.g., hybrid speciation vs. admixture vs. gene flow).
- How do we leverage SNPs from a whole chromosome to infer hybridization?

Blischak *et al*. 2018. *Syst. Biol.*
GitHub: pblischak/HyDe

Chromosome-scale hybridization detection

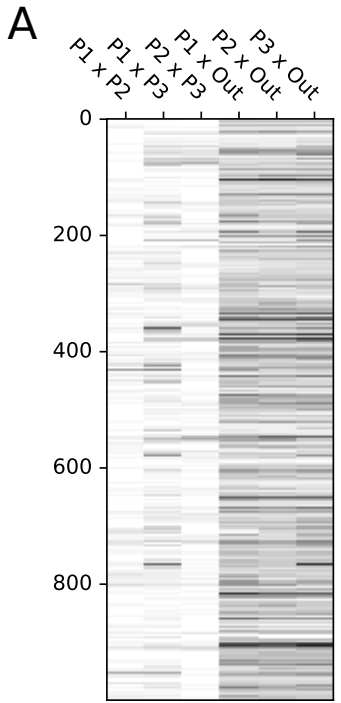
- Things that are ignored:
  - SNPs are linked.
  - The size of haplotype blocks exchanged between species can help to determine the timing of hybridization.
  - The mode of hybridization can be complex (e.g., hybrid speciation vs. admixture vs. gene flow).
- How do we leverage SNPs from a whole chromosome to infer hybridization?

Convolutional neural networks (CNNs) can capture complex correlations in image data and are well-suited to classification problems.

Blischak *et al*. 2020. *bioRxiv.*
GitHub: pblischak/hyde-cnn

Encoding genomic data as an image

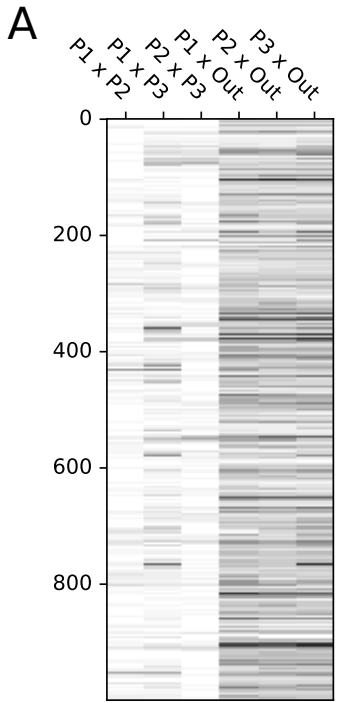
- Dimension one:
  - Calculate pairwise divergence (d_{XY}) in windows across the chromosome.



Blischak *et al*. 2020. *bioRxiv.*
GitHub: pblischak/hyde-cnn

Encoding genomic data as an image

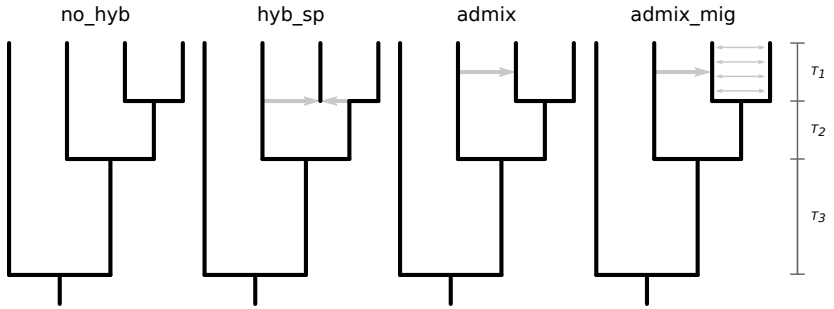
- Dimension one:
  - Calculate pairwise divergence (d_{XY}) in windows across the chromosome.
- Dimension two:
  - Order pairwise comparisons in phylogenetic order of increasing divergence.



Blischak *et al*. 2020. *bioRxiv.*
GitHub: pblischak/hyde-cnn

Generating simulated training data

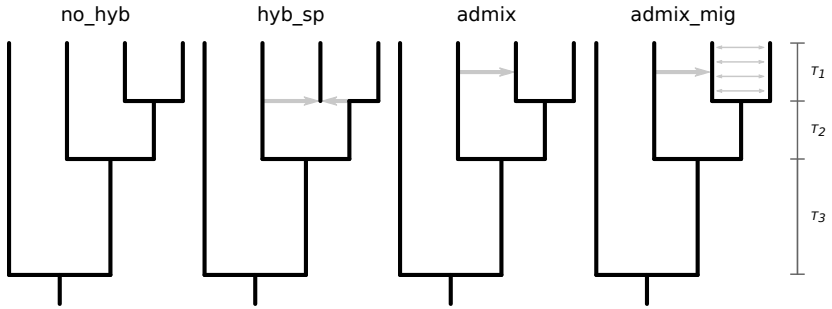
- Chromosome-scale data were generated using *msprime* under four models of hybridization.



Blischak *et al*. 2020. *bioRxiv.*
GitHub: pblischak/hyde-cnn

Generating simulated training data

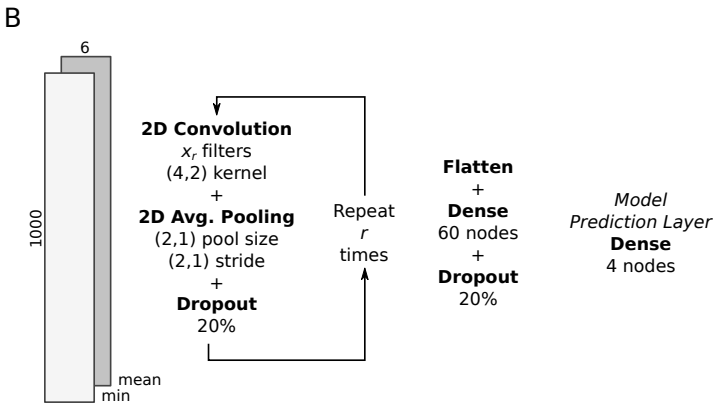
- Chromosome-scale data were generated using *msprime* under four models of hybridization.
- Sampled widely from parameter space for mutation rate, recombination rate, divergence, etc.



Blischak *et al*. 2020. *bioRxiv.*
GitHub: pblischak/hyde-cnn

CNN architecture

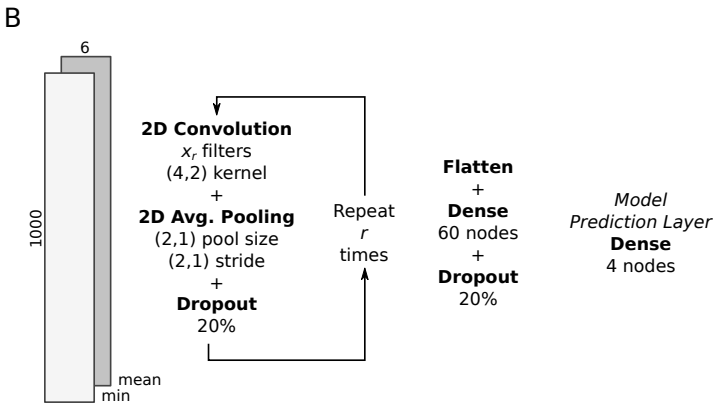
- Used a modified version of the LeNet architecture (LeCun *et al*. 1998) specified in TensorFlow 2.



Blischak *et al*. 2020. *bioRxiv.*
GitHub: pblischak/hyde-cnn

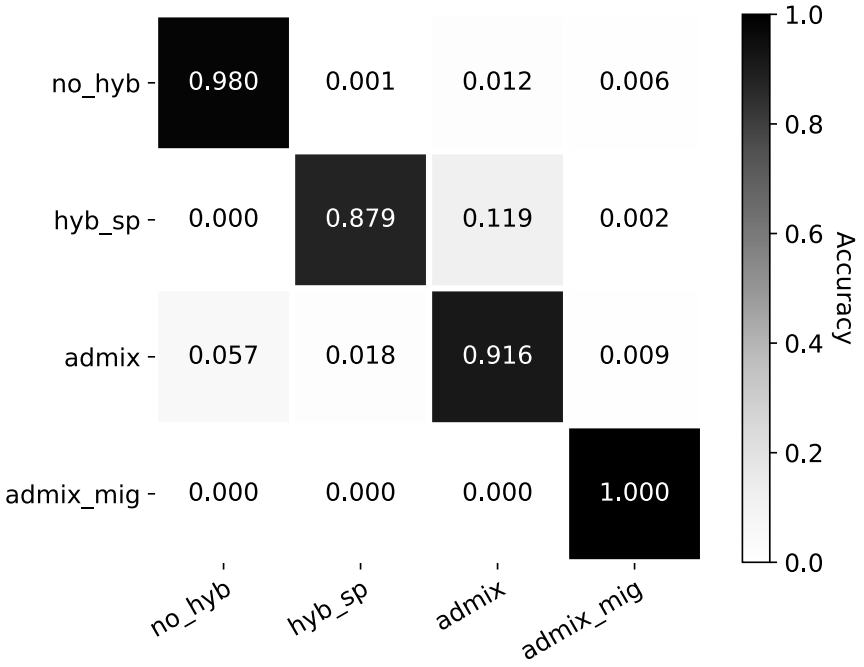
CNN architecture

- Used a modified version of the LeNet architecture (LeCun *et al*. 1998) specified in TensorFlow 2.
- Split simulated data in 70%:15%:15% for training, validation, and testing, respectively.



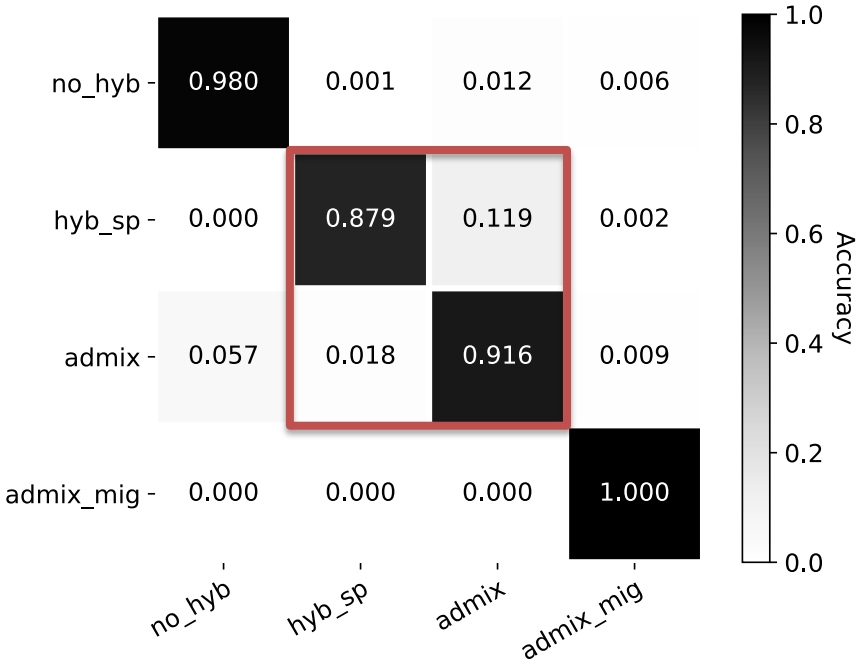
Blischak *et al*. 2020. *bioRxiv.*
GitHub: pblischak/hyde-cnn

Classification accuracy



Blischak *et al*. 2020. *bioRxiv.*
GitHub: pblischak/hyde-cnn

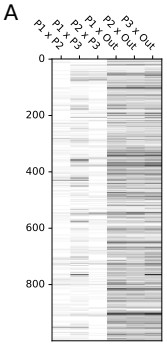
Classification accuracy



Blischak *et al*. 2020. *bioRxiv.*
GitHub: pblischak/hyde-cnn

Summary – hybridization detection

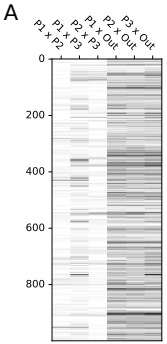
- Encoding pairwise divergence along a chromosome as an image provides powerful information for inferring patterns of hybridization with CNNs.

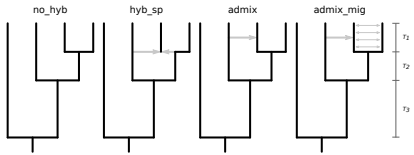


Blischak *et al*. 2020. *bioRxiv*.
GitHub: pblischak/hyde-cnn

Summary – hybridization detection

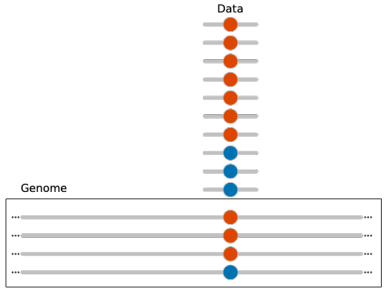
- Encoding pairwise divergence along a chromosome as an image provides powerful information for inferring patterns of hybridization with CNNs.
- Fast and flexible genome-scale simulation also makes it easy to train neural networks to make predictions.





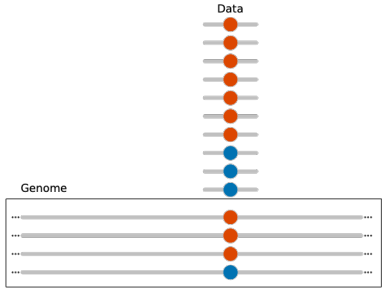
Extensions and future project ideas

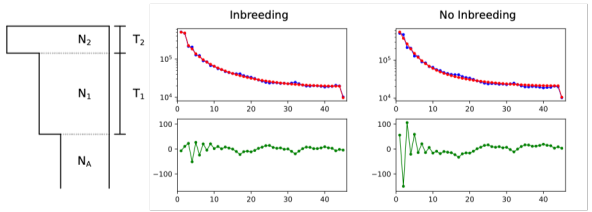
Extensions and future project ideas



- Multiallelic SNPs, haplotypes, complex variants
- Inferences on pangenomes/genome graphs
- GATK or ANGSD for polyploids

Extensions and future project ideas

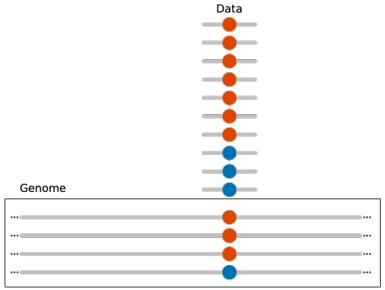


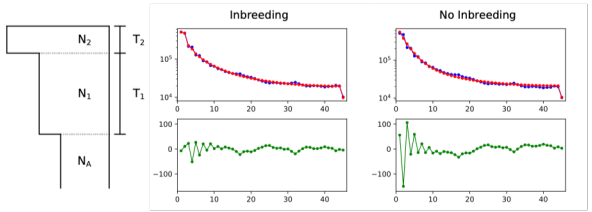


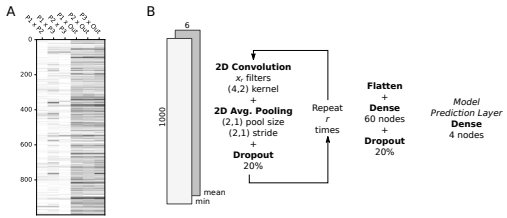
- Multiallelic SNPs, haplotypes, complex variants
- Inferences on pangenomes/genome graphs
- GATK or ANGSD for polyploids

- Incorporate inferences of selection to better ID domestication genes
- Build models for inbreeding, polyploidy, and domestication history

Extensions and future project ideas







- Multiallelic SNPs, haplotypes, complex variants
- Inferences on pangenomes/genome graphs
- GATK or ANGSD for polyploids

- Incorporate inferences of selection to better ID domestication genes
- Build models for inbreeding, polyploidy, and domestication history

- Explore additional representations and network architectures for genomic data
- Build more complex models to integrate genomic information with phenotypes, etc.

Acknowledgements

- Wolfe and Kubatko labs at The Ohio State University.
- Barker and Gutenkunst labs at University of Arizona.
- Researchers who shared data: S. Turner-Hissong (cabbage data).
- NSF Postdoc Fellowship (IOS-1811784).

Thanks!

Questions?