NC STATE UNIVERSITY

Coancestry and Inbreeding

NC STATE UNIVERSITY

Introduction and Definition of Identity by Descent

The concepts of the degree of relationship between individuals and the inbreeding coefficient were developed by Sewall Wright (1922) by use of correlation and path coefficient methodology.

Malécot (1948, 1969) developed the concepts using probability theory.

The concepts of coancestry and inbreeding will be defined for autosomal diploid loci in this section.

The method is based on Malécot' s approach modified by Kempthorne (1957).

With respect to a specific locus, say A, two alleles, x and y, are said to be alike in state if x and y are functionally indistinguishable.

In contrast, x and y are identical by descent (IBD) if one of the following three conditions holds :

1) x is a copy of y,
2) y is a copy of x, or
3) x and y are copies of the same ancestral gene.

"Copy" is synonymous with meiotic duplication.

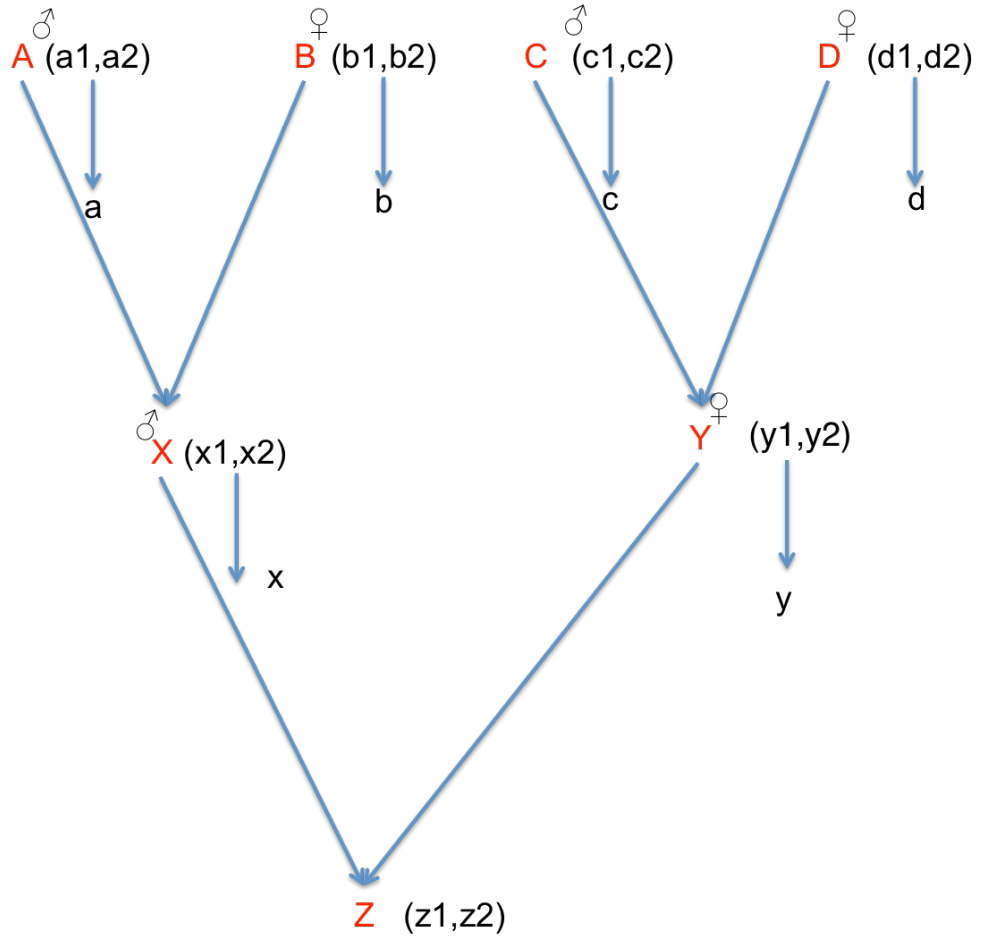
Now if x and $y4 are alleles randomly drawn from the A locus of individuals X and Y, then the coefficient of coancestry is defined as

$rXY = P(x \equiv y)$, where "≡" means IBD.

The coefficient of kinship, coefficient of consanguinity and coefficient de parente are synonymous to the coefficient of coancestry. Kempthorne (1957) used still another term for coancestry, the coefficient of parentage

NC STATE UNIVERSITY

Derivation of coancestry and Inbreeding



The coefficient of coancestry for autosomal diploids may be derived by applying the simple rules of probability. The figure above depict a pedigree showing the genes at a locus in an autosomal diploid individual. (The subscripts 1 and 2 are arbitrarily designated as alleles received from the male (1) and female (2) gametes, respectively.

Thus, $x1 = a$, $x2 = b$, $y1 = c$, $y2 = d$, $z1 = x$ and $z2 = y$.

Now let x and y be alleles at locus A sampled at random from individuals X and Y.

NC STATE UNIVERSITY

Then the coancestry is is the average of the four coancestries among the parents of X and Y:

$$rXY$$
$$= P(x = x_1 y = y_1, x_1 \equiv y_1) + P(x = x_1 y = y_2, x_1 \equiv y_2)$$
$$+ P(x = x_2 y = y_1, x_2 \equiv y_1) + P(x = x_2 y = y_2, x_2 \equiv y_2)$$
$$= \frac{1}{4}[P(a \equiv c) + P(a \equiv d) + P(b \equiv c) + P(b \equiv d)$$
$$= \frac{1}{4}(rAC + rAD + rBC + rBD)$$

If Z is a progeny of X and Y, it follows directly that the inbreeding coefficient of Z is the probability that the alleles z1 and z2 are IBD :

$$F_Z = P(z1 \equiv z_2)$$
$$= P(z_1 = x_1 z_2 = y_1, x_1 \equiv y_1) + P(z_1 = x_1 z_2 = y_2, x_1 \equiv y_2)$$
$$+ P(z_1 = x_2 z_2 = y_1, x_2 \equiv y_1) + P(z_1 = x_2 z_2 = y_2, x_2 \equiv y_2)$$
$$= \frac{1}{4}[P(X1 \equiv Y1) + P(X1 \equiv Y2) + P(X2 \equiv Y1) + P(X2 \equiv Y2)]$$
$$= \frac{1}{4}[P(a \equiv c) + P(a \equiv d) + P(b \equiv c) + P(b \equiv d)]$$
$$= r_{XY}$$

which shows that the inbreeding coefficient of an individual is equal to the coancestry between the individual' s parents.

A special case is the coancestry of an individual with itself.

Consider sampling a gene twice with replacement from individual X.

Designating these genes as x and x'

$$r_{XX} = P(x \equiv x')$$
$$= P(x = x_1 x' = x_1) + P(x = x_2 x' = x_2)$$
$$+ P(x = x_1 x' = x_2 x_1 \equiv x_2) + P(x = x_2 x' = x_1 x_2 \equiv x_1)$$
$$= (1 + F_x)/2$$

Often it is convenient to calculate coancestry between two individuals by finding the coancestry between one individual and the parents of the other.

When using this algorithm, it is essential to use the coancestry between the older individual (appearing earlier in the pedigree) and the parents of the younger individual (appearing later in the pedigree).

Otherwise, an incorrect answer may be obtained because intermediate paths of relationship may be omitted.

Referring to Figure before, gametes of C and D can form zygotes (c1 d1), (c1 d2), (c2 d1) and (c2 d2) with equal frequency of one quarter.

$$r_{XY} = r_{X(CxD)} = \frac{1}{4}\{\frac{1}{4}[P(x_1 \equiv c_1) + P(x_1 \equiv d_1) + P(x_2 \equiv c_1) + P(x_2 \equiv d_1)]$$
$$+\frac{1}{4}[P(x_1 \equiv c_1) + P(x_1 \equiv d_2) + P(x_2 \equiv c_1) + P(x_2 \equiv d_2)]$$
$$+\frac{1}{4}[P(x_1 \equiv c_2) + P(x_1 \equiv d_1) + P(x_2 \equiv c_2) + P(x_2 \equiv d_1)]$$
$$\frac{1}{4}[P(x_1 \equiv c_2) + P(x_1 \equiv d_2) + P(x_2 \equiv c_2) + P(x_2 \equiv d_2)]\}$$
$$= \frac{1}{8}\{[P(x_1 \equiv c_1) + P(x_1 \equiv c_2) + P(x_2 \equiv c_1) + P(x_2 \equiv c_2)]$$
$$+[P(x_1 \equiv d_1) + P(x_1 \equiv d_2) + P(x_2 \equiv d_1) + P(x_2 \equiv d_2)]\}$$
$$= (4r_{XC} + 4r_{XD})/8 = (r_{XC} + r_{XD})/2$$

which is the average coancestry of X with the parents of Y.

A special case occurs when we consider the coancestry of parent X with offspring Z:

$$r_{XZ} = r_{X(XxY)} = (r_{XX} + r_{XY})/2 = (1 + F_X + 2r_{XY})/4$$

If $F_X = 0$ and $r_{XY} = 0$, then the coancestry between parent and offspring is 1/4.

The coancestry between full sibs who have parents A and B is

$$r_{XY} = (2r_{AB} + r_{AA} + r_{BB})/4 = [2r_{AB} + 1(F_A + F_B)/2]/2$$

Another important probability statement is the probability that both alleles in X are IBD with both alleles in Y :

$$u_{XY} = P(x_1 \equiv y_1 x_2 \equiv y_2) + P(x_1 \equiv y_2 x_2 \equiv y_1) = r_{AC}r_{BD} + r_{AD}r_{BC}$$

For full sibs, this reduces to $u_{XY} = r_{AA}r_{BB} + r2AB$, and if $FA = FB = rAB = 0$, then $u_{XY} = 1/4$.

Clearly, uXY = 0 for parent offspring coancestry in a random mating population

NC STATE UNIVERSITY

Wright's definition of Inbreeding coefficient and coefficient of relationship

Wright (1922) defined the inbreeding coefficient of individual z as the correlation between uniting gametes

$$F_Z = \frac{1}{2}\sum_i \frac{1}{2} n_1 + n_2(1 + F_{Ai})$$

where

Ai = the ith common ancestor, $n1$ = number of generations from one parent back to the common ancestor, $n2$ = corresponding number of generations from the second parent.

Wright also defined the coefficient of relationship between X and Y as

$$R_{XY} = \frac{COv(X,Y)}{(V_X V_Y)^{1/2}} = \frac{\sum_i \frac{1}{2} n_1 + n_2(1 + F_{Ai})}{(1 + F_X)^{1/2} + (1 + F_Y)^{1/2}}$$

$$Cov(X,Y) = 2F_Z = 2r_{XY}$$

and if

$$F_X = F_Y = 0$$
$$R_{XY} = 2r_{XY}$$

NC STATE UNIVERSITY

Molecular and Pedigree Based Relationships

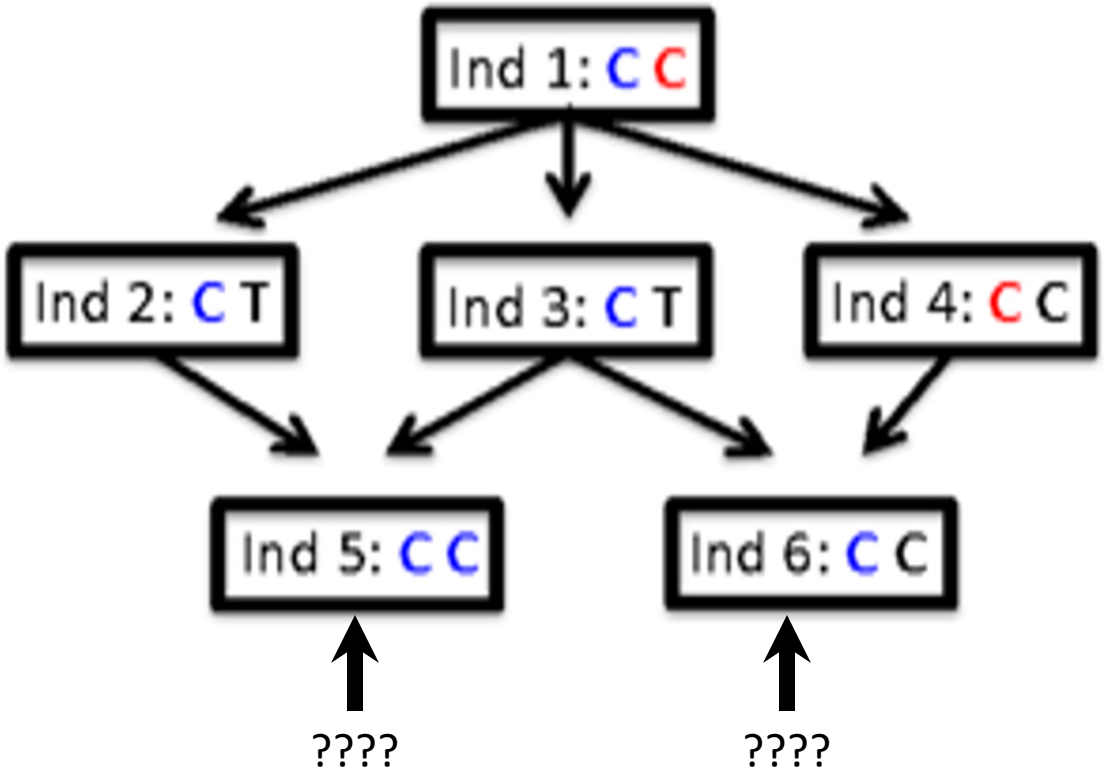
NC STATE UNIVERSITY

Outline

- The tale of two identities.
- Pedigree versus Molecular Relatedness:
  - Expectation versus Realization.
  - Inverse Calculation.
  - Computational Considerations.

IBS-IBD

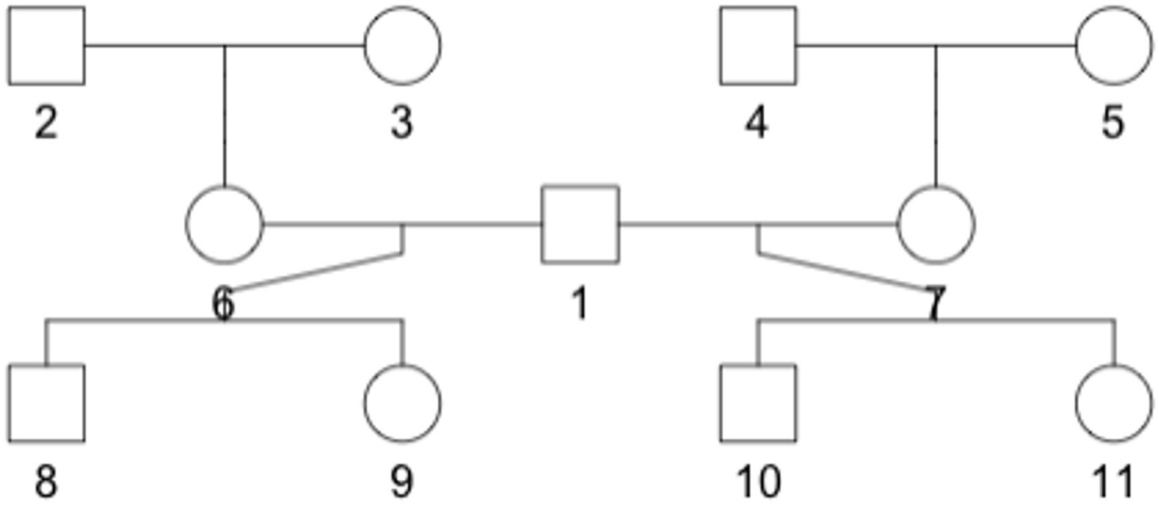
- Genotypes are derived from a random sampling process (Keep this in mind through lecture).
- IBD is a proxy to the true (unknown) IBS!!



Pedigree Relatedness

- Based on the **expected** relationship between and within individuals.
- Estimate of the **IBD** relatedness.
- Generated from a pedigree as outlined below:

Animal	Sire	Dam
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	2	3
7	4	5
8	1	6
..	..	..



Pedigree Relatedness

- Coancestry between two individuals: average coancestry between one individual and the parents of the other.
  - Always true if parents come before progeny!!

Relationship Matrix

	1	2	3	4	5	6	7	8	9	10	11
1	1+0	0	0	0	0	0	0	0.5	0.5	0.5	0.5
2		1+0	0	0	0	0.5	0	0.25	0.25	0	0
3			1+0	0	0	0.5	0	0.25	0.25	0	0
4				1+0	0	0	0.5	0	0	0.25	0.25
5					1+0	0	0.5	0	0	0.25	0.25
6						1+0	0	0.5	0.5	0	0
7							1+0	0	0	0.5	0.5
8								1+0	0.5	0.25	0.25
9									1+0	0.25	0.25
10										1+0	0.5
11											1+0

Estimate of Expected Relationship!!!

Pedigree Relatedness

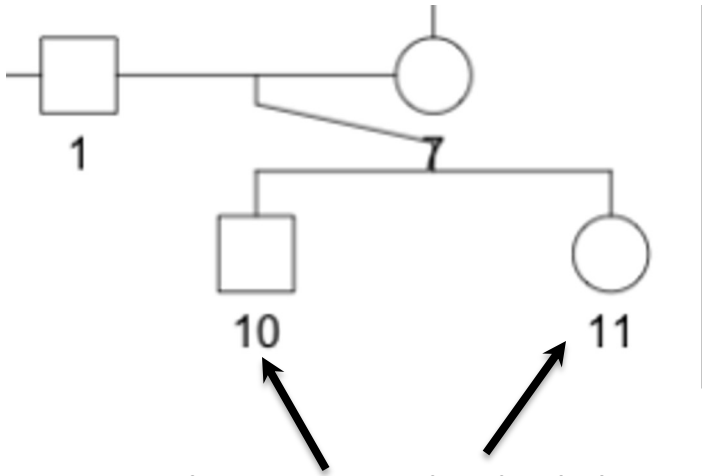
- The primary motivation in using a relationship matrix is to derive breeding values for individuals.

- Only need inverse though!!

- Henderson (1975) generated a recursive way to generate A^{-1} without have to do it brute force (i.e. inverse of A).
  - Faster methods by Quass (1976) and Meuwissen & Luo (1992).
  - Millions of animals is computationally efficient!
  - Sparsity!!

Pedigree Relatedness

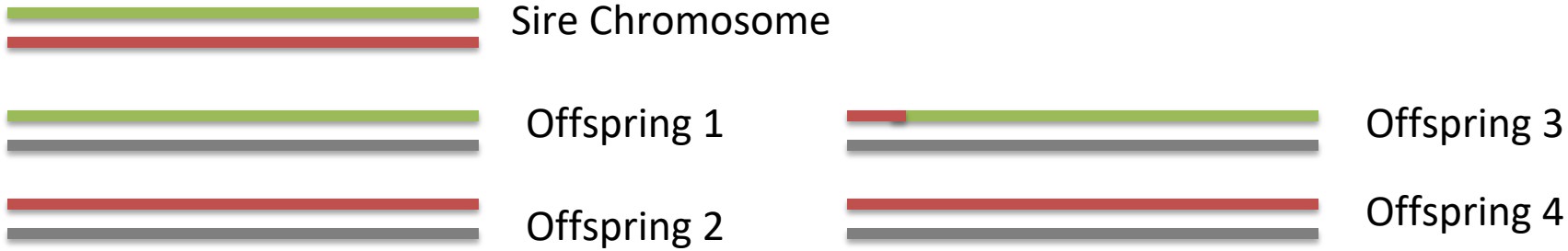
- Points on a pedigree based relationship matrix:
  - All relative to base/founder population.
  - Founder individuals relatedness.
  - Depth of pedigree.
  - What to do with multi-breed population.
  - Does not account for Mendelian Sampling!!



Do these two individuals have the same estimated breeding value?

Pedigree Relatedness

- Does not take in to account the variation in expected relationships, which is partly due to Mendelian Sampling (MS).
  - MS is a **sample from a random process** of the transmission of alleles from parent to offspring.
    - Relationship = (Expected Relationship) ± MS.
  - Sets an upper limit to the amount of information you can get from an offspring without any phenotypes!
    - Max accuracy is equal ???.



Mendelian Sampling Effect

- Example using a single locus:
  - For every A allele add a +1 to the breeding value.
  - For every a allele add a -1 to the breeding value.

		Dam	
		A	a
Sire	A	AA	Aa
	A	AA	Aa

Genomic Information
EBV = Σ allelic effects

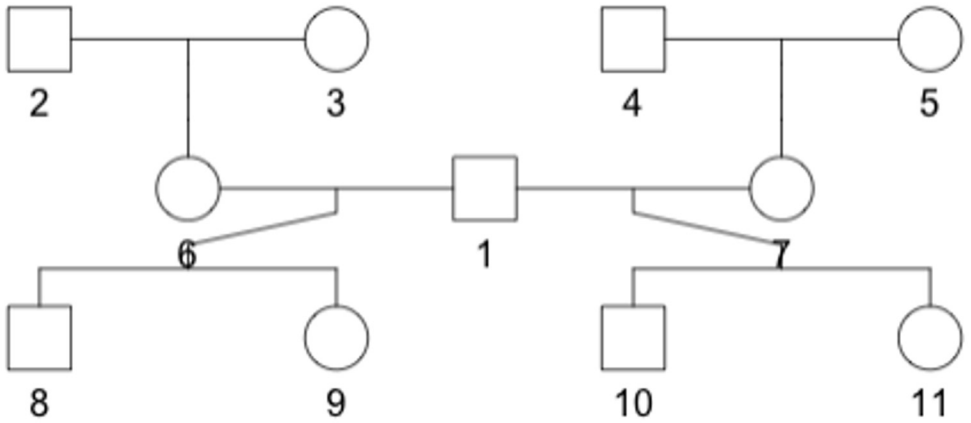
		Dam	
		A	a
Sire	A	+2	0
	A	+2	0

Pedigree Information
$EBV = 0.5*(EBV_{Sire} + EBV_{Dam})$

		Dam	
		A	a
Sire	A	+1	+1
	A	+1	+1

Molecular Relatedness

- Based on the **realized** relationship between and within individuals (i.e. Realization of the sampling process)
  - E(Molecular) = Pedigree
- Estimate of the **IBS** relatedness.
  - Proportion of shared genotypes.
- Generated from molecular data:
  - SNP, Haplotypes, Microsatellites



Molecular Relatedness

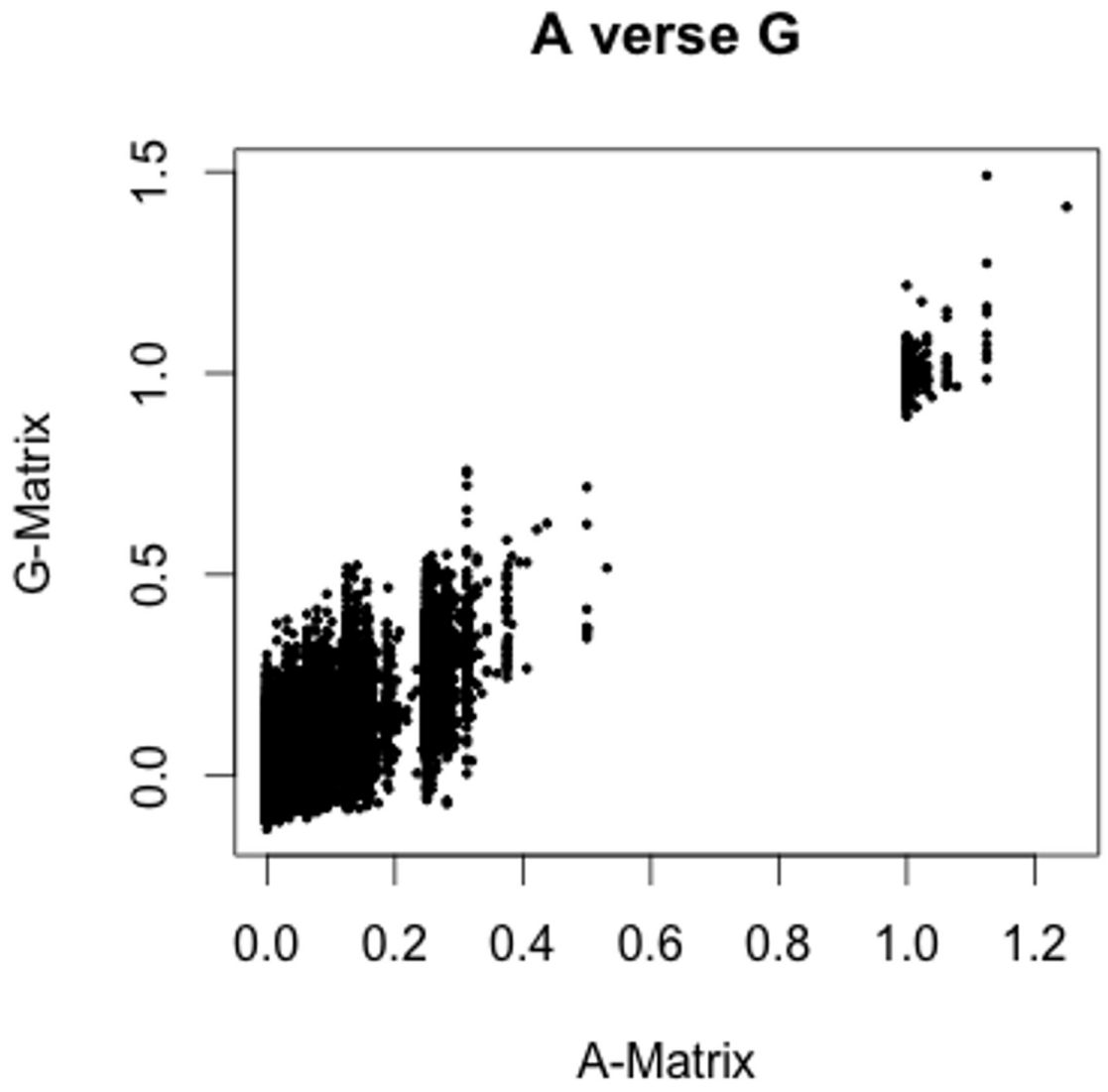
- Disregard how it is set up at this point.
- Measure of proportion of alleles shared.

Relationship Matrix

	1	2	3	4	5	6	7	8	9	10	11
1	0.96	0.12	0.15	0.43	0.12	0.07	0.05	0.5	0.5	0.5	0.5
2		0.98	0.05	0.02	0.13	0.5	0.02	0.23	0.29	0.09	0.05
3			1.02	0.08	0.05	0.5	0.06	0.28	0.22	0.4	0.03
4				1.01	0.02	0.03	0.5	0.02	0.02	0.22	0.22
5					1.05	0.04	0.5	0.07	0.09	0.29	0.27
6						1.02	0.06	0.5	0.5	0.03	0.05
7							1.01	0.06	0.07	0.5	0.5
8								1.02	0.49	0.22	0.27
9									0.98	0.28	0.22
10										0.93	0.53
11											1.04

Estimate of Realized Relationship!!!

Molecular Relatedness



Molecular Relatedness

- Points on a molecular based relationship matrix:
  - Founders are now related.
  - Accounts for Mendelian Sampling.
  - Issues related to SNP ascertainment bias.
  - Assume SNP are independent.
    - What about LD?
    - Is the amount of information generated from each SNP the same?
    - Does the information generated increase as the number of SNP increase?
  - Genotyping Errors
  - Sometimes not invertible
  - No longer Sparse

Relationship Construction

- Multiple ways to construct, but all in some form derive from this setup of:
  - W_{np} = genotype for individual “n” for marker “p”.
  - Matrix is sometimes referred to as “gene content”.
  - Dimension is number of animals by number of snp.

$$\begin{pmatrix} W_{11} & \cdots & W_{1p} \\ W_{21} & \cdots & W_{2p} \\ \vdots & & \vdots \\ W_{n1} & \cdots & W_{np} \end{pmatrix} \cdot \begin{pmatrix} W_{11} & W_{21} & \cdots & W_{n1} \\ \vdots & \vdots & & \vdots \\ W_{1p} & W_{2p} & \cdots & W_{np} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{np} W_{1i}^2 & \sum_{i=1}^{np} W_{1i} W_{2i} \\ \sum_{i=1}^{np} W_{2i} W_{1i} & \sum_{i=1}^{np} W_{2i}^2 \end{pmatrix}$$

Diagonals – Within animal summation
Off-diagonals – Across animal summation

Does SNP Coding Matter?

W(aa)	= -1
W(Aa/aA)	= 0
W(AA)	= 1

$E(W) = (p-q)$

$Var(W) = 2pq$

W(aa)	= 0
W(Aa/aA)	= 1
W(AA)	= 2

$E(W) = 2p$

$Var(W) = 2pq$

Variance Equal but mean shifts??

Traditional Way to Construct in Animal Breeding

- Set up Z which is a matrix of gene content that has been centered to set mean allele effects to 0.

Coding (0,1,2)

Z(aa) = 0 - 2p = -2p

Z(Aa/aA) = 1 - 2p = 1-2p

Z(AA) = 2 - 2p = 2(1-p)

Coding (-1,0,1)

Z(aa) = -1 - (p-q) = -2p

Z(Aa/aA) = 0 - (p-q) = 1-2p

Z(AA) = 1 -(p-q) = 2(1-p)

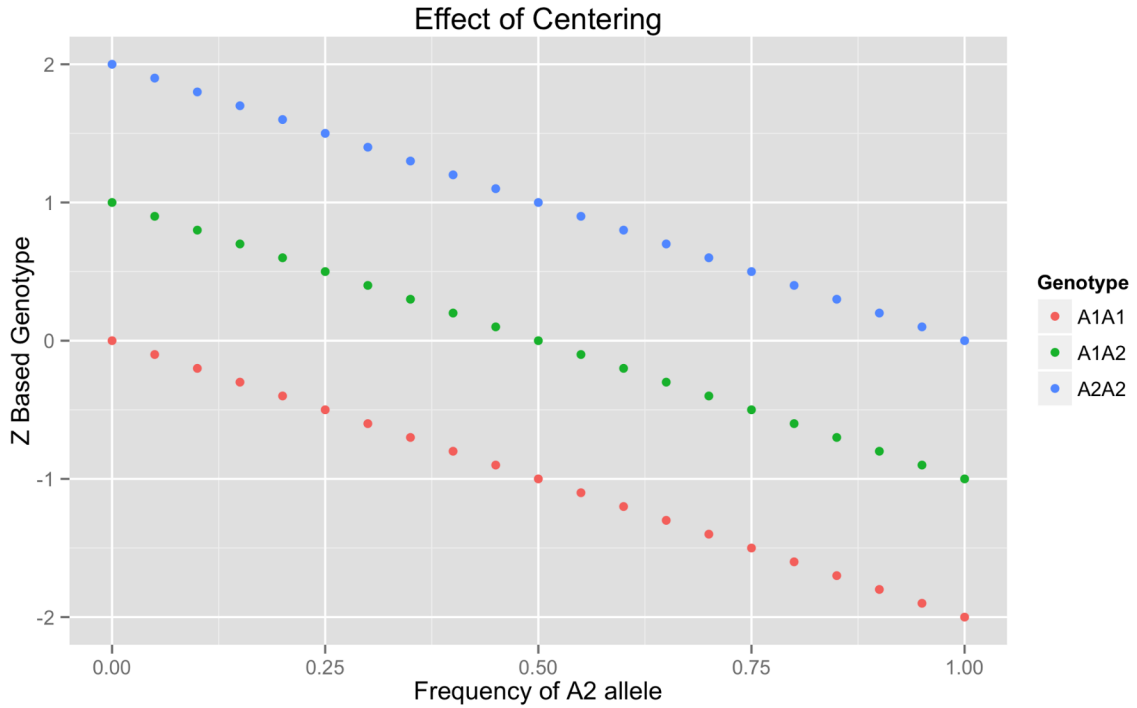
Van-Raden (2008)

Z(aa) = -1 - 2(p-0.5) = -2p

Z(Aa/aA) = 0 - 2(p-0.5) = 1-2p

Z(AA) = 1 - 2(p-0.5) = 2(1-p)

Traditional Way to Construct in Animal Breeding



- Centering:
  - More credit to rare alleles than to common ones when calculating off-diagonal relationships.
  - Genomic Inbreeding greater if individual is homozygous for rare alleles.

Traditional Way to Construct in Animal Breeding

- Once Z is created then perform the following:

$$G = ZZ' / 2 \Sigma p_i(1-p_i)$$

- Denominator scales so it is similar to A matrix.
- Properties based on HWE and Linkage Equilibrium:
  - Average Diagonals = 1.0
  - Average Off-Diagonals = 0.0

Molecular Relatedness

- As with using pedigree information the primary motivation in using a relationship matrix is to derive breeding values for individuals.

- The matrix is dense in comparison to a pedigree based matrix.
- Costly to invert
  - Misztal et al. (2014) developed a recursion or Meyer et al. (2014) developed a way to update inverse as new animals enter.

Pedigree + Molecular Relatedness

- What if you have individuals with both pedigree and genomic information??
  - Compute Pedigree based and Molecular based seperately and combine them.
  - Combine Pedigree and Genomic into one matrix (H) and is referred to as Single Step Genomic Evaluation.
    - Good paper with example pedigree:
      - Legarra et al. (2009); A relationship matrix including full pedigree and genomic information.

NC STATE UNIVERSITY

Characterization and management of homozygosity: the livestock perspective

NC STATE UNIVERSITY

Homozygosity: Difference between IBS v. IBD

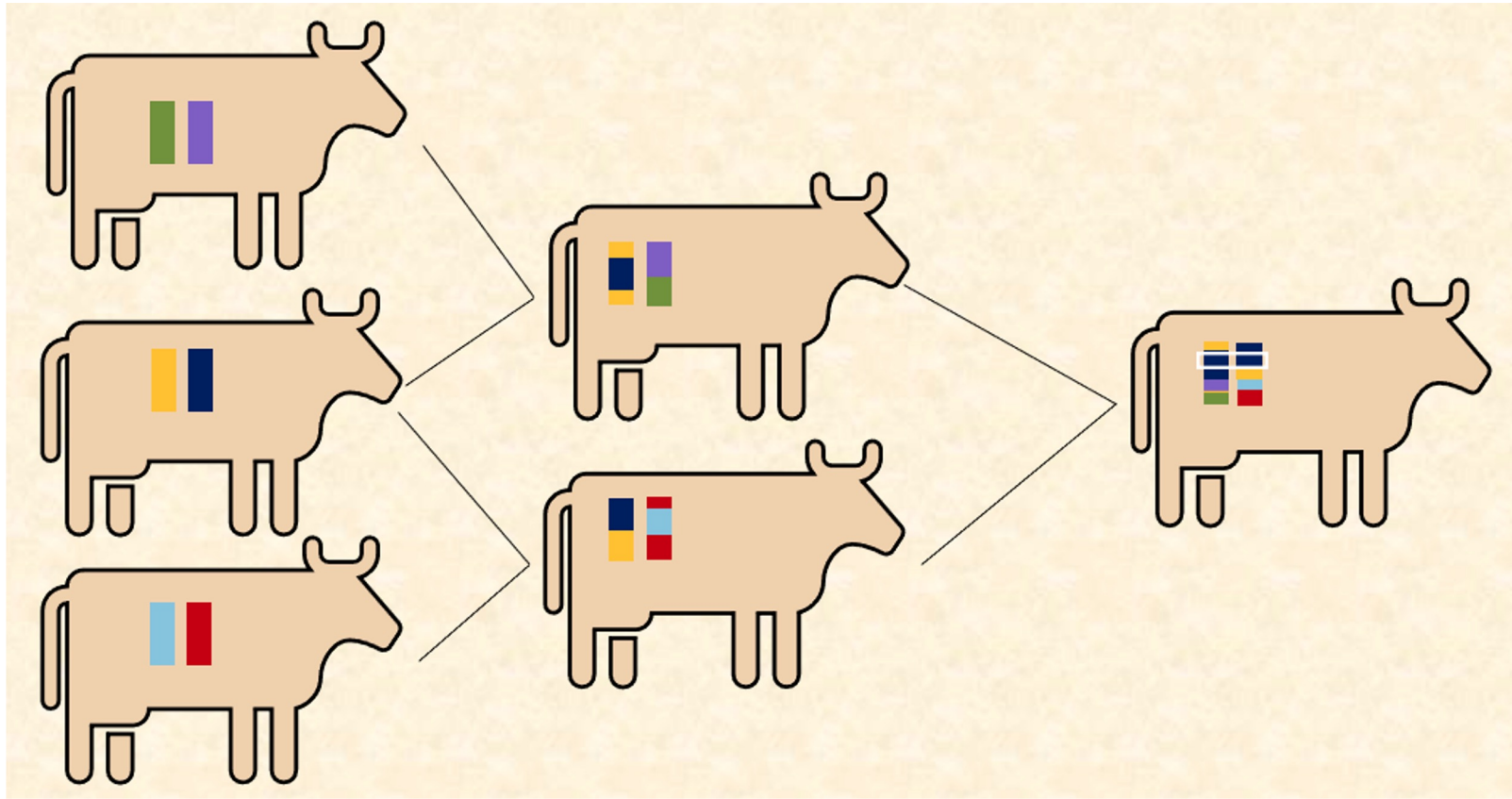
- Identity-by-state (**IBS**) = Two genes that have identical nucleotide sequences but may or may not have descended from different copies in the ancestral population.

- Identity-by-descent (**IBD**) = Identical nucleotide sequences descended from same ancestral copy.

NC STATE UNIVERSITY

Inbreeding

- Inbreeding results from the mating of related individuals (share one or more ancestors).
- The inbreeding coefficient of an individual is defined as the probability of 2 randomly chosen alleles at the same locus being IBD.



NC STATE UNIVERSITY

Pedigree inbreeding (F_{PED})

- Numerator relationship matrix (A)
  - Covariance matrix that is an estimate of the expected proportion of genes shared by individuals in the pedigree.
- Multiple methods have been developed to construct A and its inverse from pedigree information:
  - Meuwissen and Luo (1992)
  - Colleau (2002)
  - VanRaden (1992)

NC STATE UNIVERSITY

Drawbacks of using pedigree

1. Need to collect pedigree information
   - Wild populations?

1. Susceptible to missing/inaccurate information

1. It is only an expectation
   - Mendelian sampling

NC STATE UNIVERSITY

Genomic measures of inbreeding

- Largely fall into:
  - Marker-based measures
    - F_{HOM}
    - F_{GRM}
  - Segment based measures
    - F_{ROH}

NC STATE UNIVERSITY

Proportion of homozygous markers (F_{HOM})

-
- $F_{HOM} = \frac{\#\ of\ markers\ in\ homozygous\ state}{Total\ \#\ of\ markers}$

NC STATE UNIVERSITY

Genomic relationship matrix

- Using genomic data, we can build can built a matrix like A but using relationships built on the proportion of alleles shared between individuals.

- This matrix is called the genomic relationship matrix (G or GRM).
  - Realized relationship

NC STATE UNIVERSITY

Inbreeding using genomic relationship matrix (F_{GRM})

- Different ways to build G but we will focus on the first method laid out by VanRaden (2008)

$$G = \frac{(P - M)'(P - M)}{2\sum p_i(1 - p_i)}$$

- Where p is the frequency of the second allele at locus i, M is n by m matrix of marker genotypes, P is a matrix of allele frequencies expressed as a difference from 0.5 and multiplied by 2.

- F_{GRM} for animal j is obtained as $G_{jj} - 1$.

NC STATE UNIVERSITY

Drawbacks of using F_{GRM}

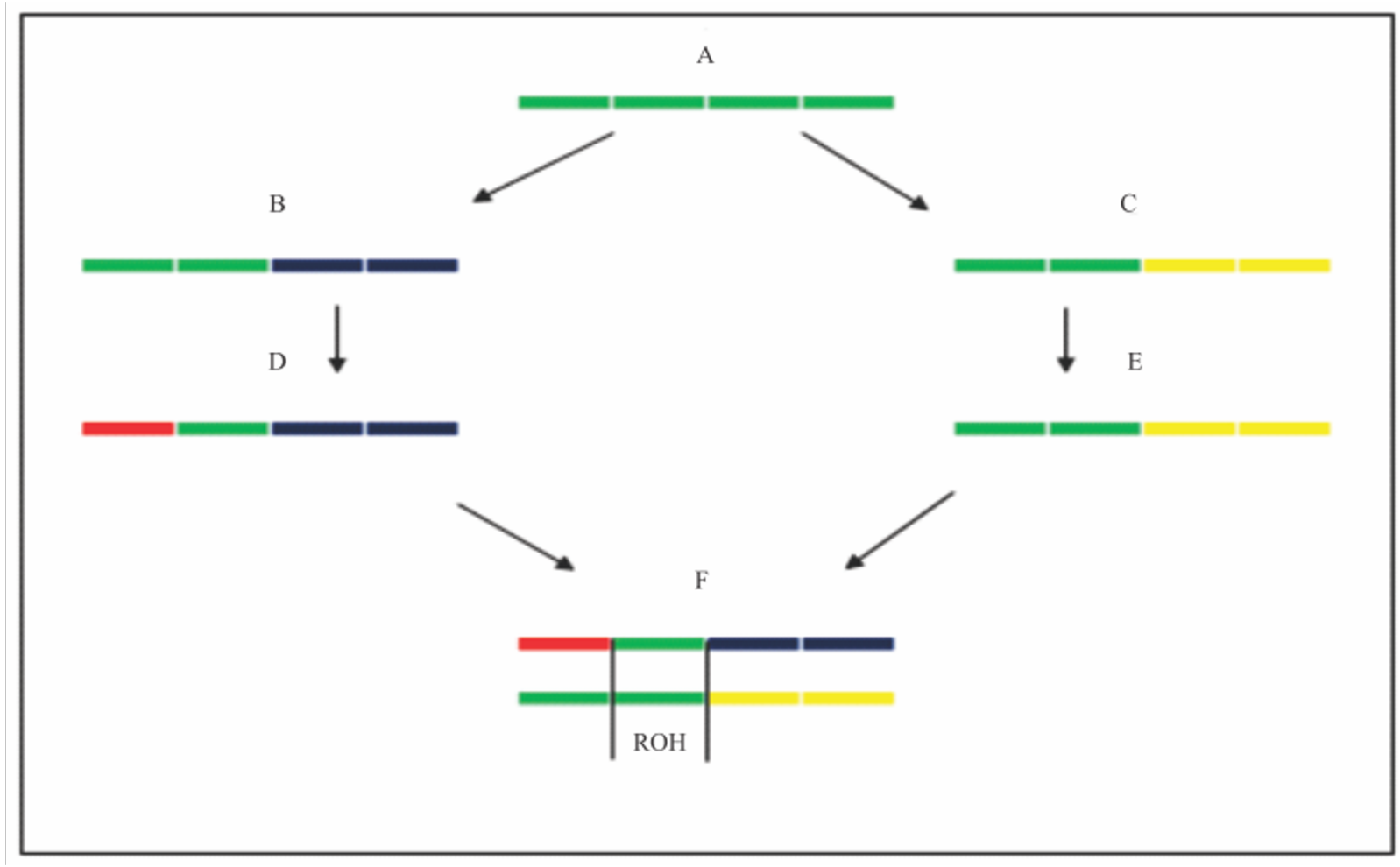
- Many methods to build G and different interpretations.
- Cannot distinguish between IBS and IBD homozygosity and therefore will tend to overestimate inbreeding.

NC STATE UNIVERSITY

Runs of homozygosity

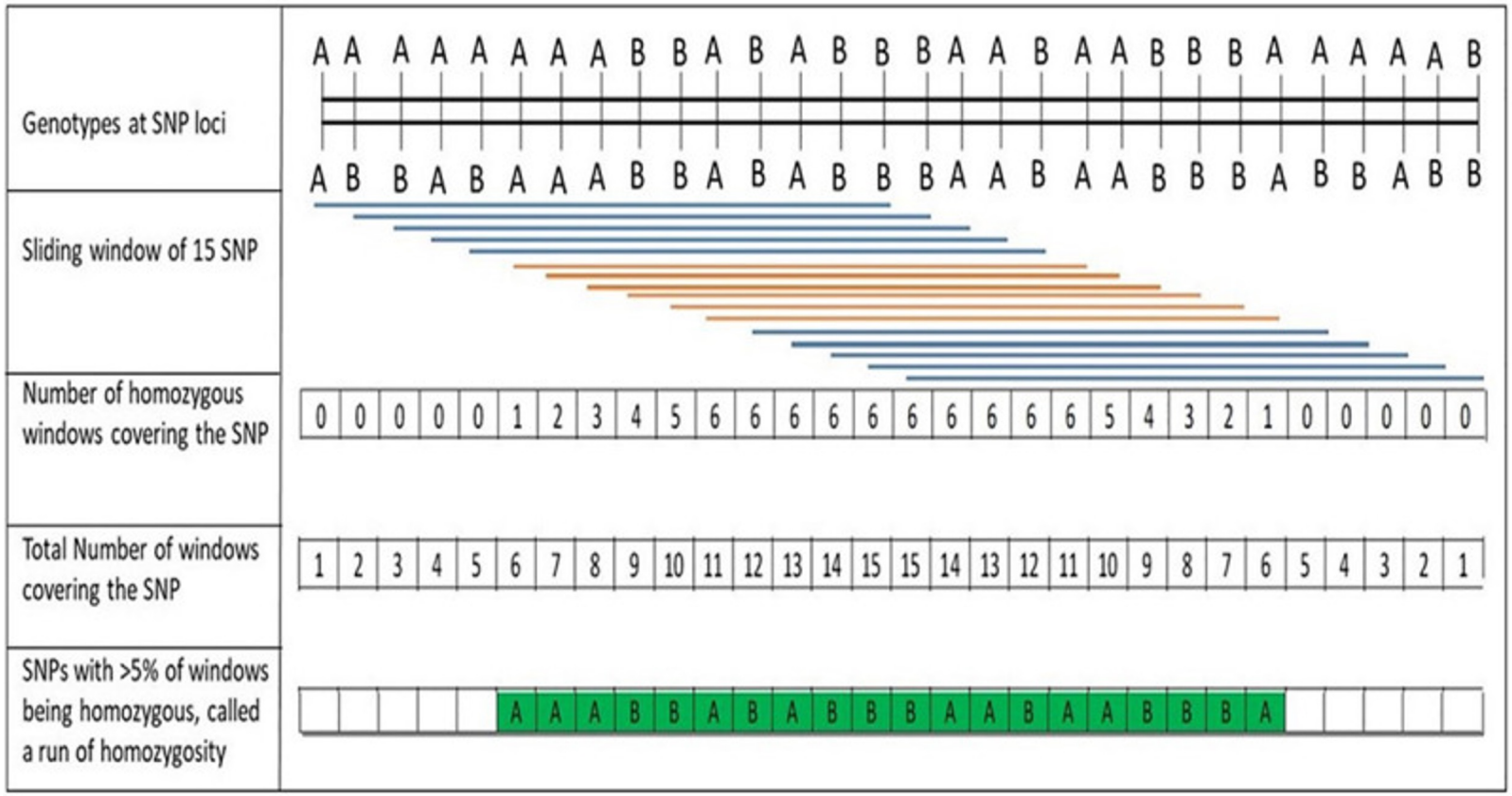
- Continuous stretches of the genome in a homozygous state are known as runs of homozygosity.

- Large homozygous segments are unlikely to be homozygous by chance.


A
B
C
D
E
F
ROH

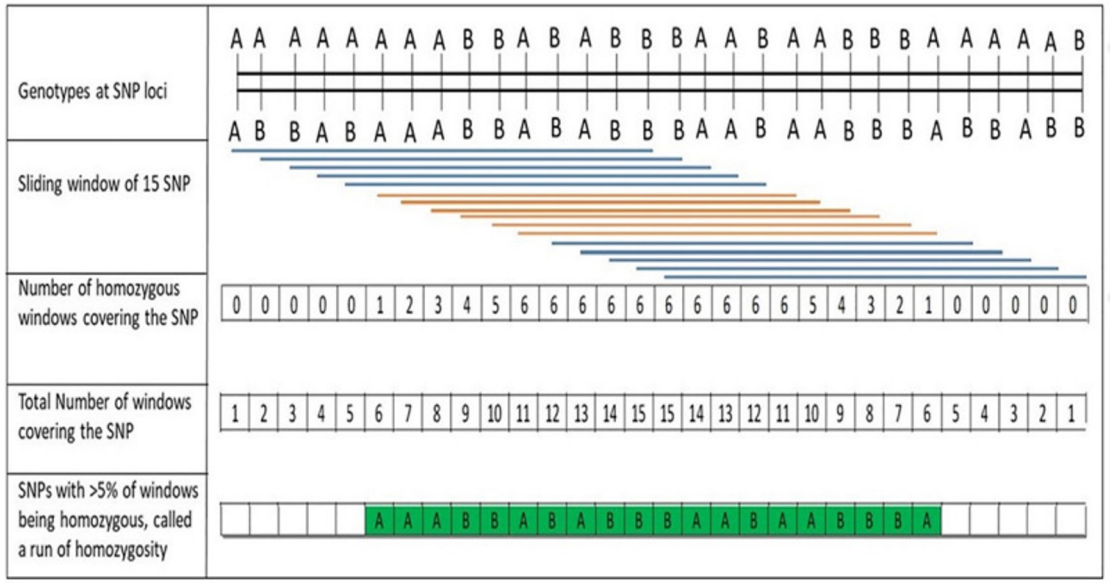
NC STATE UNIVERSITY

Detecting ROH using a sliding window approach



NC STATE UNIVERSITY

PLINK algorithm for detecting ROH

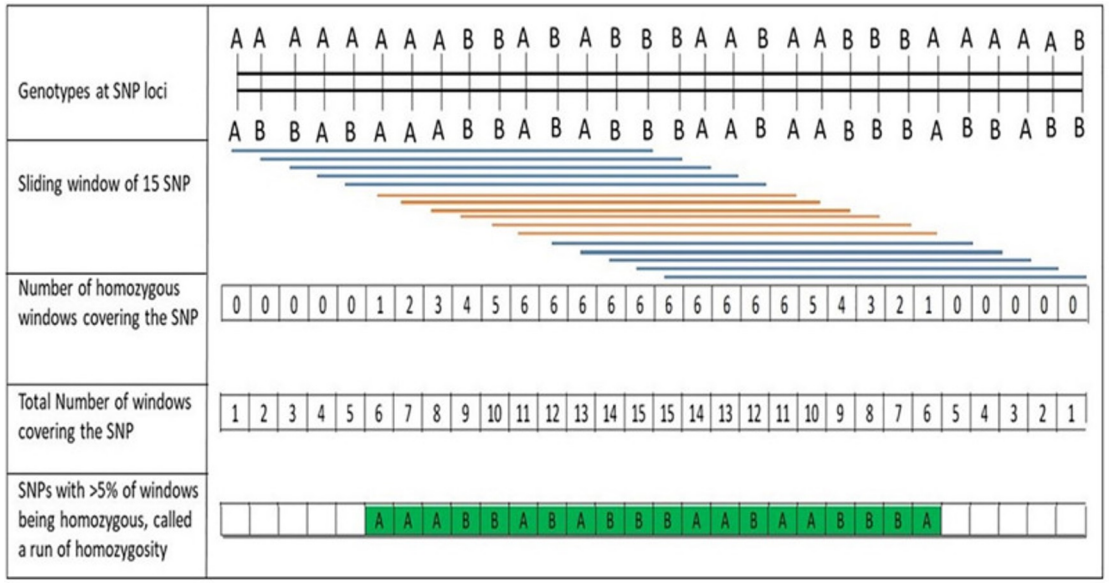


1) A scanning window is defined by a predefined number of SNPs.
- # of heterozygous SNPs
- # of missing

Then the number of homozygous windows the SNP is a part of is counted.

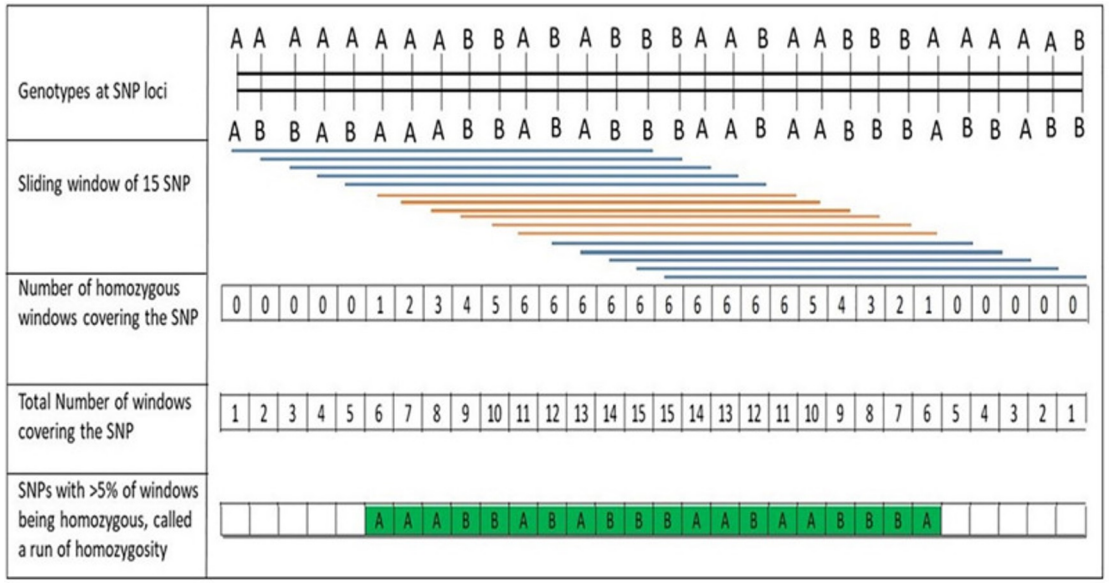
NC STATE UNIVERSITY

PLINK algorithm for detecting ROH



2) Then a threshold is imposed where the SNP needs to be a part of X amount of homozygous windows

PLINK algorithm for detecting ROH



3) ROH segments are constructed and checked against further constraints:

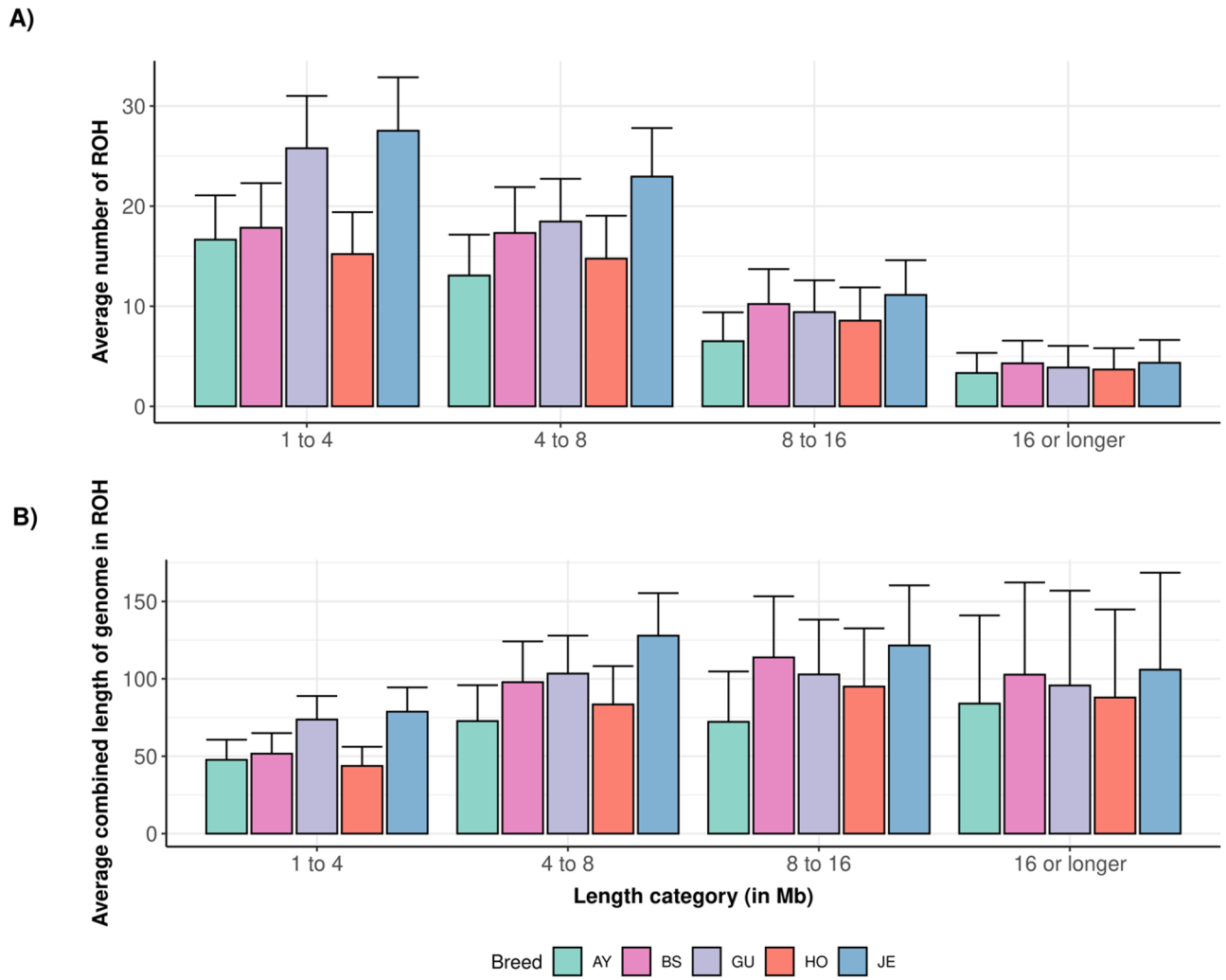
- Interval between SNPs
- Minimum number of SNPs in segment.
- Minimum segment length

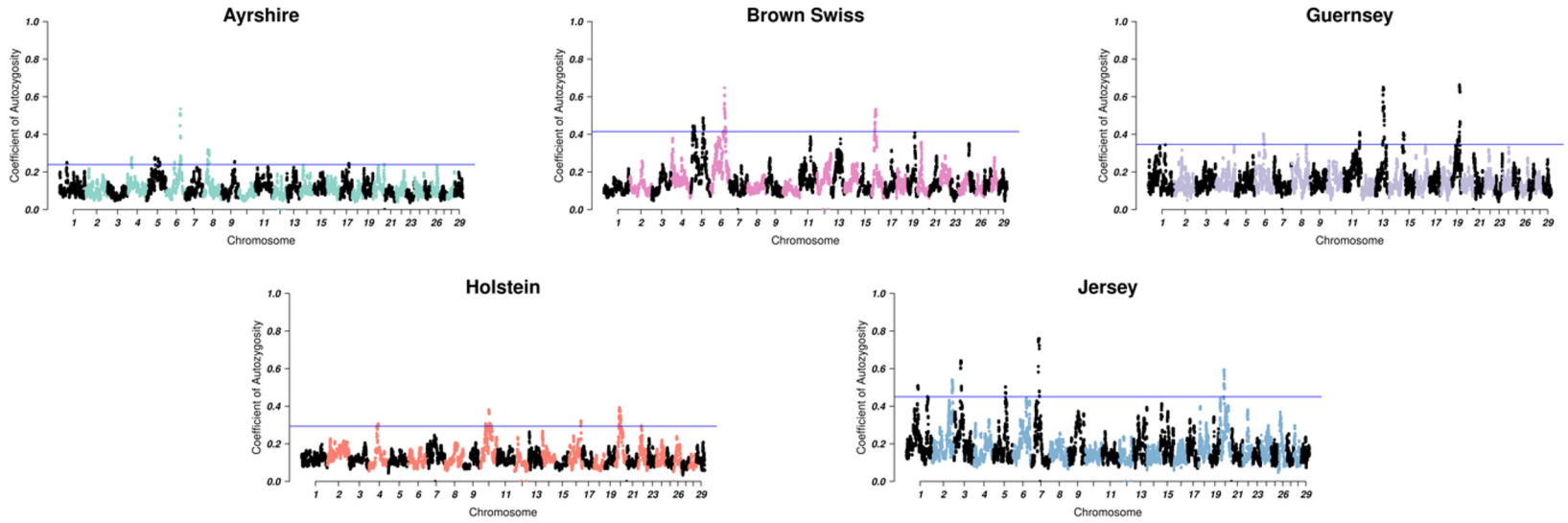
NC STATE UNIVERSITY

Characteristics of ROH

- Size of an ROH
  - The length of an ROH is expected to be related to the number of generations that have passed since the common ancestor.
- Distribution of ROH
  - Recombination
  - Selection

NC STATE UNIVERSITY




Ayrshire
Brown Swiss
Guernsey
Holstein
Jersey
Coefficient of Autozygosity
Chromosome
0.0
0.2
0.4
0.6
0.8
1.0
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 23 26 29

NC STATE UNIVERSITY

Proportion of the genome in ROH (F_{ROH})

- $$F_{ROH_i} = \frac{\sum L_{ROH_i}}{L}$$

- Where F_{ROH} is the genomic inbreeding of the ith individual using ROH; L_{ROH} is the combined length of the detected ROH in the ith individual; and L is the length of the genome covered by the array.

NC STATE UNIVERSITY

Inbreeding depression

- Reduction in the mean phenotypic value of a trait in a population.

- Assuming no epistasis, the change due to inbreeding on the phenotypic mean will be:

$$M_1 = M_0 - 2F \sum_{i=1}^{n} p_i q_i d_i$$

NC STATE UNIVERSITY

Working theories

1. Increased expression of deleterious recessive alleles.
   - As inbreeding increases the frequency of deleterious recessives increases and presents effect hidden in heterozygotes.

2. Homozygosity at loci where there is heterozygote advantage (overdominance)
   - As inbreeding increases the number of heterozygous genotypes with positive effect decrease.

NC STATE UNIVERSITY

Inbreeding depression examples:

Litter Size and lifespan

Growth and survival

Milk production and reproduction

NC STATE UNIVERSITY

Is all inbreeding the same?

- Age of inbreeding

- Location of inbreeding

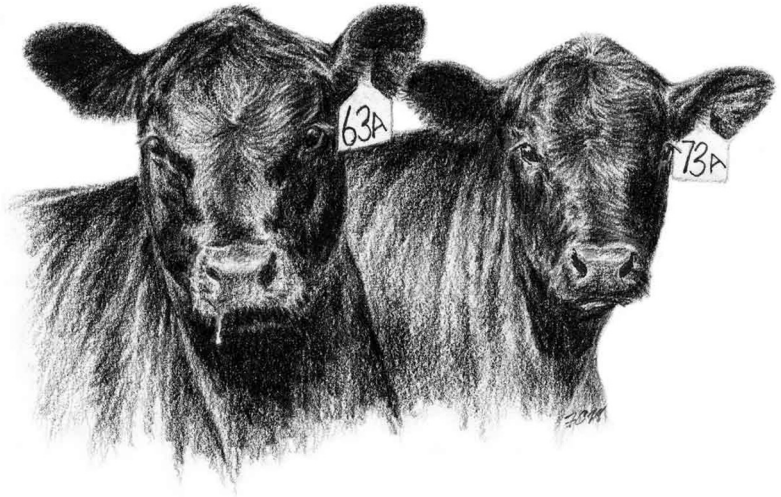
NC STATE UNIVERSITY

Case study: Inbreeding depression in American Angus cattle

- Objectives:

1. Characterize the American Angus population in terms of pedigree and genomic inbreeding levels.

1. Quantify the effect of pedigree and genomic inbreeding, as well as any moderating effects of the age of inbreeding



NC STATE UNIVERSITY

Data

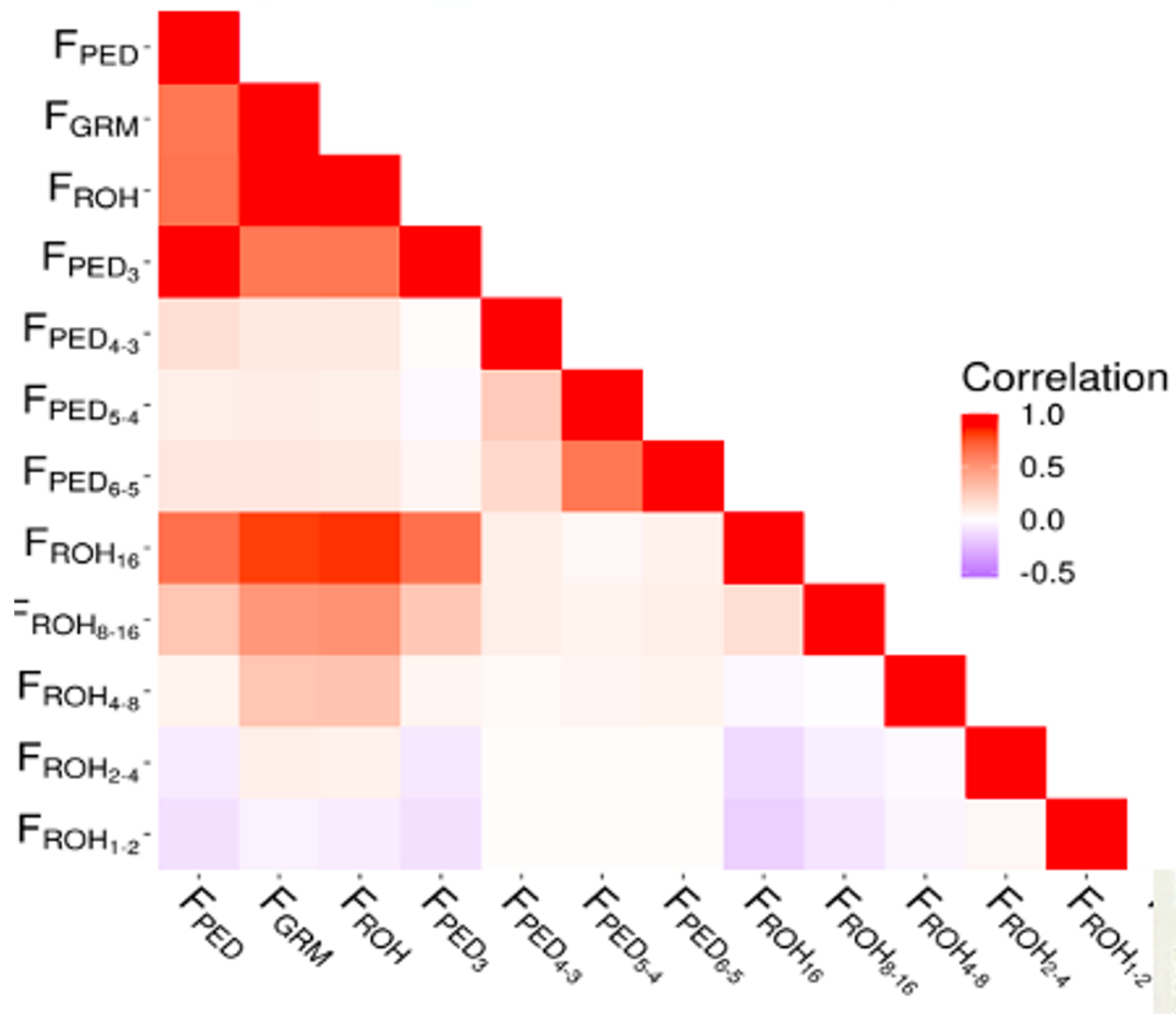
- Pedigree and genomic (BovineSNP50k v2 BeadChip) data was provided for 567,475 animals of the American Angus breed.

- Phenotypic records for heifer pregnancy (HP), birth weight (BW), weaning weight (WW), and Postweaning gain (PWG)

NC STATE UNIVERSITY

Inbreeding coefficients

- Inbreding measures considered:
  - F_{PED}
    - In order to decompose F_{PED} into age classes, we also calculated pedigree inbreeding based on the first 3 ($\mathbf{F_{PED}}$), 4 ($\mathbf{F_{PED4}}$), 5 ($\mathbf{F_{PED5}}$), and 6 ($\mathbf{F_{PED6}}$) ancestral generations.
  - F_{GRM}
  - F_{ROH}
    - In order to decompose F_{ROH} into age classes, we also calculated ROH inbreeding based on ROH of lengths 1 to 2 Mb ($F_{ROH1\text{-}2}$), 2 to 4 Mb ($F_{ROH2\text{-}4}$), 4 to 8 Mb ($F_{ROH4\text{-}8}$), 8 to 16 ($F_{ROH8\text{-}16}$), and 16 Mb or larger (F_{ROH16}).


F_{PED}
F_{GRM}
F_{ROH}
F_{PED_3}
$F_{PED_{4-3}}$
$F_{PED_{5-4}}$
$F_{PED_{6-5}}$
$F_{ROH_{16}}$
$F_{ROH_{8-16}}$
$F_{ROH_{4-8}}$
$F_{ROH_{2-4}}$
$F_{ROH_{1-2}}$
Correlation
1.0
0.5
0.0
-0.5

NC STATE UNIVERSITY

Effect (as % of trait mean) of a 1% increase in inbreeding.

Trait	Group	F_{PED}	F_{GRM}	F_{ROH}
BiW	Males	-0.08	-0.11	-0.11
	Females	-0.09	-0.15	-0.12
WW	Males	-0.16	-0.20	-0.16
	Females	-0.17	-0.21	-0.18
PWG	Males	-0.28	-0.32	-0.26
	Females	-0.32	-0.36	-0.30

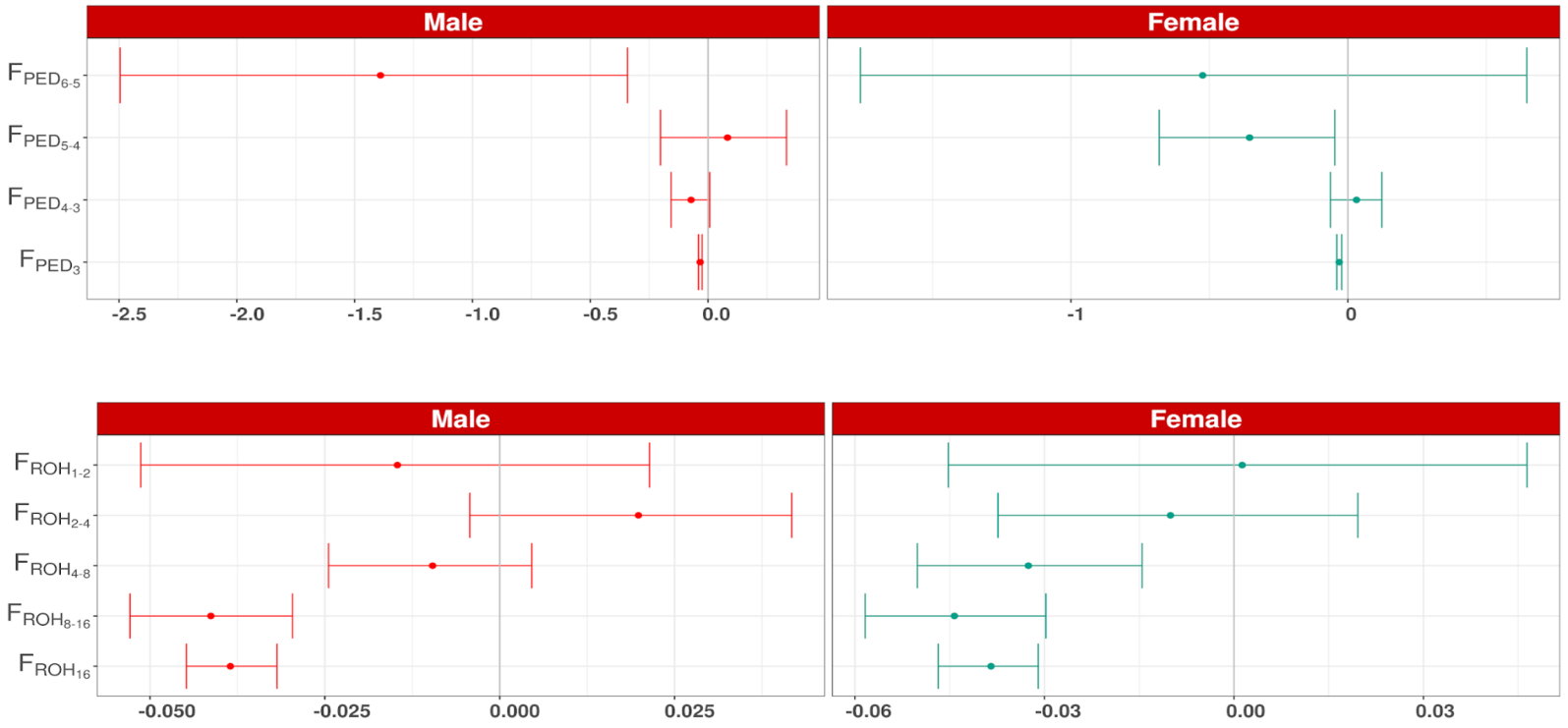
NC STATE UNIVERSITY

Projected phenotypic depression for animals with low (5th perc.) and high (95th perc.) pedigree and genomic inbreeding.

Trait	Group	F_{PED}			F_{GRM}			F_{ROH}		
		Low	High	Difference	Low	High	Difference	Low	High	Difference
BiW (kg)	Males	−0.12	−0.33	0.21	−0.15	−0.40	0.25	−0.13	−0.34	0.22
	Females	−0.11	−0.31	0.20	−0.17	−0.47	0.31	−0.14	−0.39	0.25
WW (kg)	Males	−1.75	−4.78	3.03	−2.15	−5.88	3.73	−1.81	−4.94	3.13
	Females	−1.66	−4.74	3.08	−2.07	−5.90	3.83	−1.74	−4.96	3.22
PWG (kg)	Males	−2.28	−6.10	3.82	−2.56	−6.86	4.30	−2.10	−5.62	3.52
	Females	−1.20	−3.36	2.16	−1.51	−4.22	2.71	−1.23	−3.45	2.22

NC STATE UNIVERSITY

Effect of inbreeding age on depressive effects



NC STATE UNIVERSITY

Genetic diversity

- Genetic diversity is the presence of variation in the genetic composition among individuals in a group.

NC STATE UNIVERSITY

Idealized Population

1. Random mating
1. Distinct non-overlapping generations
1. No migration
1. No mutation
1. No selection
1. Constant population size

NC STATE UNIVERSITY

Idealized population

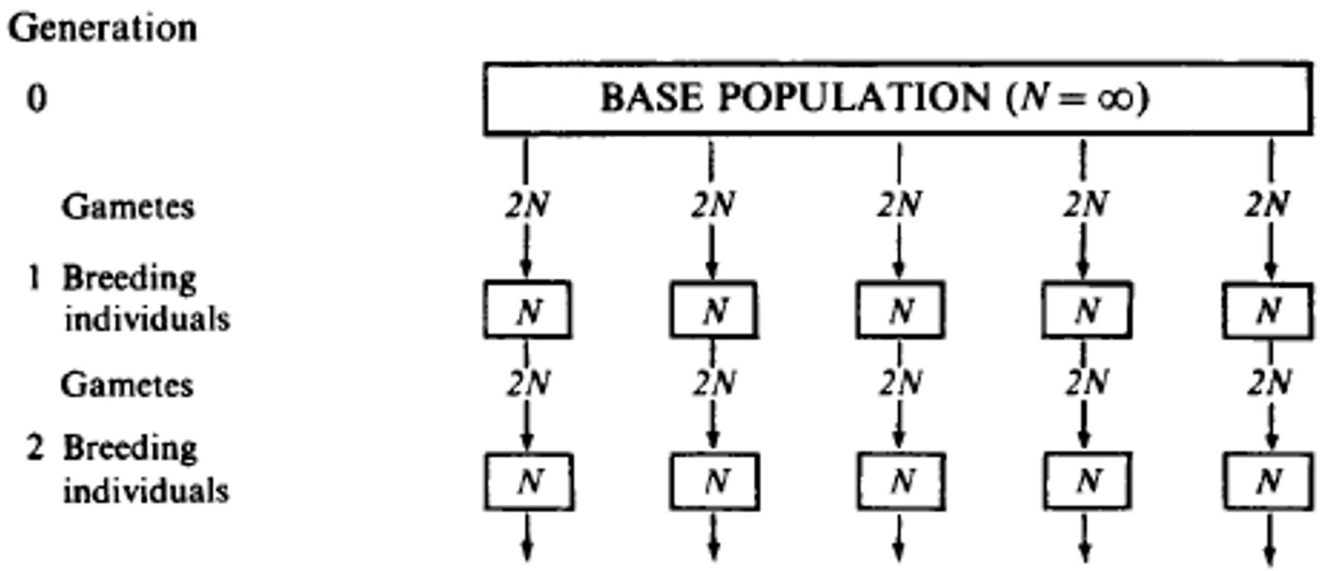


Fig. 3.1. Diagrammatic representation of the subdivision of a single large population – the base population – into a number of sub-populations, or lines.

NC STATE UNIVERSITY

Rate of Inbreeding in an idealized population

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1}$$

ΔF

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$$

NC STATE UNIVERSITY

Effective population size (N_e)

- The effective population size (N_e) is defined as the number of individuals that would give rise to the rate of inbreeding if they were bred in the manner of an ideal population.

$$N_e = \frac{1}{2\Delta F}$$

- Other definitions of effective size have been defined for cases of differing numbers of males and females, unequal number of individuals in successive generations, etc.

NC STATE UNIVERSITY

Table 1 Genealogical parameters and effective population sizes for the 140 breeds studied averaged for each species

					IBD methods					
Species	**Breed nb**	***T***	***EqG***	F_{IS} %	N_{eCi}	N_{eCt}	N_{eFi}	N_{eFt}	N_{es}	N_{ev}
Cattle	20	5.4 [4-7.2]	6.1 [3.4-8.3]	-0.45 [-1.87-1.44]	245 [55-958]	91 [27-242]	182 [58-646]	100 [35-204]	21,648 [208-133056]	934 [108-4420]
Sheep	40	3.6 [2.9-4.1]	6.0 [2.6-10.3]	-0.37 [-4.28-2.44]	189 [28-429]	68 [18-142]	191 [38-675]	95 [21-375]	1502 [30-13736]	407 [46-1812]
Horse	20	9.6 [6.8-13.7]	4.4 [1.8-7.6]	-0.1 [-1.98-2.39]	184 [33-520]	175 [44-799]	135 [22-321]	125 [33-257]	1906 [111-6349]	487 [53-2022]
Dog	60	4.1 [2.7-5.1]	5.8 [3-9.2]	1.37 [-2.87-4.7]	204 [21-692]	241 [17-1451]	89 [22-392]	80 [15-510]	1472 [37-6041]	471 [35-1443]
Total	140	4.9 [2.7-13.7]	5.7 [1.8-10.3]	0.41 [-4.28-4.7]	203 [21-958]	160 [17-1451]	138 [22-675]	93 [15-510]	4425 [30-133056]	356 [35-4420]

IBD = identity by descent; nb = number; *T* = average generation length in years; *EqG* = number of equivalent generations; F_{IS} = fixation index; N_{eCi} = method based on individual coancestry rate; N_{eCt} = method based on coancestry rate between two successive generations; N_{eFi} = method based on individual inbreeding rate; N_{eFt} = method based on inbreeding rate between two successive generations ; N_{es} = N_e method based on sex ratio; N_{ev} = method based on variance of progeny size; in brackets, minimal and maximal values.

What is a good effective population size?

NC STATE UNIVERSITY

Genetic diversity case study: American dairy cattle

Objective:

- Assess the current state of genetic diversity and changes in genetic diversity due to recent selection strategies.

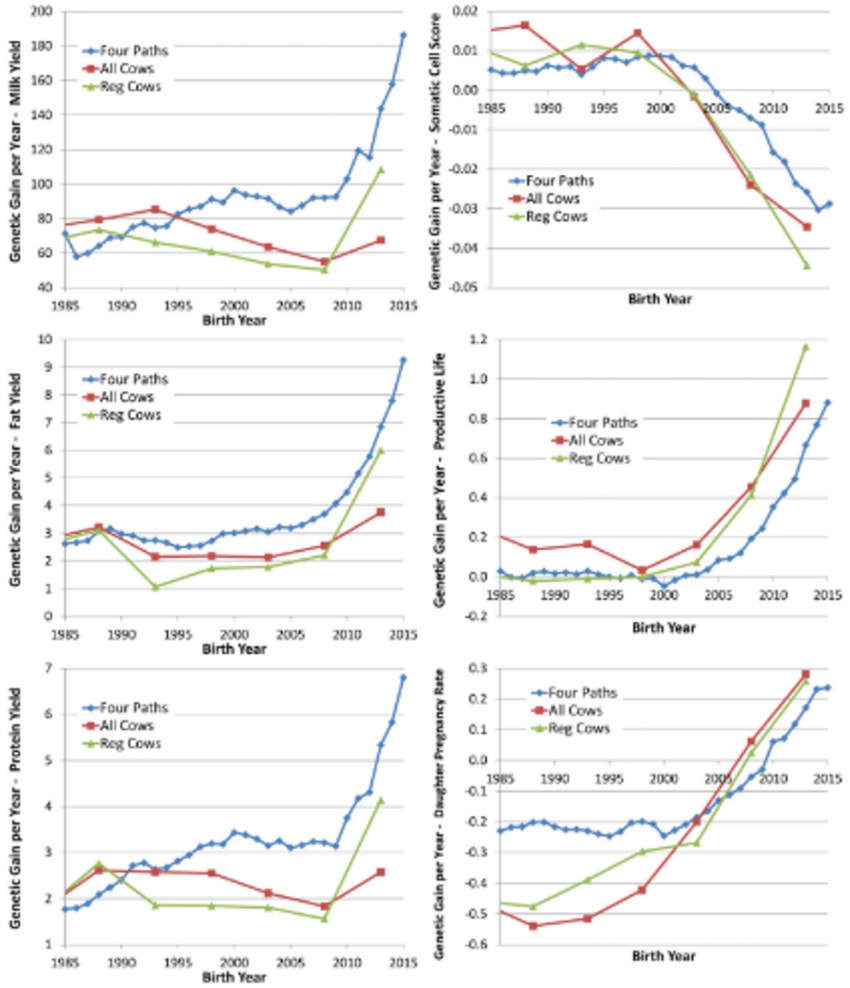
NC STATE UNIVERSITY

Genomic selection in dairy cattle

- What is genomic selection?
- When was it implemented in dairy cattle?
- What were the consequences for genetic diversity?

NC STATE UNIVERSITY

Genomic selection has considerably increased the rate of genetic gain



NC STATE UNIVERSITY

Animals

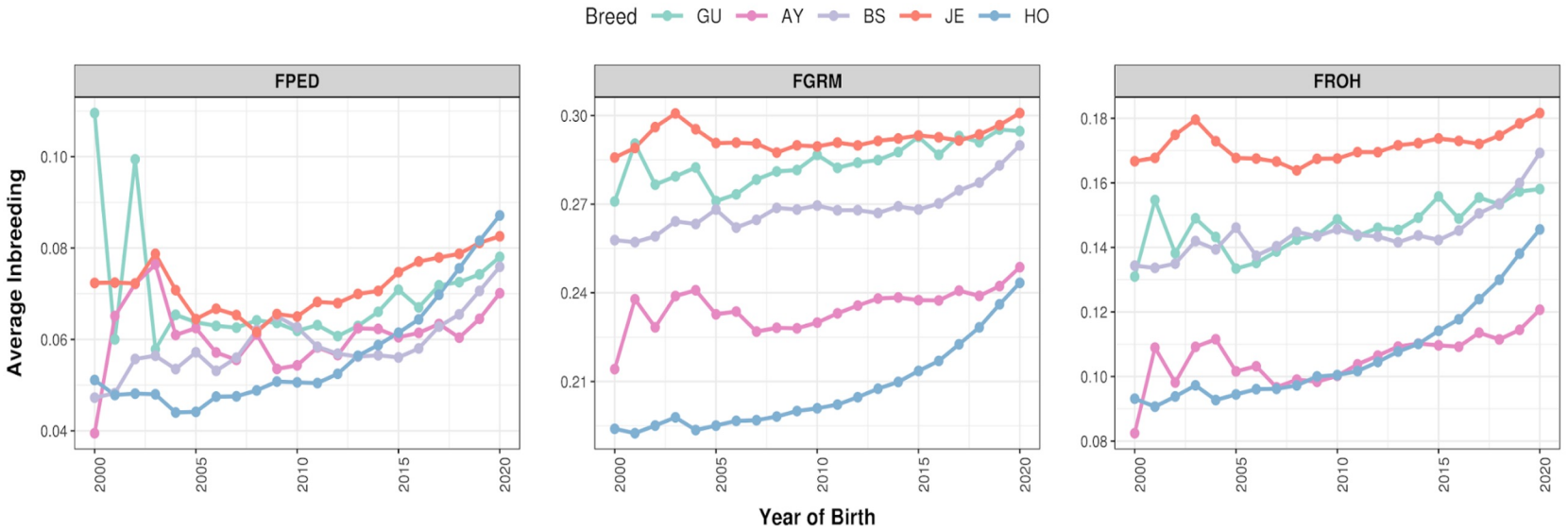
Breed	Acronym	Pedigree	Genotyped
Holstein	**HO**	6,725,679	3,649,734
Jersey	**JE**	985,959	459,784
Brown Swiss	**BS**	208,219	49,360
Ayrshire	**AY**	33,975	9,442
Guernsey	**GU**	45,792	5,359





NC STATE UNIVERSITY

Average Inbreeding by Year of Birth (2000-2020)



NC STATE UNIVERSITY

Yearly Rate of Inbreeding

- The yearly rate of inbreeding (ΔF_{yr}) was calculated for sires and dams born in three periods of interest.

1. Before the advent of genomic selection (P1; 2000-2009)
2. During the implementation of genomic selection (P2; 2010-2014)
3. After the widespread adoption of genomic selection (P3; 2015-2018)

- ΔF_{yr} was calculated using each of the inbreeding measures ($\Delta F_{PED_{yr}}$, $\Delta F_{GRM_{yr}}$, $\Delta F_{ROH_{yr}}$) by regressing the natural logarithm of 1-F on the year of birth of the animal and multiplying the slope by negative 1.

NC STATE UNIVERSITY

Effective population size estimates for US Holstein

Measure	Period	Males	Females
N_{ePED} (CI)	P1	672 (339, 37961)	75 (68, 84)
	P2	29 (27, 31)	55 (53, 57)
	P3	24 (22, 26)	29 (28, 29)
N_{eGRM} (CI)	P1	150 (116, 208)	55 (50, 60)
	P2	15 (14, 15)	42 (41, 44)
	P3	14 (14, 15)	20 (20, 20)
N_{eROH} (CI)	P1	131 (105, 174)	44 (41, 48)
	P2	16 (15, 16)	43 (42, 45)
	P3	15 (14, 16)	21 (20, 21)

NC STATE UNIVERSITY

Effective population size estimates for US Jersey

Measure	Period	Males	Females
N_{ePED} (CI)	P1	NA	NA
	P2	48 (37, 71)	117 (98, 145)
	P3	74 (44, 222)	112 (99, 130)
N_{eGRM} (CI)	P1	NA	NA
	P2	43 (29, 83)	195 (132, 369)
	P3	NA	276 (189, 512)
N_{eROH} (CI)	P1	NA	NA
	P2	35 (25, 57)	102 (82, 135)
	P3	59 (33, 279)	197 (149, 290)

NC STATE UNIVERSITY

Final remarks:

- The characterization of inbreeding has direct implications on the possible management of genetic diversity resources and avoiding inbreeding depression.

NC STATE UNIVERSITY

Covariance between relatives

NC STATE UNIVERSITY

Introduction

Quantitative genetics relies on the resemblance between relatives to estimate the genetic variances. The amount of phenotypic resemblance among relatives for the trait provides an indication of the amount of genetic variation for the trait. If trait variation has a significant genetic basis, the closer the relatives, the more similar their appearance.

Care must be taken, however, that biases due to environmental covariances between relatives are not present.In most mammalian litter - bearing species, for example, full sibs share the common environment of the litter, and thus the resemblance between them may be enhanced by the covariance due to their common environment.In humans, adopted children share a common environment with their nonbiological parents, which could cause a positive environmental covariance.In plants, competition could introduce a negative environmental covariance between sibs.

NC STATE UNIVERSITY

The covariance between relatives will be presented first using genotypic frequencies and genetic effects for one autosomal locus.Then the method of Malécot will be applied to give a general framework for deriving covariance in a random mating population.

To understand the usefulness of covariance between individuals we need to look back at the heritability. h^2 is a central concept in quantitative genetics: It explains the proportion of variation due to additive genetic values (Breeding values) and is defined as

$$h^2 = \frac{\text{VA}}{\text{VP}}$$

You should notice that while phenotypes (and hence VP) can be directly measured,

Breeding values (and hence VA) must be estimated and estimates of VA require known collections of relatives

Covariance

$$\text{Cov}(x, y) = E[x*y] - E[x]*E[y]$$

Covariance though is dependent on the unit of measure and it is not readly interpretable. A standardized measure of covariance is correlation which is

$$r = \frac{cov(x, y)}{\sqrt{V}ar(x)\sqrt{V}ar(y)}$$

correlation varies between - 1 and 1 where r = 1 implies perfect positive linear association and r = -1 perfect negative linear association

Now, consider the best (linear) predictor of y given that we know x

$$\hat{y} = \bar{y} + b\left(x - \bar{x}\right)$$

The slope of the regression is a function of the covariance

$$b = \text{cov}(x,y)/\text{Var}(x)$$

The fraction of the variation in y accounted for by knowing x, i.e,

$$Var(\hat{y} - y)$$

is

$$r^2$$

Now we can establish the relationship between correlation and regression slope as

$$r(x,y) = b * [Var(x)/Var(y)]$$

Note that If $Var(x) = Var(y)$, then

$$b(y \mid x) = b(x \mid y) = r(x, y)$$

NC STATE UNIVERSITY

Covariance (and the related measures of correlations and regression slopes) can be used to quantify the phenotypic resemblance between relatives.Quantitative genetics as a field traces back to R.A.Fisher’ s 1918 paper showing how to use the phenotypic covariances to estimate genetic variances, whereby the phenotypic covariance between relatives is expressed in terms of genetic (co)variances.

We can basically divide the relationship between individuals in two big categories. 1) Parent offspring. In this case covariance between individuals can be seen in terms of regression (which include parent offspring and mid - parent offspring regression). The slope of the (single) parent-offspring regression is estimated by

$$b(O|P) = \frac{Cov(O,P)}{Var(P)}$$

with

$$\frac{Cov(O,P)}{Var(P)} = \frac{1}{(n-1)\left(\Sigma OiPj - \text{n}\bar{O}\bar{P}\right)}$$

Notice that the regressions involves the covariance between parents and their offspring

Collateral releationships. In this case covariance between individuals can be estimated using ANOVA

With collateral relatives, the above formula for the sample covariance is not appropriate, for two reasons. First, there are usually more than two collateral relatives per family. Second, collateral relatives belong to the same class or category. In contrast, parents and offspring belong to different classes. The covariance between parents and offspring is an interclass (between-class) covariance, while the covariance between collateral relatives is an intraclass (within-class) covariance. The analysis of variance (ANOVA), first proposed in Fisher's 1918 paper, is used to estimate intraclass covariances.

Under the simplest ANOVA framework, we can consider the total variance of a trait to consists of two components:

a between-group (also called the among-group) component (for example, differences in the mean value of different families)

a within-group component (the variation in trait value within each family).

$$Var(T) = Var(B) + Var(W)$$

Variance(between groups) = covariance (within groups).

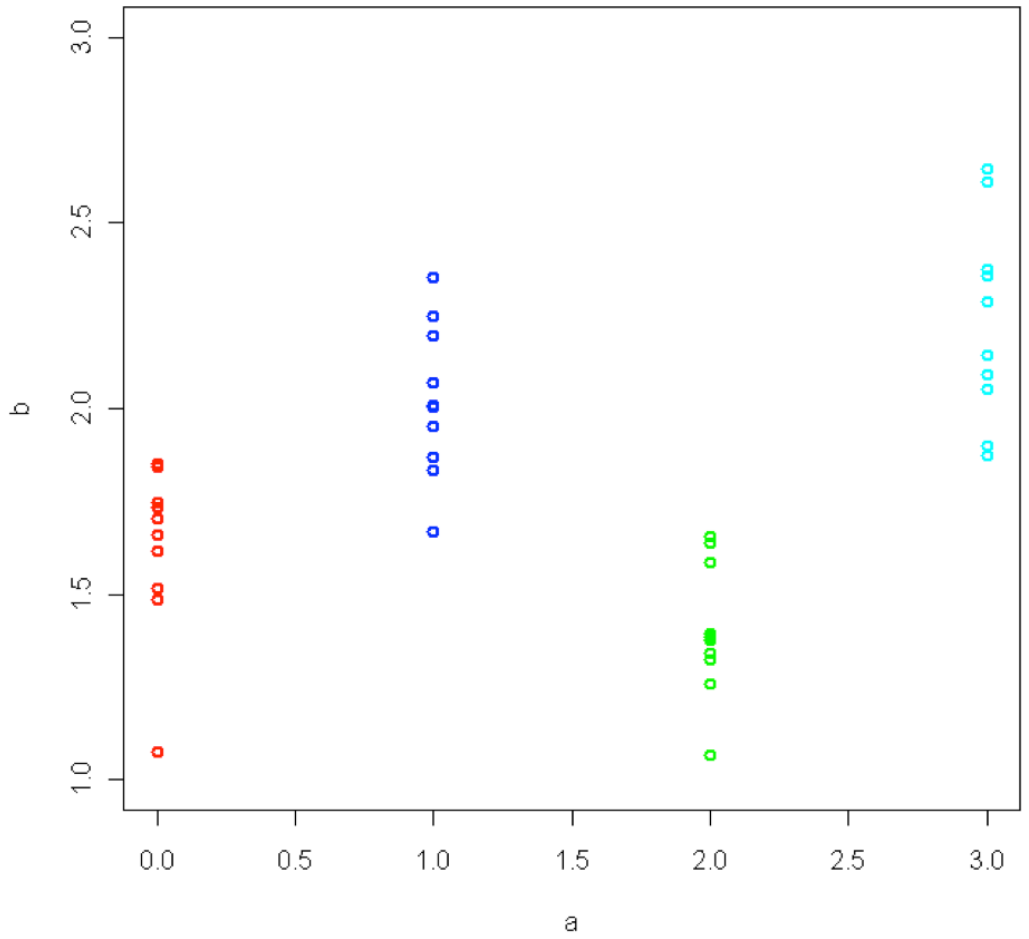
Thus, the larger the covariance between members of a family, the larger the fraction of total variation that is attributed to differences between family means

Intraclass correlation,

$$t = Var(B)/Var(T)$$

NC STATE UNIVERSITY

Situation 1

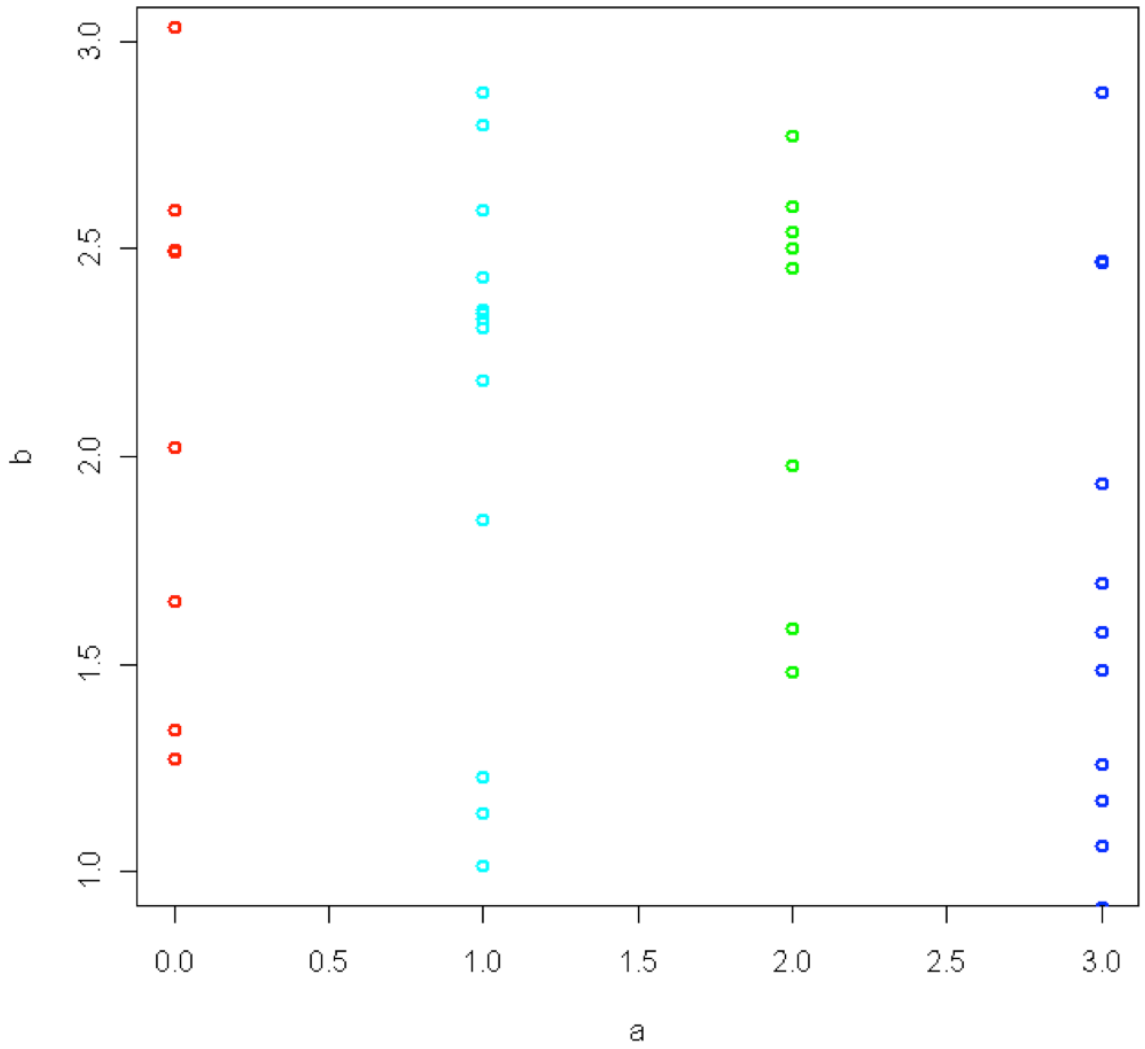


$$Var(B) = 2.5$$
$$Var(W) = 0.2$$
$$Var(T) = 2.7$$
$$t = 2.5/2.7 = 0.93$$

NC STATE UNIVERSITY

Situation 2

$$Var(B) = 0$$
$$Var(W) = 2.7$$
$$Var(T) = 2.7$$
$$t =/ 2.7 = 0$$



What happened in situation 2:

-) members of a family resemble each other no more than they do members of other families

-) there are no significant differences in average phenotype between families

-) phenotypic resemblance is low, so genetic variation is low

Note that phenotypic resemblance among relatives can equivalently be consider as a measure of the similiary among a group of relatives for the phenotype of a quantitative trait (the covariance of family members), or the difference in phenotype between different families (the between-group variance), as Cov(Within a group) = Var(Between group means).