Bioinformatics approaches in animal breeding

Summer School

July 9-11, 2025
Zagreb

University of Zagreb

University of Ljubljana

GENOMIC INBREEDING ESTIMATION

Shihabi Mario

Friday
11th July

SHORT INTRODUCTION
&
PRACTICAL WORKFLOW
&
PRACTICAL EXERCISES

SHORT INTRODUCTION

1. Inbreeding & Relatedness
2. Inbreeding & Relatedness estimation
3. Genomic Methods

Inbreeding →

4. Runs Of Homozigosity based Inbreeding (F_{ROH})
    a) ROH determination
    b) F_{ROH} estimation

Inbreeding & Relatedness →

5. Genomic Relationship Matrix (GRM)

Inbreeding

- Mating of closely related individuals
- Unavoidable due to small effective population size (Ne)

- Used also for ↑ productivity

↑ **the proportion of homozygosity (autozygosity)**

NEGATIVE CONSEQUENCES:

✓ Higher frequency of detrimental genetic disorders

✓ Inbreeding depression

– central point of population genetics/genomics –

Relatedness

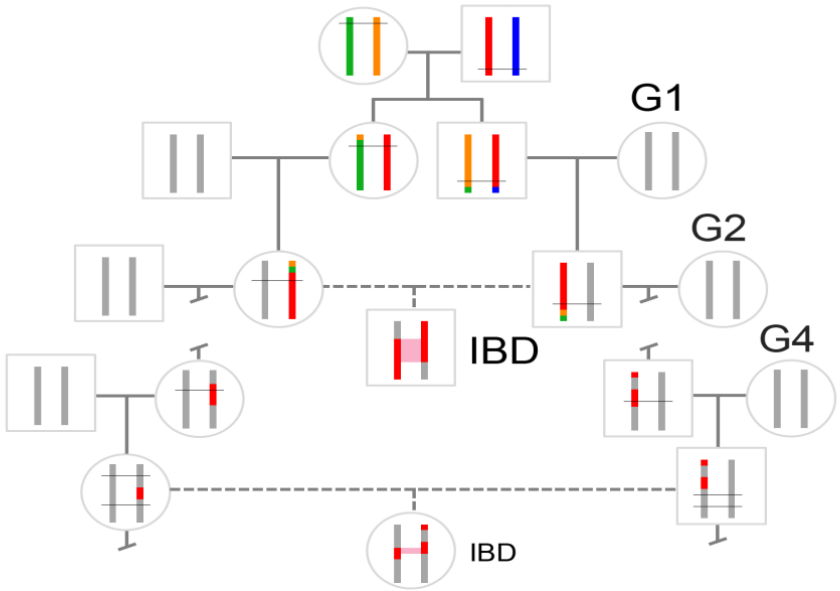
- Inbreeding -> within individuals
- Relatedness -> between individuals

- Proportion of alleles shared due to common ancestry

Key role in:

- Genetic evaluations and GWAS analyses
- Understanding population structure and dynamics
- Informing mating system strategies

Inbreeding & Relatedness estimation



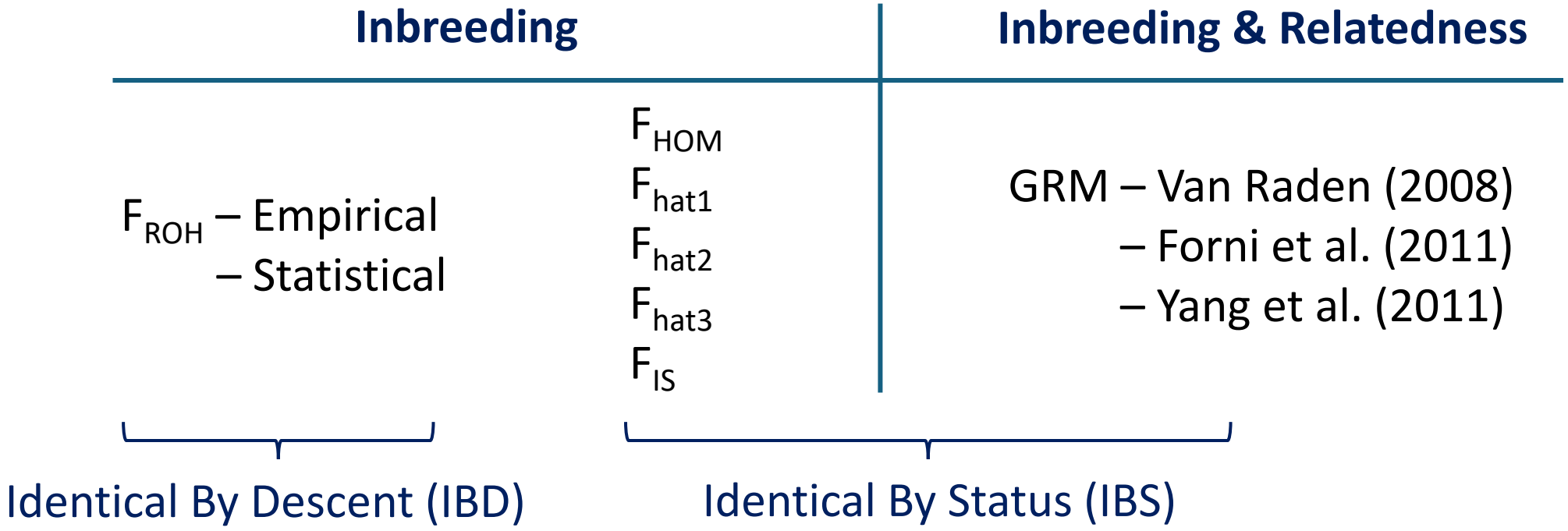
- Inbreeding coefficient: 0 - 1 (or -1 to 1)
- Relatedness coefficient: -1 to 1

THROUGH HISTORY – Through the Pedigree
MODERN AGE – **Genomic assessment**

- markers representing variable sites of the genome (SNPs)

IBD status or IBS status

Genomic Methods



Runs Of Homozigosity based Inbreeding (F_{ROH})

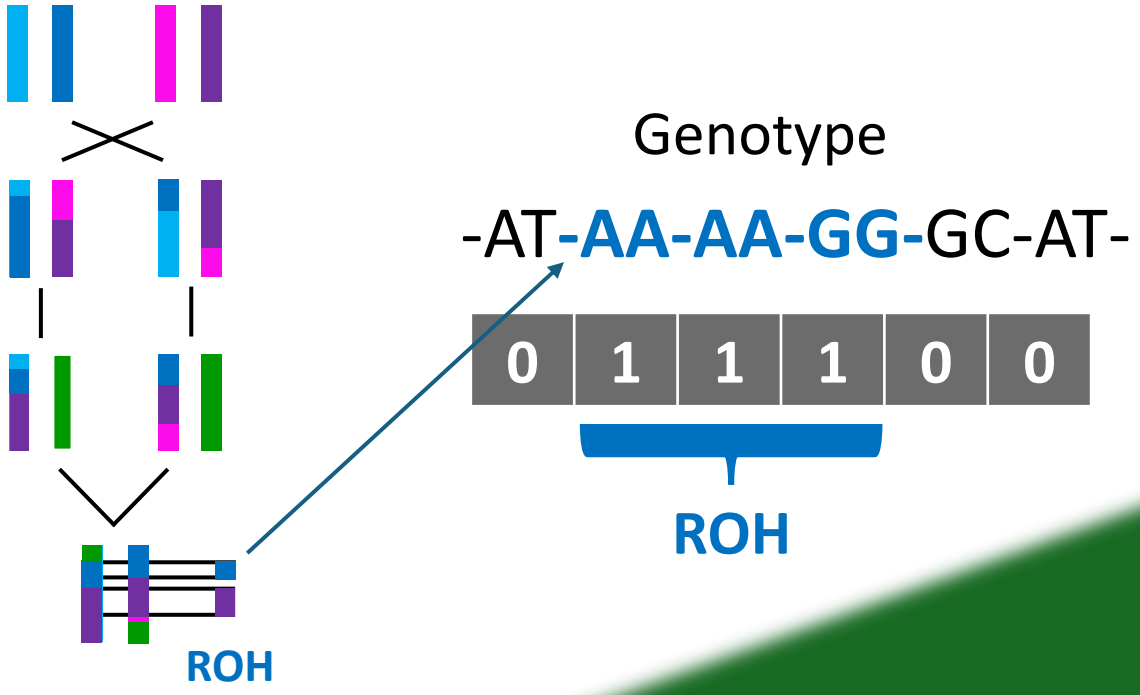
a) ROH determination

b) F_{ROH} estimation

ROH determination

ROH

- ✓ long homozygous regions in genome
- ✓ represent autozygous segments



ROH determination

a) Empirical approach

- Based on user defined tresholds

MAIN SOFTWARE:

- PLINK
- SVS Golden Helix
- detectRUNS R package
- GCTA

b) Statistical approach

- Based on statistical models

MAIN SOFTWARE:

- RZooROH R package
- bcftools
- hapROH

Empirical approach

	Parameter	Description
Main Parameter	**Min Length**	Minimum length of ROH in bps
	Min SNP	Minimum number of SNP in ROH
Due to error rate	**Max Het**	Maximum number of heterozygous SNP
	Max Missing	Maximum number of missing SNP
Optional:	**Max Gap**	Maximum distance between consecutive SNP to be still considered potential ROH
	Min Density	Minimum density of SNP per kb to define genomic region as potential ROH

Empirical approach

Parameter	50K SNP Array	HD SNP Array
Min Length	2 Mb or 4 Mb	1 Mb
Min SNP	15	15
Max Het	1 (per class if using SVS software)	1 (per class if using SVS software)
Max Missing	1 (per class if using SVS software)	1 (per class if using SVS software)
Max Gap	1 Mb	/
Min Density	/	/

1 Mb ≈ 50 generations ago
2 Mb ≈ 25 generations ago
4 Mb ≈ 12.5 generations ago
…

F_{ROH} estimation

$$F_{ROHk} = \frac{\sum_k \text{Length}(ROH_k)}{\text{Lgenome (info used)}}$$

ROH Incidence matrix

ID1	1 1 1 1 0 0 1 1 1 0 0 0 0 0 0 1 1 1	F_{ROH} = 0.56
ID2	0 0 1 1 0 0 1 1 1 0 1 1 1 0 0 0 0 0	F_{ROH} = 0.44
ID3	1 1 1 1 0 1 1 0 0 0 0 0 1 1 1 0 0 0	F_{ROH} = 0.50
ID4	0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0	F_{ROH} = 0.28

nSNPs = 18

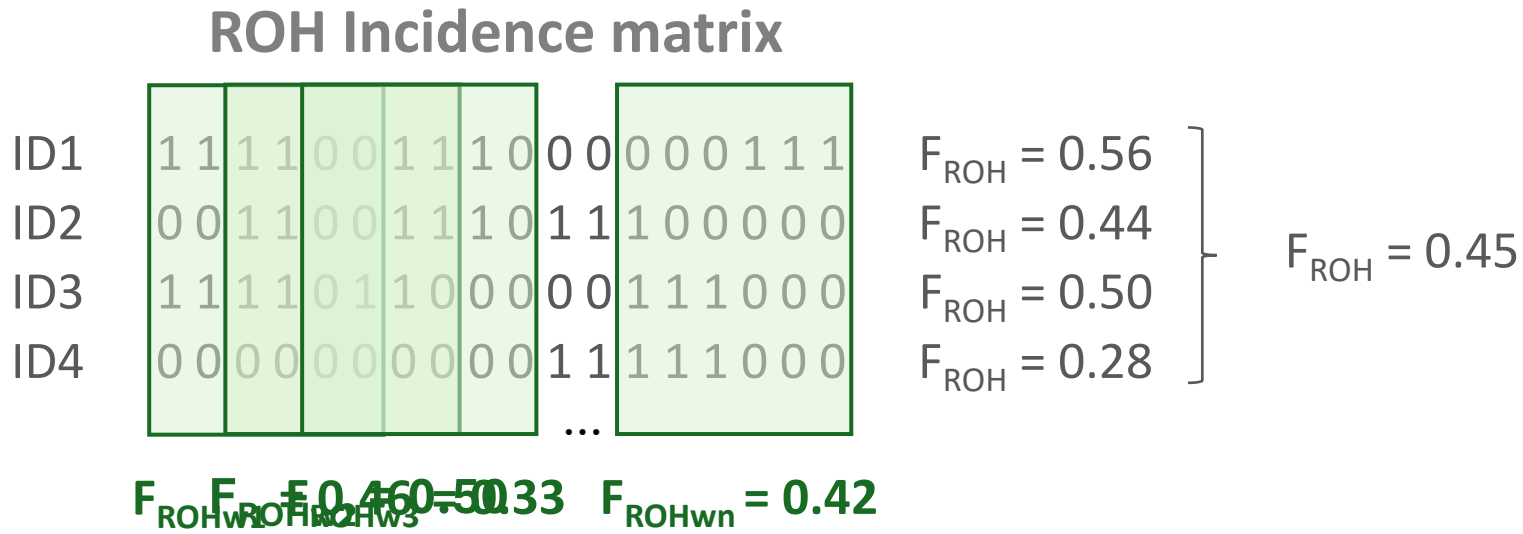
mean F_{ROH} = 0.45

1 = IBD

Small example without bps information

Regional F_{ROH} estimation

1 = IBD



This approach allows us to identify regions with different levels of inbreeding

Genomic Relationship Matrix (GRM)

- VanRaden (2008) -> centers and scales genotypes, by adjusting for allele frequencies (typically centered around 2pq) to ensure unbiased relationships

For each SNP, Formula incorporates: $(Z'Z)/\sum(2pq)$

Where:

Z'Z = matrix of genotype deviations
p and q = allele frequencies

Genomic Relationship Matrix (GRM)

-1 to 1

	ID1	ID2	ID3	ID4	ID5
ID1	0.92	-0.11	-0.22	0.16	0.13
ID2	-0.11	1.12	-0.26	0.05	0.13
ID3	-0.22	-0.26	1.33	0.16	-0.22
ID4	0.16	0.05	0.16	1.41	0.13
ID5	0.13	0.13	-0.22	0.13	0.98

Diag – 1 = Inbreeding

MAIN SOFTWARE:

- GCTA
- PLINK
- ASReml

PRACTICAL WORKFLOW

1. Import of .ped/.map files after QC
2. ROH determination
3. Import of Function.RData
4. Creation of ROH Incidence Matrix
5. F_{ROH} estimation
6. Regional F_{ROH} estimation
7. Visualisation of F_{ROH} results
8. Creation of Genomic Relationship Matrix (GRM)
9. GRM Visualisation

PRACTICAL WORKFLOW

1. Import of .ped/.map files after QC
2. ROH determination

- detectRUNS R package -> Biscarini *et al*. (2019)

❑ consecutiveRUNS.run() function

PRACTICAL WORKFLOW

1. Import of .ped/.map files after QC
2. ROH determination
3. Import of Function.RData

- 4 Internal functions:

- ❑ Start.End.Index() function
- ❑ Incidence.Matrix() function
- ❑ Froh.estimation() function
- ❑ Regional.Froh.estimation() function

PRACTICAL WORKFLOW

1. Import of .ped/.map files after QC
2. ROH determination
3. Import of Function.RData
4. Creation of ROH Incidence Matrix
5. F_{ROH} estimation
6. Regional F_{ROH} estimation
7. Visualisation of F_{ROH} results
8. Creation of Genomic Relationship Matrix (GRM)
   - Solomon Boison's function
   - ❑ calc_gnrm()

DEMONSTRATION

.R Script in R Studio

Open InbreedingAndRelatedness.R Script

Set working directory where input files are (Directory folder)

Example: .ped/.map format after QC (Ovine Infinium® HD SNP BeadChip 600K)
40 individuals of 5 Croatian native sheep breeds

PRACTICAL EXERCISES

Example 1: .ped/.map format after QC (Illumina CanineHD BeadChip 170K)
50 individuals of Labrador Retriever dog breed

Example1.R

Example 2: .ped/.map format after QC (Illumina BovineSNP50 BeadChip 50K)
97 individuals of Holstein cattle breed

Example2.R