LOUISIANA STATE UNIVERITY
College of Agriculture
School of Plant, Environmental, and Soil Sciences
AGRO 7075 Prediction-based Breeding

Quality control on marker dataset

Prof. Roberto Fritsche-Neto

rfneto@agcenter.lsu.edu

Baton Rouge, Jan 23rd, 2023

Quality control

- **Sweep samples**
- Eliminate individuals with low quality regarding genotyping

- **Call rate**
- Eliminate markers with low quality over all individuals

- **MAF**
- Eliminate monomorphic markers and with low frequency alleles (<0.05)

- **Other filters – Het, HWE and LD**

- **Inputation**
- Replace NA by a likely genotype

- **Recode**
- Recode the matrix to and incidence matrix (0, 1, or 2)

	snp1	snp2	snp3	snp4	snp5	snp6	snp7	snp8	snp9	snp10
gid 1	AA	AG	TT	CC	AC	AT	CG	CC	AA	GG
gid 2	AA	AG	TT	CT	AC	TT	CG	CC	AA	GG
gid 3	AA	NA	NA	CC	CC	NA	NA	CC	NA	AA
gid 4	AA	GG	NA	CC	CC	TT	GG	CC	AA	GG
gid 5	AA	GG	TC	CC	AA	TT	GG	GG	TT	AA
gid 6	AA	GG	CC	CC	AA	AA	CC	GG	TT	NA
gid 7	AA	AA	TC	CC	AA	AT	GG	GG	TT	AA
gid 8	AA	AA	NA	CC	AC	AT	CG	GG	TT	AA
gid 9	AA	AG	NA	CC	AA	AA	CG	GG	TT	AA
gid 10	AA	GG	CC	CC	AC	AA	GG	GG	AT	AG

	snp1	snp2	snp3	snp4	snp5	snp6	snp7	snp8	snp9	snp10
gid 1	AA	AG	TT	CC	AC	AT	CG	CC	AA	GG
gid 2	AA	AG	TT	CT	AC	TT	CG	CC	AA	GG
gid 4	AA	GG	NA	CC	CC	TT	GG	CC	AA	GG
gid 5	AA	GG	TC	CC	AA	TT	GG	GG	TT	AA
gid 6	AA	GG	CC	CC	AA	AA	CC	GG	TT	NA
gid 7	AA	AA	TC	CC	AA	AT	GG	GG	TT	AA
gid 8	AA	AA	NA	CC	AC	AT	CG	GG	TT	AA
gid 9	AA	AG	NA	CC	AA	AA	CG	GG	TT	AA
gid 10	AA	GG	CC	CC	AC	AA	GG	GG	AT	AG

	snp1	snp2	snp4	snp5	snp6	snp7	snp8	snp9	snp10
gid 1	AA	AG	CC	AC	AT	CG	CC	AA	GG
gid 2	AA	AG	CT	AC	TT	CG	CC	AA	GG
gid 4	AA	GG	CC	CC	TT	GG	CC	AA	GG
gid 5	AA	GG	CC	AA	TT	GG	GG	TT	AA
gid 6	AA	GG	CC	AA	AA	CC	GG	TT	NA
gid 7	AA	AA	CC	AA	AT	GG	GG	TT	AA
gid 8	AA	AA	CC	AC	AT	CG	GG	TT	AA
gid 9	AA	AG	CC	AA	AA	CG	GG	TT	AA
gid 10	AA	GG	CC	AC	AA	GG	GG	AT	AG

	snp2	snp5	snp6	snp7	snp8	snp9	snp10
gid 1	AG	AC	AT	CG	CC	AA	GG
gid 2	AG	AC	TT	CG	CC	AA	GG
gid 4	GG	CC	TT	GG	CC	AA	GG
gid 5	GG	AA	TT	GG	GG	TT	AA
gid 6	GG	AA	AA	CC	GG	TT	NA
gid 7	AA	AA	AT	GG	GG	TT	AA
gid 8	AA	AC	AT	CG	GG	TT	AA
gid 9	AG	AA	AA	CG	GG	TT	AA
gid 10	GG	AC	AA	GG	GG	AT	AG

	snp2	snp5	snp6	snp7	snp8	snp9	snp10	snp10
gid 1	AG	AC	AT	CG	CC	AA	GG	0
gid 2	AG	AC	TT	CG	CC	AA	GG	0
gid 4	GG	CC	TT	GG	CC	AA	GG	0
gid 5	GG	AA	TT	GG	GG	TT	AA	2
gid 6	GG	AA	AA	CC	GG	TT	AA	2
gid 7	AA	AA	AT	GG	GG	TT	AA	2
gid 8	AA	AC	AT	CG	GG	TT	AA	2
gid 9	AG	AA	AA	CG	GG	TT	AA	2
gid 10	GG	AC	AA	GG	GG	AT	AG	1