

# Imputation of genotypes in purebred and two-way crossbred pigs using low density panels

Tao Xiang

Aarhus University, Denmark  
AgroParisTech, France

Peipei Ma, Tage Ostensen, Andres Legarra, Ole Christensen

# Introduction

- **Genotype imputation generally introduced as initial step for genomic selection**
- **Factors affecting imputation performance:**
  - Genomic relationship
  - SNPs density
  - LD across markers
  - MAF
  - Size of reference panel
- **Mainly investigated in purebreds**

# Introduction

- **Crossbreeding predominant in pork production**
- **Algorithms (such as Beagle)**
  - Depend crucially on LD patterns across markers
  - Patterns may be breed specific
- **Performance of genotype imputation in crossbreds may be different than in purebreds**

# Objectives

- **Compare performance of imputation in crossbreds to that in purebreds**
- **Explore optimal strategy of imputing crossbreds from low density to medium density chip**

# Materials

## • Purebreds

8,848 Danish Landrace (LL)  
8,914 Danish Yorkshire (YY) } 42,483 SNPs (60K) genotyped

## • Crossbreds

5,679 two-way crossbred LY

- 1) 7,940 SNPs (8K) genotyped
- 2) 42,483 SNPs (60K)  
simulated based on pedigree  
not conditional on 8K

# Methods – software Beagle v3.3.2

## 5K (5,293 SNPs) to 8K (7,940 SNPs)

Validation	Reference
4682 LL	4166 LL+ 4263YY
4651 YY	4166 LL+ 4263YY
5679 crossbred	8429 LL
5679 crossbred	8429 YY
5679 crossbred	4166 LL+ 4263YY

## 8K (7,940 SNPs) to 60K (42,483 SNPs)

Validation	Reference	Data type
4682 LL	8429 LL+YY	Genotyped
4651 YY	8429 LL+YY	Genotyped
5679 crossbred	8429 LL+YY	Simulated

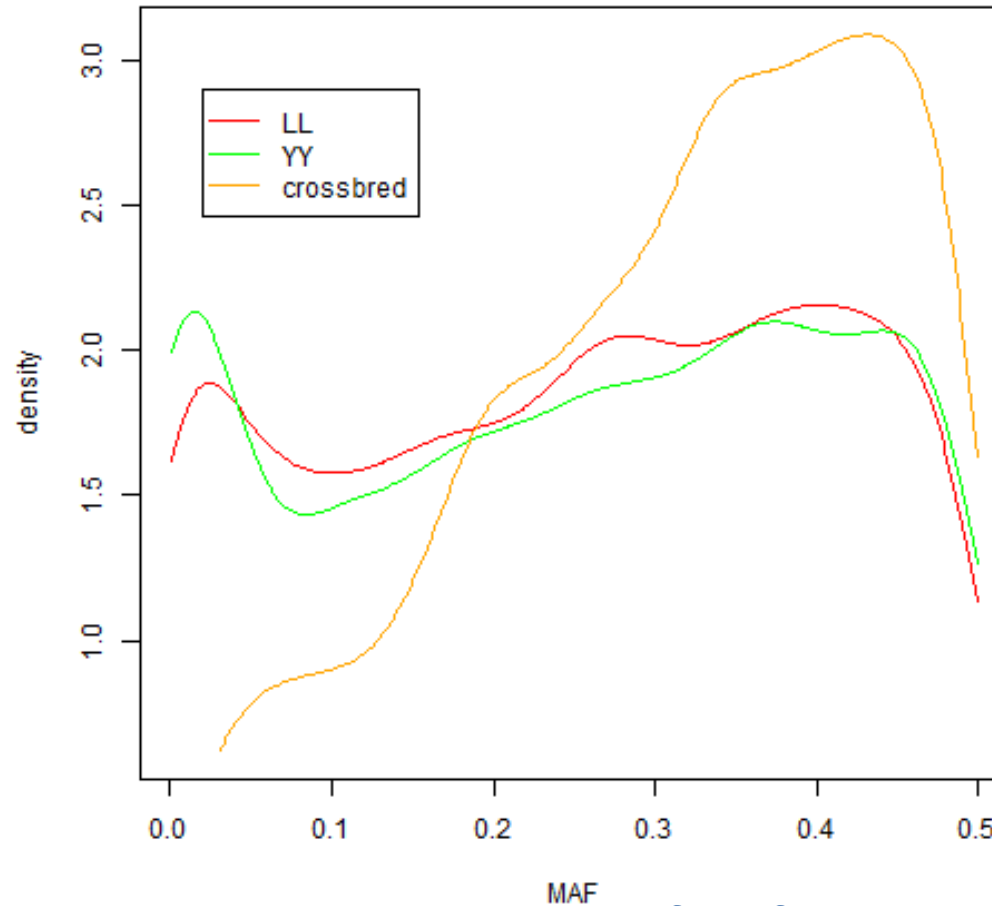
# Results and discussion

## • From 5k (5,293 SNPs) to 8k (7,940 SNPs)

Validation (size)	Ref. (size)	Correct rate	Correlation
LL (4682)	LL+YY (8429)	0.9910	0.9606
YY (4651)	LL+YY (8429)	0.9907	0.9477
LY (5679)	LL (8429)	0.9034	0.7595
LY (5679)	YY (8429)	0.8667	0.6871
LY (5679)	LL+YY (8429)	0.9850	0.9566

- Imputation in crossbreds performs as well as in purebreds, when reference panel consists of combined parental breeds.
- Reference of single breed will decrease imputation accuracies greatly.
- Landrace contributes more than Yorkshire to imputation in crossbreds.

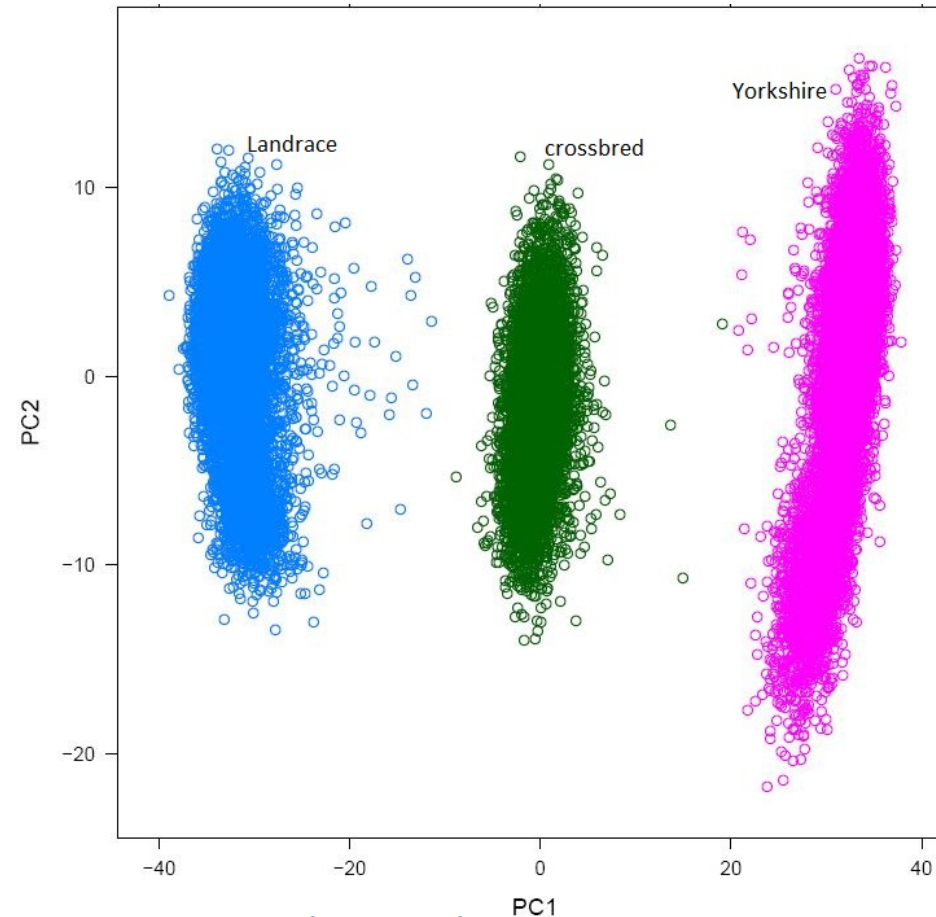
# MAF explain slight differences



- SNPs with very low MAF occur more often for purebreds
- Decreases correlation and increases correct rate for purebreds



# PCA of genomic relationship



- First principal component (x-axis) separated all three breeds
- Connection 'LL and crossbred' slightly tighter than 'YY and crossbred'

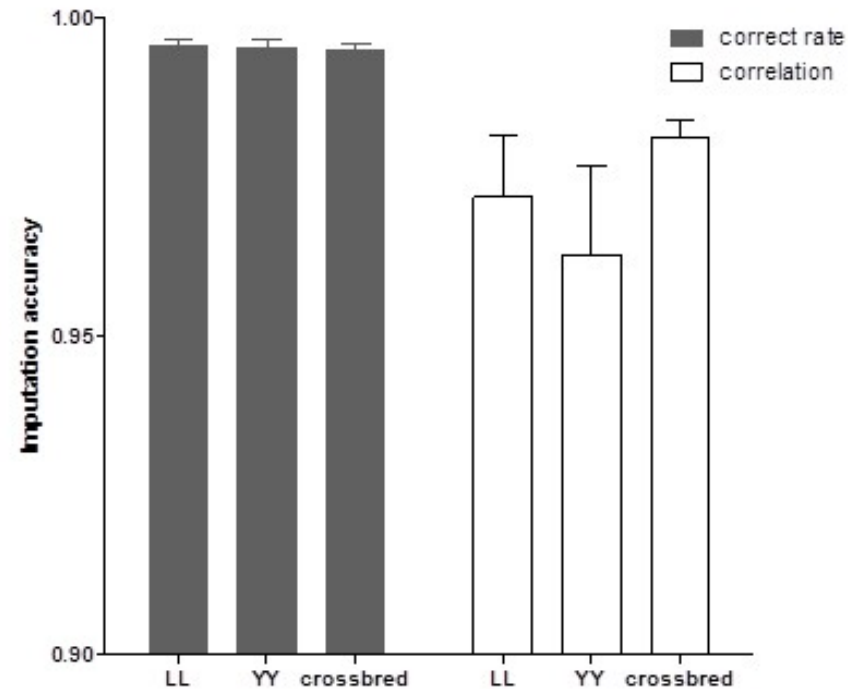
# Shared haplotypes ratio

- If the copy of one haplotype in Validation can be found in Ref, number of shared hap +1
- Ratio = number of shared haps / total haps
- Adjust the length of haps from 10 to 100 SNPs

Test	Ref	10	20	30	50	100
crossbred	LL+YY	0,9902	0,9606	0,9135	0,8092	0,5357

# Results and discussion

- From 8K (7,940) to 60K (42,483) SNP chips



- Imputation accuracies marginally increased than from 5K to 8K.
- It confirms imputation performs in crossbreds as well as in purebreds.

# Conclusions

- Imputation can perform as well in crossbreds as in purebreds
- Imputing crossbreds should include both genotyped purebred parents in the reference panel
- The ratios of shared haplotypes can be appropriate interpretation of imputation accuracy

Thank you for your attention!

# Conclusions

- Imputation can perform as well in crossbreds as in purebreds
- Imputing crossbreds should include both genotyped purebred parents in the reference panel
- The ratios of shared haplotypes can be appropriate interpretation of imputation accuracy