

# Accuracy of imputation in a sparsely-genotyped pig pedigree

Matthew A Cleveland<sup>1\*</sup>, Brian P Kinghorn<sup>2</sup> and  
John M Hickey<sup>2</sup>

<sup>1</sup>Genus plc, Hendersonville, TN USA

<sup>2</sup>School of Environmental and Rural Science, University of  
New England, Armidale, NSW Australia



# Background

- Implementation of genomic selection requires large numbers of individuals to be densely genotyped
- Candidates for selection can be genotyped to increase accuracy of selection and decrease generation interval
- Currently not cost effective to genotype all candidates in some livestock species

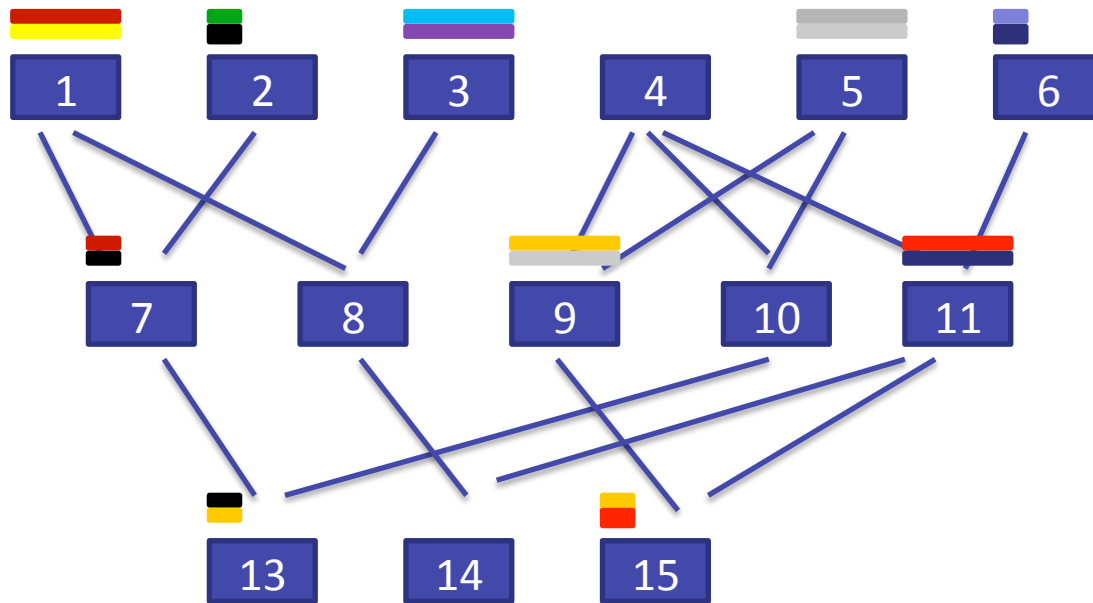
# Background

- Strategy for implementing genomic selection in pigs
  - Genotype sires and some dams at 60k
  - Genotype selected candidates at low density
    - No commercial low-density chip
  - Impute from low to high density
  - Use imputed genotypes across all traits
- Impute
  - Use pedigree information
  - Close relatives are not always genotyped at high density

# Imputation approach - AlphaImpute

- Long range phasing, segregation analysis and haplotype library imputation
- Uses information from close and distant relatives/ SNPs to impute alleles
- Imputes individuals with varying densities of genotypes and levels of relatives genotyped
- Where alleles can not be imputed allele probabilities are calculated and summed to give a probable genotype score

# Imputation approach - AlphaImpute



Genotyping strategy in terms of high density, low density and not genotyped



Haplotype library for population

- Individual's are densely, sparsely, or not genotyped
- Pedigree information available
- Single locus segregation analysis for each SNP
- Long-range phasing and haplotype libraries
- Match each pair of haplotypes with low density genotypes and allele probabilities

# Testing data

- Samples
  - Single line
  - 60k genotyped: N=3,534
  - Pedigree: N=6,473
  - Test samples
    - Most recently-born animals, no progeny: N=509
- SNPs
  - Known position on SSC01
  - Filtered for MAF, quality and call rate
  - M=4,221

N	Genotype Category
51	Both Parents
62	Sire and MGS
46	Dam and PGS
45	Sire
14	Dam
291	Other

# Imputation

- Mask genotypes to simulate low-density genotyping
- Low-density genotyped SNPs selected based on position and MAF
- Genotyping scenarios:

No. SNPs	%60k (whole-genome basis)	Whole genome density
725	18	6k
379	9	3k
184	5	1500
93	2	768
46	1	384

# Results – 6k density

## Imputation accuracy for test animals

Correct	Incorrect		Paternal	Maternal		
(%)	(%)	Corr. <sup>a</sup>	Alleles	Alleles	N	
99.4	0.2	<b>1.00</b>	99.9	99.9	51	<b>Both Parents</b>
97.7	1.4	<b>0.99</b>	99.9	99.4	62	<b>Sire and MGS</b>
98.7	0.8	<b>0.99</b>	99.7	99.7	46	<b>Dam and PGS</b>
97.2	1.7	<b>0.99</b>	99.9	99.9	45	<b>Sire</b>
96.3	2.0	<b>0.98</b>	98.8	98.8	14	<b>Dam</b>
95.1	2.9	<b>0.97</b>	99.4	99.4	291	<b>Other</b>

<sup>a</sup>Correlation between sum of allele probabilities and called genotypes



# Results – 3k density

## Imputation accuracy for test animals

Correct	Incorrect		Paternal	Maternal		
(%)	(%)	Corr. <sup>a</sup>	Alleles	Alleles	N	
			(%)	(%)		
99.2	0.4	<b>1.00</b>	99.8	99.8	51	<b>Both Parents</b>
96.6	2.4	<b>0.98</b>	88.9	99.1	62	<b>Sire and MGS</b>
98.2	1.3	<b>0.99</b>	99.8	99.8	46	<b>Dam and PGS</b>
96.1	2.4	<b>0.98</b>	99.9	98.7	45	<b>Sire</b>
93.4	4.3	<b>0.96</b>	98.0	99.9	14	<b>Dam</b>
92.8	4.5	<b>0.96</b>	99.2	98.1	291	<b>Other</b>

<sup>a</sup>Correlation between sum of allele probabilities and called genotypes

# Results – 1500 density

## Imputation accuracy for test animals

Correct (%)	Incorrect (%)	Corr. <sup>a</sup>	Paternal	Maternal	N	
			Alleles (%)	Alleles (%)		
98.9	0.6	<b>0.99</b>	99.7	99.8	51	<b>Both Parents</b>
94.9	3.7	<b>0.97</b>	99.8	98.8	62	<b>Sire and MGS</b>
97.0	2.3	<b>0.98</b>	99.6	99.7	46	<b>Dam and PGS</b>
92.7	4.3	<b>0.96</b>	99.8	97.3	45	<b>Sire</b>
92.3	4.6	<b>0.96</b>	97.2	99.7	14	<b>Dam</b>
89.1	6.8	<b>0.94</b>	98.7	97.0	291	<b>Other</b>

<sup>a</sup>Correlation between sum of allele probabilities and called genotypes

# Results – 768 density

## Imputation accuracy for test animals

Correct (%)	Incorrect (%)	Corr. <sup>a</sup>	Paternal	Maternal	N	
			Alleles (%)	Alleles (%)		
98.1	1.3	<b>0.99</b>	99.7	99.7	51	<b>Both Parents</b>
92.4	5.4	<b>0.95</b>	99.7	98.1	62	<b>Sire and MGS</b>
95.8	3.5	<b>0.97</b>	99.7	99.6	46	<b>Dam and PGS</b>
88.8	7.0	<b>0.94</b>	99.8	96.1	45	<b>Sire</b>
86.4	8.3	<b>0.92</b>	94.8	99.9	14	<b>Dam</b>
82.2	9.6	<b>0.90</b>	97.3	94.4	291	<b>Other</b>

<sup>a</sup>Correlation between sum of allele probabilities and called genotypes

# Results – 384 density

## Imputation accuracy for test animals

Correct (%)	Incorrect (%)	Corr. <sup>a</sup>	Paternal	Maternal	N	
			Alleles (%)	Alleles (%)		
97.1	2.02	<b>0.98</b>	99.5	99.7	51	<b>Both Parents</b>
88.1	8.1	<b>0.93</b>	99.6	96.7	62	<b>Sire and MGS</b>
92.6	6.5	<b>0.95</b>	99.5	99.7	46	<b>Dam and PGS</b>
80.3	11.8	<b>0.89</b>	99.8	92.3	45	<b>Sire</b>
83.6	11.0	<b>0.90</b>	94.9	99.6	14	<b>Dam</b>
72.0	12.4	<b>0.84</b>	94.7	86.0	291	<b>Other</b>

<sup>a</sup>Correlation between sum of allele probabilities and called genotypes

# Results – IMPUTE2 comparison

## Correlation between imputed and called genotypes for test animals

---

384 density		3k density			
Alpha Impute	IMPUTE2	Alpha Impute	IMPUTE2	N	
0.98	0.77	1.00	0.96	51	<b>Both Parents</b>
0.93	0.80	0.99	0.94	62	<b>Sire and MGS</b>
0.96	0.79	0.99	0.95	46	<b>Dam and PGS</b>
0.89	0.78	0.99	0.95	45	<b>Sire</b>
0.90	0.76	0.98	0.93	14	<b>Dam</b>
0.86	0.76	0.97	0.95	291	<b>Other</b>

<sup>a</sup>**IMPUTE2**: B. N. Howie, P. Donnelly and J. Marchini (2009) *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genetics* 5 (6): e1000529

# Results – Genomic breeding values using imputed genotypes

- Calculate gEBV using full SNP panel and using low-density SNP panel with imputation

## Correlation between gEBV using low-density SNPs and gEBV using all SNPs

---

<b>384 density</b>	<b>3k density</b>	
0.90	0.99	<b>Both Parents</b>
0.87	0.99	<b>Sire and MGS</b>
0.94	1.00	<b>Dam and PGS</b>
0.86	0.98	<b>Sire</b>
0.84	0.98	<b>Dam</b>
0.73	0.97	<b>Other</b>

---

# Discussion

- Improvements possible, especially for “other” category
  - In practice most animals have at least one parent genotyped
  - Recombination modeling
- Explore genotyping strategies
- Use to impute un-genotyped pedigree
  - Increase training set size
  - Simplify single step evaluation to GBLUP

# Conclusions

- Imputation method is robust and useful with incomplete genotyping
- The correlation between imputed and called genotypes was near 1 at all densities when both parents were genotyped
- Even at very low densities and when not closely related to genotyped individuals, correlations were high
- Can calculate gEBV for selection candidates using small panel



# Acknowledgements

- Lizhen Wang
- Selma Forni

