



Imputation accuracy to sequence

Deirdre Purfield¹, J.F Kearney² & D.P Berry¹

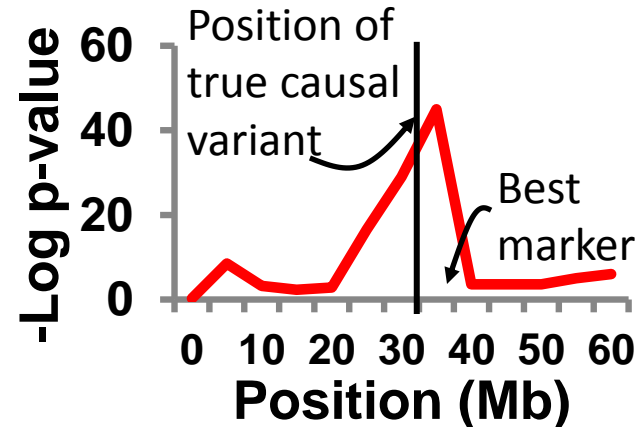
¹Teagasc, Moorepark, Ireland

²Irish Cattle Breeding Federation, Bandon, Ireland

Deirdre.purfield@teagasc.ie

Why impute to sequence?

- The most comprehensive information of an individual's variome
- Capture of **causative variants**
- Improve association studies
 - ❑ **Increased LD** → narrowed region of association
- For genomic prediction
 - ❑ **Higher accuracy** of prediction
 - ❑ Better prediction **across breeds?**
- Overcome cost by exploiting imputation

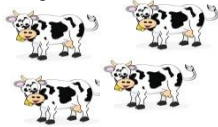


Methods

- Whole genome sequence from the 1,000 Bulls Genomes Project
 - ❑ 1147 Bos Taurus animals across 27 breeds: Run4
- **70** Holstein-Friesians **HD** genotyped and **sequenced**
- **50** Holstein-Friesians **BovineSNP50** genotyped and **sequenced**
- Imputation using **FImpute2** and **Beagle 4.0**
- Analysed BTA1, BTA15 and BTA29
- **Accuracy of imputation** (r) = correlation between the observed and imputed genotype
 - ❑ -1 opposite genotype imputed
 - ❑ +1 correct genotype imputed

Imputation Scenarios

1. Single-breed reference population

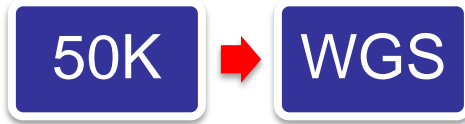


VS

Multi-breed reference population



2. BovineSNP50: **One-step** versus **two-step** imputation



VS



3. Impact of **sire** in the reference population



4. **Rare alleles** imputation

5. Position across the chromosome

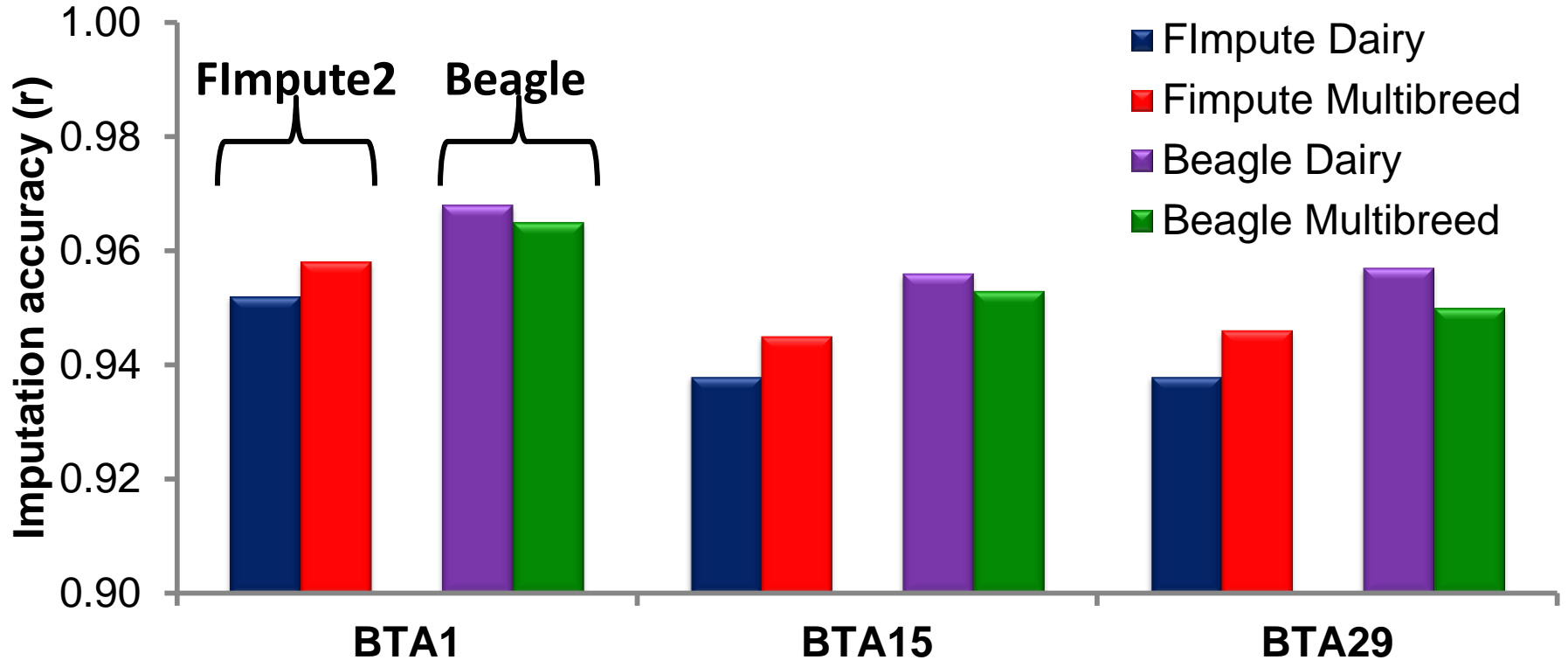
HD and sequence genotype concordance

- Strong concordance between sequence and SNP chip genotype

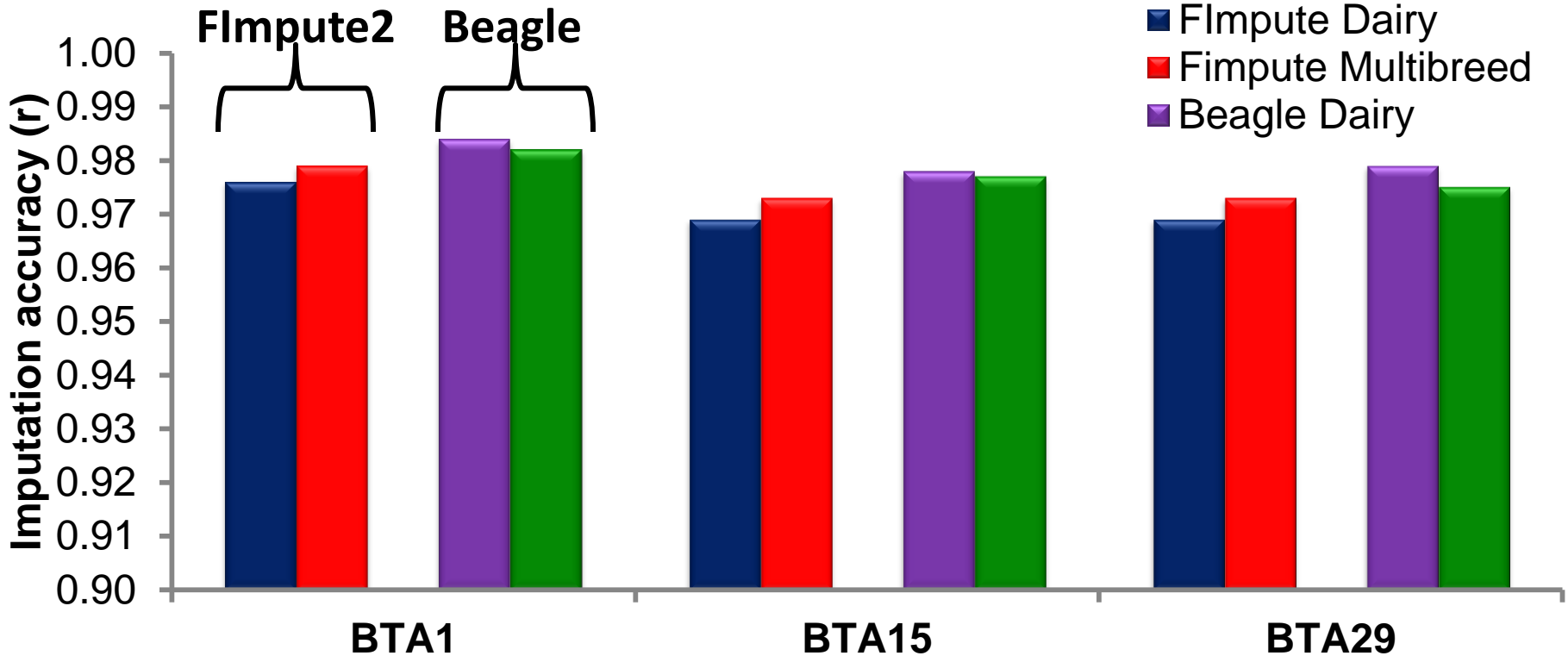
	HD
BTA1	99.08
BTA15	98.99
BTA29	98.95

- Imputation time **considerably longer** with Beagle 4.0
BTA1 60hrs:3mins vs 6hrs:42mins with FImpute2

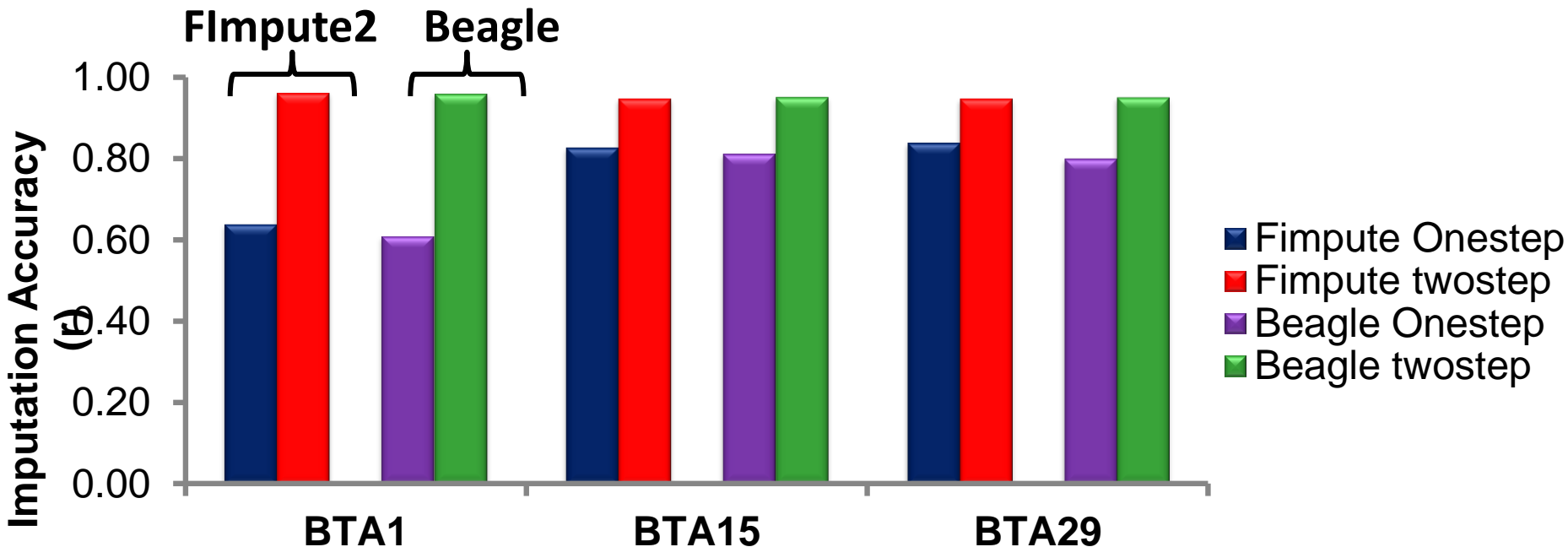
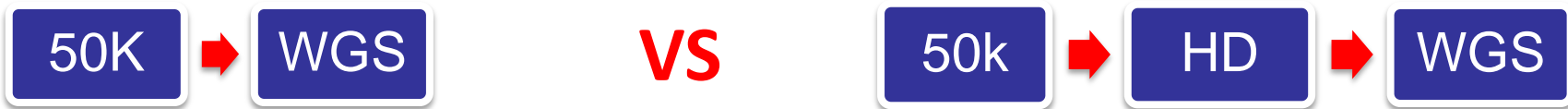
Imputation Accuracy



Allelic Concordance

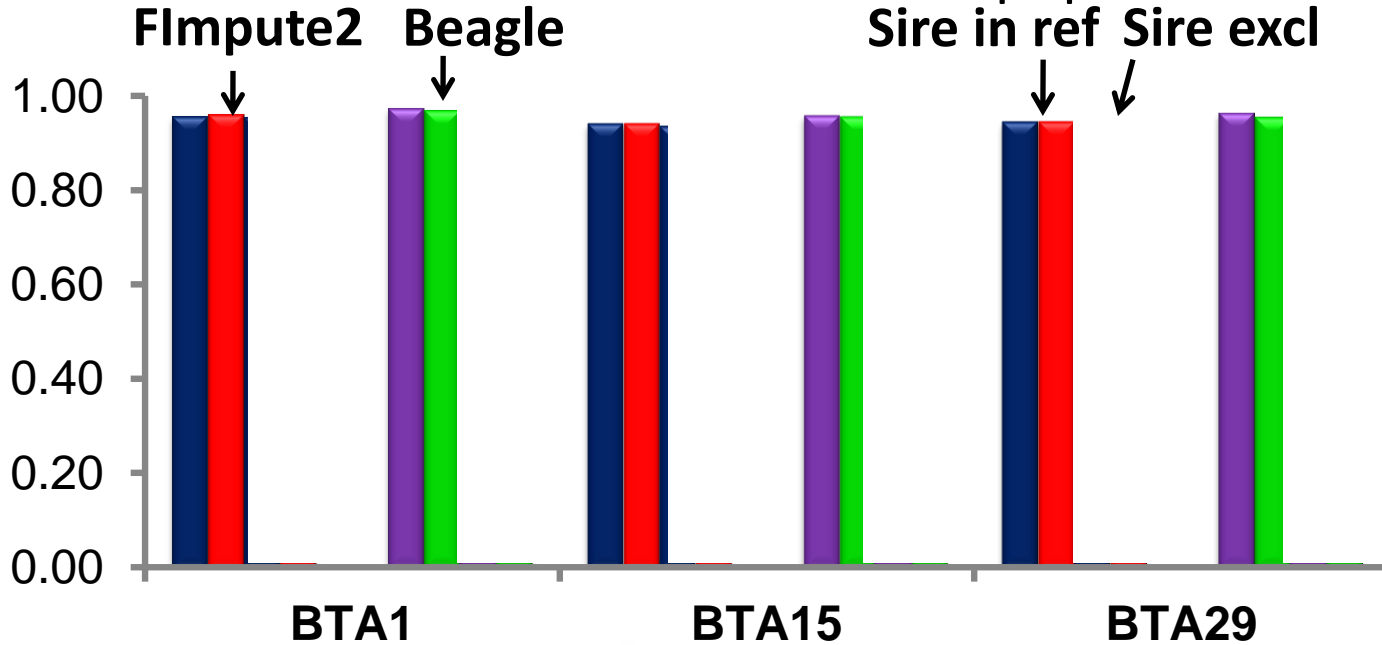


One step vs Two step



Impact of sire in the reference population

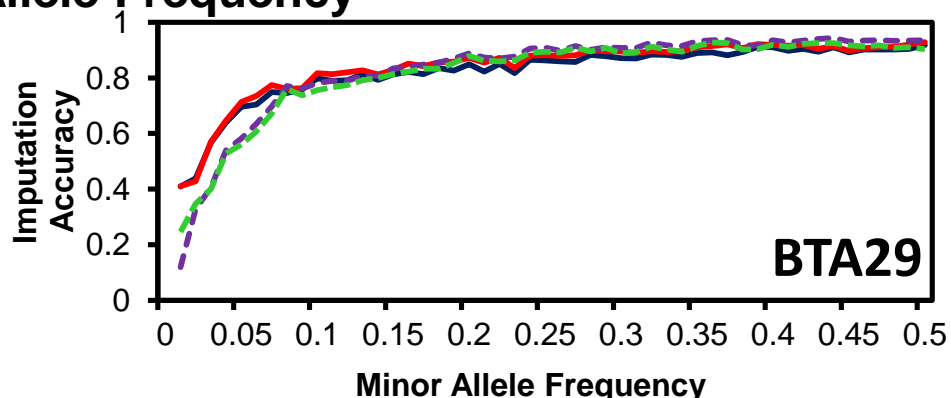
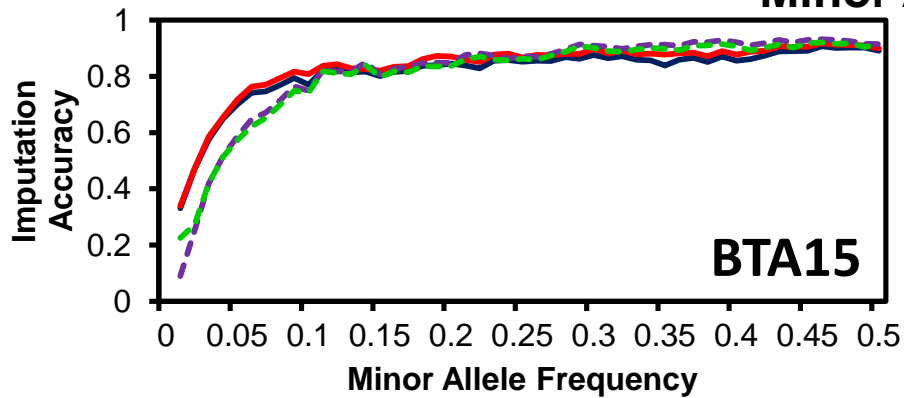
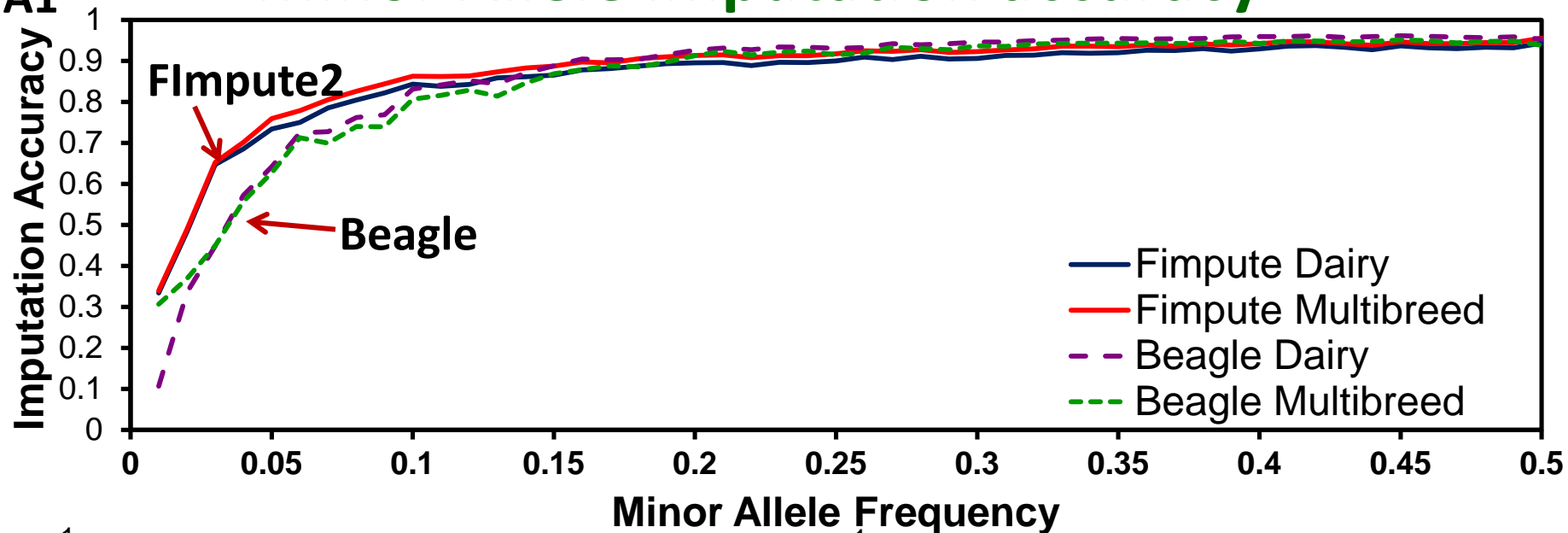
- 15 HD Hol-Fr with sires in the population
- Excluded sire from the reference population



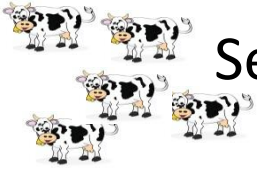
No improved
accuracy
detected

Minor Allele imputation accuracy

BTA1



Holstein Specific variants



Segregating variants in the 70 HD and sequenced animals but not segregating in any of the other breeds

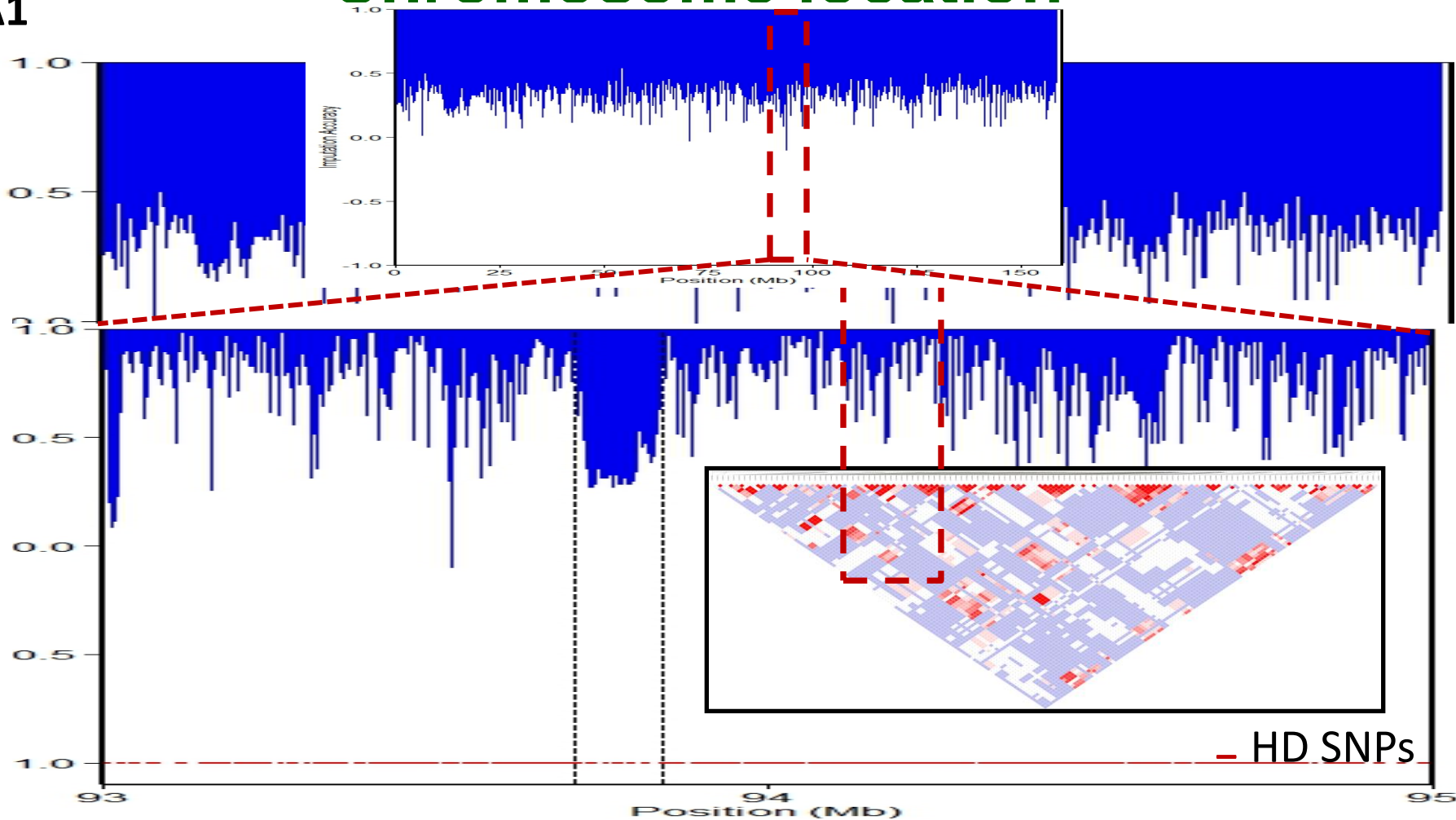
BTA	Markers	Flmpute		Beagle	
		Dairy	Multibreed	Dairy	Multibreed
1	6,741	0.971	0.971	0.973	0.968
15	4,404	0.975	0.975	0.979	0.975
29	2,439	0.969	0.973	0.974	0.969

Chromosome location

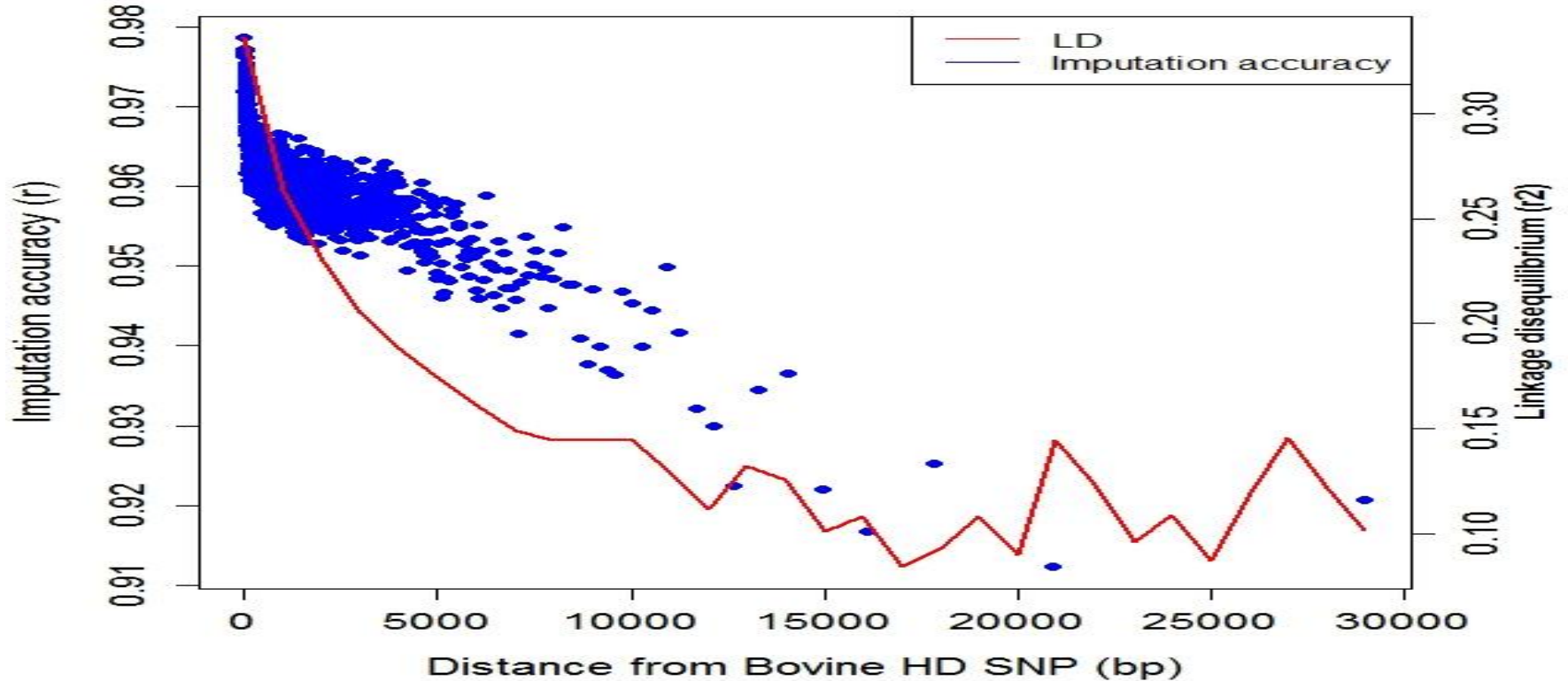
BTA1

Imputation Accuracy

Imputat



Impact of distance from nearest Bovine HD SNP



Conclusions

1. **Accurate imputation** can be achieved
2. Slightly **greater** accuracy with **BEAGLE 4.0**
3. FImpute2 is faster and better for **rare variants**
4. Imputation accuracy for rare variants has improved
5. Poor imputation in **HD SNP deficient** areas

Acknowledgements

- 1000 Bulls Genomes Project
- Research Stimulus Project
11/S/112



Department of
**Agriculture,
Food and the Marine**
An Roinn
**Talmhaíochta,
Bia agus Mara**

