



Genomic selection in crossbred cattle: the progress in Canada

Ricardo Ventura
UoGuelph – Canada
Beef Improvement Opportunitites - Canada

Flavio Schenkel – UoGuelph - Canada
Steve Miller – AgResearch – New Zealand

Topics:



Agreement between BIO and CGIL

Tools for phenotype collection

Genomic Selection of Purebred animals

Genome Canada Project

BIO Projects

Imputation

Genomic Selection using related animals

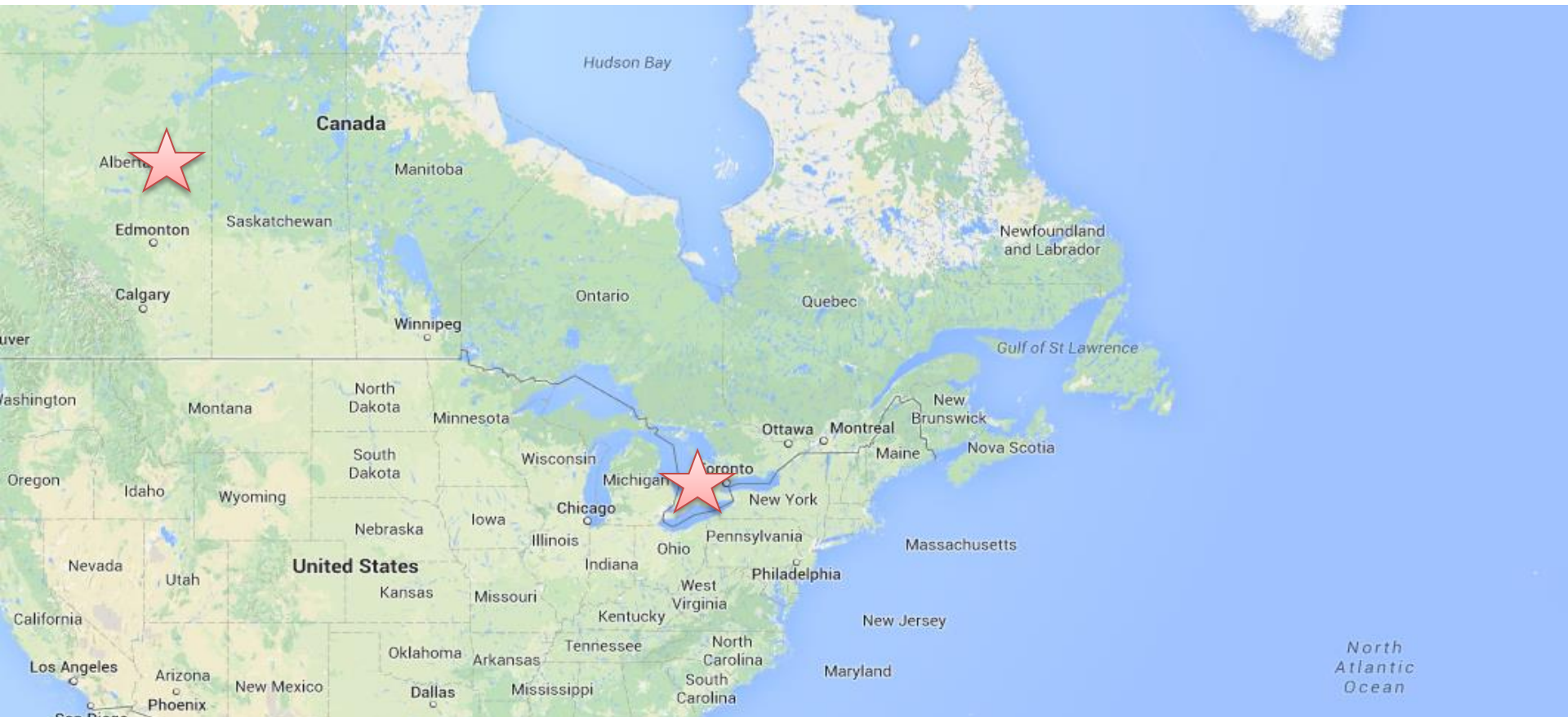
Conclusions

BIO: producer cooperative, 20 years old (ON and AB)



Genetic and Genomic Evaluations

Softwares: biotrack and biolink



BIO: producer cooperative, 20 years old (ON and AB)



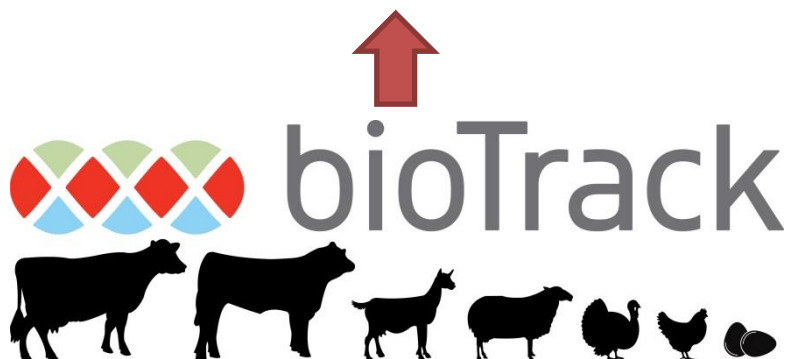
Genetic and Genomic Evaluations

Softwares: biotrack and biolink

Farmer ID



RFID tag



All cattle must be identified with an approved tag before leaving the site where they were born.

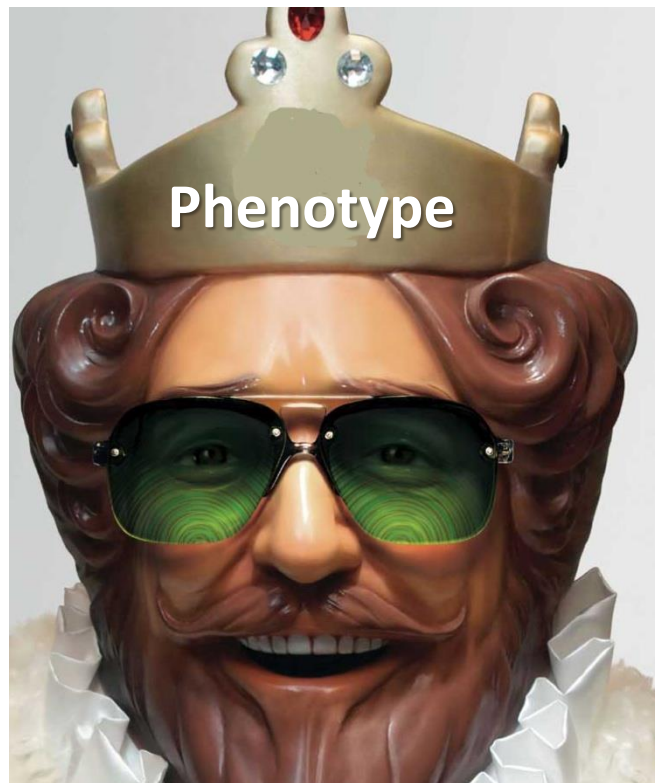
Age verify your calves automatically with data upload to CCIA every 6 hours

BIO: producer cooperative, 20 years old (ON and AB)



Genetic and Genomic Evaluations

Softwares: biotrack and biolink

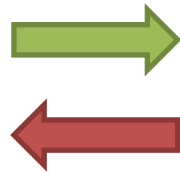


BIO: producer cooperative, 20 years old (ON and AB)



Genetic and Genomic Evaluations

Softwares: biotrack and biolink



BIXS BEEF
INFOCHANGE
SYSTEM

JBS, CARGILL ...

 **bio link**

Smaller plants

Food inventory

Tracking

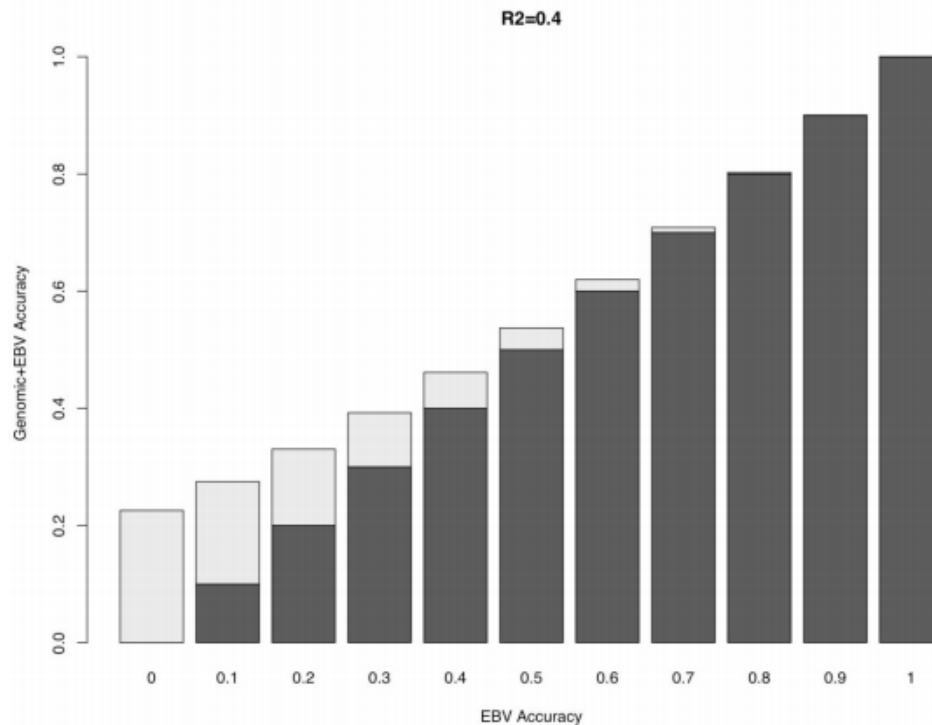
Sales

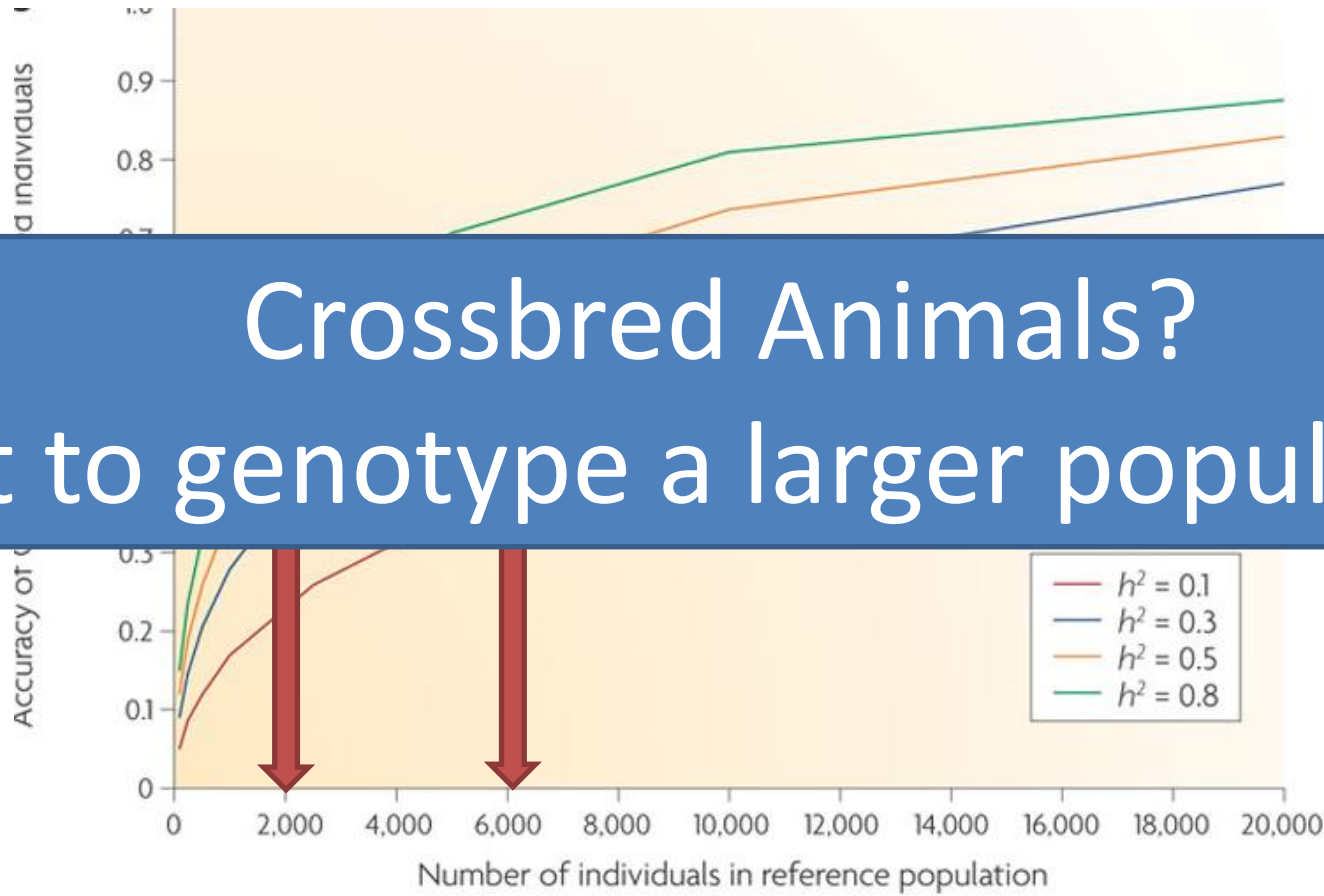


Phenotypes for Genomic Selection

Genomic Selection: Purebred animals

Increase in accuracy from integrating genomic information





Crossbred Animals?

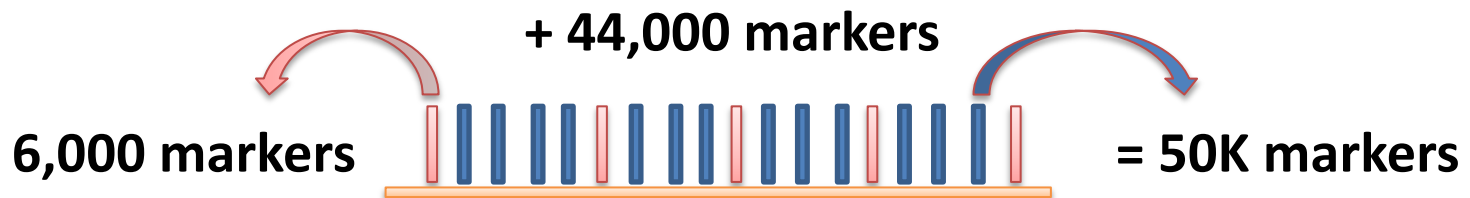
Cost to genotype a larger population?

∴ Imputation

Marker genotypes not obtained from a low density panel



Inferred from another group of animals that have been genotyped for many more markers



∴ Imputation

- Reduce costs
- Genotype more animals (Reference Populations)

1,000 animals

x \$90 = \$90,000

6K to 50K

1,000 animals

x \$50 = \$50,000

\$40,000 = **800 additional animals**

1,000 animals

x \$220 = \$220,000

50K to HD

1,000 animals

x \$90 = \$90,000

\$130,000 = **1,444 additional animals**

∴ Imputation

- Some imputation methods are based upon the **Hidden Markov model** and have computational constraints due to an **intensive sampling process**
- Cattle: **All individuals are related** and, therefore, **share haplotypes** which may differ in length and frequency based on their relationships
- The method starts with family imputation if pedigree information is available, and then exploits close relationships by searching for long haplotype matches in the reference group using overlapping sliding windows. The search continues as the window size is shrunk in each chromosome sweep in order to capture more distant relationships

∴ Imputation

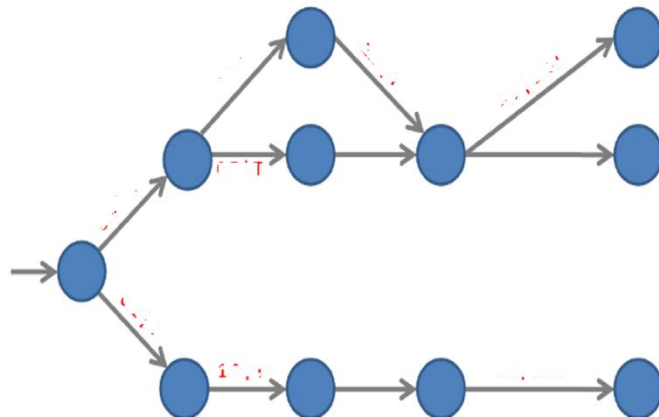
- 1) Family imputation (pedigree information)
2) Population (linkage disequilibrium) methods
- **LDP is sparse**: They are reasonably accurate
- They do not make use of close relationships directly. However, they can still capture close relationships between individuals by finding long shared haplotypes
- **Genealogy plays a very important role in phasing and imputation**
- **At the haplotype level**, close relatives share **longer haplotypes** that **have lower frequency in the population**
- Distant relatives share **shorter haplotypes** which usually have higher frequency

∴ Imputation

- Most rare variants (e.g. $MAF < 0.05$) tend to be recent and are associated with longer haplotypes (Benefit of using FIMPUTE)

FIMPUTE

BEAGLE



∴ Imputation

For denser panels, FIMPUTE performed better: Could find shared haplotypes between distant relatives with greater precision.

50:300K (HD) = FIMPUTE and IMPUTE2 performed better than BEAGLE

Inclusion of very close relative in the reference group substantially increased the imputation accuracy. The increase in imputation accuracy was more evident for 3K to 50K. As the panel become denser, the importance of having reference individuals with close relationship to the target animals decreased.

Genome Canada Project

Group 1

30 Seq
500 HD Affy
500 50K Illumina

Group 2

30 Seq
500 HD Affy
500 50K Illumina

...

Group 10

30 Seq
500 HD Affy
500 50K Illumina

Other Projects

6K
8K
Different densities

~10,000 animals - full sequence (imputation)

Imputation from LD to 50K, HD and sequence
Breed composition

Canadian Simmental Pedigree Analyses



Ancestor	Sex	Birth Year	Marginal	Cumulative	Progeny	Descendants
SIGNAL	M	1969	0.0522	0.0522	4396	563711
SIEGFRIED	M	1971	0.0413	0.0936	99	438087
MR CLEAN	M	1976	0.0405	0.1341	251	242652
LACOMBE ACHILLES	M	1969	0.0349	0.169	5502	623802
GALANT	M	1968	0.0302	0.1992	9031	749880
SIM-ROC C&B	M	1974	0.0264	0.2257	633	346346
WESTERN WBL EXTRA 1A						
WBL EXTRA 1A	M	1969	0.0237	0.2494	5040	626064
KING ARTHUR	M	1971	0.0229	0.2722	463	344840
ABR SHINAGH VIOLET	F	1973	0.0212	0.2934	4	372880
KINGFIELD ABRICOT	M	1971	0.0202	0.3136	3827	458295

BIO Projects : ALGP26 Genome Alberta

- Project: “Making DNA Analysis Results useful to Producers through the Development of EPD’s that Incorporate Genomic and Performance Information”*



BIO Projects : ALGP26

Genome Canada AND ALGP26

Imputation studies using data from several partners

Sample collection

Methods for estimation of marker effects of crossbred animals

- Not every DNA sample is of the same quality
- There are missing markers in every sample
- Quality test is needed to ensure good prediction equations



500 samples from different locations were tested in different labs

Issues regarding Nasal Swab: 47% failed

Hair sample adopted as default

Imputation results of purebred and crossbred animals



~ 1,000 animals: Angus and Charolais

~ 2,400 animals: Crossbred animals

~ 1,700 animals: Crossbred animals from different sources including project partners



5032 ANIMALS (50K panel)

Imputation results of purebred and crossbred animals



JOURNAL OF ANIMAL SCIENCE

The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science

Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle

R. V. Ventura, D. Lu, F. S. Schenkel, Z. Wang, C. Li and S. P. Miller

J ANIM SCI 2014, 92:1433-1444.

doi: 10.2527/jas.2013-6638 originally published online March 18, 2014

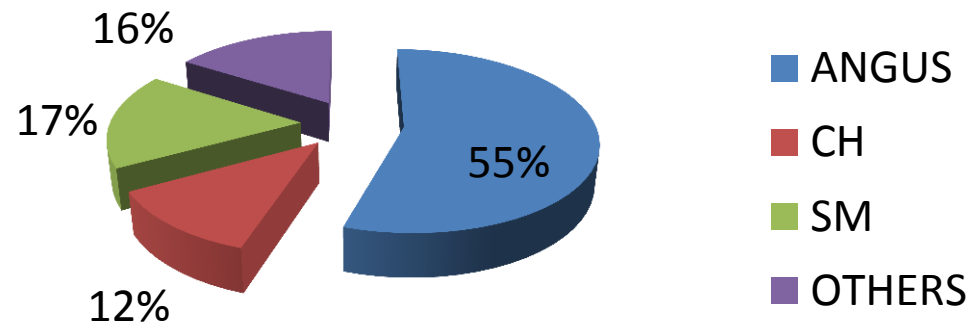
Imputation results of purebred and crossbred animals

Table 1. Description of different scenarios for imputation from 6K to 50K using different groups of purebred and crossbred animals

Scenario	No. reference animals	No. imputed animals	Reference population description	Imputed animals description ¹
A	350	100	Oldest Angus	Angus
B	313	100	Oldest Angus – 37 key closer relatives	Angus
C	700	100	Oldest Angus + oldest Charolais	Angus
D	4,932	100	All data available	Angus
E	350	100	Oldest Charolais	Charolais
F	700	100	Oldest Angus + oldest Charolais	Charolais
G	400	100	Oldest Kinsella cattle	Kinsella cattle
H	750	100	Oldest Kinsella cattle + oldest Angus	Kinsella cattle
I	750	100	Oldest Kinsella cattle + oldest Charolais	Kinsella cattle
J	700	100	Oldest Angus + oldest Charolais	Kinsella cattle
K	250	146	Oldest Elora cattle	Elora cattle
L	750	146	Oldest Elora cattle	Elora cattle
M	1,500	146	Oldest Elora cattle	Elora cattle
N	4,886	146	All data available	Elora cattle
O	350	146	Oldest Angus	Elora cattle
P	350	146	Oldest Charolais	Elora cattle

¹Youngest animals of each scenario were imputed.

Average breed composition of the reference set (ELORA)



Imputation of purebred Animals



Angus – AgriFood Canada

Charolais – AgriFood Canada



Addition of closer relatives in the reference set

Imputing Angus (AFC)

Reference Population: 345 (Angus AFC)

Imputed animals: 100 Angus (Youngest)

	Run Time	Overall Acc.
Beagle	7 m	0.956426
Fimpute	31 s	0.971443

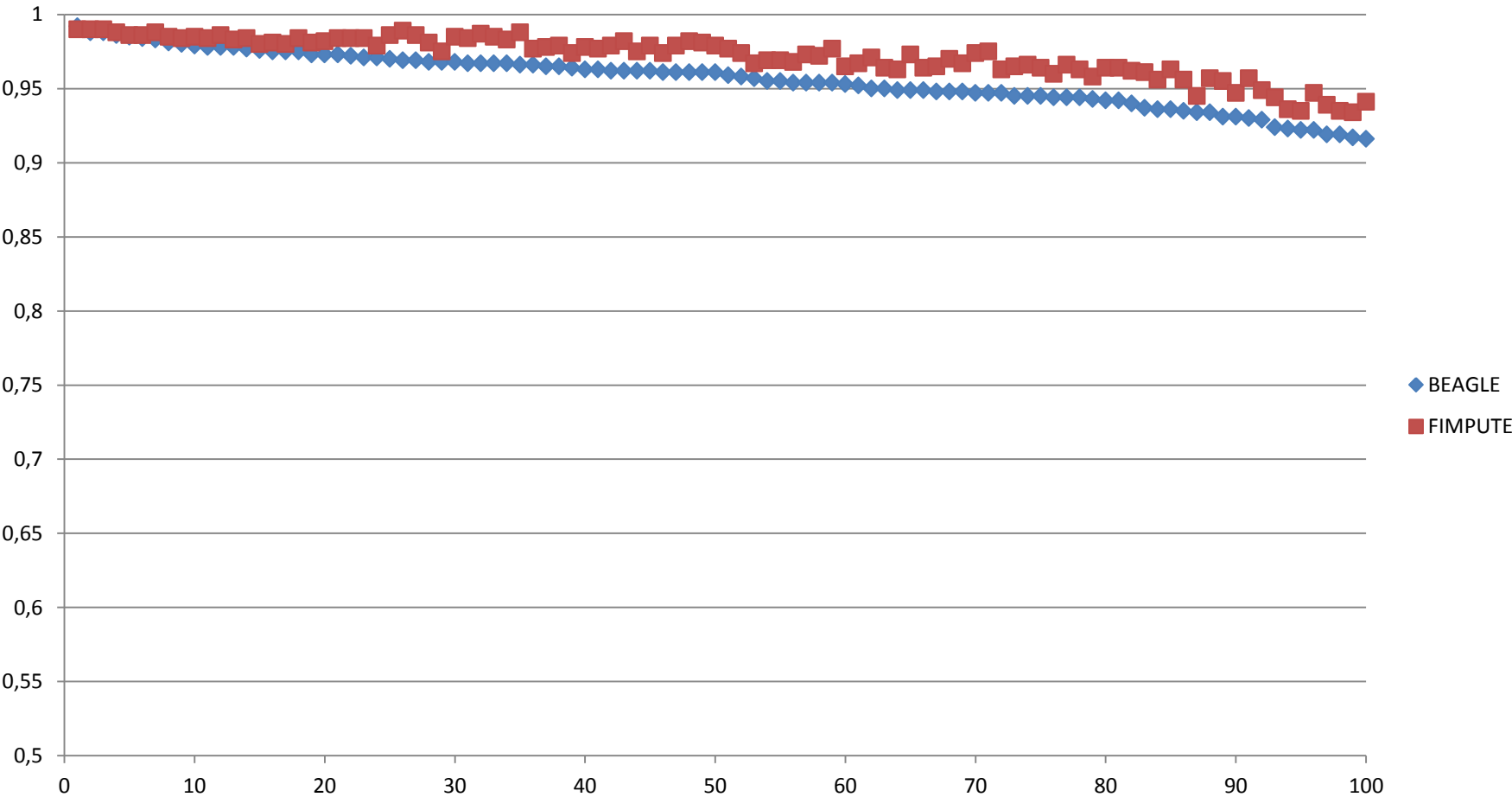


Fig: Imputation accuracy of each animal

Imputing Angus (AFC)

Reference Population: 345 (Angus AFC)

– 37 Key SIREs = 308

Imputed animals: 100 Angus (Youngest)

	Run Time	Overall Acc.
Beagle	6m36s	0.932971
FImpute	29s	0.942259
		0.956426
		0.971443

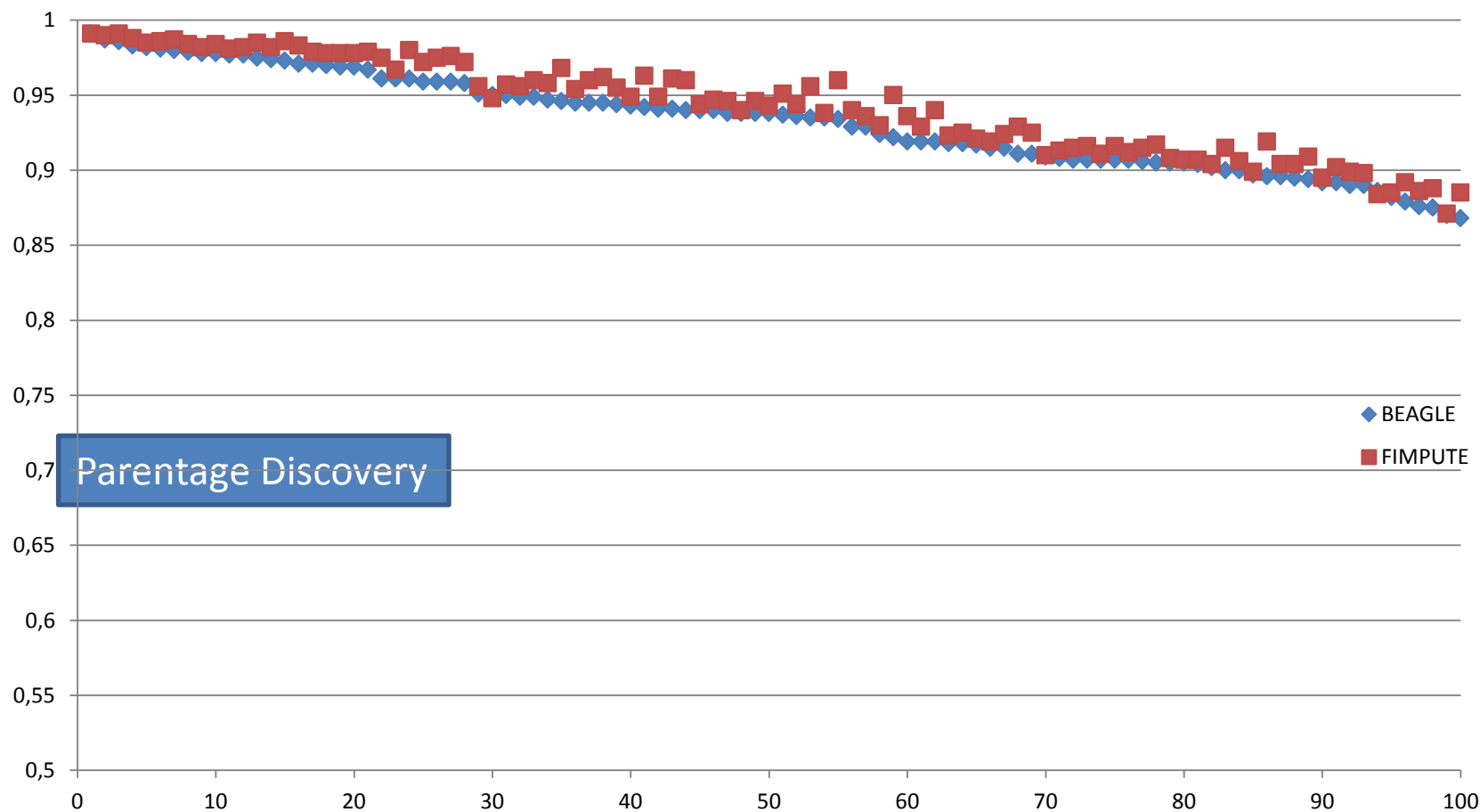


Fig: Imputation accuracy of each animal

Imputing Angus (AFC)

Reference Population: 345 (Angus AFC)
+ 350 (Charolais AFC)

	Run Time	Overall Acc.
Beagle	25m39s	0.944472
Fimpute	1m38s	0.96964

Imputed animals: 100 Angus (Youngest)

0.956426
0.971443

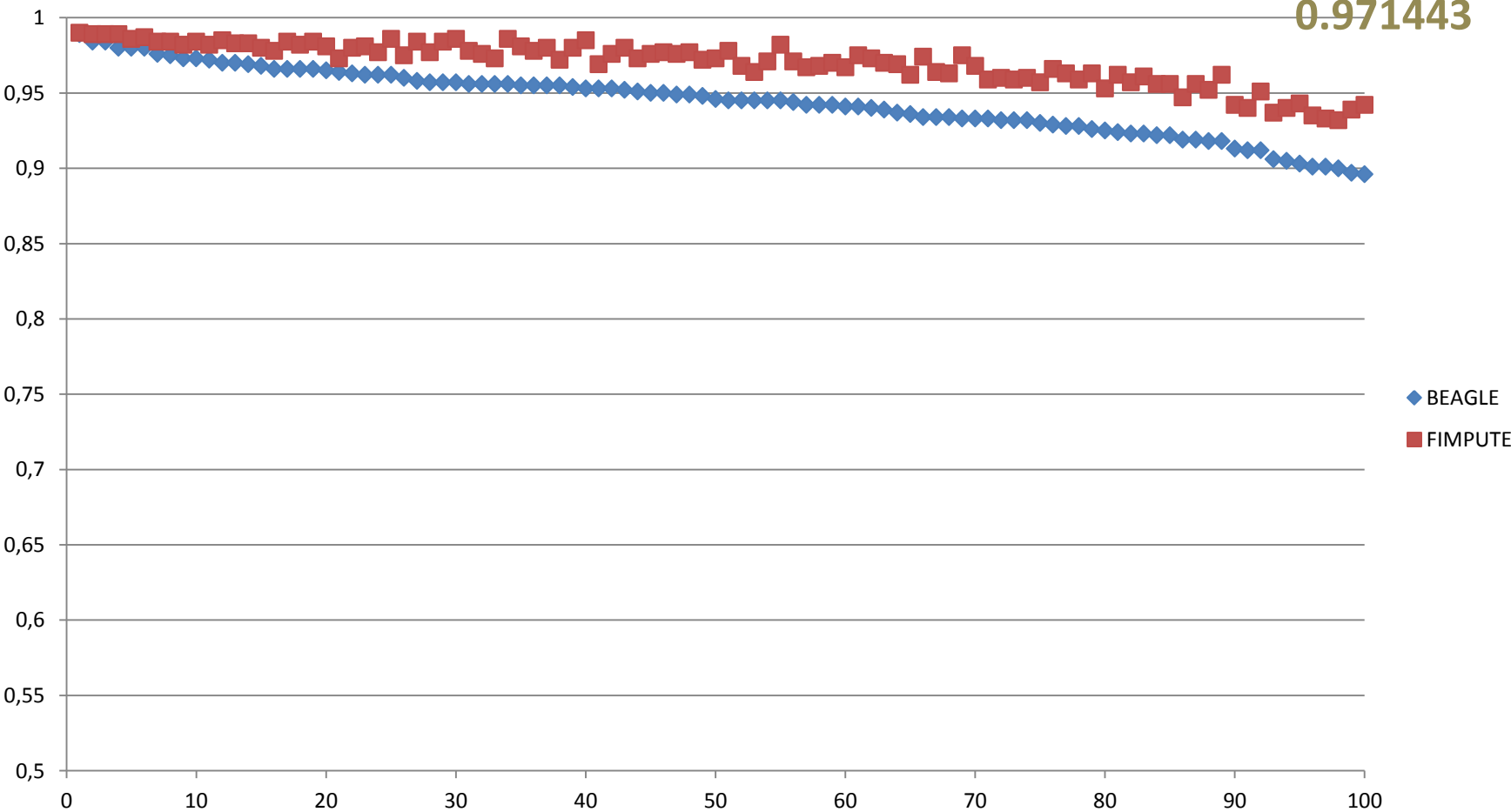


Fig: Imputation accuracy of each animal

Imputing Angus (AFC)

Reference Population: 345 (Angus AFC)
+ All data available = 4932

Imputed animals: 100 Angus (Youngest)

Run Time

Overall Acc.

Beagle 6h32m59s 0.963848

Fimpute 27m27s 0.974666

0.956426

0.971443

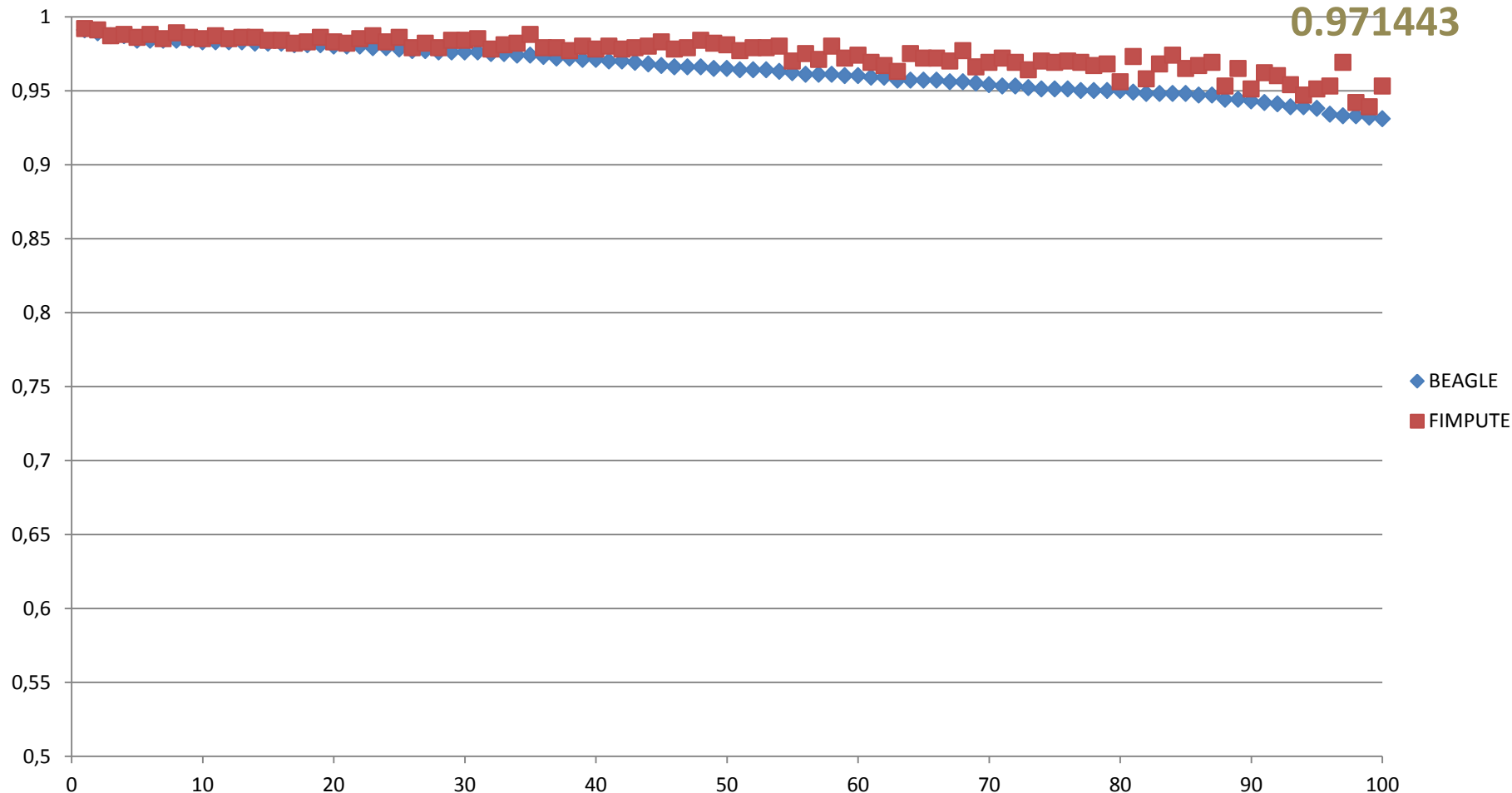
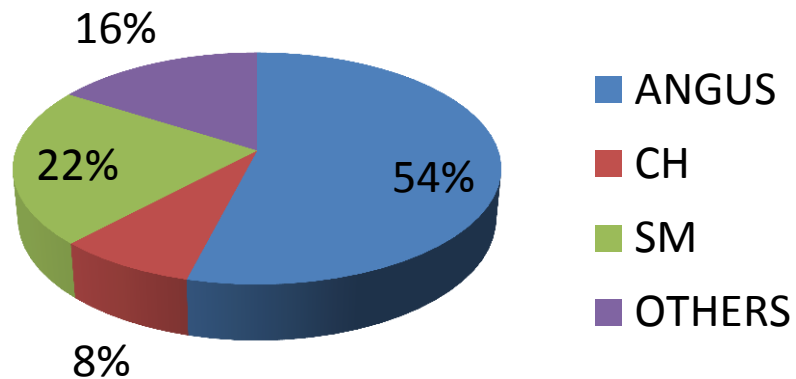


Fig: Imputation accuracy of each animal

Imputation of crossbred animals from Elora cattle – UoG

Average breed composition of Imputed animals (ELORA)



Imputing Elora UoG

Reference Population: 750 Elora

Run Time

Overall Acc.

Beagle 39m32s

0.8786

Fimpute 1m56s

0.923413

Imputed animals: 146 Elora (Youngest)

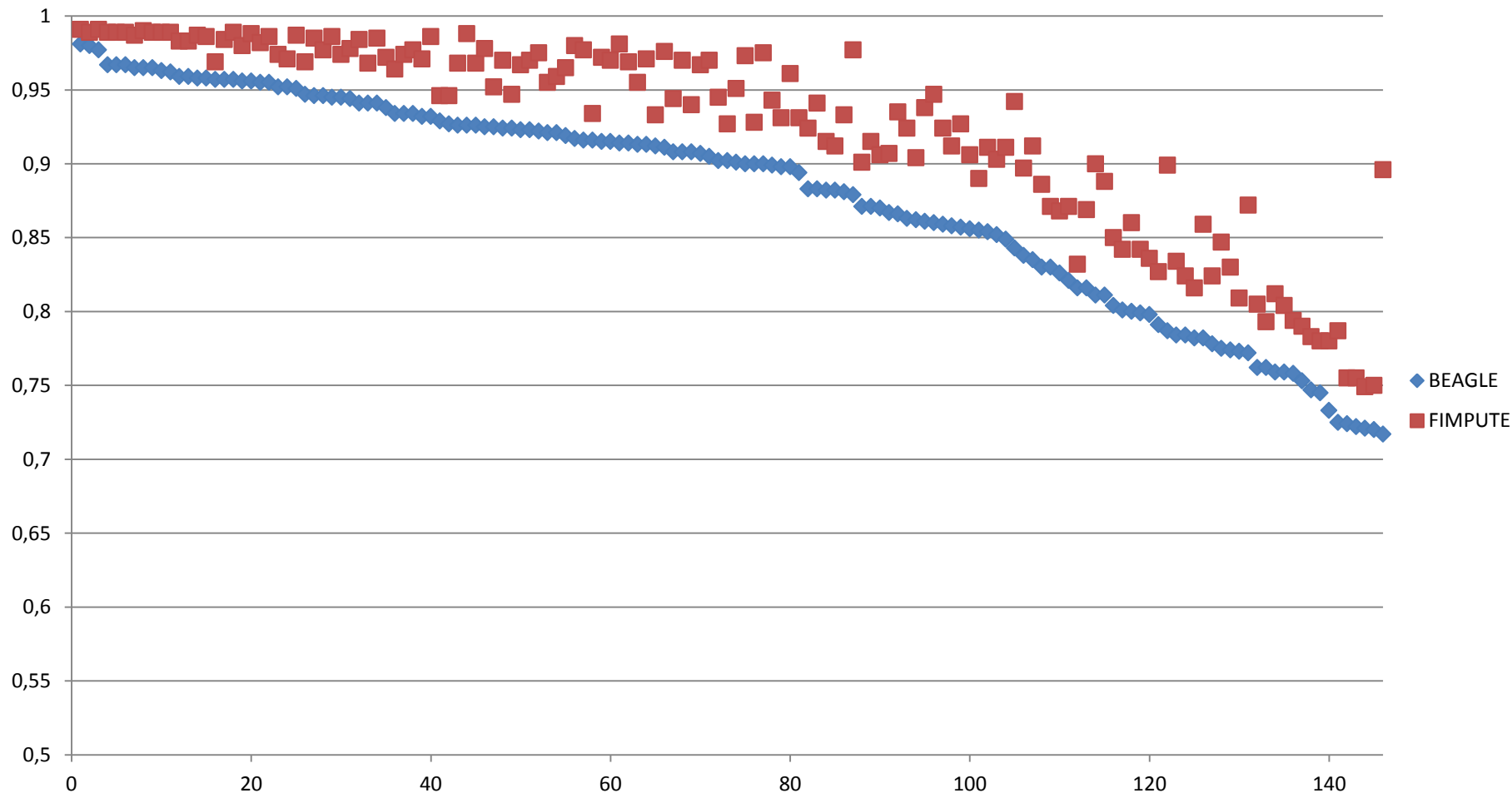


Fig: Imputation accuracy of each animal

Imputing Elora UoG

Reference Population: 1500 Elora

Imputed animals: 146 Elora (Youngest)

Run Time

Overall Acc.

Beagle 2h00m59s 0.911333

Fimpute 3m54s 0.938717

0.8786

0.923413

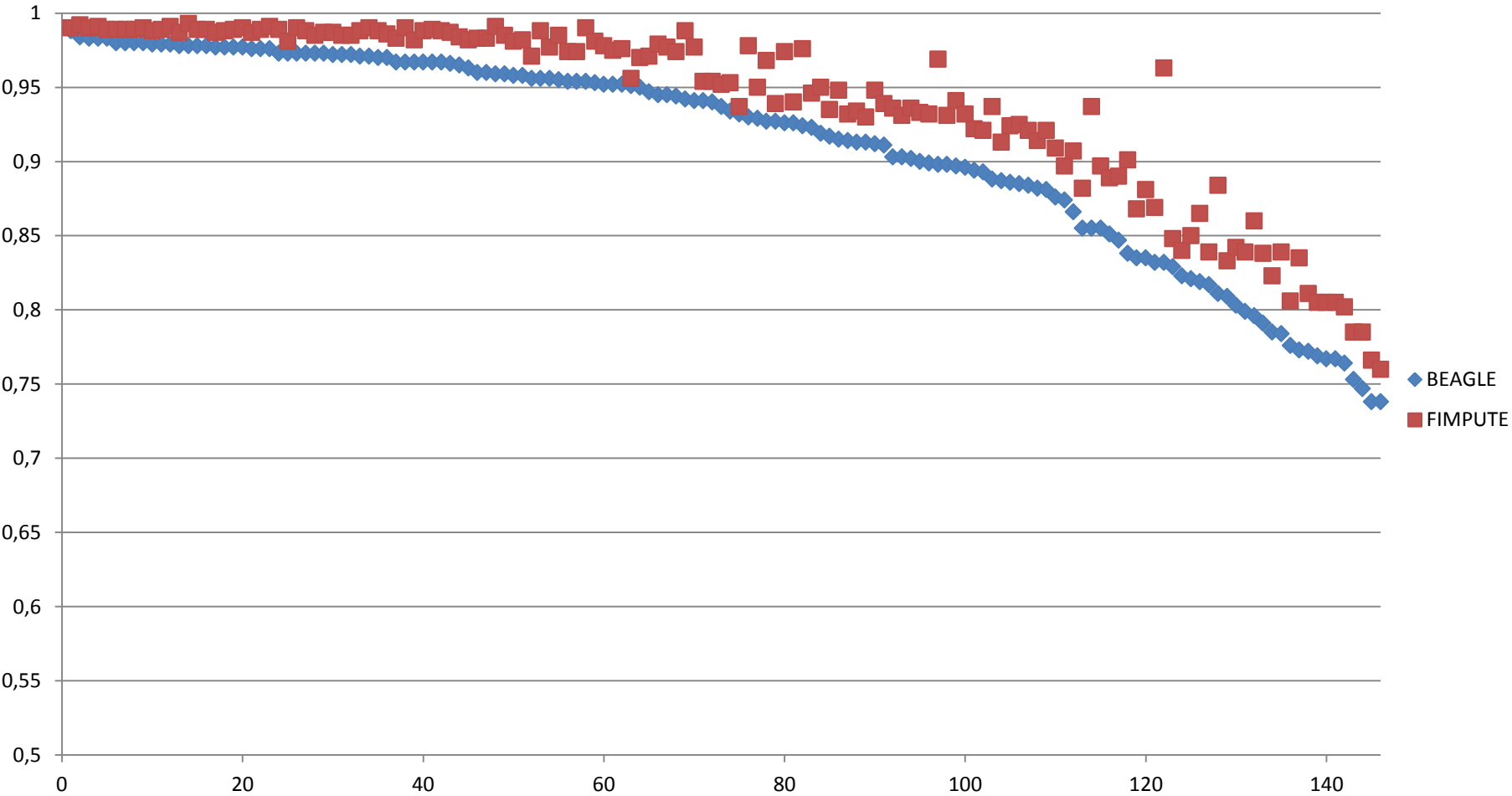


Fig: Imputation accuracy of each animal

Imputing Elora UoG

Reference Population: 1500 Elora
+ All data = 4886

Imputed animals: 146 Elora (Youngest)

Run Time

Overall Acc.

Beagle 7h35m43s 0.956468

Fimpute 28m35s 0.972171

0.911333

0.938717

0.8786

0.923413

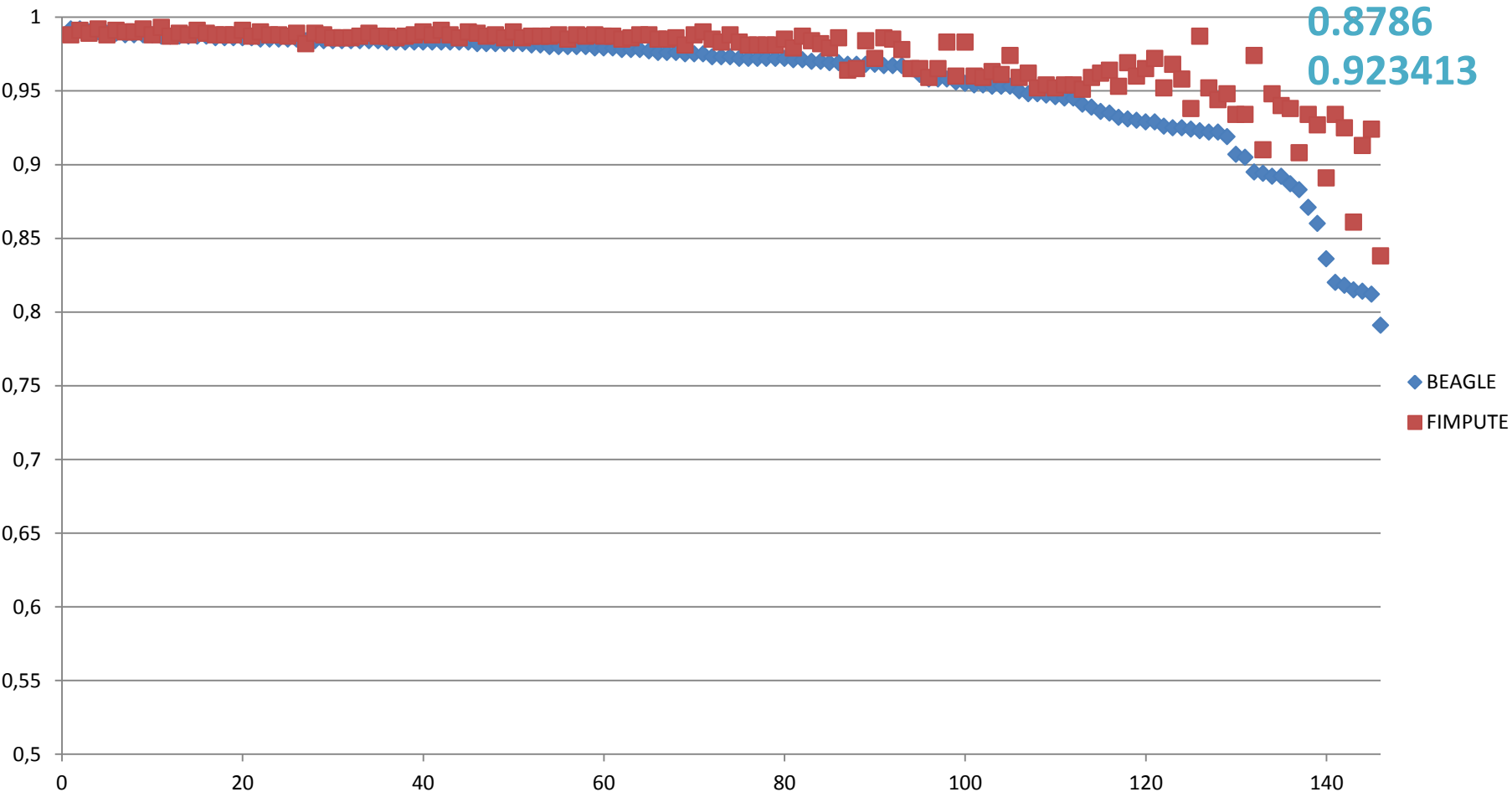


Fig: Imputation accuracy of each animal

Imputing Elora UoG

Reference Population: 350 Angus

Run Time

Beagle 17m48s

Fimpute 55s

Overall Acc.

0.698713

0.691243

Imputed animals: 146 Elora (Youngest)

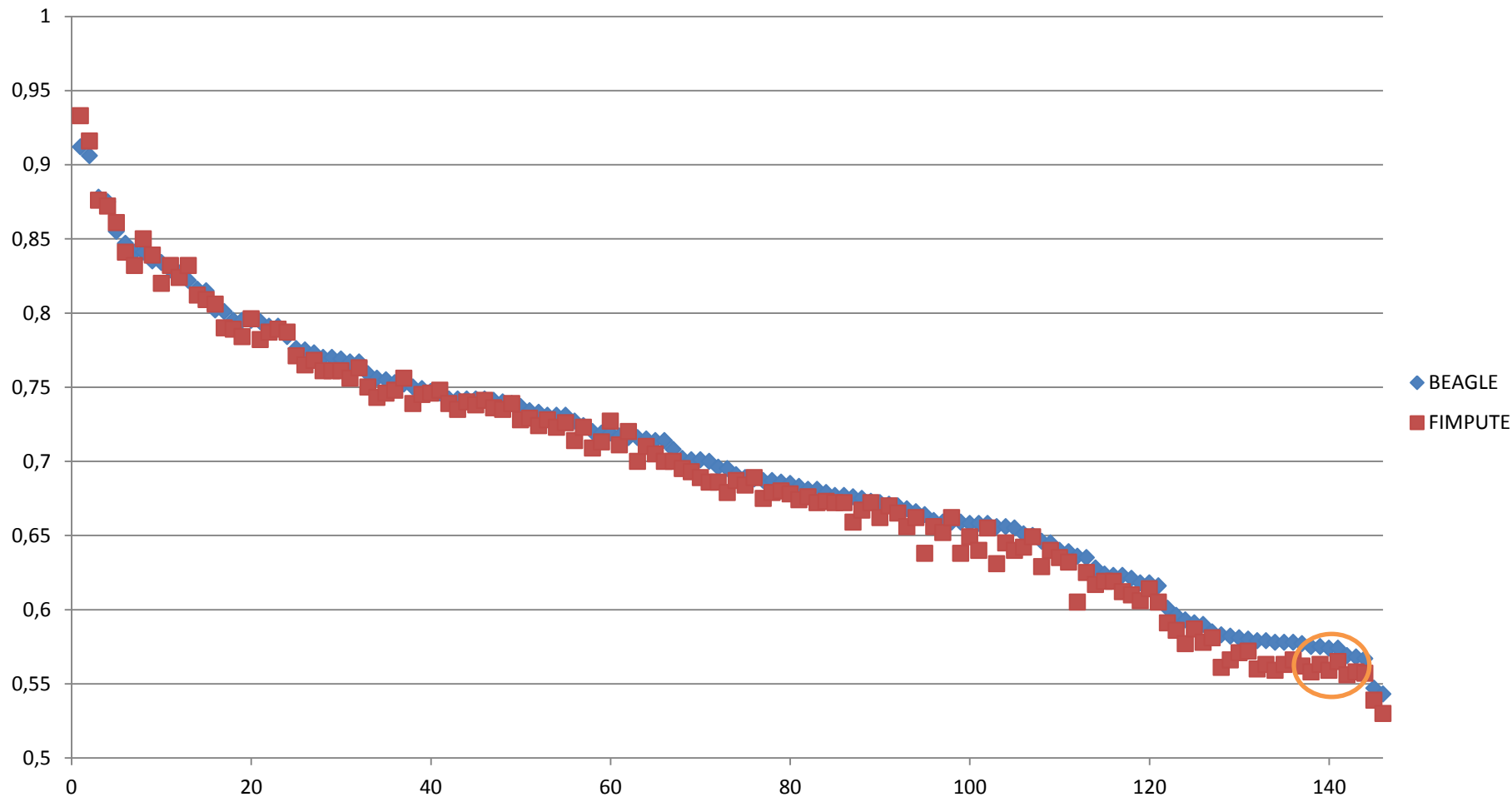


Fig: Imputation accuracy of each animal

Imputing Elora UoG

Reference Population: 350 Charolais

Run Time

Beagle 17m44s

Fimpute 1m02s

Overall Acc.

0.541534

0.54072

Imputed animals: 146 Elora (Youngest)

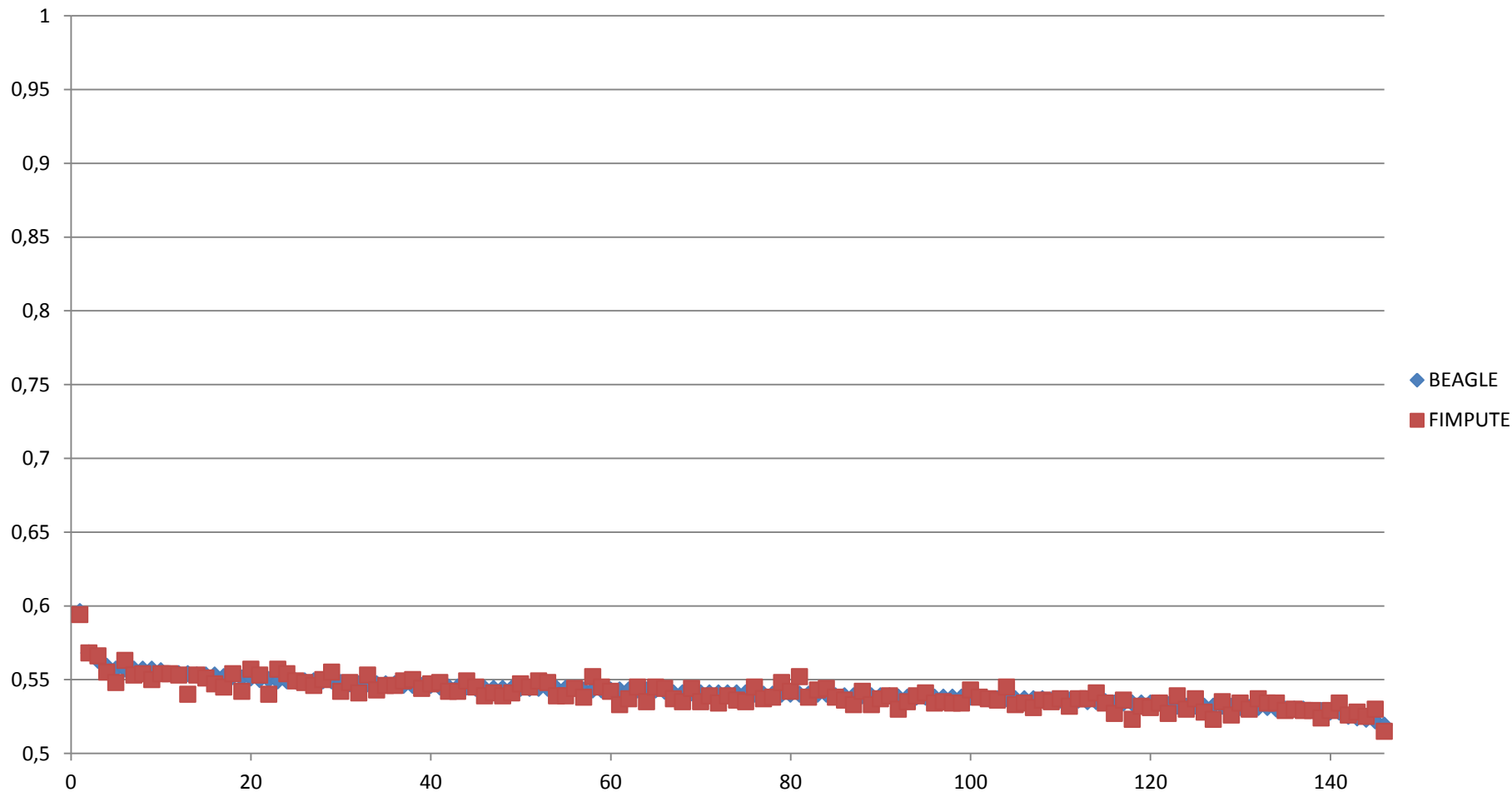
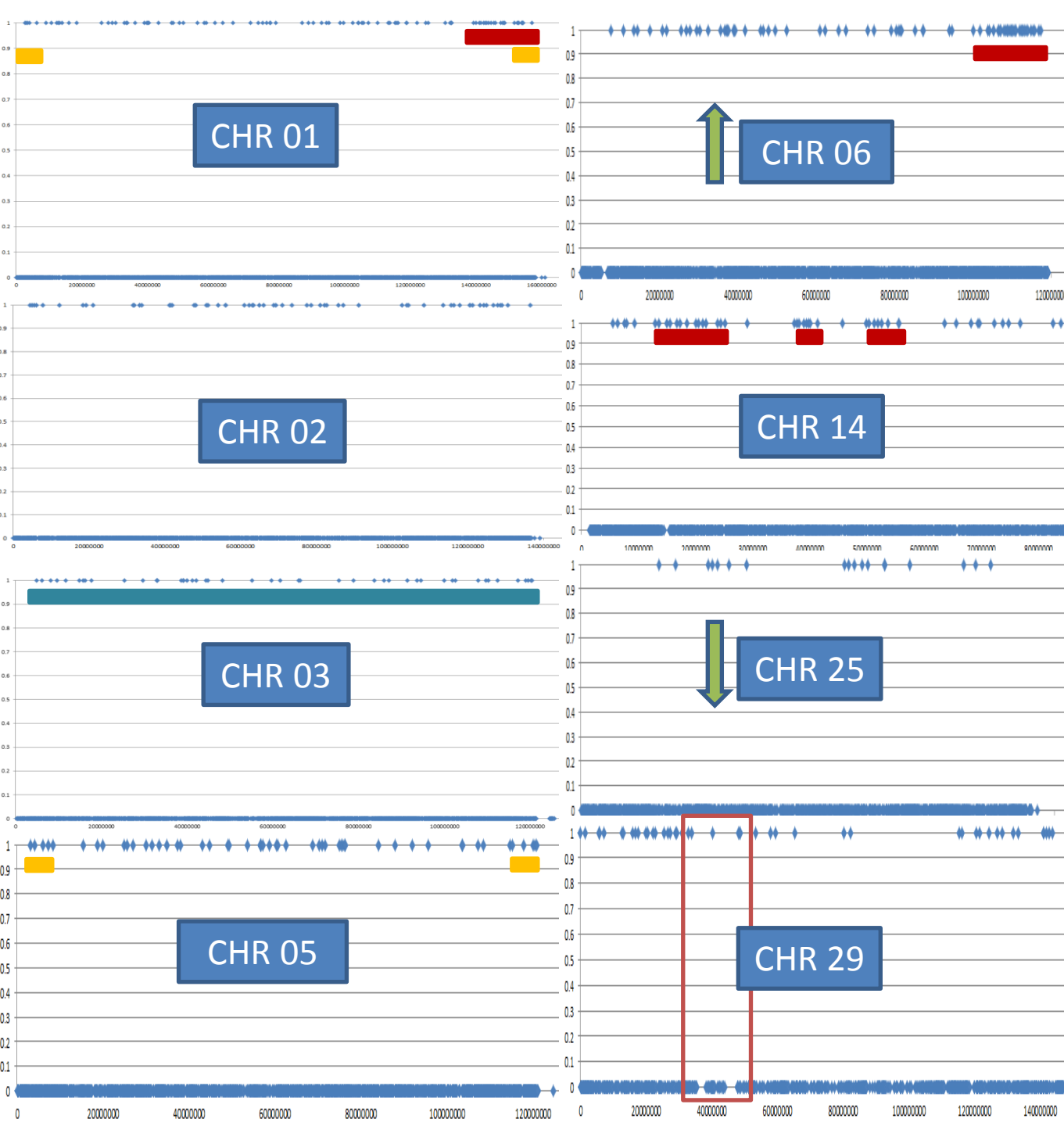


Fig: Imputation accuracy of each animal



GGP: (GeneSeek Genomic Profiler)

- Approximately 400 SNPs to enable conversion of ISAG microsatellite parentage data.
- Disease and trait markers were included
- 1174 MARKERS added for imputation

**1071/1174 (92%)
REMAINED AFTER
QUALITY CONTROL**

Reference Population: 1500 Elora

+ All data = 4886

Imputed animals: 146 Elora (Youngest)

∴ GGP

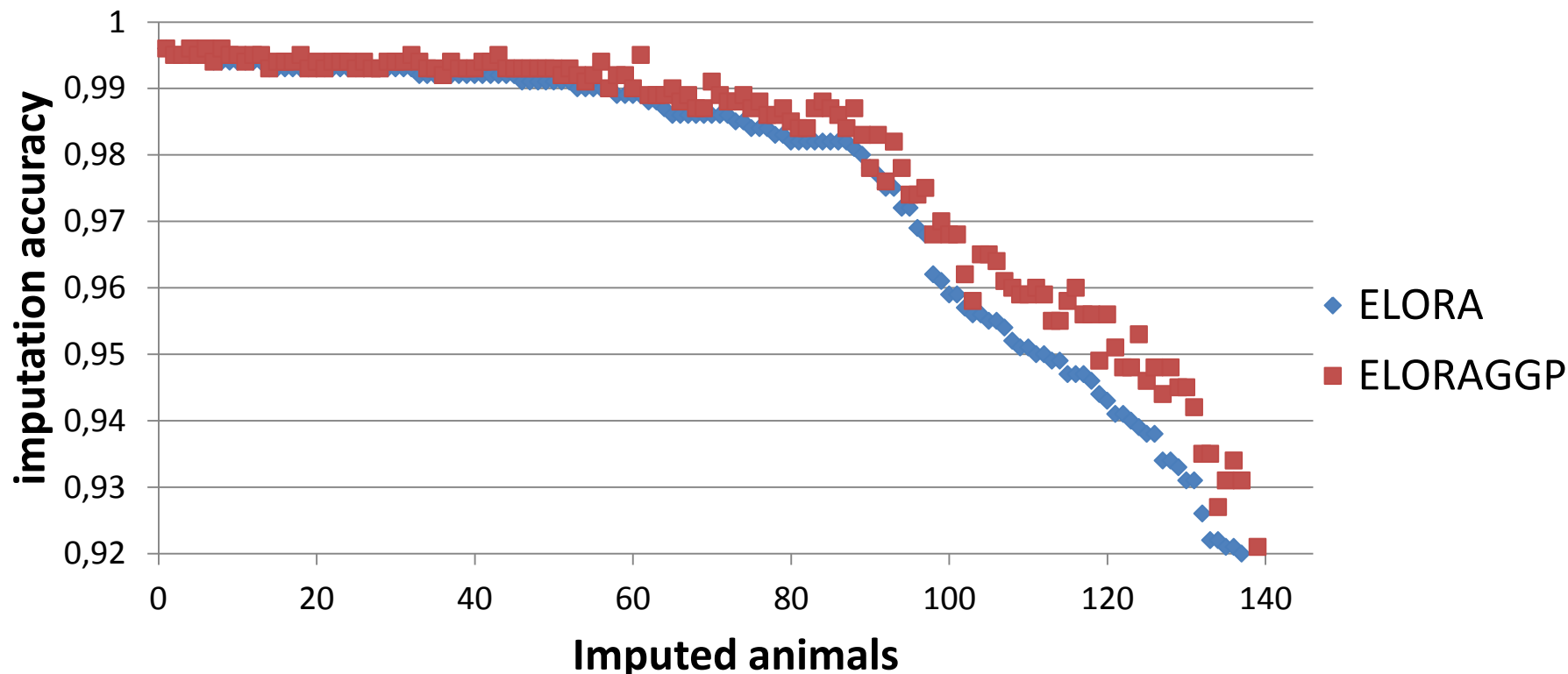
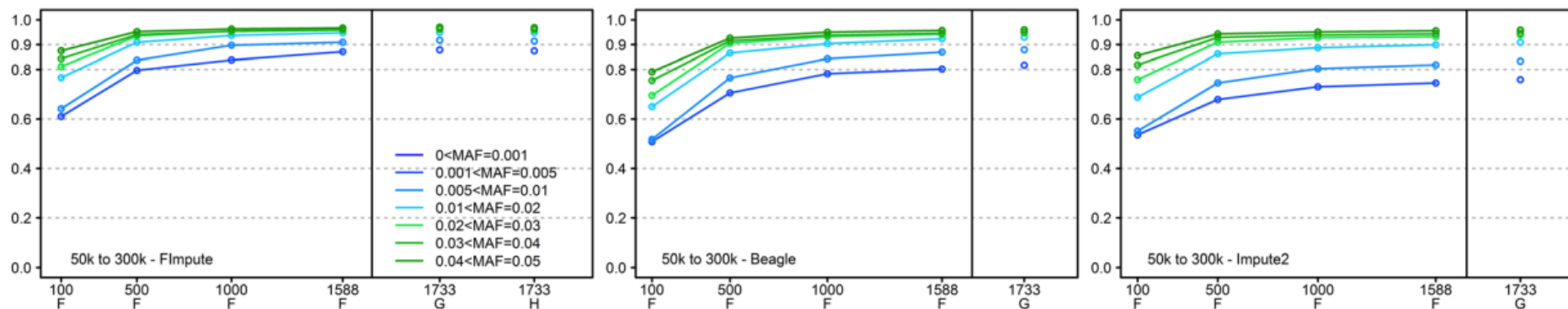
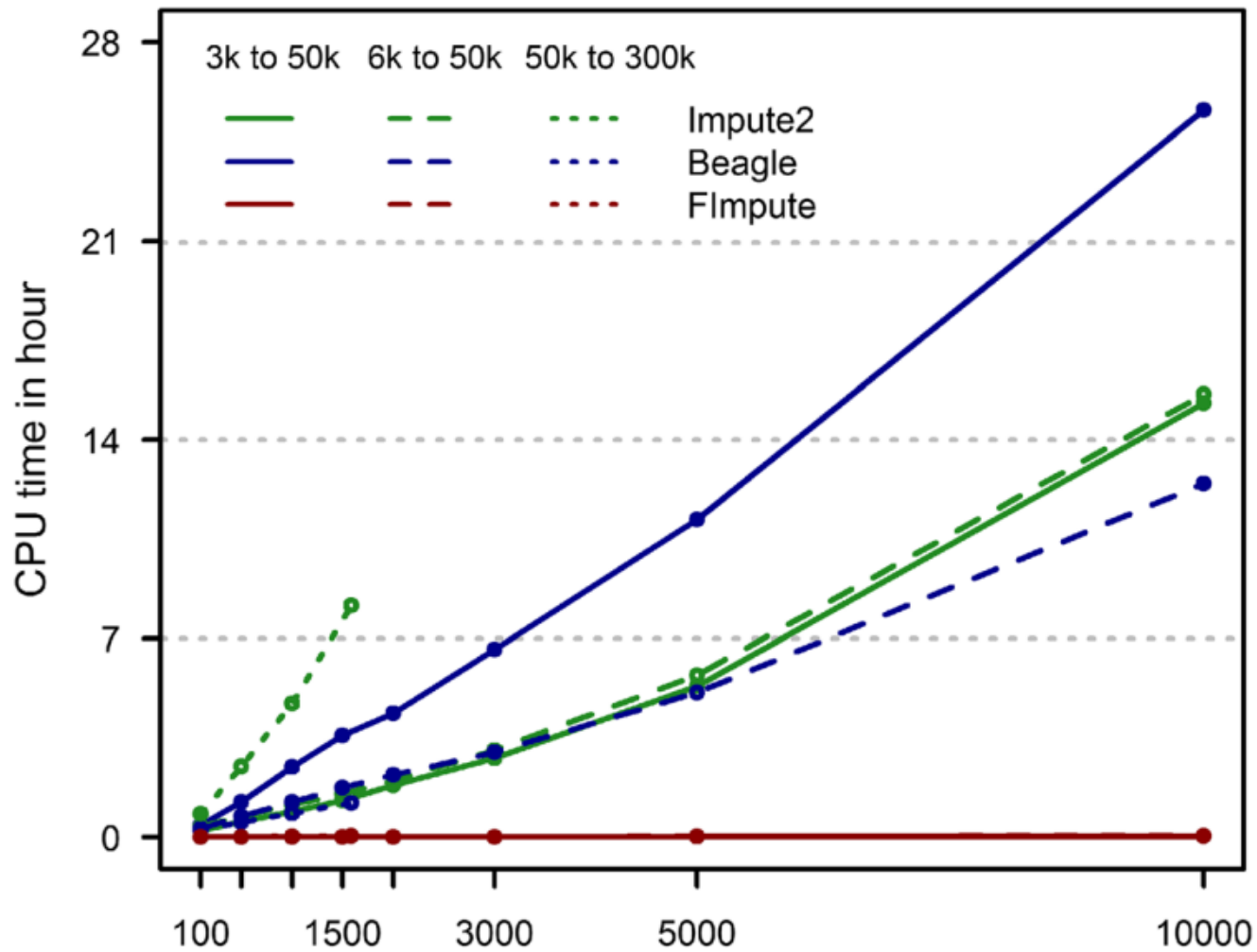


Figure 1. Imputation of 50k SNPs from 6k and 8K (GGP panel) using FIMPUTE software on crossbred animals (ELORA - UoG).

Imputação de alelos raros





Panels

1,261,128 SNPs

1,426,836 SNPs



777,962 SNPs

648,874 SNPs



165,708



777,962 SNPs

648,874 SNPs



96 animals

Multibreed population

54,609 SNPs

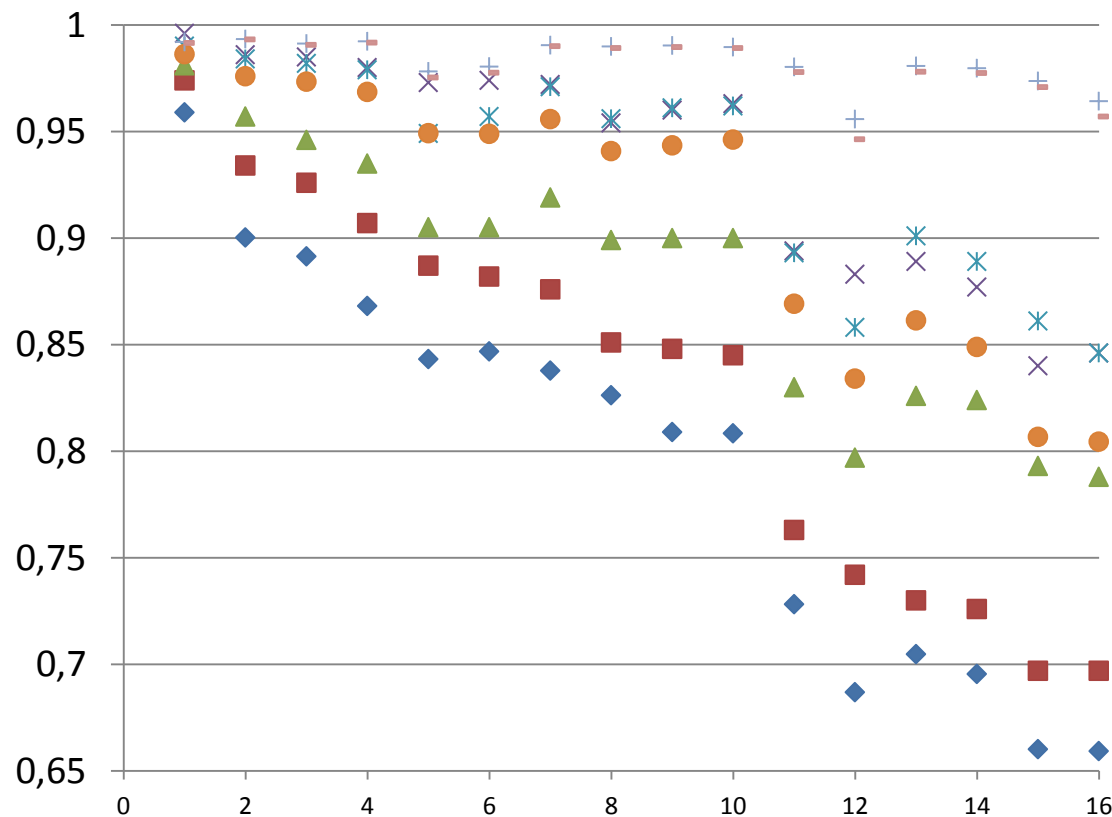


6,909 SNPs

Reference Population: 80 animals

Imputed group: 16 animals

Software: FImpute

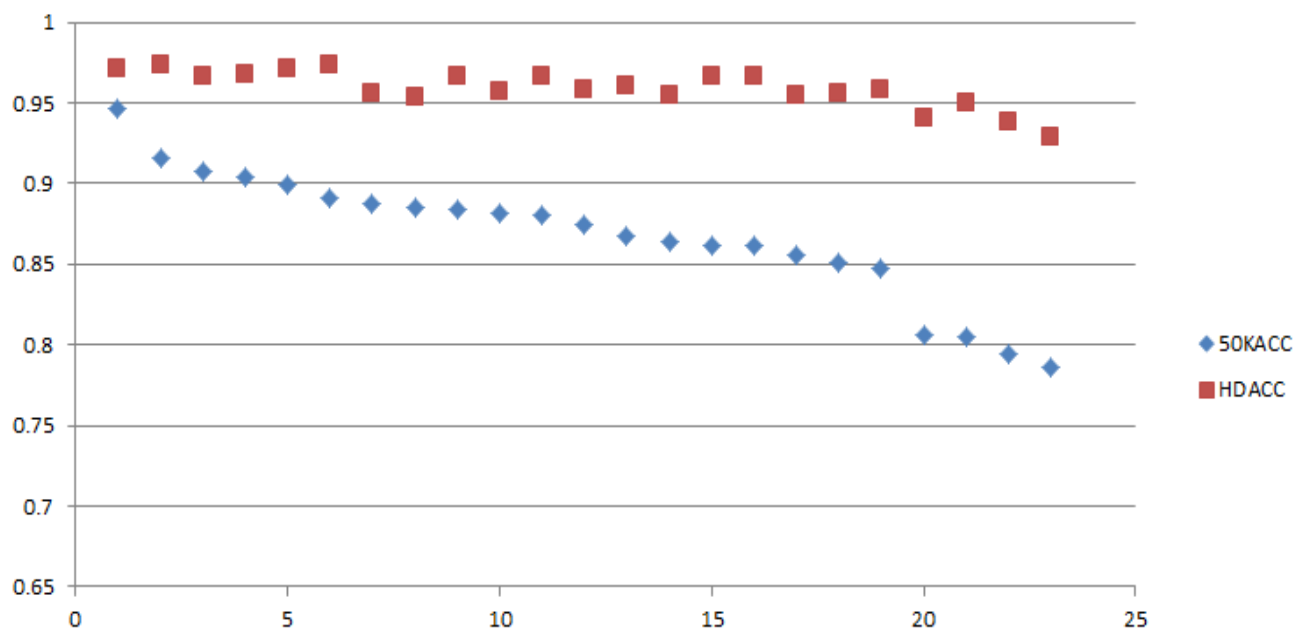


6K-Combined	79.53%
6K-Illumina	83.03%
6K-Affymetrix	88.15%
50K-Illumina	93.57%
50K-Affymetrix	93.36%
50K-Combined	91.33%
Illumina-Combined	98.26%
Affy-Combined	98.03%

Imputation from 50K and HD to Sequence (Ch27)

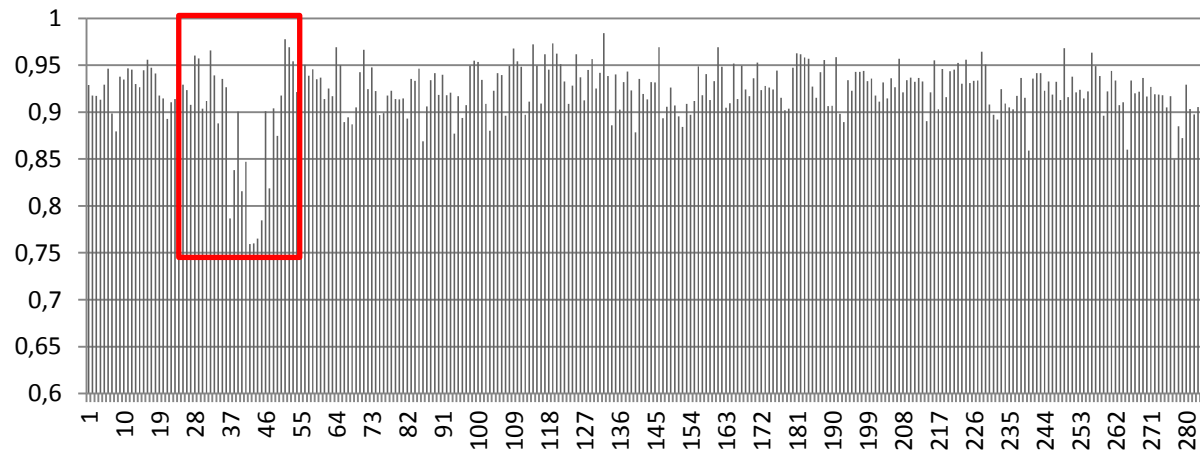
Reference population: 100 animals (66 HO, 31 SIM, 3 RED)

Imputed group: 23 animals

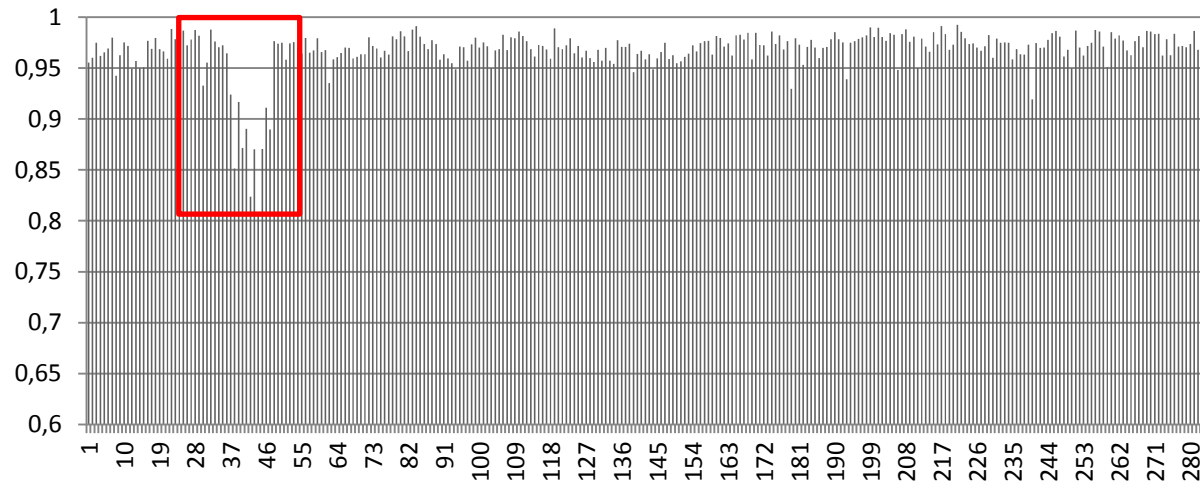


Imputation from 50K and HD to Sequence (Ch27)

SIM



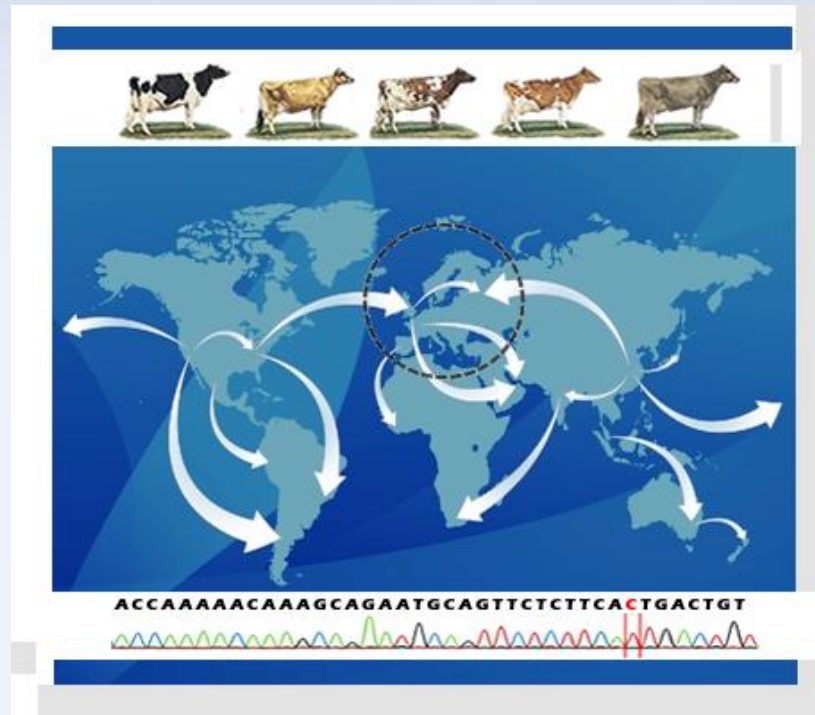
HOL



Genomic Selection of crossbred animals

- 1) How many breeds?**
- 2) Are the breeds related?**
- 2) Illumina or Affymetrix?**

Multi-breed Evaluation: HO, AY, BS and JE (Results from 2011)



Assessing admixture among 3 dairy cattle breeds using 50K SNP chip

Ricardo Ventura

Flavio Schenkel

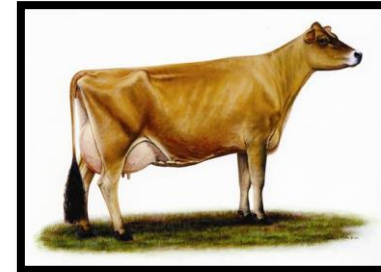
CGIL - UOGUELPH

2011

Some breeds have small number of proven sires

Purpose:

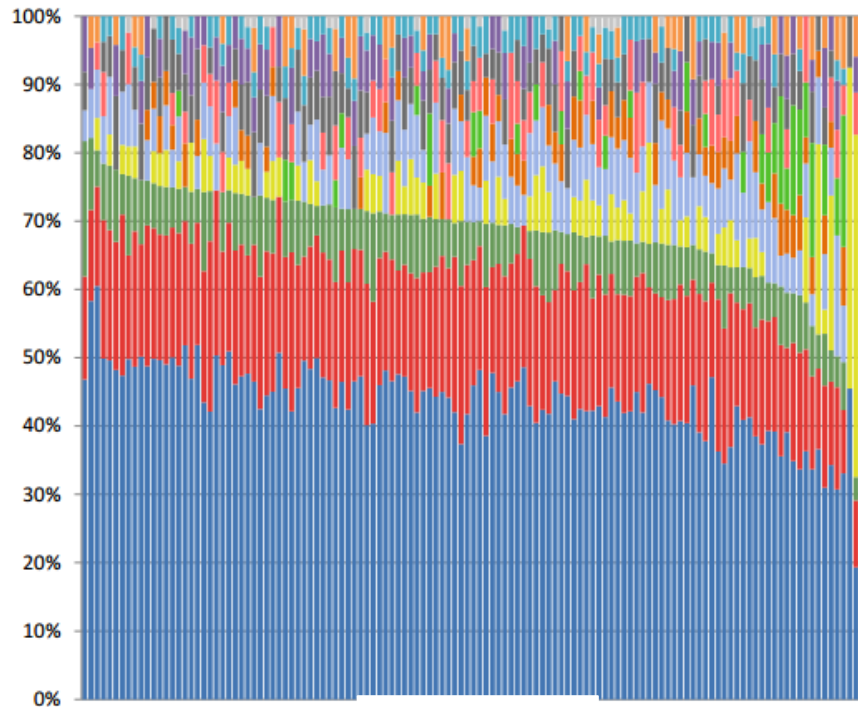
Analyse the level of **admixture** among breeds using clustering methods: Which breeds could be likely used together in a **Multibreed Genomic Evaluation**?



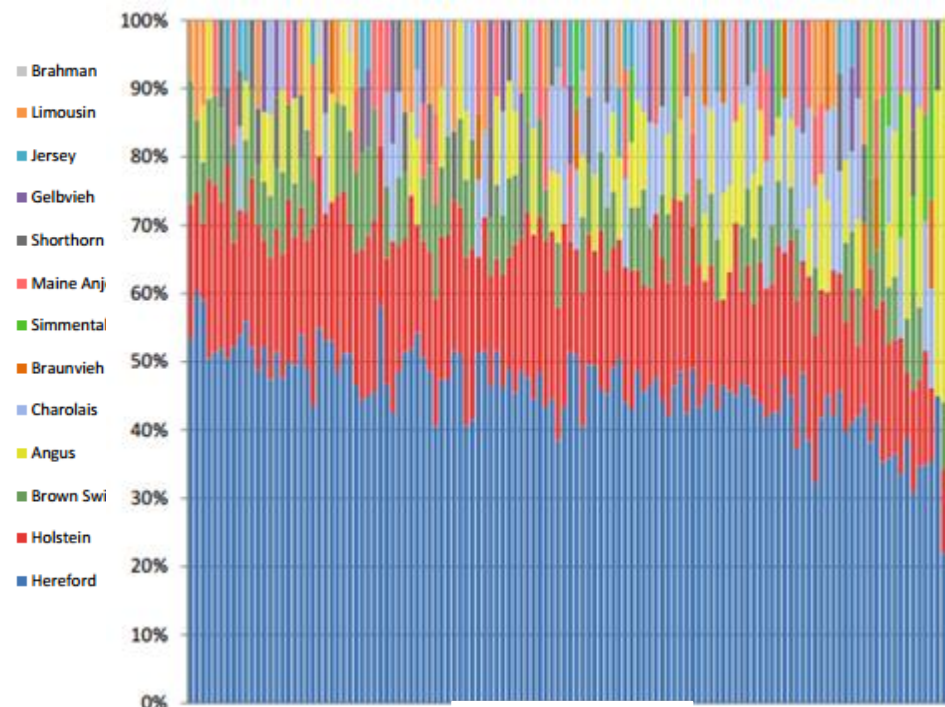
Determining Breed composition using SNP panels



- Traditional EPDs adjustments



50K



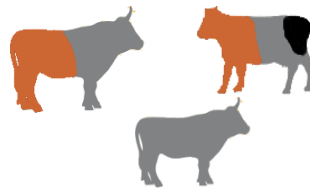
6K

Fleming et al. 2013

Comparison of high-density genomic relationship matrices in beef cattle



illumina®
777,962 SNP



N = 96

Affymetrix®
648,874 SNP

$$G = ZZ' / k \quad (\text{G_VanRaden}) \quad (\text{VanRaden, 2008})$$

With: Z center using allele frequencies estimated from the genotyped individuals

$$K = 2 \sum (p * (1 - p))$$

Euclidian Distance Matrix (EDM)

$$d_{ij} = \sqrt{\sum [(x_{(ik)} - y_{(jk)}) * (x_{(ik)} - y_{(jk)})]}$$

$$edm = \exp (-d_{ij}/h)$$

X = genotype from individual *i*

y = genotype from individual *j*

i, j = animals

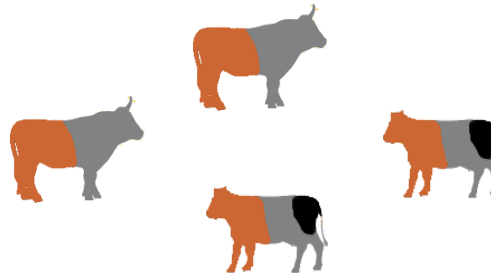
k = snps

h = tuning

Ricardo Ventura, Janusz Jamrozik, Ignacio Aguilar and Steve Miller (PAG, 2013)

Reduce number of markers: different results! Take care comparing different densities!

Parentage discovery and parentage test



1) 50K panel

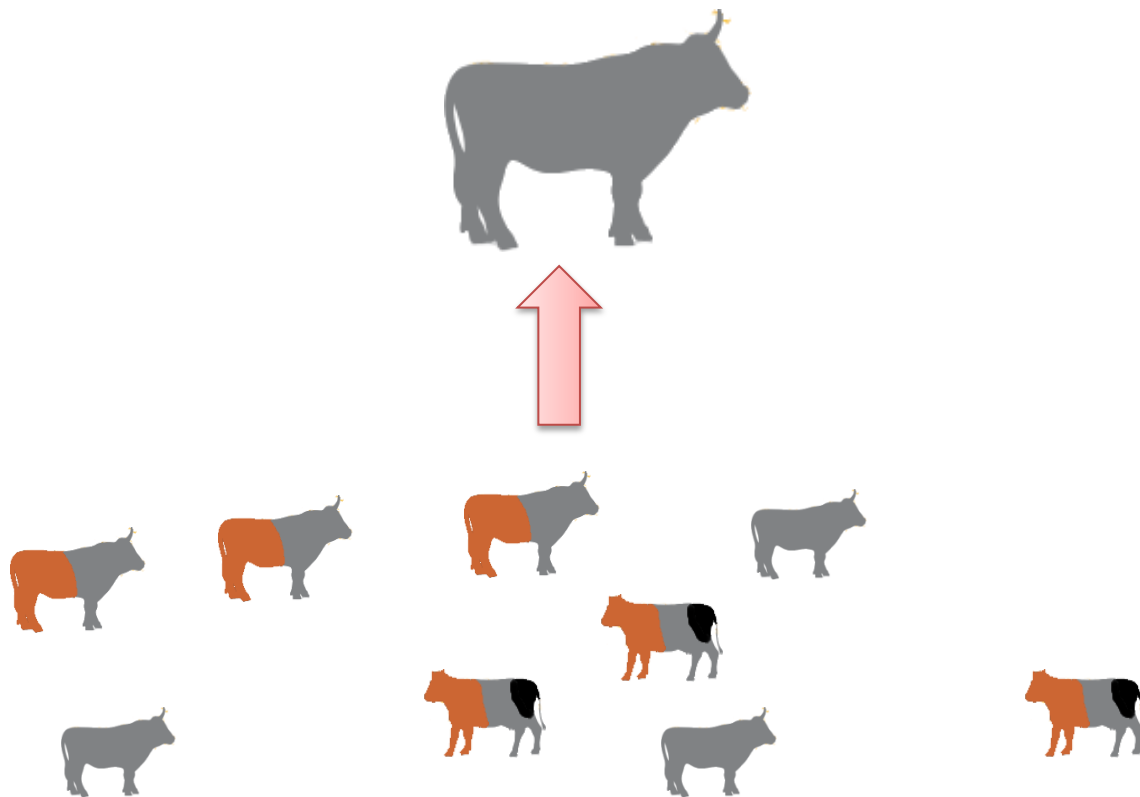


2) ~ 100 markers

BIO Protocol: Detection using 1 and confirmation using 2

<http://www.isag.us/Docs/Guideline-for-cattle-SNP-use-for-parentage-2012.pdf>

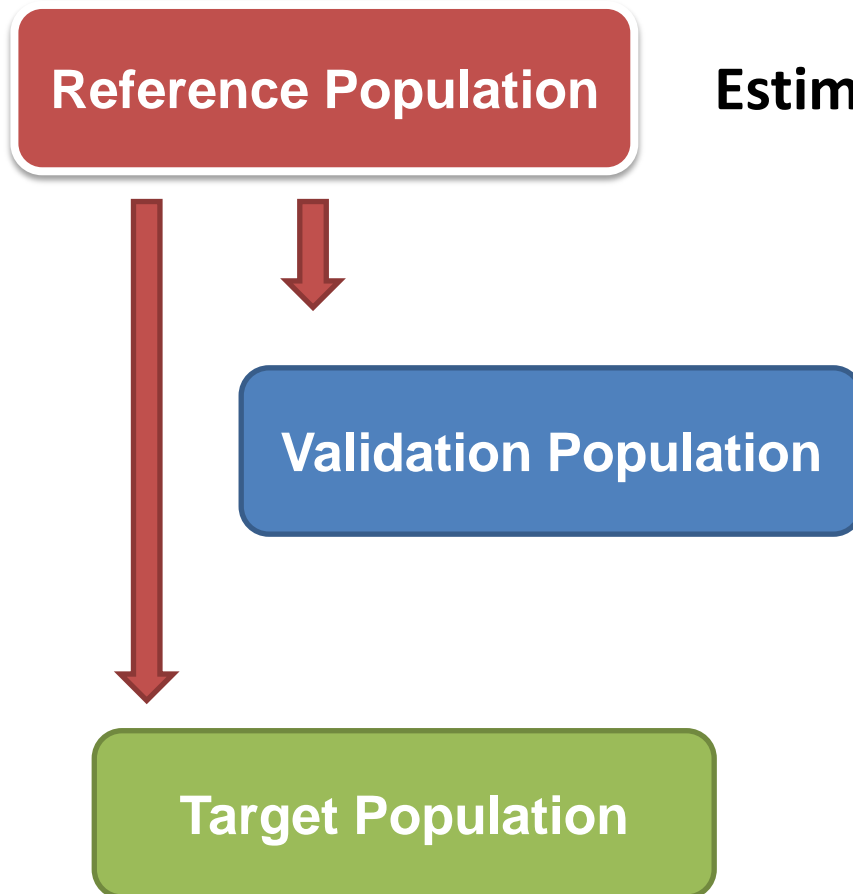
Imputation of non-genotyped animals



5 – 8 animals: not a good number for crossbred animals!

∴ Why didn't we start with new traits?

Estimation of marker effects

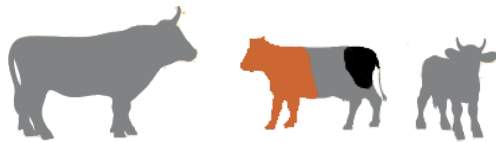


∴ Training population

LARGE NUMBER OF ANIMALS GENOTYPED



Within breed + EPDs



Purebred + Crossbred animals



Crossbred animals



ESTIMATION OF SNP EFFECTS



**MARKERS WITH ESTIMATED EFFECT
TO PREDICT MERIT OF NEW ANIMALS**

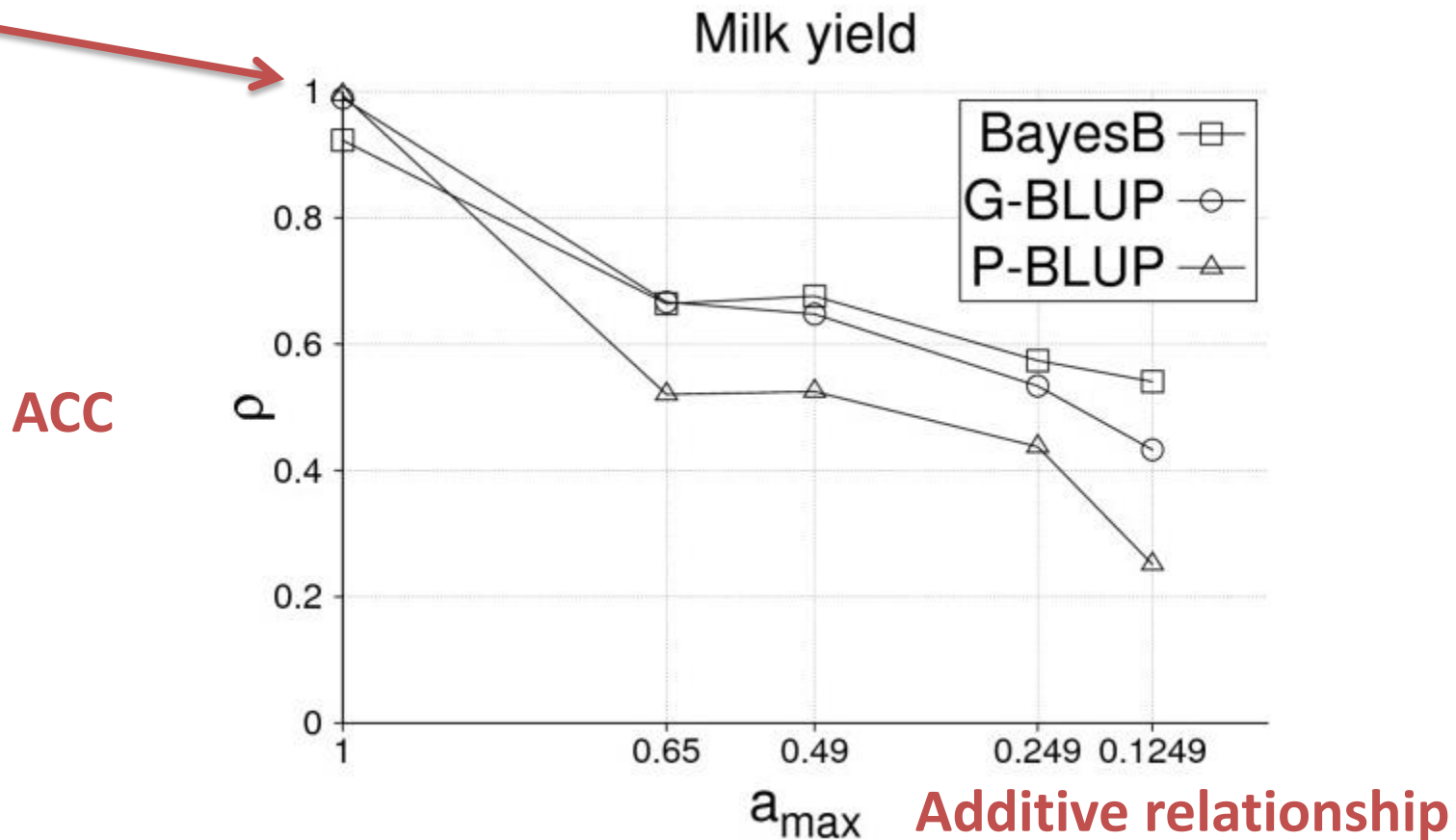
∴ Training population



~ 9000 animals for feed efficiency and 4000 for meat quality.

Markers can predict family relationships between animals, independently of LD between markers and QTL

Using the same animal for training and validation



∴ Approximate number of progeny need to reach accuracy levels for three heritabilities (h^2)

BIF	<u>Heritability Levels</u>		
	h^2 (0.1)	h^2 (0.3)	h^2 (0.5)
0.01	1	1	1
0.02	2	1	1
0.05	4	2	1
0.08	8	3	2
0.13	13	5	3
0.2	22	7	4
0.29	38	12	7
0.4	70	22	13
0.56	167	53	30
0.99	3800	1225	700

Genomic Selection using related animals

Dairy cattle

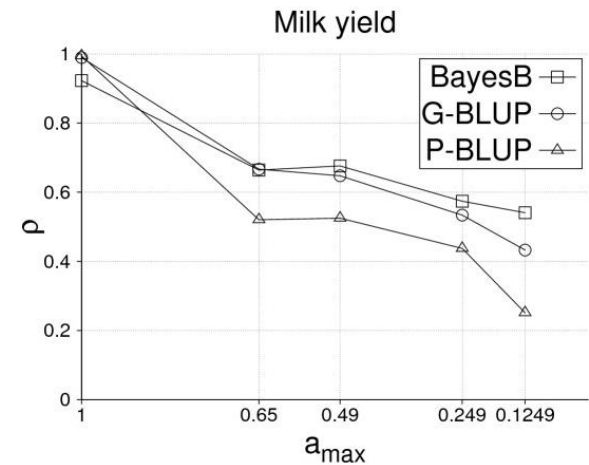


+



= 200,000 animals genotyped

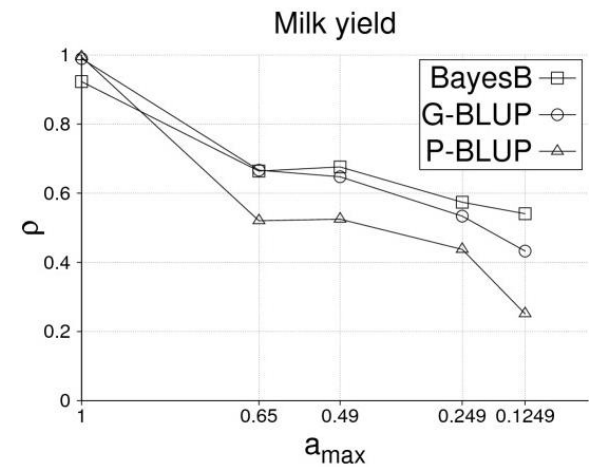
- Are all animals informative?
- Can we use only animals with higher relationship?



Genomic Selection using related animals

Beef Cattle

- Should we run gEPDs for animals with different breed composition together?



Genomic Selection using related animals

Beef Cattle

- Group animals based on their similarity without checking phenotype aspects

AABBABABBBBBBABBA
ABBAAAABBBBBBBBAA

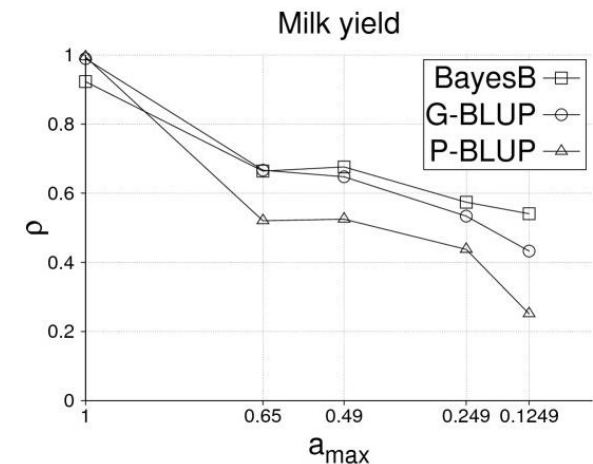
Animal 1

ABABBBABBBBBBABBA
ABAABAABBBBBBBBAA

Animal 2

AABBABABBBBBBABBA
ABBABAABBBBBBBBAA

Animal 3

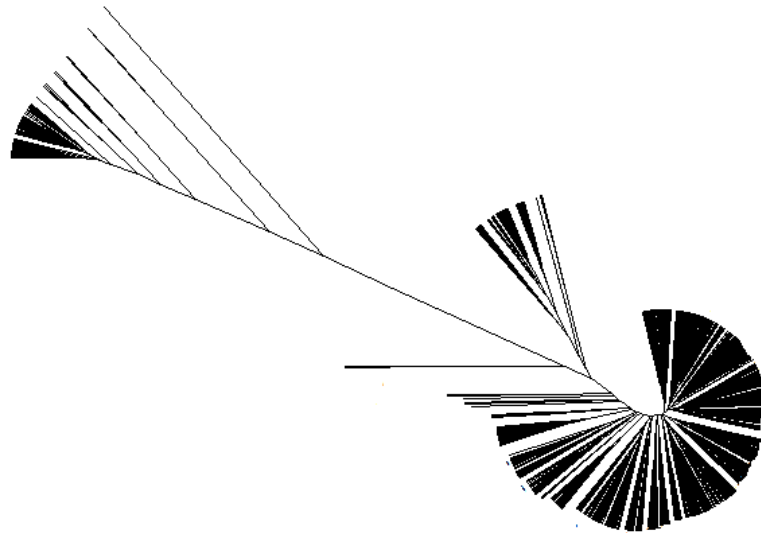


Genomic Selection using related animals

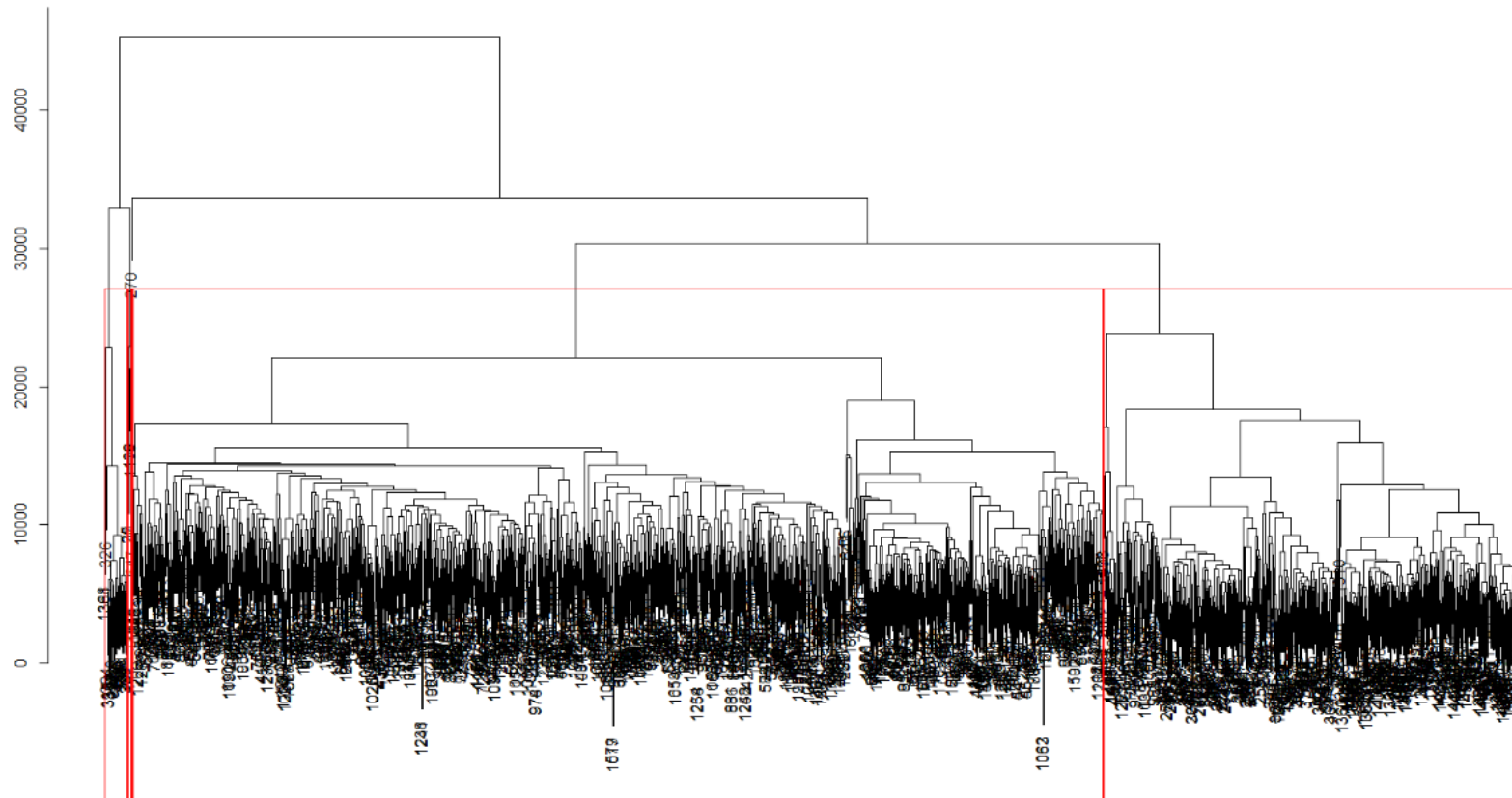
Holstein and Ayrshire

Comparison of genotypes takes forever

Computing **distance** between animals mimicking a pedigree



Genomic Selection using related animals



Dynamic Genomic Selection in Crossbred Beef Cattle Populations

R. Ventura^{1,2}, S. Larmer¹, F.S. Schenkel¹, S. Miller¹, P. Sullivan³.

¹Centre for Genetic Improvement of Livestock, University of Guelph, Ontario, Canada, ²Beef Improvement Opportunities, Guelph, Ontario, Canada, ³Canadian Dairy Network, Guelph, Ontario, Canada

Table 1. Estimates of increases in accuracy by using EDM Clustering Method compared with replacing 1/5 of the training population with animals from different clusters.

Trait	X1(212)	X2(212)	X3(212)	X4(213)	X5(211)
Scenario1	T	T	T	T	0.0734
Scenario2	T	T	T	0.0645	T
Scenario3	T	T	0.0321	T	T
Scenario4	T	0.0497	T	T	T
Scenario5	0.0654	T	T	T	T

BIO Projects



The Hard traits: Benefits from Genomics

Measured late in life

Available after slaughter

BIO Projects : RFI and Tenderness : Novel phenotypes

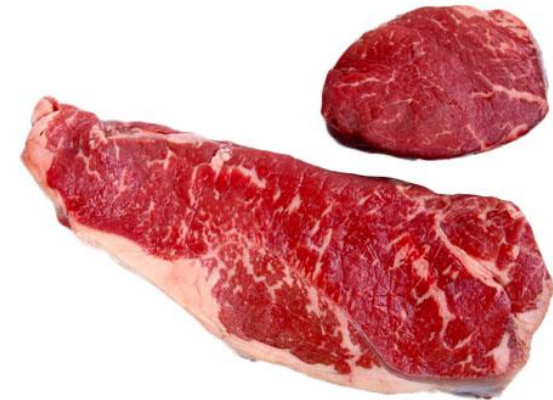
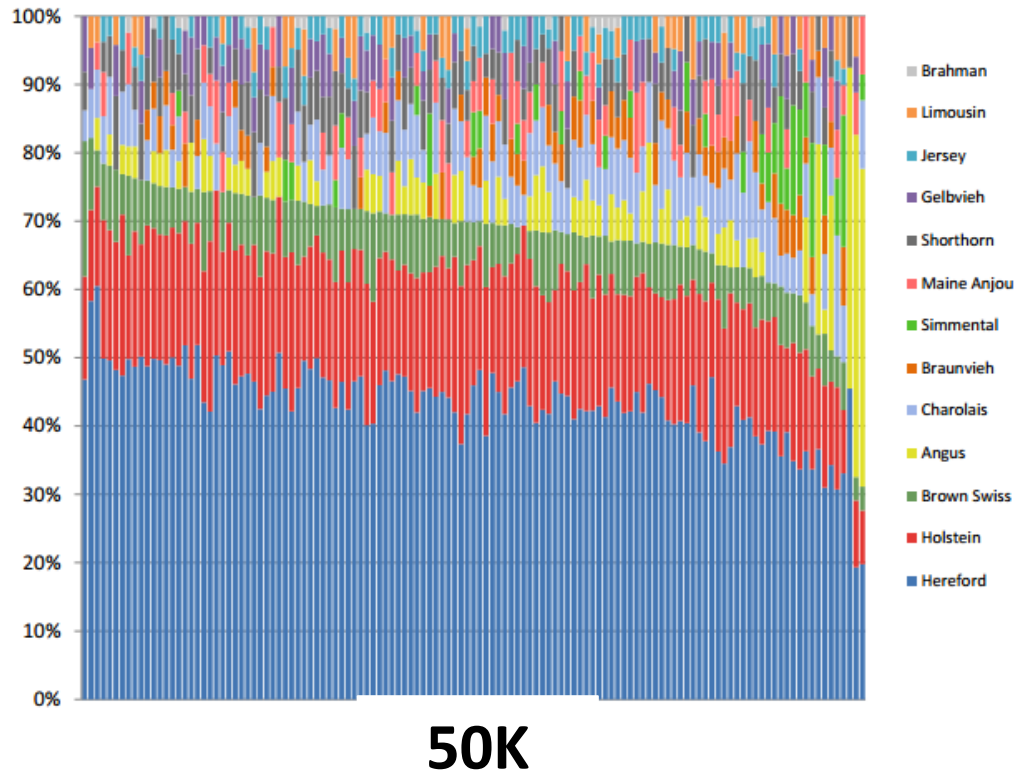


Dr. Montanholi

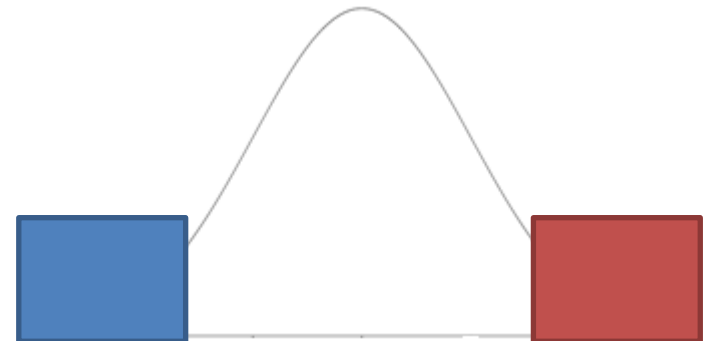


Infrared images, Fecal cortisol and several other novel phenotypes

Determining Breed composition using SNP panels



1043 samples



Conclusions

6K:50k Imputation can be successfully applied on crossbred and purebred animals

Within breed imputation of purebred animals performed slightly better

Closer relatives in the reference set: Highest accuracy

Crossbred animals: Increase the reference set

Increased accuracy (>0.057) was obtained when the training population was defined based on the clustering methodology. These preliminary results warrant further investigation.

.: Acknowledgments



Dr. Steve Miller

Dr. Flavio Schenkel

Dr. Mehdi Sargolzaei

Dr. Gordon VanderVoort

Mike McMorris

Brittney Livingston

