A new Marker-Assisted BLUP genomic evaluation for French dairy breeds

Pascal Croiseau, Aurélia Baur, David Jonas, Chris Hoze, Julie Promp, Didier Boichard, Sébastien Fritz, Vincent Ducrocq*

Genomic evaluation in France : 2009-2014

- Holstein, Normande, Montbéliarde then Brown Swiss
- A Marker-Assisted BLUP model with pre-detected QTL (via an LDLA approach) traced using haplotypes of 3 to 5 SNP
- 300 to 700 QTL per trait
- a residual polygenic effect (explaining 30-50% of total genetic variance) using pedigree information
- imputation LD ➔ 50k, with DagPhase (*Druet & George, 2010*)
- phases needed, created using DagPhase
- Reference populations consisting of bulls only

Previous model: limits

- Imputation could be improved:
  DagPhase seemed less efficient than initially
- Our research projects showed that 500 QTL did not give maximal accuracy
- Large residual polygenic variance penalized sons / daughters of bulls with no progeny test information
- Time consuming computations (old software)

Imputation

- Software comparison
- FImpute (*Sargolzaei et al*, 2011) is as accurate as Beagle in populations with a dense pedigree information
- Accuracy increased from 98.0 % to 99.2%
- Fimpute is at least 3 times faster
- Can impute the whole population in one run

=> **Switch to Fimpute** (with commercial licence)

A new genomic model

QTL size | **Genomic evaluation**

Large

Moderate

Small

traced with markers

➔ haplotype effects $\hat{h}_j$

➔ DGV = $\sum_{j=1,\dots,K}^{+} \hat{h}_j$

Tiny

Consider their sum only:

$\hat{u}=\sum_{j'}\hat{m}_{j'}\sim$ ~~N(0, ***pedigree*** relationship matrix)~~

$\sim N(0, \textbf{\textit{genomic}}\ \text{relationship matrix})$

In practice…

$$g_i = \boxed{\sum_{j=1}^{n}(h_{ij1} + h_{ij2})} \quad + \quad u_i$$

Trait dependent

Trait independent

$$g_i = \boxed{\sum_{j=1}^{n\ +}(h_{ij1} + h_{ij2})} + \sum_{j=1}^{k}(SNP_{ij1} + SNP_{ij2})$$

- Genomic relationships via EuroG10K chip: System size = constant
- New software

  to cope with very large increase in number of genotyped animals

  Strategy: read genotypes and store in memory

  preconditioned conjugate gradient with iteration on data in memory

QTL modelisation

- First step: QTL (SNP) detection with ~~GWAS~~ Bayes Cπ,

  K (=1000 | 3000 | 6000) largest SNP selected

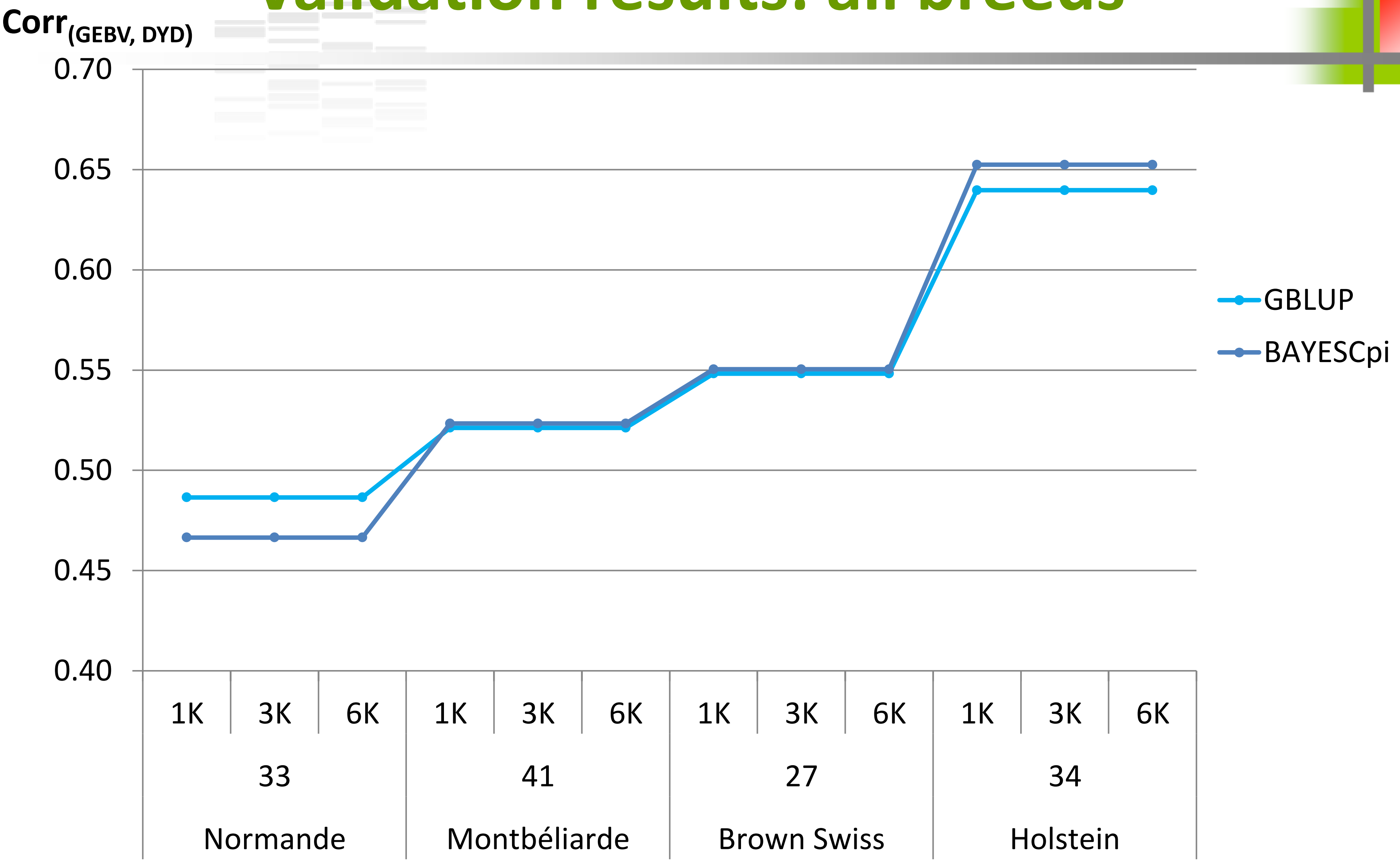
- To modelize each QTL, we use:
  - ✓SNP QTL information
  - ✓Haplotype including the SNP and flanking markers
  - ✓Haplotype including the SNP and the « best » combination of SNP in a window of 10 SNP (David Jonas et al., this congress)
    - The optimal combination is the one with the highest number of alleles and the most homogeneous distribution of these alleles in the population.

Reference populations

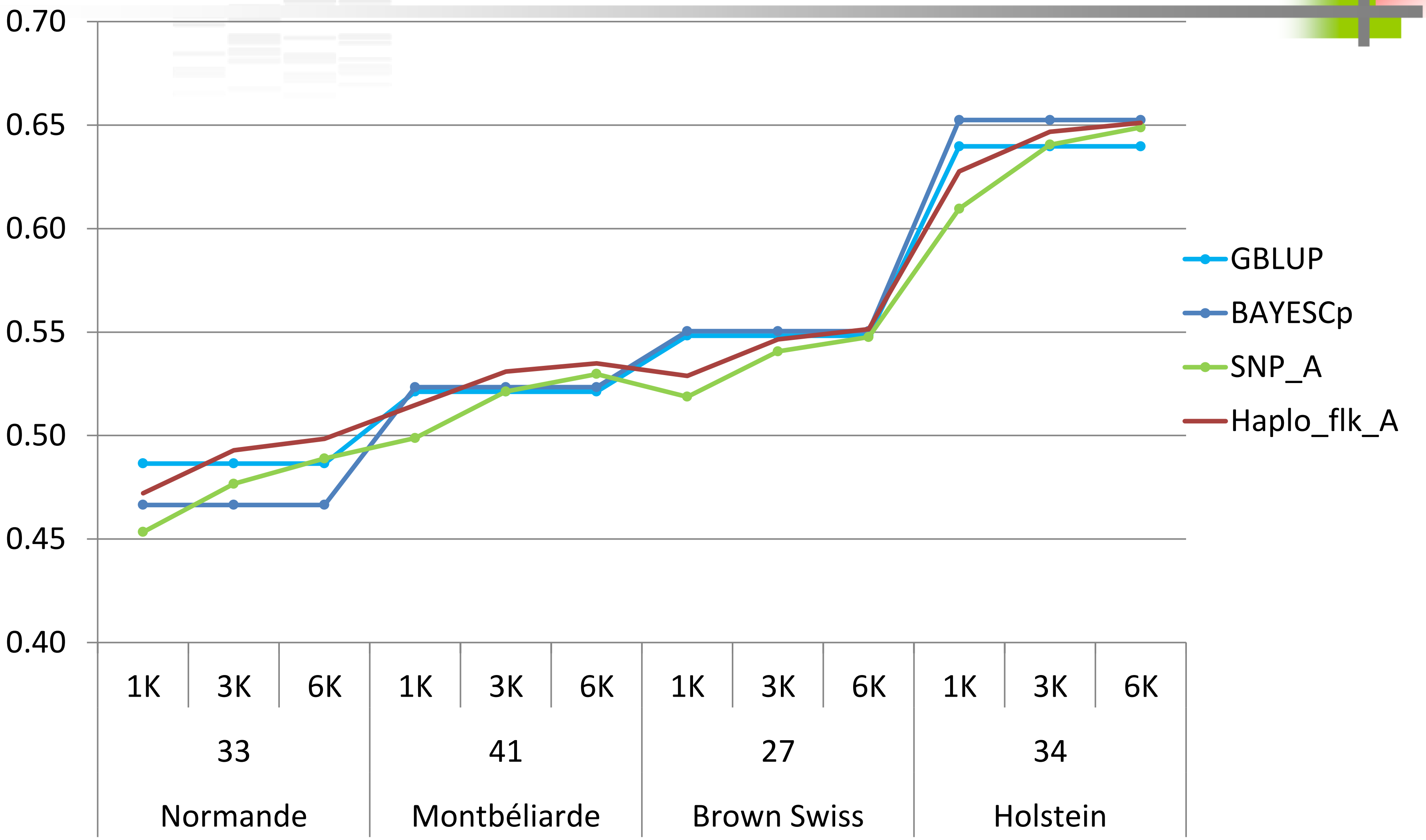
	Reference population	
	Learning	Validation
Normande	1945	385
Montbéliarde	2115	535
Brown Swiss	5557	458
Holstein	27309	3391

Comparison criteria : correlation ($GEBV_{t-4}$, DYD_t), regression slope

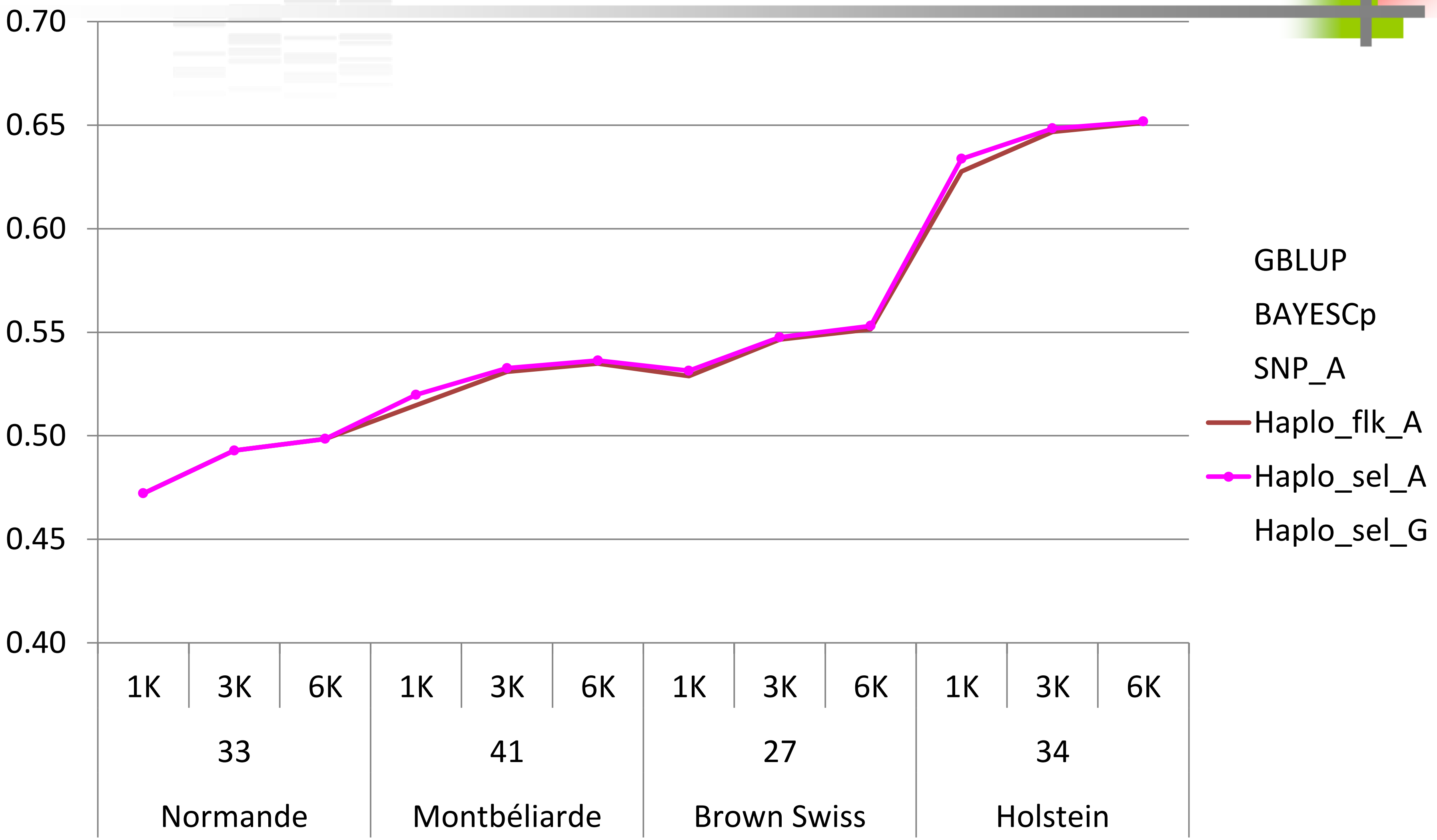
Validation results: all breeds



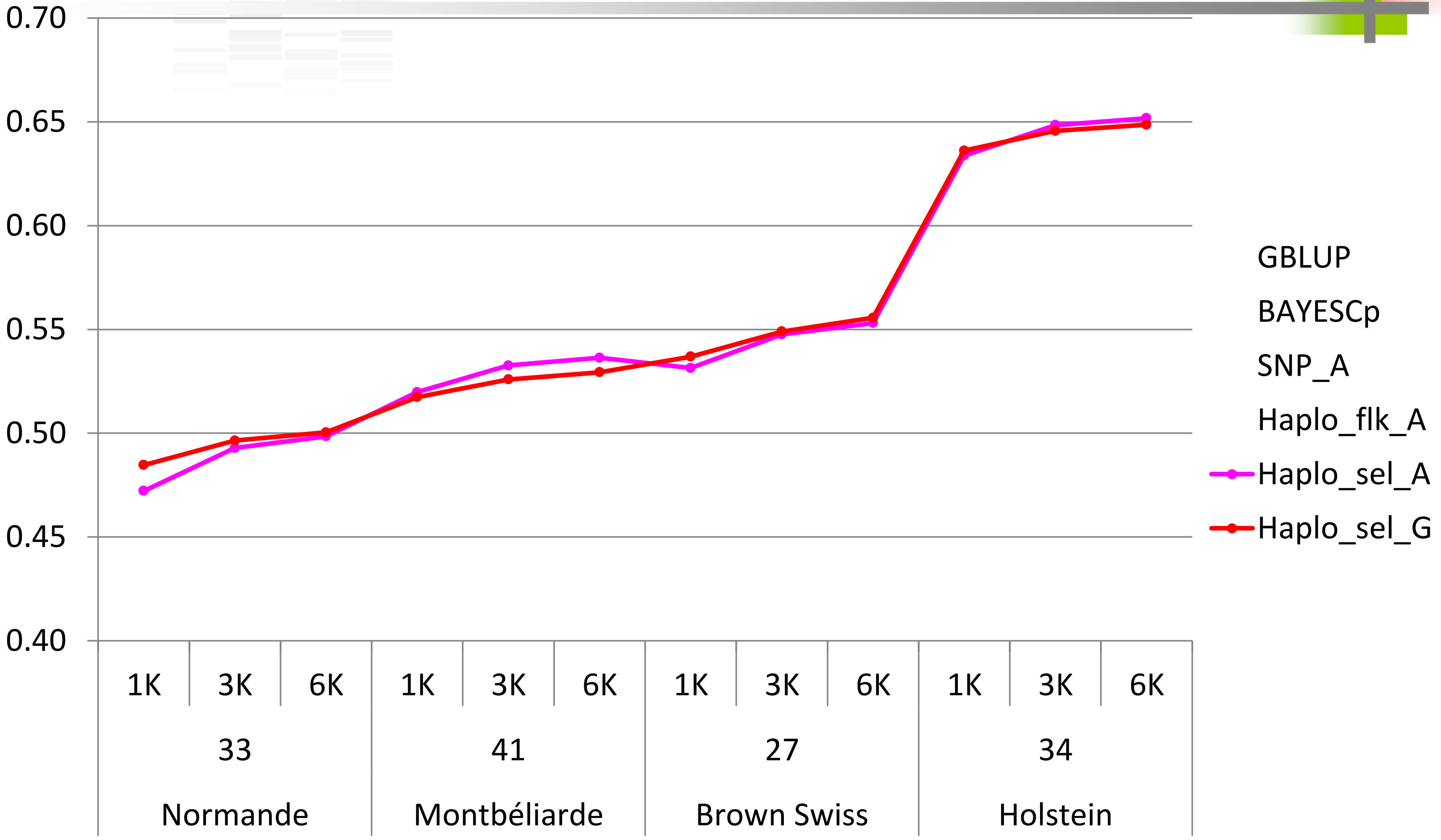
Validation results: all breeds



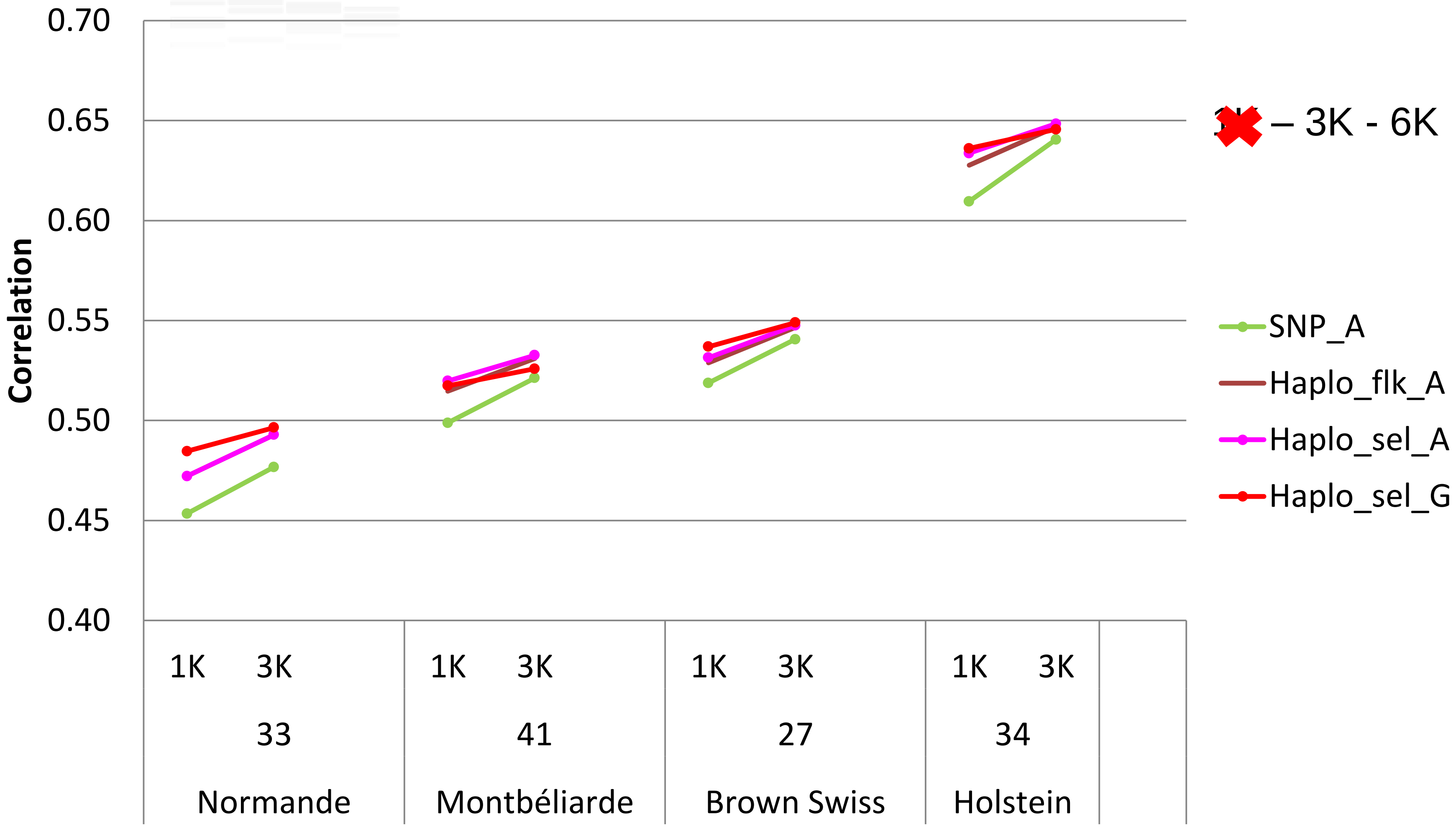
Validation results: all breeds



Validation results: all breeds

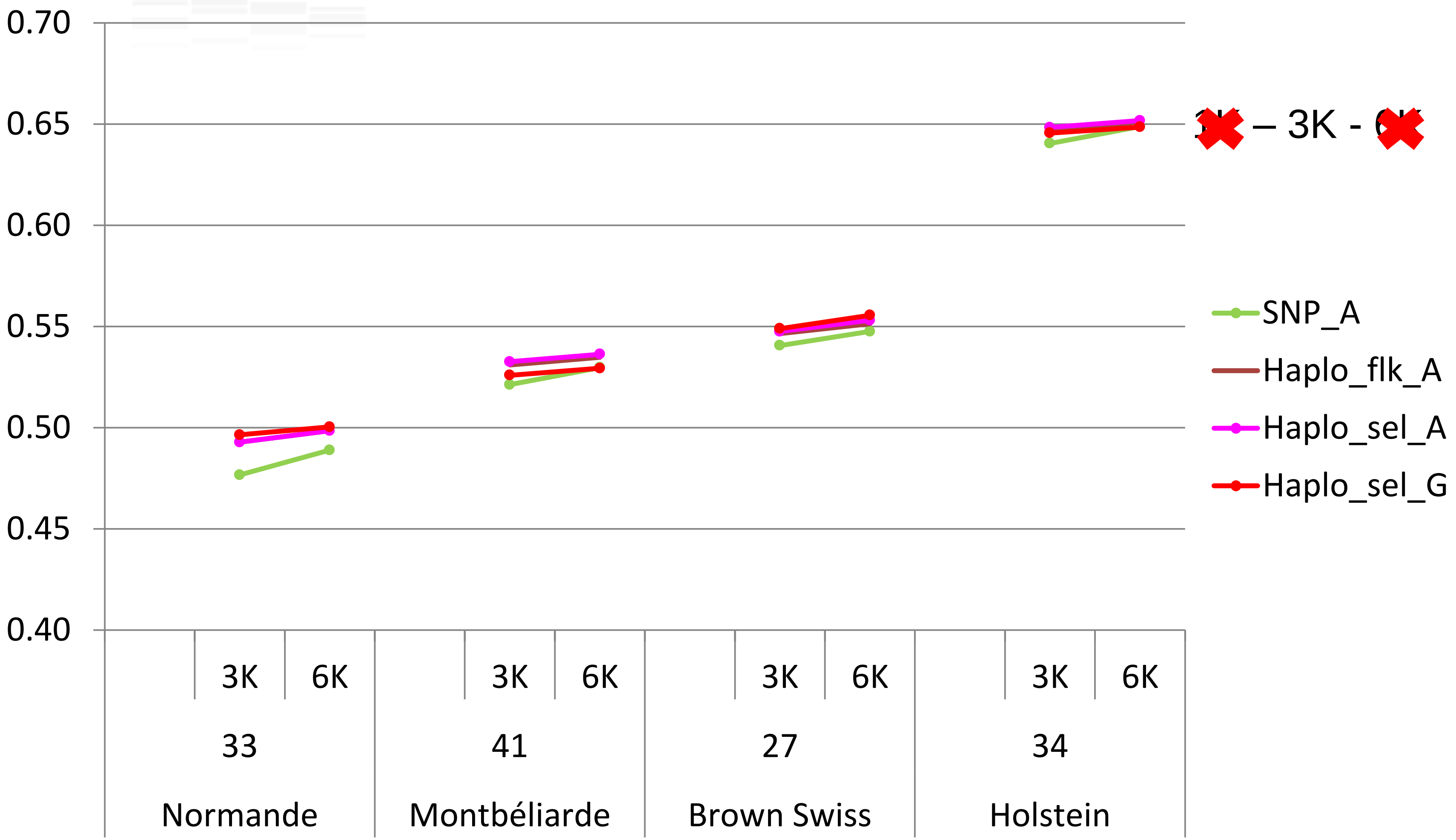


Validation results: all breeds



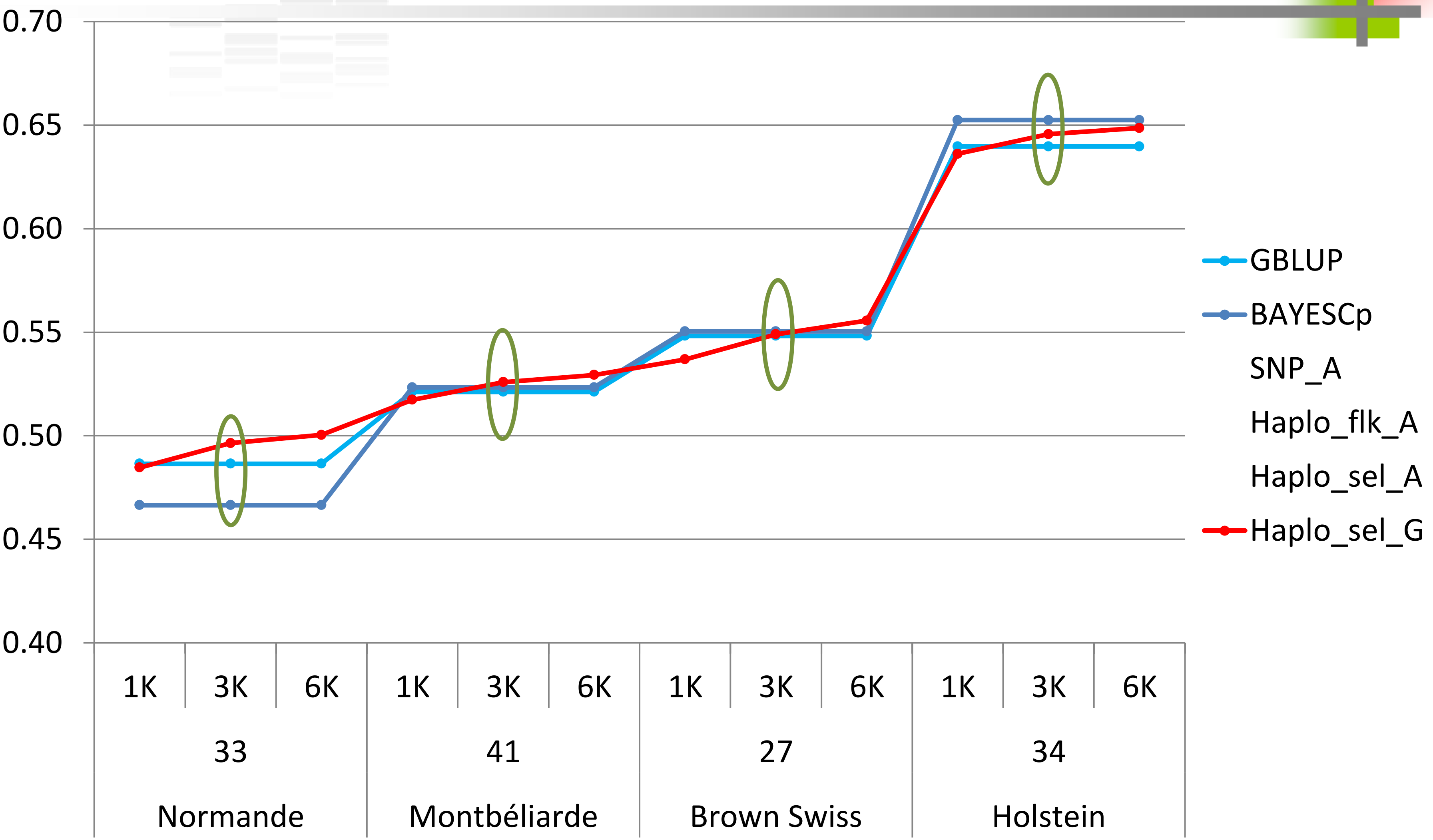
~~1K~~ – 3K - 6K

Validation results: all breeds



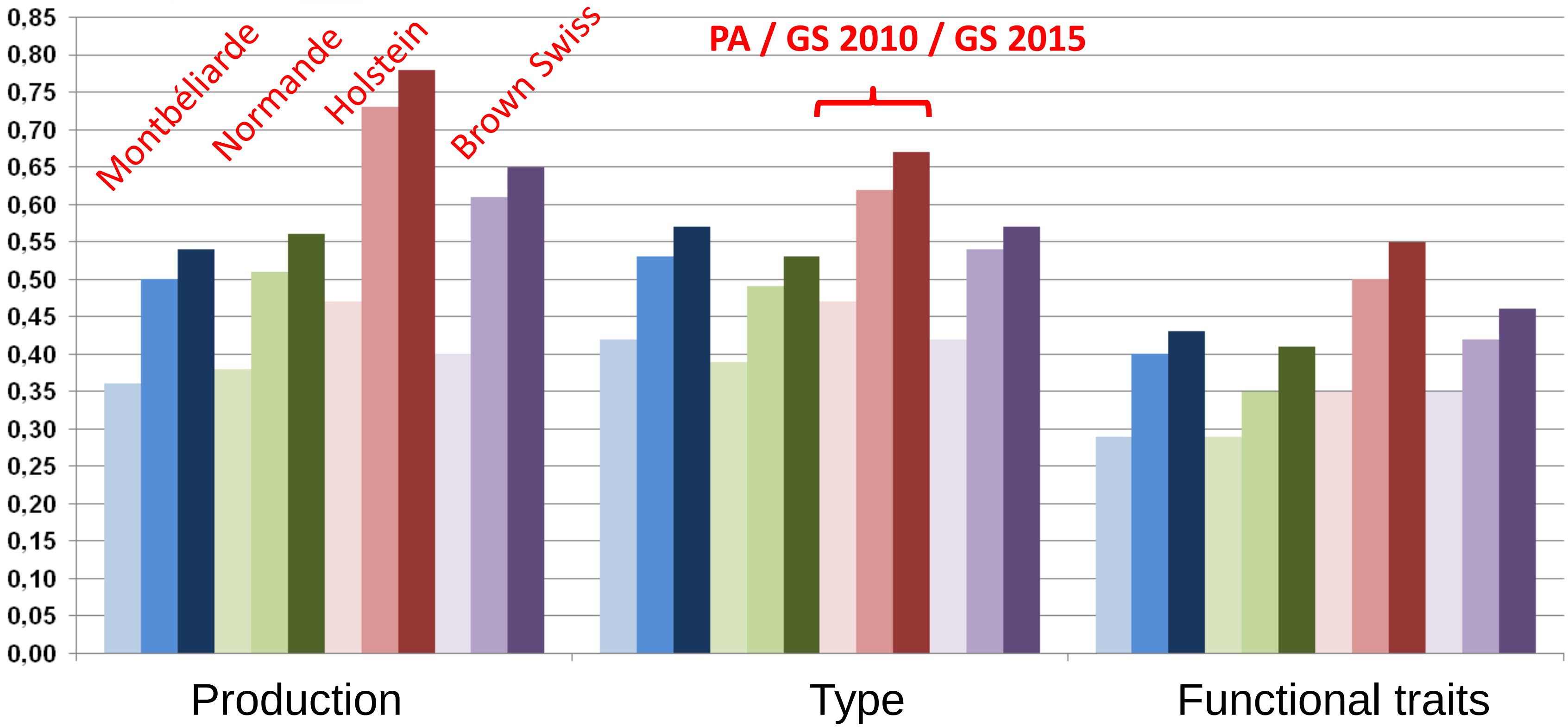
~~1K~~ – 3K - ~~6K~~

Validation results: all breeds



New evaluation: overall impact

Correlations PA/GEBV at year t-4 and DYD at year t



- Increase in reliability:

Normande: +0.11 Montbéliarde : +0.10
Holstein : +0.08 Brown Swiss : +0.11

A new genomic evaluation in Spring 2015

- Substantial gain in reliability
- New software
  - much faster
  - easy to extend to include causal mutations
- Use of genomic relationship matrix leads to
  - system size = constant
  - much less sensitive to missing pedigree information, to absence of phenotypes of sires, to absence of foreign information
- Post-processing to force regression slope for candidates to be ~1
- also being developed for other French minor breeds

A new Marker-Assisted BLUP genomic evaluation for French dairy breeds

Pascal Croiseau, Aurélia Baur, David Jonas, Chris Hoze, Julie Promp, Didier Boichard, Sébastien Fritz, Vincent Ducrocq*