

What is Single Step (and update)

Michael Lee

**Department of Mathematics & Statistics
University of Otago**



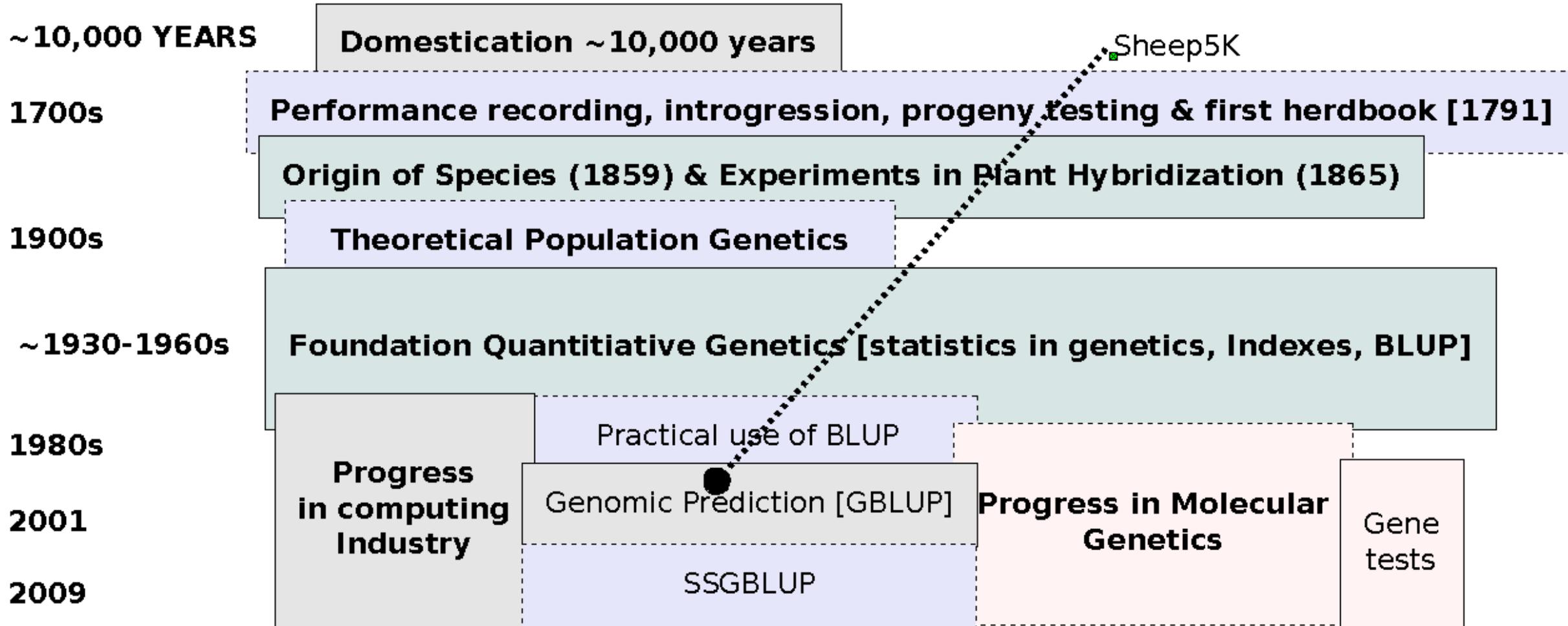
UNIVERSITY
of
OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

What the talk will cover.

- Historic perspective
- What is Single Step
- Status



Perspective



Perspective

“Using commercially available hardware and recompiling the programs would allow this to be increased eight-fold, i.e. to approximately 1.6 million animals. Thus the system should be able to handle all requirements in the foreseeable future” (S.A. NEWMAN, K.G. DODDS, J.N. CLARKE, D.J. GARRICK AND J.C. MCEWAN, 2000)

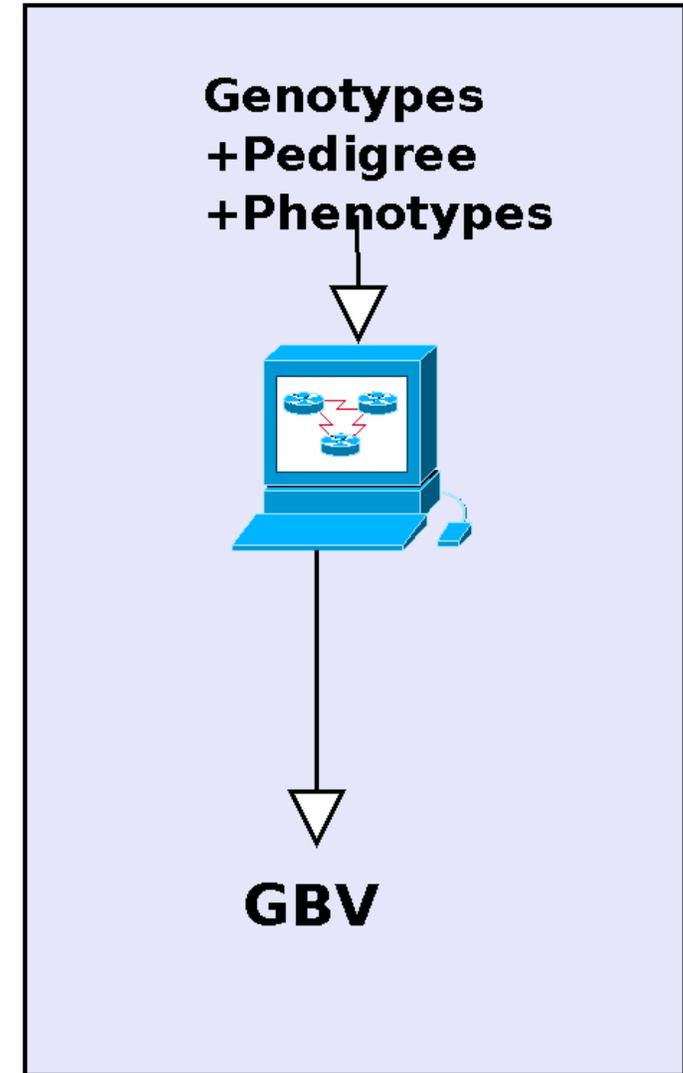
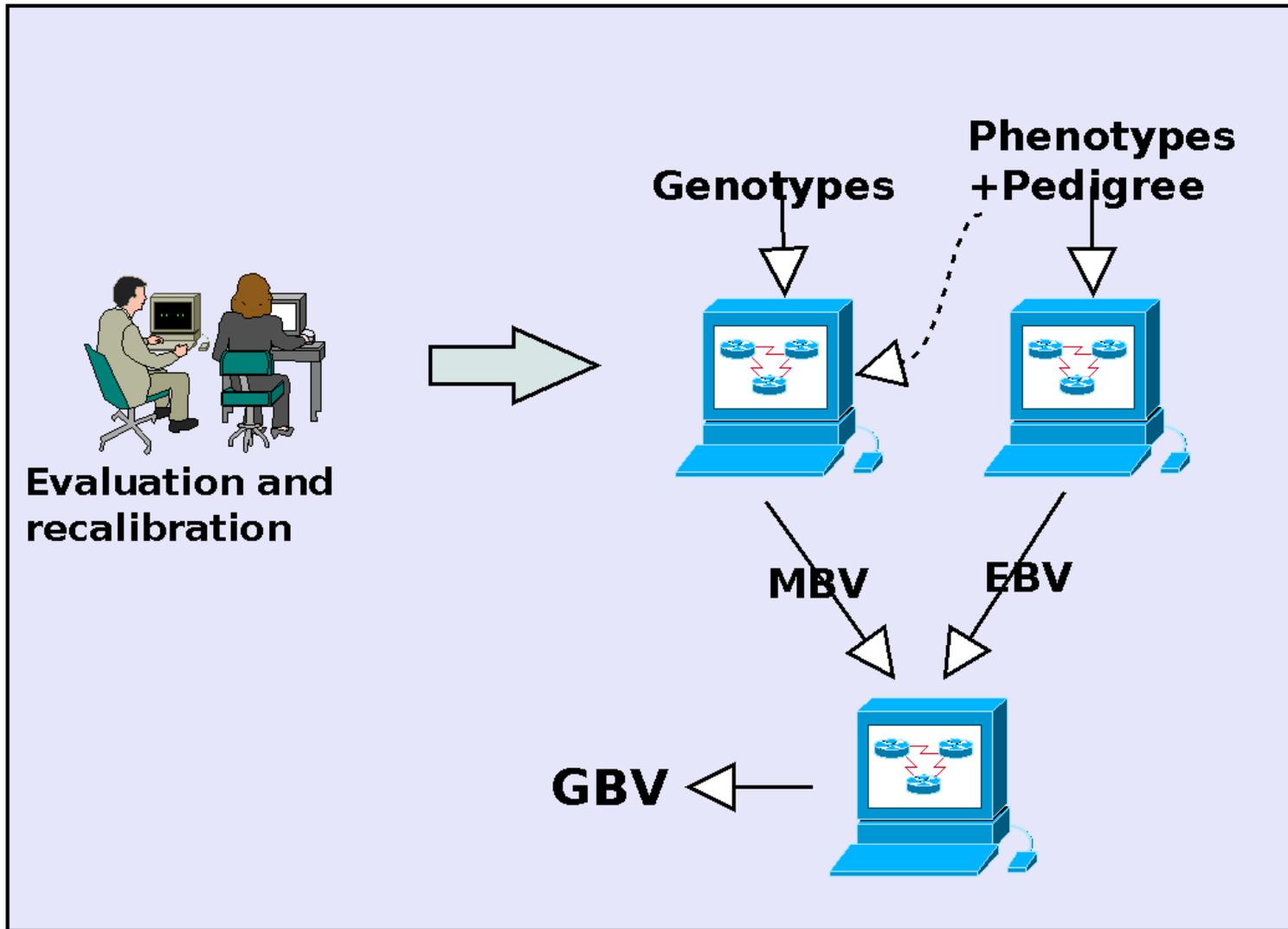


-Many million of animals; massive increase in DNA data [genotypes & sequence]; data from automated electronic data collection including from commercial animals [e.g. processors]; more complex breeding objectives



MULTI-STEP EVALUATION

SINGLE STEP



What is Single Step GBLUP

- Current system [multi-step] uses three analyses and combines three Breeding values to give a GBV [cumbersome and sub-optimal]
- Multi-step good if we don't have all the data (e.g. someone sells us MBVs and we want to combine them with our EBVs)
- Single step uses all data in one evaluation to give a breeding and uses the same maths [BLUP] as devised by Henderson in the 1950s.
- Unlike our current multi-step evaluation information is spread to non-genotyped animals.



Status: overview

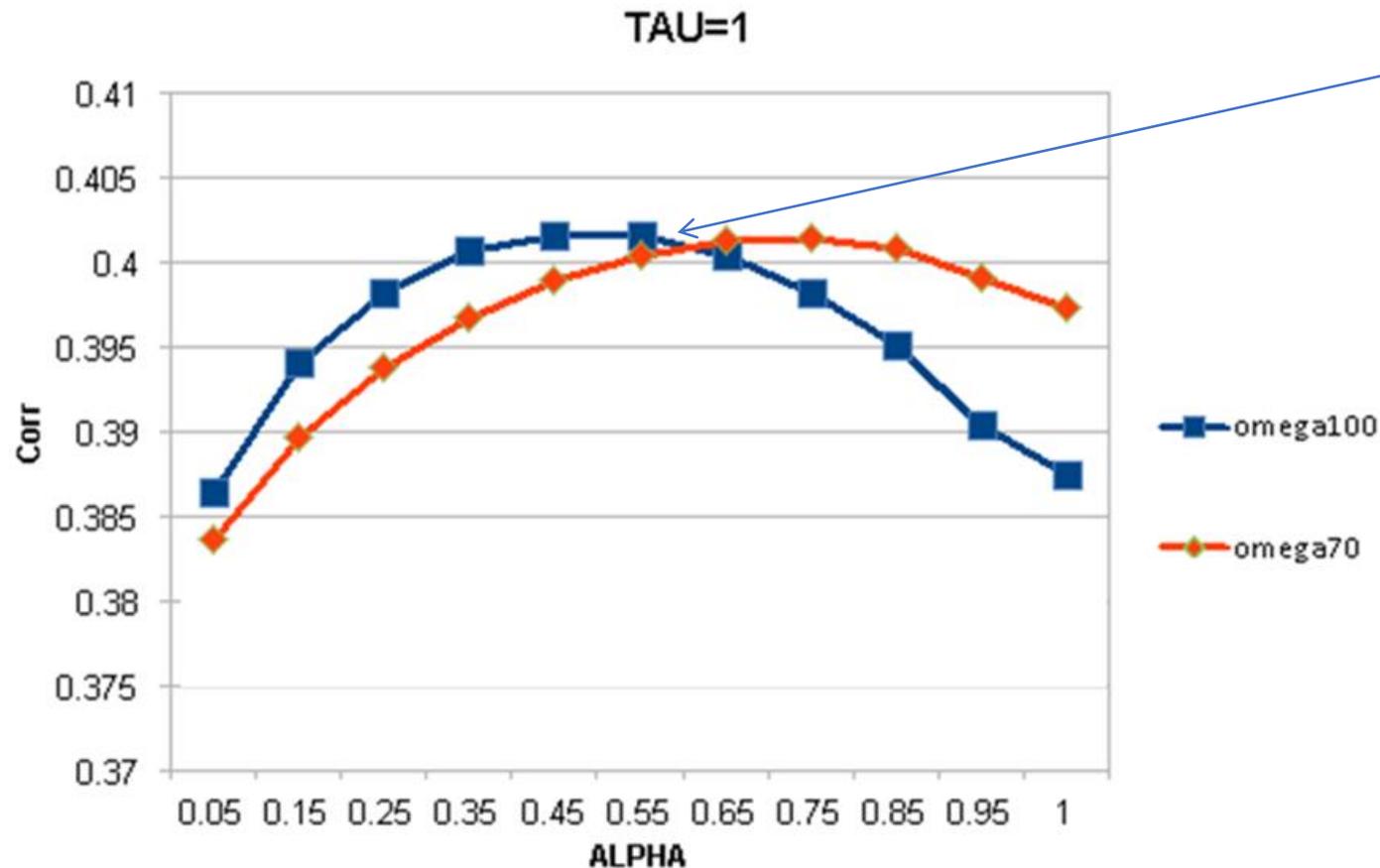
- Superiority in prediction can be generally shown for ssGBLUP over just normal evaluations with just pedigree & limited phenotype data
- As with multi-step evaluation (Sheep5K) advantage is for traits hard to measure, sex limited, expressed late in life and lowly heritable
- Advantage of single step is resolves some of the issues with multi-step evaluations [bias in statistics speech]



**Validation: accuracies probably
comparable with what is observed
in Sheep5^K**



Example from validation of Facial Eczema



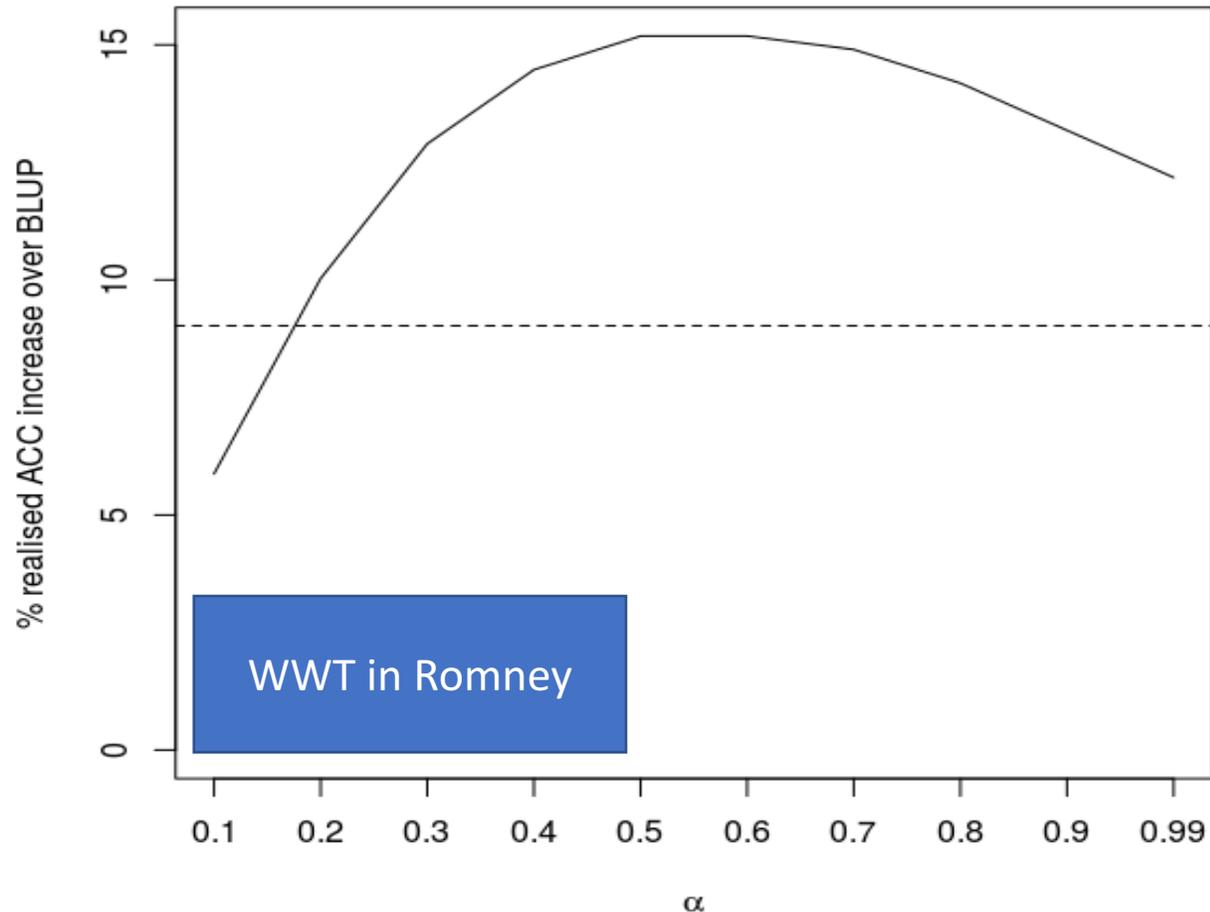
Expected accuracy of prediction without phenotype

- Accuracies are probably comparable with Sheep5K for genotyped animals

Correlation between full dataset Y-Xbeta and reduced ssGBV



Accuracy increase over No genotypes



What is really important is accuracy increase over Parent average [i.e. No genomics]

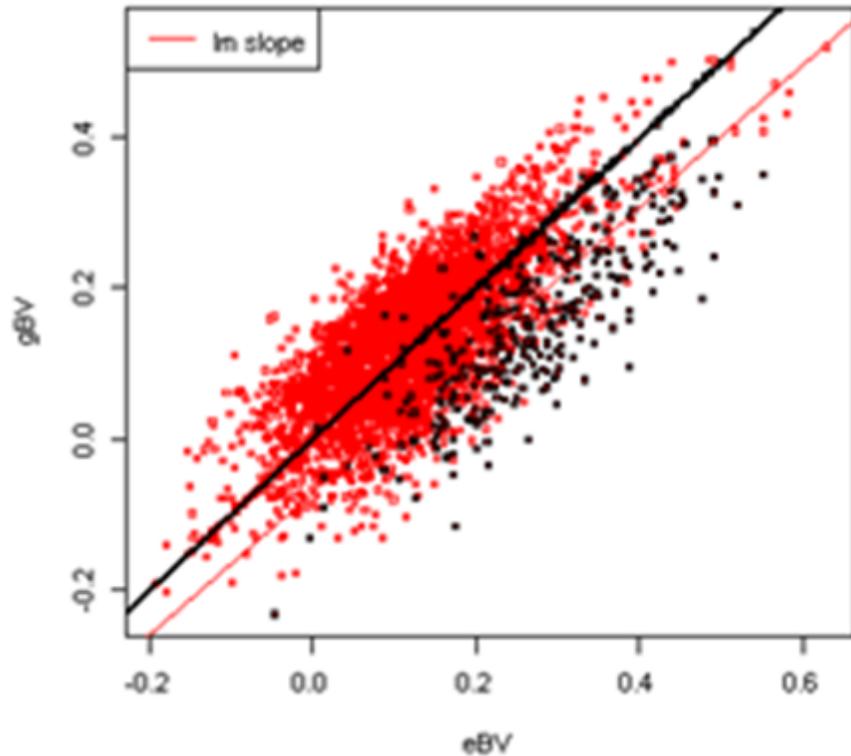
- WWT is difficult to do better than over just using Pedigree & phenotypes – but there is a slight increase over parentage average [$\sim 15\%$].
- Will be valuable [**still**] for traits typically: lowly heritable, hard to measure, sex limited & expressed late in life.



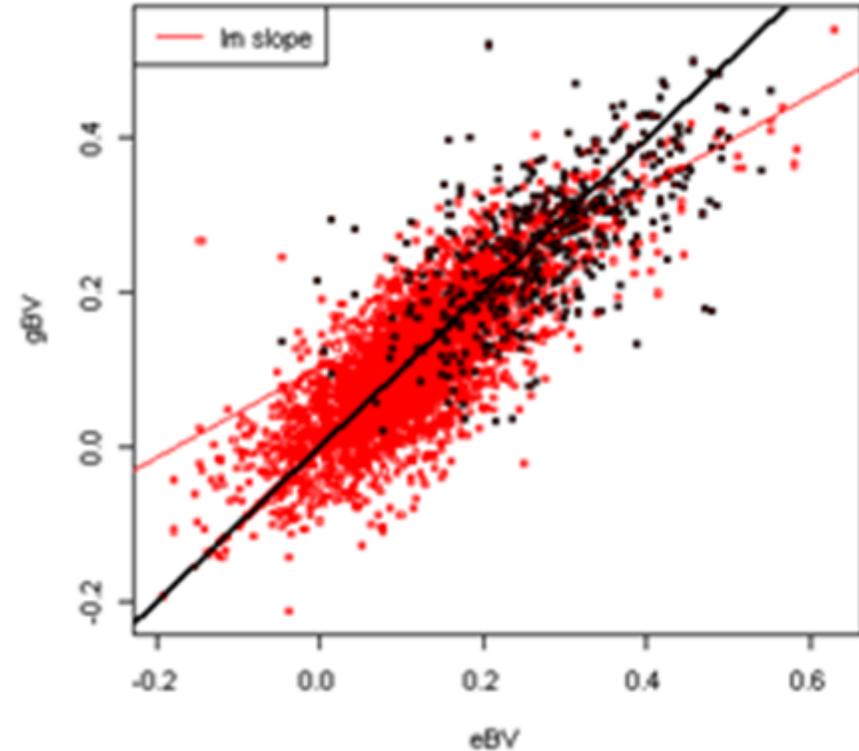
**Bias: Don't see problems that we
have needed to resolve in multi-step
[Sheep5K]**



Example of bias

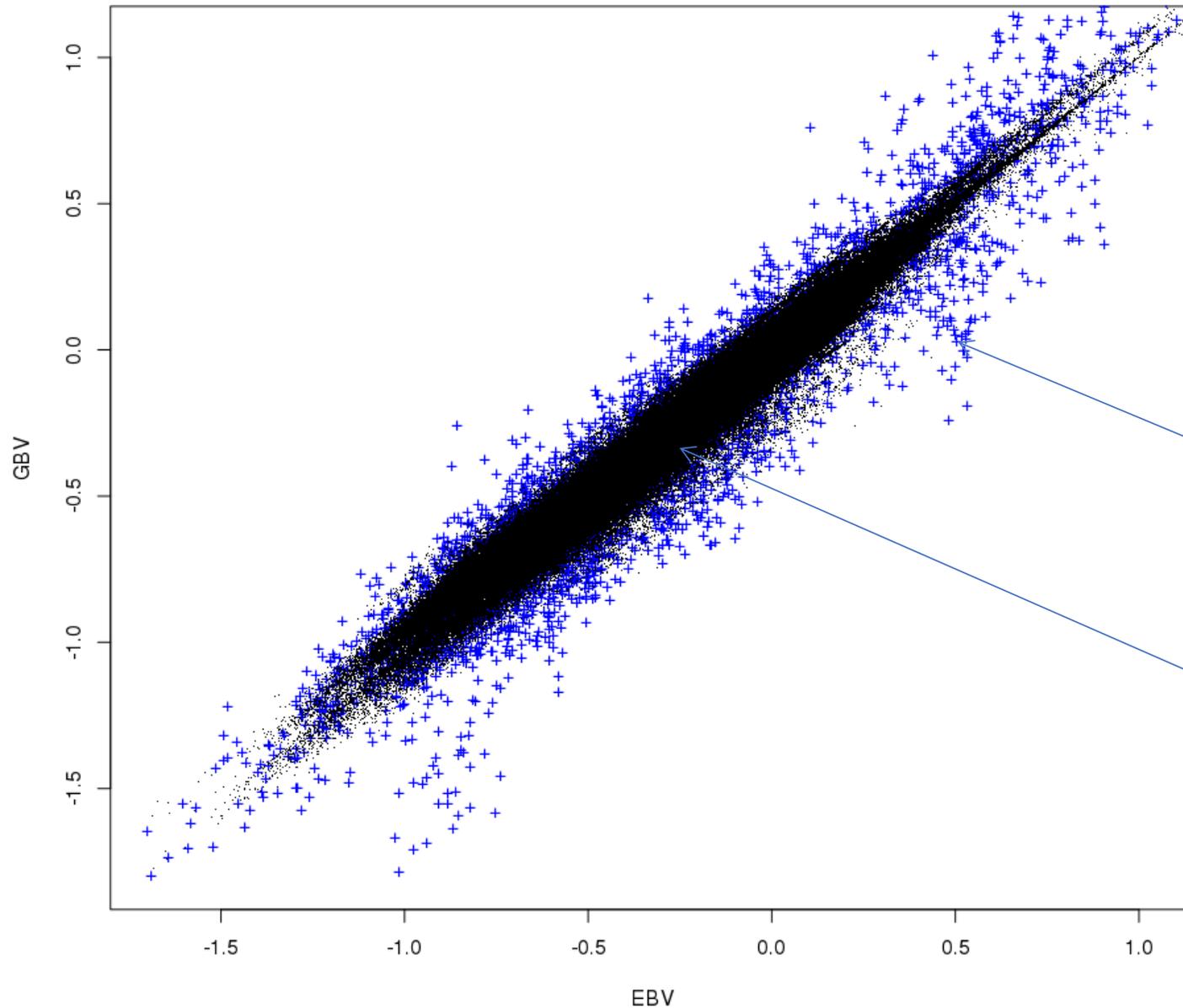


Remove
Principal
components



NLB in multi-step [Sheep5K] – issue resolved years ago,
but other issues being resolved





Graph of NZGE run with ssGBLUP

~9millions animals,
only ~5,000 genotyped animals

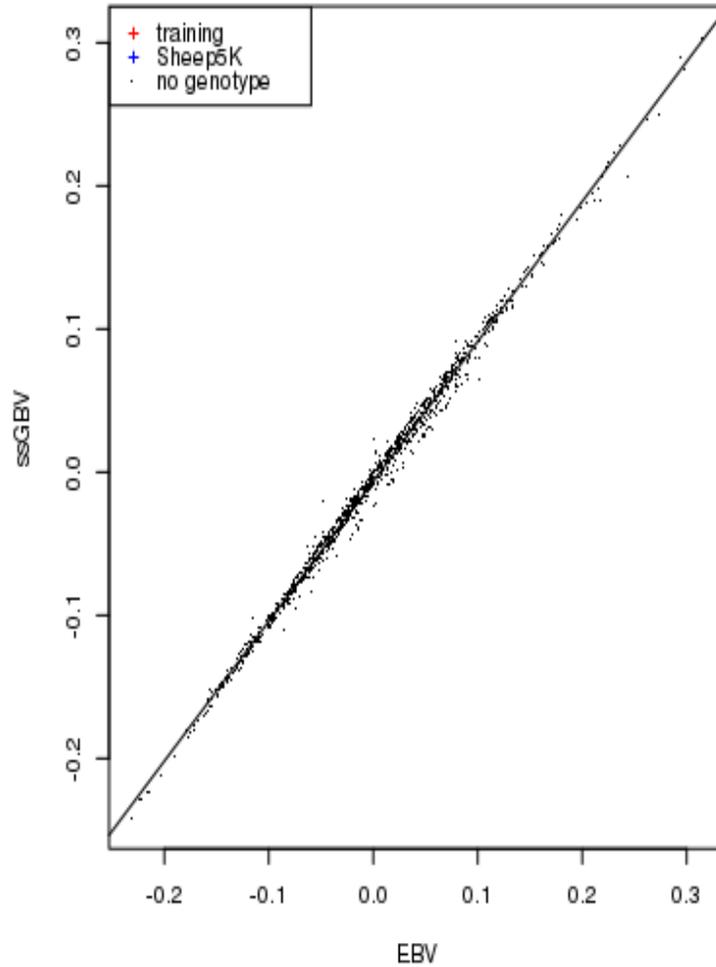
Blue is genotyped animals

Black is no genotypes
[c.f. multi-step where
this would be a
straight line

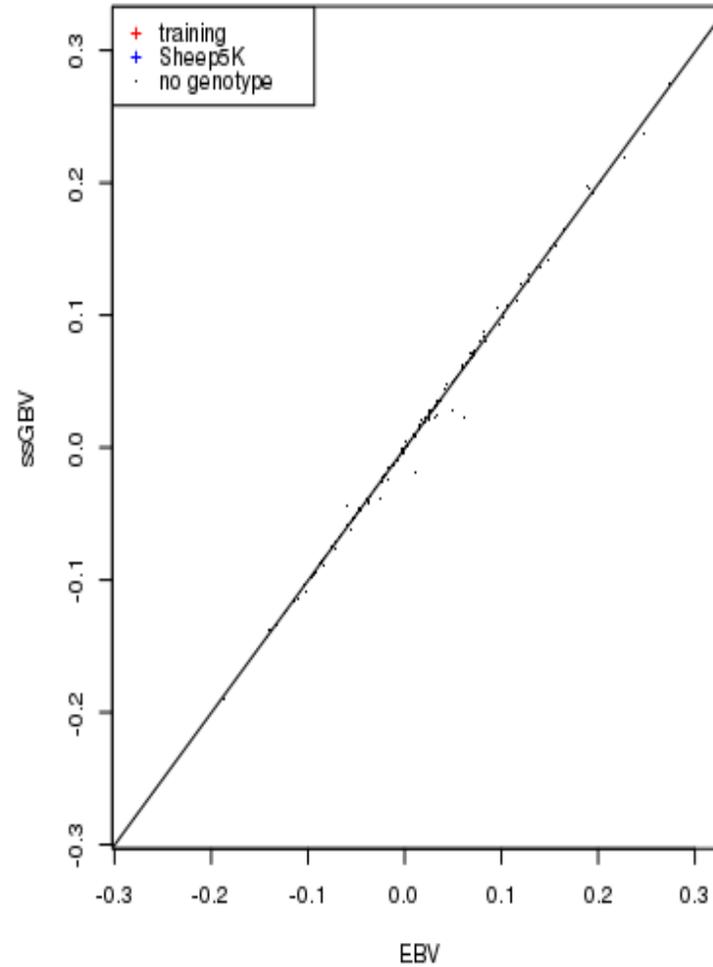


Generally flocks without any genotype data will not change much

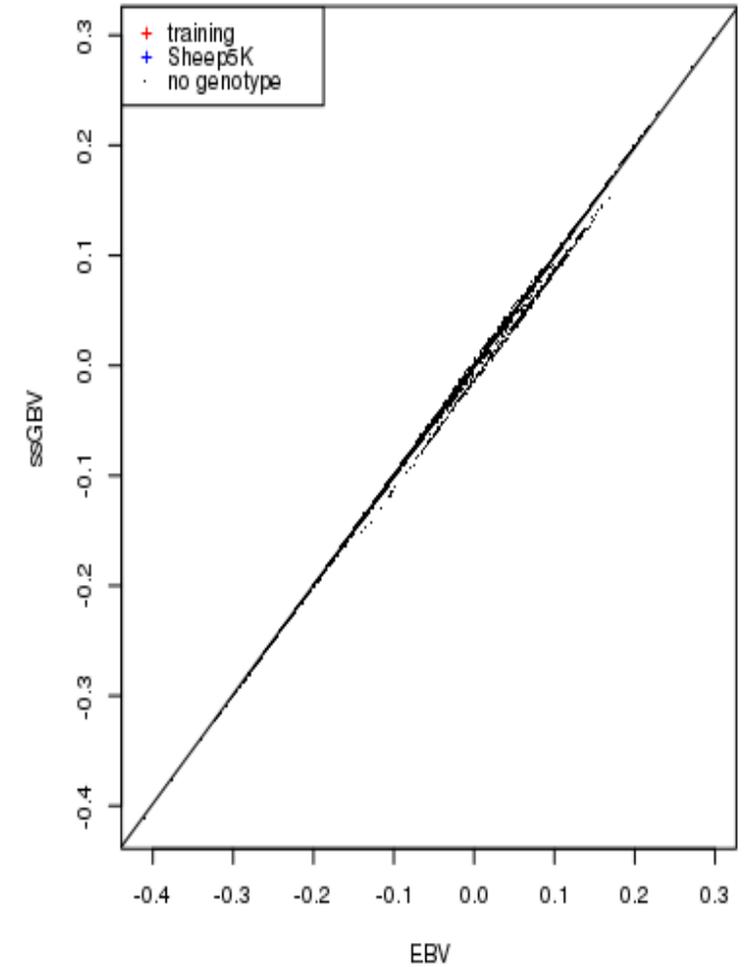
nib: flock with no genotypes



nib: flock with no genotypes



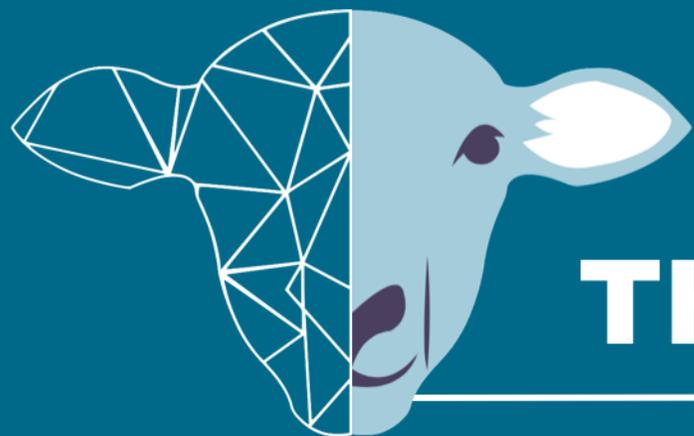
nib: flock with no genotypes



Status/progress of implementation

- Software adapted for the NZ sheep industry from software used extensively in the Dairy Industry
- Implementation via HPC using cloud computing services via the Amazon Web Services [AWS]
- Genotype Database [Rezare]
- Gearing up to demonstrate and validate the system using industry data in conjunction with enhanced parentage chip developed by AgResearch
- Run in parallel with current NZGE and work to go live 2018!!





Thank you.

Acknowledgements

Team at Mathematics & Statistics [University of Otago] & AGR Invermay – Numerical and compute;

API-Talent: AWS cloud architecture;

University Georgia (US) and Natural Resources Institute Finland – Software & support

BLG Team – Management & Mana