

ssGBLUP in practice

Daniela Lourenco

Ignacy Misztal, Shogo Tsuruta, Ignacio Aguilar, Breno Fragomeni, Yutaka Masuda, Ivan Pocrnic,
Heather Bradford, Andres Legarra

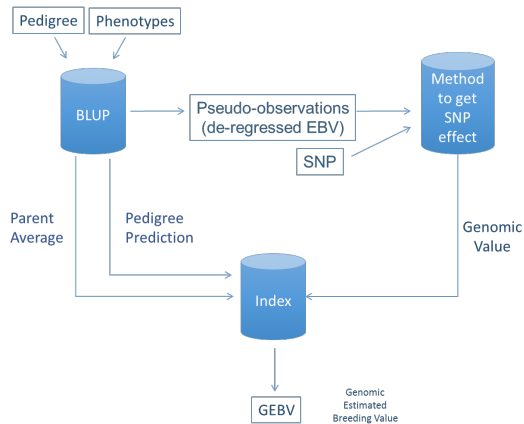
University of Georgia
June 2016

Recent questions - GS

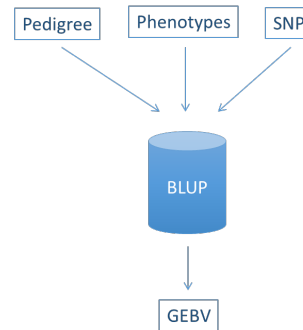
- Why to use ssGBLUP?
- Will ssGBLUP work when we have zillions of genotyped animals?

Why to use Single-step?

• Multistep



• Single-step

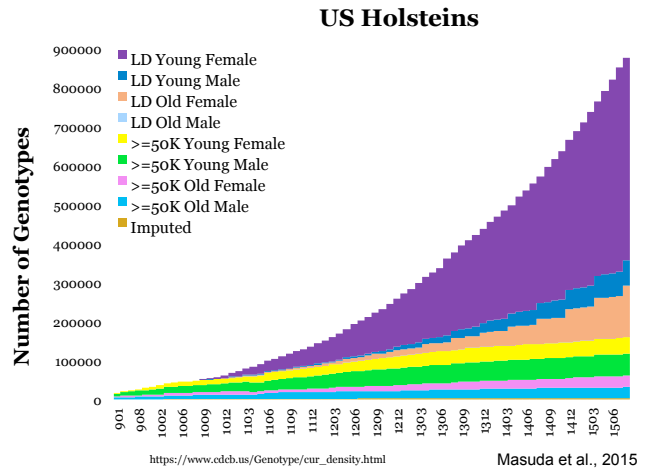
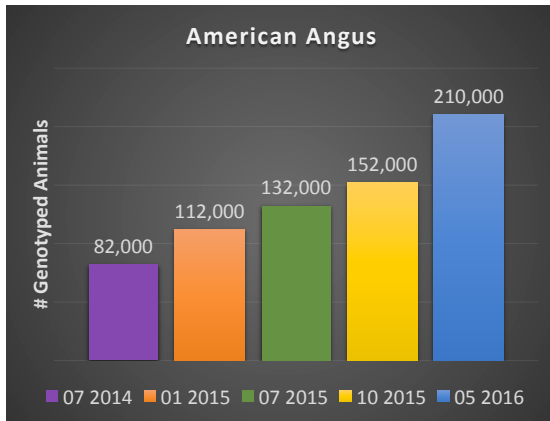


Why to use Single-step?

TAKE HOME #1

- Simplicity
- Same models
- Phenotypes, pedigree, genotypes

ssGBLUP for zillions of genotypes?



ssGBLUP for zillions of genotypes?

- ssGBLUP

- Create **G** and **A₂₂**
- Direct inversion
- Cubic cost
- Limit ~ 150,000
- 1.5h for 100,000

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

Aguilar et al., 2010

Association/Species	# Genotyped	
US Charolais	2,500	✓
Fish	2,500	✓
Pigs	30,000	✓
Chicken	25,000	✓
Angus	210,000	X
US Holstein	1,185,000	X

Limitation ~150,000

ssGBLUP for zillions of genotypes = APY

- ssGBLUP - APY

Misztal et al., 2014

CORE



<http://www.cattle.com/articles/title/Angus+Cattle.aspx>

NON-CORE



<http://www.newmanangus.com/doctors/SAVBackupMay468.html>

<https://farmcentrefrancais.wordpress.com/page/2/>



<http://www.altreacattle.com/images/2008/oct20/Let6.jpg>



<http://www.rickobackangus.com/>



<http://denfield-whiteridge.com/phil.htm>



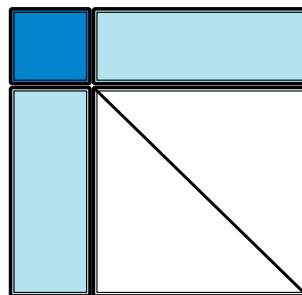
<http://www.hansenangusranch.com/>

ssGBLUP for zillions of genotypes = APY

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{G}_{pp}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{pp}^{-1} \mathbf{G}_{py} \\ \mathbf{I} \end{bmatrix} \mathbf{M}_g^{-1} \begin{bmatrix} -\mathbf{G}_{yp} \mathbf{G}_{pp}^{-1} & \mathbf{I} \end{bmatrix}$$

- ✓ 2,000 top bulls+cows
- ✓ 4,000 parents of genotyped
- ✓ 8,000 parents
- ✓ 10,000 highest accuracy
- ✓ 33,000
- ✓ Random 5k to 20k

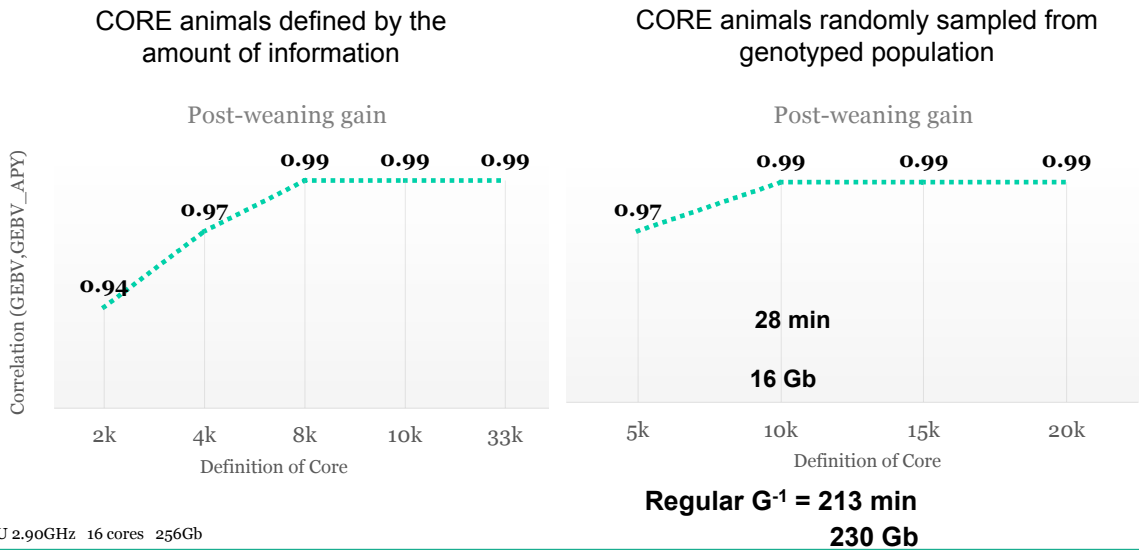
CORE
(invert)



$$m_{g,i} = g_{ii} - \mathbf{G}_{ip} \mathbf{G}_{pp}^{-1} \mathbf{G}_{pi}$$

NON-CORE

ssGBLUP for zillions of genotypes = APY



ssGBLUP for zillions of genotypes = APY

- **ssGBLUP – APY for US Holsteins**

Masuda et al. 2015

570,000 genotyped animals

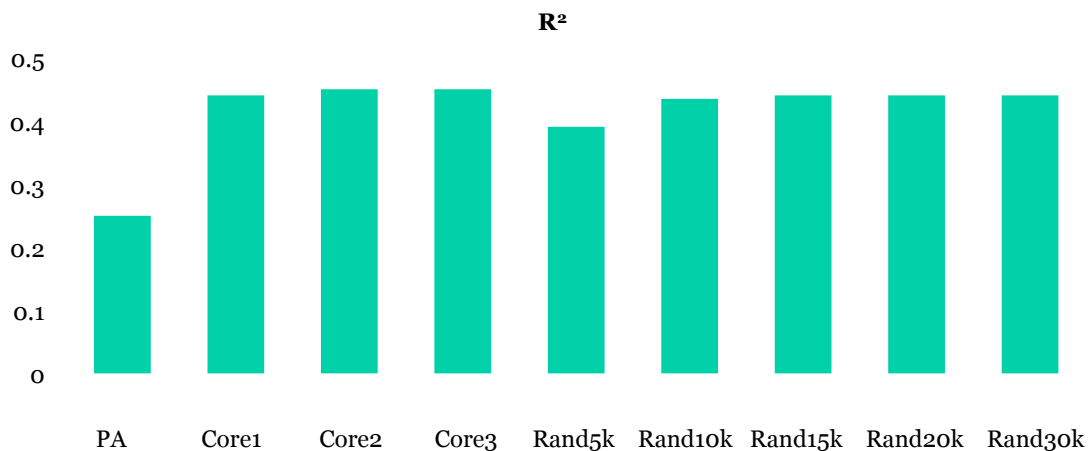
ssGBLUP for zillions of genotypes = APY

Definition - US Holsteins 570,000	Number of core animals
Core 1: Genotyped bulls with at least 1 classified daughter	9,406
Core 2: Def 1 + genotyped and classified dams of above bulls	10,458
Core 3: Def 2 + genotyped & classified cows	16,828
Rand05K, Rand10K, Rand15K, Rand20K, Rand30K randomly sampled animals from 77,066 genotyped animals born in 2009 or earlier	5,000 to 30,000

ssGBLUP for zillions of genotypes = APY

- ssGBLUP – APY for US Holsteins

Masuda et al. 2015



ssGBLUP for zillions of genotypes = APY

G-related computing	10K "core"	20K "core"
Read & store SNPs	18	18
Setting-up G_{APY}	26	48
Blend with A_{22}	26	25
Setting-up G_{APY}^{-1}	7	24
Total time	1 h 17	1 h 55

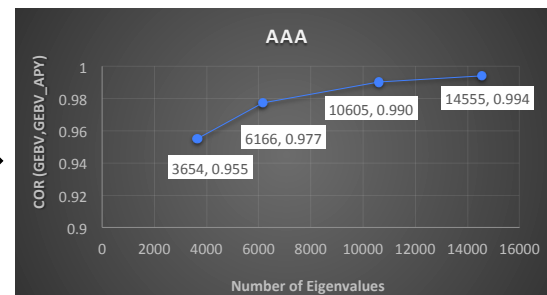
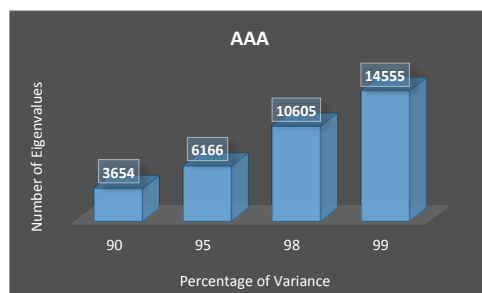
A ₂₂ -related computing	Time
Set-up the components	9
Other operations	2
Total computing time	11 min

PCG iteration	10K "core"	20K "core"
Time per round	11.7 sec	13.7 sec
Total time in 1,000 rounds	3 h 15	3 h 48

Step	10K "core"	20K "core"
Marker genotypes	16.1 GB	16.1 GB
Storage for G_{APY} and G_{APY}^{-1}	42.4 GB	84.9 GB
Peak temporary memory	6.4 GB	10.9 GB
A_{22} related	1.8GB	1.8GB
Peak requirement	66.7 GB	113.7 GB

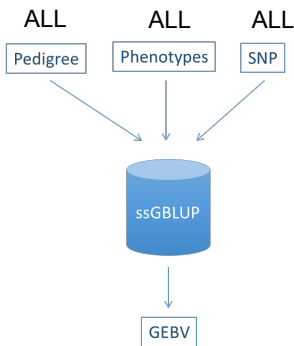
How to choose CORE?

- How to choose CORE animals?
 - Ne & Me & ESM & Eigen
 - Number of eigenvalues (Miształ 2016 & Pocrnic et al. 2016)



ssGBLUP for zillions of genotypes?

- Direct Prediction**



✓ Genomic evaluations

✓ monthly

✓ weekly

✓ New genotypes

✓ daily

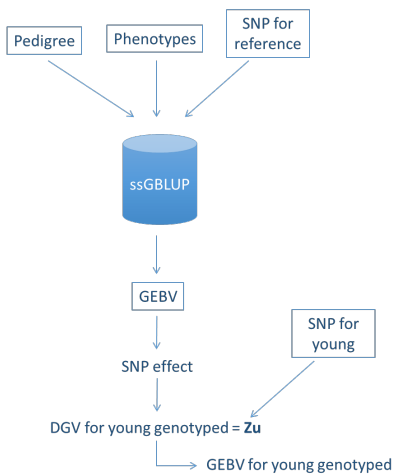
$$GEBV = w_1 PA + w_2 YD + w_3 PC + w_4 DGV - w_5 PP$$

parent average (PA), yield deviation (YD), progeny contribution (PC), direct genomic value (DGV), pedigree prediction (PP)

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

ssGBLUP for zillions of genotypes = IP

- INDIRECT Prediction**



$$\widehat{SNP}_{effect} = \hat{u} = DZ'G^{-1}(GEBV)$$

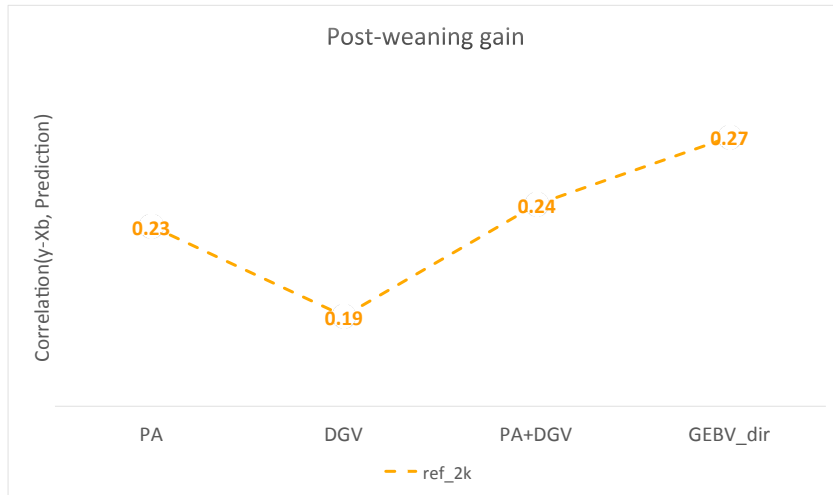
$$\widehat{SNP}_{effect} = \hat{u} = DZ'G^{-1}(DGV)$$

$$DGV = \frac{-\sum_{j,j \neq i} g^{ij} a^i}{g^{ii}}$$

$$GEBV \approx w_1 PA + w_4 DGV - w_5 PP$$

ssGBLUP for zillions of genotypes = IP

- INDIRECT Prediction



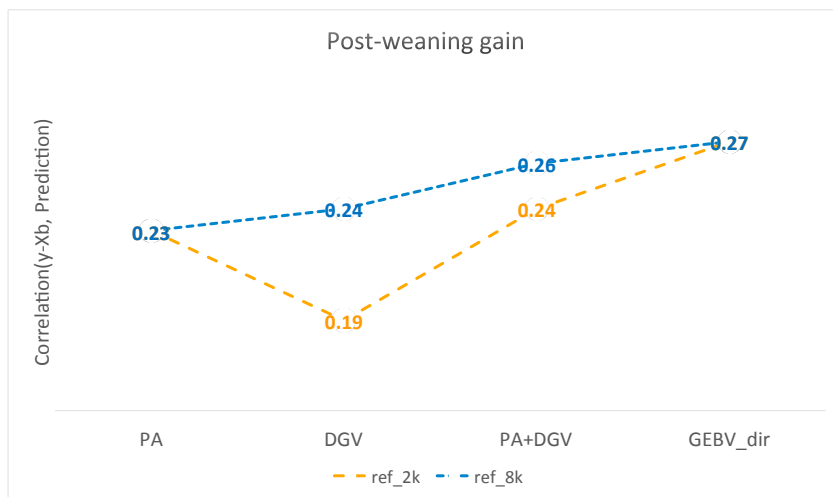
DGV = direct genomic value
= sum of SNP effects

PA = parent average
= mean of parents' EBV

GEBV_dir = GEBV from
single-step using all genotypes

ssGBLUP for zillions of genotypes = IP

- INDIRECT Prediction



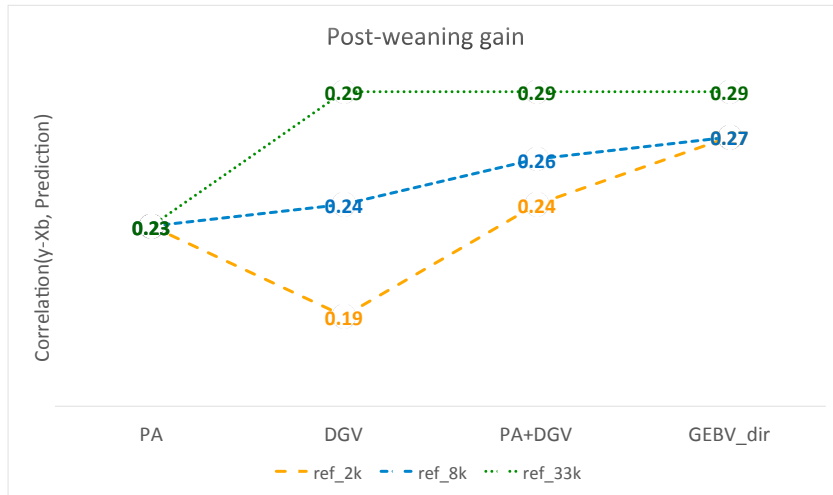
DGV = direct genomic value
= sum of SNP effects

PA = parent average
= mean of parents' EBV

GEBV_dir = GEBV from
single-step using all genotypes

ssGBLUP for zillions of genotypes = IP

- INDIRECT Prediction



few genotyped
animals - Index

$$\text{GEBV} \approx w_1 \text{PA} + w_4 \text{DGV}$$

More genotyped
animals – only DGV

$$\text{GEBV} \approx w_1 \text{PA} + w_4 \text{DGV}$$

ssGBLUP for zillions of genotypes?

TAKE HOME #2

- It is possible to run ssGBLUP for huge genotyped populations with the same accuracy as regular ssGBLUP
- APY with 10,000 – 20,000 core animals for cattle populations
- Young genotyped animals can have interim evaluations in ssGBLUP at a small computing cost