Should markers on the X chromosome be used for genomic predictions?

Guosheng Su[1], Bernt Guldbrandtsen[1], Gert P. Aamand[2], Ismo Strandén[3] and Mogens S. Lund[1]

[1]Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark

[2]Nordic Cattle Genetic Evaluation, DK-8200 Aarhus N, Denmark

[3]Biotechnology and Food Research, MTT Agrifood Research, 31600 Jokioinen, Finland

Background

➢ The X chromosome has some unique properties:

Sires do not pass X chromosome to sons.

Different relationship for males and females.

➢ EBV, DYD, DRP from a model with autosomal relationship

➢ Nordic GS use X-chr markers for RDC and Jersey, but not for Holstein.

➢ Little information on imputation accuracy of X-chr markers (Johnston et al., 2011) and effects of those markers on genomic predictions (VanRaden et al., 2009).

Objectives

➢ Investigate the accuracy of imputation of genotypes on X-chromosome

➢ Compare genomic predictions with or without X-chr markers

➢ Compare genomic predictions using **G** matrix with or without specifying X-chromosome

Data

5,643 Nordic Holstein bulls genotyped with 54K chip.

→   3,995 reference bulls born before Jan 1, 2005

       1,648 bulls in test data

Deregressed proof (DRP) of 15 traits

Datasets for analysis

54K: 54K marker data with imputation of sporadic missings

IMP_test: Bulls in test data with LD (7K) imputed to 54K

IMP_0.5ref: Half of bulls in reference data with LD imputed
       to 54K

Number of markers after editing (MAF >0.01, GC score >0.60)

Chip	Autosome	X-chromosome	
		PAR (~11cM)	X-specific
54K	43,314	133	694
LD	6,486	25	188

Imputation

Imputation of sporadic missings in 54 k data:

Beagle

Imputation from LD to 54K

Beagle: Imputation on whole X-chromosome

Findhap: PAR and X-specific segments were imputed separately.

Accuracy of imputation (allele correct rate, %) from LD to 54K

Dataset	Method	Autosomes	X-PAR	X-specific
IMP_test	Findhap	98.3	89.6	96.7
	Beagle	98.9	91.2	97.0
IMP_0.5ref	Findhap	98.0	89.9	96.2
	Beagle	98.8	91.1	96.5

➤ **Accuracy: Autosomes > X-specific > PAR**

Genomic prediction

GBLUP-A: G-matrix from autosomal markers only.

GBLUP-All: G-matrix from all markers by assuming X-specific markers as autosomal markers.

GBLUP-All$_X$: G-matrix from all markers and specifying X-chr.

GBLUP-A-X: BV divided into an autosomal component and a X-chromosomal component.

GBLUP-A-Polyg: GBLUP-A including resid. polygenic effect.

GBLUP-All$_X$-Polyg: GBLUP-All$_X$ including resid. polygenic effect.

Calculation of G-matrix correctly accounting for X-relationship (GBLUP-All$_X$ models):

Same rules as G-matrix for autosomes, but:

Element m_{ij} of **M** matrix is divided by $\sqrt{2}$, if the jth marker is X-specific and the ith individual is male.

Results

Reliability (%) of genomic predictions with or without X-chromosome, averaged over the 15 traits

Data sets	GBLUP-A	GBLUP-All
54K	38.0	38.5
IMP_test	37.9	38.3
IMP_0.5ref	37.8	38.3

➤ **X-chr slightly improve genomic prediction**
➤ **Almost no difference between imputed 54K and real 54K markers**

Reliability (%) of genomic predictions using G-matrix with or without specific calculation for X-relationship

Data sets	GBLUP-All	GBLUP-All$_X$
54K	38.5	38.5
IMP_test	38.3	38.3
IMP_0.5ref	38.3	38.3

➢ **A G-matrix with specified calculation of X-relationship did not improve predictions. Because all are males?**

Reliability (%) of genomic predictions by treating all markers as one or two components, mean over 15 traits

Data sets	GBLUP-All	GBLUP-All$_X$	GBLUP-A-X
54K	38.5	38.5	38.5
IMP_test	38.3	38.3	38.4
IMP_0.5ref	38.3	38.3	38.3

➢ **Dividing markers into two groups did not improve genomic prediction**

Reliability (%) of genomic predictions with or without including residual polygenic effect, average over 15 traits

Data sets	GBLUP-A	GBLUP-All$_X$	GBLUP-A-Pol	GBLUP-All$_X$-Pol
54K	38.0	38.5	38.9	39.3
IMP_test	37.9	38.3	38.9	39.2
IMP_0.5ref	37.8	38.2	38.8	39.1

➢ **Resid. polygenic effect improves genetic prediction (mainly for longevity and other-diseases).**

➢ **Using such model, X-chr still slightly improve prediction**

Reliability (%) of genomic predictions with or without X-chr for each trait (GBLUP-A vs GBLUP-A-X, real 54K data)

Traits	N	GBLUP-A	GBLUP-A-X	Difference	Var-Xchr
Milk	1159	48.7	48.9	0.2	0.9
Fat	1159	47.1	47.6	0.5	1.3
Protein	1159	45.9	46.2	0.3	1.5
Fertility	1158	40.7	42.6	1.9	3.6
Birth index	1642	32.5	32.7	0.2	0.8
Calving index	1239	30.3	30.5	0.2	0.7
Udder health	1204	39.5	40.1	0.6	2.7
Other diseases	1050	36.3	38.2	1.9	4.1
Body conform.	1156	27.6	27.4	-0.3	2.2
Feet & legs	1150	33.2	33.7	0.6	1.5
Udder conform.	1156	44.0	44.5	0.5	1.8
Growth	1351	47.2	47.2	0.0	0.0
Milking ability	1155	47.1	47.4	0.3	1.2
Temperament	1142	18.3	18.3	0.0	2.5
Longevity	817	31.1	31.8	0.6	0.8
Average	**1180**	**38.0**	**38.5**	**0.5**	**1.7**

Conclusions

➢Imputation accuracy for X-chr lower than autosomes, but still high (95%)

➢Including X-chr improves genomic predictions slightly (0.3-0.5%)

➢Recommend using X-chr for genomic prediction