



Increasing accuracy of genomic predictions: from SNP chips to sequence data

Daniela Lourenco

Shogo Tsuruta and Ignacy Misztal

6/12/2020 52nd Beef Improvement Federation Research Symposium and Convention



Genomics in animal and plant breeding

Theor Appl Genet (1983) 67:25-33

Genetic polymorphism in varietal identification and genetic improvement*

M. Soller¹ and J. S. Beckmann²

¹ Department of Genetics, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

² Institute of Field and Garden Crops, Agricultural Research Organization, The Volcani Center 50250 Bet Dagan, Israel

Received July 14, 1982; Accepted July 3, 1983 Communicated by A. Robertson

Summary. New sources of genetic polymorphisms promise significant additions to the number of useful genetic markers in agricultural plants and animals, and prompt this review of potential applications of polymorphic genetic markers in plant and animal breeding. Two major areas of application can be distinguished. The first is based on the utilization of genetic markers to determine genetic relationships. These applications include varietal identification, protection of breeder's rights, and parentage determination. The second area of application is based on the use of genetic markers to identify and map loci affecting quantitative traits and to monitor these loci during introgression or selection programs. A variety of breeding applications based on

Use of DNA markers

- Construct genetic relationships
- Parentage determination
- Identification of genes/QTL





SNP - Single nucleotide polymorphisms



http://www.thinnergene.com/about-thinnergene/genetics-101/

"SNPs have become the bread-and-butter of DNA sequence variation" (Stonecking, 2001)





What are SNPs doing?



- Identify genes
- Track relationships
- Parentage determination



How many SNP in the genomic test?

• Livestock, poultry, aquaculture: 3,000 - 800,000 SNP









https://www.lihs.cuhk.edu.hk/en-us/corefacilitiesandresources/corefacilities/geneexpressionprofiling,analysisandgenotyping.aspx



Genomic Selection





Adoption of genomic testing

3,400,000



http://www.holsteinusa.net/programs_services/backgrounds.html

780,000



http://www.angus.org/AGI/default.aspx

460,000



https://www.usjersey.com/AJCA-NAJ-JMS/AJCA/AnimalIdentificationServices/HerdRegister.aspx

50,000 - 70,000



www.sheepcrc.org.au/

~50,000/line



http://sesenfarm.com/raising-pigs/

~50,000/line







https://www.cobb-vantress.com/

Benefits of genomic testing in beef cattle

Average gain in accuracy

52k animals = 25%

335k animals = 36%

Genotyped for 50k SNP

The gain in accuracy is because genomics is a non-redundant piece of information!



Angus

0.47

UNIVERSITY OF



Benefits of genomic testing

Trait	PE	Trait	PE
Calving Ease Direct	26	Heifer Pregnancy	17
Birth Weight	23	Calving Ease Maternal	20
Weaning Weight	27	Milk	36
Yearling Weight	23	Mature Weight	15
Dry Matter Intake	12	Mature Height	9
Yearling Height	17	Carcass Weight	15
Scrotal Circumference	15	Carcass Marbling	11
Docility	12	Carcass Ribeye	17
Claw Angle	10	Carcass Fat	14
Foot Angle	10		

What's a SNP genotype worth for fertility?



Cole, 2020

https://www.angus.org/AGI/GenomicEnhancedEPDs.pdf



Benefits of genomic testing

$\Delta \boldsymbol{G} = \frac{i \, \boldsymbol{r} \, \sigma_a}{\boldsymbol{L}}$

Genetic gain = (selection intensity * accuracy * genetic SD) / generation interval



Benefits of genomic testing in dairy cattle



Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection

Adriana García-Ruiz^{a,b}, John B. Cole^b, Paul M. VanRaden^b, George R. Wiggans^b, Felipe J. Ruiz-López^a, and Curtis P. Van Tassell^{b,1}

^aCentro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Ajuchitlán, Querétaro, 76280, México; and ^bAnimal Genomics and Improvement Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705

Genotyped for ~ 60k SNP

- Increase in genetic gain
 - ~50 to 100% for yield traits
 - 3- to 4-fold for traits with low heritability (fertility)



Benefits of genomic testing in pigs





Benefits of genomic testing

- Increase in accuracy
 - Number of genotyped animals
 - Amount of performance records
 - Heritability of the trait
 - How informative the genomic test is
 - 50k SNP in beef cattle

Is the genomic test informative enough?

Do we have enough SNPs in the test?

Are we looking at the right SNPs?



About the right SNP

- Is the genomic testing informative enough?
- Do we have enough SNPs in the test?
- Are we looking at the right SNPs?

illumina

54k SNP cattle –	 38k SNP
54k SNP pigs –	 44k SNP
54k SNP chicken	 39k SNP



Sequence information

21		31	41	51	61		1 #1	92	101	111	131	131	141	151	161	171	181	191	- 24	1	
CARCETAT	TABOTAS	Transfer to the	CTOCO DE CO			Concerner of	TATION AND CONTRACT	An excitation of the second	TOPACTORIA	1000000000	COLUMN THE PARTY			CATEGO LITT	THE REAL PROPERTY.	and the second	TATTACASCO	100000	-	Contractor of	-
CACCO II	THECHO	Circulation of	0 001 00	11100			In the occurrence	all second decide	100 0 00 1	CINCOL	GACC G	6 64	330 33	CA ODEA A	A COBAGE	ALC: CAN	A MENGEO	CLICKTAC	10 100	5 6 C	ALL ALL
	ANCENE	CHEGOGAGE.				e aaaa	A GENERALING	AT BEGRANDER	Tel-Gridero		616	CELLING.			TTTATO CHECK		tattacegoto			statett	
	AVOCVC	IENEGOGNEE!	CTOCA CO			C-000000	TA TECACOCEAT ACC	A TECONORICES	THE REAL PROPERTY.	CACCCUAT							Compensate	SOLD BE		101011	
1 300	AAOCAC	CAEGGGGAGE	C CC1 CC	6 <u>66</u>	4 <u>5</u> 5	200000	ATGCACOCGA AGC	AT BEGAGARARD	105-07-06-0	CAOOC AT	G	5 S	300 33	CA ODCA A	A COEACE	CG CM	AT ACADOCO	ALEA AC	15	5.6 C	
CEE A	ANDENE	ICAL COCKOC	C CEA LC	1 22		Contraction of the second	A TOTAL DE DE DA TADE	A THE GAS CON	100 01 00 0	CACEC A	6		00 600	CATOREA IA	A COCACE	ALC: CAN	A TACKDORD		10 100		
CCC 8	AACEAC	CHEGGEAGE	C CEA CO	35 8	A CG	COCCCC	A TECHCOCCA TACC	A GEOLEVICE	100 01 000 0	CACECTAT	G	i i	CC 600	CA COCA A	A COCACC	ALC CA		MPARE	10 1.00	C 10 10	0 - D
	ABECHE	CHESSENSES	C CCI 10	INNER	ettereg	ET+00000	INTERACOCONTACT	ATTROOMETERS	10010010010	040001010			10001000	EATODECATTA	TTA TOGCALCO		TATTACAGECO			6161611	
	AVECVC	ICALOGONOC				<	INTECNCOCONTINCC	AT TOCONONCOR	100-01-00-0			1011100					terteceppop			states	
	ANDENE	CAEGGGAGE	C CEA CC	A 66	A EG	000000	A TOCACOCOA TADO	AT BEGAGAEOR	Jee et cere	CACCE: A G							TATTACAGECO	AACATAC		6 G G	
5 .	AACCAC	CAEGGGAGE	C CEA CO	1 99	4 <u>9</u>	600000	A GENEGERA ADD	AT DECASALSO	100-00-00-0	CACEE A G		110	CC 000	CH DOCH IN	E COCICC	- See		ADC DOD C	16 110	1 9 9 1	-
1	AMOUNT	I AT DECADE	C CCA DO		1 12	- denter	TATES ACCOUNTS OF	ATTIC: GACATO	100 0 00 0	CACCOLLE			Les mer	CA COLL IN	Laborate.						
	ABCCAC	CACODODOC	E 1004 80	111001	at the long		a locacocoal ago	etteconances	100-01-00-0	CARGE IN IS			10001000	CALODCALLA	A COCHEE	-	TRITING PORCE		40 144	digitant.	100
COLUMN T	ARCONC	COCOCOCOCOC	C 1006 80	ALC: UNK	accessory.		centechcecost rec	all aconances	101-0-01-0	CACCOLATE			ITECIOCE!	Concecuta	ITT AT COCHECT	ACC CAR	CARD ACTING	and a first	Car Land	sterets	
		1000000000				C 000000	INTEGROCONTINC	al lacaratess	10070440070						A COCACE		TATIA AGON			6-6-6-1	
	URCH	10102020121	E 664 BS			8 000000	A IBCACGCON ABC	* BCOHORCOR	B89464.0849	CPCCC18191			2229 222	CATODCATTA	A CECACO	ACC CAN	AT ACAGECO	ALC: NO	46 142	6.6.6	
	AMPEAC	The balance	C CCA DO	1 St.	1 1	000000	A THE REDUCT ADD	AT DEGREATE	TOCION CONTO	CALLE ATE	1		TO DEC	CATOCOLATIA	A COCACE	ALC: CAM			- 200		
	AND COMPANY	Concession of the	1 114 H	e			In the second second		Transie on the				10 000	CALCOCAL IA	A COCACE	1 CA	In the story of		10 110	6 6 6	
	I BRECHC	Care and a lot	E DES 10			1.0000	(e) accesses lend	al second state	105-0-00-0	LOCEC	0.10		3331 331	CA COLA A	A COLNEC	ALC CA	THTRACKCOCC	ALCA NO	10 100	6 6 6	Minut
	ROLLOG	LICENSIDE	C LLCS OF	COLUMN 1	11110	B-1007031	In Interestant and	rt actioner	100-01-00-0	LOCCEPTO	010		Decigee)		Intercepter (Collected to	ABC PERCE		states	
		11.000000000				Concess:	Internessentates	al managements	100-04-00-0								inthe segue			etolett	
		Teneggage	C CCA BC	11.881	111691	6 0000000	angenegege age	UT BEGAGACES	100 9 99 9	CADDE # 9	E CHI I A		000	CA DECK A	A COCACC	NEC CAN	I AT ACAGGCO	ANCA AC	10 100	6 6 6	AATE AA
		CARGOGAGE	C CEA OF	6 <u>66</u>	4	Coccess.	A BEARGEST ADD			CACCE A G	DOCAG A C G			CA COCA IA	A COCACC		A ACAGO C		10 100	5 6 6	
		ACCOUNTER 1	C 00.4 00	2		1	a lacacecea labo	A CONDECCO	100 00 00 00	CACES IN G	COCHE A C C			DEEA A	A COLACC	ALC CAN	LATIAL MOST O		46		
		GAGE	E DEA 66	Anneer	A	a anna	A DEPARTMENT NOC	A THE CALL OF	100 01 00 0	CACCETATE	COCAL A C L			CCC4T14	Intelligence	ACT CAR	Competences			101011	100
		32	5 554185			5 222222	(afgenegegatings	at laconances	10010550010	cessel alg	egeodiate ig	tettigs			It alogeace	obg CAR	Calificopper			allakatt	
						6 999999	to to cacacao tage	a seconder be	199996499999		ogcagtatetg						Latincaggeg			at grant	
			C CC3 85			C 200333	INTREACOCOST POL	el lacararca:	10810-00810	CHEES NTO	Caces a Cill				A LICKER	OCAL CAR	I I MERSELS		BCLER	1.0.0	
			C CEA DO		1 10	0000000	A GEALGEGE AGE	A DEGREATER	105-101 (05-10	CACCE A G	DECASIA C G	1 2	00		A COEARE	ACC CAN	A ACADER				
			C CEA DE	1 60	A 26	autore a	A GENERICA ME	ATTRACKS PER	100 11 100 1	CACCOLATE	COCAG A E G	8 a	THEC		La Cacaca		Contraction of the				
			CTCCA TEC			a cettere	ATGCACOCGATAGE	A BEGAGIEGO	TOC CLICCOLD	CACCCTATE	COCAG A C O				IN A LODGERCO	ACC COL	CATLACADOLS	and share	No. toos	enoners.	
				atttegt		E 1000000	a focacocoa tage	a gegagebge	10040640940		ogcegta cig	1211104			(cociecie)		Tert acagoo			at a latt	
						E1+50555	INTOCHESSON FOR	AT BESTRICSS	Tesessergensi						COCHCE!	ACC CAR	Callacapper	ABCOTOC		enerert.	
				1 (QQ)	a	E 000033	IN GENEREDA NOC	A BCRARKES	00000000000	CHEER # 12	59549 4 6 8		20, 33		00400	MEG CAA	ACAGGET	ALC: AT AC	16 LU	5 6 6	
				1.1		1000	a percessi ape		100 3 20 1	CODEC IN C	0,00,00	1 2		CATEC	CHCC	10 St.			10 144		
				5	1 8	COLORADO	A TECATOCEA ADD	A BECKLEREN	12. 2. 2. 2	CACCO A TO	COCAG A C G	1 2	100 000	CATCODAT		100 LEAR	In The sector	And a line	1.0.1.0.0.0	eneligin	
							A TOCACOCOA TAGE	A BEGAGALON	100-01-00-0		COCAGTA TO TO					HOG I COM	fattacagory	SHCPTOC		ang ng hit	
						C.600000	la tecnogoga tago	alleconseces	199-94499-94		eacedia:c.d.	tertings	thee toes				CATTACA9909			sighter .	
						6 500000	In GENERES INCO	of scondrood	1951544810	CACCE N B	69699 9 6 9	1.0	DE DEE	CALCOCAL IN		VEC CVV	T. WWW	ANDARA	10 144	6.6.6	
					50	1000000	a genegena ant	a <u>sconger</u> oo	100-5- 00-5	100000 8 9	0.001.0.0	5 8	CC DCC	CD DECK		C.81	B THEB STOR	OPC OF THE			
					100	1 1111	a lacer of an last			LINE A	Contract of the local sectors	1	60 000	CALCULA IN			Terra and				
					100		a locacocda lago	allipepapatos	100 0 00 0	cecccie igi	cocagia (cilo)	101110	ttecigeet	calcocetta.		CAN	TATIACAGECO	ALCAULT	TREAM	CIG:GIT	ALC: NAME
						C 600000	ATECACOCOAT NOC	ATTECENERCE	TOG ACK COGAC		COCAG A C G	GIGA					TATI ACASSIS	ALCATAC		5.0.5	
						0000	IN TECHNOLOGIA I NEK	V. BUCKNENCO	105-6-06-6	CACCELATE	COUVER 14 10			CATOOCATTA	TELATOSCACC:		CATTACA6600	ALC: NO		6161C	
							A TECHCECEN THEC	Y CONCINCION	166-sector-o	CACCOLATE	COCAG A C G	e ci	00.000	CA ODCA A	A COCACE	YCC CYN	A MENERCO	ALC AT ALC	10 1.11	6 6 6 F	
							67 67 57 55 1967	A CAMPAGE		ACCOUNTS ON A	CORAL A TO D	1 3	10 000	CA SECTION A	1 2 2420		10010				
							Grandrana	L'ECCLEMENT	The server is	CACCOLLIN	COLAG A COL	TE THE	THEFT	EATECEATA	ATCOCACT	ACCTICA	The second second		No. Inc.		
							ACCOST LABO	al Inconsecor	100-00-00-00	CHECK 19101	coces of city	101110	tice ace	CR SECC118	IT I COCHCC	ALC: CH	HEREE	ODC DECK	the laws	10.9	a attant
							HERCONSIDE.	AT INCOMPANY	100-01100-0	CARECTINIO	00400101010	141110	TRE'QCC		TTA COCACC	acg to car	-			101011	
							es gegel ago	a	00-01-00-0								900			a. a. a	
							HERCON HOC	B CONSTRUCT	22 2 2 2	0.0000	caced a c d	9	20 22	C8 0004 18	a cocace	SE CH	CC	ALLAC	AC 444	666	100
							COLORIAGE	ATTER EXEMPTER	Tec-cr cc-r	CACCO A G	COCAD ATT O	1 1	TEC TOOL	CA COTA LA	A COCACE	A24 644	-	ALC: NO	40 444	2.2.2	
							the second se	the second se	the second se				the second se								

http://mtc.science/playing-around-with-ngs-step-by-step

50k SNP

30M SNP



SNPs capture relationships at genes



- Expected relationship between a bull and its grandparents?
- Observed relationships based on proportion of alleles shared between a bull and its paternal grandsire and maternal grandsire?

With sequence data, we may find SNPs that give more precise information about genes because they may be closer to genes



Sequence information



- Trying to get a more precise information about the genes
 - More SNPs to cover larger areas of the DNA
 - Select SNP that are closer to genes for traits of interest



Sequencing animals

- Becoming cheap
 - Sequence a small portion of animals
 - Imputation from 50k to sequence (filling the gaps)

RESEARCH ARTICLE



Accuracy of whole-genome sequence imputation using hybrid peeling in large pedigreed livestock populations

Roger Ros-Freixedes^{1,2*}, Andrew Whalen¹, Ching-Yi Chen³, Gregor Gorjanc¹, William O. Herring³, Alan J. Mileham⁴ and John M. Hickey¹

- 18,000 to 107,000 pigs
- Sequenced 2% of the animals
- Imputation to 20M to 30M
- Imputation accuracy of 94% 98%



Largest sequence data in pigs





RESEARCH ARTICLE

Open Access



Roger Ros-Freixedes^{1,2*}, Andrew Whalen¹, Ching-Yi Chen³, Gregor Gorjanc¹, William O. Herring³, Alan J. Mileham⁴ and John M. Hickey¹



- Gain in accuracy
- Multibreed
- Persistency of accuracy
- Up to 100,000 sequenced/imputed pigs
- 20M to 30M SNP

Do we work with all ~30M SNP?



Mining sequence data



YOU COULD THINK OF IT AS IF SOME SNPS ARE CARRYING A TINY CAMPAIGN SIGN SUGGESTING WHICH GENE THEY'RE ASSOCIATED WITH.



Mining sequence data

Birth weight in Angus



- Causative SNP
- SNP give information about birth weight

Aguilar et al., 2019

Chromosome



- Issues
 - Effective population size
 - Number of animals with sequence and phenotypes
 - Large data more confidence in finding causative SNP
 - Small data more difficult to identify

- more susceptible to errors













- Small data: not enough power to discover which SNP are truly causative
- Methods are also important







Using selected SNP to compute GEPD

- Hanwoo cattle from South Korea
- 545k Marbling score records
- 1.3M animals in pedigree
- 1160 genotyped animals
- Imputation: 50k to 777k to 11.1M SNP
- 321k SNP selected out of 11.1M sequence SNP



Jang et al., 2020

en.prnasia.com



Gain in accuracy with selected SNP

Jang et al., 2020





Gain in accuracy with selected SNP

- VanRaden et al. (2017)
 - 27k US Holstein bulls
 - Selected 16.6k SNP based on effect size on 33 traits
 - Added to the 60k SNP chip





Gain in accuracy with selected SNP

- Data from VanRaden et al. (2017) Bayes A
 - 27k US Holstein bulls
 - Selected 16.6k SNP based on effect size on 33 traits
 - Added to the 60k SNP chip
- Fragomeni et al. (2019) ssGBLUP



Stature in US Holsteins



Simulated data – real genes

- Simulated population
 - 180,000 animals
 - 30,000 genotyped
 - 60,000 SNP
 - 100 genes (QTN)
- Single-step genomic predictions based on
 - SNP
 - SNP + QTN
 - QTN

Fragomeni *et al. Genet Sel Evol (2017) 49:59* DOI 10.1186/s12711-017-0335-0



RESEARCH ARTICLE



() CrossMark

Incorporation of causative quantitative trait nucleotides in single-step GBLUP

Breno O. Fragomeni^{1*}, Daniela A. L. Lourenco¹, Yukata Masuda¹, Andres Legarra² and Ignacy Misztal¹

Accuracy with simulated genes





Fragomeni et al. (2017)



Accuracy with simulated genes



60k + 100 QTN – ssGBLUP APY – True weights



SNP closely linked to genes

SNP that turn genes on and off

Amount of information they give

Is it possible to increase accuracy of genomic EPD with sequence data in real cattle populations?

We don't know all the genes Knowing a bit may help a bit Lots of work in progress



Acknowledgements













Smithfield







USDA United States Department of Agriculture Agricultural Research Service

Warmwater Aquaculture Research Unit







