

# **Application of a custom SNP chip: Microsatellite imputation, parentage SNP imputation, genomic evaluations, and across-breed nation-wide genetic disease prevalence with the International Beef and Dairy SNP chip**

Matthew C. McClure<sup>1</sup>, Michael P. Mullen<sup>2</sup>, J. Francis Kearney<sup>1</sup>, Andrew R. Cromie<sup>1</sup>, Mags Treacy<sup>3</sup>, Paul Flynn<sup>3</sup>, Rebecca Weld<sup>3</sup>, Donagh P. Berry<sup>4</sup>

<sup>1</sup> *Irish Cattle Breeding Federation, Bandon, Co. Cork,* <sup>2</sup> *Animal & Grassland Research and Innovation Centre, Teagasc, Athenry, Co. Galway,* <sup>3</sup> *Weatherbys DNA Laboratory, Johnstown, Co. Kildare,* <sup>4</sup> *Animal and Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland*

## **Abstract**

Since spring, 2013 over 24,500 Irish beef and dairy cattle have been genotyped on the custom International Beef and Dairy (IDB) SNP chip for 16,562 single nucleotide polymorphisms (SNP). The IDB has been used in genomic evaluations, parentage verification and assignment, as well as determination of animal genetic disease and major gene status. When the putative sire had no SNP genotypes available, parentage verification was undertaken on >10,000 cattle using 1,002 IDB SNP to impute ISAG microsatellite (MS) alleles with >98% accuracy for MS-based parental verification. Any animal with >1 MS concordance error between its imputed MS profile and its listed sire was directly MS genotyped to maintain stringent verification quality levels. Additionally we present here that imputation of parentage verification SNP from >20 genotyped progeny is achievable at >99% accuracy. MS imputation is aiding, and progeny-based SNP imputation will aid, the transition from MS- to SNP-based parentage verification. MS imputation has saved the Irish cattle industry >€200,000, as the average MS genotype cost parentage verification is €20 per animal. Diagnostic SNP for genetic diseases and major genes were included on the IDB to provide more accurate breeding advice. Direct genomic values (DGVs) for dairy cattle are as accurate when the IDB is used as when Illumina 50K data is used. Overall, the IDB has had a major impact on the Irish cattle population enhancing DGV predictions, assisting in the rapid transition to SNP-based parentage verification, and facilitating precision genomic matings to avoid unwanted allele frequencies in lethal mutations and major genes.

## Introduction

As the livestock industry moves into the genomic era a major financial hindrance is the transition from microsatellite (MS)- to single nucleotide polymorphisms (SNP)-based parentage verification. To address this, the custom International Beef and Dairy (IDB) SNP chip was designed to include SNP for imputation of MS alleles (Mullen et al., 2013, McClure et al., 2013). MS markers have been the standard for genetic verification of an animal's pedigree for the past few decades, while SNP are quickly becoming the preferred standard. When compared to MS, SNP have higher genotyping accuracies, are easier to automate, have standard allele calling across platforms and labs, have lower error rates, and have greater power for pedigree verification when sufficient numbers are used. Less than 45 SNP are needed to provide  $\geq 2$  orders of magnitude greater matching probability than 11 MS markers for parental verification in cattle (Allen et al., 2010, Fernández et al., 2013). Therefore, the International Society of Animals Genetics (ISAG) recommended set of 100 SNP ([http://www.isag.us/Docs/ISAG-ICAR\\_Cattle\\_Comparison\\_Test\\_100\\_SNP\\_panel.xlsx](http://www.isag.us/Docs/ISAG-ICAR_Cattle_Comparison_Test_100_SNP_panel.xlsx), (Heaton et al., 2002)) represents a large improvement in accuracy for parental verification over the 12 MS markers that comprise the ISAG recommended MS bovine parentage panel ([www.isag.us/Docs/CattleMMPTest\\_CT.pdf](http://www.isag.us/Docs/CattleMMPTest_CT.pdf)).

While SNP are more powerful, if a producer is only wanting parentage verification MS are still more economical, although the price gap between genotyping for  $\geq 100$  SNP and 12 MS is narrowing. Traditional agriculture SNP panel designs were focused on genomic breeding value calculations, SNP-based pedigree verification, SNP imputation, and for research use. For all the advantages of SNPs, many livestock consumers that wish to utilize SNP technology will still need to spend ~€20 on MS genotypes for parentage verification if their animal's parents only have MS genotypes.

The Irish Cattle Breeding Federation (ICBF) in collaboration with Teagasc and Weatherbys Ireland, recently developed a custom SNP chip called the International Dairy and Beef (IDB) for dairy and beef cattle breeding (Mullen et al., 2013). The IDB is a a low cost custom SNP genotyping platform suitable for *Bos taurus* cattle populations which was designed to 1) have improved imputation to higher density SNP panels for genomic selection (especially in beef cattle), 2) provide highly accurate pedigree verification, 3) screen for known lethal recessives and genes of major effect, and 4) impute MS alleles for the ISAG recommended MS bovine parentage panel.

Traditionally, when a livestock industry transitions to a new technology for parentage verification the cost of genotyping the transition generation with both the new and old technology has been absorbed by the producer or breed association. For example, during the previous transition between blood typing to microsatellites based analysis for parentage, all transition animals were blood typed and microsatellite genotyped at a tremendous cost to the livestock industry. The IDB incorporates an approach developed to impute MS alleles from specific SNP haplotypes, thereby eliminating the requirement and cost for the transition generation to be genotyped for both MS and SNP (McClure et al., 2013). This MS imputation has been successfully implemented in Ireland and has greatly aided the transition to SNP-based parentage verification for the Irish cattle population.

While MS imputation can be a major financial help while transitioning from SNP-based pedigree verification it can represent a workflow bottleneck. MS data can be used to predict the most likely sire but only if an animal's dam has been MS genotyped. Due to the low number of dams with MS data the producer must often incur the additional cost of MS genotyping the dam when MS-based sire predictions occur. Sire prediction with SNP data does not require the dam to be genotyped, saving the producer time and money. If a historic sire who was previously MS genotyped has no available genetic material for SNP genotyping all of his progeny must be MS genotyped or have MS imputed for them to be progeny verified.

Recently Berry et al. (2014) showed that a parent's genotypes could be imputed from 12 progeny at 95% accuracy. While this imputation accuracy could be acceptable for genomic breeding value calculations, it is too low for parentage verification which requires  $\leq 1\%$  misdiscordance rates between and animal and its progeny. We evaluated the accuracy of imputing a sire's genotype from much higher progeny numbers than Berry et al. (2014) for just SNP used for parentage verification or identification.

Currently in Ireland, for both beef and dairy animals, the ISAG bovine standard 100 parentage SNP (ISAG core) and a set of 1,970 autosomal SNP (2K) that are common across all commercial Illumina bovine SNP chips and the IDB are used for are used for initial sire verification or to predict the most likely sire when an animal's true sire is unknown, respectively. If a historic sire could have his ISAG core or 2K genotypes imputed at  $\geq 99\%$  accuracy from his verified offspring, via imputed MS genotypes, then his future progeny could be verified via SNP and he could also be included in sire prediction pipelines.

## Material and Methods

**Genetic disease and major gene mutations.** Causative mutations for multiple lethal recessive, congenital disorders, and genes with large effect on a quantitative trait were included on the IDB using published probe information or data from labs that held the mutation's intellectual property (Mullen et al., 2013). DNA from confirmed carriers was obtained and used to validate the diagnostic probes.

**MS imputation SNP.** In McClure et al. (2013) 892 of 3,744 SNP from the Illumina HDSNP chip (Illumina Inc, 2010) that are located within 500kb of 12 MS markers (*BM1818*, *BM1824*, *BM2113*, *ETH3*, *ETH10*, *ETH225*, *INRA023*, *SPS115*, *TGLA53*, *TGLA122*, *TGLA126*, and *TGLA227*) were identified as the minimum set required for accurate imputation of MS alleles. To allow for the possibility of haplotype adjustments 2,176 of those 3,744 SNP were included on the IDB which included all of the 892 minimum SNP. Additional SNP were chosen in increasing distance from the MS. Only the original 892 minimum SNP were initially used for MS imputation.

**MS reference data sets.** The original publically available 7,130 animals from McClure et al., (2013) were initially used as the sole MS imputation reference dataset. For each MS marker reference data set only animals with a genotype for that MS were included. On regular intervals additional Irish animals that had MS and IDB genotypes were added to the reference dataset, with the current reference dataset including 8,258 animals. The reference datasets were fully rephased each time additional animals were added using all possible SNP that are within 500KB of each MS.

**IDB genotyping.** In Ireland >24,500 commercial and pedigree beef and dairy animals representing 12 breeds have been genotyped with the IDB chip at Weatherbys Ireland (Naas, Ireland). DNA was extracted primarily from hair roots and also from ear punches, blood, and semen as required. Extractions were performed using standard laboratory practices. Genotyping with the IDB chip was performed using the Illumina recommended protocols. Any animal with call rate <0.90 had its genotype removed and a replacement tissue sample was requested from the animal's owner.

**MS imputation.** Animal's genotyped on the IDB had BEAGLE (Browning and Browning, 2007) input files created for the MS imputation SNPs. In BEAGLE each MS was separately imputed using the current MS reference dataset.

**Parentage verification.** Each animal was initially parent verified using the ISAG core SNP (Heaton et al., 2002) if their reported sire also had SNP genotypes. If  $\geq 2$  SNP were

in misdiscordance the reported sire failed verification. Next all SNP failed animals had their sires and dams predicted using all ICBF SNP genotyped animals (N >35,000) using 1,970 autosomal SNP that are present on all commercial Illumina bovine SNP panels and the IDB. Any sire or dam with  $\leq 5$  SNP misdiscordances were considered to be the animal's true parent and  $>5$  but  $\leq 10$  misdiscordances were considered for further investigation as the potential true parent.

Animals whose parentage could not be SNP-verified had imputed MS used for parentage verification. If  $\geq 2$  misdiscordances were present between an animal's imputed MS profile and its reported sire the animal was directly MS genotyped at Weatherbys DNA Laboratory (Johnstown, Co. Kildare, Ireland) and used for sire verification or identification. This is in line with previous publications that recommended  $\geq 2$  MS misdiscordances for an animal to be excluded from parentage verification (Bonin et al., 2004, Weller et al., 2004, Baruch and Weller, 2008).

**MS imputation SNP haplotype size adjustments.** During the fall of 2013, it was observed that 1-2 specific alleles for *BM2113*, *TGLA122*, and *TGLA126* were consistently being imputed incorrectly, mainly in a breed specific manner. In October, 2013 the haplotype sizes were reanalysed and 100 additional SNP were added to the original 892 minimum MS imputation SNP set. While the majority of the 100 SNP were added to the genomic regions of the 3 mentioned MS, most MS genomic regions had a few SNP added as a measure to capture any unique MS-SNP haplotype that might appear in future reference data sets.

**TGLA53 exclusion.** Some labs do not use *TGLA53* in their parentage testing due to well documented issues where its larger alleles do not amplify well and thus can cause problems with parentage verification (John Flynn, Weatherbys Ireland, Personal Communication 12/04/2013). During the October, 2013 reanalysis *TGLA53*'s performance in 4,100 animals was examined to determine if it should be included in imputed MS based parentage verifications.

**Parentage SNP imputation.** All available Irish genotypes from March 28, 2014 were used: 4,360 Illumina HD, 4,260 50K (Matukumalli et al., 2009), 24,490 IDB, and 2,987 LD (Illumina Inc, 2011a), genotypes. Table 1 compares the autosomal SNP shared among the commercial Illumina SNP chips and the IDB. SNP were filtered on call rate  $\geq 0.90$  and autosomal placement on the UMD3.1 bovine assembly (Zimin et al., 2009), animals were filtered on passing parentage verification; any animal genotyped on multiple chip platforms had genotypes only from the higher density chip retained. After filtering 35,616 animals and

740,354 SNP remained, table 2 shows genotype by breed. A 6 generation pedigree from the ICBF database was generated to allow for pedigree based imputation methods. FImpute v2.2 (Sargolzaei et al., 2011) was used to ensure that no pedigree errors were present between a genotyped parent and their offspring at a 0.001 misconcordance rate and to ensure that no duplicate genotypes were present.

SNP genotyped sires had their SNP genotyped progeny counted and 8 progeny groups were identified that balanced the number of offspring genotyped versus the number of sires present in the group. The groups (and count of sires in each) were 8-9 progeny (97 sires), 10-14 (124), 15-19 (80), 20-29 (66), 30-39 (42), 40-59 (37), 60-99 (29), and  $\geq 100$  (26), Table 2 shows this sire count by breed. The 498 sires were not used across multiple progeny groups. FImpute v2.2 was used to impute each sire group separately at 4 separate SNP densities: 2K, 3K, or LD, where 3K is the Illumina GoldenGate Bovine3K beadchip (Illumina Inc, 2011b). For each imputation the sires in the selected group had their genotypes fully excluded and all animals had their genotypes filtered down to the selected SNP density. For instance, for the 8-9 progeny imputation at the LD level, only autosomal SNP present on the LD chip were used, and only sires that had a maximum of 8-9 genotyped progeny had their genotypes fully excluded.

FImpute v2.2 was used for imputation of each sire group by SNP density combination. Imputed genotypes for the ISAG core and the 2K SNPs were then compared to the sire's original genotype. Genotype accuracy was calculated as  $(\text{SNP}_{\text{max}} - \text{SNP}_{\text{imputed}}) / \text{SNP}_{\text{max}}$  where  $\text{SNP}_{\text{max}}$  was 100 for ISAG core and 1970 for 2K and  $\text{SNP}_{\text{imputed}}$  was the number of accurately imputed SNP in the ISAG core or 2K SNP set.

Table 1. Comparison of SNPs with placement on the UMD3.1 chromosome 1-29 on commercial Illumina chips, the IDB, and the 2K parentage panel.

	HD	50K	IDBv1	LD	3K	2K
HD	<b>73,529<sup>a</sup></b>	90.40% <sup>b</sup>	100.00%	100.00%	100.00%	100.00%
50K	47,810 <sup>c</sup>	<b>52,887</b>	75.58%	99.79%	98.20%	99.95%
IDBv1	14,073	10,637	<b>14,073</b>	100.00%	77.85%	100.00%
LD	6,651	6,637	6,651	<b>6,651</b>	77.85%	100.00%
3K	2722	2673	2119	2119	<b>2722</b>	99.24%
2K	1985	1984	1985	1985	1970	<b>1985</b>

a) Diagonal is SNP count on a chip

b) Above diagonal is the percentage of SNP shared between chips

c) Below diagonal is the count of SNP shared between chips

Table 2. Count of genotyped animals by breed and SNP chip, post filtering.

Breed	All Animals					Total	Imputed Sires
	LD	IDB	50K	HD			
AAN	79	1,365	.	521	1,965		24
AUB	1	1	.	.	2		.
AYS	.	.	.	2	2		.
BAQ	5	8	.	3	16		.
BBL	7	184	.	331	522		7
BRF	.	.	80	89	169		23
CHA	10	3,790	.	1,016	4,817		62
DNR	.	.	.	1	1		.
HER	1	1,314	.	343	1,658		24
HOL	2,745	12,721	3,852	436	19,754		264
IRM	.	93	.	.	93		.
JER	3	.	4	81	88		1
KER	.	.	.	1	1		.
LIM	119	3,983	.	1,030	5,132		68
MON	.	.	1	33	34		.
MRY	.	.	4	11	15		.
MSH	.	31	.	3	34		.
NWR	2	.	1	18	21		.
PAR	.	77	.	1	78		.
PIE	.	2	.	.	2		.
RBT	.	.	.	1	1		.
SAL	2	12	.	2	16		1
SIM	10	784	.	400	1,194		24
SWR	.	.	.	1	1		.

Breed abbreviations: AAN (Angus), AUB (Aubrac), AYS (Ayshire), BAQ (Blonde d'Aquitaine), BBL (Belgian Blue), BRF (Friesian), CHA (Charolais), DNR (Danish Red), HER (Hereford), HOL (Holstein), IRM (Irish Maol), JER (Jersey), KER (Kerry), LIM (Limousin), MON (Montbeliarde), MRY (Murray Grey), MSH (Milking Shorthorn), NWR (Norwegian Red), PAR (Parthenaise), PIE (Piedmontese), RBT (Rotbunte), SAL (Saler), SIM (Simmental), SWR (Swedish Red)

## Results and Discussion

**Genetic Disease and Major Gene Frequencies.** Analysis of validated disease and major gene SNP on the IDB identified the prevalence of carriers in the Irish Holstein-Friesian population of the lethal recessive diseases, brachyspina (2%) and complex vertebral malformation (CVM, 4%); congenital disorders citrullinaemia, osteopetrosis and syndactyly carriers were < 1%; while homozygous carriers were detected for A2  $\beta$ -casein (33%), K-casein (12%) and DGAT1 (15%). Myostatin gene variations, nt821, F94L and Q204X were

identified to be segregating in Irish Angus, Belgian Blue, Charolais, Limousin and Simmental populations. The *KRT71* causative mutation for hypotrichosis (hairlessness) was present in 2% of Irish Hereford cattle. Finally, no deficiency of uridine monophosphate synthase (DUMPS) or Curly Calf carriers were identified in Irish cattle.

**MS Imputation.** Since March, 2013 over 24,500 Irish cattle have been genotyped on the IDB SNP chip. Overall 55% of the animals had their pedigree verified using SNP data with the remaining having MS imputed profiles used. Early on it was decided that every animal with an IDB genotype would have its MS imputed as this was operationally easier and does not affect the MS imputation accuracy.

While initially all imputed MS were used for parental verification, *TGLA53* was excluded from this list in October, 2013. Analysis of its performance revealed that 1/3 of single MS misconcordances were due to the *TGLA53*. Of the animals that were directly MS genotyped because they had 2 MS misconcordances, 40% of them would have had only 1 misconcordance if *TGLA53* was excluded. Due to this, and because traditionally Weatherbys did not use *TGLA53* in their standard MS parentage testing, the decision was made to keep imputing the MS marker allele, but not to use it for parentage verification.

On average 280 new animals were added to the MS reference data primarily on a monthly basis (Table 3). Currently >1,100 animals have been added to the original publically available reference data set. With the increased reference data set, adjusted SNP haplotype sizes, and exclusion of *TGLA53* the imputation of MS allele has consistently been >98% accurate since November 2013 (Figure 1).

**Parentage SNP Imputation.** Imputation accuracy of parentage SNP genotypes (ISAG core or 2K) for pedigree verification was >90% for all offspring groups when imputing to the 3K or LD density, while it was  $\leq 40\%$  when imputing to just the 2K density. When >20 progeny was used the imputation of the ISAG core parentage SNP was >99% for 3K and LD densities, while >30 progeny were needed for >99% accuracy of the 2K parentage SNP (Table 4 and Figure 2).



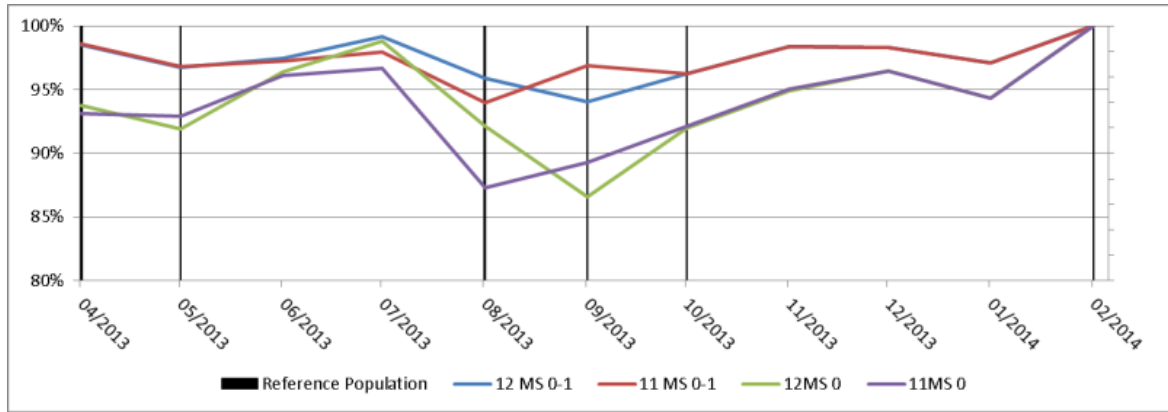


Figure 1. Accuracies of MS imputations by date of imputation and if 12 or 11 MS markers were used for parentage verification. The blue and red lines are accuracies when up to 1 MS misdiscordance is allowed, while the green and purple lines have 0 misdiscordances allowed. Black bars show when MS reference population were updated.

Table 3. Count of animals in MS imputation reference population by marker and date.

Marker	Original	5/2013	8/2013	9/2013	10/2013	2/2014
<i>BM1824</i>	7,035	7,314	7,477	7,665	7,821	8,157
<i>BM2113</i>	7,069	7,345	7,510	7,699	7,855	8,190
<i>INRA023</i>	7,006	7,287	7,451	7,641	7,797	8,130
<i>ETH10</i>	6,875	7,153	7,316	7,506	7,662	8,000
<i>ETH225</i>	7,055	7,332	7,497	7,686	7,842	8,174
<i>SPS115</i>	6,971	7,237	7,399	7,588	7,744	8,074
<i>TGLA53</i>	2,902	3,137	3,237	3,381	3,526	3,644
<i>TGLA227</i>	7,053	7,329	7,492	7,677	7,833	8,166
<i>ETH3</i>	3,850	4,129	4,293	4,482	4,638	4,972
<i>TGLA126</i>	6,808	7,078	7,243	7,433	7,589	7,925
<i>TGLA122</i>	7,056	7,334	7,497	7,687	7,843	8,171
<i>BM1818</i>	1,463	1,578	1,741	1,927	2,083	2,442
Average	5,929	6,188	6,346	6,531	6,686	7,004
Total	7,130	7,411	7,576	7,706	7,861	8,258

Table 4. Accuracy of parentage SNP imputation.

Parentage SNP <sup>b</sup>	Progeny <sup>c</sup>	2K <sup>a</sup>			3K			LD		
		Ave <sup>d</sup>	STD <sup>e</sup>	Wrong <sup>f</sup>	Ave	STD	Wrong	Ave	STD	Wrong
2K	8 to 9	0.3657	0.0144	1242	0.8932	0.0274	209	0.9250	0.0214	147
	10 to 14	0.3576	0.0159	1258	0.9339	0.0251	129	0.9502	0.0190	97
	15 to 19	0.3551	0.0121	1263	0.9692	0.0113	60	0.9749	0.0088	49
	20 to 29	0.3536	0.0111	1266	0.9866	0.0060	26	0.9896	0.0043	20
	30 to 39	0.3566	0.0195	1260	0.9942	0.0048	11	0.9953	0.0039	9
	40 to 59	0.3565	0.0134	1260	0.9970	0.0020	6	0.9978	0.0014	4
	60 to 99	0.3589	0.0129	1255	0.9962	0.0028	7	0.9977	0.0013	5
	100+	0.3566	0.0107	1260	0.9988	0.0012	2	0.9991	0.0007	2
ISAG core	8 to 9	0.4018	0.0484	60	0.8978	0.0444	10	0.9279	0.0376	7
	10 to 14	0.3924	0.0502	61	0.9431	0.0352	6	0.9594	0.0276	4
	15 to 19	0.3991	0.0549	60	0.9700	0.0199	3	0.9760	0.0176	2
	20 to 29	0.3920	0.0570	61	0.9903	0.0094	1	0.9929	0.0081	1
	30 to 39	0.3838	0.0584	62	0.9948	0.0082	1	0.9957	0.0088	0
	40 to 59	0.3811	0.0431	62	0.9981	0.0039	0	0.9981	0.0039	0
	60 to 99	0.3888	0.0590	61	0.9954	0.0063	0	0.9973	0.0052	0
	100+	0.3665	0.0567	63	0.9988	0.0032	0	0.9992	0.0027	0

a) SNP density was used for imputation

b) Parentage SNP density analysed for imputation accuracy

c) Number of genotyped progeny used to impute a parent's genotype profile

d) Average genotype imputation accuracy

e) Standard deviation of the imputation accuracy

f) Average number of incorrectly imputed genotypes given the accuracy and parentage SNP density

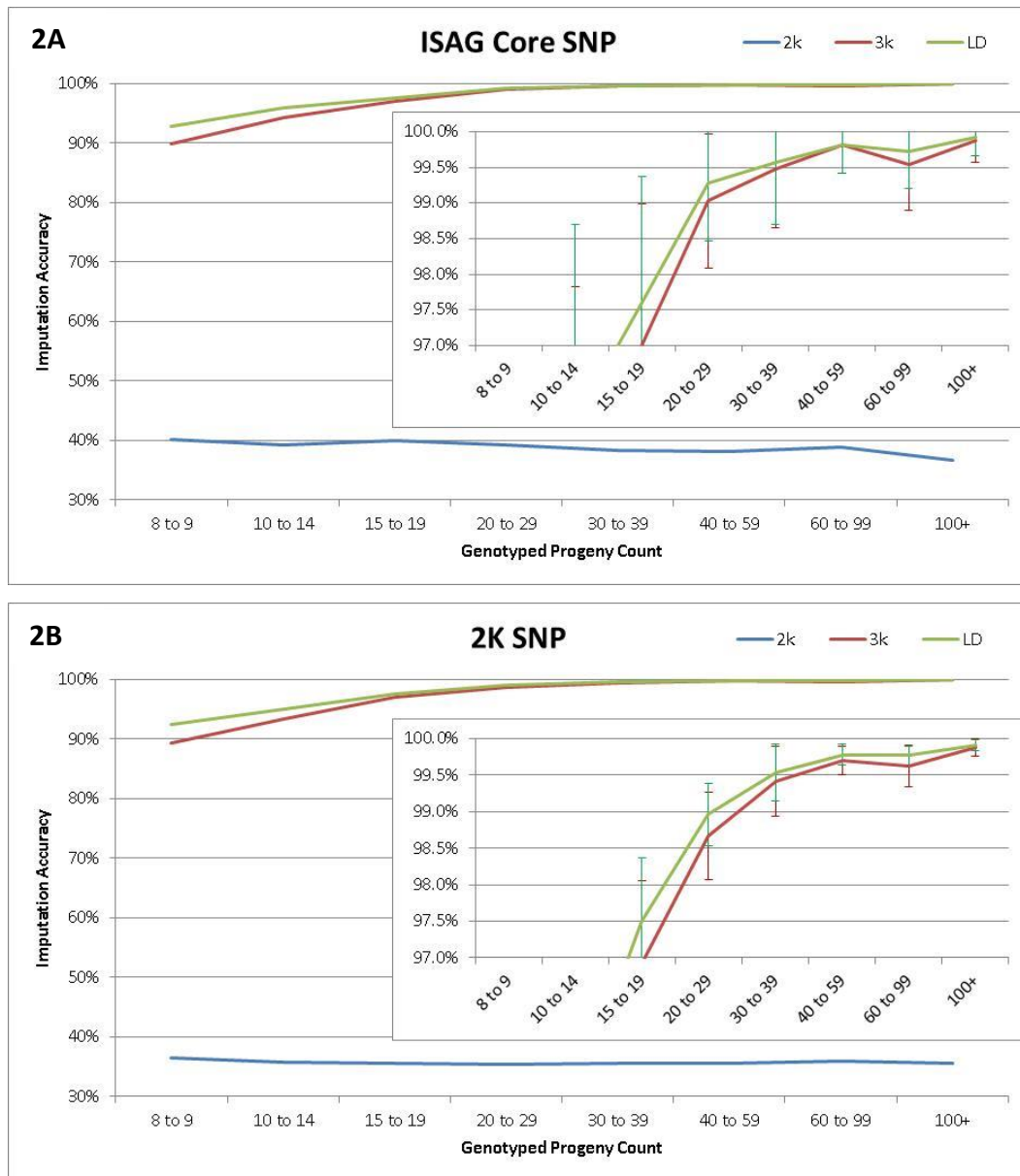


Figure 2. Imputation accuracies for parentage verification and identification SNP. 2A is the accuracy when the ISAG core and 2B is when the 2K SNP set was used to determine imputation accuracy. The lines represent which SNP density was used for imputation, 2K =blue, 3K=red, and LD=green. Imputation accuracy standard deviations are shown in the same colour as the line.

## Conclusion

The use of microsatellite imputation has resulted in considerable cost savings (>€200,000) to the Irish cattle farmer as the industry moves to SNP based technologies. As the MS reference data set increases and SNP haplotype sizes are reanalysed the MS imputation accuracy will increase and will also allow for the capture of rare MS-SNP haplotypes. The imputation has been proved to be highly accurate (>98%) across multiple beef and dairy breeds and is now a routine procedure for parentage verification in Ireland.

Our results show that parentage SNP imputation at >99% accuracy is also possible given that one has >20 verified progeny that have been genotyped on the LD or a higher density chip (Table 4 and Figure 2). If one only wants to perform sire verification with the ISAG core parentage SNP then using >20 progeny genotyped at 3K or higher density will allow you to impute the sire's parentage SNP at >99% accuracy and have  $\leq 1$  incorrectly imputed SNP genotype. For performing sire prediction, using >40 progeny genotyped at LD or higher density will allow one to impute the sire's 2K parentage SNP at >99% accuracy and have  $\leq 4$  incorrectly imputed SNP genotype.

Using both MS and parentage SNP imputation via a 2 step process will greatly accelerate the cattle industry's movement from MS- to SNP-based parentage verification. Currently 3 chip types contain the necessary SNP for MS imputation: the IDB, the GeneSeek Genomic Profiler (GGP, (Wiggans et al., 2013)), and Illumina's HD SNP chip. One can genotype a calf with one of those 3 chips, impute its MS profile, and then pedigree verify the animal with SNP or MS depending on what genotype its parents have. For sires that only have a MS profile they can have their ISAG core SNP imputed once >20 parentage verified offspring have been IDB, GGP, or HD genotyped. From that point on any future progeny from the sire can be verified using the ISAG core SNP.

Recently an additional 100 SNP have been recommend to be added to the ISAG bovine core SNP parentage panel to increase the power of exclusion to 0.9999999 ([http://www.isag.us/Docs/1\\_Workshop\\_report\\_2012-CMMPT.pdf](http://www.isag.us/Docs/1_Workshop_report_2012-CMMPT.pdf)). While those additional SNP are present on the LD/IDB/50K/HD beadchip they were left off the imputation list in this study by mistake. Additional analysis will be performed to determine the imputation accuracy of these additional parentage SNP, and to determine the imputation accuracy of a parent's SNP from progeny when imputing to IDB, 50K, or HD density.

## Literature Cited

- ALLEN, A., TAYLOR, M., MCKEOWN, B., CURRY, A., LAVERY, J., MITCHELL, A., HARTSHORNE, D., FRIES, R. & SKUCE, R. 2010. Compilation of a panel of informative single nucleotide polymorphisms for bovine identification in the Northern Irish cattle population. *BMC genetics*, 11, 5.
- BARUCH, E. & WELLER, J. I. 2008. Estimation of the number of SNP genetic markers required for parentage verification. *Animal Genetics*, 39, 474-479.
- BERRY, D. P., MCPARLAND, S., KEARNEY, J. F., SARGOLZAEI, M. & MULLEN, M. P. 2014. Imputation of ungenotyped parental genotypes in dairy and beef cattle from progeny genotypes. *Animal*, doi:10.1017/S1751731114000883.
- BONIN, A., BELLEMAIN, E., BRONKEN EIDENSEN, P., POMPANON, F., BROCHMANN, C. & TABERLET, P. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular ecology*, 13, 3261-73.

- BROWNING, S. R. & BROWNING, B. L. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, 81, 1084-97.
- FERNÁNDEZ, M. E., GOSZCZYNSKI, D. E., LIRÓN, J. P., VILLEGAS-CASTAGNASSO, E. E., CARINO, M. H., RIPOLI, M. V., ROGBERG-MUÑOZ, A., POSIK, D. M., PERAL-GARCÍA, P. & GIOVAMBATTISTA, G. 2013. Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genetics and Molecular Biology*, 36, 185-191.
- HEATON, M. P., HARHAY, G. P., BENNETT, G. L., STONE, R. T., GROSSE, W. M., CASAS, E., KEELE, J. W., SMITH, T. P. L., CHITKO-MCKOWN, C. G. & LAEGREID, W. W. 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome*, 13, 272-281.
- ILLUMINA INC. 2010. *BovineHD Genotyping Beadchip* [Online]. Available: [http://www.illumina.com/Documents/products/datasheets/datasheet\\_bovineHD.pdf](http://www.illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf) [Accessed 6/12/2011 2011].
- ILLUMINA INC. 2011a. *BovineLD Genotyping BeadChip* [Online]. Available: [http://res.illumina.com/documents/products/datasheets/datasheet\\_bovineld.pdf](http://res.illumina.com/documents/products/datasheets/datasheet_bovineld.pdf) [Accessed August 31, 2012].
- ILLUMINA INC. 2011b. *GoldenGate Bovine3K Genotyping BeadChip* [Online]. Available: [http://www.illumina.com/Documents/products/datasheets/datasheet\\_bovine3K.pdf](http://www.illumina.com/Documents/products/datasheets/datasheet_bovine3K.pdf) [Accessed 5/16/2012].
- MATUKUMALLI, L. K., LAWLEY, C. T., SCHNABEL, R. D., TAYLOR, J. F., ALLAN, M. F., HEATON, M. P., O'CONNELL, J., MOORE, S. S., SMITH, T. P., SONSTEGARD, T. S. & VAN TASSELL, C. P. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*, 4, e5350.
- MCCLURE, M. C., SONSTEGARD, T. S., WIGGANS, G. R., VAN EENENNAAM, A. L., WEBER, K. L., PENEDO, M. C. T., BERRY, D., FLYNN, J., GARCIA, J. F., CARMO, A. S., REGITANO, L. C. A., ALBUQUERQUE, M., SILVA, M. V. G. B., MACHADO, M. A., COFFEY, M., MOORE, K., BOSCHER, M.-Y., GENESTOUT, L., MAZZA, R., TAYLOR, J. F., SCHNABEL, R. D., SIMPSON, B., MARQUES, E., MCEWAN, J., CROMIE, A., COUTINHO, L. L., KUEHN, L., KEELE, J., PIPER, E., COOK, J., WILLIAMS, R. & VAN TASSELL, C. 2013. Imputation of Microsatellite Alleles from Dense SNP Genotypes for Parentage Verification Across Multiple *Bos taurus* and *Bos indicus* breeds. *Frontiers in Genetics*, 4, 1-11.
- MULLEN, M. P., MCCLURE, M. C., KEARNEY, J. F., WATERS, S. M., WELD, R., FLYNN, P., CREEVEY, C. J., CROMIE, A. R. & BERRY, D. P. Year. Development of a custom SNP chip for dairy and beef cattle breeding, parentage, and research. *In: INTERBULL*, Augustt 23-25, 2013 2013 Nantes, Franch. Bullentin No. 47.
- SARGOLZAEI, M., CHESNAIS, J. & SCHENKEL, F. 2011. FImpute-An efficient imputation algorithm for dairy cattle populations. *J. Dairy. Sci*, 94, 421. (Abstr.).
- WELLER, J. I., FELDMESSER, E., GOLIK, M., TAGER-COHEN, I., DOMOCHOVSKY, R., ALUS, O., EZRA, E. & RON, M. 2004. Factors Affecting Incorrect Paternity Assignment in the Israeli Holstein Population. *Journal of Dairy Science*, 87, 2627-2640.
- WIGGANS, G. R., COOPER, T. A., VAN TASSELL, C. P., SONSTEGARD, T. S. & SIMPSON, E. B. 2013. Technical note: Characteristics and use of the Illumina BovineLD and GeneSeek Genomic Profiler low-density bead chips for genomic evaluation. *Journal of Dairy Science*, 96, 1258-63.
- ZIMIN, A. V., DELCHER, A. L., FLOREA, L., KELLEY, D. R., SCHATZ, M. C., PUIU, D., HANRAHAN, F., PERTEA, G., VAN TASSELL, C. P., SONSTEGARD, T. S., MARCAIS, G., ROBERTS, M., SUBRAMANIAN, P., YORKE, J. A. & SALZBERG, S. L. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biology*, 10, R42.