



IRISH CATTLE BREEDING FEDERATION

SNP Selection for Nationwide Parentage Verification and Identification in Beef and Dairy Cattle



M.C. McClure, J. McCarthy, P. Flynn, R. Weld, M. Keane, K. O'Connell,
M.P. Mullen, S. Waters, J.F. Kearney

ICAR Technical Workshop,
Krakow, Poland, 9th June, 2015

IDB SNP chip (International Dairy and Beef)

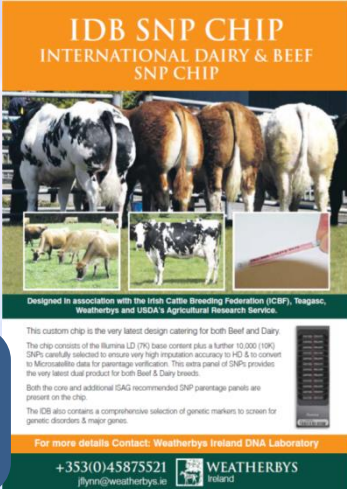
Parentage Verification / Prediction

Microsatellite Imputation

Sex Verification

Genomic Breeding Value

Disease/Major Gene Status



IDB SNP CHIP
INTERNATIONAL DAIRY & BEEF
SNP CHIP

Designed in association with the Irish Cattle Breeding Federation (ICBF), Teagasc, Weatherbys and USDA's Agricultural Research Service.

This custom chip is the very latest design catering for both Beef and Dairy. The chip consists of the Illumina LD™ (low density) core plus a further 10,000 (10K) SNPs carefully selected to ensure very high imputation accuracy to HD & to convert to Microsatellite data for parentage verification. This extra panel of SNPs provides the very latest dual product for both Beef & Dairy breeds.

Both the core and additional 10K recommended SNP parentage panels are present on the chip.

The IDB also contains a comprehensive selection of genetic markers to screen for genetic disorders & major genes.

For more details Contact: Weatherbys Ireland DNA Laboratory
+353(0)45875521
jlynn@weatherbys.ie

WEATHERBYS
Ireland

illumina
SERVED PROGRESS

ICBF
www.icbf.com

capac
www.capac.com

ISAG
www.isag.com

Lethal recessives

- 1 DM1 (Polyomavirus Infection)
- 2 DM2
- 3 BSE/Variant Creutzfeldt-Jakob
- 4 BLAD

Congenital disorders

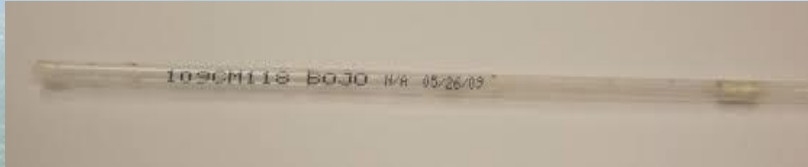
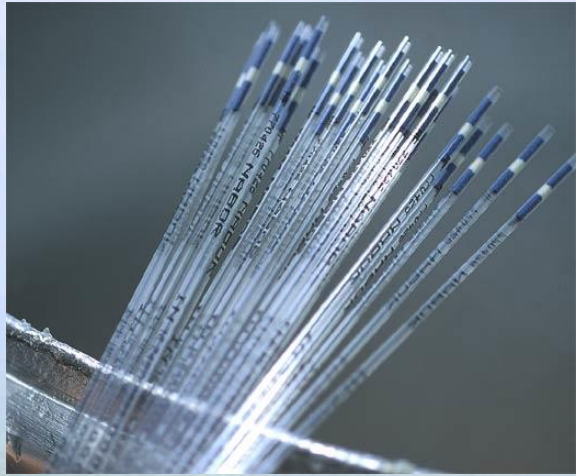
- 1 Bovine Spongiform Encephalitis
- 2 Fawn Coat Syndrome or Contractural Alopecia/Scaly*
- 3 Hypertrophic Osteodystrophy
- 4 Hypertrophic Osteodystrophy, HCDK & DAP
- 5 Hypertrophic Osteodystrophy
- 6 SpinaLink (COL3A1 gene, Bovine Spina)
- 7 SpinaLink (COL3A1 gene, Bovine Spina)
- 8 Polystyrene
- 9 Milk Fever
- 10 Toxicoinfection (EPEC)
- 11 Black/Red Coat Crossed Factor
- 12 Blue Scurf (Blue Scurf) (Collagen VI gene defect)
- 13 Blue Coat (BLU)
- 14 Blue Coat (BLU)
- 15 Blue Coat (BLU)
- 16 Blue Coat (BLU)
- 17 Blue Coat (BLU)
- 18 Blue Coat (BLU)
- 19 Blue Coat (BLU)
- 20 Blue Coat (BLU)
- 21 Blue Coat (BLU)
- 22 Blue Coat (BLU)
- 23 Blue Coat (BLU)
- 24 Blue Coat (BLU)
- 25 Blue Coat (BLU)
- 26 Blue Coat (BLU)
- 27 Blue Coat (BLU)
- 28 Blue Coat (BLU)
- 29 Blue Coat (BLU)
- 30 Blue Coat (BLU)
- 31 Blue Coat (BLU)
- 32 Blue Coat (BLU)

Major genes

- 1 DM1
- 2 DM2
- 3 A1/A2 beta casein
- 4 Kappa Casein (KAP)
- 5 Kappa Casein (KAP)
- 6 Kappa Casein (KAP)
- 7 A1/A2
- 8 GDF14 and GDF15 (Mating, growth, meat)
- 9 IGF1
- 10 IGF1
- 11 IGF1
- 12 IGF1
- 13 IGF1
- 14 IGF1
- 15 IGF1
- 16 IGF1
- 17 IGF1
- 18 IGF1
- 19 IGF1
- 20 IGF1
- 21 IGF1
- 22 IGF1
- 23 IGF1
- 24 IGF1
- 25 IGF1
- 26 IGF1
- 27 IGF1
- 28 IGF1
- 29 IGF1
- 30 IGF1
- 31 IGF1
- 32 IGF1

* Usually seen only aspy

Parentage Validation/Prediction



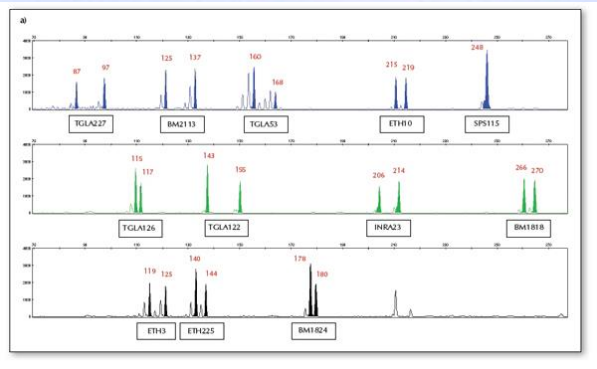
Parentage Validation

How many markers are needed?

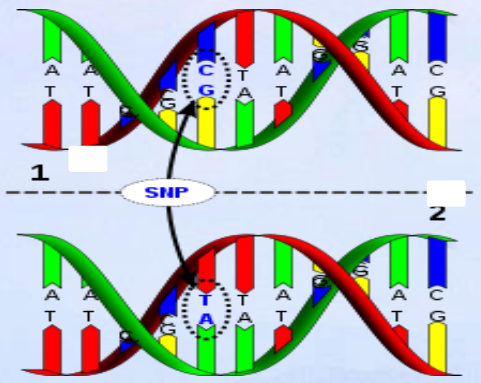
Recipient's blood			Reactions with donor's red blood cells			
ABO antigens	ABO antibodies	ABO blood type	Donor type O cells	Donor type A cells	Donor type B cells	Donor type AB cells
None	Anti-A Anti-B	O				
A	Anti-B	A				
B	Anti-A	B				
A & B	None	AB				

Compatible Not compatible

Blood Typing



Microsatellite Markers



SNPs

Microsatellites

ISAG Panel: BM1824, BM2113, INRA023, SPS115, TGLA122, TGLA126, TGLA227, ETH10, ETH225, ETH3, SPS115, TGLA53

Additional panel of MGTG4B, CSRM60, SPS113, ILSTS006, RM067, CSSM66

SNP: ISAG100
ISAG200
More???

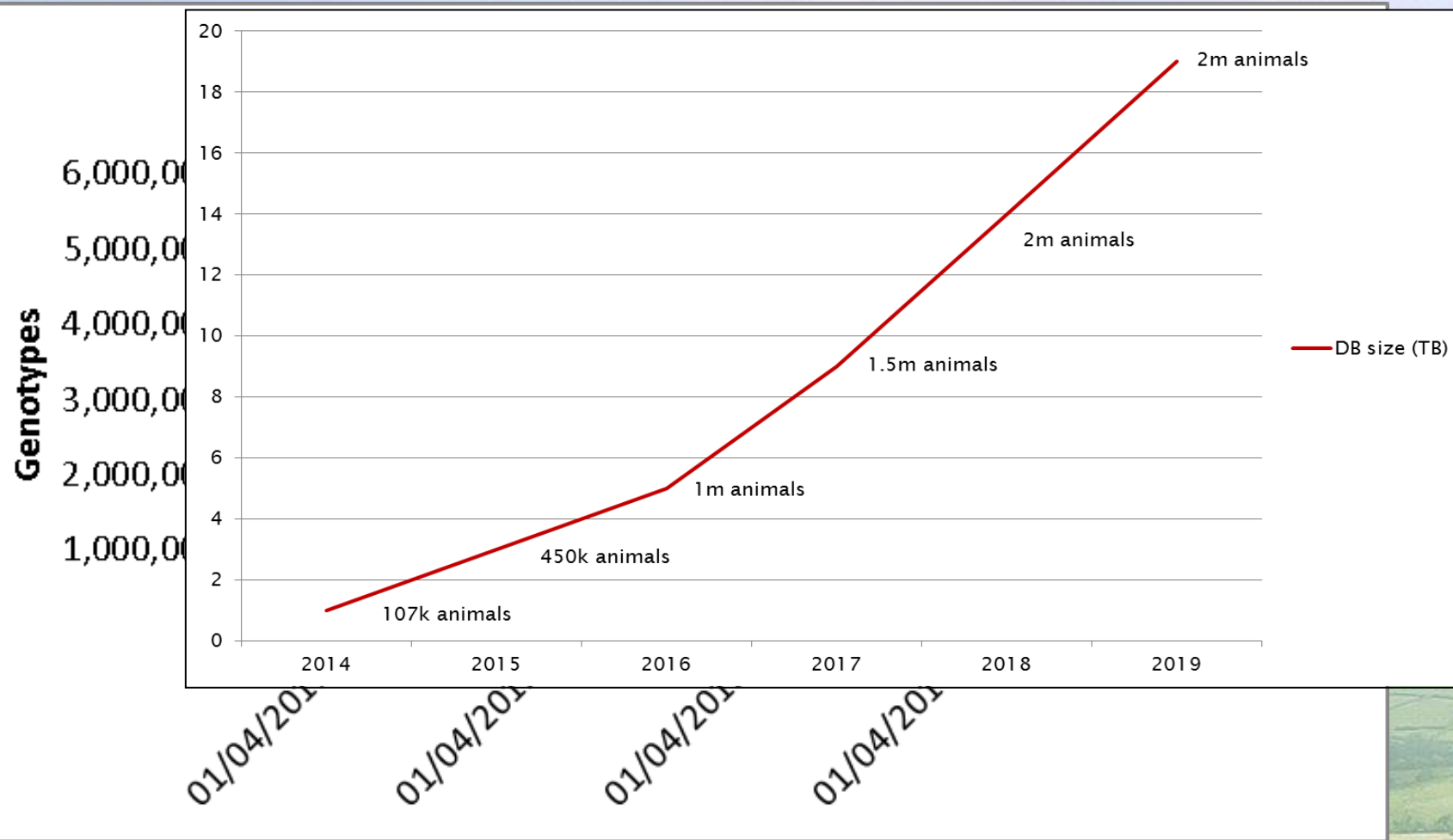
121 Validation



How Many SNP to Validate 1 Sire



ICBF Actual/Predicted Growth



Small SNP Sets: Genotype Errors

- Fail Sire A with 121 SNP (ISAG200 subset)
 - ≥ 2 misconcordance = fail
- Predict Sire A with 2,000 SNP
 - ≤ 10 misconcordance = pass



Why validate with small and predict large SNP?

Map	itt	sire_itt	ARS-USMARC-PARENT-EF093511-RS2901
1	HERAUSM00000GRKA584	1	B_B
2	HERIRLM141607070601	HERAUSM00000GRKA584	A_A
3	HERIRLM151225330379	HERAUSM00000GRKA584	A_A
4	HERIRLM151225340413	HERAUSM00000GRKA584	A_A
5	HERIRLM151225360423	HERAUSM00000GRKA584	A_A
6	HERIRLM151225370399	HERAUSM00000GRKA584	A_A
7	HERIRLM151225380425	HERAUSM00000GRKA584	A_A
8	HERIRLM151225380441	HERAUSM00000GRKA584	A_A
9	HERIRLM191141760568	HERAUSM00000GRKA584	A_A
10	HERIRLM231305310207	HERAUSM00000GRKA584	A_A
11	HERIRLM241245190385	HERAUSM00000GRKA584	A_A
12	HERIRLM331493150493	HERAUSM00000GRKA584	A_A
13	HERIRLM361055830685	HERAUSM00000GRKA584	A_A
14	HERIRLM361055880665	HERAUSM00000GRKA584	A_A
15	HERIRLM121182590084	HERAUSM00000GRKA584	A_B
16	HERIRLM121201450915	HERAUSM00000GRKA584	A_B
17	HERIRLM121201460899	HERAUSM00000GRKA584	A_B
18	HERIRLM141643520372	HERAUSM00000GRKA584	A_B
19	HERIRLM151225310435	HERAUSM00000GRKA584	A_B
20	HERIRLM151225330412	HERAUSM00000GRKA584	A_B
21	HERIRLM151225390376	HERAUSM00000GRKA584	A_B
22	HERIRLM151313160329	HERAUSM00000GRKA584	A_B
23	HERIRLM151654470902	HERAUSM00000GRKA584	A_B
24	HERIRLM211463520049	HERAUSM00000GRKA584	A_B
25	HERIRLM211463580046	HERAUSM00000GRKA584	A_B
26	HERIRLM231305370204	HERAUSM00000GRKA584	A_B
27	HERIRLM241842660174	HERAUSM00000GRKA584	A_B
28	HERIRLM281110080573	HERAUSM00000GRKA584	A_B
29	HERIRLM291166240641	HERAUSM00000GRKA584	A_B
30	HERIRLM301207290630	HERAUSM00000GRKA584	A_B
31	HERIRLM311122080052	HERAUSM00000GRKA584	A_B
32	HERIRLM311317670240	HERAUSM00000GRKA584	A_B
33	HERIRLM331340371248	HERAUSM00000GRKA584	A_B
34	HERIRLM331358460991	HERAUSM00000GRKA584	A_B
35	HERIRLM121182560081	HERAUSM00000GRKA584	B_B
36	HERIRLM141607090603	HERAUSM00000GRKA584	B_B
37	HERIRLM151225340421	HERAUSM00000GRKA584	B_B
38	HERIRLM151225370424	HERAUSM00000GRKA584	B_B
39	HERIRLM151313130326	HERAUSM00000GRKA584	B_B

Small SNP Sets: Validate Wrong Sire

Animal	Listed sire		
LIMIRLF.....554	LIMIRLM.....050		
SNP set	#_mismatches	#SNP_both_genotyped	%_SNP_mismatch
121 SNP	0	118	0.00%
LD_base	267	6848	3.90%
CHAIRLF.....639	CHAIRLM.....163		
SNP set	#_mismatches	#SNP_both_genotyped	%_SNP_mismatch
121 SNP	1	113	0.88%
LD_base	336	6729	4.99%
CHAIRLF.....355	CHAFRAM.....993		
SNP set	#_mismatches	#SNP_both_genotyped	%_SNP_mismatch
121 SNP	1	114	0.88%
LD_base	227	6761	3.36%
LIMIRLF.....500	LIMIRLM.....959		
SNP set	#_mismatches	#SNP_both_genotyped	%_SNP_mismatch
121 SNP	0	112	0.00%
LD_base	0	6711	0.00%

Sire Validation Errors with Small SNP Sets

- Fail with 121 then predict with 2,000 SNP
- Validate the wrong sire?
- **Why validate with small and predict large SNP**



ICAR-2014: How Many SNP are used to Predict Parentage?



How Many SNP to Predict 1 Correct Sire

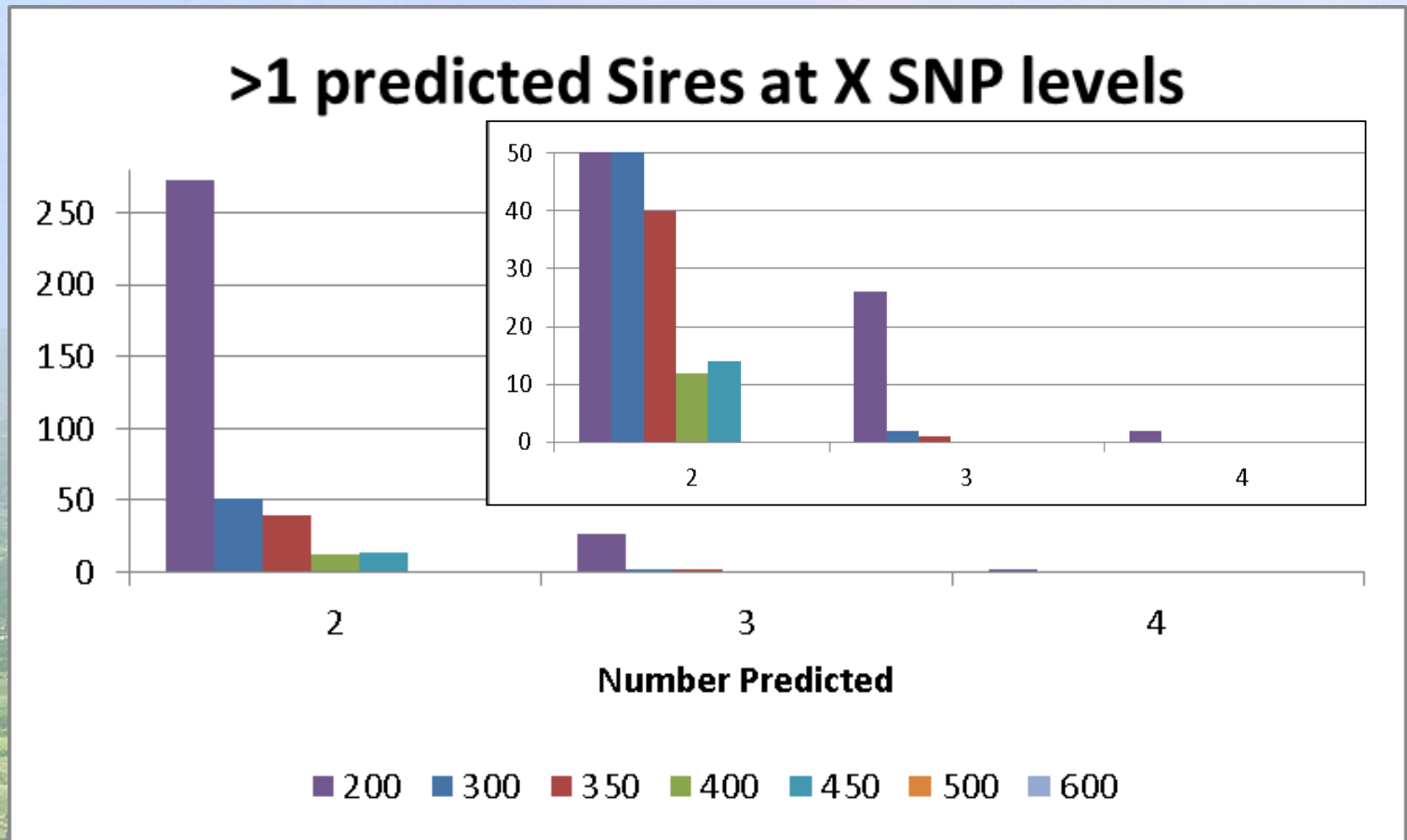


How Many SNP to Predict 1 Correct Sire

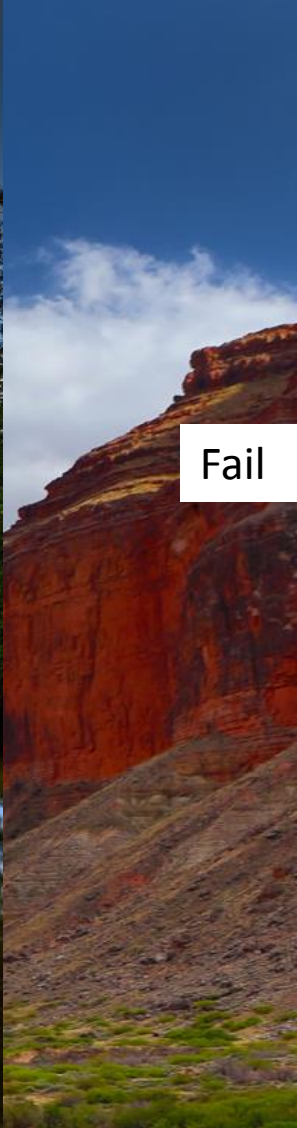
- 56,000 genotyped animals
- SNP densities tested
 - ISAG200
 - Plus top 100, 200,1800 MAF SNP
- 9,000 sire validated animals
- Predict sire/dams (1% misconcordance)

Animals	56,000
breed	Jun-14
HOL	68.65%
LIM	7.94%
CHA	7.41%
AAN	4.42%
SIM	2.35%
HER	3.06%
BBL	1.01%
MSH	0.06%
SAL	0.04%
JER	0.17%
PAR	0.17%
LMS	1.91%
BAQ	0.04%
AUB	0.00%
CHL	1.68%
PIE	0.55%
HFD	0.17%
MON	0.14%
IRM	0.00%
NWR	0.09%
RED	0.01%

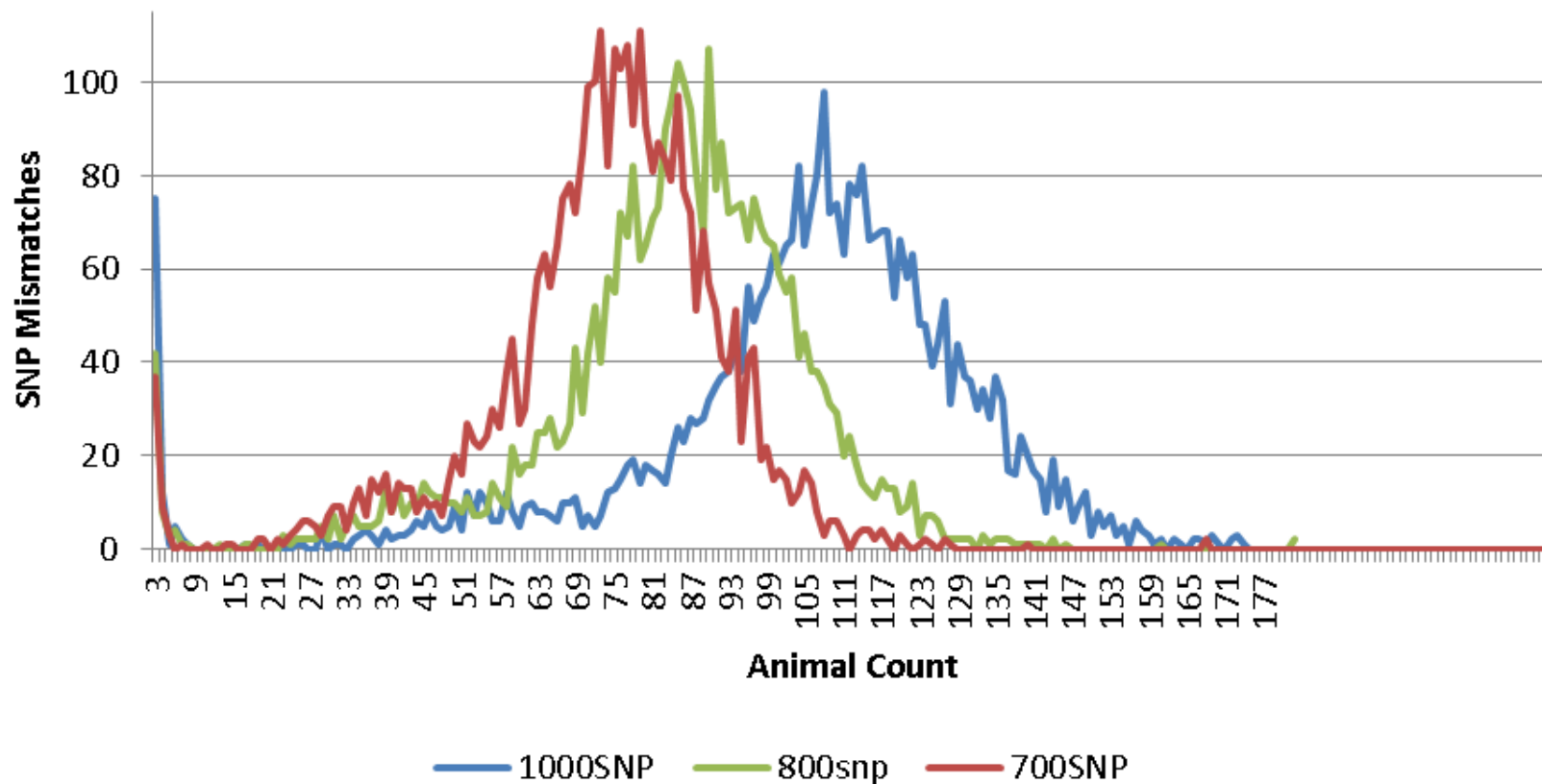
Sire Prediction with 200, 300,... SNP



Space between Validation and Fail



SNP Mismatches for sire validations 114,000 animals



Recommend: Use ≥ 500 SNP for Validation

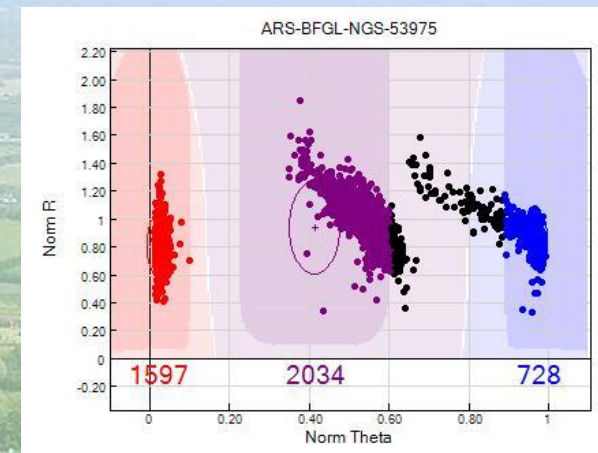
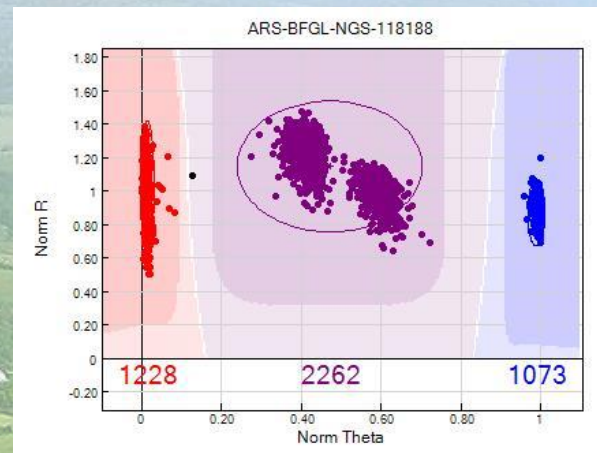
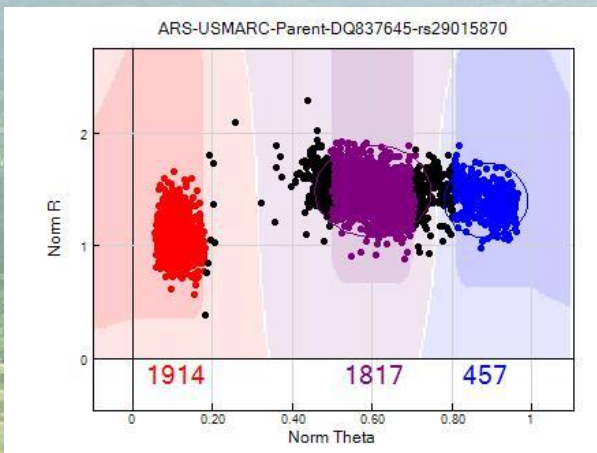
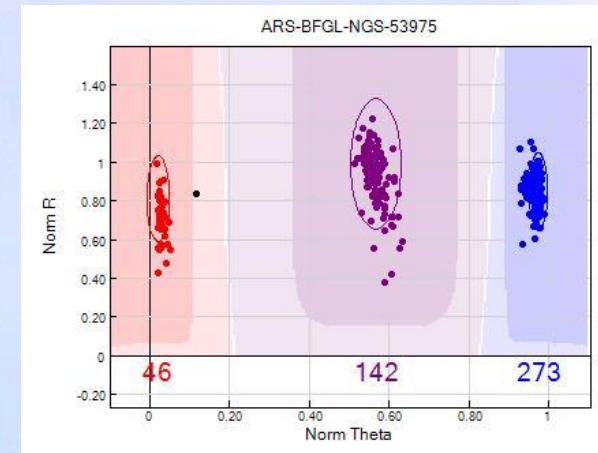
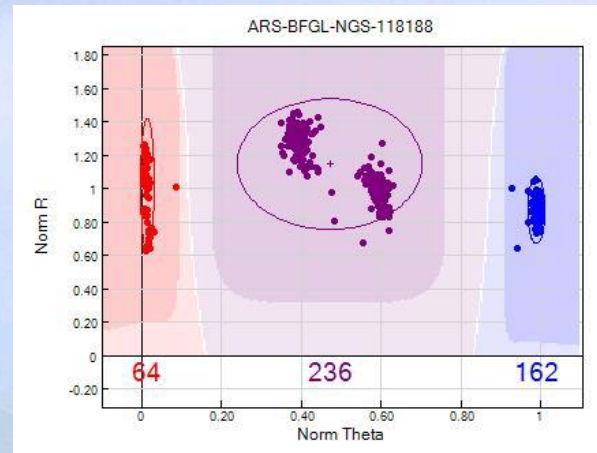
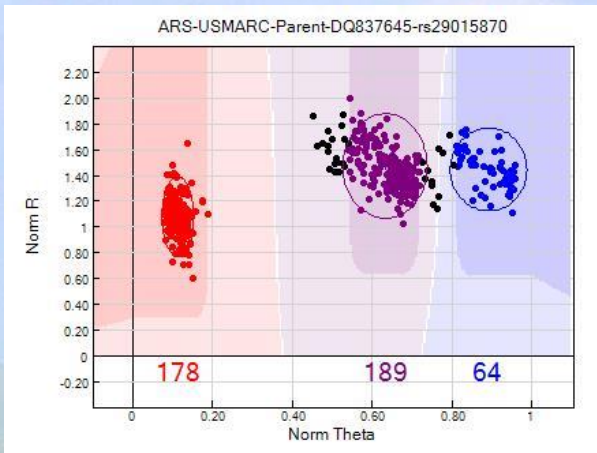
- ICBF Switched to 800 SNP July 2014
- Refinement in February 2015
 - More breeds
 - Less Dairy heavy
 - SNP QC

800 SNP list at

http://www.icbf.com/wp/wp-content/uploads/2013/07/ICBF_Parentage_SNP_Selection.csv

Animals	56,000	180,000
breed	Jun-14	Feb-15
HOL	68.65%	30.08%
LIM	7.94%	22.34%
CHA	7.41%	18.68%
AAN	4.42%	7.22%
SIM	2.35%	6.77%
HER	3.06%	4.66%
BBL	1.01%	2.75%
MSH	0.06%	1.57%
SAL	0.04%	0.92%
JER	0.17%	0.67%
PAR	0.17%	0.64%
LMS	1.91%	0.59%
BAQ	0.04%	0.59%
AUB	0.00%	0.53%
CHL	1.68%	0.52%
PIE	0.55%	0.17%
HFD	0.17%	0.07%
MON	0.14%	0.05%
IRM	0.00%	0.03%
NWR	0.09%	0.03%
RED	0.01%	0.03%

SNP Refinement QC Quality: Clusters when processing 400 vs 4K / week



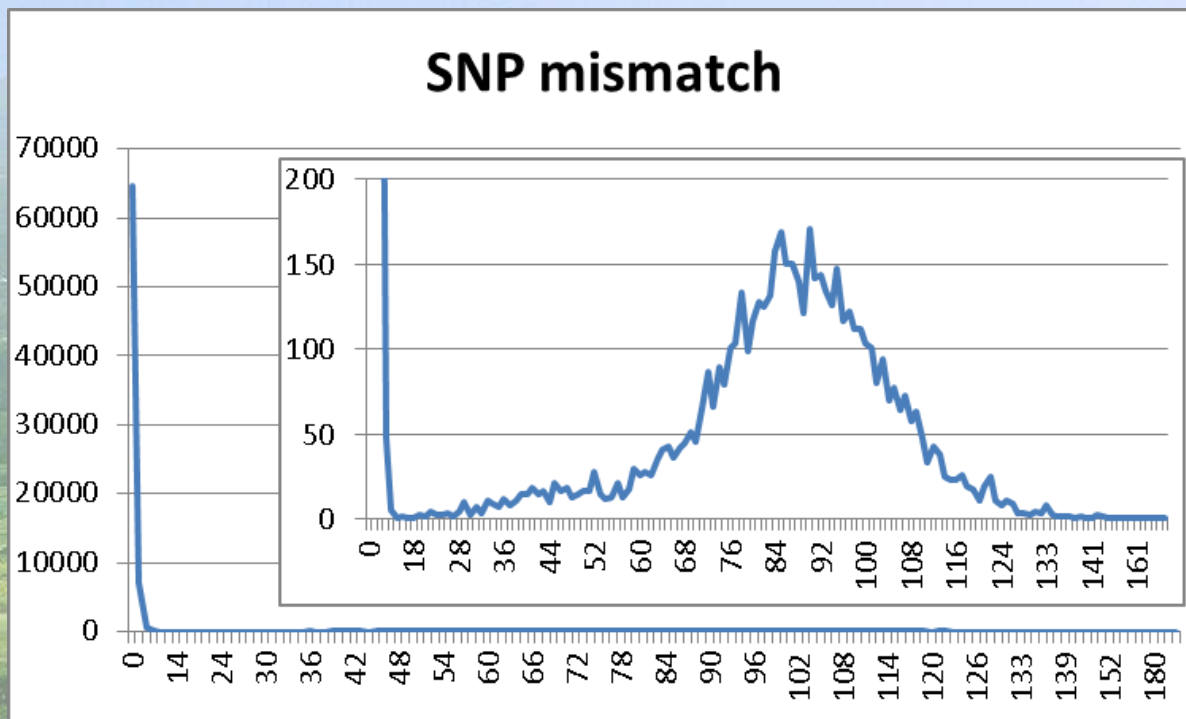
SNP Excluded from ICBF Parentage Analysis Due to

- 1) Call Rate (CR<0.3)
- 2) Minor Allele Frequency (MAF<0.01)
- 3) Probe Clustering issues

SNPID	RS#	ISAG100	ISAG200	CR	MAF	Cluster
ARS-BFGL-NGS-103099	rs110896475	0	0	0.952	0.496	1
ARS-BFGL-NGS-112652	rs109504357	0	0	0.994	0.499	1
ARS-BFGL-NGS-11469	rs111032185	0	0	0.955	0.496	1
ARS-BFGL-NGS-118188	rs42156449	0	0	0.993	0.485	1
ARS-BFGL-NGS-3547	rs110206613	0	0	1	0.493	1
ARS-BFGL-NGS-43361	rs109934313	0	0	0.998	0.494	1
ARS-BFGL-NGS-53975	rs109456438	0	0	0.988	0.484	1
ARS-BFGL-NGS-62906	rs110146023	0	0	0.998	0.484	1
ARS-BFGL-NGS-66558	rs109817790	0	0	0.991	0.495	1
ARS-BFGL-NGS-76191	rs110846944	0	1	0.998	0.385	1
ARS-USMARC-PARENT-DQ786764-NO-RS	rs109943112	1	1	0.952	<0.001	-
ARS-USMARC-PARENT-DQ837645-RS29015870	rs29015870	1	1	0.276	0.389	1
ARS-USMARC-PARENT-EF034087-NO-RS	rs110665639	1	1	0.984	0.005	-
BTA-100621-NO-RS	rs41611675	0	1	0.928	0.452	1
BTB-00147175	rs43356919	0	0	0.99	0.491	1
BTB-01834338	rs42942345	0	0	0.999	0.498	1
HAPMAP47324-BTA-55159	rs41586638	0	0	0.984	0.496	1
UA-IFASA-9571	rs41659357	0	0	0.985	0.485	1

Recommend: Use >500 SNP for Validation

- ICBF Switched to 800 SNP July 2014
- Refinement in February 2015



Animals	56,000	180,000
breed	Jun-14	Feb-15
HOL	68.65%	30.08%
LIM	7.94%	22.34%
CHA	7.41%	18.68%
AAN	4.42%	7.22%
SIM	2.35%	6.77%
HER	3.06%	4.66%
BBL	1.01%	2.75%
MSH	0.06%	1.57%
SAL	0.04%	0.92%
JER	0.17%	0.67%
PAR	0.17%	0.64%
LMS	1.91%	0.59%
BAQ	0.04%	0.59%
AUB	0.00%	0.53%
CHL	1.68%	0.52%
PIE	0.55%	0.17%
HFD	0.17%	0.07%
MON	0.14%	0.05%
IRM	0.00%	0.03%
NWR	0.09%	0.03%
RED	0.01%	0.03%

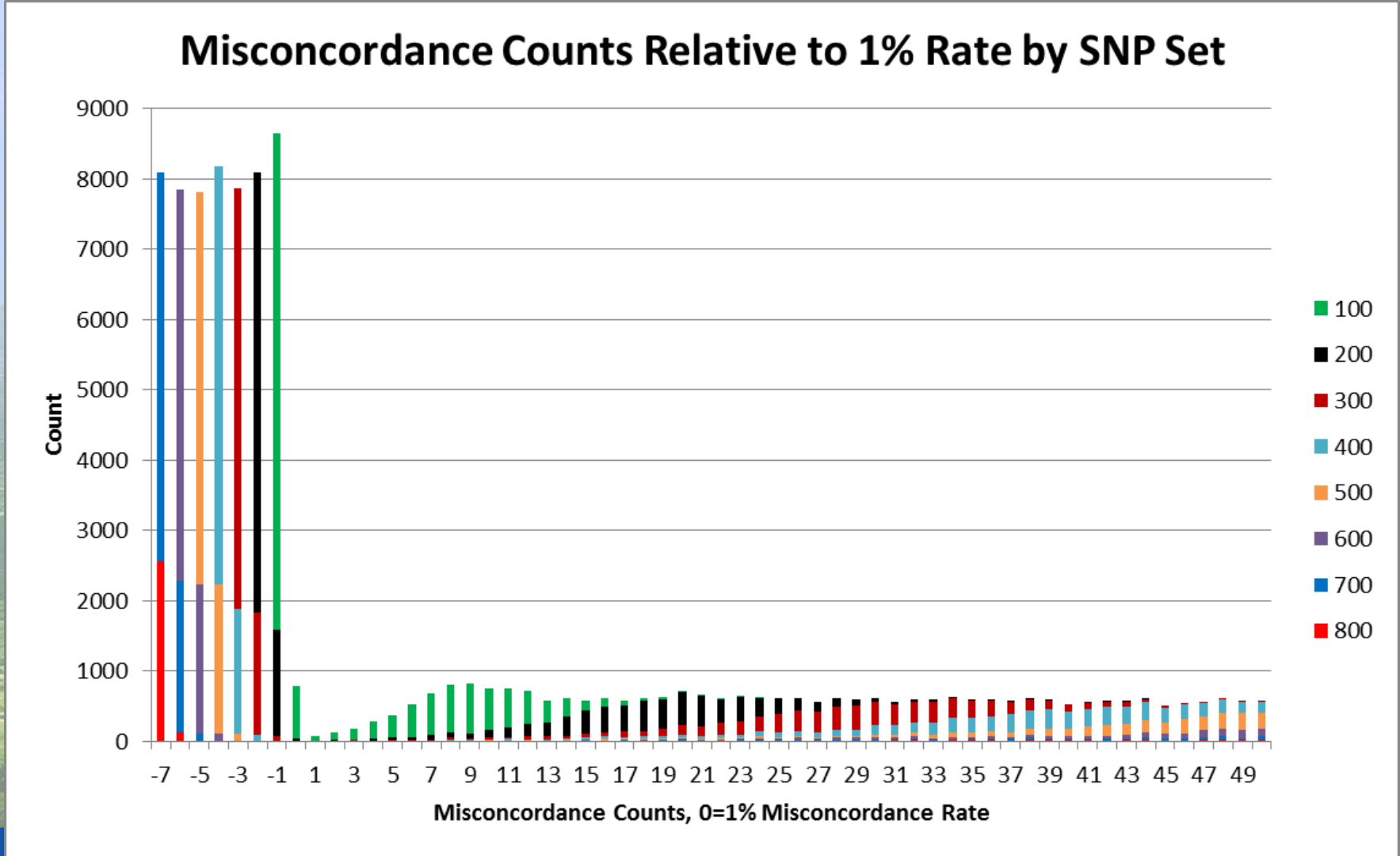


Can <800 SNP be Used

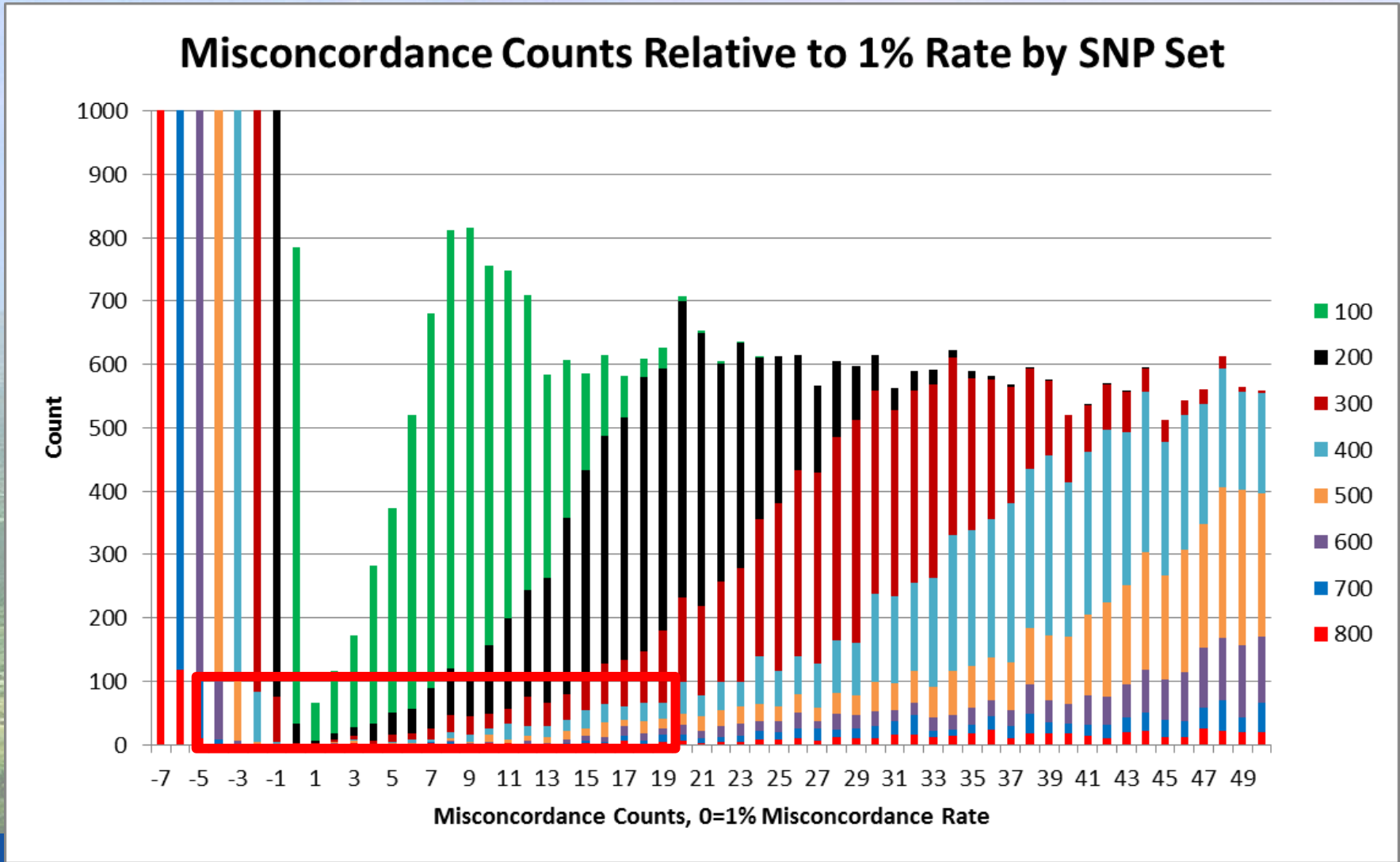
- 13,748 animals with ≥ 1 misconcordance (800 SNP)
- Reran validation process at:
700, 600, 500, 400, 300, 200, 100 SNP
- Count number of misconcordances <1% rate>
- 1% misconcordances = 0
- Accuracy
- “Gap” between pass/fail

	Fail on 800	Pass on 800
SNP#	Pass on X	Fail on X
700	-	-
600	-	-
500	-	-
400	1	-
300	-	-
200	2	-
100	17	14

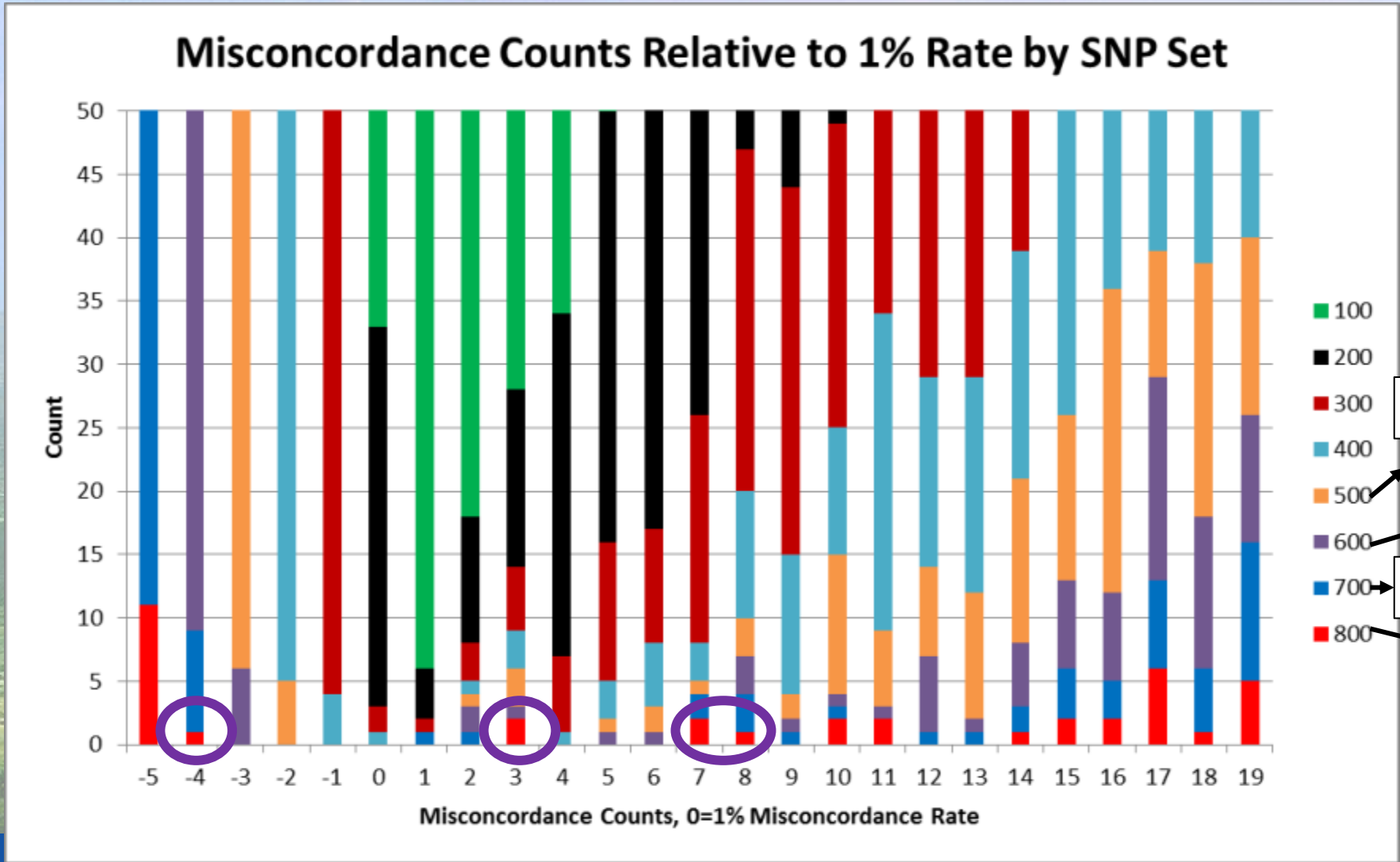
Can <800 SNP be Used



Can <800 SNP be Used



Can <800 SNP be Used



Final Thoughts

- Some ISAG parentage SNP should be dropped
- Advise using ≥ 500 SNP for parentage
- Exact number of SNP to use depends on each groups:
 - Parentage situation (farm level vs national level)
 - Acceptable risk level
 - SNP error rate for your genotyping platform(s)
 - Pass/Fail misconcordance gap they are comfortable with
- “The exact SNP don’t matter as long as you have enough quality SNP”—George Wiggans

Acknowledgements/Funding

- ICBF
- Teagsac
- Teagasc RMIS 6079
- Weatherbys
- Irish farmers and herdbooks
- DAFM RSF
 - Molecular Biomarkers of Fertility 11/S/104,
 - GenoTrace 13/F/403



Thank You



800 SNP list at

http://www.icbf.com/wp/wp-content/uploads/2013/07/ICBF_Parentage_SNP_Selection.csv