

## **Identifying the genomic basis of Johne's disease in the New Zealand dairy population.**

***Sarah Loker, Hinrich Voges, Ric Sherlock and Richard Spelman.***

### **Executive summary**

Over 1800 Johnes affected dairy animals have been identified, DNA sampled and genotyped over the past 3 years. The animals were identified from a milk based ELISA test and verified from a blood sample. Genotyping was undertaken over the Illumina high density (777K markers) marker panel. Statistical analysis has commenced with three different statistical models being employed; i) single marker association, ii) Bayesian Lasso approach that fits all the markers simultaneously. For both of these methods the control population is sampled from the other 25,000 animals that have been genotyped through the LIC genomic programme. The third method is a gene drop approach where the pedigree of the affected animal is used to calculate the expectation of the allele frequencies using the sire genotypes from the LIC genomic programme.

Statistical analyses have identified regions of the genome that are significantly different between Johnes affected group and the control. However, at this stage of the analysis there is moderate level of concordance between the different statistical approaches. Despite this there are regions that have been identified where there are interesting candidates that underlie the genomic association. We think there is another 4 months of work required to fully analyse the data set and identify and define the genomic regions of interest. We also think that we should analyse the data in another manner where we exploit the apparent greater susceptibility of the Jersey breed. This will focus on the crossbred population and identify if there are regions that are more Jersey than other regions in the affected animals. Once these analyses have concluded we think that it would be appropriate to undertake a meta-analysis with other groups that have undertaken genomic studies on dairy cattle.

### **Introduction**

JDRC full programme milestone 1: An archive of ruminant DNA related to *M. paratuberculosis* infection for genome wide association studies in sheep, cattle and deer.

During the next three years, collect and store a DNA sample of at least 200ug of purified DNA from ruminants that have been phenotyped as accurately as possible regarding their disease status including their immune response and the identity of the *M. paratuberculosis* pathogen causing the disease.

### Genomics approach for the Bovine DNA archive:

- Identify 2000 Johnes's disease positive cows using ELISA on milk herd test samples, followed by serum confirmation.
- Apply a phenotype breed definition of greater than 13/16<sup>th</sup> Holstein Friesian or Jersey.
- Use the existing LIC population data (23,000 genotypes) as control population.

- Utilise the Illumina HD and imputed 50k SNP panels to find genes for resistance and susceptibility to Johne's disease.

The timeline that was assigned to this project was:

- Year 1: Dec 08 – June 09 Pilot trial to test screening and collection process (completed).
- Year 2: June 09/10 First season large scale screening (completed).
- Year 3: June 10/11 Second season large scale screening (completed).
- Year 4: June 11/12 Genomic analysis for markers (in progress).

## **Methodology**

*Identification of affected animals:* The Animal Health Lab at LIC validated a commercial ELISA for use with milk samples (both bulk and individual cow samples) to identify Johne's disease positive cows. A screening process to efficiently identify affected cows was devised; using dairy company vat test samples (a bulk milk sample for that herd) to identify herds that had an indication of high Johnes level. These herds were then screened when they submitted their herd test samples. Ten animals were pooled together and if the pool was positive for Johnes then the individual animals were screened. The positive animals were then verified from a blood sample.

*Genotyping:* The animals were genotyped over the Illumina 777K marker panel at the GeneSeek laboratory. Animals that had a call rate less than 90% were excluded from the analysis. Additionally animals that had a heterozygosity level greater than 30% were also excluded from the analysis. Individual markers were screened for departure from Hardy-Weinberg equilibrium based on these 1843 samples and also another 1500 animals that LIC genotyped over the HD panel. Approximately 23000 cows from the NZ population for which LIC has genotypes were used to provide an appropriate control population. About 1500 of these were HD genotypes, the remainder having been generated using the Illumina 50k SNP panel.

*Imputation:* All 50k genotypes were imputed to High Density (HD) using Beagle (Browning and Browning, 2009). The reference population of HD genotypes used to impute the ungenotyped markers consisted of 2782 animals including the Johnes-positive animals. HD genotypes were also imputed to fill in any sporadic missing data. LIC has previously shown that imputation of HD genotypes from 50k has an accuracy of 99.5%.

## **Statistical methods**

A number of statistical methods have been investigated to identify the set of SNPs that is most highly associated with susceptibility to Johnes' disease.

*PLINK:* A simple genome-wide association study was carried out by comparing the allele frequencies between cases (Johnes-positive animals) and controls (23000 cows representing the the NZ dairy cow population) using PLINK 1.07. The model fitted included breed 16ths and year-of-birth.

*Bayesian Lasso:* This method differs from the simple GWAS in that it fits all the SNPs to the phenotypic data at once. Multi-SNP analysis should better apportion the total variance among the SNPs rather than attributing the same effects multiple times to individual SNPs as is possible in a single SNP analysis. The Bayesian Lasso analysis was repeated four times with a control chosen randomly but matched for breed percentage from the total set of 23000 control animals. The results of the four analyses were averaged to the final result.

*GeneDrop:* This method compares the frequency of alleles in the Johnes-positive animals to the expected allele frequencies based on knowledge of their ancestors' genotypes. This method uses the genotypes available for ancestors as the control.

## Results

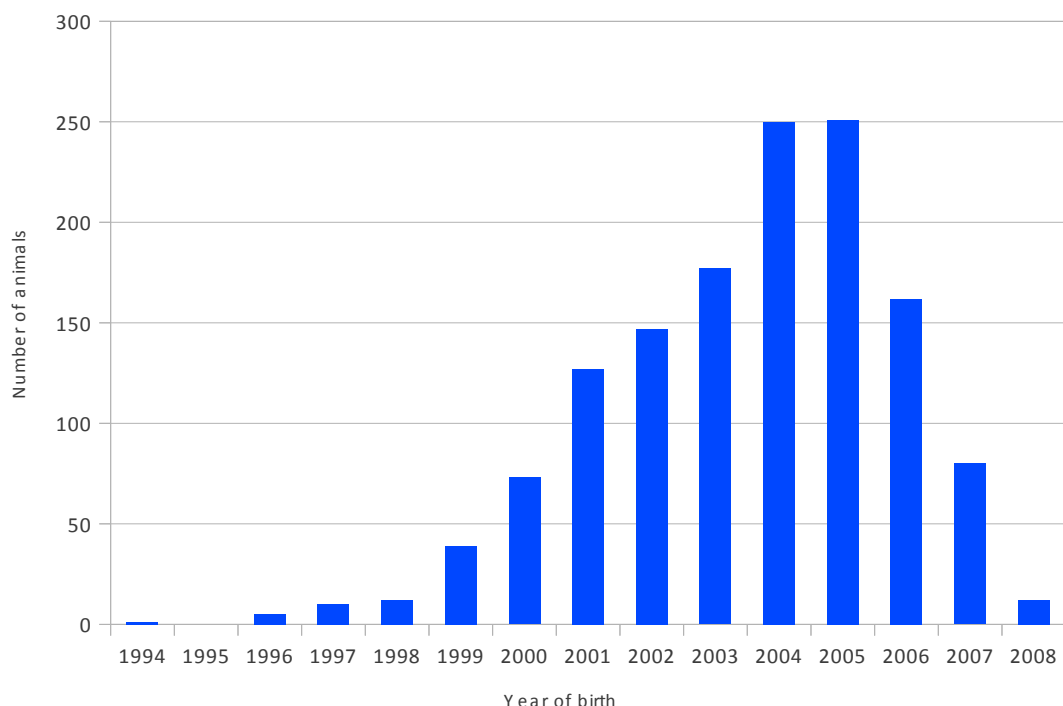
The breed proportion of the 1,842 Johnes affected cows and the 22,452 controls is shown in Table 1. There is quite a difference in the breed composition of the two groups and thus some of the statistical models have had breed included to adjust for this and other controls have been matched for breed composition.

**Table 1.** Breed composition in the Johnes and control groups.

	Johnes	Controls
% HF	33.59	65.77
% JE	65.55	33.19
% AY	0.72	0.97
% other	0.14	0.07

The Johnes affected animals range from being born in 1994 to 2008 as shown in figure 1.

**Figure 1.** Numbers of Johnes-positive samples by year of birth



**Plink analysis**

At the time of the analysis, there were 1,430 high-density Johnes genotyped animals available. These were analyzed with all 22,452 control dams, most of which were 50k genotypes imputed up to high-density. Table 2 contains the top 50 SNP with the greatest effect. See the Appendix for manhattan plots of SNP effects for each chromosome both before and after adjusting for fixed effects of breed percentage and year of birth.

**GeneDrop analysis**

At the time of the analysis, there were 1,430 high-density Johnes genotypes available. The analysis was performed across breeds and with Jerseys only. Tables 3 and 4 contain the top 50 SNP with the greatest effect for the analysis across breeds, and Jersey only analysis, respectively.

**Table 2.** PLINK results for the top 50 SNP with the greatest effect on Johne's Disease occurrence

CHR	BP	P	$-\log_{10}(P)$
11	53667205	5.771e-309	308.24
1	37302581	5.798e-301	300.24
9	84407907	4.739e-298	297.32
1	130651474	1.491e-294	293.83
17	6096279	5.851e-294	293.23
2	104835932	1.159e-292	291.94
5	80169191	2.653e-285	284.58
2	104838590	2.472e-278	277.61
8	29615889	8.58e-278	277.07
20	6402124	2.405e-275	274.62
2	104824189	1.302e-274	273.89
21	52971607	5.877e-271	270.23
8	89710756	3.929e-267	266.41
7	326553	9.587e-266	265.02
7	314077	1.802e-264	263.74
16	29377815	2.072e-260	259.68
27	21300127	1.66e-259	258.78
13	34279710	4.501e-259	258.35
20	27955583	1.282e-257	256.89
26	51550459	1.008e-256	256.00
8	13402990	1.718e-249	248.76
5	87064339	9.646e-249	248.02
2	104857458	8.699e-240	239.06
7	23855608	3.716e-239	238.43
1	133415401	3.714e-238	237.43
9	53044764	2.382e-237	236.62
2	121106581	9.274e-236	235.03
4	8400250	9.718e-236	235.01
7	24282757	5.956e-229	228.23
4	17065384	1.669e-228	227.78
1	50694989	4.954e-227	226.31
9	89140849	5.986e-226	225.22
8	82358868	1.512e-225	224.82
1	49352176	1.808e-223	222.74
1	49356559	1.808e-223	222.74
15	5975694	3.582e-223	222.45
19	63374904	6.293e-223	222.20
20	69161252	4.48e-222	221.35
26	2097464	7.349e-220	219.13
4	21217532	1.041e-219	218.98
1	50694060	1.235e-219	218.91
14	47510390	2.748e-218	217.56
5	13793056	1.09e-217	216.96
2	86680839	2.845e-216	215.55
26	24671624	2.904e-216	215.54
26	23333769	1.017e-215	214.99
2	26647403	1.89e-212	211.72
7	46984331	1.421e-211	210.85
7	46983491	1.58e-211	210.80
7	46986222	1.58e-211	210.80

**Table 3.** GeneDrop results across all breeds for the top 50 SNP with the greatest effect on Johne's Disease occurrence

CHR	BP	P	$-\log_{10}(P)$
21	52971607	1.174055079e-234	233.93
1	20782171	7.701710254e-181	180.11
2	86680839	9.53703065e-180	179.02
6	70136705	5.962924312e-166	165.22
1	40605078	1.146840698e-142	141.94
12	25544396	9.08535394e-128	127.04
6	24860898	1.133600961e-114	113.95
17	57691623	5.44151421e-112	111.26
11	32447991	3.123218305e-97	96.51
6	52064261	1.001851378e-95	95.00
11	63027162	2.877980904e-91	90.54
22	32796544	1.093714603e-88	87.96
1	133834146	8.022765922e-86	85.10
12	49878698	2.805828015e-85	84.55
1	61779936	4.452213963e-83	82.35
27	1163789	5.859700123e-83	82.23
1	17863776	4.96863057e-82	81.30
5	80169191	1.99015567e-80	79.70
9	55965650	4.695999219e-74	73.33
11	28420313	7.266820044e-73	72.14
11	34977466	4.78636702e-70	69.32
7	10092268	9.930612102e-69	68.00
20	27009784	8.21881774e-68	67.09
10	65943603	2.875478518e-64	63.54
3	52163849	5.769915525e-64	63.24
1	84799179	4.261567337e-60	59.37
17	27416064	5.071569093e-54	53.29
28	8720872	7.304875424e-52	51.14
27	43873735	1.681084783e-47	46.77
21	52642002	2.033019992e-47	46.69
27	35042130	2.252576214e-47	46.65
21	35510518	3.533230395e-47	46.45
20	37979437	9.355839278e-46	45.03
15	45828017	5.022686997e-45	44.30
9	15663564	8.072056849e-45	44.09
13	55299036	3.260964267e-44	43.49
3	53784243	5.168239284e-44	43.29
20	58329179	5.293360836e-44	43.28
15	45858786	9.400465611e-44	43.03
13	5690382	1.949408685e-43	42.71
13	55300785	6.5882432e-43	42.18
11	1013822	8.582113603e-43	42.07
13	5737984	1.198557696e-42	41.92
14	25505663	5.807240814e-42	41.24
14	25506575	5.807240814e-42	41.24
14	25507730	5.807240814e-42	41.24
14	25510859	5.807240814e-42	41.24
14	25529645	5.950885958e-42	41.23
14	25536019	5.950885958e-42	41.23
14	25537252	5.950885958e-42	41.23

**Table 4.** GeneDrop results for Jerseys for the top 50 SNP with the greatest effect on Johne's Disease occurrence

CHR	BP	P	$-\log_{10}(P)$
21	52642002	1.36948447e-252	251.86
17	18451699	1.396922196e-216	215.85
1	40605078	1.013267377e-207	206.99
11	32447991	2.991826204e-200	199.52
11	48806126	2.068881112e-175	174.68
3	57551250	8.361290948e-171	170.08
10	55506736	1.093220351e-162	161.96
20	54839930	2.360721905e-162	161.63
1	17863776	7.082245017e-140	139.15
28	8720872	4.270762844e-139	138.37
7	41870969	8.195262731e-136	135.09
2	55305358	6.27249016e-133	132.20
2	103674928	3.002462511e-130	129.52
27	1163789	3.010573715e-115	114.52
23	48095660	9.553493039e-112	111.02
27	43873735	2.121359902e-109	108.67
28	9023118	1.872000755e-107	106.73
18	5074252	1.641900886e-104	103.78
6	24860898	8.148769237e-104	103.09
17	57691623	1.061065623e-91	90.97
1	84799179	6.906284835e-82	81.16
29	38983875	5.091679825e-76	75.29
27	19375259	6.744031136e-75	74.17
15	11147347	9.852113622e-73	72.01
20	27009784	4.554007679e-71	70.34
12	38609164	1.359050514e-68	67.87
22	47740297	1.14927869e-67	66.94
10	10519944	4.197975196e-64	63.38
9	54005390	4.373982759e-63	62.36
12	25544396	1.764452802e-61	60.75
7	49479134	2.288220436e-61	60.64
23	24386889	1.900579969e-58	57.72
5	80169191	2.204985778e-58	57.66
4	76768665	2.483929893e-58	57.60
12	46109141	5.205730451e-56	55.28
11	34977466	2.493589881e-55	54.60
10	65943603	3.467567368e-55	54.46
21	37145945	6.057382057e-51	50.22
7	10092268	7.729497029e-51	50.11
4	35426434	8.949971017e-50	49.05
12	41410035	1.329726588e-48	47.88
7	94158211	9.811374985e-48	47.01
9	54886468	1.397011294e-47	46.85
11	56337652	1.699044477e-46	45.77
14	77279332	3.639625661e-46	45.44
5	78027505	1.295537894e-45	44.89
14	77280386	1.391111917e-45	44.86
14	77281908	1.391111917e-45	44.86
14	77284832	1.391111917e-45	44.86
14	77289804	1.391111917e-45	44.86

### **Bayesian Lasso results**

At the time of the analysis, there were 1,842 high-density Johnes genotypes available. Animals were removed that were not either Holstein Friesian, Jersey or Holstein-Friesian-Jersey cross. This left 1,346 Johnes animals. Of the 22,452 control dams, 1,346 were randomly selected to match the breed percentages found in the Johnes group. This resulted in 2,692 animals per sample. R's BLR package was used (de los Campos and Rodriguez, 2010) for the analysis. Fixed effects breed percentage and year of birth were included in the model. The analysis was repeated over 4 samples, and the results of the samples were averaged. Table 5 contains the top 50 SNP with the greatest effect. See the Appendix for manhattan plots of SNP effects for each chromosome. The average genotypic and allelic frequencies (across the 4 samples) for Johnes animals and controls were calculated for the average top 50 SNP, and are shown in Tables 6 and 7.



**Table 5.** Bayesian Lasso results for the top 50 SNP with the greatest effect on Johne's Disease occurrence

CHR	SNP	BP	BL estimate
19	ARS-BFGL-NGS-105532	52264019	0.000451
21	BovineHD2100008456	29453554	0.000327
6	Hapmap34404-BES10_Contig556_782	33097354	0.000326
8	ARS-BFGL-NGS-7009	1406173	0.000298
28	BTB-02015177	35668756	0.000292
13	ARS-BFGL-NGS-100178	30062520	0.000273
26	ARS-BFGL-NGS-104867	14512486	0.000258
14	BovineHD1400016243	58525430	0.000257
3	BovineHD0300018189	60597021	0.000238
4	BovineHD0400010006	35635120	0.000232
1	BovineHD0100017649	62224652	0.000223
16	BovineHD1600010424	36381373	0.000213
4	ARS-BFGL-NGS-116672	34794353	0.000208
1	BovineHD0100010687	37302581	0.000206
7	BovineHD0700012107	41663063	0.000205
29	BovineHD2900014594	49581850	0.000205
14	Hapmap61004-rs29017348	35518739	0.000204
2	BovineHD0200011906	40914004	0.000200
22	BovineHD2200007736	26828581	0.000199
10	BovineHD1000016471	55570835	0.000198
15	BovineHD1500012722	45377058	0.000193
9	BovineHD0900003822	14693542	0.000191
8	BovineHD0800019652	65684394	0.000190
4	BovineHD0400031099	108843618	0.000177
11	BovineHD1100015649	53667205	0.000176
11	BovineHD1100015061	51260862	0.000172
8	BTB-01563190	48070965	0.000169
28	BovineHD2800006070	23262813	0.000167
2	BovineHD0200029738	103525780	0.000167
11	BovineHD1100016346	55933757	0.000166
8	BTB-00362651	84896726	0.000165
27	BovineHD2700006098	21613004	0.000165
7	BovineHD0700012366	42478275	0.000163
5	BovineHD0500024676	87064339	0.000160
5	BovineHD0500005662	19588556	0.000160
5	BovineHD0500027960	98105788	0.000158
5	BovineHD0500024398	86123665	0.000156
26	BovineHD2600006687	25589629	0.000155
1	BovineHD0100031358	110907400	0.000154
2	BovineHD0200020209	70283460	0.000150
5	BovineHD0500022126	77899012	0.000149
10	BovineHD1000028530	98637473	0.000148
26	BovineHD2600000314	2097464	0.000147
13	BovineHD1300000940	3448480	0.000145
10	BovineHD1000026437	91731772	0.000144
5	BovineHD0500009004	30812280	0.000144
3	BovineHD0300020381	68996341	0.000144
23	BovineHD2300001680	6736346	0.000142
11	ARS-BFGL-NGS-11940	7346147	0.000141
25	BovineHD2500003940	14101539	0.000141

**Table 6.** Genotypic frequencies for Johnes and control animals averaged across the 4 BLR samples.

CHR	SNP	BP	Controls			Johnes		
			AA	AB	BB	AA	AB	BB
19	ARS-BFGL-NGS-105532	52264019	0.16	0.45	0.39	0.00	0.14	0.86
21	BovineHD2100008456	29453554	0.19	0.48	0.33	0.03	0.33	0.64
6	Hapmap34404-BES10_Contig556_782	33097354	0.25	0.45	0.29	0.26	0.01	0.73
8	ARS-BFGL-NGS-7009	1406173	0.32	0.42	0.26	0.35	0.01	0.64
28	BTB-02015177	35668756	0.41	0.48	0.12	0.01	0.85	0.15
13	ARS-BFGL-NGS-100178	30062520	0.49	0.37	0.13	0.48	0.01	0.51
26	ARS-BFGL-NGS-104867	14512486	0.51	0.35	0.13	0.52	0.01	0.47
14	BovineHD1400016243	58525430	0.96	0.04	0.00	0.56	0.44	0.00
3	BovineHD0300018189	60597021	0.65	0.32	0.03	0.27	0.64	0.09
4	BovineHD0400010006	35635120	0.64	0.32	0.04	0.24	0.68	0.08
1	BovineHD0100017649	62224652	0.61	0.34	0.05	0.24	0.70	0.06
16	BovineHD1600010424	36381373	0.54	0.40	0.07	0.21	0.68	0.10
4	ARS-BFGL-NGS-116672	34794353	0.64	0.33	0.04	0.33	0.57	0.10
1	BovineHD0100010687	37302581	0.59	0.36	0.04	0.23	0.69	0.07
7	BovineHD0700012107	41663063	0.72	0.25	0.04	0.35	0.61	0.04
29	BovineHD2900014594	49581850	0.81	0.18	0.01	0.45	0.54	0.01
14	Hapmap61004-rs29017348	35518739	0.63	0.31	0.06	0.55	0.01	0.44
2	BovineHD0200011906	40914004	0.98	0.02	0.00	0.64	0.36	0.00
22	BovineHD2200007736	26828581	0.08	0.41	0.51	0.03	0.25	0.72
10	BovineHD1000016471	55570835	0.56	0.39	0.04	0.22	0.69	0.09
15	BovineHD1500012722	45377058	0.58	0.36	0.06	0.26	0.67	0.08
9	BovineHD0900003822	14693542	0.51	0.41	0.07	0.19	0.67	0.14
8	BovineHD0800019652	65684394	0.68	0.29	0.03	0.31	0.68	0.02
4	BovineHD0400031099	108843618	0.55	0.40	0.05	0.25	0.69	0.07
11	BovineHD1100015649	53667205	0.69	0.28	0.03	0.36	0.61	0.04
11	BovineHD1100015061	51260862	0.93	0.06	0.00	0.72	0.27	0.01
8	BTB-01563190	48070965	0.58	0.33	0.09	0.63	0.01	0.36
28	BovineHD2800006070	23262813	0.52	0.41	0.08	0.23	0.66	0.11
2	BovineHD0200029738	103525780	0.73	0.26	0.02	0.43	0.54	0.02
11	BovineHD1100016346	55933757	0.52	0.41	0.07	0.23	0.66	0.11
8	BTB-00362651	84896726	0.62	0.30	0.08	0.64	0.01	0.35
27	BovineHD2700006098	21613004	0.87	0.13	0.00	0.55	0.44	0.01
7	BovineHD0700012366	42478275	0.51	0.35	0.14	0.28	0.57	0.15
5	BovineHD0500024676	87064339	0.74	0.25	0.01	0.46	0.50	0.04
5	BovineHD0500005662	19588556	0.51	0.41	0.08	0.22	0.67	0.11
5	BovineHD0500027960	98105788	0.76	0.23	0.02	0.51	0.47	0.02
5	BovineHD0500024398	86123665	0.91	0.09	0.00	0.64	0.35	0.01
26	BovineHD2600006687	25589629	0.37	0.47	0.16	0.15	0.67	0.18
1	BovineHD0100031358	110907400	0.38	0.48	0.14	0.15	0.66	0.19
2	BovineHD0200020209	70283460	0.75	0.23	0.02	0.61	0.33	0.06
5	BovineHD0500022126	77899012	0.39	0.48	0.12	0.15	0.67	0.18
10	BovineHD1000028530	98637473	0.55	0.40	0.05	0.28	0.66	0.06
26	BovineHD2600000314	2097464	0.76	0.23	0.02	0.49	0.47	0.04
13	BovineHD1300000940	3448480	0.46	0.44	0.11	0.19	0.66	0.15
10	BovineHD1000026437	91731772	0.96	0.04	0.00	0.69	0.31	0.00
5	BovineHD0500009004	30812280	0.61	0.34	0.05	0.31	0.64	0.05
3	BovineHD0300020381	68996341	0.49	0.43	0.08	0.20	0.71	0.09
23	BovineHD2300001680	6736346	0.09	0.46	0.44	0.12	0.21	0.68
11	ARS-BFGL-NGS-11940	7346147	0.60	0.31	0.09	0.60	0.01	0.39
25	BovineHD2500003940	14101539	0.79	0.19	0.01	0.54	0.44	0.02

**Table 7.** Allelic frequencies for Johnes and control animals averaged across the 4 BLR samples

CHR	SNP	BP	Controls		Johnes	
			A	B	A	B
19	ARS-BFGL-NGS-105532	52264019	38.15	61.85	6.97	93.03
21	BovineHD2100008456	29453554	42.67	57.33	19.73	80.27
6	Hapmap34404-BES10_Contig556_782	33097354	48.00	52.00	26.41	73.59
8	ARS-BFGL-NGS-7009	1406173	53.21	46.79	35.54	64.46
28	BTB-02015177	35668756	64.41	35.59	43.05	56.95
13	ARS-BFGL-NGS-100178	30062520	68.09	31.91	48.63	51.37
26	ARS-BFGL-NGS-104867	14512486	69.12	30.88	52.82	47.18
14	BovineHD1400016243	58525430	98.17	1.83	77.85	22.15
3	BovineHD0300018189	60597021	80.91	19.09	58.76	41.24
4	BovineHD0400010006	35635120	80.10	19.90	58.06	41.94
1	BovineHD0100017649	62224652	77.74	22.26	59.43	40.57
16	BovineHD1600010424	36381373	73.55	26.45	55.65	44.35
4	ARS-BFGL-NGS-116672	34794353	80.18	19.82	61.75	38.25
1	BovineHD0100010687	37302581	77.54	22.46	57.95	42.05
7	BovineHD0700012107	41663063	84.05	15.95	65.68	34.32
29	BovineHD2900014594	49581850	90.37	9.63	72.07	27.93
14	Hapmap61004-rs29017348	35518739	78.51	21.49	55.79	44.21
2	BovineHD0200011906	40914004	98.78	1.22	81.88	18.12
22	BovineHD2200007736	26828581	28.29	71.71	15.10	84.90
10	BovineHD1000016471	55570835	76.09	23.91	56.06	43.94
15	BovineHD1500012722	45377058	76.09	23.91	59.12	40.88
9	BovineHD0900003822	14693542	72.07	27.93	52.58	47.42
8	BovineHD0800019652	65684394	82.66	17.34	64.46	35.54
4	BovineHD0400031099	108843618	74.83	25.17	58.78	41.22
11	BovineHD1100015649	53667205	83.07	16.93	66.08	33.92
11	BovineHD1100015061	51260862	96.54	3.46	85.59	14.41
8	BTB-01563190	48070965	74.23	25.77	63.51	36.49
28	BovineHD2800006070	23262813	71.94	28.06	56.03	43.97
2	BovineHD0200029738	103525780	85.69	14.31	70.60	29.40
11	BovineHD1100016346	55933757	72.16	27.84	55.59	44.41
8	BTB-00362651	84896726	76.87	23.13	64.45	35.55
27	BovineHD2700006098	21613004	93.06	6.94	77.07	22.93
7	BovineHD0700012366	42478275	68.25	31.75	56.31	43.69
5	BovineHD0500024676	87064339	86.33	13.67	70.87	29.13
5	BovineHD0500005662	19588556	71.60	28.40	55.76	44.24
5	BovineHD0500027960	98105788	87.00	13.00	74.28	25.72
5	BovineHD0500024398	86123665	95.41	4.59	81.85	18.15
26	BovineHD2600006687	25589629	60.74	39.26	48.24	51.76
1	BovineHD0100031358	110907400	62.14	37.86	48.38	51.62
2	BovineHD0200020209	70283460	86.90	13.10	77.41	22.59
5	BovineHD0500022126	77899012	63.55	36.45	48.71	51.29
10	BovineHD1000028530	98637473	74.83	25.17	61.25	38.75
26	BovineHD2600000314	2097464	87.04	12.96	72.34	27.66
13	BovineHD1300000940	3448480	67.49	32.51	52.32	47.68
10	BovineHD1000026437	91731772	97.85	2.15	84.28	15.72
5	BovineHD0500009004	30812280	78.05	21.95	63.06	36.94
3	BovineHD0300020381	68996341	70.12	29.88	55.24	44.76
23	BovineHD2300001680	6736346	32.61	67.39	21.92	78.08
11	ARS-BFGL-NGS-11940	7346147	75.33	24.67	60.18	39.82
25	BovineHD2500003940	14101539	88.81	11.19	76.05	23.95

## Discussion

Three different statistical models were used to analyse the genotypic data. Each of the methods have their pros and cons. The single marker association mapping (PLINK) is quicker than the other three methods but does not account for SNP effects that have been estimated in other parts of the genome. Bayesian Lasso is computationally demanding but has the advantage of fitting all SNP effects at the same time. The Genedrop method removes the need to have a control population, which can generate ascertainment bias. However it is dependent on the pedigree being recorded correctly, which is an issue as the parentage error rate in the NZ dairy population is approximately 20%. This can be overcome utilizing genomic data to identify close ancestors to the animal of interest. This step has yet to be undertaken with the dataset and the results are based on the recorded pedigree.

Concordance between the 3 methods is moderate. Of the top 10 markers from the Bayesian Lasso, two are in the top 50 for the PLINK analysis and 1 for the Genedrop method (Table 8). For the other two methods the level of concordance is higher (Table 8).

**Table 8.** Number of the top 10 markers that are in the top 50 for the other statistical methods.

	Method for the top 10 hits		
	PLINK	Bayesian Lasso	Gene drop
PLINK	-	2	4
Bayesian Lasso	4	-	1
Gene drop	4	1	-

The PLINK and Gene drop methods have to be re-run with the larger dataset of 1841 animals to complete the analysis and also to be more comparable with the other two methods. Identifying genomic regions that are consistent across the three methods will increase the likelihood that they are of biological significance.

The analysis is dependent on the control population that is chosen. As shown in Appendix one fitting breed in the model reduces the level of background significance and thus increases the clarity of the markers that have statistical significance. The Bayesian Lasso method uses a subsample of the control animals but matching by breed is undertaken to try and reduce the breed effect. Repeated sampling of the control and re-running of the analysis is an attempt to reduce the stochastic effect of the sampling but is limited by the computational demands of the method. The Genedrop method is the method that is most robust to the control especially if genomic relatedness is used to derive the pedigree.

The allelic and genotypic frequencies for the 50 most significant markers from the Bayesian Lasso analysis shows quite large differences between the affected and the controls (Tables 6 and 7). Some of the markers for the affected animals show very unusual genotypic frequencies eg. third, fourth and sixth most significant markers where there is a paucity of heterozygous markers. Further investigation is underway to see if this is being generated by admixture of the two breeds.

Taking the most significant regions from the Bayesian Lasso method there are various genes that look biologically interesting. For the most significant marker on chromosome 19 (ARS-BFGL-NGS-105532) genes that appear interesting include SLC38A10 (BP 52045219..52086316), SLC26A11 (BP 53023436..53041077), SLC16A3 (BP 51243972..51255385), and SOCS-3 (BP 54458856..5445955). Solute Carrier (SLC) family genes are often expressed in the intestines, are linked to immunity, and are often the subject of candidate gene analyses for Johne's disease. The SOCS-3 gene is a suppressor of cytokine signaling and has been linked with inflammatory bowel disease. Additionally, chromosome 21's BovineHD2100008456 marker (the marker with the second greatest effect) is located near IL16 (BP 27590364..27606711), a gene that influences immunity. Further investigation of the positional candidate genes will be undertaken. This may include identifying new SNPs from LIC's sequencing project and further genotyping in the Johne's population to see if the statistical significance of the signal can be increased.

### **Future Direction**

The analysis is not complete and it is expected that there are 3-4 months of more work required. A number of methods have been used as we attempt to work with a quite unique dataset. The analysis may be extended to investigate analysis methods that use haplotypes rather than the single SNPs that have been used in the three methods used to date. Further analysis that utilize the crossbred animals in the population and exploit the greater susceptibility of the Jersey population may also be investigated. Furthermore alternative Bayes ((Bayes B, Bayes N)) algorithms that only fit a selection of SNPs at once or fit physical sections of the chromosome may perform better

Once a set of markers is decided on as being predictive of Johne's disease cross-validation methods will be used with the affected and control animals to test the predictive ability of the markers. It is expected that the markers will have moderate predictive ability and will benefit from combining this dataset with others that have been generated in dairy cow populations.

### **References**

B L Browning and S R Browning (2009) A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 84:210-223. doi:10.1016/j.ajhg.2009.01.005.

S R Browning and B L Browning (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am J Hum Genet* 81:1084-1097. doi:10.1086/521987

de los Campos and Rodriguez, (2010) <http://cran.r-project.org/web/packages/BLR/BLR.pdf>

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.