Haplotype-based genomic prediction in Atlantic salmon

Solomon Antwi Boison

Serap Gonen, Ashie Norris and Matt Baranski

Mowi ASA, PO Box 4102 Sandviken, 5835 Bergen, Norway

Introduction

- GS >> widely adopted in Atlantic salmon breeding populations

- increase in genetic gain comes from
  - ⬆ *intensity* of selection (within family selection)
  - ⬆ *accuracy* of selection (full-sib accuracy ~ 0.71)

$$\boxed{\Delta G = \frac{i * r * \sigma_a}{L}}$$

Introduction

- Currently results with SNP alleles suggest a substantial increase in accuracy of selection

Trait type	Trait	Pedigree	Genomic	Gain (%)
Growth	Bodyweight	0.50	0.70	**40**
	Fillet Colour	0.48	0.67	**40**
	Smolt weight	0.72	0.90	**25**

Introduction

- Currently results with SNP alleles suggest a substantial increase in accuracy of selection

Trait type	Trait	Pedigree	Genomic	Gain (%)
Growth	Bodyweight	0.50	0.70	**40**
	Fillet Colour	0.48	0.67	**40**
	Smolt weight	0.72	0.90	**25**
Diseases	AGD	0.50	0.65	**30**
	CMS	0.42	0.52	**25**
	PD	0.63	0.83	**32**
	SL	0.49	0.59	**20**

Introduction

- Haplotype based GS approach has been shown to further increase accuracy (Villumsen and Janss, 2009, BMC Proc; Cuyabano et al., 2015, JDS; Hess et al., 2017, GSE; Karami Z., 2018, PhD Thesis)

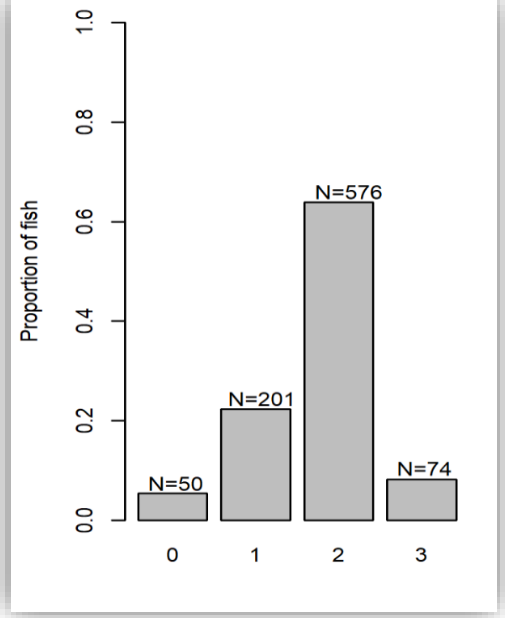
Introduction

- Haplotype based GS approach has been shown to further increase accuracy (Villumsen and Janss 2009, BMC Proc; Cuyabano et al. 2015, JDS; Hess et al. 2017, GSE; Karami Z. 2018, PhD Thesis)

Objective : Investigate the potential of using haplotype-based GS approaches in Atlantic salmon population to further increase accuracy of selection

Materials – sample information and phenotypes

- 32 sires and 73 dams >>> mating ratio 1:2 (65%) and 1:3 (35%)
- 1,095 fish (15 eye-eggs per full-sibling family)
- Phenotype:
  - Challenge tested for cardiomyopathy syndrome (CMS) disease
  - Histopathology of the heart (atrium score)
  - Estimated $h^2_{CMS} = 0.39 - 0.47\ (se = 0.056 - 0.088)$



Methods – phasing and haplotype analysis

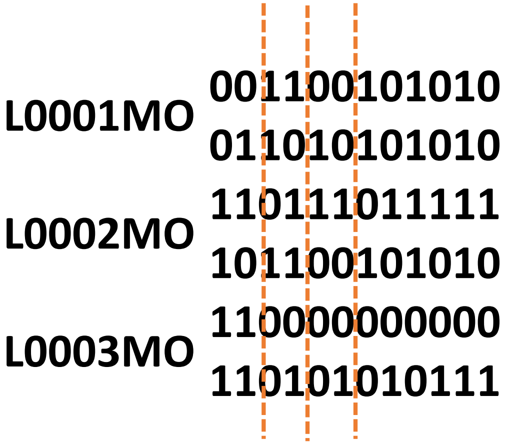
- Phasing >> Beagle v4.1
    - parental and offspring genotypes (53,678 SNPs)

- Haplotypes were constructed as moving windows
    - Fixed length in kilobase pairs (100, 250, 500, 1000)
    - Fixed length in number of SNPs (5, 10, 20)

Methods – phasing and haplotype analysis

- Phasing >> Beagle v4.1
  - parental and offspring genotypes (53,678 SNPs)

- Haplotypes were constructed as moving windows
  - Fixed length in kilobase pairs (100, 250, 500, 1000)
  - Fixed length in number of SNPs (5, 10, 20)



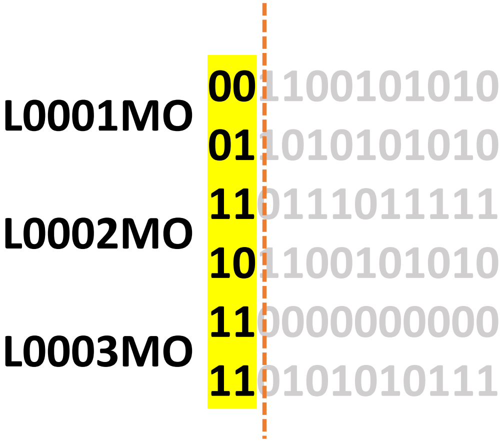
L0001MO
001100101010
011010101010

L0002MO
110111011111
101100101010

L0003MO
110000000000
110101010111

Methods – phasing and haplotype analysis

- Haplotypes were converted to bi-allelic SNPs



L0001MO	**00**1100101010 **01**1010101010
L0002MO	**11**0111011111 **10**1100101010
L0003MO	**11**0000000000 **11**0101010111

$$H = \{h_{00}, h_{01}, h_{10}, h_{11}\}$$

$$H = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

- $HAP_{freq} \geq 0.02$

Methods – pedigree and genomic predictions

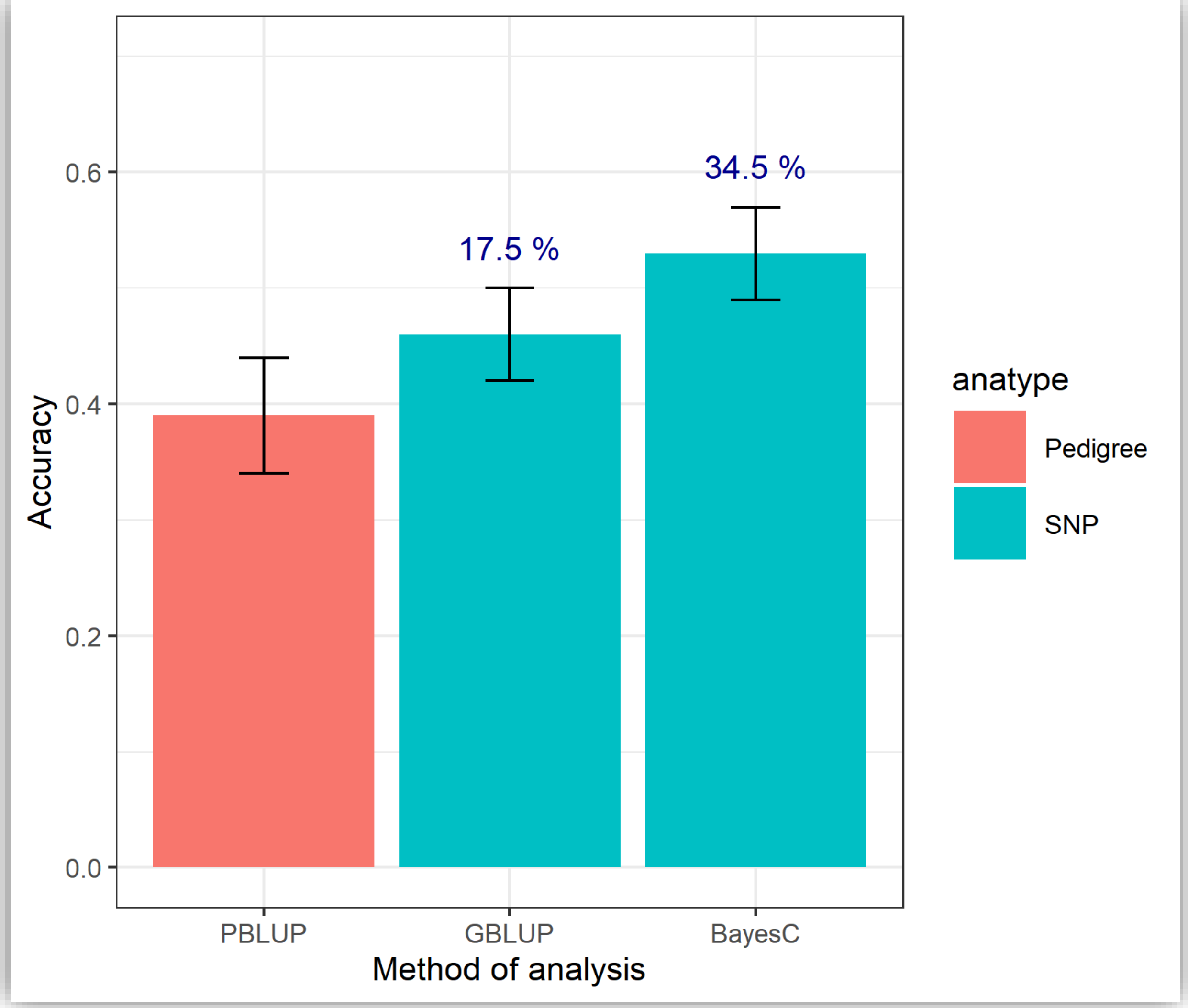
Information	**Method**	**Model**
Pedigree	PBLUP	$y = \mu + Tu + e;\ a\sim N(0, A\sigma_u^2);\ e\sim N(0, I\sigma_e^2)$ --- ASREML
SNP	GBLUP	$y = \mu + Tg + e; g\sim N\left(0, \frac{ZZ'}{2*\sum_i^{nSNP} p_i q_i}\sigma_g^2\right);\ e\sim N(0, I\sigma_e^2)$ --- ASREML
Haplotypes + singleton SNPs	GBLUP	
SNP	BayesCπ = 0.95	$y = \mu + Za + e;\ a\sim N(0, \chi^{-2});\ e\sim N(0, \chi^{-2}); g = Z\hat{a}$ --- GS3
Haplotypes + singleton SNPs	BayesCπ = 0.95	

Methods – pedigree, genomic predictions and cross validation

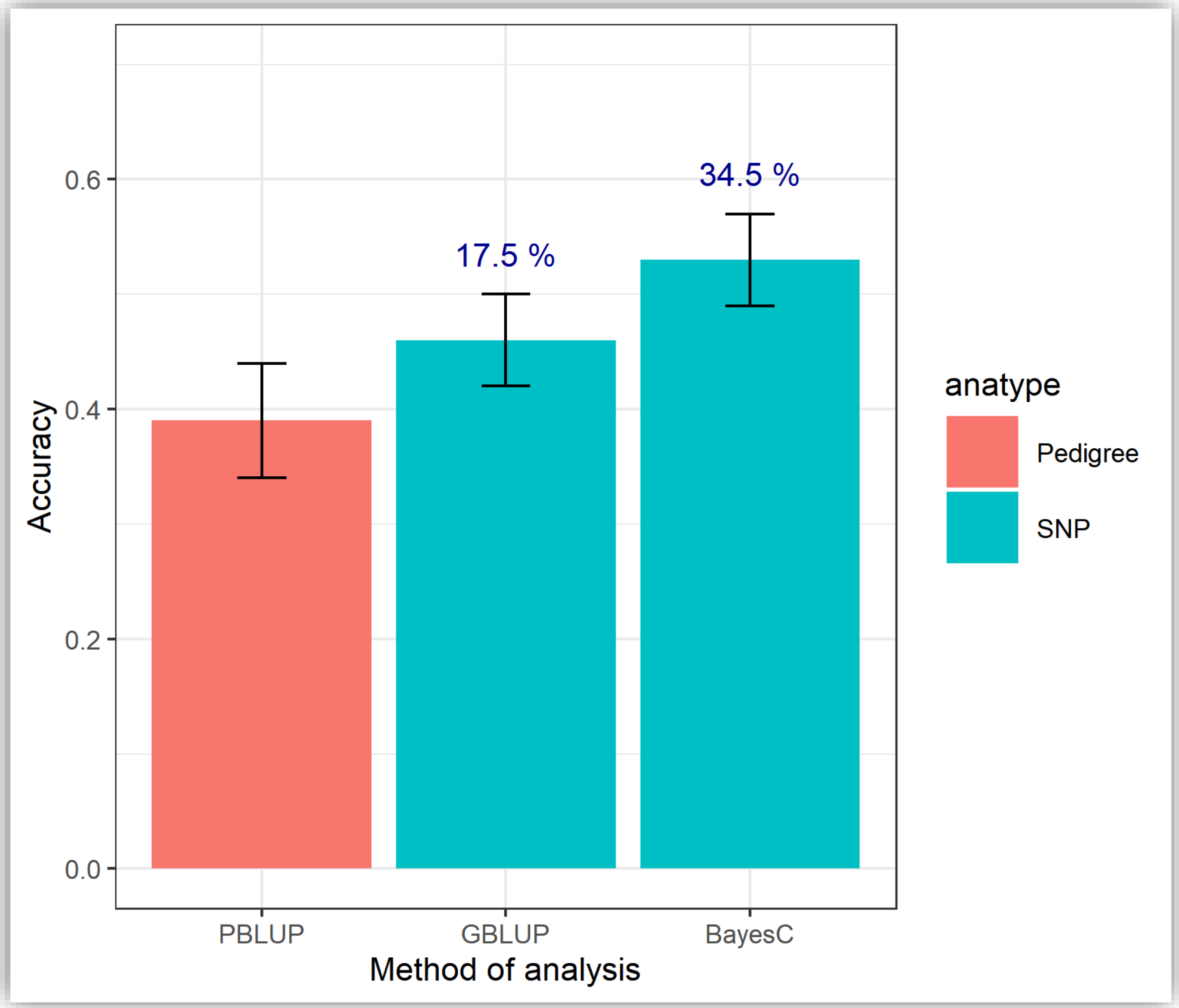
- Family - based validation scheme
  - Two full sibs from each family were randomly selected as validation (n=150) and the rest of the siblings as training (n=757) (20 replicates)

- Accuracy ($r = \frac{\rho(g[p]EBV, y)}{\sqrt{h^2}}$), inflation(bias) = $\frac{\text{cov}(\text{y}, g[p]EBV)}{\sigma^2_{g[p]EBV}}$

Results – pedigree vs. genomic (accuracy)



Results – pedigree vs. genomic (bias)



Type of dataset	Method of analysis	Bias (±SD)	% relative to PBLUP
Pedigree	PBLUP	1.10 (0.18)	
SNP	GBLUP	1.00 (0.13)	9.1 %
SNP	BayesCπ	1.03 (0.16)	6.7 %

Results – haplotype statistics

Window size (kbs/SNP)	N_{HAP} blocks	N-singletons (SNPs)
100 – kb	8,288	14,291
250 – kb	6,986	2,663
500 – kb	4,256	519
1000 – kb	2,246	189

Results – haplotype statistics

Window size (kbs/SNP)	N_{HAP} blocks	N-singletons (SNPs)	Total N_{HAP}	Total N_{HAP_freq} ≥0.02
100 – kb	8,288	14,291	71,392	38,081
250 – kb	6,986	2,663	109,197	38,960
500 – kb	4,256	519	121,792	25,949
1000 – kb	2,246	189	115,571	13,577

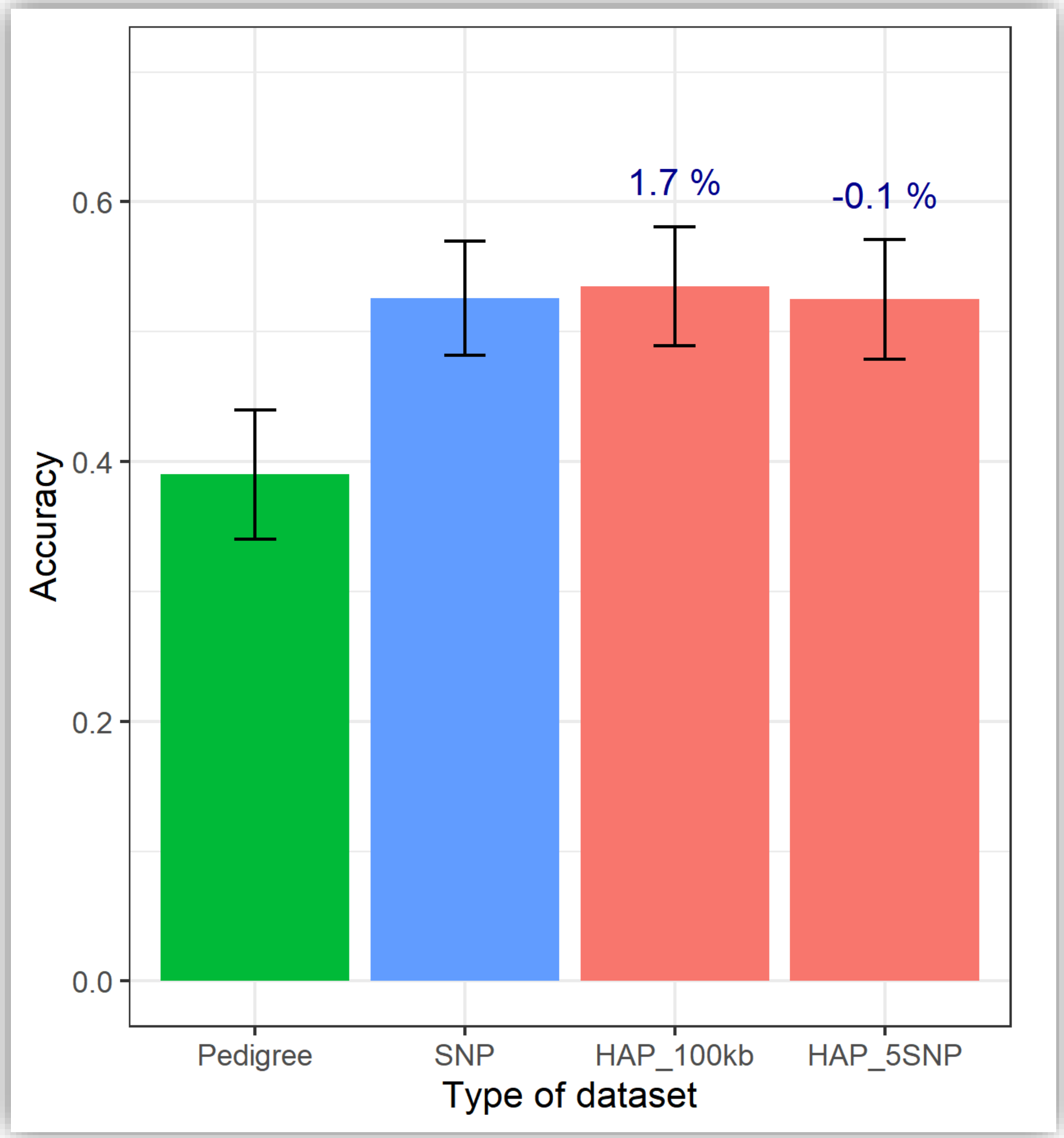
Results – haplotype statistics

Window size (kbs/SNP)	N_{HAP} blocks	N-singletons (SNPs)	Total N_{HAP}	Total N_{HAP_freq} ≥0.02
100 – kb	8,288	14,291	71,392	38,081
250 – kb	6,986	2,663	109,197	38,960
500 – kb	4,256	519	121,792	25,949
1000 – kb	2,246	189	115,571	13,577
5 – SNP	10,696		112,911	56,664
10 – SNP	5,341		137,676	33,915
20 – SNP	2,662		129,201	16,373

Results – comparing window sizes (kb/SNP)

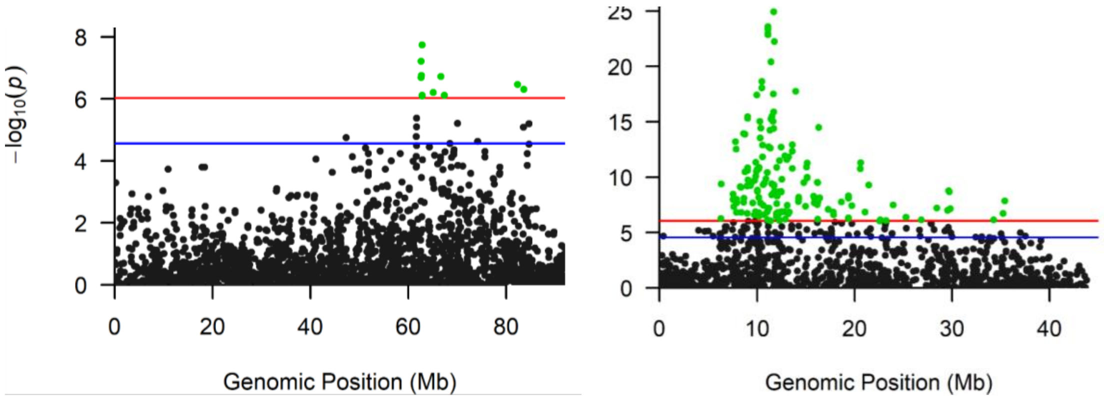
Window size (kbs/SNP)	Mean	SD
100 – kb	**0.534**	**0.05**
250 – kb	**0.529**	**0.05**
500 – kb	**0.515**	**0.05**
1000 – kb	**0.503**	**0.05**
5 – SNP	**0.525**	**0.05**
10 – SNP	**0.521**	**0.05**
20 – SNP	**0.518**	**0.05**

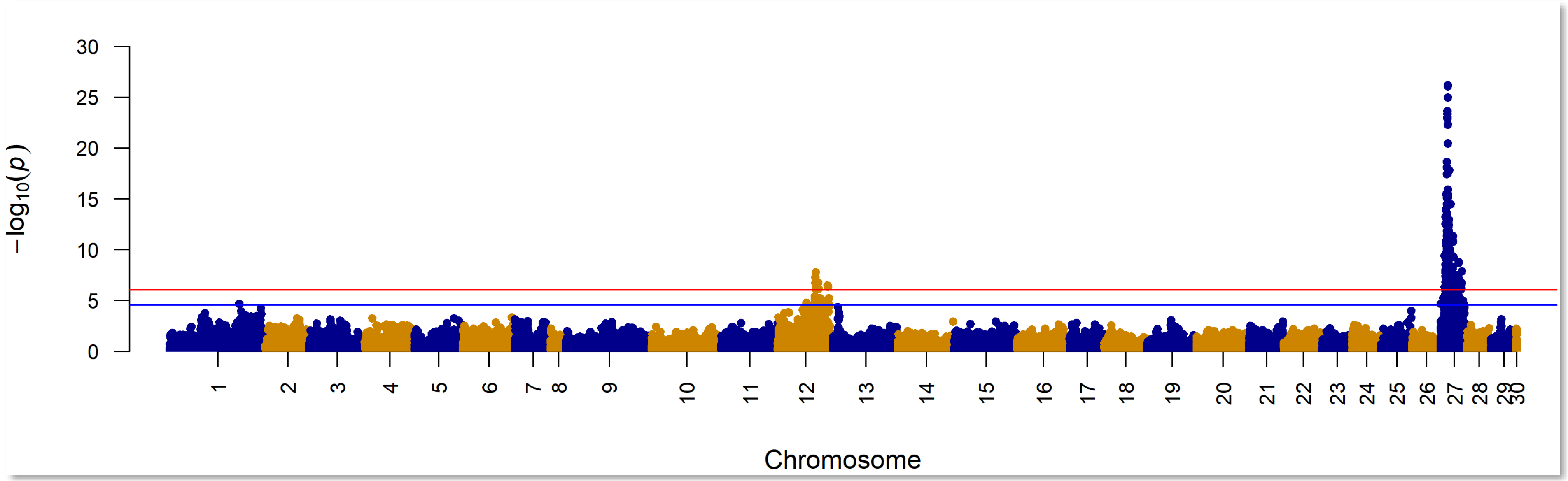
Results – comparing haplotypes and SNPs



Discussion

- Increase in accuracy with BayesC$\pi_{0.95}$ large >>> GBLUP
  - Reason: Large QTL effect that explains 40 - 50% of the genetic variance



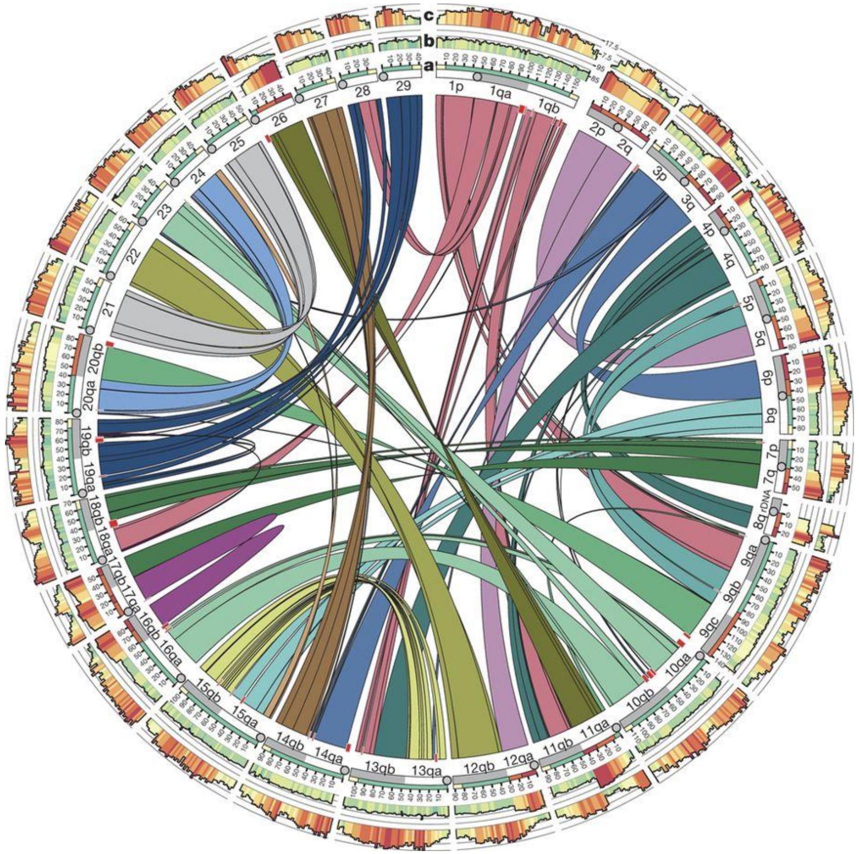


Discussion

- Haplotype based approaches **NOT** increasing accuracy

Reasons

- Genome assembly needs improvement

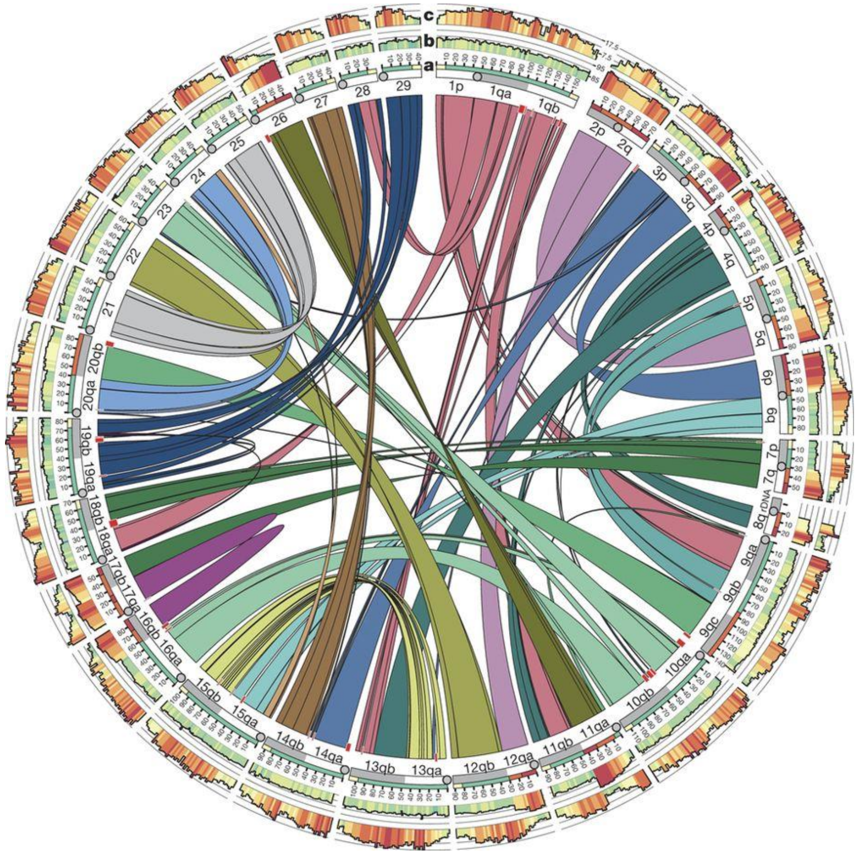


Discussion

- Haplotype based approaches **NOT** increasing accuracy

Reasons

- Genome assembly needs improvement
- LD in salmon populations relatively low
- Low sample size >>> will test with large dataset in next study



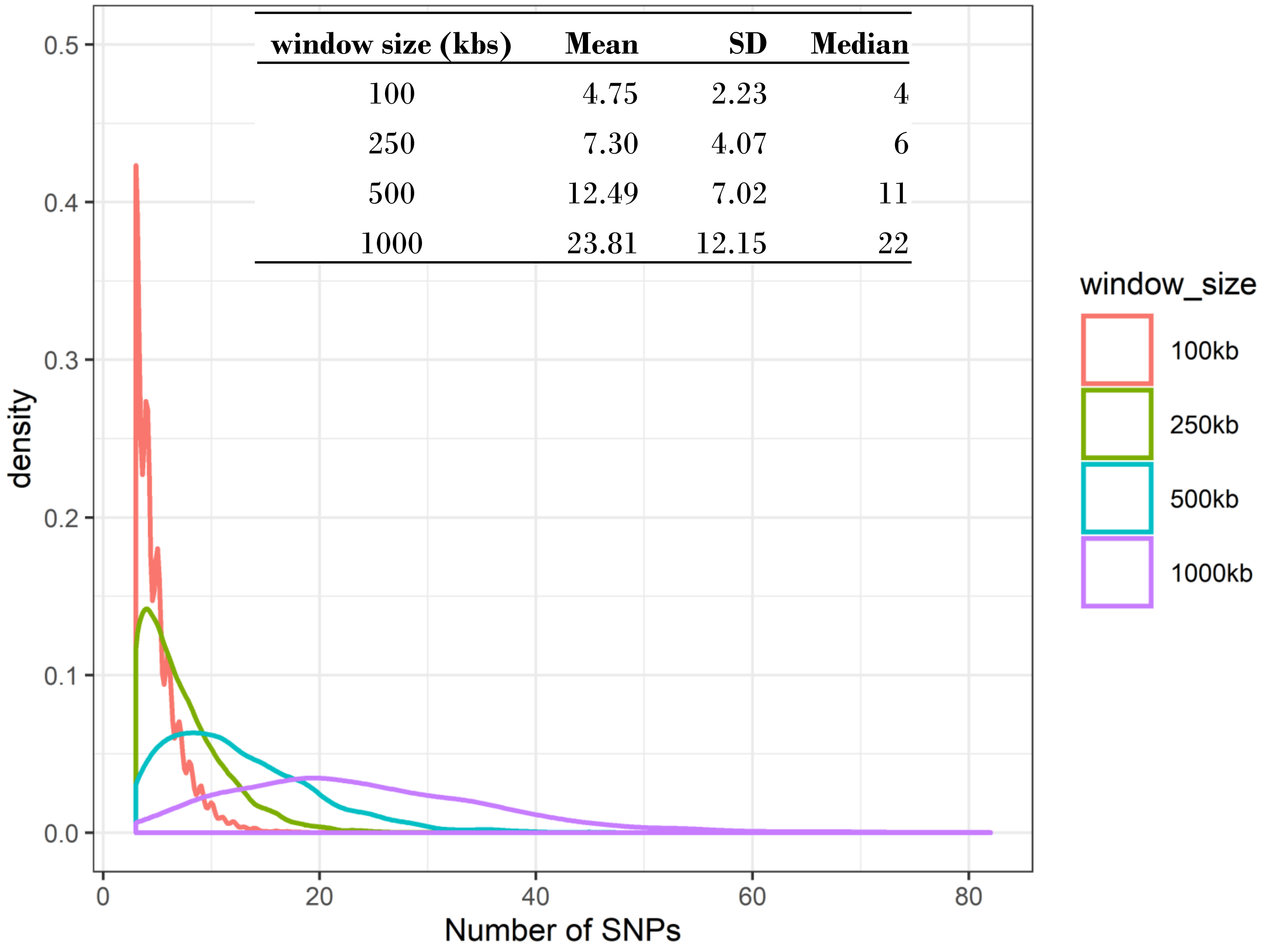
Conclusion

- Genomic information increases accuracy of selection in Atlantic salmon for CMS (17 - 34 %)
- As expected, for large effect QTL **Bayesian**/weighting markers differently results in increase in accuracy
- Haplotype based GS did not improve accuracy for CMS in this population

Thank you

Solomon.boison@mowi.com

Results – haplotype statistics



window size (kbs)	Mean	SD	Median
100	4.75	2.23	4
250	7.30	4.07	6
500	12.49	7.02	11
1000	23.81	12.15	22

ASE from BayesC