

Genomic Prediction Using Haplotype Blocks Built From High Density Marker Map

Beatriz Castro Dias Cuyabano

Main Supervisor: Guosheng Su

Project Supervisor: Mogens Sandø Lund

Center for Quantitative Genetics and Genomics
Graduate School of Science and Technology - Aarhus University, Denmark

December 11, 2012

Introduction

Introduction

- High Density Markers - Why?

Introduction

- High Density Markers - Why?
 - High Density (HD) marker chip (up to $\approx 770k$ markers) is now available;

Introduction

- High Density Markers - Why?
 - High Density (HD) marker chip (up to $\approx 770k$ markers) is now available;
 - Accuracy of genomic prediction can be further increased using HD marker or re-sequence information (Meuwissen and Goddard, 2010);

Introduction

- High Density Markers - Why?
 - High Density (HD) marker chip (up to $\approx 770k$ markers) is now available;
 - Accuracy of genomic prediction can be further increased using HD marker or re-sequence information (Meuwissen and Goddard, 2010);
 - Increasing degree of Linkage Disequilibrium (LD) between single-nucleotide polymorphism (SNP) and Quantitative Trait Loci (QTL);

Introduction

- High Density Markers - Why?
 - High Density (HD) marker chip (up to $\approx 770k$ markers) is now available;
 - Accuracy of genomic prediction can be further increased using HD marker or re-sequence information (Meuwissen and Goddard, 2010);
 - Increasing degree of Linkage Disequilibrium (LD) between single-nucleotide polymorphism (SNP) and Quantitative Trait Loci (QTL);
 - LD phase between breeds in the 50k chip is very low;

Introduction

- High Density Markers - Why?
 - High Density (HD) marker chip (up to $\approx 770k$ markers) is now available;
 - Accuracy of genomic prediction can be further increased using HD marker or re-sequence information (Meuwissen and Goddard, 2010);
 - Increasing degree of Linkage Disequilibrium (LD) between single-nucleotide polymorphism (SNP) and Quantitative Trait Loci (QTL);
 - LD phase between breeds in the 50k chip is very low;
 - HD maps will be very useful in genomic prediction across breeds due to the higher LD phase;

Introduction

- High Density Markers - Why?
 - High Density (HD) marker chip (up to $\approx 770k$ markers) is now available;
 - Accuracy of genomic prediction can be further increased using HD marker or re-sequence information (Meuwissen and Goddard, 2010);
 - Increasing degree of Linkage Disequilibrium (LD) between single-nucleotide polymorphism (SNP) and Quantitative Trait Loci (QTL);
 - LD phase between breeds in the 50k chip is very low;
 - HD maps will be very useful in genomic prediction across breeds due to the higher LD phase;
 - The challenge: genomic prediction with HD maps, using popular methods applied to low-medium density maps, have not presented significant improvement so far.

Haplotyping

Haplotyping

- First step: deal with the large amount of variables;

Haplotyping

- First step: deal with the large amount of variables;
- Build haplotype blocks within HD SNP-marker sequences:

Haplotyping

- First step: deal with the large amount of variables;
- Build haplotype blocks within HD SNP-marker sequences:

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 1 2 1 1 2 2 1 1 2 ...

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 2 1 1 2 2 1 1 1 2 ...

Haplotyping

- First step: deal with the large amount of variables;
- Build haplotype blocks within HD SNP-marker sequences:

... 1 1 2 1 2 2 2 1 1 2 ... haplotype 1

... 1 1 2 1 1 2 2 1 1 2 ...

... 1 1 2 1 2 2 2 1 1 2 ... haplotype 1

... 1 2 1 1 2 2 1 1 1 2 ...

Haplotyping

- First step: deal with the large amount of variables;
- Build haplotype blocks within HD SNP-marker sequences:

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 1 2 1 1 2 2 1 1 2 ... haplotype 2

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 2 1 1 2 2 1 1 1 2 ...

Haplotyping

- First step: deal with the large amount of variables;
- Build haplotype blocks within HD SNP-marker sequences:

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 1 2 1 1 2 2 1 1 2 ...

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 2 1 1 2 2 1 1 1 2 ... haplotype 3

Haplotyping

- First step: deal with the large amount of variables;
- Build haplotype blocks within HD SNP-marker sequences:

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 1 2 1 1 2 2 1 1 2 ...

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 2 1 1 2 2 1 1 1 2 ...

6 markers → 3 haplotypes: data/variable reduction!

Haplotyping

- First step: deal with the large amount of variables;
- Build haplotype blocks within HD SNP-marker sequences:

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 1 2 1 1 2 2 1 1 2 ...

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 2 1 1 2 2 1 1 1 2 ...

6 markers → 3 haplotypes: data/variable reduction!

- Where in the sequence should we build the blocks?

Haplotyping

- First step: deal with the large amount of variables;
- Build haplotype blocks within HD SNP-marker sequences:

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 1 2 1 1 2 2 1 1 2 ...

... 1 1 2 1 2 2 2 1 1 2 ...

... 1 2 1 1 2 2 1 1 1 2 ...

6 markers → 3 haplotypes: data/variable reduction!

- Where in the sequence should we build the blocks?
- How big should a block be?

Haplotyping

- Set LD threshold to define a haplotype block;

Haplotyping

- Set LD threshold to define a haplotype block;
 - $D' \geq d$ between any two markers $\Rightarrow$ Haplotype block;

Haplotyping

- Set LD threshold to define a haplotype block;
 - $D' \geq d$ between any two markers $\Rightarrow$ Haplotype block;
 - Threshold d varying in six values (0.25, 0.35, 0.45, 0.55, 0.65 and 0.75);

Haplotyping

- Set LD threshold to define a haplotype block;
 - $D' \geq d$ between any two markers $\Rightarrow$ Haplotype block;
 - Threshold d varying in six values (0.25, 0.35, 0.45, 0.55, 0.65 and 0.75);
- Our data:

Haplotyping

- Set LD threshold to define a haplotype block;
 - $D' \geq d$ between any two markers $\Rightarrow$ Haplotype block;
 - Threshold d varying in six values (0.25, 0.35, 0.45, 0.55, 0.65 and 0.75);
- Our data:
 - 5214 animals from the Nordic Holstein population;

Haplotyping

- Set LD threshold to define a haplotype block;
 - $D' \geq d$ between any two markers $\Rightarrow$ Haplotype block;
 - Threshold d varying in six values (0.25, 0.35, 0.45, 0.55, 0.65 and 0.75);
- Our data:
 - 5214 animals from the Nordic Holstein population;
 - $\approx 500k$ SNP markers.

Building Haploblocks

Building Haploblocks

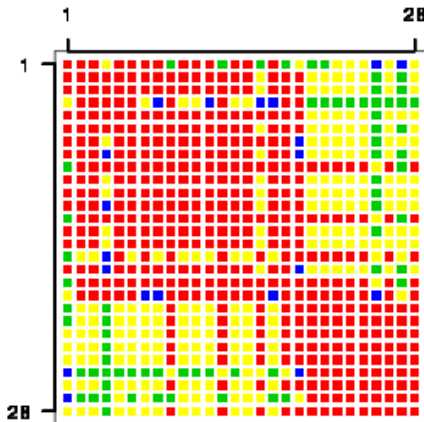


Figure: Toy example - LD map between all the markers

Building Haploblocks

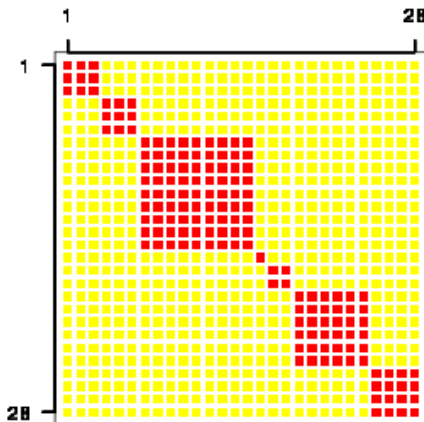


Figure: Toy example - Outlined haplotype blocks for $D' \geq 0.75$

Building Haploblocks

Table: Description of Haploblocks built based on the different D' thresholds.

	D' threshold					
	0.25	0.35	0.45	0.55	0.65	0.75
Non-blocked SNPs	3 513	5 399	7 744	10 280	13 207	16 812
# haploblocks	55 513	62 309	68 318	73 928	79 154	84 634
Haploblocks var.	338 460	346 938	353 221	358 461	362 455	366 167
Mean length	9	8	7	7	6	6
Max. length	80	74	78	78	78	70
Mean variants	6	6	5	5	5	4
Max. variants	18	18	16	15	13	12

Genomic Prediction - Models

Genomic Prediction - Models

- BLUP (with polygenic effect);

Genomic Prediction - Models

- BLUP (with polygenic effect);
- 4 Distribution Mixture Model (with polygenic effect):
Set prior proportions 0.889, 0.1, 0.01, 0.001 to variable effects.

Genomic Prediction - Results

Genomic Prediction - Results

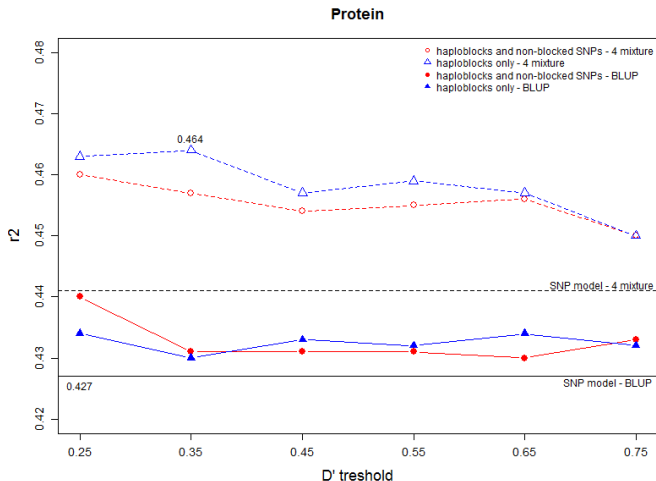


Figure: Correlation between predicted and true breeding values

Genomic Prediction - Results

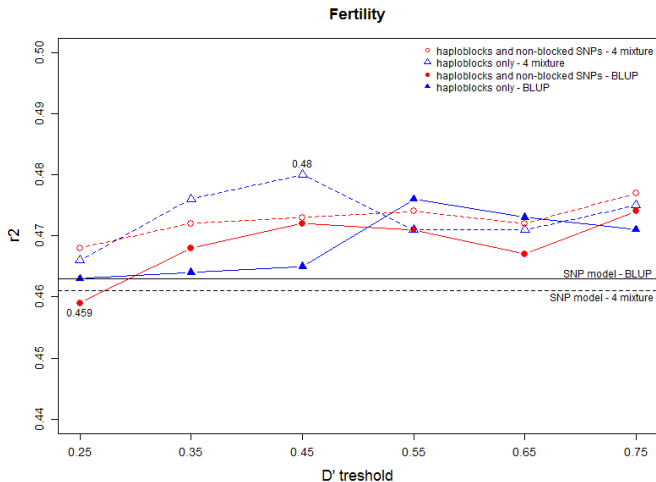


Figure: Correlation between predicted and true breeding values

Genomic Prediction - Results

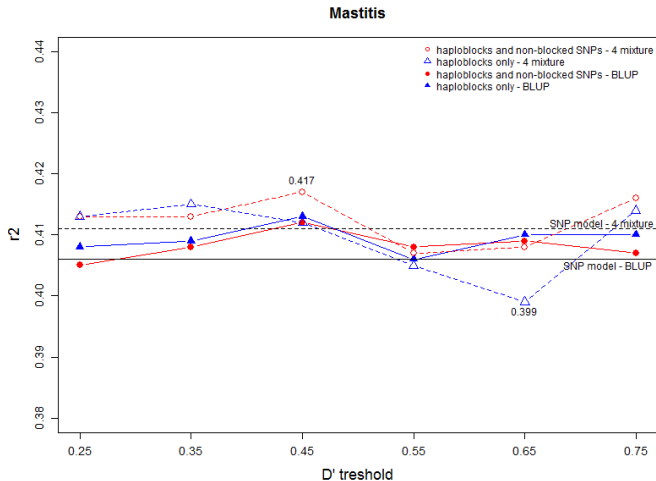


Figure: Correlation between predicted and true breeding values

Some Conclusions and Work in Progress

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;
- 4 mixture models with polygenic effect have superior prediction accuracy than BLUP models;

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;
- 4 mixture models with polygenic effect have superior prediction accuracy than BLUP models;
- Improvement of prediction compared to GBLUP models:

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;
- 4 mixture models with polygenic effect have superior prediction accuracy than BLUP models;
- Improvement of prediction compared to GBLUP models:
 - 3.7% for protein;

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;
- 4 mixture models with polygenic effect have superior prediction accuracy than BLUP models;
- Improvement of prediction compared to GBLUP models:
 - 3.7% for protein;
 - 1.7% for fertility;

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;
- 4 mixture models with polygenic effect have superior prediction accuracy than BLUP models;
- Improvement of prediction compared to GBLUP models:
 - 3.7% for protein;
 - 1.7% for fertility;
 - 1.1% for mastitis.

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;
- 4 mixture models with polygenic effect have superior prediction accuracy than BLUP models;
- Improvement of prediction compared to GBLUP models:
 - 3.7% for protein;
 - 1.7% for fertility;
 - 1.1% for mastitis.
- Next steps:

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;
- 4 mixture models with polygenic effect have superior prediction accuracy than BLUP models;
- Improvement of prediction compared to GBLUP models:
 - 3.7% for protein;
 - 1.7% for fertility;
 - 1.1% for mastitis.
- Next steps:
 - Select haploblocks to perform prediction - haploblocks that contain the main SNP effects from the SNP models;

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;
- 4 mixture models with polygenic effect have superior prediction accuracy than BLUP models;
- Improvement of prediction compared to GBLUP models:
 - 3.7% for protein;
 - 1.7% for fertility;
 - 1.1% for mastitis.
- Next steps:
 - Select haploblocks to perform prediction - haploblocks that contain the main SNP effects from the SNP models;
 - Perform haploblocks models in an across-breed design;

Some Conclusions and Work in Progress

- Haplotype approach presented improvement in genomic prediction from HD marker maps;
- 4 mixture models with polygenic effect have superior prediction accuracy than BLUP models;
- Improvement of prediction compared to GBLUP models:
 - 3.7% for protein;
 - 1.7% for fertility;
 - 1.1% for mastitis.
- Next steps:
 - Select haploblocks to perform prediction - haploblocks that contain the main SNP effects from the SNP models;
 - Perform haploblocks models in an across-breed design;
 - Work on different building methods of haploblock to use in prediction.

Thank You!