A novel method to visualize local LD
-- a study on Danish pig breeds

Lei Wang
David Edwards, Peter Sørenson, Luc Janss and Tage Ostersen
Linkage disequilibrium (LD)

Degree of non-random association of alleles at different loci (correlation between alleles)

LD matrix

SNP1   SNP2   SNP3   SNP1   SNP2   SNP3   ...
P1      1      1      0      1      0.13   0.42   ...
M1      0      1      1      1      1      0.38   ...
P2      0      0      0      0      0.42   1      ...
M2      1      0      1      1      1      1      ...
...
Conventional method to visualize LD in color blocks

It’s difficult to compare LD across breeds or between regions!
Can we quantify LD block as a single value?

Example:

\[ E(r^2) \]

5 markers → 10 LD values
Regress LD ($r^2$) on distance ($d$)

**Sliding window**
- window size: 50, 100, 200 markers
- 1 marker per step

**The model:**

$$r^2_{ij} = 1/(1 + \alpha \times d_{ij})$$

![Diagram]

1. fit $r^2$ and $d$ into the model, estimate $\hat{\alpha}$ for each window
2. substitute $\hat{\alpha}$ and $d$, calculate $E(r^2)$ for each window
3. plot $E(r^2)$ against middle-SNP physical position

"local" effective population size
single value represent LD at each marker locus

Example:

10 $r^2$ values
Data

3 Danish pig breeds:

<table>
<thead>
<tr>
<th>Breed</th>
<th>No. of animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duroc</td>
<td>4249</td>
</tr>
<tr>
<td>Landrace</td>
<td>1979</td>
</tr>
<tr>
<td>Yorkshire</td>
<td>2123</td>
</tr>
</tbody>
</table>

Illumina PorcineSNP60 BeadChip

<table>
<thead>
<tr>
<th>Breed</th>
<th>No. of markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duroc</td>
<td>29567</td>
</tr>
<tr>
<td>Landrace</td>
<td>33719</td>
</tr>
<tr>
<td>Yorkshire</td>
<td>33907</td>
</tr>
</tbody>
</table>
local LD map versus heatmap

Duroc, Chromosome 1, N=100
Compare local LD across breeds

(b)

- Duroc
- Landrace
- Yorkshire

$r^2$

physical position, in Mb

0 50 100 150 200 250 300
Discussion

• The robustness of the method is being tested
  $\alpha$ is population specific

• Interpret $\alpha$ as ”local effective population size”

Tortereau et al, BMC Genomics, 2012
Conclusion

• New method for visualizing local LD developed

• Method facilitates comparison of genetic variation across breeds

• Prospects of the method:
  • Assist genomic prediction across breeds
  • Selection of markers for prediction
Thank you for your attention !!!
Measure Persistence of phase

- correlation of $r$ in two populations
- common markers between two populations

$$cor(r_{A,B}) = \frac{\sum_{(i,j) \in P} (r_{ij}(A) - \bar{r}_A)(r_{ij}(B) - \bar{r}_B)}{S_A S_B}$$

1. $r_{12}$
2. $r_{23}$
3. $r_{34}$
4. $r_{35}$
5. $r_{45}$

Duroc

Landrace
Persistence of Phase over distance

• Persistence of phase:
  • group $r$ the same way as LD
  • calculate correlation of $A(r_{12}, r_{23}, r_{13}, ...)$ with $B(r_{12}, r_{23}, r_{13}, ...)$

• Two scenarios:
  • Scenario I: 17,918 all 3 breeds
  • Scenario II:
    • Duroc-Landrace 22,142;
    • Duroc-Yorkshire 22,347;
    • Landrace-Yorkshire 26,505
local PoP in two scenarios

(a) Scenario I

(b) Scenario II
Data

• **3 breeds**
  • Duroc, Landrace, Yorkshire

• **Illumina PorcineSNP60 BeadChip**
  • Genotype missing rate < 10% (animals)
  • CallRate > 0.9
  • MAF < 0.05
  • Hardy Weinberg Equilibrium \((p < 1 \times 10^{-7})\)

<table>
<thead>
<tr>
<th></th>
<th>Duroc</th>
<th>Landrace</th>
<th>Yorkshire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>4249</td>
<td>1979</td>
<td>2123</td>
</tr>
<tr>
<td>SNP number</td>
<td>29567</td>
<td>33719</td>
<td>33907</td>
</tr>
<tr>
<td>Aver. Space</td>
<td>83 kb</td>
<td>72 kb</td>
<td>72 kb</td>
</tr>
</tbody>
</table>
sliding window size $N$
compare local LD across breeds

(a)

(b)