Profiles of causative SNP in GWAS

Ignacy Misztal, Ivan Pocrnic*, Daniela Lourenco
University of Georgia
*now Roslin Institute
Dimensionality of genomic information

• Farm animals have small effective population size
  • Cattle – 100 to 150
  • Pigs and chickens – around 50

• Limited number of independent chromosome segments
  • 5000 to 15,000
  • 25M SNP compresses between 5000 to 15000
  • Average size of segment 200 kb to 600 kb

• Genomic selection mostly by estimation of chromosome segments
  • With large data – all chromosome segments well estimated - high accuracies of GEBV
Do segments affect GWAS?

• If each segment 200k-500k, SNP may point to segment not QTL
• Usually Manhattan plots shown as 1-5M averages

• If QTL in data, can it be identified by GWAS?

• Assume only 100 QTL per trait
  • If all known and their values known – 100% accuracy
  • If all known and their values estimated – perhaps 99% accuracy
  • If not known, exhaustive search for ~ 100% accuracy
GWAS and sequence data

• What is expected response (profile) of single QTL in Manhattan plots?

• Does it depend on effective population size and amount of data?
Simulation study

• AlphaSimR (Faux et al., 2016)

• 100 causative SNP with same effect and equally spaced
• 10 generations with 2000 animals in each
• Last 3 generations genotyped for 50k SNP – causative SNP included

• 3 data sets
  • Ne=60
  • Ne=600
  • Ne=60_3x --3 times more data (6000 per generation)

• Analyzes by ssGBLUP with p-value option
Manhattan plots for SNP effects

Ne=60

Ne=60 3X

Ne=600
Manhattan plots for P-values

Ne=60

Ne=60 3X

Ne=600
Manhattan plot for first chromosome

Ne=60                                    Ne=60 3X                                             Ne=600
Plots averaged for all causative SNP

Ne=60

Ne=60 3X

Ne=600
Hypothetical profile of causative SNP

Small Ne

Large Ne

"noise"
Profiles of causative SNP in various analyzes

- Unweighted Causative SNP present
- Unweighted Causative SNP absent
- Weighted Correct variance
- Weighted Incorrect variance

“noise”
Cases for data/methods

- Causative SNP likely present in data ➔ weigh single SNP
- Causative SNP likely absent ➔ weigh region

- Methods based on sequential SNP estimation (BayesACR...) ➔ SNPs adjacent to causative SNP reduced or eliminated automatically

- Methods based on joint SNP estimation (NonlinearAB, GBLUP) ➔ require ad-hoc elimination of SNPs adjacent to causative SNP
Comments

• Some peaks in GWAS due to relationships (tagging important individuals)

• Likely few causative SNP with detectable level present

• SNP selection makes commercial multitrait analyses hard
Conclusions

• Profile of causative SNP wide, with sharp peak for causative SNP

• Profile narrower with higher effective populations size
Response to QTL in methods with sequential SNP estimation (e.g., BayesB)

Response to QTL with method with joint estimation (NonlinearA, ssGWAS) and variance elevated for a large region

Profile in SNP BLUP or GBLUP with small (....) and large (---) data

Response to QTL with method with joint estimation (NonlinearA, ssGWAS) and variance elevated for a small region and nulling of adjacent SNP
Response to QTL in SNP BLUP or GBLUP with small (...) and large (---) data.

Response to QTL with method with joint estimation (NonlinearA, ssGWAS) and variance elevated for a large region.

Response to QTL in methods with sequential SNP estimation (e.g., BayesB).

Response to QTL with method with joint estimation (NonlinearA, ssGWAS) and variance elevated for a small region and nulling of adjacent SNP.