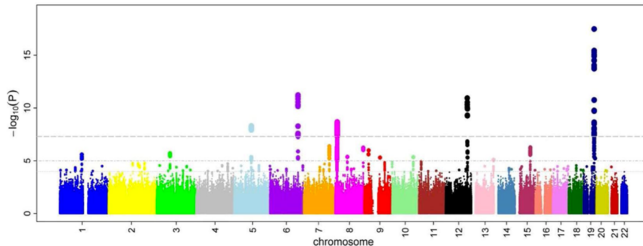


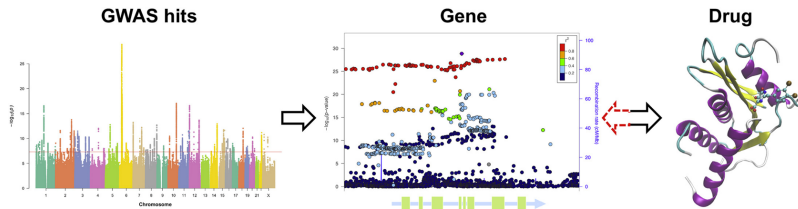
Genome-wide association studies in natural populations: managing expectations and avoiding error

Susan E. Johnston & Lewis G. Spurgin

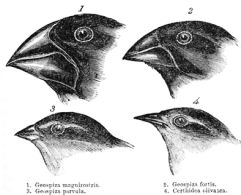


Genome **W**ide **A**ssociation **S**tudy (GWAS)

Experimental design used to detect associations between genetic variants and traits in population samples



Trait	Gene with GWAS hits	Known or candidate drug
Type 2 Diabetes	<i>SLC30A8/KCNJ11</i>	ZnT-8 antagonists/Glyburide
Rheumatoid Arthritis	<i>PADI4/IL6R</i>	BB-CI-amidine/Tocilizumab
Ankylosing Spondylitis(AS)	<i>TNFR1/PTGER4/TYK2</i>	TNF-inhibitors/NSAIDs/fostamatinib
Psoriasis(Ps)	<i>IL23A</i>	Risankizumab
Osteoporosis	<i>RANKL/ESR1</i>	Denosumab/Raloxifene and HRT
Schizophrenia	<i>DRD2</i>	Anti-psychotics
LDL cholesterol	<i>HMGCR</i>	Pravastatin
AS, Ps, Psoriatic Arthritis	<i>IL12B</i>	Ustekinumab



Key factors to consider in GWAS

Key factors to consider in GWAS

- ▶ Have I missed any variants associated with my trait? (**False negatives**)
- ▶ Have I identified any spurious associations with my trait? (**False positives**)

Key factors to consider in GWAS

Factors affecting false positive and false negative rates

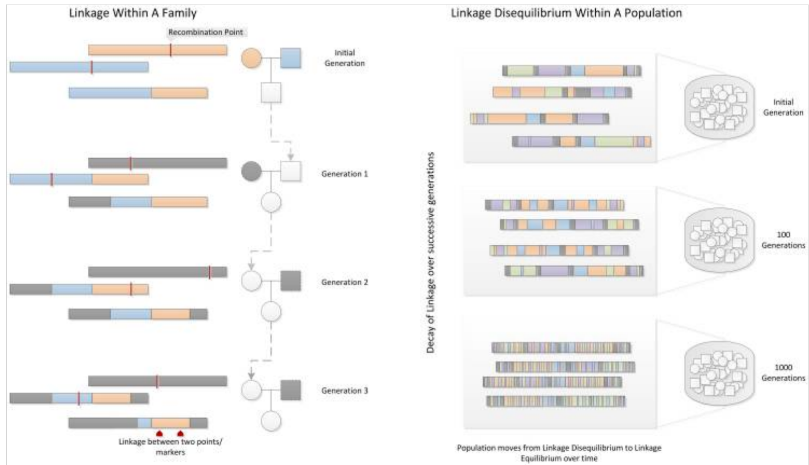
- ▶ Architecture of the trait
- ▶ Number of individuals sampled
- ▶ Number of genetic markers sampled
- ▶ Extent of linkage disequilibrium between sampled markers and causal variants

Key factors to consider in GWAS

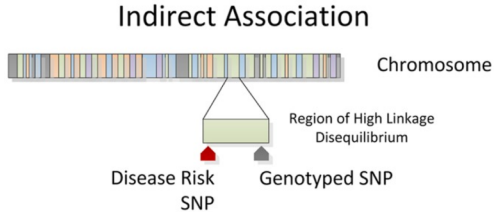
Factors affecting false positive and false negative rates

- ▶ Architecture of the trait
- ▶ Number of individuals sampled
- ▶ Number of genetic markers sampled
- ▶ **Extent of linkage disequilibrium between sampled markers and causal variants**

Linkage disequilibrium (LD) and GWAS



Linkage disequilibrium (LD) and GWAS



Bush and Moore 2012, PLoS Comput. Biol.

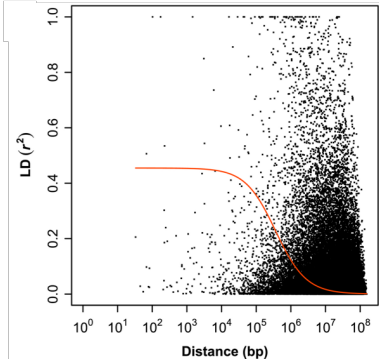
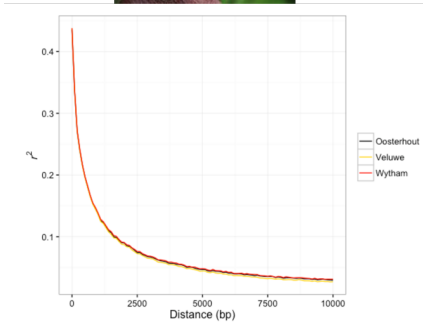
Linkage disequilibrium (LD) and GWAS

Present day patterns of LD in a population are the product of:

1. Historical evolutionary processes (e.g. recombination rate, demography, selection etc. - **Ancestral LD**)
2. Nature of the population sample (e.g. family structure, substructure, sample size - **Sample LD**)

Both ancestral LD and sample LD will vary widely across studies of natural populations

Linkage disequilibrium (LD) and GWAS



GWAS in natural populations

- ▶ The power GWAS will vary enormously across natural populations and species with different evolutionary histories
- ▶ Sample size (number of individuals/markers) will therefore need to be adjusted accordingly
- ▶ This should be considered **before** undertaking sampling and genotyping

GWAS in natural populations: a simulation approach

Study aims

Develop a simulation and analysis pipeline to:

- ▶ Illustrate some of the key issues affecting the power and promise of GWAS in natural populations
- ▶ Allow other researchers to estimate power in their system before undertaking costly fieldwork/genotyping

Simulation approach