A combined coalescence forward in time simulator software for pedigreed populations undergoing selection for complex traits.

Jeremy T. Howard[1] Franceso Tiezzi[1] Jennie E. Pryce[2,3]
Christian Maltecca[1]

[1]North Carolina State University
[2]Department of Economic Development, Victoria, Australia
[3]La Trobe University, Australia

jthoward@ncsu.edu

July 20, 2016

Current State of Simulation Programs.

NC STATE UNIVERSITY

Introduction

A large number of simulation programs have been developed that are suitable for testing alternative selection and/or mating strategies that are primarily based on the additive genetic effects for a quantitative trait.

- QMsim (Sargolzaei and Schenkel, 2009).
- AlphaSim Suite (Hickey et al., 2014).
- ms2gs (Pérez-Enciso and Legarra, 2016).
- FREGENE (Chadeau-Hyam et al., 2008).
- XSim (Cheng et al. 2015)
- etc.....

Introduction

In the context of animal breeding there is currently a lack of simulation programs tailored towards:

- Identifying "best practice" management decisions to manage a population at the genetic level in the form of:
  - Genetic Diversity
  - Fitness Effects
  - Additive and Dominance Effects
- The optimal use of dense marker information to manage a population at the genomic for populations that are routinely genotyped.

Introduction

Due to this we have developed a simulation tool that:

- Generates quantitative and/or fitness traits with additive and dominance effects.
- Utilizes computationally efficient routines to generate dense marker based relationship matrices and their associated inverse.

Geno-Diver Overview

NC STATE UNIVERSITY

Geno-Diver Overview

Historical Population

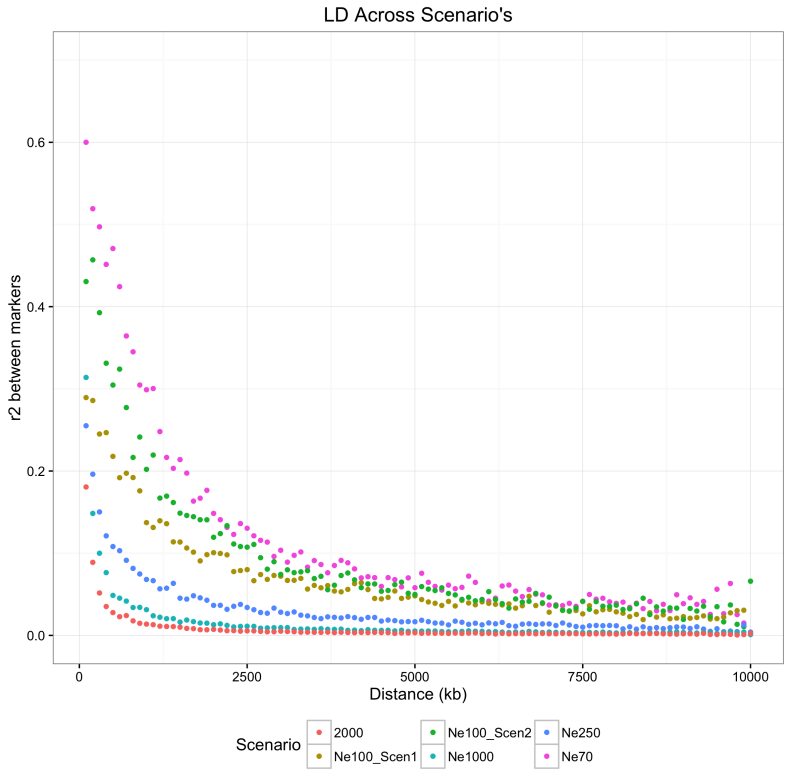
- Use MaCS (Chen et al. 2009) to generate founder sequences.
- Generate QTL architecture based on founder sequences.

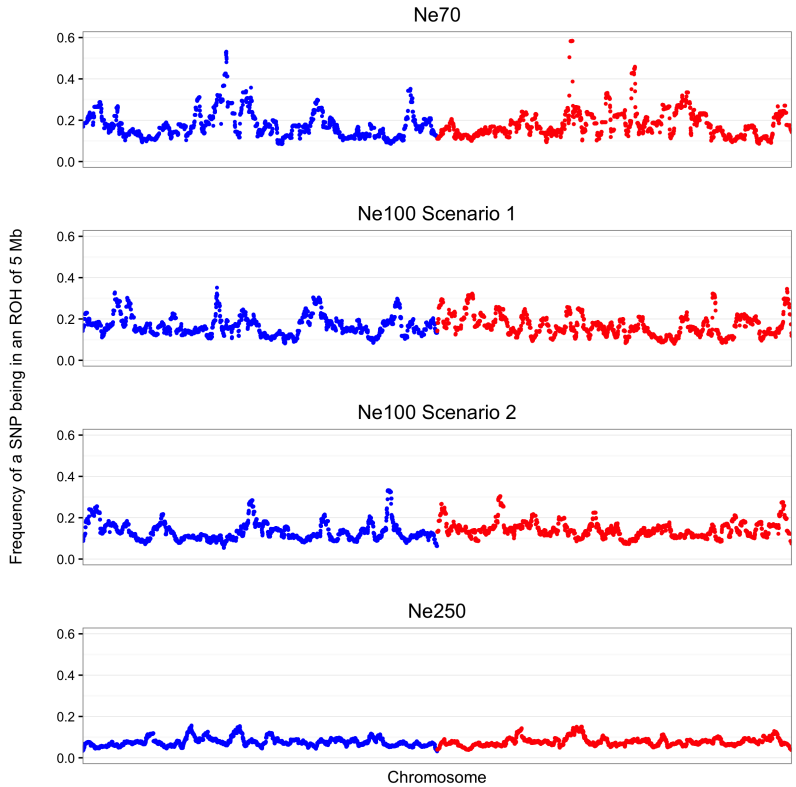
Recent Population

- Select progeny.
- Generate gametes.
- Generate progeny.
- Cull parents.

Historical Population

- The founder population is created by calling MaCS which allows for a variety of linkage disequilibrium (LD) decay scenarios to be generated.





Geno-Diver Overview

Genetic Architecture

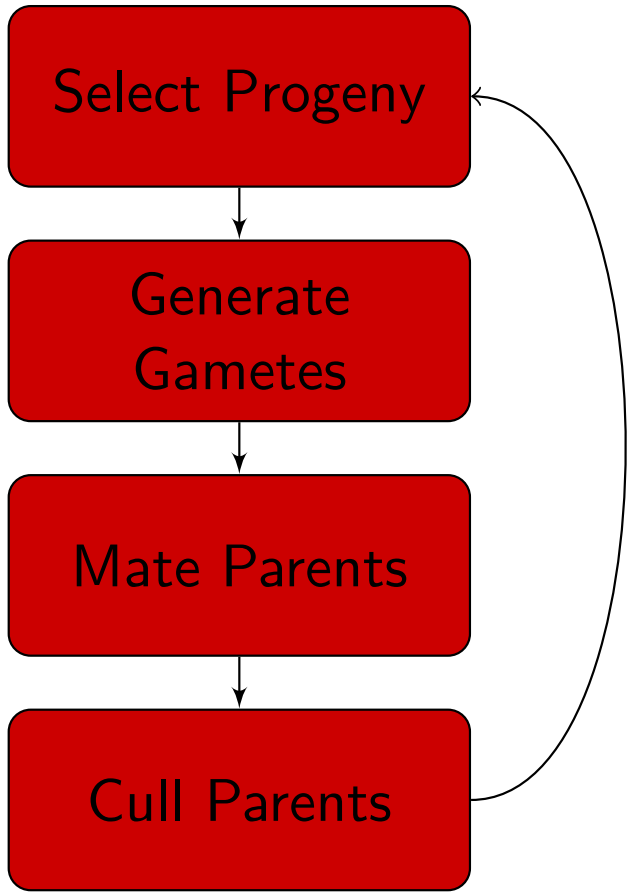
- After the founder sequences have been created QTLs are assigned to a random set of SNP
  - Quantitative Trait (Quan):
    - ★ Additive effect (Add): Gamma with equal proportion positive or negative.
    - ★ Dominance effect (Dom): |Additive effect| * Degree of Dominance (Normal).
  - Fitness Trait (Fit):
    - ★ Lethal (higher "s", lower "h").
    - ★ Sub-lethal (lower "s", higher "h").
    - ★ Animal survives to breeding if multiplicative fitness value is greater than a uniform deviate ranging from 0 to 1.
  - Relationship between Quantitative and Fitness effects
    - ★ Proportion with quantitative and fitness effects.
    - ★ Correlation.

$$Quan(Add) \leftrightarrow Quan(Add + Dom) \leftrightarrow Quan(Add + Dom) + Fit \leftrightarrow Fit$$

Geno-Diver Overview

Recent Population

- Selection and culling based on either estimated breeding value (EBV), true breeding value, phenotype or random.
- EBV generated from pedigree or genomic relationship matrix.
- A marker array is generated from SNP that aren't QTL.
- Multiple options are available to make the simulation more realistic:
  - Maximum number of full-sibling kept within a family.
  - Differential mate allocation by age of sire.
  - Avoidance matings.



Summary Statistics

A number of summary statistics are created within each generation relating to:

- LD decay metrics.
- Mean phenotype and genetic values.
- QTL frequency.
- Number of founder or new mutations fixed or still segregating.
- Inbreeding metrics based on genomic and pedigree.
- Mean number of lethal or sub-lethal genotypes in the homozygous or heterozygous state.
- Mean fitness value of an animal and lethal equivalents.

Geno-Diver Overview

Computing Procedures

- Intel MKL libraries for matrix multiplications
- Allows for multithreading.
- Generates SNP-by-SNP relationship matrices based on strategies outlined by Aguilar et al. (2011).
- Generates inverse by updating previous generation based on either Meyer et al. (2013) or Misztal et al. (2016).
- Input is sequence information and has been tested for 1,000,000+ marker panel.

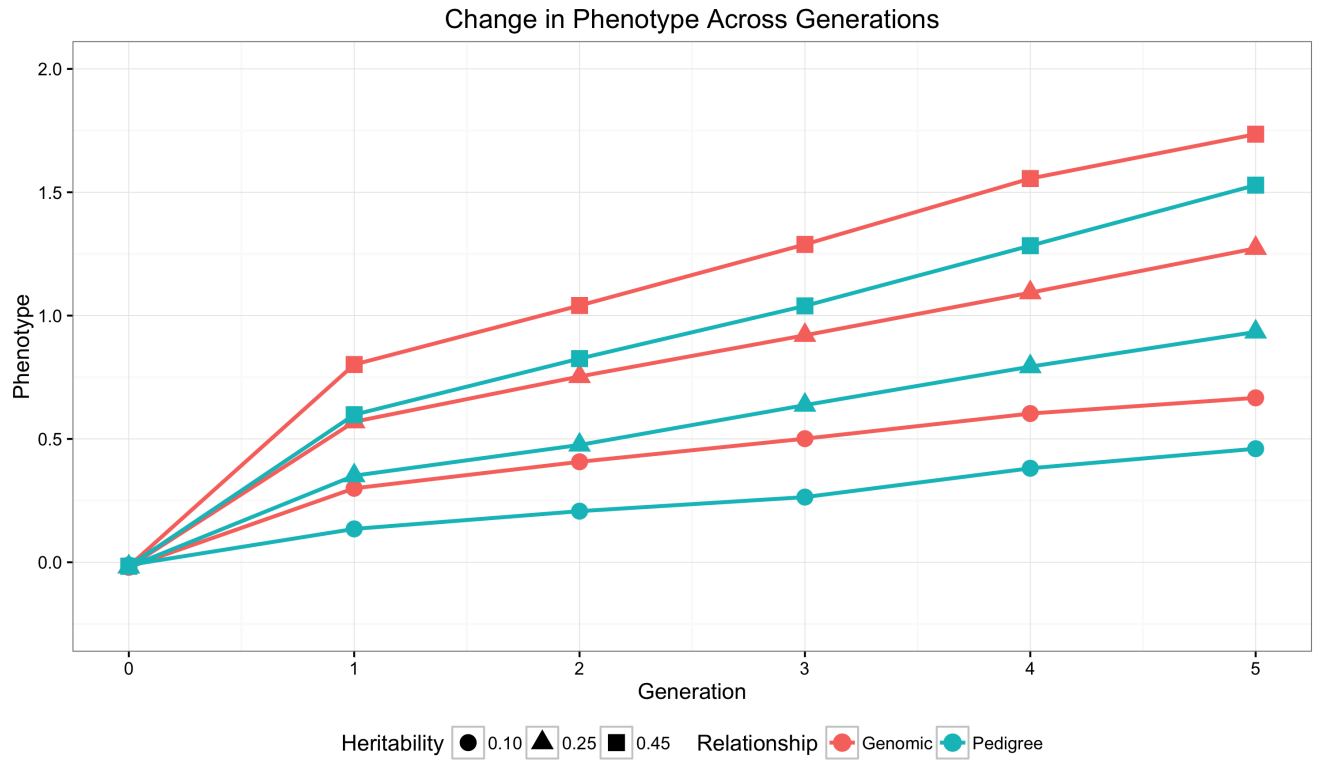
Simulation Examples

NC STATE UNIVERSITY

Examples

Comparison of Pedigree Versus Genomic BLUP Selection

- After sequences were generated the average time to simulate 5 generations:
  - Pedigree: 1.52 minutes.
  - Genomic: 2.55 minutes.

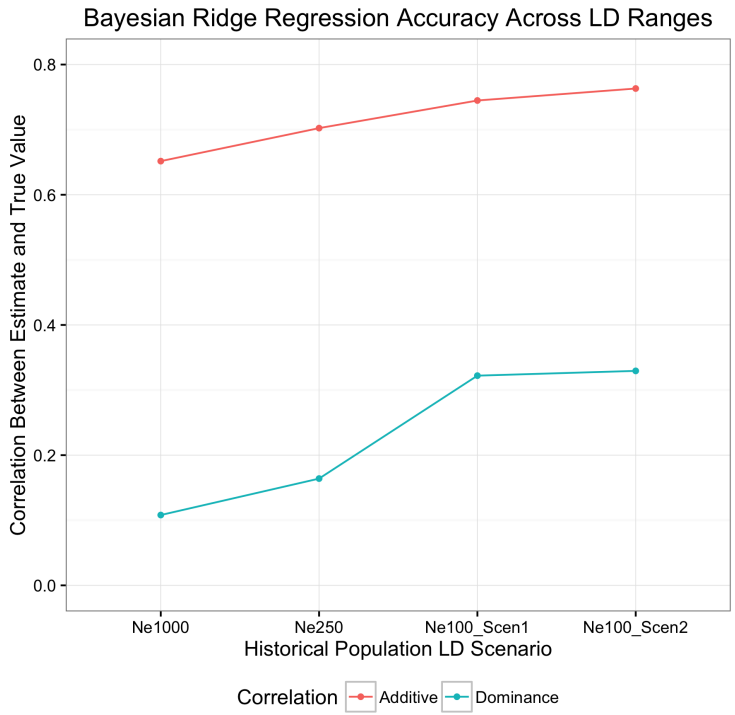


Examples

Cattle Example

- Example of prediction accuracy across different LD ranges.
- Trained on Generation 7,8 and 9 and Predicted on 10.

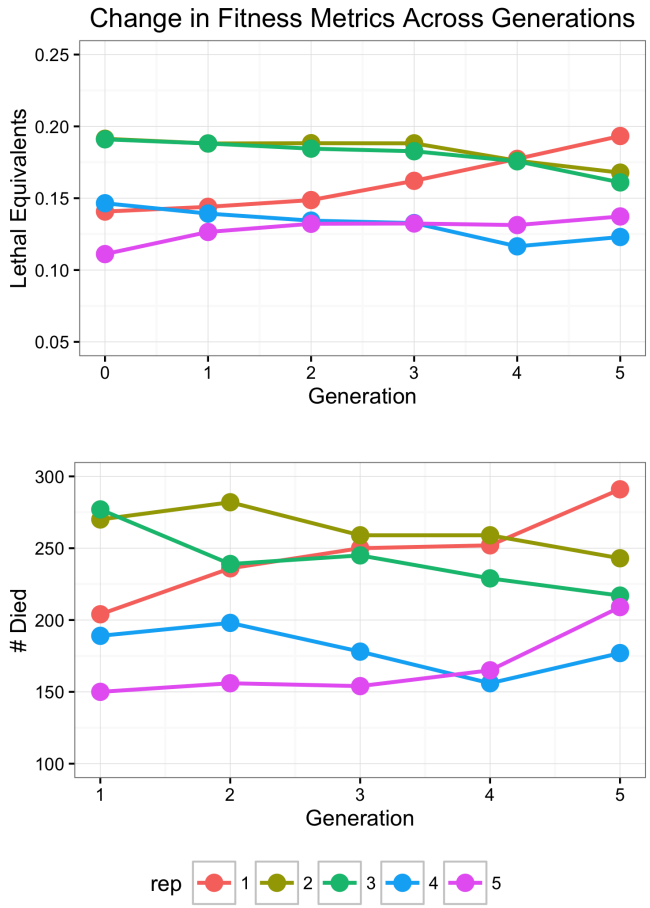
```
-----|    Parameter file for Geno-Diver    |-----
-----|              Starting Point         |-----
START: founder
OUTPUTFOLDER: GenoDiverFiles
SEED: 1500
NTHREAD: 10
-----| Genome and Marker Information  |-----
CHR: 3
CHR_LENGTH: 150 150 150
QTL: 50
QUANTITATIVE_MAF: 0.075
NUM_MARK: 4000 4000 4000
-----|      Population Characteristics      |-----
FOUNDER_Effective_Size: Ne100_Scen2
VARIANCE_A: 0.35
VARIANCE_D: 0.05
-----| Selection and Mating Parameters  |-----
GENERATIONS: 10
INDIVIDUALS: 50 0.2 600 0.2
PROGENY: 1
SELECTION: ebv high
SOLVER_INVERSE: pedigree pcg cholesky
MATING: random5
CULLING: ebv 5
-----|            Output Options           |-----
OUTPUT_LD: yes
GENOTYPES: yes 0
```



Swine Example

- Overview of Fitness Based Statistics Across Generations.

```
-----|    Parameter file for Geno-Diver    |-----
-----|              Starting Point         |-----
START: founder
OUTPUTFOLDER: GenoDiverFiles
SEED: 1500
NTHREAD: 10
-----| Genome and Marker Information  |-----
CHR: 3
CHR_LENGTH: 150 150 150
QTL: 50
FIT_LETHAL: 25
FIT_SUBLETHAL: 100
QUANTITATIVE_MAF: 0.075
NUM_MARK: 4000 4000 4000
-----|      Population Characteristics      |-----
FOUNDER_Effective_Size: Ne100_Scen2
VARIANCE_A: 0.20
VARIANCE_D: 0.05
COVAR: 0.5 0.2
-----| Selection and Mating Parameters  |-----
GENERATIONS: 5
INDIVIDUALS: 100 0.2 600 0.2
PROGENY: 8
MAXFULLSIB: 2
SELECTION: ebv high
SOLVER_INVERSE: pedigree pcg cholesky
MATING: random5
CULLING: ebv 5
-----|           Output Options            |-----
OUTPUT_LD: yes
GENOTYPES: yes 0
```



Modules to Incorporate in the Future

NC STATE UNIVERSITY

Modules to Incorporate in the Future

- Minimize relationships across a variety of relationship matrices.
- Simulate sex-limited traits.
- Incorporate real genotype data as a tool to manage the population.
- Incorporate external breeding value predictions.
- Incorporate the use of advanced reproductive technologies.

Source Code and Executable

- Source code for Geno-Diver can be found at:
  - "https://github.com/jeremyhoward"
- A linux executable is available.
- A comprehensive user manual with examples are also included.
- Any questions or inquiries can be directed to:
  - "jthoward@ncsu.edu"

Questions?

NC STATE UNIVERSITY

Citations

- Sargolzaei and Schenkel. 2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics, 25:680-681.
- Hickey et al. 2014. AlphaMPSim: flexible simulation of multi-parent crosses. Bioinformatics, 30:2686-2688.
- Pérez-Enciso and Legarra. 2016. A combined coalescence gene-dropping tool for evaluating genomic selection in complex scenarios (ms2gs). Journal of Animal Breeding and Genetics, 133:85-91.
- Chadeau-Hyam et al. 2008. Fregene: Simulation of realistic sequence-level data in populations and ascertained samples. BMC Bioinformatics 9:364.
- Cheng et al. 2015. XSim: Simulation of Descendants from Ancestors with Sequence Data. G3 (Bethesda). 5(7):1415-1417.

Citations

- Aguilar et al. 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. J. Anim. Breed. Genet. 128: 422?428.
- Meyer et al. 2013. Technical note: Updating the inverse of the genomic relationship matrix. J. Anim. Sci. 91:2583?2586.
- Misztal et al. 2014. Using recursion to compute the inverse of the genomic relationship matrix.