Breeding against Infectious Disease

Estimating Gene Effects on Susceptibility and Infectivity

EAAP 2016 Belfast
Floor Biemans
Mart de Jong
Piter Bijma
floor.biemans@wur.nl
Breeding against Infectious Disease

- Overall objective: Reduce prevalence
  - Prevalence = fraction of the population infected

- Current approach: \( y = Xb + Za + e \)
  - Connect disease status to own breeding value
  - Captures genetic variation in host susceptibility only

- We miss part of the genetic variation: Host Infectivity
  - Infectivity = propensity to infect others
Traits affecting prevalence

- $R_0$ determines prevalence

- $R_0$: Basic reproduction ratio
  - $R_0 = \text{"Number of new cases due to a case"}$

- Anche et al. 2012: $R_0 = c \times \text{susceptibility} \times \text{infectivity}$

Reduce prevalence → we should also consider infectivity
Objective

- Develop methods to estimate single gene-effects on susceptibility and infectivity

**Challenges:**

- Infectivity:
  - Generalized linear models

- 0/1 trait (disease status)
  - Generalized linear models

- Time dynamics
  - Time-series data
Epidemiological & Genetic Model

- **Endemic disease**
- **SIS-model**
  - Susceptible – Infected - Susceptible
- **Genetic model**
  - 2 loci, each with 2 alleles
  - Susceptibility locus; alleles $g$ and $G$
    - Three genotypes: $gg$, $gG$, $GG$
  - Infectivity locus; alleles $f$ and $F$
    - Three genotypes: $ff$, $fF$, $FF$
Simulated data

- 10 herds of ~100 individuals
- Within-herd endemics
- 11 observation moments per herds
  - Time series data on disease status (0/1 = S/I)
Simulated time-series data

Individual disease status (0/1) recorded at each time point
Data

- Disease status of each individual at each time point (0/1 = S/I)
  - Which susceptibles have become infected
  - Which infecteds (may) have done it
- Genotyped individuals
- Length of the time interval
Methods

Model the probability of transmission in a time interval: $P(S \rightarrow I)_{t \rightarrow t+1}$

- Binary data from a Poisson process
  - GLM with complementary log-log link function

- Transmission probability depends on:
  - Overall average transmission rate ($c$)
  - Susceptibility genotype of the (focal) individual
  - Number of infectious herd-mates at time $t$
  - Infectivity genotype of those herd mates
Results: Generalized Linear Model

\[ cloglog \left( \varepsilon \left( \frac{c}{s} \right) \right) = c_0 + c_1 \times \text{Index}G + c_2 \times \text{fraction}F + \log \left( \frac{I_{tot}}{N} \right) \times \Delta t \]

GLM with complementary log-log link-function
Results: Generalized Linear Model

cloglog \left( \varepsilon \left( \frac{C}{S} \right) \right) = c_0 + c_1 \cdot \text{IndexG} + c_2 \cdot \text{fractionF} + \log \left( \left( \frac{I_{tot}}{N} \right) \cdot \Delta t \right)

Expected number of cases for each susceptibility genotype, in the interval \( t \to t+1 \)
Results: Generalized Linear Model

\[
cloglog \left( \varepsilon \left( \frac{C}{S} \right) \right) = c_0 + c_1 \cdot \text{IndexG} + c_2 \cdot \text{fractionF} + \log \left( \left( \frac{I_{\text{tot}}}{N} \right) \cdot \Delta t \right)
\]

Susceptibility genotype of susceptibles at time t

= Allele count (0, 1, or 2)
Results: Generalized Linear Model

\[ cloglog \left( \varepsilon \left( \frac{C}{S} \right) \right) = c_0 + c_1 \times \text{IndexG} + c_2 \times \text{fractionF} + \log \left( \frac{I_{tot}}{N} \right) \Delta t \]

Infectivity genotype of infectious herd mates at time t

= Average allele count of those herd mates
Results: Generalized Linear Model

$$cloglog \left( \varepsilon \left( \frac{c}{s} \right) \right) = c_0 + c_1 \cdot IndexG + c_2 \cdot fractionF + \log \left( \left( \frac{I_{tot}}{N} \right) \cdot \Delta t \right)$$

Offset

- Fraction of herd mates infected at time $t$
- Length of the time interval
Results: Generalized Linear Model

\[
cloglog \left( \varepsilon \left( \frac{C}{S} \right) \right) = c_0 + c_1 \times \text{IndexG} + c_2 \times \text{fractionF} + \log \left( \frac{I_{\text{tot}}}{N} \right) \Delta t
\]

Solutions: \( \hat{c}_0, \hat{c}_1, \hat{c}_2 \)

Estimates of interest:

- Susceptibility effect G-allele = \( e^{\hat{c}_1} \)
- Infectivity effect F-allele = \( e^{\hat{c}_2} \)
Results: Estimates (at optimum recording interval)
Results: effect recording interval

For a fixed total number of recordings

Optimum

Estimated effect

Recording interval (% average infectious time)
Conclusions

- Generalized LM for GWAS
  - Susceptibility estimates are unbiased
  - Infectivity estimates tend to be biased downwards (conservative)
  - Optimum recording interval ~1/3 of infectious period

- Ongoing
  - Application to digital dermatitis in dairy cattle (Mortellaro’s disease)

- Extensions
  - Mixed models and Genomic Prediction

- Alternatives: Bayesian models (Anacleto et al. 2015)