A 2-step strategy to infer genome-wide associations for endoparasite traits in local DSN cattle

Katharina May¹,², Carsten Scheper¹, Kerstin Brügemann¹, Tong Yin¹, Christina Strube², Paula Korkuć³, Gudrun A. Brockmann³, Sven König¹

¹Institute of Animal Breeding and Genetics, Justus-Liebig-University of Gießen, 35390 Gießen, Germany,
²Institute for Parasitology, University of Veterinary Medicine Hannover, Buenteweg 17, 30559 Hanover, Germany,
³Department for Animal Breeding Biology and Molecular Genetics, Faculty of Live Science, Humboldt-Universität of Berlin, 10115 Berlin, Germany
Endoparasite infections in dairy cows

• Infections with endoparasites lead to reduced milk production and detrimental impacts on fertility (e.g., Charlier et al., 2005; Dank et al., 2015)
  → high economic losses

• Increase of anthelmintic resistance (e.g., Gasbarre et al., 2014; Sutherland et al., 2011)

• Increasing interest in local breeds being best adapted to various environmental conditions
Why genomic studies in endangered DSN cattle?

- "Deutsches Schwarzbuntes Niederungsrind" /Black Pied cattle
- One of the founding breeds of Holstein Friesian

Better fertility and productive life (Biedermann et al., 2005)

More resistant to metabolic disease (Jaeger et al., 2016)

More robust under harsh environmental conditions (Al-Kanaan et al., 2016)

Less susceptibility to endoparasite infections?
Background and preliminary work

• Faecal examination of 1166 dairy cows (2006 observations), 17 herds
• 4 different Black and White cattle selection lines
• HF milk, HF pasture, HF New Zealand, dual-purpose cows (DSN)

1. Quantitative-genetic parameters for endoparasite traits:

<table>
<thead>
<tr>
<th>Endoparasite trait</th>
<th>$h^2 \pm SE$</th>
<th>$r_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastrointestinal nematodes (GIN)</td>
<td>0.05 – 0.06 ± 0.04</td>
<td>1.00</td>
</tr>
<tr>
<td>Lungworm larvae (<em>Dictyocaulus viviparus</em>)</td>
<td>0.05 ± 0.04</td>
<td>-0.10</td>
</tr>
<tr>
<td>Liver fluke (<em>Fasciola hepatica</em>)</td>
<td>0.33 ± 0.06</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Genomic scales for endoparasite traits

2. Genome-wide association study (GWAS) for endoparasite traits in DSN

a) Identification of SNP markers being associated with endoparasite resistance

b) Analysis of potential candidate genes

c) Pathway analyses
Selective genotyping approach

• Collection of blood samples from 148 dual-purpose DSN cows
• 3 herds

Selection criteria:

1. Herd prevalence for gastrointestinal nematodes

2. Selection of most resistant and most susceptible cows within herds
   • 50% of cows being infected
   • 50% being not infected

3. Minimization of genetic relationships within and between groups
Genotyping, Imputing and Quality control

148 cows
587,615 SNP

Genotyped with BovineSNP50 Bead Chip
Imputed to Illumina HD Bead Chip level (700 k)

Exclusion of SNP markers

MAF < 0.05%
Call rate per individual < 95%
Call rate per SNP < 95%
Deviation from Hardy-Weinberg-Equilibrium ($P < 10^{-6}$)

148 cows
423,654 SNP

Blood sampling

Software PLINK, Purcell et al., 2007
Pre-correction of endoparasite traits

1. Calculation of residuals for endoparasite traits using the whole dataset of 1166 cows

\[ y_{ijklm} = \mu + \text{Farm}_i + \text{Parity}_j + \text{GenLin}_k + \text{SP}_l + \text{Lstage}_m + e_{ijklm} \]

- \( y_{ijklm} \) = observations for endoparasite traits (n = 2006)
- \( \mu \) = overall mean effect
- \( \text{Farm}_i \) = fixed effect of the \( i^{th} \) farm
- \( \text{Parity}_j \) = fixed effect of the \( j^{th} \) parity number
- \( \text{GenLin}_k \) = fixed effect of the \( k^{th} \) genetic line
- \( \text{SP}_l \) = fixed effect of the \( l^{th} \) sampling period
- \( \text{Lstage}_m \) = fixed effect of the \( m^{th} \) lactation stage
- \( e_{ijklm} \) = random residual effect
Association analysis in GCTA (Yang et al., 2011)

2. Implementation of the GWAS in GCTA using residuals

\[ y_i = a_i + \text{SNP}_i + e_i \]

- \( y_i \) = corrected phenotypic values (residuals of endoparasite traits)
- \( a_i \) = random polygenetic additive-genetic effect on the basis of genomic relationship matrix
- \( \text{SNP}_i \) = fixed single-locus SNP effect
- \( e_i \) = random residual effect
Manhattan-Plot for gastrointestinal nematodes

- Bonferroni-corrected genome-wide significance threshold ($p = 4.47 \times 10^{-7}$)
- Suggestive candidate threshold ($p = 1e-04$)

- 68 variants reached suggestive candidate threshold ($p_{\text{Cand}} = 1 \times 10^{-4}$)
- 11 candidate genes
Manhattan-Plot for liver flukes (*F. hepatica*)

- Bonferroni-corrected genome-wide significance threshold ($p = 4.47 \times 10^{-7}$)
- Suggestive candidate threshold ($p = 1e-04$)

- 3 variants reached $p_{\text{Bonf}}$ and 53 variants reached $p_{\text{Cand}}$
- 7 candidate genes
Gene annotation: Liver fluke (*F. hepatica*)

- Candidate threshold ($p = 1e-04$)
- Bonferroni corrected genome-wide threshold ($p = 4.47 \times 10^{-7}$)

Activated leukocyte cell adhesion molecule (*ALCAM*) gene

Immune response mechanisms, e.g. cell adhesion interactions of T cells
Pathway-analyses (DAVID, Huang et al., 2008)

Cell adhesion interactions of T cells

KCNJ3 (BTA2)
EGFR (BTA 22)

CDH2 (BTA 24)

Important functions in cellular immune response, involved in host defense mechanisms during infections with *Neospora caninum*
SNP effect correlations within endoparasite traits

ROI für GIN

ROI for liver flukes

ROI for lungworms

SNP effect correlation

-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0

rFEC-GIN  rFEC-FH  rFLC-DV

2: 41,311,137 - 41,496,744 (n = 50)

24: 61,563,972 - 61,797,092 (n = 53)

1: 50,282,272 - 50,608,841 (n = 72)

17: 19,679,366 - 19,858,282 (n = 41)

5: 101,531,511 - 101,668,950 (n = 24)

13: 32,494,524 - 32,694,278 (n = 28)
... and between endoparasite and test-day traits
Conclusions

• The 2-step approach using pre-corrected phenotype data based on a larger dataset is a valid strategy for a small number of phenotypes from small-sized cattle populations

• 53 potential candidate genes were related to endoparasite resistance

• 7 pathways associated with immune response mechanisms to endoparasite infections or involved in host-pathogen interactions for candidate genes located on BTA 1, 2, 6, 13, 22, 24 and 28

• Joint genetic basis for the two nematodes GIN and D. viviparus (high correlations on genetic and genomic scales, same genes (NAV3) and pathways)

• Negative correlations between GIN or F. hepatica and SCS within most genomic regions
We acknowledge the financial support for this project provided by transnational funding bodies, being partners of the FP7 ERA-net project, CORE Organic Plus, and the cofund from the European Commission.

We gratefully thank the H. Wilhelm Schaumann Stiftung for the provision of a PhD grant.
Manhattan-Plot for bovine lungworms

- Bonferroni-corrected genome-wide significance threshold ($p = 4.47 \times 10^{-7}$)

- Suggestive candidate threshold ($p = 1e-04$)

- 41 variants reached $p_{\text{Bonf}}$ and 311 variants reached $p_{\text{Cand}}$

- 35 candidate genes
Genotyping and Imputing

- Genotyped with BovineSNP50 Bead Chip (50k SNP chip)
- Imputing to Illumina HD Bead Chip level (i.e. 700,000 SNPs) using a multi-breed reference panel of 2188 animals

- The reference panel comprised:
  - 48 DSN animals genotyped with the Illumina HD Bead Chip array (Illumina Inc., San Diego, CA, USA)
  - 2140 sequenced animals from the 1000 bull genome project database (Daetwyler et al., 2014) downscaled to Illumina HD Bead Chip density including 30 sequenced DSN animals
Immune response mechanisms during endoparasite infections

- Activated T lymphocyte
- Cytokines
- Antigen
- Interleukin
- B lymphocyte
- Marked antigen
- Antibody or immune globulines
- Plasma cell: antibody production
- B cell development
- IgG1, IL4, IL10
- IgA, IgG, IgM
- IgE

- B cell development
- Antibody or immune globulines
PCA for the three herds
PCA plotted by GIN
PCA plotted by liver flukes (*F. hepatica*)