Supplementary Material

# Supplementary Material and Methods

We used single trait linear random regression models (STM) to estimate the genetic parameters for host tolerance to different parasites and the genetic parameters for body weight across a parasite burden trajectory. The model was previously described in the Material and Methods, and the analysis was accomplished through Bayesian methods by using Gibbs sampler.

For the STM, the prior assumptions and distributions were:

*y* was assumed as , where: *G0* is the covariance matrix for the genetic additive effects, which were previously described in the Material and Methods; R1 is the matrix of the residual effects, from which was considered homogeneity of variance so that . *X* and *Z* are the incidence matrices for the systematic (*b*) and genetic additive effects (*a*), respectively; *N* is the normal distribution; and *I* is an identity matrix with order equal to the number of observations.

The prior distribution for the effects were: ; ; and ; and Scaled inverse chi-squared distributions were assumed for () and (, where A is the relationship matrix, , and , are the hyperparameters for the scaled inverse chi-squared distributions. Non-informative priors were used. Information about a posteriori complete conditional distributions are also provided in Sorensen and Gianola [1].

Multiple trait linear random regression models (MTM) were used to estimate the genetic correlation between the same parasite (Ticks, gastrointestinal nematodes, and *Eimeria* spp.) in different ages. For this, the ages were grouped in two-trait analysis. The MTM assumptions can be described as: and ; where is the expectation of , is the vector of traits, is the incidence matrix for the systematic effects; is the vector with solutions for the systematic effects and ; is the covariance matrix among traits for additive genetic effect - previously described at Material and Methods. R2 is the covariance matrix of residual effects among traits, in which is:

where  are the residual variances for each evaluated trait; is the covariance between the residual effects of each trait. For the covariance matrices G1 and R2 for genetic and residual effects, respectively, inverse Wishart distributions were assumed and and , where and are the hyper parameters of the inverse Wishart distribution and and are degrees of freedom of inverse Wishart distribution of genetic and residual random effects, respectively. Information about a posteriori complete conditional distributions are provided in Sorensen and Gianola [1].

To estimate the genetic variances for intercept and slopes of a same age (, and the covariance between them , the samples obtained in the different two trait analysis were grouped, and the means and high posterior density intervals with 90% of samples (HPD90) were calculated based in the grouped samples. The covariances between intercepts and slopes of different ages () and () were directly calculated from the output of the analysis, for each two by two combination.

For both STM and MTM, the samples of the complete conditional distributions were obtained through the Gibbs sampler using the software GIBBS3F90 [2], with a chain of 1100000 iterations, discard of the 100000 first and sampling each 100 cycles. All the analysis were processed in sagarana HPC cluster, CEPAD-ICB-UFMG. The chain length was defined according to the method of Raftery and Lewis [3] in preliminary analysis, which is available in the BOA package [4] of the software R [5]. The convergence of the chains for each parameter of the model was evaluated by the criteria of Geweke [6] and Heidelberger and Welch [7], which are available in the same software and by visual inspection of the sampled values. For each parameter of the models the posterior means and high posterior density intervals with 90% of samples (HPD90) were calculated. The HPD90 was considered as a measure of uncertainty of the parameter estimate.

For the estimates obtained through MTM, covariances for intercept and HT to each parasite in each ED were obtained in more than on analysis. To calculate the posterior mean and HPD90 of these parameters, the samples of each analysis were grouped, so that the parameters were obtained based on 40000 samples.

# Supplementary references

1. Sorensen D, Gianola D. Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Springer-Verlag; 2002.

2. Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. Athens: University of Geogia; 2015.

3. Raftery AE, Lewis SM. One long run with diagnostics: implementation strategies for Markov Chain Monte Carlo. Stat Sci. 1992;7:493–7.

4. Smith BJ. Bayesian output analysis program (BOA) version 1.1 user’s manual. Dept Biostat Univ Iowa Coll Public Health [Internet]. 2005 [cited 2016 Oct 25]; Available from: http://www.public-health.uiowa.edu/boa/BOA.pdf

5. R Core Team. R: a language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2019 [cited 2016 Oct 25]. Available from: https://www.R-project.org/

6. Geweke J. Evaluation the accuracy of sampling-Based Approaches to the calculation of posterior moments. Oxford, UK: Oxford University Press; 1991.

7. Heidelberger P, Welch PD. Simulation run length control in the presence of an initial transient. Oper Res. 1983;31:1109–44.

# Supplementary Figures

**Supplementary Figure 1.** Distributions of body weight information (BW-a), ticks (TICK-b), gastrointestinal nematodes eggs (GIN-c), and *Eimeria* spp. oocysts (EIM-d) counts in each measurement event (ME). 331, 385, 443, 498, and 555 represent the mean ages of the animals in each ME.

**Supplementary Figure 2.** Genetic additive variances () and heritability () estimates for body weight (BW) in differently infested environments. The parasite burden levels represent the observed median counts of ticks (TICK), gastrointestinal nematodes eggs (GIN), and *Eimeria* spp. oocysts (EIM) observed in each measurement event (ME). 331, 385, 443, 498, and 555 represent the mean ages in days that animals had in each ME.

**Supplementary Figure 3.** Manhattan plots for the genome-wide association studies for body weight measured at different measurement events (ME). 331, 385, 443, 498, and 555 are the mean ages (in days) that the animals had at the moment of evaluation. The dotted line (y=5.64) indicates the threshold for statistical significance. The dashed line (y=4.00) indicates the threshold for suggestive evidence of association.

**Supplementary Figure 4.** Manhattan plots for the genome-wide association studies for host tolerance to ticks evaluated at different measurement events (ME). 331, 385, 443, 498, and 555 are the mean ages (in days) that the animals had at the moment of evaluation. The dotted line (y=5.64) indicates the threshold for statistical significance. The dashed line (y=4.00) indicates the threshold for suggestive evidence of association.

**Supplementary Figure 5** Manhattan plots for the genome-wide association studies for host tolerance to gastrointestinal nematodes evaluated at different measurement events (ME). 331, 385, 443, 498, and 555 are the mean ages (in days) that the animals had at the moment of evaluation. The dotted line (y=5.64) indicates the threshold for statistical significance. The dashed line (y=4.00) indicates the threshold for suggestive evidence of association.

**Supplementary Figure 6** Manhattan plots for the genome-wide association studies for host tolerance to *Eimeria* spp. evaluated at different measurement events (ME). 331, 385, 443, 498, and 555 are the mean ages (in days) that the animals had at the moment of evaluation. The dotted line (y=5.64) indicates the threshold for statistical significance. The dashed line (y=4.00) indicates the threshold for suggestive evidence of association.

# Supplementary Tables

Table S1. Number of repeated measurements per animal

|  |  |
| --- | --- |
| Number of measurements | Number of animals |
| 1 | 30 |
| 2 | 74 |
| 3 | 232 |
| 4 | 447 |
| 5 | 929 |

Table S2. Keywords used to construct the trained list of genes for body weight (BW) and host tolerance ticks (HT.TICK), gastrointestinal nematodes (HT.GIN) and *Eimeria* spp. (HT.EIM)

|  |  |
| --- | --- |
| Trait | GUILDify Keywords |
| BW | Body weight, Growth, Obesity, Protein, Muscle, Fat, Growth factors, Height |
| HT.TICK | Immunity, Immune response, Inflammation, Ectoparasite, Cytokines, Tick, Infection, Tolerance |
| HT.GIN | Immunity, Immune response, Inflammation, Endoparasite, Cytokines, Nematodes, Infection, Tolerance |
| HT.EIM | Immunity, Immune Response, Inflammation, Endoparasite, Cytokines, Eimeria, Infection, Tolerance |

Table S3. Genetic parameters1 for multiple trait linear random regression models coefficients when the body weight was evaluated in function of the direct (intercept) and host tolerance to ticks (TICK); gastrointestinal nematodes (GIN) and *Eimeria* spp. (slope) effects in different measurement event (ME)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ME2 |  |  |  |  |  |
| TICK | | | | | |
| 331 | 127.95  (0.78;223.6) | 2.09  (0.18;3.94) | -7.54  (-15.65;0.86) | -0.29  (-0.67;0.31) | 368.07  (294.1;476.2) |
| 385 | 122.43  (1;235.3) | 1.41  (0.19;2.54) | -7.73  (-16.4;0.56) | -0.45  (-0.77;0.07) | 440.61  (372.6;513.3) |
| 443 | 51.62  (0.65;140.4) | 1.14  (0.24;2.46) | -2.69  (-13.45;2.2) | -0.05  (-0.78;0.32) | 497.38  (443.2;557.4) |
| 498 | 54.90  (0.97;151.2) | 1.36  (0.27;2.68) | -1.53  (-5.99;1.71) | 0.00  (-0.47;0.45) | 569.53  (482.3;643.3) |
| 555 | 62.92  (0.22;223.2) | 1.53  (0.4;2.95) | -3.30  (-14.91;1.7) | -0.05  (-0.69;0.4) | 598.53  (506;680.4) |
| GIN | | | | | |
| 331 | 181.32  (80.2;280.7) | 5.8  (2.2;9.25) | -14.83  (-25.5; -4.47) | -0.45  (-0.58; -0.32) | 315.25  (244.8;393.4) |
| 385 | 124.51  (4.15;296.4) | 5.32  (0.46;13.95) | -12.26  (-38.22;3.58) | -0.18  (-0.66;0.36) | 417.19  (309.8;505.2) |
| 443 | 48.23  (3.48;125.7) | 1.82  (0.47;2.83) | 0.57  (-4.5;4.38) | 0.20  (-0.3;0.52) | 465.89  (407.3;523.8) |
| 498 | 180.72  (10.9;383.4) | 8.18  (1.47;17.2) | -24.12  (-57.81;4.96) | -0.44  (-0.83;0.31) | 530.34  (445.6;604.3) |
| 555 | 36.15  (1.26;83.59) | 3.39  (1.14;5.48) | 2.00  (-1.73;5.23) | 0.19  (-0.36;0.59) | 587.01  (516.7;664.2) |
| EIM | | | | | |
| 331 | 139.88  (63.9;237.9) | 4.24  (1.58;6.83) | -8.46  (-15.62;0.39) | -0.32  (-0.54; -0.11) | 334.38  (260.4;399.9) |
| 385 | 180.26  (29.89;292.5) | 11.07  (1.55;18.15) | -26.69  (-48.93;4.66) | -0.44  (-0.79;0.27) | 385.95  (319.1;453.3) |
| 443 | 70.7  (9.26;141.9) | 4.58  (0.8;11.5) | -2.05  (-14.46;5.78) | 0.05  (-0.46;0.41) | 459.59  (391.3;530.1) |
| 498 | 65.8  (6.92;145.1) | 5.41  (2.11;8.89) | -2.68  (-11.88;6.09) | -0.05  (-0.62;0.46) | 557.82  (483.8;625.3) |
| 555 | 40.3  (1.42;90.39) | 8.99  (4.01;13.63) | 1.07  (-5.27;8.46) | 0.12  (-0.24;0.49) | 592.29  (529.4;660.9) |

1 **=** genetic additive variance for the intercept;  **=** genetic additive variance for the slope; **=** genetic additive covariance between intercept and slope; **=** genetic correlation between intercept and slope;  **=** residual variance. 2 331, 385, 443, 498 and 555 represent the mean ages in days that the animals had in each ME.

Table S4. Summary statistics of genes submits to candidate genes prioritization analysis for host tolerance to ticks

|  |  |  |  |
| --- | --- | --- | --- |
| Gene symbol | Gene name | Average score | Overall *P-value* |
| CDC42 | cell division cycle 42 | 0.831 | 0.000 |
| CFH | complement factor H | 0.818 | 0.000 |
| ECE1 | endothelin converting enzyme 1 | 0.790 | 0.001 |
| WNT4 | Wnt family member 4 | 0.776 | 0.001 |
| HSPG2 | heparan sulfate proteoglycan 2 | 0.786 | 0.001 |
| SERPINF1 | serpin family F member 1 | 0.567 | 0.001 |
| YWHAE | tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein epsilon | 0.672 | 0.001 |
| CRK | CRK proto-oncogene, adaptor protein | 0.653 | 0.003 |
| SERPINF2 | serpin family F member 2 | 0.674 | 0.003 |
| RAP1GAP | RAP1 GTPase activating protein | 0.707 | 0.006 |
| ALPL | alkaline phosphatase, biomineralization associated | 0.650 | 0.008 |
| INPP5K | inositol polyphosphate-5-phosphatase K | 0.648 | 0.008 |
| MYO1C | myosin IC | 0.483 | 0.013 |
| SRR | serine racemase | 0.606 | 0.017 |
| RILP | Rab interacting lysosomal protein | 0.471 | 0.023 |
| KCNT2 | potassium sodium-activated channel subfamily T member 2 | 0.559 | 0.027 |
| MNT | MAX network transcriptional repressor | 0.352 | 0.028 |
| WDR81 | WD repeat domain 81 | 0.542 | 0.028 |
| SCARF1 | scavenger receptor class F member 1 | 0.471 | 0.030 |
| RTN4RL1 | reticulon 4 receptor like 1 | 0.473 | 0.032 |
| RPA1 | replication protein A1 | 0.482 | 0.048 |
| PITPNA | phosphatidylinositol transfer protein alpha | 0.524 | 0.060 |
| RPH3AL | rabphilin 3A like (without C2 domains) | 0.330 | 0.079 |
| HIC1 | HIC ZBTB transcriptional repressor 1 | 0.404 | 0.093 |
| SLC43A2 | solute carrier family 43 member 2 | 0.350 | 0.118 |
| DOC2B | double C2 domain beta | 0.289 | 0.141 |
| PRPF8 | pre-mRNA processing factor 8 | 0.360 | 0.149 |
| MIR22 | microRNA 22 | 0.333 | 0.162 |
| SGSM2 | small G protein signaling modulator 2 | 0.427 | 0.177 |
| CELA3B | chymotrypsin like elastase 3B | 0.209 | 0.221 |
| TLCD2 | TLC domain containing 2 | 0.335 | 0.244 |
| TSR1 | TSR1 ribosome maturation factor | 0.256 | 0.244 |
| OVCA2 | OVCA2 serine hydrolase domain containing | 0.212 | 0.254 |
| DPH1 | diphthamide biosynthesis 1 | 0.167 | 0.274 |
| USP48 | ubiquitin specific peptidase 48 | 0.201 | 0.291 |
| SMG6 | SMG6 nonsense mediated mRNA decay factor | 0.137 | 0.413 |
| METTL16 | methyltransferase like 16 | 0.114 | 0.495 |
| SMYD4 | SET and MYND domain containing 4 | 0.046 | 0.686 |

Table S5. Summary statistics of genes submits to candidate genes prioritization analysis for host tolerance to ticks

|  |  |  |  |
| --- | --- | --- | --- |
| Gene symbol | Gene name | Average score | Overall *P-value* |
| CXCL12 | C-X-C motif chemokine ligand 12 | 1.000 | 0.000 |
| AGT | angiotensinogen | 0.794 | 0.000 |
| PTK2 | protein tyrosine kinase 2 | 0.773 | 0.000 |
| ROS1 | ROS proto-oncogene 1, receptor tyrosine kinase | 0.722 | 0.002 |
| KCNQ3 | potassium voltage-gated channel subfamily Q member 3 | 0.527 | 0.004 |
| AGO2 | argonaute RISC catalytic component 2 | 0.642 | 0.006 |
| KCNK9 | potassium two pore domain channel subfamily K member 9 | 0.476 | 0.015 |
| GOPC | golgi associated PDZ and coiled-coil motif containing | 0.452 | 0.015 |
| KPNA5 | karyopherin subunit alpha 5 | 0.532 | 0.034 |
| LRRC6 | leucine rich repeat containing 6 | 0.517 | 0.039 |
| TFAM | transcription factor A, mitochondrial | 0.441 | 0.058 |
| TRAPPC9 | trafficking protein particle complex 9 | 0.445 | 0.070 |
| NUS1 | NUS1 dehydrodolichyl diphosphate synthase subunit | 0.380 | 0.085 |
| RFX6 | regulatory factor X6 | 0.276 | 0.138 |
| EFR3A | EFR3 homolog A | 0.360 | 0.158 |
| CHRAC1 | chromatin accessibility complex subunit 1 | 0.234 | 0.234 |
| COG2 | component of oligomeric golgi complex 2 | 0.233 | 0.257 |
| ZUP1 | zinc finger containing ubiquitin peptidase 1 | 0.262 | 0.257 |
| RSPH4A | radial spoke head component 4A | 0.200 | 0.316 |
| DCBLD1 | discoidin, CUB and LCCL domain containing 1 | 0.221 | 0.373 |
| TMEM71 | transmembrane protein 71 | 0.167 | 0.383 |
| SULT1C4 | sulfotransferase family 1C member 4 | 0.174 | 0.384 |
| MIR151A | microRNA 151a | 0.184 | 0.431 |
| PHF20L1 | PHD finger protein 20 like 1 | 0.158 | 0.487 |
| ZNF239 | zinc finger protein 239 | 0.156 | 0.496 |
| VGLL2 | vestigial like family member 2 | 0.081 | 0.505 |
| HHLA1 | HERV-H LTR-associating 1 | 0.151 | 0.511 |
| FAM162B | family with sequence similarity 162 member B | 0.062 | 0.606 |
| ZNF32 | zinc finger protein 32 | 0.069 | 0.650 |

Table S6. Summary statistics of genes submits to candidate genes prioritization analysis for host tolerance to ticks

|  |  |  |  |
| --- | --- | --- | --- |
| Gene symbol | Gene name | Average score | Overall *P-value* |
| CXCL9 | C-X-C motif chemokine ligand 9 | 1.000 | 0.000 |
| CXCL10 | C-X-C motif chemokine ligand 10 | 1.000 | 0.000 |
| CXCL11 | C-X-C motif chemokine ligand 11 | 1.000 | 0.000 |
| TNFSF13B | TNF superfamily member 13b [ | 0.771 | 0.000 |
| SCARB2 | scavenger receptor class B member 2 | 0.718 | 0.001 |
| IRS2 | insulin receptor substrate 2 | 0.639 | 0.001 |
| SPINK5 | serine peptidase inhibitor Kazal type 5 | 0.727 | 0.001 |
| SPINK1 | serine peptidase inhibitor Kazal type 1 | 0.710 | 0.002 |
| HTR5A | 5-hydroxytryptamine receptor 5A | 0.607 | 0.005 |
| COL4A1 | collagen type IV alpha 1 chain | 0.547 | 0.005 |
| LIG4 | DNA ligase 4 | 0.598 | 0.006 |
| DPP6 | dipeptidyl peptidase like 6 | 0.618 | 0.007 |
| PTPRT | protein tyrosine phosphatase receptor type T | 0.563 | 0.015 |
| DPYSL3 | dihydropyrimidinase like 3 | 0.504 | 0.020 |
| NAAA | N-acylethanolamine acid amidase | 0.544 | 0.034 |
| STBD1 | starch binding domain 1 | 0.360 | 0.042 |
| NUP54 | nucleoporin 54 | 0.391 | 0.064 |
| SPINK6 | serine peptidase inhibitor Kazal type 6 | 0.389 | 0.084 |
| PAXIP1 | PAX interacting protein 1 | 0.378 | 0.099 |
| ART3 | ADP-ribosyltransferase 3 | 0.277 | 0.144 |
| PPEF2 | protein phosphatase with EF-hand domain 2 | 0.383 | 0.145 |
| SCGB3A2 | secretoglobin family 3A member 2 | 0.348 | 0.191 |
| MYO16 | myosin XVI | 0.274 | 0.222 |
| STK32A | serine/threonine kinase 32A | 0.270 | 0.224 |
| SHROOM3 | shroom family member 3 | 0.285 | 0.313 |
| JAKMIP2 | janus kinase and microtubule interacting protein 2 | 0.217 | 0.324 |
| SDAD1 | SDA1 domain containing 1 | 0.238 | 0.364 |
| SOWAHB | sosondowah ankyrin repeat domain family member B | 0.171 | 0.412 |
| ABHD13 | abhydrolase domain containing 13 | 0.110 | 0.500 |
| CCDC158 | coiled-coil domain containing 158 | 0.000 | 0.763 |