Journal Title Here, YYYY, pp. 1–4

doi: DOI HERE
Advance Access Publication Date: Day Month Year
Paper

PAPER

Tolerance of hybrid hosts against infections, TAC Meeting December 2022

Fay Webster,[1,2,*] Lubomír Bednář[1,2,†] and Emanuel Heitlinger[1,2,‡]

[1]Department of Molecular Parasitology, Institute for Biology, Humboldt University Berlin (HU), Philippstraße 13, Berlin, 10115, Germany and [2]Research Group Ecology and Evolution of Molecular Parasite-Host Interactions, Leibniz-Institut for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, Berlin, 10315, Germany
[*]Corresponding author. fay.webster@hu-berlin.de[†]Current email address: fay.webster@hu-berlin.de))[‡]
FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

Parasites in hybrid zones can give insight into species barriers, as they are modulating the fitness of hybrid hosts. Recent findings have demonstrated lower infection intensities with parasites in hybrids in the European House Mouse Hybrid zone (HMHZ), indicating higher disease resistance. However, tolerance has not yet been addressed in depth, as it is impractical to measure in wild populations. In an attempt to predict and evaluate the health impact of parasite infections and extrapolate tolerance in the HMHZ, we use a machine learning method. A random forest model was trained on immune parameters measured in experimental lab infections with Eimeria and then applied to data obtained from field sampling. Our predictions revealed that these infections are more detrimental to hybrid male mice. This approach represents an initial step in assessing tolerance in field studies.

Key words: hybrids; parasites

1. Introduction

1.1. Introduce the topic

2. Methods

2.1. Laboratory infection experiments

2.1.1. Mouse strains (Luke)

In order to gain a better understanding of tolerance in hybrid mice we established a laboratory model of experimental lab infections with the Eimeria spp. Our experimental setup is a variation of the framework in Balard et al., 2020. The mice used are four wild-derived inbred mouse strains and their generated F1 hybrids. The mouse strains are fully inbred, as they have pased through at least 20 generations of sibling pairing. From the fours strains, two were used as a representation of the M. m. domesticus: SCHUNT (Locality: Schweben, Hessen, Germany [N: 5°0 26', E: 9°36'] Martincová et al. (2019)) and STRA (Locality: Straas, Bavaria, Germany [N: 50°11', E: 11°46'] (Piálek et al. (2008)). The two following strains were in turn derived from M. m. musculus: BUSNA (Locality: Buškovice, Bohemia, Czech Republic [N: 5°0 14', E: 1°3 22'] (pialek2008development)) and PWD (Locality: Kunratice, Bohemia, Czech Republic [N: 5°0 01', E: 14 2°9'] (Gregorova et al. (2000)). In our setup there are two two intersubspecific hybrids (STRAxBUSNA and SCHUNTxPWD) and two intrasubspecific hybrids (SCHUNTxSTRA and PWDxBUSNA). The mice were between 5.6 and 21.4 weeks. The mice were acquired from the Institute of Vertebrate Biology of the Czech Academy of Sciences in Studenec (license number 61974/2017-MZE-17214; for further details on strains see https://housemice.cz/en).

Infections with the parasite Eimeria induce a protective immune reaction in the host against reinfection (Rose et al., 1992a; Smith & Hayday, 2000). The feces of the naive mice were tested to ensure that the mice were Eimeria spp., prior to infection, following the methods of Balard Balard et al. (2020b).

2.1.2. Infections with Eimeria spp. (Luke)

The procedure used is as described in in Balard et al., 2020. During the infections mice were housed solo in cages. We infected the mice orally with 150 sporulated oocysts of one Eimeria isolate suspended in 100 µl phosphate-buffered saline (PBS). The mice had access to food and water ad libitum SNIFF, Rat/Mouse maintenance feed 10 mm and were observed

daily for 8 days until their sacrifice by cervical dislocation. In the case that individual mice showed severe adverse health effects or extreme weigh loss of more than 18% relative to their weight at the start of experiment, were then sacrificed earlier at defined humane end points (experiment license Reg. 0431/17). Daily measurements of weight were recorded and fecal matter was gathered. Collected feces were supspended in 2% potassium dichromate and paaracite oocysts were retrieved by NaCl flotation.

To enable a consistent distribution between experimental groups, mice were allocated at random, while ensuring a similar distribution of age and sex between groups.

2.1.3. Gene expression high-thoughput qPCR (Luke)

Homogenized caecum tissue was processed for RNA using the filter based innuPREP RNA Mini Kit 2.0 (Jena analytik, Germany) according to manufacturer instructions. Extracted RNA was quantified using NanoDrop 2000c (Thermo Scientific, Waltham, USA) and transcribed into cDNA with iScript (Bio-rad Laboratories, Hercules, California, United States), following manufacturer protocol. Gene expression was ascertained via hight-throughput qPCR of cDNA from the caecum extracted material, plated onto Fluidigm IFC (integrated fluidic circuit), initialized by the Juno controller and read in the Fluidigm BioMark HD (PN 100-7222 C1, Fluidigm, South San Francisco, California, United States). After cDNA conversion mentioned above, the samples went through the recommended preparation steps, carried out in sterile extractor hood, with reagents being kept at 4 °C when in use, and at -20 °C when stored overnight. All pipetting was done using sterile filter tips. Table of wet-lab tested primers can be found in the appendix file 20614FDGP21T1_DESIGN

2.1.4. Specific target amplification (STA) using Fluidigm PreAmp master mix (Luke)

ll primers at 100µM were pooled together in a micro centrifuge tube and resulting solution was diluted to a final concentration of 500 nM, with a 10 mM Tris-HCl (pH 8,0), 0.1 mM EDTA TE Buffer (12090-015, Invitrogen, Waltham, Massachusetts, United States). Fluidigm Preamp Master Mix (Fluidigm PN 100-5580) was added to the solution according to the manufacturer instructions, to create delta gene assay Preamp Master Mix. Sample cDNA (including a non-template control (NTC)) was plated onto 96-well plates and the master mix added to each well. Sample plate was then gently vortexed for 5 seconds and spun down at 1,000 x g for 1 minute. The amplification reactions were carried out in the Biometra TOne 96 (846-2-070-301, Analytic Jena, Jena, Germany) at the following cycling conditions: Hold at 95 ºC for 2 minutes, 95 ºC for 15 seconds, then 60 ºC for 4 minutes (15x), Hold at 4 ºC for infinity.

2.1.5. Primers and sample preparation: (Luke)

1.5 µL of primer (wet lab tested) at 100 µM ,13.5 µL of TE Buffer and 15 µL of 2X Assay Loading Reagent (PN 100-7611, Fluidigm), were combined to create 10X Assay solutions. Sample reaction mixes were then created by combining 1.8 µL of preamplified and Exonuclease I treated cDNA, 2 µL of 2X SsoFast EvaGreen Supermix with low ROX™ (Bio-Rad, PN 172-5211) and 192.24 Delta Gene Sample Reagent (PN 100-6653, Fluidigm), to create 4 µL stock per 1 sample reaction. This was scaled to accommodate sample repeat runs. Prepared samples and primers were stored at -20ºC until the IFC chips were primed.

2.1.6. IFC qPCR runs (Luke)

The 192.24 IFC was plated and treated as per manufacturer instructions (PN 100-7222 C1), consisting of control line fluid injection into accumulator 2 slot, removal of protective film, pipetting of samples and 10X assay mixes in 3 µL volumes, adding 150 µL of Actuation Fluid (PN 100-6250) into the P1 port, 150 µL of Pressure Fluid (PN 100-6249) into the P2 and P3 ports, 20 µL of Pressure Fluid into the P4 and P5 ports and finally initializing the IFC in the Juno controller, using the Load Mix 192.24 GE script. After the initialization completed, the IFC was transferred into the BioMark HD and the assay properties were set as follows: Application type: Gene Expression, Passive reference: ROX, Assay: Single probe, Probe type: EvaGreen, using the GE 192x24 PCR+Melt v2.pcl protocol, on Auto-exposure.

2.2. Wild mice

During the years 2016 to 2019, we sampled 1889 mice in the House Mouse Hybrid zone. We used live traps to capture mice in farms and houses, during September every year. To catch mice of *M. musculus musculus*, *M. musculus domesticus* and hybrid origin, we selected a large geographic area in Brandenburg and Neuruppin. For each mouse, we have collected information during dissections, on their body length, their weight and their pathophysiology. Tissue samples were used to genotype the hosts. Further information on the set-up of the experiment can be found in previous publications of the group Balard et al. (2020a) From 1889 mice, 336 were selected with a varying genotype for Fluidigm BioMark assay. Another 95 were selected for fluorescence-activated cell sorting.

2.3. Statistical Analysis

2.3.1. Imputation of missing data

To make the most of our data collection, we aimed to resolve missingness. Missing data were imputed using multiple imputations by chained equations. We used the package MICE in R Van Buuren and Groothuis-Oudshoorn (2011), with five imputed data sets and five iterations. Data generated by FACS or the Gene Expression / Biomarker assay were regarded as missing if each mouse had measurements for some variables. For each continuous variable, we specified a predictive mean matching model. All the remaining variables were used as predictors in the imputation. To control the quality of our imputations, we evaluated the distribution plot of the existing data and the imputed data for all measurements 1 2. Further, we tested for convergence. We assume data is "missing completely at random" or "missing at random". For both types of missingness, multiple imputation is a suggested method to impute missing variables Van Buuren (2018).

Questions / To-dos:

1. Should I log-transform the data prior to imputation?
2. Increasing produced data sets / iterations
3. sensitivity analyses using complete cases only

2.3.2. Random forest

In an attempt to predict the health impact of infections in mice, we used a random forest model Breiman (2001). We chose the maximum weight loss during the experimental

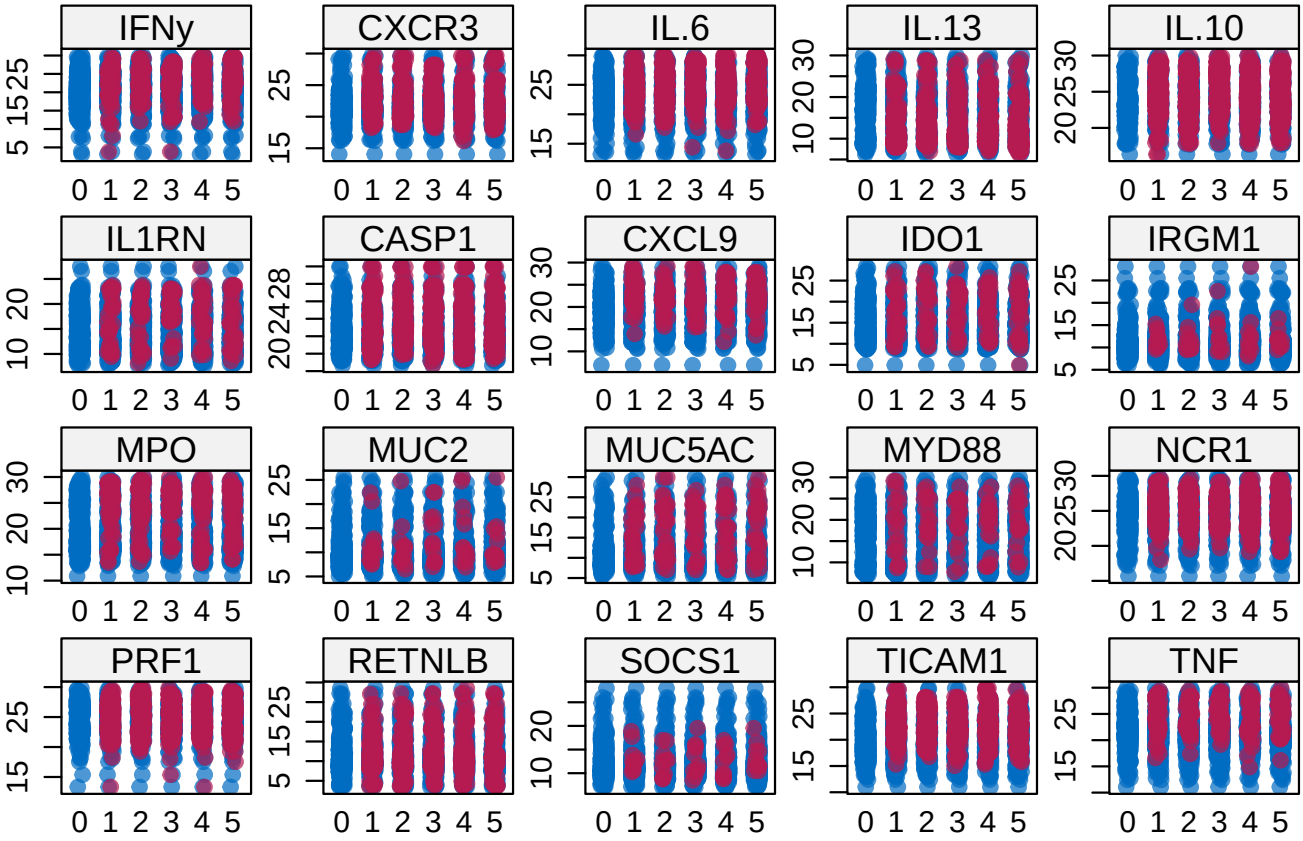


Fig. 1. Stripplot of observed and imputed data

laboratory infections of mice with the parasite *Eimeria spp.* as a response variable. Maximum weight loss is here used as a proxy describing the health impact caused by infections. The random forest was constructed utilizing the expression data of 20 genes, obtained by the biomarker assay (utilizing the R package "randomForest," ntree = 1000). The data set was split into a training data set of 70 % and to a testing data set of 30 % to avoid over-fitting and to assess the performance of the model on "unseen" data.

As a quality assessment, we used k- fold cross validation, where the data set is divided into k subsets. Each time, the model is assessed by using one of the k subsets as the test data set and the other as the training data set. The model was then fitted on each k-subset and afterwards evaluated on the test set. Last the evaluation score was noted and the model then discarded. Further, a permutation test cross-validation was implemented (using the function "rf.crossValidation" (using the R package rfUtilities, version 2.1-5). The percent variance explained from the specified fit model was 27.7%. Moreover, the mean squared error from each bootstrapped model was 44.14. Next, the variable importance was calculated according to the total decrease in node impurities from splitting on each variable, averaged over all trees. In this case of regression, the node impurity is measured by residual sum of squares. Variables with comparetively higher importance have a greater impact on the predictions of the model.

```
## running: regression cross-validation with 99 iterations
```

```
## [1] 27.7
```

```
## [1] 44.13772
```

```
## [1] 78
```

```
## [1] 6.543936
```

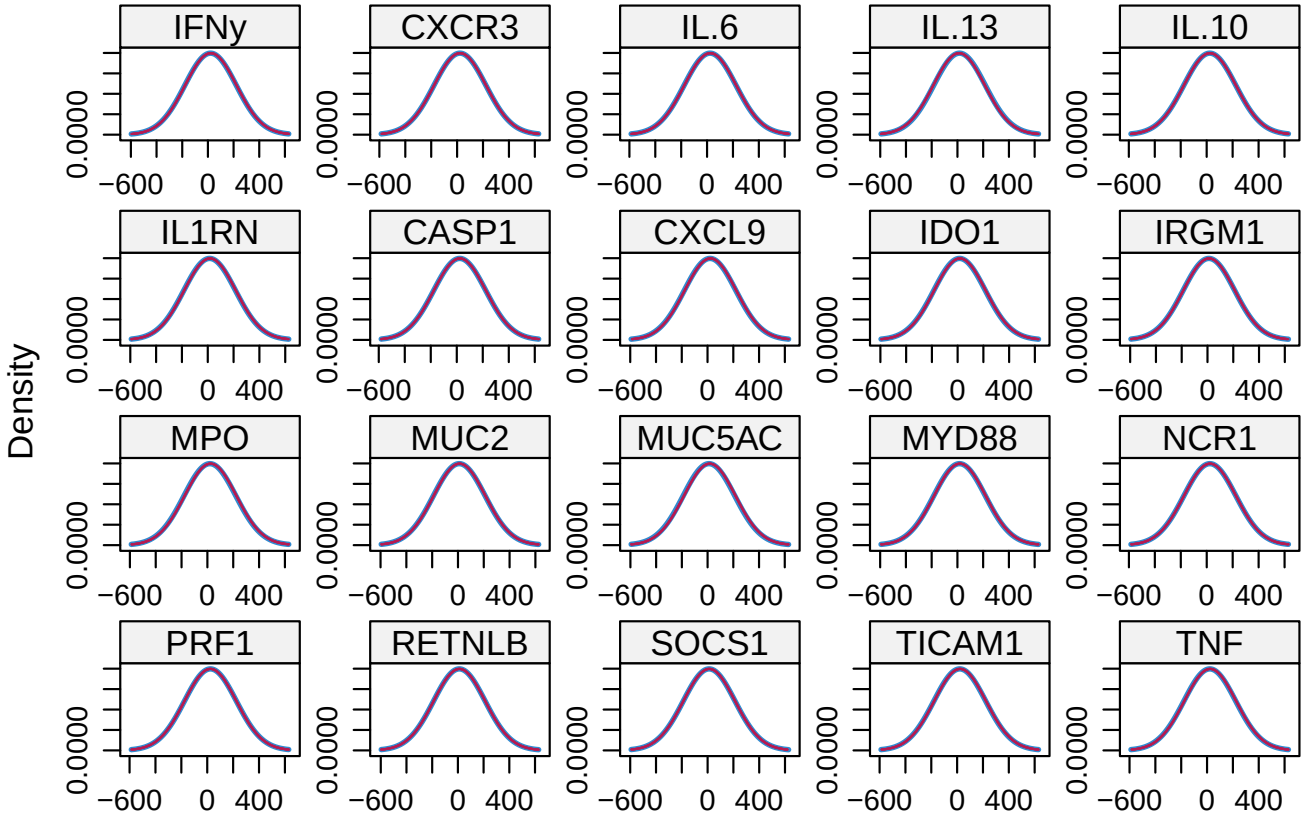


Fig. 2. Density plot of observed and imputed data

```
##         IncNodePurity Var.Names
## SOCS1        582.5419     SOCS1
## TICAM1       549.4166    TICAM1
## CXCL9        538.1206     CXCL9
## IL.6         459.0768      IL.6
## IFNy         395.9684      IFNy
## IL1RN        327.9611     IL1RN
## RETNLB       289.5686    RETNLB
## NCR1         239.8289      NCR1
## TNF          226.3735       TNF
## MPO          214.1609       MPO
## IL.13        201.9333     IL.13
## MUC2         200.2868      MUC2
## MYD88        196.5789     MYD88
## IRGM1        178.2120     IRGM1
## CASP1        170.0408     CASP1
## CXCR3        169.9418     CXCR3
## IDO1         162.3620      IDO1
## PRF1         161.2846      PRF1
## IL.10        156.3871     IL.10
## MUC5AC       149.1062    MUC5AC
```

3. Results

4. Discussion

5. Conclusion

6. Literature citations

7. Competing interests

There are no competing interest.

8. Author contributions statement

To be worked on

References

A. Balard, V. H. Jarquín-Díaz, J. Jost, I. Martincová, L. Ďureje, J. Piálek, M. Macholán, J. Goüy de Bellocq, S. J. Baird, and E. Heitlinger. Intensity of infection with intracellular eimeria spp. and pinworms is reduced in hybrid mice compared to parental subspecies. *Journal of Evolutionary Biology*, 33(4):435–448, 2020a.

A. Balard, V. H. Jarquín-Díaz, J. Jost, V. Mittné, F. Böhning, L. Ďureje, J. Piálek, and E. Heitlinger. Coupling between tolerance and resistance for two related eimeria parasite species. *Ecology and evolution*, 10(24):13938–13948, 2020b.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

S. Gregorova, J. Forejt, et al. Pwd/ph and pwk/ph inbred mouse strains of mus m. musculus subspecies–a

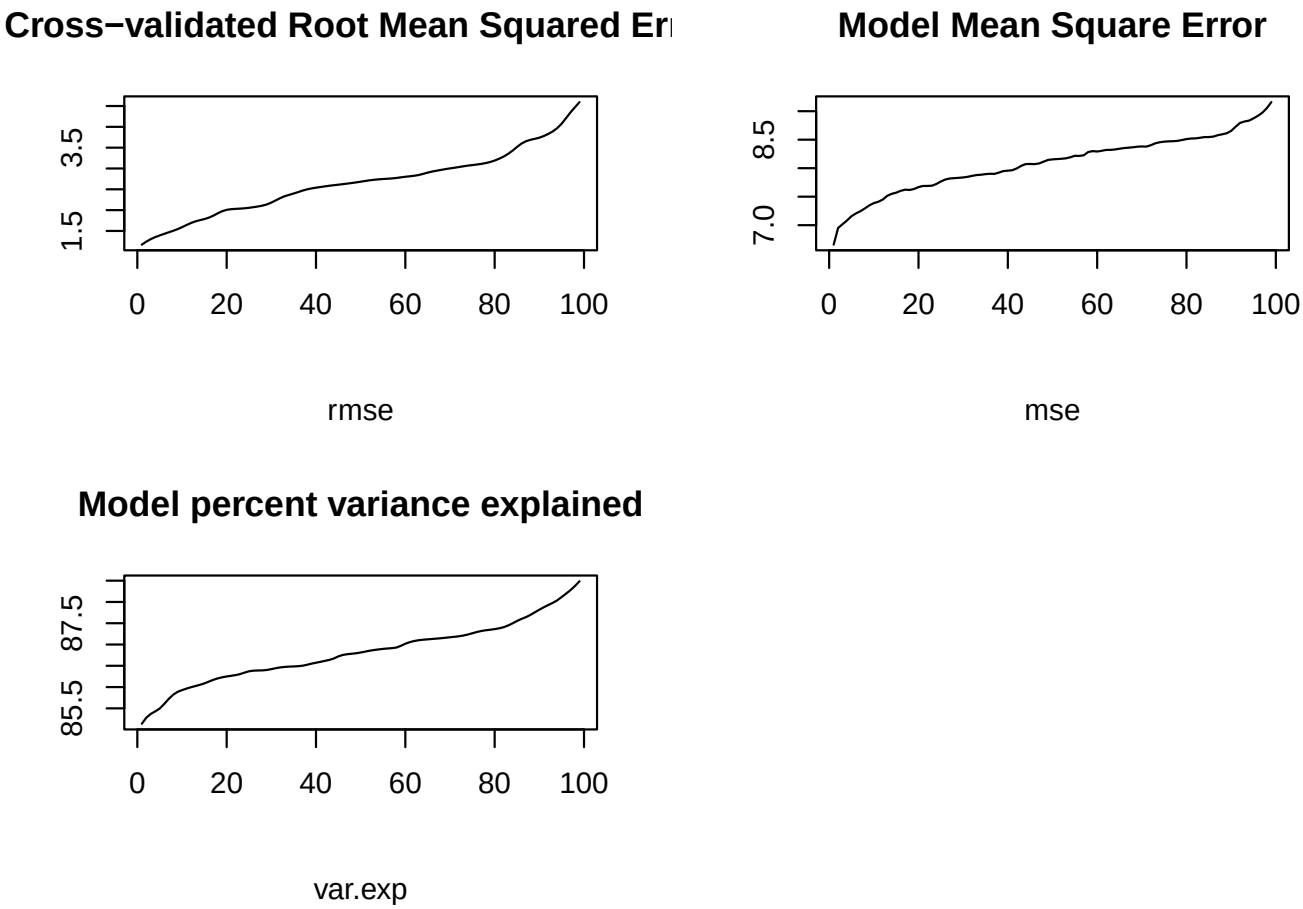


Fig. 3. Variance explained and Root Mean Squared Error

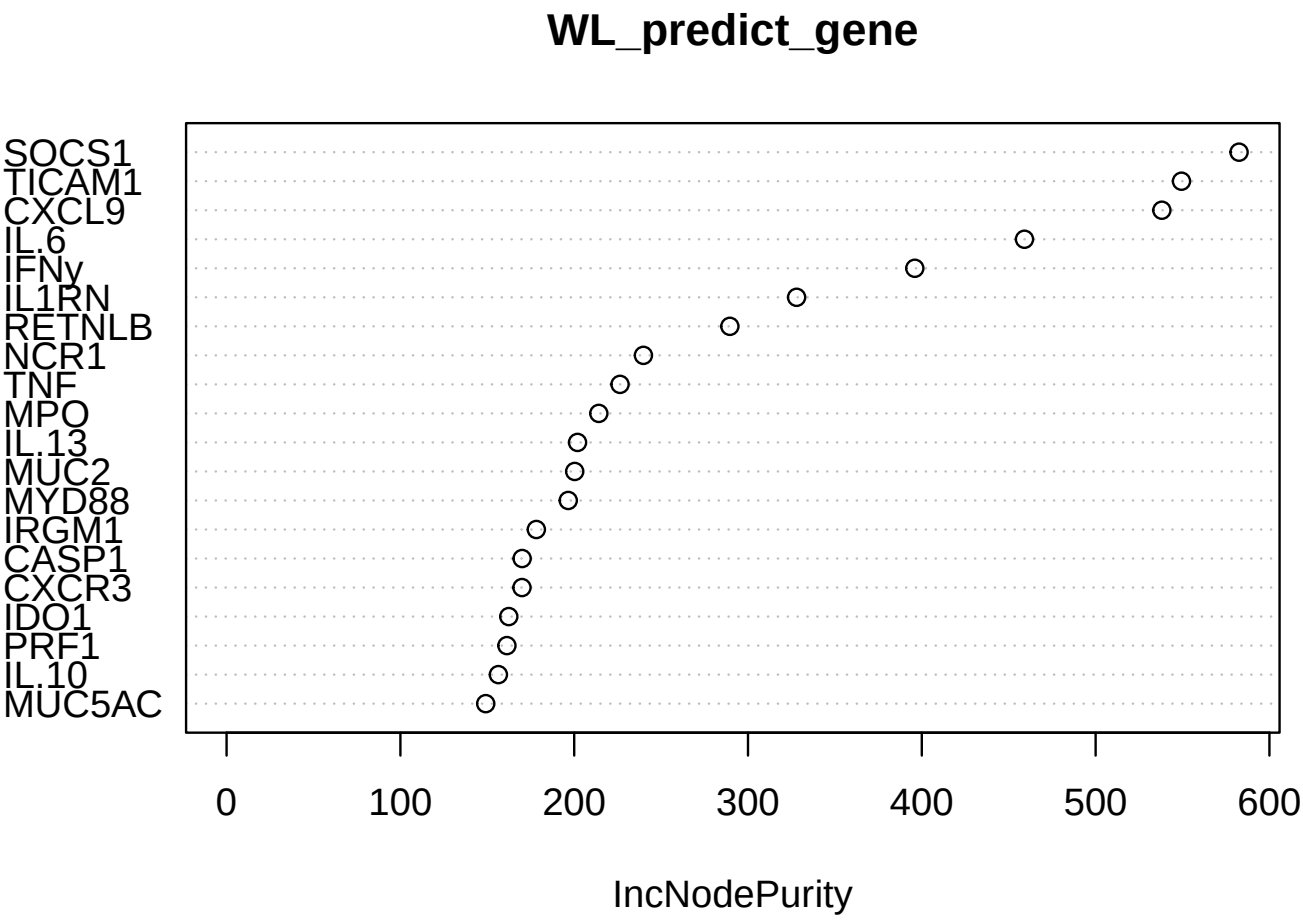
valuable resource of phenotypic variations and genomic polymorphisms. *Folia Biol (Praha)*, 46(1):31–41, 2000.

I. Martincová, L. Ďureje, J. Kreisinger, M. Macholán, and J. Piálek. Phenotypic effects of the y chromosome are variable and structured in hybrids among house mouse recombinant lines. *Ecology and evolution*, 9(10):6124–6137, 2019.

J. Piálek, M. Vyskočilová, B. Bímová, D. Havelková, J. Piálková, P. Dufková, V. Bencová, L. Ďureje, T. Albrecht, H. C. Hauffe, et al. Development of unique house mouse resources suitable for evolutionary studies of speciation. *Journal of Heredity*, 99(1):34–44, 2008.

S. Van Buuren. *Flexible imputation of missing data*. CRC press, 2018.

S. Van Buuren and K. Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 45:1–67, 2011.


WL_predict_gene
SOCS1
TICAM1
CXCL9
IL.6
IFNy
IL1RN
RETNLB
NCR1
TNF
MPO
IL.13
MUC2
MYD88
IRGM1
CASP1
CXCR3
IDO1
PRF1
IL.10
MUC5AC
0
100
200
300
400
500
600
IncNodePurity