A candidate tolerance gene identified in a natural population of field voles (Microtus agrestis)

Wanelik, KM, Begon, M, Birtles, RJ, Bradley, JE, Friberg, IM, Jackson, JA, Taylor, CH, Thomason, AG, Turner, AK and Paterson, S

http://dx.doi.org/10.1111/mec.14476

A candidate tolerance gene identified in a natural population of field voles (*Microtus agrestis*)

Klara M. Wanelik[1*], Mike Begon[1], Richard J. Birtles[2], Janette E. Bradley[3], Ida M. Friberg[2], Joseph A. Jackson[2], Christopher H. Taylor[3], Anna G. Thomason[2], Andrew K. Turner[1], Steve Paterson[1]

[1] Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom

[2] School of Environment and Life Sciences, University of Salford, Salford, United Kingdom

[3] School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

[*] Corresponding author

Email: kwanelik@liverpool.ac.uk

Keywords

Disease ecology; eco-immunology; immune strategy; Gata3; Fcer1a

Running title

A candidate tolerance gene in a wild rodent

Abstract

The animal immune response has hitherto been viewed primarily in the context of resistance only. However, individuals, can also employ a tolerance strategy to maintain good health in the face of on-going infection. To shed light on the genetic and physiological basis of tolerance, we use a natural population of field voles, *Microtus agrestis*, to search for an association between the expression of the transcription factor Gata3, previously identified as a marker of tolerance in this system, and polymorphism in 84 immune and non-immune genes. Our results show clear evidence for an association between Gata3 expression and polymorphism in the Fcer1a gene, with the explanatory power of this polymorphism being comparable to that of other non-genetic variables previously identified as important predictors of Gata3 expression. We also uncover the possible mechanism behind this association using an existing protein-protein interaction network for the mouse model rodent, *Mus musculus*, which we validate using our own expression network for *M. agrestis*. Our results suggest that the polymorphism in question may be working at the transcriptional level, leading to changes in the expression of the Th2-related genes, Tyrosine-protein kinase BTK and Tyrosine-protein kinase TXK, and hence potentially altering the strength of the Th2 response, of which Gata3 is a mediator. We believe our work has implications for both treatment and control of infectious disease.

Introduction

Tolerance, like resistance, is an active response to infection involving the activation of molecular and physiological mechanisms. Unlike resistance though, rather than preventing or clearing an infection, a tolerance response minimises the disease pathology caused by infection (Caldwell, Schafer, Compton, & Patterson, 1958; Schafer, 1971). This strategy may be favoured where infection is a daily occurrence, or infection is persistent (Restif & Koella, 2004). In these cases, the costs of constantly mounting an immune response in terms of damage to host tissue (immunopathology) may be worse than those of infection itself (Medzhitov, Schneider, & Soares, 2012). A resistant strategy, on the other hand, might be associated with acute exposure (Restif & Koella, 2004), where the costs of infection outweigh those of mounting an immune response. Tolerance of infection is now attracting considerable interest in the immunological and ecological literature (Medzhitov et al., 2012; Råberg, Graham, & Read, 2009) and provides a new perspective to help understand how the immune response in animals functions following infection, which has hitherto been viewed primarily in the context of resistance only.

Individuals in apparently similar circumstances differ in their responses to infection, and some are worse than others at either resisting or tolerating infection (Arriero et al., 2017; Buehler, Piersma, Matson, & Tieleman, 2008; Kluen, Siitari, & Brommer, 2013). Beyond recognising that such variation exists in natural populations, though, we understand little of the genetic and physiological basis of this variation but this is a key step towards predicting which individuals are most vulnerable to infectious disease (Råberg, 2014). Genetic variation for tolerance has been previously demonstrated in inbred strains of lab mice (Raberg, Sim, &

Read, 2007) and, to a more limited extent, in natural systems (Regoes et al., 2014). However, knowledge of specific genes controlling tolerance, and hence potentially driving this heritable variation in strategy in the wild, is lacking. Candidate genes include those involved in limiting immunopathology and/or regulation of the immune response (Medzhitov et al., 2012; Råberg et al., 2009). In the laboratory, a genetic locus on mouse chromosome 11 (Ctrq3) has been shown to influence tolerance to *Chlamydia psittaci* infection, with circumstantial evidence for candidate genes belonging to the family of immunity-related GTPases (Miyairi et al., 2012). Another study has also identified a signalling protease required for melanisation in *Drosophila melanogaster* (CG3066) as being of importance (Ayres & Schneider, 2008). Finally, in humans, an association between HLA-B genotype and degree of tolerance to HIV has been shown (Regoes et al., 2014).

Our own work has previously identified the expression of a particular master transcription factor, Gata3, as a marker of tolerance in mature male field voles, *Microtus agrestis.* This work showed that macroparasite infection in these mature voles gave rise to elevated levels of Gata3 expression, which in turn gave rise to improved body condition and enhanced survival (Jackson et al., 2014). This fits with the known role of Gata3 as a mediator of the Th2 response, and the role of the Th2 immune system in tissue repair (Allen & Wynn, 2011). Furthermore, we have shown consistent differences between individuals in their typical level of Gata3 expression, after other measured sources of variation have been taken into account (Arriero et al., 2017). Together, our results imply consistent difference between individuals in the strength of their tolerance response.

Here, we address the contribution of genotype to consistent individual differences in the expression of Gata3, a marker of tolerance. We use a natural population of wild *M. agrestis* to search for an association between Gata3 expression and polymorphism in 84 immune and non-immune genes. We find Gata3 expression associated with polymorphism at the Fcer1a gene (which encodes the alpha chain of the high affinity receptor for immunoglobulin epsilon, IgE), and show that the proportion of variation in Gata3 expression explained by this polymorphism is comparable to that explained by other environmental and physiological variables. We also shed light on the possible mechanism behind this association by constructing a protein-protein interaction network for the mouse model rodent, *Mus musculus*, which we validate using our own expression network for *M. agrestis*.

Materials & Methods

Field design and animals

We studied *M. agrestis* in Kielder Forest, Northumberland (55º13' N, 2º3' W) using live-trapping to access individual animals from natural populations. Our studies were designed to permit the analysis of individual variation in host condition and survival, infection status, and the expression of immune genes (for full details of all methods below see Jackson et al. 2011, 2014). The studies were divided into longitudinal and cross-sectional components.

Initial survey

We repeated our field design at two spatially separate sites (BLB and SQC) in 2008–2009, and a further two (SCP and KTH) in 2009–2010. Each site contained a central trapping grid

(~0.375 ha) of 150 (10 x 15) regularly spaced traps (3–5 m intervals) which was used in a capture-recapture study (reported elsewhere). The cross-sectional component reported here utilised curvilinear transects of 100 live traps arranged at 5–10 m intervals which were placed around the margins of each habitat.

Ten voles per month were destructively sampled from the transects between February and November (2008–2009) or April and November (2009–2010). In November (2008 and 2009) and March (2009 and 2010), larger numbers of animals were sampled both from the transects and from the central grid habitats. These samples are used here to carry out a haplotype association analysis.

On capture, each animal was examined for ectoparasites (see below). Only results for male *M. agrestis* are reported here given the focus of previous work (Jackson et al., 2014). Males were classified as either immature (non-mating with undeveloped testes) or mature (mating with large testes and expanded seminal vesicles). Some biometric data were also collected, including body weight (g) and snout-vent length (mm). All animal procedures carried out as part of this initial survey were performed with approval from the University of Liverpool Animal Welfare Committee and under a UK Home Office license (PPL 40/3235 to MB).

Parasite assays

On capture, ectoparasite infections were recorded, as direct counts of ticks (*Ixodes* spp.) and small flea species (*Ctenophthalmus nobilis*, *Peromyscopsylla spectabilis*, *Megabothris walkeri*, *Malaraeus penicilliger*, *Rhadinopsylla pentacantha*). Captured animals were then returned to the laboratory where they were killed by an overdose of chloroform followed by

exsanguination, and dissection in order to take a more comprehensive set of infection measurements. This included a direct count of adult cestodes found in the gut (*Anoplocephaloides dentata* aff., *Paranoplocephala* sp., *Rodentolepis asymmetrica*, *Arostrilepis horrida*). We collected infection metrics for these macroparasites because they are the most common species that would be expected to be in strong contact with the host immune system (Jackson et al., 2014).

Follow-up survey

In 2015, we collected samples at four sites (GRD, CHE, SCP, COL). Similarly to the initial survey, each site contained a trapping grid of 150-197 regularly spaces traps (at approx. 5 m intervals) but this was used both for cross-sectional and longitudinal components (not reported here). Sixty-four voles were also destructively sampled from the grids between July and October 2015 to assay expression by RNA-seq. Both females and males were included in order to maximise sample size. In this study, voles were killed by a rising concentration of CO_2, followed by exsanguination. These samples were shown to be comparable in terms of weight, age and sex to the population sampled in the initial survey (Table S1) and are used here to construct an expression network for *M. agrestis*. All animal procedures carried out as part of this survey were performed with approval from the University of Liverpool Animal Welfare Committee and under a UK Home Office license (PPL ??? to SP).

Haplotype association study

Immunological assays

We used two-step reverse transcription quantitative real-time PCR (Q-PCR) to measure messenger RNA (mRNA) accumulation of Gata binding protein 3 (Gata3; a transcription factor associated with the Th2 response) from splenocyte cultures stimulated with mitogen phytohaemagglutinin (PHA). Gata3 has previously been identified as a marker of tolerance in mature male voles (Jackson et al., 2014). PHA preferentially activates and stimulates proliferation of CD4+ helper T-cells in vitro (O'Donovan, Johns, & Wilcox, 1995). Here, we use that observed expression profile as a measure of the potential responsiveness of the immune system in vivo.

SNP identification and genotyping

We identified 288 single nucleotide polymorphisms (SNPs) in 85 immune-related genes and 25 non-immune genes. Immune genes included cytokine genes and other genes known to be involved in pathogen resistance. The Immunome database version 1.1. (http://structure.bmc.lu.se/idbase/Immunome/index.php), a manually curated database containing information on 893 genes considered essential to the human immune system, was used a starting point for identifying a list of candidate immune genes (Ortutay & Vihinen, 2006). First, we excluded all those genes in this database with no known orthologue in house mice. We then applied a heuristic approach to ensure that those genes which were most likely to be of interest given our previous work (e.g. Jackson *et al.* 2014) were represented in our list, and excluded those genes with no known polymorphisms in *M. agrestis*. We also chose a set of non-immune genes to act as a control for spurious associations, caused, for example, by demographic effects. This set was composed solely of metabolic genes, as these are far less likely to be involved in host-pathogen interactions (see Table S2 for full list of immune and non-immune genes identified).

DNA was extracted from the livers of voles that had been destructively sampled as part of the cross-sectional study and for which Gata3 expression levels were available (n = 221) using DNeasy Blood and Tissue Kit (Qiagen). Genotyping was then performed by KBiosciences (Hoddesdon, UK; http://www.kbioscience.co.uk) using the KASPar SNP genotyping system. This included negative controls (water) and duplicate samples to validate reproducibility.

Statistical analyses

All analyses were carried out in R statistical software version 3.4.0 (R Core Team, 2016). The SNP genotyping data were checked in a number of ways. We used the SNPassoc package (González, Armengol, Guinó, Solé, & Moreno, 2014) to test for deviations from Hardy-Weinberg equilibrium using exact tests. Because of the large number of exact tests performed, the Benjamini and Hochberg method of correction was applied to the resulting *p*-values (Benjamini & Hochberg, 1995). The degree of linkage disequilibrium (LD) between SNPs was analysed using the genetics package (Warnes, Gorjanc, Leisch, & Man, 2013). Pairwise scaled LD estimates (D´) were computed for each pair of SNPs to test (a) whether SNPs within the same gene demonstrate high LD and therefore are more appropriately used to identify phenotypic associations in combination, and (b) whether SNPs within different genes demonstrate low LD, indicative of the independence of genetic loci.

The SNP genotyping data were (a) converted into haplotype data for each gene and (b) tested for associations with mitogen-stimulated Gata3 expression while controlling for other known covariates, using the hapassoc package (Burkett, Graham, & McNeney, 2006; Burkett, McNeney, & Graham, 2004). This software allows likelihood inference of trait associations

with SNP haplotypes and other attributes, adopts a generalized linear model framework and estimates parameters using an expectation-maximization algorithm. If the haplotype combination of an individual cannot be inferred from its genotyping data (a) because it is heterozygous at two or more markers or (b) because it has missing data for a single marker, the approach implemented in hapassoc is to consider all possible haplotype combinations for that individual. Standard errors accounting for this added uncertainty are calculated using the Louis' method (Louis, 1982).

We assumed an additive genetic model, where Gata3 expression is linearly related to the number of copies of a haplotype present and we pooled together all those haplotypes with frequencies below 5%. Gata3 expression values were Box-Cox transformed to achieve approximately normal residuals. Other non-genetic covariates included in this model were site (BLB, SQC, SCP & KTH), maturity (either immature or mature male), residual weight (adjusted for body size) and the first principal component from a PCA summarising the macroparasites measured. This component explained 47% of the variation in macroparasite burden and showed high positive loadings for all three macroparasite groups (ticks: 0.56, fleas: 0.57 and adult cestodes: 0.60). Grouping of ectoparasites and endoparasites in this way is in line with previous work that shows that both ectoparasites (V. D. Boppana, Thangamani, Alarcon-Chaidez, Adler, & Wikel, 2009; V. Boppana, Thangamani, AJ, & Wikel, 2009) and endoparasites (Anthony et al., 2007; Harris & Gause, 2011) stimulate the Th2 response, which has been suggested to act "as an adaptive tissue repair mechanism that quickly heals the wounds they inflict" (Allen & Wynn, 2011). These variables have previously been identified as important predictors of Gata3 expression (Jackson et al., 2011, 2014). All non-

genetic covariates were tested for independence (Pearson correlation coefficients = -0.2 - 0.35).

As required by the hapassoc package, we excluded all genes with a single SNP and all monomorphic SNPs (see Table S2 for these), resulting in a total of 238 SNPs in 62 immune-related genes and 22 non-immune genes being included in the analysis (see Table S3 for final list of immune and non-immune genes). We also excluded those subjects for which more than one single-locus genotype had missing data. Because of the large number of association tests performed, the Benjamini and Hochberg method of correction was applied to all *p*-values, with the false discovery rate set to 0.1 (Benjamini & Hochberg, 1995). Resulting *q*-values (FDR-corrected *p*-values) were checked for a uniform distribution.

We were unable to include any random variables or interaction terms in the initial trait association analysis, as the hapassoc package does not have his capability. So, following this analysis, a linear mixed effects model (LMM) for Gata3 expression was constructed to confirm these results. This included season [four levels, designated as: spring (March to May), summer (June to August), autumn (September to November) and winter (December to February)], assay plate number and site as random effects. It also included previously identified interactions between maturity and macroparasitic load, as well as maturity and residual weight (Jackson *et al.* 2014). Three haplotypes were identified at this locus: GCC, ACC and ACT. GCC was found to be the haplotype of interest in relation to Gata3 expression in the initial trait association analysis (see Results). Genotype was therefore coded as a the number of GCC copies. This was treated as a continuous variable because only five individuals were found to have two copies of the GCC haplotype, making it difficult to make

reliable comparisons between factor levels. Treatment of genotype as a continuous variable also reduced the number of degrees of freedom by one. Only those individuals whose combination of haplotypes or 'haplotype phase' could be determined with certainty were included in this analysis, but this was the majority of individuals (n = 191; 86%). The contribution of genotype relative to other predictors in explaining variance in Gata3 expression was assessed by calculating the marginal R^2 using the MuMIn package (Barton, 2016) for (a) the full LMM, and (b) the LMM with each of the fixed effects (as well as any associated interaction terms) removed individually.

RNA-Seq study

RNA preparation

PHA-stimulated splenocyte cultures from the 64 voles collected in 2015 were used in the RNA-Seq experiment. RNA was extracted using Invitrogen PureLink kits. cDNA sequencing libraries were prepared using Illumina RiboZero kits to deplete rRNA followed by library construction with NEBNext Ultra directional RNA library prep kit according to manufacturers protocols. Samples were sequenced to produce 2 x 75bp paired-end reads on an Illumina HiSeq4000 platform. Adaptor sequences were removed using Cutadapt version 1.2.1 and further trimmed with Sickle version 1.200 with a minimum window quality score of 20. This resulted in a mean library size of 18 million (range = 5 – 50 million) paired-end reads.

Read mapping

High quality reads were mapped against a draft genome for *M. agrestis* (GenBank Accession no: LIQJ00000000), using TopHat version 2.1.0 (Trapnell, Pachter, & Salzberg, 2009).

BRAKER2 was used to generate a set of predicted gene models using mapped reads to guide Augustus (Hoff, Lange, Lomsadze, Borodovsky, & Stanke, 2015). Mapped reads were then counted using featureCounts (Liao, Smyth, & Shi, 2014). Further analysis of gene count data was performed in R version 3.4.0 (R Core Team, 2016) using the edgeR package (Robinson, McCarthy, & Smyth, 2010). Count data were filtered to remove those genes with fewer than 3 counts per million across all samples to avoid convergence problems later on. Following filtering, library sizes were recalculated, data were normalised and MDS plots were generated to check for any unusual patterns in the data.

Protein-protein interaction network construction

The STRING database version 10 (Szklarczyk et al., 2015) for *M. musculus* was used to construct a network of proteins known to interact with either Gata3 or Fcer1a using the stringApp in Cytoscape version 3.3.0 (Shannon et al., 2003). The default confidence score cut-off of 0.4 was used to extract those interactions that were well supported. The application PesCa version 3.0.8 (Scardoni, Tosadori, Pratap, Spoto, & Laudanna, 2016) was then used to extract the shortest paths between Fcer1a and Gata3 from this network.

Expression network construction

To validate the *M. musculus* network, which included seven genes (including Fcer1a and Gata3; see Results), we constructed a network for the same seven genes using the normalised count data. Spearman rank correlation coefficients were calculated for each combination of these genes, and associated *p*-values deduced from a null distribution composed of 2×10^8 coefficients generated from a randomised version of the dataset. Only statistically significant correlations ($p < 0.05$) were reported and included in the network. Two paralogous vole genes

were found for the mouse gene, Btk, but these were summarised as a single node in the vole network. This resulted in one pair of duplicated edges between these Btk paralogues and Jun - the more significant edge is presented in the network.

Results

The majority of SNPs were found to be in Hardy-Weinberg equilibrium (n = 259; 90%) and only four genes were found to have all SNPs departing from Hardy-Weinberg equilibrium: Gucy2f, Il13ra1, Tlr13, Tlr7 and Tlr8 (see Table S2 for summary of all loci). High LD was detected between SNPs within the same genes (mean D´ = 0.76; 95% CI = 0.72 – 0.81) but not between SNPs located in different genes (mean D´ = 0.28; 95% CI = 0.28 – 0.28).

Gata3 expression is associated with polymorphism in Fcer1a

Of the 84 immune and non-immune genes tested, only polymorphism in the gene Fcer1a was found to be significantly associated with Gata3 expression (q = 0.07; FDR cut-off = 0.1). Three haplotypes were identified at this locus: GCC, ACC and ACT at frequencies of 0.12, 0.76 and 0.07 respectively. The GCC haplotype was associated with lower expression levels of Gata3 than the ACC and ACT haplotypes (p = 0.003; 0.01; Fig. 1). This was confirmed by the LMM (p = 0.002; Table 1). No significant association was found between polymorphism in the Gata3 gene itself and Gata3 expression (q = 1.00).

The Fcer1a polymorphism is comparable in explanatory power to non-genetic variables previously identified as important predictors of Gata3 expression

The percentage variance in Gata3 expression explained by the fixed effects in the full model (or marginal R^2), including genotype was 10%. This dropped to about 5% when genotype, macroparasites or maturity were removed (individually) and to 8% when maturity × macroparasites was removed, indicating that genotype was comparable in explanatory power to other non-genetic variables previously identified as important predictors of Gata3 expression (Table 2). Furthermore, the greatest increase in AICc (relative to the full model) was observed when genotype was removed (ΔAICc = 7.7). However, a degree of overlap or multicollinearity between the variables was evident from these estimates.

Both Fcer1a and Gata3 are associated with Btk and Txk in the mouse model and vole

The *M. musculus* network included seven nodes (the proteins Fcer1a and Gata3, as well as Txk, Btk, Jun, Fos and Itk) and 18 edges (Fig. 2a). The *M. agrestis* network included six of these nodes connected by 10 edges (Fig. 2b). Itk could not be included as it was not annotated in the vole genome. Nine out of 18 of the edges in the *M. musculus* network were identified, in addition to a significant edge between Btk and Txk ($\rho = -0.32$; $p < 0.01$). Btk was found to be significantly correlated with both Fcer1a ($\rho = 0.26$, $p = 0.02$) and Gata3 ($\rho = -0.41$, $p < 0.001$), as was Txk (Fcer1a: $\rho = -0.23$, $p = 0.03$; Gata3: $\rho = 0.43$, $p < 0.001$).

Discussion

In this study, we have found an association between polymorphism in the gene Fcer1a and the expression of the transcription factor Gata3, which has previously been identified as a marker of tolerance to infection in this system. We have also shown that this polymorphism is

comparable in explaining power to other non-genetic variables previously identified as important predictors of Gata3 expression (Jackson et al., 2014).

Our results indicate that genotype has the potential to play an important role in driving consistent individual differences in immune gene expression in the wild (Arriero et al., in press). This suggests that individuals are, to a significant, detectable degree, hard-wired to respond in a certain way to challenges from parasites and pathogens. However, little is known about how natural selection acts on tolerance. Previous studies have found evidence for tolerance being less costly than resistance (Howick & Lazzaro, 2014). Under this scenario, one may expect tolerance to evolve more quickly and to to have lower levels of genetic variation than resistance (Råberg, 2014). Indeed, some evidence for positive directional selection on tolerance already exists (Hayward et al., 2014). However, genetic variation may also be maintained by temporal shifts in the strengths and directions of selection pressures. This may lead to low frequencies of individual haplotypes, as observed here.

Our results also shed light on the potential molecular and physiological mechanisms driving tolerance in the wild, which hitherto have been neglected. We find no effect of polymorphism in the Gata3 gene on its own expression, but rather a trans-acting effect of Fcer1a on Gata3 expression. By starting with an existing mouse PPI network and subsequently validating this using a novel vole expression dataset, we have also found evidence for a functionally relevant mechanism for this association. Fcer1a encodes the alpha chain of the high affinity receptor for immunoglobulin epsilon (IgE). This receptor is expressed on basophils, mast cells and eosinophils. When activated by an antigen interacting with Fcer1-bound IgE these cells promote a cascade of anti-macroparasitic Th2 responses, of which Gata3 is also a mediator.

This is reflected by the fact that, among other proteins, both Gata3 and Fcer1a are known to interact with two non-receptor kinases: Tyrosine-protein kinase BTK (Btk) and Tyrosine-protein kinase TXK (Txk). Btk plays a key role in B cell development, differentiation and signalling (Maas & Hendriks, 2001), and Txk exerts its effects on Th cell differentiation and function (Sahu et al., 2008). We were able to validate both of these interactions using our own expression network for *M. agrestis*. This suggests that the polymorphism in question may be working at the transcriptional level, leading to changes in the expression of Th2-related genes and hence potentially altering the strength of the Th2 response.

We focus here on tolerance, as this is a neglected area of study, but a diversity of immune strategies have been identified in natural populations (Abolins, Pocock, Hafalla, Riley, & Viney, 2011; Buehler et al., 2008). In our own study population of voles, we have shown a link between Gata3 expression and macroparasite resistance in immature male voles (Jackson et al., 2014), suggestive of an important role for Gata3 not just as a marker of tolerance, but more generally, of the immune strategy adopted by an individual. Indeed, this is consistent with previous work in a laboratory setting, which shows that polymorphism at a single locus can confer both resistance and tolerance (Ayres & Schneider, 2008; Miyairi et al., 2012). In the context of tolerance though, these results could have important implications for controlling the spread of disease, as high levels of tolerance can be associated with neutral or even positive effects on parasite prevalence (Miller, White, & Boots, 2006; Roy & Kirchner, 2000) and tolerant individuals can act as ‘superspreaders’, responsible for a large proportion of transmission events (Lloyd-Smith *et al.* 2005). In general, the identification of tolerance genes or haplotypes could facilitate the identification of such high-risk individuals, enabling more targeted control and helping to prevent the spread of disease in the wild. On the other

hand, tolerance is also associated with good health and condition despite infection, which could act as a potential pathway for the development of new treatments for infectious disease (Medzhitov et al., 2012; Råberg, 2014). Mapping out the network mediating the effects of a tolerance gene is a first step towards this. For these reasons, we believe this is an exciting and rare example of a candidate tolerance gene in a natural population, which we hope to continue monitoring to shed further light not only on tolerance, but on immune strategy more generally, in the wild.

Acknowledgements

The authors wish to thank the many individuals involved in obtaining and processing samples from the field, Kim Clarke for his technical advice and Deborah Dawson for her help with setting up the genotyping. They also wish to thank the Forestry Commission for access to the study sites and the Centre for Genomic Research for sequencing samples. This research was funded by the Natural Environment Research Council (NERC) award NE/E015131/1 to MB and NE/L013452/1 to SP.

References

Abolins, S. R., Pocock, M. J., Hafalla, J. C., Riley, E. M., & Viney, M. E. (2011). Measures of immune function of wild mice, Mus musculus. *Molecular Ecology*, *20*(5), 881–892. http://doi.org/10.1111/j.1365-294X.2010.04910.x

Allen, J. E., & Wynn, T. A. (2011). Evolution of Th2 Immunity: A Rapid Repair Response to Tissue Destructive Pathogens. *PLoS Pathogens*, *7*(5), 5–8. http://doi.org/10.1371/journal.ppat.1002003

Anthony, R. M., Rutitzky, L. I., Urban Jr., J. F., Stadecker, M. J., Gause, W. C., Urban, J. F., … Gause, W. C. (2007). Protective immune mechanisms in helminth infection. *Nature Reviews. Immunology*, *7*(12), 975–987. http://doi.org/10.1038/nri2199

Arriero, E., Wanelik, K. M., Birtles, R. J., Bradley, J. E., Jackson, J. A., Paterson, S., & Begon, M. (2017). From the animal house to the field: Are there consistent individual differences in immunological profile in wild populations of field voles (Microtus agrestis)? *PLoS ONE*, *12*(8), e0183450. http://doi.org/10.1371/journal.pone.0183450

Ayres, J. S., & Schneider, D. S. (2008). A signaling protease required for melanization in Drosophila affects resistance and tolerance of infections. *PLoS Biology*, *6*(12), 2764–2773. http://doi.org/10.1371/journal.pbio.0060305

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Boppana, V. D., Thangamani, S., Alarcon-Chaidez, F. J., Adler, A. J., & Wikel, S. K. (2009). Blood feeding by the Rocky Mountain spotted fever vector, Dermacentor andersoni, induces interleukin-4 expression by cognate antigen responding CD4+ T cells. *Parasites & Vectors*, *2*, 47. http://doi.org/10.1186/1756-3305-2-47

Boppana, V., Thangamani, S., AJ, A., & Wikel, S. (2009). SAAG-4 is a Novel Mosquito Salivary Protein that Programs Host CD4+ T Cells to Express IL-4. *Parasite Immunology*, *31*(6), 287–295. http://doi.org/10.1016/j.jneumeth.2010.08.011.Autogenic

Buehler, D. M., Piersma, T., Matson, K., & Tieleman, B. I. (2008). Seasonal redistribution of immune function in a migrant shorebird: Annual-cycle effects override adjustments to thermal regime. *The American Naturalist*, *172*(6), 783–96. http://doi.org/10.1086/592865

Burkett, K., Graham, J., & McNeney, B. (2006). hapassoc: Software for likelihood inference of trait associations with SNP haplotypes and other attributes. *Journal of Statistical Software*, *16*(2), 1–19.

Burkett, K., McNeney, B., & Graham, J. (2004). A note on inference of trait associations with SNP haplotypes and other attributes in generalized linear models. *Human Heredity*, *57*, 200–206. http://doi.org/https://doi.org/10.1159/000081447

Caldwell, R., Schafer, J., Compton, L., & Patterson, F. (1958). Tolerance to Cereal Leaf Rusts. *Science 128*, *128*, 714–715.

González, J. R., Armengol, L., Guinó, E., Solé, X., & Moreno, V. (2014). SNPassoc: SNPs-based whole genome association studies. R package version 1.9-2. Retrieved from https://cran.r-project.org/package=SNPassoc

Harris, N., & Gause, W. (2011). B cell function in the immune response to helminths. *Trends in Immunology*, *32*(2), 80–88. http://doi.org/10.1016/j.it.2010.11.005.B

Hayward, A. D., Nussey, D. H., Wilson, A. J., Berenos, C., Pilkington, J. G., Watt, K. A., … Graham, A. L. (2014). Natural selection on individual variation in tolerance of gastrointestinal nematode infection. *PLoS Biology*, *12*(7), e1001917. http://doi.org/10.1371/journal.pbio.1001917

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2015). BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, *32*(5), 767–769. http://doi.org/10.1093/bioinformatics/btv661

Howick, V. M., & Lazzaro, B. P. (2014). Genotype and diet shape resistance and tolerance across distinct phases of bacterial infection. *BMC Evolutionary Biology*, *14*, 56. http://doi.org/10.1186/1471-2148-14-56

Jackson, J. A., Begon, M., Birtles, R., Paterson, S., Friberg, I. M., Hall, A., … Bradley, J. E. (2011). The analysis of immunological profiles in wild animals: A case study on immunodynamics in the field vole, Microtus agrestis. *Molecular Ecology*, *20*, 893–909. http://doi.org/10.1111/j.1365-294X.2010.04907.x

Jackson, J. A., Hall, A. J., Friberg, I. M., Ralli, C., Lowe, A., Zawadzka, M., … Begon, M. (2014). An immunological marker of tolerance to infection in wild rodents. *PLoS Biology*, *12*(7), e1001901. http://doi.org/10.1371/journal.pbio.1001901

Kluen, E., Siitari, H., & Brommer, J. E. (2013). Testing for between individual correlations of personality and physiological traits in a wild bird. *Behavioral Ecology and Sociobiology*, *68*(2), 205–213. http://doi.org/10.1007/s00265-013-1635-1

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. http://doi.org/10.1093/bioinformatics/btt656

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., & Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, *438*, 355–359. http://doi.org/10.1038/nature04153

Louis, T. (1982). Finding the observed information matrix when using the EM Algorithm. *Journal of the Royal Statistical Society B*, *44*, 226–233.

Maas, A., & Hendriks, R. W. (2001). Role of Bruton's tyrosine kinase in B cell development. *Developmental Immunology*, *8*(3–4), 171–81. http://doi.org/10.1155/2001/28962

Medzhitov, R., Schneider, D. S., & Soares, M. P. (2012). Disease tolerance as a defense strategy. *Science*, *335*(6071), 936–941. http://doi.org/10.1126/science.1214935

Miller, M. R., White, A., & Boots, M. (2006). The Evolution of Parasites in Response To Tolerance in Their Hosts: the Good, the Bad, and Apparent Commensalism. *Evolution*, *60*(5), 945–956. http://doi.org/10.1554/05-654.1

Miyairi, I., Ziebarth, J., Laxton, J. D., Wang, X., van Rooijen, N., Williams, R. W., … Cui, Y. (2012). Host genetics and Chlamydia disease: Prediction and validation of disease severity mechanisms. *PLoS ONE*, *7*(3), e33781. http://doi.org/10.1371/journal.pone.0033781

O'Donovan, M. R., Johns, S., & Wilcox, P. (1995). The effect of PHA stimulation on lymphocyte sub-populations in whole-blood cultures. *Mutagenesis*, *10*(4), 371–374. http://doi.org/10.1093/mutage/10.4.371

Ortutay, C., & Vihinen, M. (2006). Immunome: A reference set of genes and proteins for systems biology of the human immune system. *Cellular Immunology*, *244*(2), 87–89. http://doi.org/10.1016/j.cellimm.2007.01.012

Råberg, L. (2014). How to live with the enemy: Understanding tolerance to parasites. *PLoS Biology*, *12*(11), e1001989. http://doi.org/10.1371/journal.pbio.1001989

Råberg, L., Graham, A. L., & Read, A. F. (2009). Decomposing health: Tolerance and resistance to parasites in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1513), 37–49. http://doi.org/10.1098/rstb.2008.0184

Raberg, L., Sim, D., & Read, A. (2007). Disentangling genetic variation for resistance and

tolerance to infectious diseases in animals. *Science*, *318*(5851), 812–814. http://doi.org/0.1126/science.1148526

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. Vienna, Austria. Retrieved from https://www.r-project.org/

Regoes, R. R., McLaren, P. J., Battegay, M., Bernasconi, E., Calmy, A., Günthard, H. F., … Fellay, J. (2014). Disentangling human tolerance and resistance against HIV. *PLoS Biology*, *12*(9), e1001951. http://doi.org/10.1371/journal.pbio.1001951

Restif, O., & Koella, J. C. (2004). Concurrent Evolution of Resistance and Tolerance to Pathogens. *The American Naturalist*, *164*(4), E90–E102. http://doi.org/10.1086/423713

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–40. http://doi.org/10.1093/bioinformatics/btp616

Roy, B. a, & Kirchner, J. W. (2000). Evolutionary dynamics of pathogen resistance and tolerance. *Evolution Int J Org Evolution*, *54*(1), 51–63. http://doi.org/10.1554/0014-3820(2000)054[0051:EDOPRA]2.0.CO;2

Sahu, N., Venegas, A. M., Jankovic, D., Mitzner, W., Gomez-Rodriguez, J., Cannons, J., … August, A. (2008). Selective expression rather than specific function of Txk and Itk regulate TH1 and TH2 responses. *Journal of Immunology*, *181*(9), 6125–6131.

Scardoni, G., Tosadori, G., Pratap, S., Spoto, F., & Laudanna, C. (2016). Finding the shortest path with PesCa: A tool for network reconstruction [version 2; referees: 2 approved, 2 approved with reservations]. *F1000Research*, *4*, 484. http://doi.org/10.12688/f1000research.6769.1

Schafer, J. F. (1971). Tolerance to Plant Disease. *Annual Review of Phytopathology*, *9*(1), 235–252. http://doi.org/10.1146/annurev.py.09.090171.001315

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. http://doi.org/10.1101/gr.1239303.metabolite

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-cepas, J., … Mering, C. Von. (2015). STRING v10: Protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*, 447–452. http://doi.org/10.1093/nar/gku1003

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105–1111. http://doi.org/10.1093/bioinformatics/btp120

Warnes, G., Gorjanc, G., Leisch, F., & Man, M. (2013). genetics: Population genetics. R package version 1.3.8.1. Retrieved from https://cran.r-project.org/package=genetics

Data Accessibility

RNA-Seq data will be deposited in the European Nucleotide Archive (ENA) on acceptance. SNP data and field data will be deposited in Dryad.

Author Contributions

MB, JAJ, RJB, JEB and SP designed the initial experiment (in 2008-2010) and MB, JAJ, JEB and SP designed the follow-up experiment (in 2015). AKT and SP were responsible for performing the genotyping experiment. IMJ and AGT performed RNA extractions. KMW was responsible for writing the manuscript. Both KMW and SP were responsible for analysing the data, and CHT produced Table S1. All authors contributed to editing the manuscript.

Tables

Table 1 Parameter estimates, standard errors and associated significance from LMM for Gata3 expression, including all fixed terms, random terms and interactions.

	Estimate	**SE**	***t***	***p***
(Intercept)	0.186	0.025	7.289	< 0.001
Maturity	-0.024	0.027	-0.911	0.367
Residual Weight	-0.003	0.007	-0.422	0.673
Macroparasites	-0.037	0.022	-1.655	0.101
Genotype	-0.065	0.021	-3.070	0.002
Maturity × Residual Weight	0.004	0.008	0.442	0.659
Maturity × Macroparasites	0.059	0.024	2.436	0.016

Table 2 Marginal and conditional R^2 estimates for LMM for Gata3 expression, with different variables removed (as well as associated interactions).

Model	Variable removed	Marginal R^2	Conditional R^2	AICc
Genotype, Maturity, Residual weight, Macroparasites, Maturity × Residual weight, Maturity × Macroparasites	NA	0.10	0.28	-193.68
Maturity, Residual weight, Macroparasites, Maturity × Residual weight, Maturity × Macroparasites	Genotype	0.05	0.22	-185.98
Genotype, Maturity, Residual weight, Macroparasites, Maturity × Macroparasites	Maturity × Residual weight	0.10	0.28	-195.71
Genotype, Maturity, Residual weight, Macroparasites, Maturity × Residual weight	Maturity × Macroparasites	0.08	0.25	-190.24
Genotype, Residual weight, Macroparasites	Maturity	0.05	0.21	-189.25
Genotype, Maturity, Macroparasites, Maturity × Macroparasites	Residual Weight	0.10	0.28	-197.94
Genotype, Maturity, Residual weight, Maturity × Residual weight	Macroparasites	0.06	0.21	-190.02

Figures

Fig. 1 Predicted Gata3 expression level for each haplotype (Parameters Site BLB, Maturity Immature, Residual Weight and Macroparasite Load set to 0; Error bars represent 95% confidence intervals; Gata3 expression levels are Box-Cox transformed).

Fig. 2 Panel figure showing (a) *M. musculus* protein-protein interaction network and (b) *M. agrestis* gene expression network. Edge weights represent (a) confidence scores (range = 0.50 – 1.00), or (b) *p*-values (range = < 0.001 – 0.03).