University of Salford MANCHESTER

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

Title	A candidate tolerance gene identified in a natural population of field voles (Microtus agrestis)
Authors	Wanelik, KM, Begon, M, Birtles, RJ, Bradley, JE, Friberg, IM, Jackson, JA, Taylor, CH, Thomason, AG, Turner, AK and Paterson, S
Type	Article
URL	This version is available at: http://usir.salford.ac.uk/id/eprint/44525/
Published Date	2017

USIR is a digital collection of the research output of the University of Salford. Where copyright permits, full text material held in the repository is made freely available online and can be read, downloaded and copied for non-commercial private study or research purposes. Please check the manuscript for any further copyright restrictions.

For more information, including our policy and submission procedure, please contact the Repository Team at: usir@salford.ac.uk.

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.

46 ## Introduction

47

48 Tolerance, like resistance, is an active response to infection involving the activation of
49 molecular and physiological mechanisms. Unlike resistance though, rather than preventing or
50 clearing an infection, a tolerance response minimises the disease pathology caused by
51 infection (Caldwell, Schafer, Compton, & Patterson, 1958; Schafer, 1971). This strategy may
52 be favoured where infection is a daily occurrence, or infection is persistent (Restif & Koella,
53 2004). In these cases, the costs of constantly mounting an immune response in terms of
54 damage to host tissue (immunopathology) may be worse than those of infection itself
55 (Medzhitov, Schneider, & Soares, 2012). A resistant strategy, on the other hand, might be
56 associated with acute exposure (Restif & Koella, 2004), where the costs of infection outweigh
57 those of mounting an immune response. Tolerance of infection is now attracting considerable
58 interest in the immunological and ecological literature (Medzhitov et al., 2012; Råberg,
59 Graham, & Read, 2009) and provides a new perspective to help understand how the immune
60 response in animals functions following infection, which has hitherto been viewed primarily
61 in the context of resistance only.

62

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

70 Read, 2007) and, to a more limited extent, in natural systems (Regoes et al., 2014). However,
71 knowledge of specific genes controlling tolerance, and hence potentially driving this heritable
72 variation in strategy in the wild, is lacking. Candidate genes include those involved in limiting
73 immunopathology and/or regulation of the immune response (Medzhitov et al., 2012; Råberg
74 et al., 2009). In the laboratory, a genetic locus on mouse chromosome 11 (Ctrq3) has been
75 shown to influence tolerance to *Chlamydia psittaci* infection, with circumstantial evidence for
76 candidate genes belonging to the family of immunity-related GTPases (Miyairi et al., 2012).
77 Another study has also identified a signalling protease required for melanisation in
78 *Drosophila melanogaster* (CG3066) as being of importance (Ayres & Schneider, 2008).
79 Finally, in humans, an association between HLA-B genotype and degree of tolerance to HIV
80 has been shown (Regoes et al., 2014).
81
82 Our own work has previously identified the expression of a particular master transcription
83 factor, Gata3, as a marker of tolerance in mature male field voles, *Microtus agrestis.* This
84 work showed that macroparasite infection in these mature voles gave rise to elevated levels of
85 Gata3 expression, which in turn gave rise to improved body condition and enhanced survival
86 (Jackson et al., 2014). This fits with the known role of Gata3 as a mediator of the Th2
87 response, and the role of the Th2 immune system in tissue repair (Allen & Wynn, 2011).
88 Furthermore, we have shown consistent differences between individuals in their typical level
89 of Gata3 expression, after other measured sources of variation have been taken into account
90 (Arriero et al., 2017). Together, our results imply consistent difference between individuals in
91 the strength of their tolerance response.
92

93 Here, we address the contribution of genotype to consistent individual differences in the
94 expression of Gata3, a marker of tolerance. We use a natural population of wild *M. agrestis* to
95 search for an association between Gata3 expression and polymorphism in 84 immune and
96 non-immune genes. We find Gata3 expression associated with polymorphism at the Fcer1a
97 gene (which encodes the alpha chain of the high affinity receptor for immunoglobulin epsilon,
98 IgE), and show that the proportion of variation in Gata3 expression explained by this
99 polymorphism is comparable to that explained by other environmental and physiological
100 variables. We also shed light on the possible mechanism behind this association by
101 constructing a protein-protein interaction network for the mouse model rodent, *Mus musculus*,
102 which we validate using our own expression network for *M. agrestis*.
103

104 Materials & Methods

105

106 Field design and animals

107

108 We studied *M. agrestis* in Kielder Forest, Northumberland (55º13' N, 2º3' W) using live-
109 trapping to access individual animals from natural populations. Our studies were designed to
110 permit the analysis of individual variation in host condition and survival, infection status, and
111 the expression of immune genes (for full details of all methods below see Jackson et al. 2011,
112 2014). The studies were divided into longitudinal and cross-sectional components.
113

114 Initial survey

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

117 (~0.375 ha) of 150 (10 x 15) regularly spaced traps (3–5 m intervals) which was used in a
118 capture-recapture study (reported elsewhere). The cross-sectional component reported here
119 utilised curvilinear transects of 100 live traps arranged at 5–10 m intervals which were placed
120 around the margins of each habitat.
121
122 Ten voles per month were destructively sampled from the transects between February and
123 November (2008–2009) or April and November (2009–2010). In November (2008 and 2009)
124 and March (2009 and 2010), larger numbers of animals were sampled both from the transects
125 and from the central grid habitats. These samples are used here to carry out a haplotype
126 association analysis.
127
128 On capture, each animal was examined for ectoparasites (see below). Only results for male *M.*
129 *agrestis* are reported here given the focus of previous work (Jackson et al., 2014). Males were
130 classified as either immature (non-mating with undeveloped testes) or mature (mating with
131 large testes and expanded seminal vesicles). Some biometric data were also collected,
132 including body weight (g) and snout-vent length (mm). All animal procedures carried out as
133 part of this initial survey were performed with approval from the University of Liverpool
134 Animal Welfare Committee and under a UK Home Office license (PPL 40/3235 to MB).
135
136 ***Parasite assays***
137 On capture, ectoparasite infections were recorded, as direct counts of ticks (*Ixodes* spp.) and
138 small flea species (*Ctenophthalmus nobilis*, *Peromyscopsylla spectabilis*, *Megabothris*
139 *walkeri*, *Malaraeus penicilliger*, *Rhadinopsylla pentacantha*). Captured animals were then
140 returned to the laboratory where they were killed by an overdose of chloroform followed by

141 exsanguination, and dissection in order to take a more comprehensive set of infection
142 measurements. This included a direct count of adult cestodes found in the gut
143 (*Anoplocephaloides dentata* aff., *Paranoplocephala* sp., *Rodentolepis asymmetrica*,
144 *Arostrilepis horrida*). We collected infection metrics for these macroparasites because they
145 are the most common species that would be expected to be in strong contact with the host
146 immune system (Jackson et al., 2014).
147

148 **Follow-up survey**

149 In 2015, we collected samples at four sites (GRD, CHE, SCP, COL). Similarly to the initial
150 survey, each site contained a trapping grid of 150-197 regularly spaces traps (at approx. 5 m
151 intervals) but this was used both for cross-sectional and longitudinal components (not
152 reported here). Sixty-four voles were also destructively sampled from the grids between July
153 and October 2015 to assay expression by RNA-seq. Both females and males were included in
154 order to maximise sample size. In this study, voles were killed by a rising concentration of
155 CO_2, followed by exsanguination. These samples were shown to be comparable in terms of
156 weight, age and sex to the population sampled in the initial survey (Table S1) and are used
157 here to construct an expression network for *M. agrestis*. All animal procedures carried out as
158 part of this survey were performed with approval from the University of Liverpool Animal
159 Welfare Committee and under a UK Home Office license (PPL ??? to SP).
160

161 Haplotype association study

162

163 **Immunological assays**

164 We used two-step reverse transcription quantitative real-time PCR (Q-PCR) to measure
165 messenger RNA (mRNA) accumulation of Gata binding protein 3 (Gata3; a transcription
166 factor associated with the Th2 response) from splenocyte cultures stimulated with mitogen
167 phytohaemagglutinin (PHA). Gata3 has previously been identified as a marker of tolerance in
168 mature male voles (Jackson et al., 2014). PHA preferentially activates and stimulates
169 proliferation of CD4+ helper T-cells in vitro (O'Donovan, Johns, & Wilcox, 1995). Here, we
170 use that observed expression profile as a measure of the potential responsiveness of the
171 immune system in vivo.
172

173 **SNP identification and genotyping**

174 We identified 288 single nucleotide polymorphisms (SNPs) in 85 immune-related genes and
175 25 non-immune genes. Immune genes included cytokine genes and other genes known to be
176 involved in pathogen resistance. The Immunome database version 1.1.
177 (http://structure.bmc.lu.se/idbase/Immunome/index.php), a manually curated database
178 containing information on 893 genes considered essential to the human immune system, was
179 used a starting point for identifying a list of candidate immune genes (Ortutay & Vihinen,
180 2006). First, we excluded all those genes in this database with no known orthologue in house
181 mice. We then applied a heuristic approach to ensure that those genes which were most likely
182 to be of interest given our previous work (e.g. Jackson *et al.* 2014) were represented in our
183 list, and excluded those genes with no known polymorphisms in *M. agrestis*. We also chose a
184 set of non-immune genes to act as a control for spurious associations, caused, for example, by
185 demographic effects. This set was composed solely of metabolic genes, as these are far less
186 likely to be involved in host-pathogen interactions (see Table S2 for full list of immune and
187 non-immune genes identified).

188

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

194

195 **Statistical analyses**

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

207

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

212 with SNP haplotypes and other attributes, adopts a generalized linear model framework and
213 estimates parameters using an expectation-maximization algorithm. If the haplotype
214 combination of an individual cannot be inferred from its genotyping data (a) because it is
215 heterozygous at two or more markers or (b) because it has missing data for a single marker,
216 the approach implemented in hapassoc is to consider all possible haplotype combinations for
217 that individual. Standard errors accounting for this added uncertainty are calculated using the
218 Louis' method (Louis, 1982).
219
220 We assumed an additive genetic model, where Gata3 expression is linearly related to the
221 number of copies of a haplotype present and we pooled together all those haplotypes with
222 frequencies below 5%. Gata3 expression values were Box-Cox transformed to achieve
223 approximately normal residuals. Other non-genetic covariates included in this model were site
224 (BLB, SQC, SCP & KTH), maturity (either immature or mature male), residual weight
225 (adjusted for body size) and the first principal component from a PCA summarising the
226 macroparasites measured. This component explained 47% of the variation in macroparasite
227 burden and showed high positive loadings for all three macroparasite groups (ticks: 0.56,
228 fleas: 0.57 and adult cestodes: 0.60). Grouping of ectoparasites and endoparasites in this way
229 is in line with previous work that shows that both ectoparasites (V. D. Boppana, Thangamani,
230 Alarcon-Chaidez, Adler, & Wikel, 2009; V. Boppana, Thangamani, AJ, & Wikel, 2009) and
231 endoparasites (Anthony et al., 2007; Harris & Gause, 2011) stimulate the Th2 response,
232 which has been suggested to act "as an adaptive tissue repair mechanism that quickly heals
233 the wounds they inflict" (Allen & Wynn, 2011). These variables have previously been
234 identified as important predictors of Gata3 expression (Jackson et al., 2011, 2014). All non-

235 genetic covariates were tested for independence (Pearson correlation coefficients = -0.2 -
236 0.35).
237
238 As required by the hapassoc package, we excluded all genes with a single SNP and all
239 monomorphic SNPs (see Table S2 for these), resulting in a total of 238 SNPs in 62 immune-
240 related genes and 22 non-immune genes being included in the analysis (see Table S3 for final
241 list of immune and non-immune genes). We also excluded those subjects for which more than
242 one single-locus genotype had missing data. Because of the large number of association tests
243 performed, the Benjamini and Hochberg method of correction was applied to all *p*-values,
244 with the false discovery rate set to 0.1 (Benjamini & Hochberg, 1995). Resulting *q*-values
245 (FDR-corrected *p*-values) were checked for a uniform distribution.
246
247 We were unable to include any random variables or interaction terms in the initial trait
248 association analysis, as the hapassoc package does not have his capability. So, following this
249 analysis, a linear mixed effects model (LMM) for Gata3 expression was constructed to
250 confirm these results. This included season [four levels, designated as: spring (March to
251 May), summer (June to August), autumn (September to November) and winter (December to
252 February)], assay plate number and site as random effects. It also included previously
253 identified interactions between maturity and macroparasitic load, as well as maturity and
254 residual weight (Jackson *et al.* 2014). Three haplotypes were identified at this locus: GCC,
255 ACC and ACT. GCC was found to be the haplotype of interest in relation to Gata3 expression
256 in the initial trait association analysis (see Results). Genotype was therefore coded as a the
257 number of GCC copies. This was treated as a continuous variable because only five
258 individuals were found to have two copies of the GCC haplotype, making it difficult to make

259 reliable comparisons between factor levels. Treatment of genotype as a continuous variable
260 also reduced the number of degrees of freedom by one. Only those individuals whose
261 combination of haplotypes or 'haplotype phase' could be determined with certainty were
262 included in this analysis, but this was the majority of individuals (n = 191; 86%). The
263 contribution of genotype relative to other predictors in explaining variance in Gata3
264 expression was assessed by calculating the marginal R^2 using the MuMIn package (Barton,
265 2016) for (a) the full LMM, and (b) the LMM with each of the fixed effects (as well as any
266 associated interaction terms) removed individually.
267
268 **RNA-Seq study**
269
270 **RNA preparation**
271 PHA-stimulated splenocyte cultures from the 64 voles collected in 2015 were used in the
272 RNA-Seq experiment. RNA was extracted using Invitrogen PureLink kits. cDNA sequencing
273 libraries were prepared using Illumina RiboZero kits to deplete rRNA followed by library
274 construction with NEBNext Ultra directional RNA library prep kit according to
275 manufacturers protocols. Samples were sequenced to produce 2 x 75bp paired-end reads on an
276 Illumina HiSeq4000 platform. Adaptor sequences were removed using Cutadapt version 1.2.1
277 and further trimmed with Sickle version 1.200 with a minimum window quality score of 20.
278 This resulted in a mean library size of 18 million (range = 5 – 50 million) paired-end reads.
279
280 **Read mapping**
281 High quality reads were mapped against a draft genome for *M. agrestis* (GenBank Accession
282 no: LIQJ00000000), using TopHat version 2.1.0 (Trapnell, Pachter, & Salzberg, 2009).

283 BRAKER2 was used to generate a set of predicted gene models using mapped reads to guide
284 Augustus (Hoff, Lange, Lomsadze, Borodovsky, & Stanke, 2015). Mapped reads were then
285 counted using featureCounts (Liao, Smyth, & Shi, 2014). Further analysis of gene count data
286 was performed in R version 3.4.0 (R Core Team, 2016) using the edgeR package (Robinson,
287 McCarthy, & Smyth, 2010). Count data were filtered to remove those genes with fewer than 3
288 counts per million across all samples to avoid convergence problems later on. Following
289 filtering, library sizes were recalculated, data were normalised and MDS plots were generated
290 to check for any unusual patterns in the data.
291
292 **Protein-protein interaction network construction**
293 The STRING database version 10 (Szklarczyk et al., 2015) for *M. musculus* was used to
294 construct a network of proteins known to interact with either Gata3 or Fcer1a using the
295 stringApp in Cytoscape version 3.3.0 (Shannon et al., 2003). The default confidence score
296 cut-off of 0.4 was used to extract those interactions that were well supported. The application
297 PesCa version 3.0.8 (Scardoni, Tosadori, Pratap, Spoto, & Laudanna, 2016) was then used to
298 extract the shortest paths between Fcer1a and Gata3 from this network.
299
300 **Expression network construction**
301 To validate the *M. musculus* network, which included seven genes (including Fcer1a and
302 Gata3; see Results), we constructed a network for the same seven genes using the normalised
303 count data. Spearman rank correlation coefficients were calculated for each combination of
304 these genes, and associated *p*-values deduced from a null distribution composed of 2×10^8
305 coefficients generated from a randomised version of the dataset. Only statistically significant
306 correlations ($p < 0.05$) were reported and included in the network. Two paralogous vole genes

307 were found for the mouse gene, Btk, but these were summarised as a single node in the vole
308 network. This resulted in one pair of duplicated edges between these Btk paralogues and Jun -
309 the more significant edge is presented in the network.
310

311 Results

312

313 The majority of SNPs were found to be in Hardy-Weinberg equilibrium (n = 259; 90%) and
314 only four genes were found to have all SNPs departing from Hardy-Weinberg equilibrium:
315 Gucy2f, Il13ra1, Tlr13, Tlr7 and Tlr8 (see Table S2 for summary of all loci). High LD was
316 detected between SNPs within the same genes (mean D´ = 0.76; 95% CI = 0.72 – 0.81) but
317 not between SNPs located in different genes (mean D´ = 0.28; 95% CI = 0.28 – 0.28).
318

319 **Gata3 expression is associated with polymorphism in Fcer1a**
320 Of the 84 immune and non-immune genes tested, only polymorphism in the gene Fcer1a was
321 found to be significantly associated with Gata3 expression (q = 0.07; FDR cut-off = 0.1).
322 Three haplotypes were identified at this locus: GCC, ACC and ACT at frequencies of 0.12,
323 0.76 and 0.07 respectively. The GCC haplotype was associated with lower expression levels
324 of Gata3 than the ACC and ACT haplotypes (p = 0.003; 0.01; Fig. 1). This was confirmed by
325 the LMM (p = 0.002; Table 1). No significant association was found between polymorphism
326 in the Gata3 gene itself and Gata3 expression (q = 1.00).
327

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

330 The percentage variance in Gata3 expression explained by the fixed effects in the full model

331 (or marginal R^2), including genotype was 10%. This dropped to about 5% when genotype,

332 macroparasites or maturity were removed (individually) and to 8% when maturity ×

333 macroparasites was removed, indicating that genotype was comparable in explanatory power

334 to other non-genetic variables previously identified as important predictors of Gata3

335 expression (Table 2). Furthermore, the greatest increase in AICc (relative to the full model)

336 was observed when genotype was removed (ΔAICc = 7.7). However, a degree of overlap or

337 multicollinearity between the variables was evident from these estimates.

338

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

340 The *M. musculus* network included seven nodes (the proteins Fcer1a and Gata3, as well as

341 Txk, Btk, Jun, Fos and Itk) and 18 edges (Fig. 2a). The *M. agrestis* network included six of

342 these nodes connected by 10 edges (Fig. 2b). Itk could not be included as it was not annotated

343 in the vole genome. Nine out of 18 of the edges in the *M. musculus* network were identified,

344 in addition to a significant edge between Btk and Txk ($\rho = -0.32$; $p < 0.01$). Btk was found to

345 be significantly correlated with both Fcer1a ($\rho = 0.26$, $p = 0.02$) and Gata3 ($\rho = -0.41$, $p <$

346 0.001), as was Txk (Fcer1a: $\rho = -0.23$, $p = 0.03$; Gata3: $\rho = 0.43$, $p < 0.001$).

347

348 ## Discussion

349

350 In this study, we have found an association between polymorphism in the gene Fcer1a and the

351 expression of the transcription factor Gata3, which has previously been identified as a marker

352 of tolerance to infection in this system. We have also shown that this polymorphism is

353 comparable in explaining power to other non-genetic variables previously identified as
354 important predictors of Gata3 expression (Jackson et al., 2014).
355
356 Our results indicate that genotype has the potential to play an important role in driving
357 consistent individual differences in immune gene expression in the wild (Arriero et al., in
358 press). This suggests that individuals are, to a significant, detectable degree, hard-wired to
359 respond in a certain way to challenges from parasites and pathogens. However, little is known
360 about how natural selection acts on tolerance. Previous studies have found evidence for
361 tolerance being less costly than resistance (Howick & Lazzaro, 2014). Under this scenario,
362 one may expect tolerance to evolve more quickly and to to have lower levels of genetic
363 variation than resistance (Råberg, 2014). Indeed, some evidence for positive directional
364 selection on tolerance already exists (Hayward et al., 2014). However, genetic variation may
365 also be maintained by temporal shifts in the strengths and directions of selection pressures.
366 This may lead to low frequencies of individual haplotypes, as observed here.
367
368 Our results also shed light on the potential molecular and physiological mechanisms driving
369 tolerance in the wild, which hitherto have been neglected. We find no effect of polymorphism
370 in the Gata3 gene on its own expression, but rather a trans-acting effect of Fcer1a on Gata3
371 expression. By starting with an existing mouse PPI network and subsequently validating this
372 using a novel vole expression dataset, we have also found evidence for a functionally relevant
373 mechanism for this association. Fcer1a encodes the alpha chain of the high affinity receptor
374 for immunoglobulin epsilon (IgE). This receptor is expressed on basophils, mast cells and
375 eosinophils. When activated by an antigen interacting with Fcer1-bound IgE these cells
376 promote a cascade of anti-macroparasitic Th2 responses, of which Gata3 is also a mediator.

377 This is reflected by the fact that, among other proteins, both Gata3 and Fcer1a are known to
378 interact with two non-receptor kinases: Tyrosine-protein kinase BTK (Btk) and Tyrosine-
379 protein kinase TXK (Txk). Btk plays a key role in B cell development, differentiation and
380 signalling (Maas & Hendriks, 2001), and Txk exerts its effects on Th cell differentiation and
381 function (Sahu et al., 2008). We were able to validate both of these interactions using our own
382 expression network for *M. agrestis*. This suggests that the polymorphism in question may be
383 working at the transcriptional level, leading to changes in the expression of Th2-related genes
384 and hence potentially altering the strength of the Th2 response.
385
386 We focus here on tolerance, as this is a neglected area of study, but a diversity of immune
387 strategies have been identified in natural populations (Abolins, Pocock, Hafalla, Riley, &
388 Viney, 2011; Buehler et al., 2008). In our own study population of voles, we have shown a
389 link between Gata3 expression and macroparasite resistance in immature male voles (Jackson
390 et al., 2014), suggestive of an important role for Gata3 not just as a marker of tolerance, but
391 more generally, of the immune strategy adopted by an individual. Indeed, this is consistent
392 with previous work in a laboratory setting, which shows that polymorphism at a single locus
393 can confer both resistance and tolerance (Ayres & Schneider, 2008; Miyairi et al., 2012). In
394 the context of tolerance though, these results could have important implications for
395 controlling the spread of disease, as high levels of tolerance can be associated with neutral or
396 even positive effects on parasite prevalence (Miller, White, & Boots, 2006; Roy & Kirchner,
397 2000) and tolerant individuals can act as 'superspreaders', responsible for a large proportion
398 of transmission events (Lloyd-Smith *et al.* 2005). In general, the identification of tolerance
399 genes or haplotypes could facilitate the identification of such high-risk individuals, enabling
400 more targeted control and helping to prevent the spread of disease in the wild. On the other

401 hand, tolerance is also associated with good health and condition despite infection, which
402 could act as a potential pathway for the development of new treatments for infectious disease
403 (Medzhitov et al., 2012; Råberg, 2014). Mapping out the network mediating the effects of a
404 tolerance gene is a first step towards this. For these reasons, we believe this is an exciting and
405 rare example of a candidate tolerance gene in a natural population, which we hope to continue
406 monitoring to shed further light not only on tolerance, but on immune strategy more
407 generally, in the wild.
408

409 ## Acknowledgements

410 The authors wish to thank the many individuals involved in obtaining and processing samples
411 from the field, Kim Clarke for his technical advice and Deborah Dawson for her help with
412 setting up the genotyping. They also wish to thank the Forestry Commission for access to the
413 study sites and the Centre for Genomic Research for sequencing samples. This research was
414 funded by the Natural Environment Research Council (NERC) award NE/E015131/1 to MB
415 and NE/L013452/1 to SP.
416
417
418
419
420
421
422
423
424

425 ## References

426

427 Abolins, S. R., Pocock, M. J., Hafalla, J. C., Riley, E. M., & Viney, M. E. (2011). Measures
428 of immune function of wild mice, Mus musculus. *Molecular Ecology*, *20*(5), 881–892.
429 http://doi.org/10.1111/j.1365-294X.2010.04910.x
430 Allen, J. E., & Wynn, T. A. (2011). Evolution of Th2 Immunity: A Rapid Repair Response to
431 Tissue Destructive Pathogens. *PLoS Pathogens*, *7*(5), 5–8.
432 http://doi.org/10.1371/journal.ppat.1002003
433 Anthony, R. M., Rutitzky, L. I., Urban Jr., J. F., Stadecker, M. J., Gause, W. C., Urban, J. F.,
434 … Gause, W. C. (2007). Protective immune mechanisms in helminth infection. *Nature*
435 *Reviews. Immunology*, *7*(12), 975–987. http://doi.org/10.1038/nri2199
436 Arriero, E., Wanelik, K. M., Birtles, R. J., Bradley, J. E., Jackson, J. A., Paterson, S., &
437 Begon, M. (2017). From the animal house to the field: Are there consistent individual
438 differences in immunological profile in wild populations of field voles (Microtus
439 agrestis)? *PLoS ONE*, *12*(8), e0183450. http://doi.org/10.1371/journal.pone.0183450
440 Ayres, J. S., & Schneider, D. S. (2008). A signaling protease required for melanization in
441 Drosophila affects resistance and tolerance of infections. *PLoS Biology*, *6*(12), 2764–
442 2773. http://doi.org/10.1371/journal.pbio.0060305
443 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and
444 powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*
445 *(Methodological)*, *57*(1), 289–300.
446 Boppana, V. D., Thangamani, S., Alarcon-Chaidez, F. J., Adler, A. J., & Wikel, S. K. (2009).
447 Blood feeding by the Rocky Mountain spotted fever vector, Dermacentor andersoni,
448 induces interleukin-4 expression by cognate antigen responding CD4+ T cells. *Parasites*
449 *& Vectors*, *2*, 47. http://doi.org/10.1186/1756-3305-2-47
450 Boppana, V., Thangamani, S., AJ, A., & Wikel, S. (2009). SAAG-4 is a Novel Mosquito
451 Salivary Protein that Programs Host CD4+ T Cells to Express IL-4. *Parasite*
452 *Immunology*, *31*(6), 287–295. http://doi.org/10.1016/j.jneumeth.2010.08.011.Autogenic
453 Buehler, D. M., Piersma, T., Matson, K., & Tieleman, B. I. (2008). Seasonal redistribution of
454 immune function in a migrant shorebird: Annual-cycle effects override adjustments to
455 thermal regime. *The American Naturalist*, *172*(6), 783–96. http://doi.org/10.1086/592865
456 Burkett, K., Graham, J., & McNeney, B. (2006). hapassoc: Software for likelihood inference
457 of trait associations with SNP haplotypes and other attributes. *Journal of Statistical*
458 *Software*, *16*(2), 1–19.
459 Burkett, K., McNeney, B., & Graham, J. (2004). A note on inference of trait associations with
460 SNP haplotypes and other attributes in generalized linear models. *Human Heredity*, *57*,
461 200–206. http://doi.org/https://doi.org/10.1159/000081447
462 Caldwell, R., Schafer, J., Compton, L., & Patterson, F. (1958). Tolerance to Cereal Leaf
463 Rusts. *Science 128*, *128*, 714–715.
464 González, J. R., Armengol, L., Guinó, E., Solé, X., & Moreno, V. (2014). SNPassoc: SNPs-
465 based whole genome association studies. R package version 1.9-2. Retrieved from
466 https://cran.r-project.org/package=SNPassoc
467 Harris, N., & Gause, W. (2011). B cell function in the immune response to helminths. *Trends*
468 *in Immunology*, *32*(2), 80–88. http://doi.org/10.1016/j.it.2010.11.005.B
469 Hayward, A. D., Nussey, D. H., Wilson, A. J., Berenos, C., Pilkington, J. G., Watt, K. A., …
470 Graham, A. L. (2014). Natural selection on individual variation in tolerance of
471 gastrointestinal nematode infection. *PLoS Biology*, *12*(7), e1001917.
472 http://doi.org/10.1371/journal.pbio.1001917

473 Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2015). BRAKER1:
474 Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and
475 AUGUSTUS. *Bioinformatics*, *32*(5), 767–769.
476 http://doi.org/10.1093/bioinformatics/btv661
477 Howick, V. M., & Lazzaro, B. P. (2014). Genotype and diet shape resistance and tolerance
478 across distinct phases of bacterial infection. *BMC Evolutionary Biology*, *14*, 56.
479 http://doi.org/10.1186/1471-2148-14-56
480 Jackson, J. A., Begon, M., Birtles, R., Paterson, S., Friberg, I. M., Hall, A., … Bradley, J. E.
481 (2011). The analysis of immunological profiles in wild animals: A case study on
482 immunodynamics in the field vole, Microtus agrestis. *Molecular Ecology*, *20*, 893–909.
483 http://doi.org/10.1111/j.1365-294X.2010.04907.x
484 Jackson, J. A., Hall, A. J., Friberg, I. M., Ralli, C., Lowe, A., Zawadzka, M., … Begon, M.
485 (2014). An immunological marker of tolerance to infection in wild rodents. *PLoS*
486 *Biology*, *12*(7), e1001901. http://doi.org/10.1371/journal.pbio.1001901
487 Kluen, E., Siitari, H., & Brommer, J. E. (2013). Testing for between individual correlations of
488 personality and physiological traits in a wild bird. *Behavioral Ecology and Sociobiology*,
489 *68*(2), 205–213. http://doi.org/10.1007/s00265-013-1635-1
490 Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose
491 program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–
492 930. http://doi.org/10.1093/bioinformatics/btt656
493 Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., & Getz, W. M. (2005). Superspreading and
494 the effect of individual variation on disease emergence. *Nature*, *438*, 355–359.
495 http://doi.org/10.1038/nature04153
496 Louis, T. (1982). Finding the observed information matrix when using the EM Algorithm.
497 *Journal of the Royal Statistical Society B*, *44*, 226–233.
498 Maas, A., & Hendriks, R. W. (2001). Role of Bruton's tyrosine kinase in B cell development.
499 *Developmental Immunology*, *8*(3–4), 171–81. http://doi.org/10.1155/2001/28962
500 Medzhitov, R., Schneider, D. S., & Soares, M. P. (2012). Disease tolerance as a defense
501 strategy. *Science*, *335*(6071), 936–941. http://doi.org/10.1126/science.1214935
502 Miller, M. R., White, A., & Boots, M. (2006). The Evolution of Parasites in Response To
503 Tolerance in Their Hosts: the Good, the Bad, and Apparent Commensalism. *Evolution*,
504 *60*(5), 945–956. http://doi.org/10.1554/05-654.1
505 Miyairi, I., Ziebarth, J., Laxton, J. D., Wang, X., van Rooijen, N., Williams, R. W., … Cui, Y.
506 (2012). Host genetics and Chlamydia disease: Prediction and validation of disease
507 severity mechanisms. *PLoS ONE*, *7*(3), e33781.
508 http://doi.org/10.1371/journal.pone.0033781
509 O'Donovan, M. R., Johns, S., & Wilcox, P. (1995). The effect of PHA stimulation on
510 lymphocyte sub-populations in whole-blood cultures. *Mutagenesis*, *10*(4), 371–374.
511 http://doi.org/10.1093/mutage/10.4.371
512 Ortutay, C., & Vihinen, M. (2006). Immunome: A reference set of genes and proteins for
513 systems biology of the human immune system. *Cellular Immunology*, *244*(2), 87–89.
514 http://doi.org/10.1016/j.cellimm.2007.01.012
515 Råberg, L. (2014). How to live with the enemy: Understanding tolerance to parasites. *PLoS*
516 *Biology*, *12*(11), e1001989. http://doi.org/10.1371/journal.pbio.1001989
517 Råberg, L., Graham, A. L., & Read, A. F. (2009). Decomposing health: Tolerance and
518 resistance to parasites in animals. *Philosophical Transactions of the Royal Society B:*
519 *Biological Sciences*, *364*(1513), 37–49. http://doi.org/10.1098/rstb.2008.0184
520 Raberg, L., Sim, D., & Read, A. (2007). Disentangling genetic variation for resistance and

521 tolerance to infectious diseases in animals. *Science*, *318*(5851), 812–814.
522 http://doi.org/0.1126/science.1148526
523 R Core Team. (2016). R: A language and environment for statistical computing. R
524 Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
525 Vienna, Austria. Retrieved from https://www.r-project.org/
526 Regoes, R. R., McLaren, P. J., Battegay, M., Bernasconi, E., Calmy, A., Günthard, H. F., …
527 Fellay, J. (2014). Disentangling human tolerance and resistance against HIV. *PLoS*
528 *Biology*, *12*(9), e1001951. http://doi.org/10.1371/journal.pbio.1001951
529 Restif, O., & Koella, J. C. (2004). Concurrent Evolution of Resistance and Tolerance to
530 Pathogens. *The American Naturalist*, *164*(4), E90–E102. http://doi.org/10.1086/423713
531 Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package
532 for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1),
533 139–40. http://doi.org/10.1093/bioinformatics/btp616
534 Roy, B. a, & Kirchner, J. W. (2000). Evolutionary dynamics of pathogen resistance and
535 tolerance. *Evolution Int J Org Evolution*, *54*(1), 51–63. http://doi.org/10.1554/0014-
536 3820(2000)054[0051:EDOPRA]2.0.CO;2
537 Sahu, N., Venegas, A. M., Jankovic, D., Mitzner, W., Gomez-Rodriguez, J., Cannons, J., …
538 August, A. (2008). Selective expression rather than specific function of Txk and Itk
539 regulate TH1 and TH2 responses. *Journal of Immunology*, *181*(9), 6125–6131.
540 Scardoni, G., Tosadori, G., Pratap, S., Spoto, F., & Laudanna, C. (2016). Finding the shortest
541 path with PesCa: A tool for network reconstruction [version 2; referees: 2 approved, 2
542 approved with reservations]. *F1000Research*, *4*, 484.
543 http://doi.org/10.12688/f1000research.6769.1
544 Schafer, J. F. (1971). Tolerance to Plant Disease. *Annual Review of Phytopathology*, *9*(1),
545 235–252. http://doi.org/10.1146/annurev.py.09.090171.001315
546 Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T.
547 (2003). Cytoscape: A software environment for integrated models of biomolecular
548 interaction networks. *Genome Research*, *13*(11), 2498–2504.
549 http://doi.org/10.1101/gr.1239303.metabolite
550 Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-cepas, J., …
551 Mering, C. Von. (2015). STRING v10: Protein–protein interaction networks, integrated
552 over the tree of life. *Nucleic Acids Research*, *43*, 447–452.
553 http://doi.org/10.1093/nar/gku1003
554 Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with
555 RNA-Seq. *Bioinformatics*, *25*(9), 1105–1111.
556 http://doi.org/10.1093/bioinformatics/btp120
557 Warnes, G., Gorjanc, G., Leisch, F., & Man, M. (2013). genetics: Population genetics. R
558 package version 1.3.8.1. Retrieved from https://cran.r-project.org/package=genetics
559
560
561
562
563
564
565
566
567
568

569 ## Data Accessibility

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

572

573 ## Author Contributions

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

580
581
582
583
584
585
586
587
588
589
590
591
592

593 # Tables
594
595 **Table 1** Parameter estimates, standard errors and associated significance from LMM for
596 Gata3 expression, including all fixed terms, random terms and interactions.
597

	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

598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623

624 **Table 2** Marginal and conditional R^2 estimates for LMM for Gata3 expression, with different
625 variables removed (as well as associated interactions).
626
627

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

628
629
630
631
632
633
634
635
636
637

638
639
640 ## Figures
641
642 **Fig. 1** Predicted Gata3 expression level for each haplotype (Parameters Site BLB, Maturity
643 Immature, Residual Weight and Macroparasite Load set to 0; Error bars represent 95%
644 confidence intervals; Gata3 expression levels are Box-Cox transformed).
645
646 **Fig. 2** Panel figure showing (a) *M. musculus* protein-protein interaction network and (b) *M.*
647 *agrestis* gene expression network. Edge weights represent (a) confidence scores (range = 0.50
648 – 1.00), or (b) *p*-values (range = < 0.001 – 0.03).
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685