bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A locus conferring tolerance to Theileria infection in African cattle

Wragg, D.[1,2§], Cook, E.A.J.[3,4§], Latre de Late, P.[3,4], Sitt, T.[4], Hemmink, J.D.[1,3,3,4], Makau, M.[3,4], Njeru, R.[3,4], Powell, J.[2], Paxton, E.[2], Callaby, R.[2,5], Talenti, A.[2], Miyunga, A.A.[3,4], Ndambuki, G.[3,4], Mwaura, S.[4], Auty, H.[6], Matika, O.[2], Hassan, M.[1,2], Marshall, K.[3,4], Connelley, T.[1,2], Morrison, L.J.[1,2], Bronsvoort, B.M.deC.[1,2], Morrison, W.I.[1,2], Toye, P.[3,4†*], Prendergast, J.G.D.[1,2†*]

[1] Centre for Tropical Livestock Genetics and Health (CTLGH), Easter Bush Campus, EH25 9RG, UK

[2] The Roslin Institute, University of Edinburgh, UK

[3] Centre for Tropical Livestock Genetics and Health (CTLGH), ILRI Kenya, P.O. Box 30709, Nairobi 00100, Kenya

[4] ILRI Kenya, P.O. Box 30709, Nairobi 00100, Kenya

[5] The Epidemiology, Economics and Risk Assessment (EEA) Group, Easter Bush Campus, EH25 9RG, UK

[6] Institute of Biodiversity Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, G12 8QQ

[§] David Wragg and Elizabeth Cook contributed equally to this work

[†] Philip Toye and James Prendergast contributed equally to this work

[*] Corresponding authors: James G.D. Prendergast james.prendergast@roslin.ed.ac.uk

Philip Toye p.toye@cgiar.org

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

East Coast fever, a tick-borne cattle disease caused by the *Theileria parva* parasite, is among the biggest natural killers of cattle in East Africa, leading to over 1 million deaths annually. Here we report on the genetic analysis of a cohort of Boran cattle demonstrating heritable tolerance to infection by *T. parva* (h^2 = 0.65, s.e. 0.57). Through a linkage analysis we identify a 6 Mb genomic region on *Bos taurus* chromosome 15 that is significantly associated with survival outcome following *T. parva* exposure. Testing this locus in an independent cohort of animals replicates this association with survival following *T. parva* infection. A stop gained polymorphism in this region was found to be highly associated with survival across both related and unrelated animals, with only one of the 20 homozygote carriers (T/T) of this change succumbing to the disease in contrast to 44 out of 97 animals homozygote for the reference allele (C/C). Consequently, we present a genetic locus linked to tolerance of one of Africa's most important cattle diseases, raising the promise of marker-assisted selection for cattle that are less susceptible to infection by *T. parva*.

Introduction

East Coast fever (ECF), a tick-borne lymphoproliferative disease caused by the protozoan parasite *Theileria parva*, is among the biggest natural causes of death in cattle across eastern, central and southern Africa, and is estimated to be responsible for at least 1.1 million deaths per year (1). Direct economic losses attributed to deaths from ECF have been estimated at ~ US$600 million annually (2, 3), and disproportionately affect small-scale low-income households (4). The potential of *T. parva* to extend its range beyond current areas is likely to increase in the next 100 years due to climate change, as conditions become more favourable for its vector, the brown ear tick *Rhipicephalus appendiculatus (5)*. Cattle breeds not native to East Africa are highly susceptible to ECF, and mortality rates can exceed 95% (6). The disease is consequently a major barrier to the wider introduction of productive European cattle breeds into large parts of Africa. However, as ECF has exerted an unusually strong selective pressure on East African cattle since their introduction to the region approximately 5000 years ago (7), various local cattle breeds have developed high levels of heritable tolerance to the disease (8, 9).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As early as the 1950s, the innate genetic tolerance of *Bos indicus* zebu cattle from ECF-endemic areas of Kenya had been experimentally characterised. In 1953, Barnett (9) transported zebu from ECF-endemic and non-endemic areas of Kenya to a farm free of ECF. Calves subsequently born to these animals were then each exposed to a single *T. parva* infected tick. The offspring of zebu from endemic and non-endemic areas of Kenya showed substantial differences in their ability to survive this infection. Colostral immunity was not the cause, with subsequent studies also suggesting that colostral immunity plays little to no role in inherited tolerance to ECF (10). More recently in 2005, offspring of East African Shorthorn Zebu (EASZ) originating from endemic areas of Kenya, that were born in a non-endemic area and therefore previously uninfected, showed a significantly higher tolerance to challenge with a *T. parva* stabilate than not only Friesian and Boran cattle, but also EASZ originating from non-endemic areas (8). Whereas all of the EASZ from the ECF-endemic area of Kenya survived infection, only 60-70% of animals from nearby ECF-non-endemic areas survived the same challenge. These studies highlight that among cattle breeds there is a marked difference in tolerance to infection with *T. parva*, suggesting the frequency of natural tolerance alleles differs between cattle derived from ECF-endemic and ECF-non-endemic areas.

Despite this clear evidence for genetic tolerance among certain local breeds, there has been little work to map the genetic loci driving such tolerance. In this study, we analysed an extended Boran pedigree, the members of which display markedly higher tolerance to *T. parva* infection than other cattle of the same breed. Through a linkage study, we identify a genomic region significantly associated with this increased tolerance.

Results

Tolerance to infection by Theileria parva in a Bos indicus pedigree

The initial indication of tolerance among related Boran animals was detected in a vaccine trial undertaken in 2013 (11), in which four out of six animals that did not succumb to infection by *T. parva* were observed to be first generation descendants of the same Boran bull (3167). This included all of the three unvaccinated

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

survivors in this study. This trial was followed by two further natural field challenge trials in 2014 and 2015 using unvaccinated 1st generation progeny of the same sire alongside unrelated animals, and two subsequent natural field challenge trials in 2017 and 2018 examining 2nd generation offspring (the progeny of male first generation descendants of sire 3167). The survival data for each trial are shown in Table 1. In summary, 67.9% of the first generation offspring of 3167 and 51.1% of the second generation animals survived the field challenge, compared to 8.7% of the unrelated cattle of the same breed across all field trials. As shown in Figure 1, body temperatures typically peaked around day 15 following exposure, consistent with the animals showing signs of infection and clinical symptoms (see methods for further details). Across the five field trials, the offspring of sire 3167 were observed to be 3.2 times more likely to survive the exposure to *T. parva* than unrelated controls obtained from the same farm. We calculated the heritability (h^2) of the tolerance phenotype among these animals by fitting survival status to sex and trial as fixed effects with sire as a random effect. This model returned an estimated h^2 of 0.65 (s.e. = 0.57).

Table 1. Summary of survival outcomes for control and pedigree animals over the course of five field trials

	Controls				**Sire 3167 offspring**			
Trial*	Died	Euthanised	Treated	Survived	Died	Euthanised	Treated	Survived
2013	8	0	1	0	0	0	0	3
2014	9	4	0	1	3	0	3	4
2015	8	0	2	0	2	0	1	12
2017†	2	1	3	1	4	0	3	17
2018†	4	0	0	2	12	2	2	7
Total	31 (67.4%)	5 (10.9%)	6 (13%)	4 (8.7%)	21 (28%)	2 (2.7%)	9 (12%)	43 (57.3%)

* The first three field trials involved 1st generation offspring of sire 3167, whilst the field trials in 2017 and 2018 (indicated by a †) involved 2nd generation offspring.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

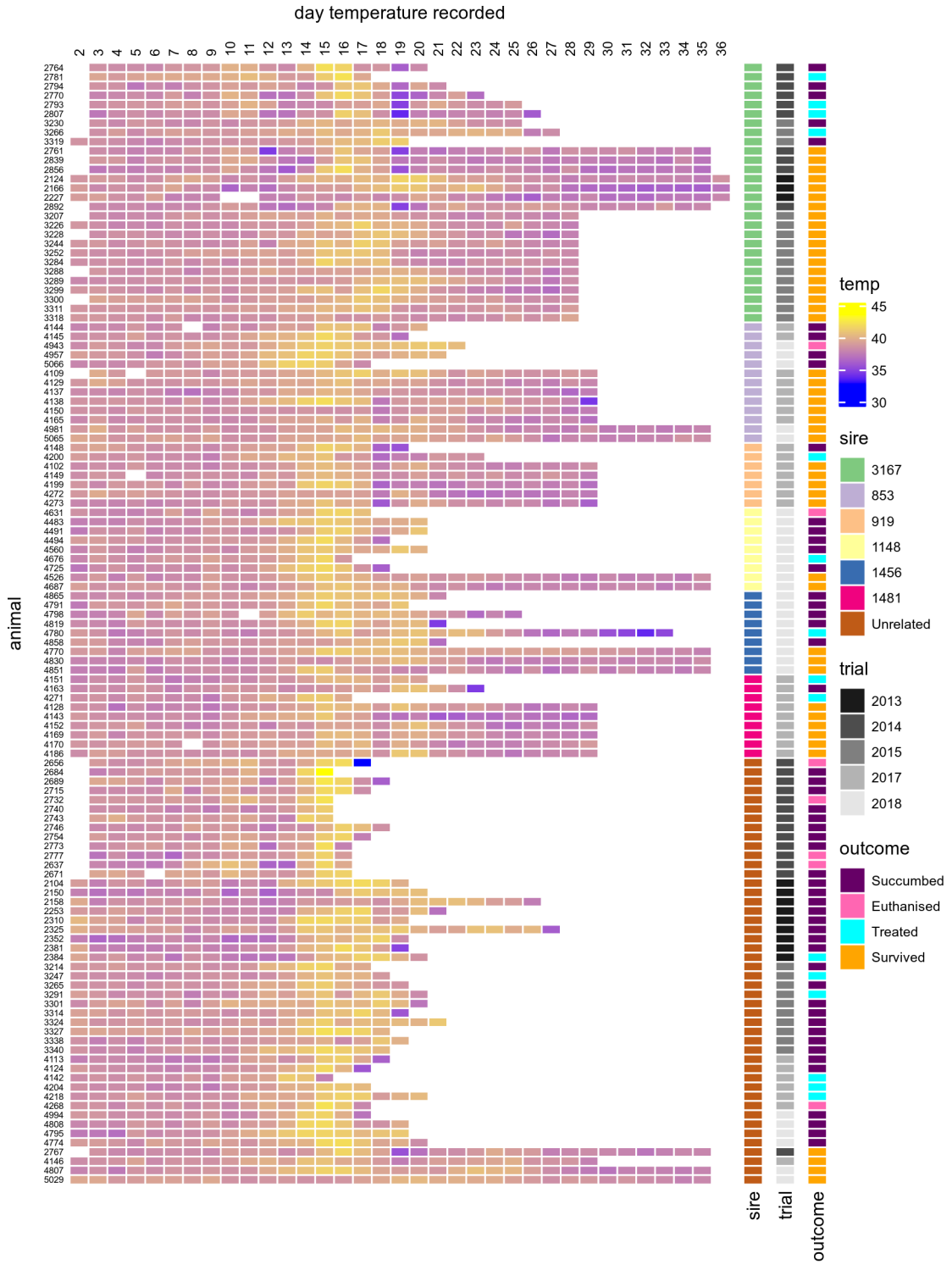

Figure 1. Heatmap of daily body temperature (℃) recordings throughout the field trials. Animals 853, 919, 1148, 1456 and 1481 were all male progeny of sire 3167. Temperature observations commenced on day 2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Survival outcome is significantly associated with a region on chromosome 15

We sought to identify loci underlying the heritable tolerance to *T. parva* within this pedigree. The founder of the pedigree, sire 3167, was no longer available, and so his genotype probabilities (GP) were calculated from the genotype frequencies of his offspring (see methods). Briefly, we generated whole-genome sequence (WGS) data for 43 animals which included 28 1st generation animals, four animals subsequently identified as likely 3rd degree relatives, and 11 unrelated animals. We genotyped a further 78 animals using the Illumina BovineHD array, bringing the total number of animals to 121 comprising 28 1st generation, 47 2nd generation, and 46 unrelated animals. After merging the datasets, filtering and correcting allele mismatches between the WGS and BovineHD data, simulations were performed to calculate the GP of sire 3167 from the genotypes of all 1st generation offspring. In total 465,938 SNPs were analysed and calculated to have a GP > 0.98, of which 97% (n = 452168) had a GP > 0.9999, i.e. could be inferred with very high confidence. These GP were used to impute genotypes for sire 3167, which were then merged with the genotype data of the pedigree and unrelated animals and collectively phased while accounting for the known pedigree information.

The assumption in this study is that sire 3167 is carrying at least one genetic locus that confers tolerance to *T. parva* infection. The pedigree could therefore inherit a tolerance haplotype from this founding bull, but we did not want to exclude the possibility of it also sometimes being inherited down the maternal line. To do this, haplotype analyses were conducted by first partitioning the phased genotype data into 1 Mb windows, for which the Hamming distance of each animal's paternal and maternal haplotypes to each of those of sire 3167 was calculated, resulting in four distance values per window for each animal. We fitted a regression model with binomial survival outcome as a response to the four haplotype distances alongside sex, trial ID and sire to account for these potential confounders. This allowed us to determine whether animals carrying particular haplotypes derived that from, or that were highly similar to those of, sire 3167 were more likely to survive exposure to *T. parva* infection.

We identified a genome-wide significant association ($p = 4.12 \times 10^{-07}$, adjusted for multiple testing via permutations p = 0.027, see methods) on chromosome 15 (15:49-50 Mb; Figure 2), followed by two neighbouring haplotypes (15:46-47 Mb unadjusted $p = 9.1 \times 10^{-6}$; 15:47-48 Mb unadjusted $p = 1.43 \times 10^{-5}$) which although also nominally significant - were not significant genome-wide after correction (15:46-47 Mb

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

corrected p = 0.23; 15:47-48 Mb corrected p = 0.32). The paternally-derived 1 Mb haplotype windows of surviving animals in this region (15:46-50 Mb) were more similar to both the A (median Hamming distance *DA* = 0.069 ± 0.165) and B (*DB* = 0 ± 0.160) haplotypes of sire 3167 than those of the animals that succumbed to infection (*DA* = 0.254 ± 0.250; *DB* = 0.278 ± 0.258).

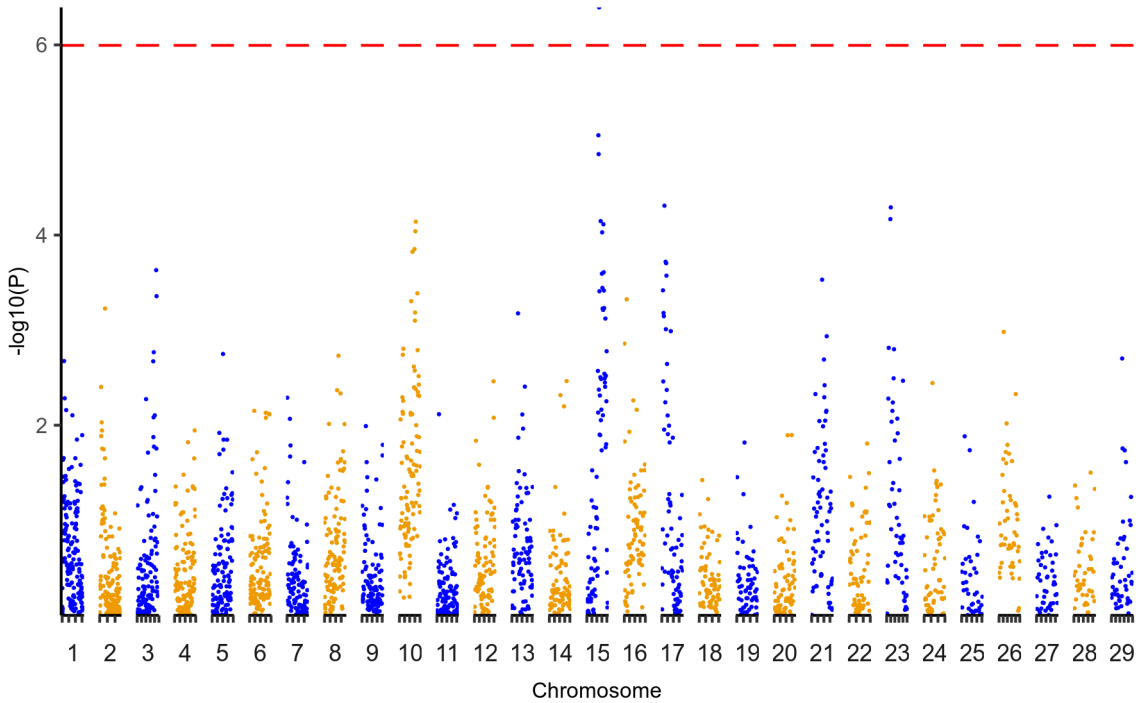

Figure 2. Regression analyses on haplotype distances. Results of regression analyses of haplotypes versus survival status. A single window crossed the corrected significance threshold determined from 1000 permutations, indicated by the dashed red line at p < 0.05.

Characterisation of variants within the tolerance locus

Having identified a genomic region significantly associated with survival outcome we next sought to identify potential candidate functional variants within the locus for further investigation. To do this we characterised both the impact of variants on coding regions as well as any potential link to the expression of nearby genes. To define potential regulatory variants (eVariants) we sequenced mRNA extracted from white blood cells collected from 29 animals (n = 23 pedigree, n = 6 unrelated) involved in the 2018 field trial at day 0 (i.e.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

before their transport to Ol Pejeta Conservancy), day 7 and day 15. Within the target region, 720 variants showed a nominally significant association between their genotype and the expression level of a nearby gene on at least one of the timepoints (Supplementary Table 4), and a further 587 variants were tagged as potential response eQTLs, where the association between genotype and gene expression differed between timepoints (Supplementary Table 5). However, none of these associations were significant after applying a false discovery rate correction to account for multiple testing.

We next characterised the impact of the variants in this region on predicted coding regions. The allele frequency and functional consequence of all SNPs spanning 1 Mb upstream and downstream of the identified wider region (15:45-51 Mb) were characterised from the whole-genome sequence (WGS) data. After filtering to retain high-quality variants, we identified 58066 variants and 235 genes in this region. These included 57119 single nucleotide variants, 500 indels (an insertion or deletion, affecting two or more nucleotides), and 447 sequence alterations (sequence ontology SO:0001059). A summary of all predicted variant consequences is provided in the supplementary material (Supplementary Data 1). From these, we selected 55 variants predicted to have a functional consequence with a high impact. As any variant conferring tolerance to infection is unlikely to be present in non-native breeds, as evidenced by their extremely high mortality when exposed to *T. parva*, we removed variants with an alternate allele frequency (aAF) > 0.1 in European taurine breeds, leaving 31 variants. Similarly, we expected to observe the variant at a high frequency in the 1st generation offspring of sire 3167 that survived infection, and so removed variants with aAF < 0.5 in these animals, resulting in 4 variants (Figure 3A): 1 frameshift variant, 2 stop gained and 1 start lost, each affecting a different gene. Each of these variants had an alternate allele frequency < 0.04 in European taurine, < 0.091 in African taurine breeds, but ranged from 0.22 to 0.54 in African and Asian indicine breeds, and from 0.53 to 0.78 in the pedigree of 3167 (Figure 3B).

Of the genes associated with these variants, two are olfactory receptors (OR51H1 and OR51G1) and two are novel genes - the first is a paralogue of haemaglobin subunit beta (HBB) and the second a paralogue of FAS-associated factor 1 (FAF1). It should be noted we could not find any evidence of any of these genes being expressed in white blood cell RNA-seq data, suggesting this is not their tissue of action. Olfactory receptors (ORs) have been found to be involved in immune responses associated with the olfactory bulb, and there is

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

increasing evidence of OR expression in non-olfactory tissues, however there remains a significant gap in our knowledge concerning the functional importance of ectopic ORs (12). HBB is associated with erythrocytes and variants in HBB have previously been linked to protection to human malaria, caused by another pathogenic Apicomplexan protozoon, *Plasmodium falciparum*. However, as infection of lymphocytes is responsible for the pathology of disease caused by *T. parva* and infected lymphocytes are also the target of the protective acquired immune response, with infection of erythrocytes having little or no role in either of these, the role of HBB tolerance to *T. parva* infection is not immediately obvious. In contrast, FAS is a member of the TNF-receptor superfamily involved in apoptosis, and *T. parva*-infected cells have previously been shown to be resistant to FAS-induced cell death (13). Thus, we focused on further examining the association of the variant in the FAF1 paralog with tolerance to *T. parva* infection.

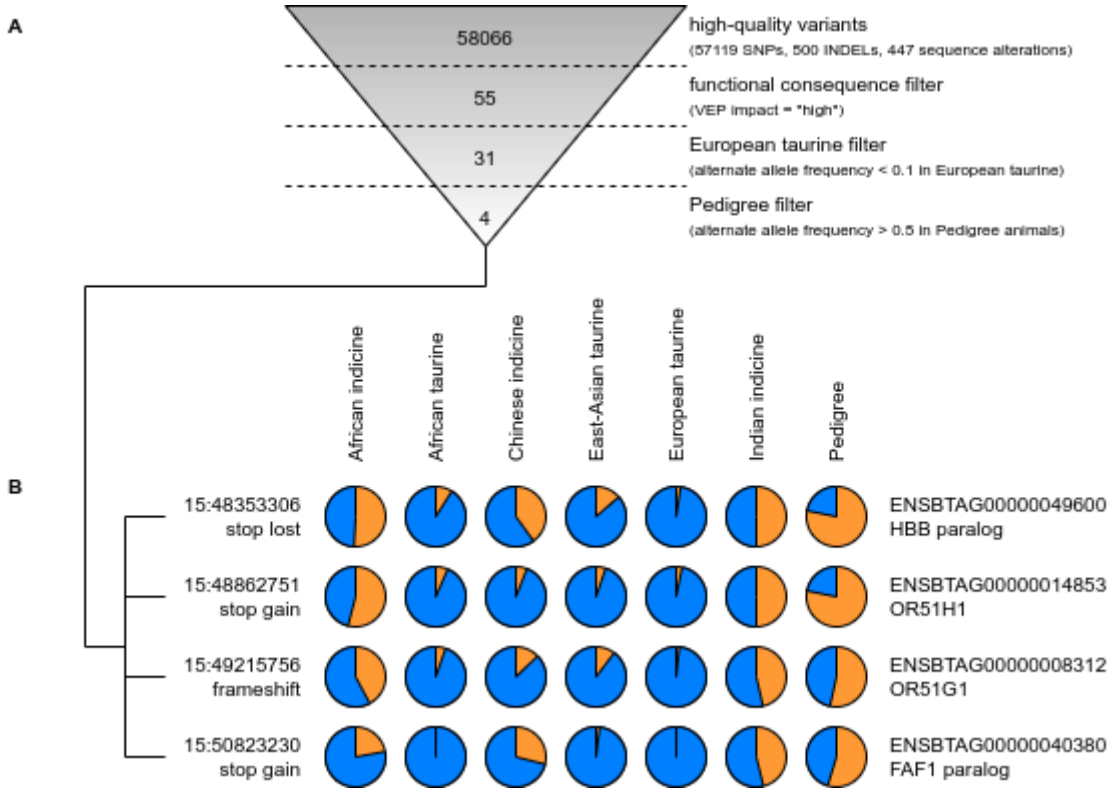

Figure 3. Prioritisation of variants. (A) Successive filters were applied to the high-quality variants in the region identified as significantly associated with tolerance to *T. parva* infection (15:45-51 Mb). This reduced the number of variants from 93960 to four predicted to have a high functional consequence, a low alternate

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

allele frequency (aAF < 0.1) in European taurine animals, and a high alternative allele frequency in Pedigree animals (aAF > 0.5). (B) The allele frequencies (reference allele in blue, alternative allele in orange) of these four variants were subsequently calculated in different populations, indicating the alternative allele is potentially of an indicine origin.

Genotyping of the FAF1 paralog confirms a significant association with survival outcome

The FAF1 paralog shares 96.75% identity with 68.77% of FAF1's nucleotide sequence. The stop gain variant (15:50823230 on ARS-UCD1.2) in the FAF1 paralog (referred to here as FAF1B) converts an arginine codon (cga) to a stop codon (tga) which would result in a truncated protein. Within FAF1 itself there is no variant recorded in dbSNP at the homologous base position (3:95786449) to the FAF1B variant (15:50823230). The FAF1B variant is not present on the BovineHD array used to genotype the majority of the animals. To establish if there is a significant association between genotype and survival outcome across the wider set of animals, we designed primers to amplify and sequence this variant in 57 trial animals that were not previously whole-genome sequenced, bringing our genotype data for FAF1B to 100 animals. These animals include 77 pedigree (43 survived, 34 succumbed to infection) and 23 unrelated individuals (4 survived, 19 succumbed to infection). In all, 88% of the animals homozygous for the variant allele (T/T) survived the field challenge (7 out of 8, the final animal having been treated rather than having died from the infection), compared to 62% (36 out of 58) of heterozygotes (C/T) and only 12% (4 out of 34) of those homozygous for the reference allele (C/C; Table 2). Analysis by logistic regression of survival status versus genotype at this variant was significant across all animals (likelihood ratio test $p = 2.48 \times 10^{-04}$), within unrelated animals (likelihood ratio test $p = 5.38 \times 10^{-04}$), and within the pedigree (likelihood ratio test $p = 6.34 \times 10^{-04}$), when controlling for sex, trial, sire (where pedigree animals involved), including 3rd degree relatives in the pedigree, and classifying animals that received treatment or were euthanised as having died.

Considering explicitly the unrelated and pedigree (1st and 2nd generation) animals that survived or succumbed to infection, a comparison of survival curves fitting survival time versus genotype at the FAF1B variant while accounting for degree of relatedness returns a significant association (log-rank $p = 4.13 \times 10^{-13}$; Figure

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4A). Further to its association with survival, among these animals that succumbed to the disease, those heterozygous at the variant succumbed significantly (Wilcoxon $p = 7.3 \times 10^{-04}$) later on average than homozygous reference animals (mean of 21.27 ± 2.25 days versus 18.55 ± 1.7 days; Figure 4B). An animal's genotype at this variant was also observed to be associated with their temperature over the course of the infection (Figure 4C), with T/T animals generally showing a lower increase in temperature, and in particular at days 15 and 16 (ANOVA F test P of 2.44×10^{-06} and 4.43×10^{-07} respectively when accounting for relatedness, sex and trial. Figure 4D).

Table 2. The genotype counts of the FAF1B variant

Genotype	Status	Unrelated	3rd degree*	2nd gen.	1st gen.	Total
C/C	Succumbed	13	2	7	0	22 (64.7%)
C/C	Euthanised	2	0	1	0	3 (8.8%)
C/C	Treated	3	0	2	0	5 (14.7%)
C/C	Survived	0	0	4	0	**4 (11.8%)**
C/T	Succumbed	1	0	9	5	15 (25.9%)
C/T	Euthanised	0	0	1	0	1 (1.7%)
C/T	Treated	0	0	3	3	6 (10.3%)
C/T	Survived	3	2	16	15	**36 (62.1%)**
T/T	Succumbed	0	0	0	0	0 (0%)
T/T	Euthanised	0	0	0	0	0 (0%)
T/T	Treated	0	0	0	1	1 (12.5%)
T/T	Survived	1	0	2	4	**7 (87.5%)**

* Upon genotyping, four controls were determined to in fact likely be 3rd degree relatives of sire 3167, these are separated out in the above table. Note there were no homozygote C/C first generation animals as sire 3167 was T/T at this variant.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

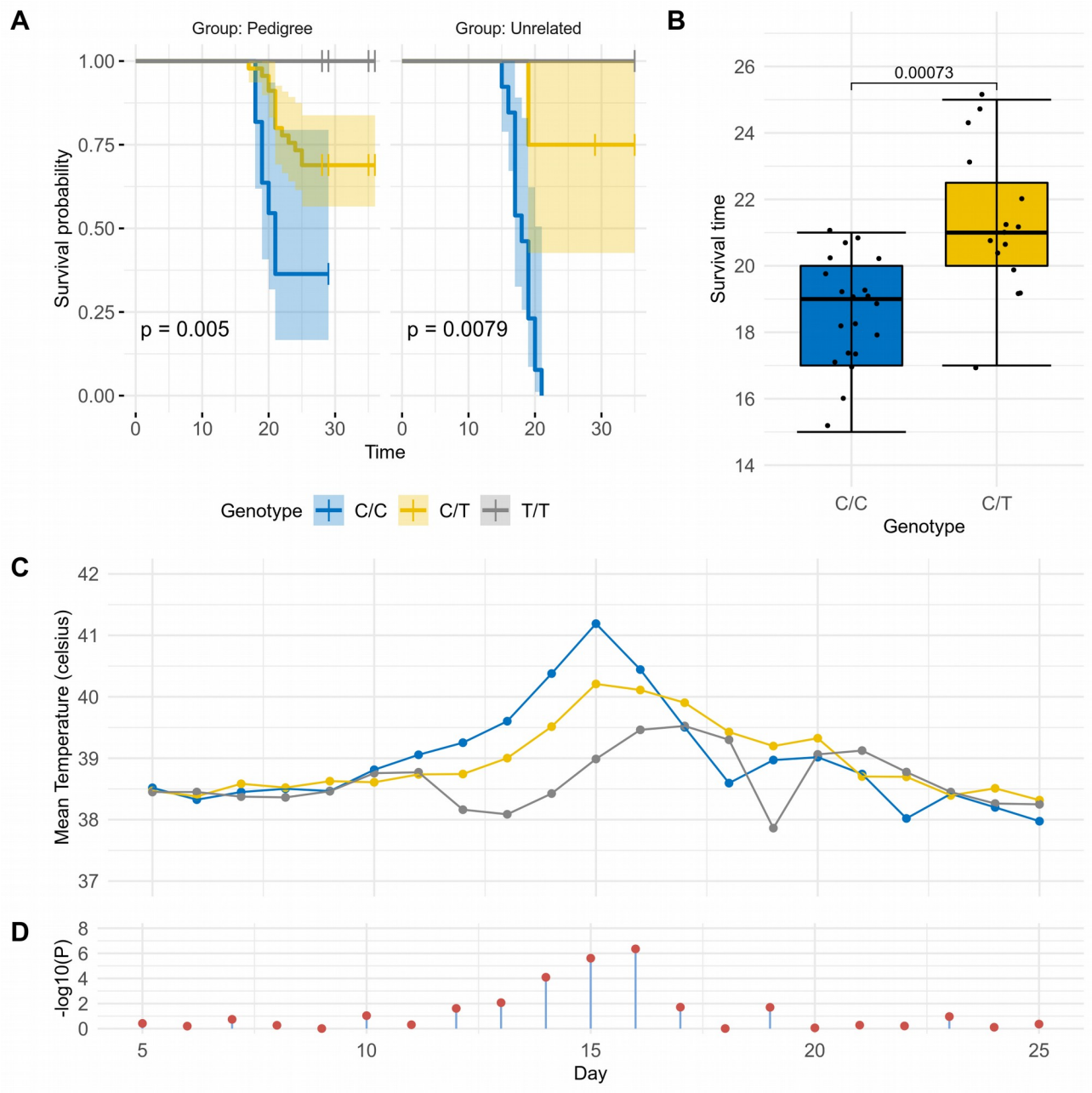

Figure 4. Survival analyses for the FAF1B variant in pedigree (1st and 2nd generation) and unrelated animals. (A) Analysis of survival curves in animals that succumbed or survived *T. parva* infection reveals significant associations between survival probability and genotype. Log-rank p values are reported. Comparing survival curves when fitting survival time against genotype whilst accounting for relatedness (unrelated, 1st generation, 2nd generation) returns log-rank $p = 4.13 \times 10^{-13}$. (B) Comparison of survival time versus genotype for animals that succumbed to infection. The mean survival time for C/C animals was 18.55 ± 1.7 days, while mean survival time for C/T animals was 21.27 ± 2.25 days. A Wilcoxon p value is reported. (C)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mean temperature versus field trial day for all animals regardless of survival outcome and relatedness, grouped by genotype. Temperature peaked at day 15 in C/C and C/T animals, and at day 16 in T/T animals. (D) Fitting T allele count against temperature whilst accounting for relatedness, sex and trial, returns significant associations from days 13 to 16 (ANOVA F-test $p < 0.01$) .

To determine how much variance in heritability is explained by the FAF1B SNP we ran linear mixed effect models using residual maximum likelihood to estimate parameters. Having calculated the heritability (h^2) based on phenotypic data alone as 0.65 (s.e. = 0.57) as stated above, we subsequently accounted for genetic effects by fitting the genotype at the FAF1B SNP as a fixed effect. This model returned a significant additive SNP effect of $\alpha = 0.43 \pm 0.083$ for a single copy of allele T. This is highly significant assuming an additive model $t = 5.2$ (Student's left-tailed t-distribution $p = 2.51 \times 10^{-06}$), with no evidence for a dominant effect (dom = 0.12 ± 0.09) $t = 1.25$ (Student's left-tailed t-distribution $p = 0.18$).

FAF1B genotype-associated tolerance replicated in independent population

We next investigated any evidence for replication of this association in the independent Infectious Disease of East African Livestock (IDEAL) cohort. The IDEAL study was conducted in a *T. parva*-endemic area in western Kenya in 2007 in a smallholder system free of buffalo, and involved the intensive monitoring of 548 East African Shorthorn Zebu (EASZ) calves for the first year of life. The study generated data on the frequency and clinical signs of infections and their impact on health and growth(14), with 32 calves determined, by clinical signs and post mortem observations, to have died of ECF. There were a number of important differences between this study and the field trials reported above. The animals in the IDEAL study were spread across farms and different environmental exposures and, unlike the Boran breed, EASZ are believed to exhibit generally elevated tolerance to ECF - supported by the IDEAL project reporting a 6% mortality due to *T. parva* infection despite 76% of animals becoming infected in their first year of life(15). Furthermore, transmission of *T. parva* in the IDEAL study will have been via tick transmission from cattle to cattle, rather than transmission to cattle from ticks that had previously fed on buffalo as in the field trials. We

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genotyped 130 EASZ at the FAF1B SNP using the same primers as for the Boran. This included the 32 calves that died of ECF, and 98 randomly selected survivors from the IDEAL study. Consistent with the field trial results suggesting the T allele is associated with tolerance, no homozygote T/T animals in the study succumbed to ECF (Table 3). Logistic regression of survival status versus genotype while accounting for sex was significant in this cohort of 130 animals (likelihood ratio test p = 0.022) with the same direction of effect as in the 3167 pedigree.

Table 3. IDEAL cohort genotypes

Genotype	**Succumbed to ECF (n = 32)**	**Survived**[†]	
		Episode (n = 31)	**No-episode (n = 67)**
C/C	16 (50%)	21 (68%)	30 (45%)
C/T	16 (50%)	8 (26%)	27 (40%)
T/T	0 (0%)	2 (6.5%)	10 (14.9%)

[†] Randomly selected animals that survived are further categorised according to whether or not they presented any evidence of a clinical episode associated with ECF, which include: pyrexia, and hypertrophic parotid and prescapular lymph nodes.

FAF1B genotype is associated with live cell count and ex vivo infection level

The expansion of *T. parva*-infected cells has been found to be reduced in cattle surviving *in vitro T. parva* infection compared to animals that succumbed. We therefore sought to determine if there is also a difference in the *in* vitro expansion of *T. parva*-infected cells based on the FAF1B genotype of the animals. Peripheral blood mononuclear cells (PBMCs) were cultured and infected by incubation with salivary glands dissected from *R. appendiculatus* fed on animals infected with *T. parva*. The level of infection in ticks was estimated by counting the number of infected acini in a sample of dissected salivary glands, and the sporozoite suspension adjusted to a concentration equivalent to 2000 infected acini per ml (see 16). Equal volumes of sporozoites and cells (2 x 10^7 PBMCs) were cultured from 12 cattle unrelated to 3167 carrying the C/C

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genotype at the FAF1B SNP, and 12 cattle carrying the T/T genotype, which comprised of six animals from the tolerant pedigree and a further six unrelated animals for comparison. Counting of cells stained with trypan blue every 2 days over a period of 12 days revealed a significant association between live cell count and genotype when fitting day as an interaction term and accounting for relatedness (see methods, F-test p = 1.42×10^{-07}; Figure 5A). To determine if there was a difference in infectivity associated with the FAF1B SNP, cells were stained with a fluorophore directed at the polymorphic immunodominant molecule (PIM) expressed on the *T. parva* schizont's surface. As with the live cell count, a significant association between the proportion of infected (PIM+) cells and genotype was observed when fitting day as an interaction term and accounting for relatedness (F-test p = 7.56×10^{-09}; Figure 5B).

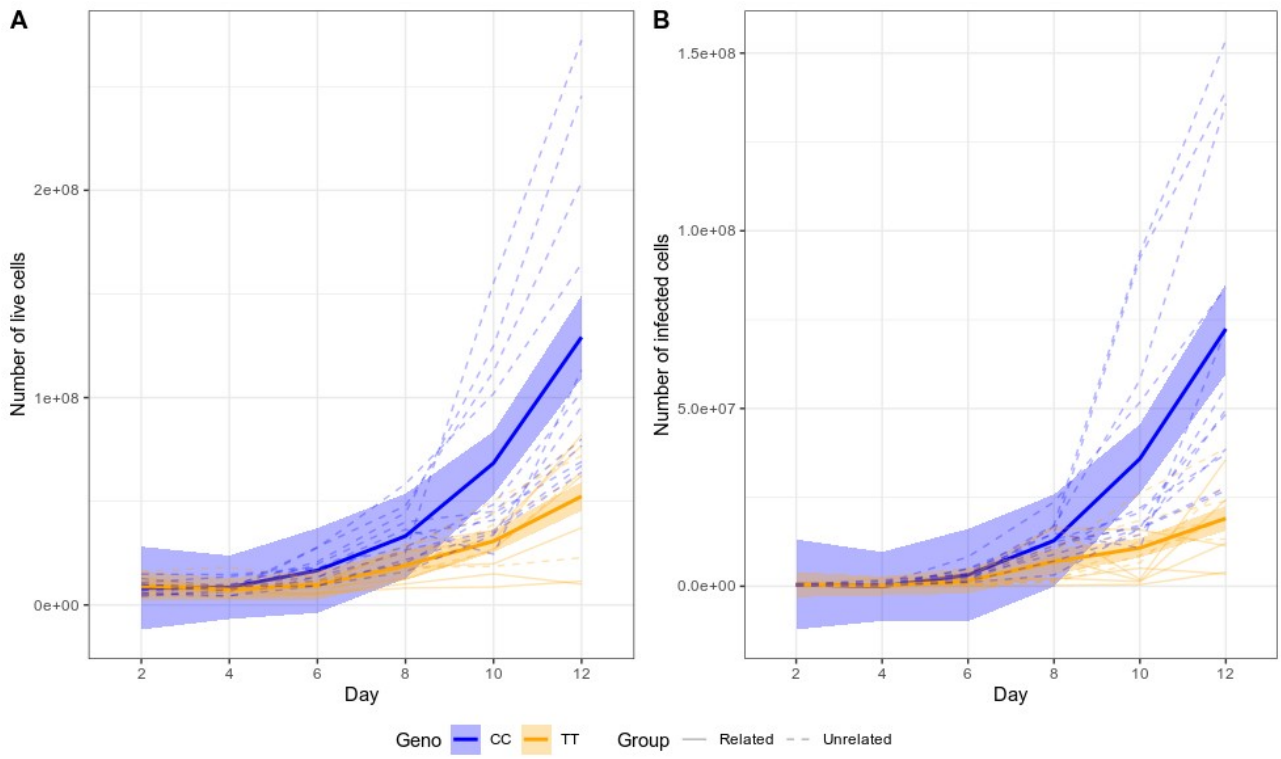

Figure 5. Cell expansion and infectivity levels in cultured cells infected with *T. parva.* (A) Animals possessing a C/C genotype at the FAF1B SNP exhibit higher live counts throughout the course of infection compared to T/T animals. The association between live cell count and genotype when fitting day as an interaction term and accounting for relatedness is significant (F-test p = 1.42×10^{-07}). (B) In addition, animals possessing a C/C genotype exhibit a higher proportion of infected (PIM+) live cells, which is also

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

significant (F-test p = 7.56×10^{-09}). Thin dashed lines indicated unrelated animals, solid, thin lines indicate animals from the ECF-tolerant pedigree. The blue and orange ribbons and thick blue and orange lines indicate the 95% confidence interval around the mean for C/C and T/T genotypes, respectively.

Consequently, these data suggest the association of the FAF1B locus with survival in the 3167 and IDEAL cohorts may also be reflected in different capacities of *T. parva*-infected cells to establish and expand *in vitro*.

Discussion

The progeny of sire 3167 demonstrate a clear, heritable tolerance to buffalo-derived *T. parva* infection, which haplotype-based analyses reveal to be significantly associated with an extended 6 Mb genomic region on chromosome 15. Notably, this same region was one of a small number showing evidence of between-breed selective sweeps in both cattle and water buffalo in a recent study (17), and both species are susceptible to a closely related *Theileria* species, *T. annulata*. This, along with the elevated non-reference allele frequency among indicine cattle breeds at variants in this region, raises the possibility that the tolerance phenotype first arose as a response to *T. annulata*, which is transmitted by a different tick and is present in a region extending from north Africa and southern Europe into the Middle East and Asia. This was then introduced into Africa in one of the waves of introduction of *Bos indicus* breeds. Within this genomic region we identified a putative stop-gained polymorphism within a paralogue of the FAF1 gene, FAF1B, of which the genotype is significantly associated with survival outcome. As FAS-induced apoptosis has previously been shown to be targeted by the parasite to facilitate the transformation of T cells (13), this gene is a plausible candidate for a central role in mediating tolerance. We also find the FAF1B SNP to be associated with cell expansion and infection levels *in vitro*. A key hallmark of *T. parva* infection is it leads to the uncontrolled proliferation of T-cells and, although further work is required, these *in vitro* findings potentially support the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

idea that the tolerance phenotype is linked to better management of the uncontrolled proliferation of infected cells.

It should be noted we could find no evidence that FAF1B is expressed in white blood cells (WBCs) before or after infection, though its action could be in other tissues. Although we found little evidence for any eQTLs in this region potentially driving the phenotype, this analysis had limitations in that the sample size was small, blood is potentially confounded by cell-type composition differences, and, although a reasonable candidate, blood may not be the tissue of action. Consequently, further work is required to elucidate any causal relationship between this FAF1 paralogue and the enhanced tolerance to *T. parva* linked to this locus. Irrespective of whether it is driving tolerance though, we have demonstrated the putative stop gained variant in this gene is strongly linked to the tolerance phenotype and importantly provides a potential effective genetic marker for breeding for tolerance. Notably, of the combined total of 20 animals that were homozygote for the alternate allele at this locus across the Boran pedigree, controls and EASZ, none succumbed to *T. parva* infection (though one was treated). This is in contrast to the 53% of animals which were homozygote for the reference allele (n = 44) that succumbed (n=83). Although a number of these animals were related, the lack of deaths among homozygote carriers adds support to the promise of potential breeding for tolerance using this allele. It is also worth noting that of the four identified 3rd degree relatives, only the two that possessed a T allele survived infection. Furthermore, those heterozygote carriers that succumbed to the disease were observed to survive several days longer on average than animals carrying no copies of the variant, and to display some evidence of potentially showing fewer clinical episodes in the IDEAL cohort. *T. parva* can be transmitted via ticks that have previously fed on either infected cattle or infected African buffalo. The challenge in the field trials was buffalo-derived, as cattle had been absent from the trial site prior to the introduction of the trial animals. In contrast the challenge in the IDEAL study was cattle-derived. Consequently, the evidence points towards this locus being potentially protective against disease caused by both forms of the parasite.

We have therefore validated a genetic marker which is significantly associated with survival outcome. This marker can consequently be applied to support selective breeding programmes with a view to improving the tolerance of cattle populations, in order to reduce the substantial impact of *T. parva* on sub-Sahara African

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

countries. Confirmation of the functional variant in this region will facilitate the application of modern gene editing technologies to formally demonstrate the functional role of the variant and ultimately to rapidly increase its frequency in imported, productive cattle throughout Africa.

Methods

The study protocols were approved by ILRI's Institutional Animal Care and Use Committee (References 2013–03, 2014 -32, 2015-29, 2017-02, 2018-10).

Field trials

Field trials involved transporting animals from the Kapiti Research Station in Machakos county, a region largely free of ECF, to the Ol Pejeta Conservancy (Kenya), a region endemic for *T. parva*. The study site where animals remained for up to 35 days, was free of other cattle but populated by African buffalo (11). The 2013, 2014 and 2015 field trials involved 1st generation progeny of sire 3167 while the 2017 and 2018 field trials involved progeny of males sired by 3167; in all trials unrelated animals were also included, and researchers conducting the phenotyping were unaware as to which animals were pedigree or unrelated. Although pedigree records were maintained throughout, for the 2017 and 2018 field trials animals were genotyped in advance to verify their relatedness. Genetic analyses (described below) identified 21 individuals that were incorrectly documented as belonging to the pedigree of sire 3167; eight pedigree individuals whose sire was incorrectly recorded, but which remained within the pedigree; and seven individuals thought to be unrelated but which were found to be 3rd degree relatives from four sibships. Animals in all field trials were screened in advance to ensure they tested negative by ELISA (18) for antibodies against *T. parva* and, for the last four trials, *T. mutans*, and in addition in the 2013 (11), 2017 and 2018 field trials by p104 PCR (19) to test for active *T. parva* infection. During the field trials body temperature recordings were taken on a daily basis. On a small number of occasions a recording could not be taken due to the temperament of the animal. The first four days of temperature recordings were excluded

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

from downstream analyses as these typically exhibited unusual deviations due to the stress of transport. A summary of phenotypic data for field trial animals analysed in this study is provided in Supplementary Table 1, along with temperature observations in Supplementary Table 2.

The most likely cause of death for the first study was reported to be buffalo-derived *T. parva* infection, based on clinical signs and post mortem observations (11). Similar observations were made in the subsequent studies. All animals exhibited pyrexia except one (4150; Fig. 1), and swollen parotid and prescapular lymph nodes, whilst laboured breathing, nasal discharge and corneal opacity were commonly observed. Microscopic examination of aspirates from the parotid or prescapular lymph nodes revealed *T. parva* macroschizonts in at least one lymph node of every animal. Routine post mortem observations included frothy exudate in the trachea and pulmonary oedema. Seroconversion to anti-*T. parva* antibodies was detected in 46 of the 47 surviving animals and in 42 of the 74 animals that succumbed.

Sampling

Blood was collected from the animals in EDTA tubes and extracted using DNEasy® Blood and Tissue Kits (Qiagen) according to manufacturer's instructions. For genotyping on the Illumina BovineHD BeadChip, DNA integrity was first assessed by electrophoresis on a 1.5% agarose gel, and quantitated by NanoDropTM spectrophotometer for dilution to a concentration of 50 ng/µl in 40µl of elution buffer. DNA samples were shipped to Edinburgh Genomics where quality was assessed by PicoGreenTM assay prior to genotyping. For whole-genome sequencing DNA was extracted using the MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche Diagnostics GmbH).

Survival analysis

Survival analyses were performed and plotted in R (20) (v3.4.4) using the survminer (v0.4.6) package. Briefly, a survival object was constructed excluding animals that were treated or euthanised fitting survival time, defined in days as either the end of the field trial or the day on which the animal died, against genotype whilst accounting for degree of relatedness (unrelated, 1st or 2nd generation). P values were calculated from the fitted survival curves using the log-rank method.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Processing of whole-genome sequence data

Sequencing was performed on the Illumina HiSeq X platform using 150bp pair end reads with 550bp insert, TruSeq PCR free libraries, to a target depth of 30X. Sequence reads were aligned to the ARS-UCD1.2 assembly using BWA-MEM (21) (v0.7.17), and processed using GATK (22, 23) (v4.0.8.0) applying MappingQualityNotZeroReadFilter, MarkDuplicates, and base quality score recalibration (BQSR) using known sites from the 1000bulls genome project (http://www.1000bullgenomes.com/). Variants were called using HaplotypeCaller (24, 25) to generate gVCF files, from which joint variant calling was performed across samples, followed by variant quality score recalibration (VQSR).

Processing of BovineHD Genotyping data

Genomic coordinates for the Illumina BovineHD Genotyping BeadChip were updated from the UDM3.1 reference assembly to the newer ARS-UCD1.2 assembly using data provided by Robert Schnabel from the University of Missouri (UMC_marker_names_180910.zip, available at https://www.animalgenome.org/repository/cattle/UMC_bovine_coordinates/). The array data was converted from PED to VCF format using Plink (26) (v1.90blg). A Python script was written to cross-reference the alleles for SNPs from the genotype data against those from the sequence data, to correct for discrepancies in reference allele and/or strand between the different datasets. SNPs with alleles whose concordance could not be validated (A/T, T/A, C/G, G/C) were discarded. The VCFs from genotyping and sequencing were subsequently merged using bcftools (27) (v1.2), and filtered to remove SNPs with a minor allele frequency (MAF) < 0.05 and call rate < 0.95.

Analysis of relatedness

To cross-validate the written breeding records and ensure that individuals were not assigned to the pedigree of sire 3167, incorrectly, the genotype data were analysed using VCFtools (28) (v0.1.13) to calculate relatedness as per Manichaikul et al. (29). Relatedness values range from $\Phi < 0.0442$ for unrelated individuals, $0.0442 < \Phi < 0.0884$ for 3rd degree relatives, $0.0884 < \Phi < 0.177$ for 2nd degree relatives, and $\Phi > 0.1777$ for 1st degree relatives.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Imputation of founding sire 3167

Genotypes from the 1st generation offspring were phased with SHAPEIT2 (30) (v2r837) using a genetic map estimated from physical SNP positions assuming 1 cM/Mb, employing a window size (-W) of 5 Mb, and the duoHMM (31) (v0.1.7) algorithm to take into account the available pedigree information. A Python script was written to estimate sire 3167's genotype probabilities from the 1st generation genotypes. For each SNP, 1 million simulations were performed to generate genotypes for 34 individuals (the number of 1st generation offspring genotyped). During each iteration, for each individual a 'paternal' allele was randomly chosen from each of sire 3167's possible genotypes {0/0, 0/1, 1/1}, and combined with a random 'maternal' allele {0, 1}. Thus, for each iteration we recorded the counts of offspring for each combination of alleles {0/0, 0/1, 1/1} for each of sire 3167's possible genotypes. The simulations incorporated a 5% error rate which would cause the paternal allele to be switched. From these simulations, the offspring genotype counts that matched those of the 1st generation data were retrieved for each of sire 3167's possible genotypes, and the frequency of each recorded as sire 3167's genotype probabilities. Sire 3167's genotypes were then imputed from these probabilities and the 1st generation phased genotype data in 5 Mb windows using IMPUTE2 (32) (v2.3.2). The imputed genotypes for sire 3167 and the 1st generation genotype data were then combined with the 2nd generation genotype data, and collectively phased using SHAPEIT2 and duoHMM.

Haplotype association analyses

Assuming sire 3167 carries one or more haplotypes associated with tolerance to *T. parva*, we partitioned the genotype data into 1 Mb 'haplotype blocks' and calculated the Hamming (33) distance for each progeny's paternal (A) and maternal (B) haplotypes (h) to those imputed for sire 3167 (r). This resulted in four haplotype distance metrics per individual $\{A_h,A_r; A_h,B_r; B_h,A_r; B_h,B_r\}$ which were divided by the number of SNPs in the haplotype to correct for the varying SNP density of haplotypes. The analysis was performed in R using the proxy (v0.4-22) package. Multiple regression was performed in R to test binomial survival status against the four haplotype distances with sex, field trial and sire fitted as covariates. This model was then compared to a reduced model without the four haplotype distances included using ANOVA. Individuals that received veterinary intervention due to their severe symptoms, or were euthanised, were treated as non-survivors for the purpose of this binary phenotype. To determine significance thresholds that account for

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

multiple testing 1000 permutations were performed where the set of four distances were permuted between individuals. For all blocks on the same chromosome the distances were swapped between the same sets of animals to maintain their relationship across regions. Thus, the phenotypic metadata remained associated with the correct ID and any linkage between haplotypes along a chromosome remained intact. The minimum p value observed across all blocks in the genome for each permutation was recorded, from which a 0.05 significant threshold was determined as the value where just 5% of these values were smaller (Supplementary Table 3).

Expression quantitative trait loci (eQTL) analyses

RNA was extracted from white blood cells of animals sampled during the 2018 trial prior to being transported to the field site (day 0) and on days 7 and 15 of the field trial. These animals included 15 that succumbed to infection, 9 that survived, 2 that were euthanised and 2 that were treated. Following phenol chloroform extraction, mRNA was sequenced on the Illumina HiSeq platform to generate ~70M x 50 bp reads per sample. RNA sequencing reads were aligned to ARS-UCD1.2 using STAR (34) (v2.7.1a;). Stranded fragments per kilobase of exon model per million reads mapped (FPKM) values were calculated for exon features from the alignments using Htseq-count (35) (v0.11.2;). FPKM values were tabulated and filtered in R for each sampling day to retain only genes where at least 50% of samples had FPKM ≥ 3. Response expression quantitative trait loci (reQTL) analyses were performed in R as follows. Genes within the region 15:44-52 Mb were identified, and for each day the FPKM of each gene for each individual was retained, along with the allele dosages for *cis* variants - these included any bi-allelic SNP with MAF > 0.1 within 1 Mb upstream of the gene's start position to 1 Mb downstream of its end position. SNP allele dosages were derived from the genotype data generated using the BovineHD array described above. Animals were assigned a group value of 'pedigree' or 'control', and for each day we regressed a gene's FPKM values against the allele dosages of each of its *cis* variants independently, while accounting for an individual's sex and group assignment. To identify reQTL we used a beta-comparison approach. Here we performed pairwise comparisons of regression slopes for an eQTL at the different time points in a z-test:

$$Z = \frac{\beta_x - \beta_y}{\sqrt{\delta_x^2 + \delta_y^2}}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To test which genes exhibited significantly different expression between survivors and those that succumbed to infection by *T. parva*, for each gene within the region we ran logistic regression fitting binomial survival status against FPKM adding sex and group as covariates. This was compared to a reduced model without FPKM, by ANOVA and the likelihood ratio test.

Prioritisation of candidate variants

Publicly available WGS data for 421 cattle (17) were processed as described above, combined with the WGS data for 43 cattle generated for this study, and variants jointly called. Additional filtering beyond that described above included extracting the target interval 15:45095457-51095457, the removal of singletons and of variants with a genotype quality (GQ) < 30, a minor allele frequency (MAF) < 0.01 or missingness ≥ 0.05. This resulted in the removal of 11 cattle and 1 variant due to missingness, and 277 variants due to low MAF. After filtering, stratified allele frequencies were calculated using Plink for the remaining 93682 variants across 453 cattle that represented different backgrounds (Boran pedigree, African taurine, African indicine, Chinese indicine, East-Asian indicine, European taurine, Indian indicine, and cattle from the Middle East). The functional consequence of variants was determined using Ensembl's Variant Effect Predictor (36) (VEP).

Primer design and PCR amplification

Primers to genotype the candidate variant were designed using Primer3Plus (37) (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi) and checked for specificity by a BLAT search against the reference genome (ARS-UCD1.2). The primers Pair1_L (5'GCTTGGGATCTGACAAAGGA3') and Pair1_R(5'TGGCCTCACGTTCTTCTTCT3'), synthesized at Macrogen Europe, amplified a 382bp fragment. Genomic DNA was extracted from blood samples using DNeasy Blood & Tissue Kit (Qiagen, Germany) according to the manufacturer's instructions. A 25µl PCR mix was prepared using 12.5µl OneTaq® Quick-Load® 2X Master Mix (New England Biolabs) with Standard Buffer, 9.5µl of nuclease-free water, 0.5µl of each 10µM primer (Pair1_L and Pair1_R) and 2µl of genomic DNA extract. The PCR was performed using an AllInOneCycler™ (Bioneer) with the following conditions: initial denaturation at 94°C for 30 seconds; 30 cycles of denaturation at 94°C, annealing at 58°C and extension at 68°C for 30 seconds at

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

each step; the final extension at 68°C for 5minutes. PCR products were sent to Macrogen Europe (Amsterdam, Netherlands) for sequencing.

Heritability analyses

The heritability analysis was performed using ASReml (v4.2; https://asreml.kb.vsni.co.uk/) by fitting the fixed effects of sex and field trial year, with sire as a random effect. Genotype was included as a fixed effect when accounting for genetic effects. We also estimated the additive and dominance effects of each SNP. Defining AA, BB and AB to be the predicted trait values for each genotype class, p and q to be the SNP allele frequencies, the genetic effects were then calculated as follows: additive effect, a = (AA - BB)/2 and dominance effect, d = AB - [(AA + BB)/2].

Cell in vitro expansion and infection

Peripheral blood mononuclear cells (PBMC) were isolated from venous blood and cryopreserved in liquid nitrogen. Cells were suspended in complete RPMI culture medium and infected with *T. parva* by incubation with freshly dissected salivary glands from *R. appendiculatus* fed on animals infected with *T. parva* Muguga, stabilate 3087, as previously described (38). Tick infection levels were estimated by counting the number of infected acini in a sample of dissected salivary glands, and the sporozoite suspension adjusted to a concentration equivalent to 2000 infected acini per ml (16). During each experiment, cells from C/C and T/T animals were infected with the same batch of sporozoites. Equal volumes of sporozoites and cells (2×10^7 PBMCs) were mixed and incubated at 37°C for 90 min with periodic mixing. Cells were centrifuged, washed, and resuspended in culture medium as described. Cells were maintained in T25 flasks with fresh culture medium added every 2 to 3 days. Live cells in cultures were quantified by trypan blue staining every two days for 12 days. To quantify live, infected cells in culture, the infrared dye LIVE/DEAD™ Fixable Near-IR Dead Cell Stain (ThermoFisher Scientific) was first used to identify dead cells for removal, and the remaining live cells were stained with monoclonal antibody IL-S40.2, which recognises the polymorphic immunodominant molecule (PIM) expressed on the schizont surface. Fluorescence data were acquired for 105 cells per sample using a BD FACSCanto II flow cytometer (Becton Dickinson, Belgium) ,and analysed using Flow Jo software (FlowJo, LLC, Oregon, USA).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

This research was conducted as part of the CGIAR Research Program on Livestock. ILRI is supported by contributors to the CGIAR Trust Fund. CGIAR is a global research partnership for a food-secure future. Its science is carried out by 15 Research Centers in close collaboration with hundreds of partners across the globe (www.cgiar.org). Some of the work described in this paper was supported by grant BB/H009515/1 awarded jointly by the then UK Department for International Development and the UK Biotechnology and Biological Sciences Research Council (BBSRC) under the Combating Infectious Diseases of Livestock for International Development (CIDLID) program. This research was funded in part by the Bill & Melinda Gates Foundation and with UK aid from the UK Foreign, Commonwealth and Development Office (Grant Agreement OPP1127286) under the auspices of the Centre for Tropical Livestock Genetics and Health (CTLGH), established jointly by the University of Edinburgh, SRUC (Scotland's Rural College), and the International Livestock Research Institute (ILRI). The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation nor the UK Government. This work was also supported by funding from the BBSRC (BBS/E/D/30002275). The authors are very grateful to the management of the Ol Pejeta Conservancy, Private Bag, Nanyuki 10400, Kenya for allowing us to perform the field studies. The authors would also like to thank Mingyan Yu and Cynthia Onzere of ILRI, for preparing the DNA samples for the genomic analyses.

Author contributions

JGDP, WIM and PGT conceived the study and secured funding. EAJC, TS, JDH, MM, RN, AM, GN, SM, HA, KM, TC, MH, LJM, EJP, DW, JGDP and PT were involved in field trial design, clinical observations and sampling. JPR, EP, and PLdL processed DNA and RNA extractions. AM and PGT undertook the SNP PCR genotyping analyses. AT developed the pipeline for processing sequence data. PLdL conducted *in* vitro *T. parva*-infected cell assays. RC, PGT and BMdCB conducted the IDEAL project, and provided samples and clinical data. DW and JGDP performed statistical and bioinformatic analyses. DW, EAJC, JGDP and PGT wrote the manuscript, with feedback from all coauthors.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data availability

Whole-genome sequencing data generated from the field trial animals is available to download from the European Nucleotide Archive (ENA) under project accession PRJEB39210. The publicly available WGS data is available from the ENA under project accessions: PRJEB12739, PRJEB14552, PRJEB18113, PRJEB1829, PRJEB28185, PRJEB39282, PRJEB39330, PRJEB39352, PRJEB39924, PRJEB5462, PRJNA176557, PRJNA210519, PRJNA210523, PRJNA256210, PRJNA312138, PRJNA343262, PRJNA379859, PRJNA379859, PRJNA422979, PRJNA431934, PRJNA432125; and from the China National GenBank Database (CNGBdb) under project accession: CNP0000189. Illumina BovineHD genotype data is available to download from Edinburgh DataShare, https://doi.org/10.7488/ds/2985.

Competing interests

The Centre for Tropical Livestock Genetics and Health (CTLGH) has filed a patent application regarding the use of this locus in improving cattle through methods such as marker assisted breeding.

References

1. J. Gachohi, R. Skilton, F. Hansen, P. Ngumi, P. Kitala, Epidemiology of East Coast fever (Theileria parva infection) in Kenya: past, present and the future. *Parasit. Vectors* **5**, 194 (2012).

2. R. McLeod, P. Kristjanson, Economic Impact of Ticks and Tick-Borne Diseases to Livestock in Africa, Asia and Australia. *ACIAR - Aust. Cent. Int. Agric. Res.* (2017) (February 1, 2019).

3. , East Coast Fever. *GALVmed* (February 12, 2021).

4. B. Minjauw, A. McLeod, Tick-borne diseases and poverty. The impact of ticks and tickborne diseases on the livelihood of small-scale and marginal livestock owners in India and eastern and southern Africa. (2003) (February 1, 2019).

5. G. S. Cumming, D. P. V. Vuuren, Will climate change affect ectoparasite species ranges? *Glob. Ecol. Biogeogr.* **15**, 486–497 (2006).

6. W. O. Neitz, Theileriosis, gonderioses and cytauxzoonoses: a review. *Onderstepoort J. Vet. Res.* **27** (1957).

7. H. Epstein, *The origin of the domestic animals of Africa* (Africana Publishing Corporation, 1971) (February 1, 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

8. S. G. Ndungu, C. G. D. Brown, T. T. Dolan, In vivo comparison of susceptibility between Bos indicus and Bos taurus cattle types to Theileria parva infection. *Onderstepoort J. Vet. Res.* **72**, 13–22 (2005).

9. S. F. Barnett, Theileriosis control. *Bull Epizoot Afr* **5**, 343–357 (1957).

10. A. D. Irvin, M. P. Cunningham, A. S. Young, Eds., *Advances in the Control of Theileriosis: Proceedings of an International Conference held at the International Laboratory for Research on Animal Diseases in Nairobi, 9–13th February, 1981* (Springer Netherlands, 1981) (February 5, 2019).

11. T. Sitt, *et al.*, Exposure of vaccinated and naive cattle to natural challenge from buffalo-derived Theileria parva. *Int. J. Parasitol. Parasites Wildl.* **4**, 244–251 (2015).

12. A. Sepahi, *et al.*, Olfactory sensory neurons mediate ultrarapid antiviral immune responses in a TrkA-dependent manner. *Proc. Natl. Acad. Sci.* **116**, 12428–12436 (2019).

13. P. Küenzi, P. Schneider, D. A. E. Dobbelaere, Theileria parva-transformed T cells show enhanced resistance to Fas/Fas ligand-induced apoptosis. *J. Immunol. Baltim. Md 1950* **171**, 1224–1231 (2003).

14. B. M. de Clare Bronsvoort, *et al.*, Design and descriptive epidemiology of the Infectious Diseases of East African Livestock (IDEAL) project, a longitudinal calf cohort study in western Kenya. *BMC Vet. Res.* **9**, 171 (2013).

15. R. Callaby, *et al.*, Reflections on IDEAL: What we have learnt from a unique calf cohort study. *Prev. Vet. Med.* **181**, 105062 (2020).

16. W. I. Morrison, N. D. MacHugh, P. A. Lalor, Pathogenicity of Theileria parva is influenced by the host cell type infected by the parasite. *Infect. Immun.* **64**, 557–562 (1996).

17. P. Dutta, *et al.*, Whole genome analysis of water buffalo and global cattle breeds highlights convergent signatures of domestication. *Nat. Commun.* **11**, 4739 (2020).

18. J. Katende, *et al.*, An enzyme-linked immunosorbent assay for detection of Theileria parva antibodies in cattle using a recombinant polymorphic immunodominant molecule. *Parasitol. Res.* **84**, 408–416 (1998).

19. D. O. Odongo, J. D. Sunter, H. K. Kiara, R. A. Skilton, R. P. Bishop, A nested PCR assay exhibits enhanced sensitivity for detection of Theileria parva infections in bovine blood samples from carrier animals. *Parasitol. Res.* **106**, 357 (2009).

20. R Development Core Team, R: A Language and Environment for Statistical Computing. **1**, 409 (2011).

21. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* (2013) (January 6, 2014).

22. A. McKenna, *et al.*, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

23. M. A. DePristo, *et al.*, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

24. R. Poplin, *et al.*, Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178 (2018).

25. G. A. V. der Auwera, *et al.*, From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1-11.10.33 (2013).

26. C. C. Chang, *et al.*, Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4** (2015).

27. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinforma. Oxf. Engl.* **27**, 2987–2993 (2011).

28. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinforma. Oxf. Engl.* **27**, 2156–2158 (2011).

29. A. Manichaikul, *et al.*, Robust relationship inference in genome-wide association studies. *Bioinforma. Oxf. Engl.* **26**, 2867–2873 (2010).

30. O. Delaneau, J. Marchini, J.-F. Zagury, A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).

31. J. O'Connell, *et al.*, A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLOS Genet.* **10**, e1004234 (2014).

32. B. N. Howie, P. Donnelly, J. Marchini, A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLOS Genet.* **5**, e1000529 (2009).

33. R. W. Hamming, Error detecting and error correcting codes. *Bell Syst. Tech. J.* **29**, 147–160 (1950).

34. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).

35. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

36. W. McLaren, *et al.*, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

37. A. Untergasser, *et al.*, Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**, W71–W74 (2007).

38. B. M. Goddeeris, W. I. Morrison, Techniques for the generation, cloning, and characterization of bovine cytotoxic T cells specific for the protozoan Theileria parva. *J. Tissue Cult. Methods* **11**, 101–110 (1988).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.26.453771; this version posted July 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1. Heatmap of daily body temperature (°C) recordings throughout field trials

Figure 2. Survival analysis of unrelated versus pedigree animals

Figure 3. Regression analyses on haplotype distances

Figure 4. Prioritisation of variants

Figure 5. Survival time versus genotype

Figure 6. Cell expansion and infectivity levels in cultured cells infected with *T. parva.*

Supplementary Materials

Supplementary Table 1. Phenotypic data for animals involved in field trials

Supplementary Table 2. Daily temperature observations per animal during field trials.

Supplementary Table 3. Results of modeling to fit survival status in response to haplotype distances

Supplementary Table 4. Results of eQTL modeling

Supplementary Table 5. Results of eQTL beta comparisons to identify response eQTL

Supplementary Table 6. Variant Effect Predictor output for variants identified in the genomic interval associated with tolerance to *Theileria parva* infection, 15:45-51 Mb