Genomic heterozygosity is associated with parasite abundance, but the effects are not mediated by host condition

Sarah A. Budischak1 · Sarah Halvorsen2 · Findley Finseth1

Received: 31 October 2021 / Accepted: 17 March 2022 / Published online: 7 April 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract
Whether, when, and how genetic diversity buffers individuals and populations against infectious disease risk is a critical and open question for understanding wildlife disease and zoonotic disease risk. Several, but not all, studies have found negative relationships between infection and heterozygosity in wildlife. Since they can host multiple zoonotic infections, we sampled a population of wild deer mice (Peromyscus maniculatus), sequenced their genomes, and examined their fecal samples for coccidia and nematode eggs. We analyzed coccidia infection status, abundance, and coinfection status in relation to per-locus and per-individual measures of heterozygosity, as well as identified SNPs associated with infection status. Since heterozygosity might affect host condition, and condition is known to affect immunity, it was included as a co-variate in the per-individual analyses and as response variable in relation to heterozygosity. Not only did coccidia-infected individuals have lower levels of genome-wide per-locus diversity across all metrics, but we found an inverse relationship between genomic diversity and severity of coccidia infection. We also found weaker evidence that coinfected individuals had lower levels of private allelic variation than all other groups. In the per-individual analyses, relationships between heterozygosity and infection were marginal but followed the same negative trends. Condition was negatively correlated with infection, but was not associated with heterozygosity, suggesting that effects of heterozygosity on infection were not mediated by host condition in this system. Association tests identified multiple loci involved in the inflammatory response, with a particular role for NF-κB signaling, supporting previous work on the genetic basis of coccidia resistance. Taken together, we find that increased genome-wide neutral diversity, the presence of specific genetic variants, and improved condition positively impact infection status. Our results underscore the importance of considering host genomic variation as a buffer against infection, especially in systems that can harbor zoonotic diseases.

Sarah A. Budischak
sbudischak@kecksci.claremont.edu

1 W.M. Keck Science Department, Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, USA
2 Harvey Mudd College, Claremont, USA
**Introduction**

The spread of infectious diseases among humans, wildlife, and domestic animals continues to impact human health, the global economy, and wildlife conservation (Daszak et al. 2000; Cunningham et al. 2017). Approximately 75% of emerging infectious diseases of humans are zoonotic (Jones et al. 2008), including multiple coronavirus epidemics, which transferred from bats to intermediate hosts to humans (Frutos et al. 2021). As a result of devastating wildlife epidemics (e.g. chytridiomycosis, a fungal infection that caused declines and extinctions of numerous amphibian species (Fisher and Garner 2020)), infectious diseases are increasingly recognized as important for wildlife management and conservation (Blanchong et al. 2016). Across multiple study systems and at both the population and individual scales, heterozygosity has been associated with reduced infection prevalence and abundance. Yet, not all studies have found such associations (reviewed in (Blanchong et al. 2016)) and the relationship between heterozygosity and infection can be context-dependent (Gompper et al. 2011; Ruiz-López et al. 2012). Thus, whether, when, and how genome-wide genetic diversity buffers individuals and populations against infectious disease risk is a critical and open question for understanding and managing infectious diseases in wildlife, and also has implications for public health (Blanchong et al. 2016; DeCandia et al. 2018).

Correlations between genome-wide patterns of heterozygosity and fitness-related traits are the product of at least three non-exclusive hypotheses: 1) the direct fitness benefits of the loci studied, 2) the fitness benefits of linked loci, or 3) the general fitness costs of homozygosity throughout the genome (Hansson and Westerberg 2002). Indeed, both candidate gene and neutral genetic diversity correlated with parasite resistance in Mediterranean mouflon (Portanier et al. 2019). Adding infection into the heterozygosity-fitness paradigm seems simple at first; parasite infection can mediate inbreeding depression because less heterozygous individuals may have weaker resistance to infection (Acevedo-Whitehouse et al. 2005; Reid et al. 2007), incur higher infection rates (Luikart et al. 2008; Voegeli et al. 2012), or experience higher parasite-induced morbidity (e.g. reduced condition) (Acevedo-Whitehouse et al. 2006; Murray et al. 2013). Tolerance, an organism’s ability to limit the fitness costs of infection, also is known to have a genetic basis (Råberg et al. 2007). However, these mechanisms are physiologically linked within hosts, with bi-directional interactions well established among immunity, infection, condition, and reproduction (French et al. 2009). Indeed, a study of wild Alpine ibex used a path analysis approach to infer that the positive relationship between heterozygosity and horn growth, a fitness-related trait, was mediated by helminth abundance and body mass (Brambilla et al. 2015). Similarly, more inbred Soay sheep experienced higher overwinter mortality only during periods of high population density, but deworming ameliorated this effect (Coltman et al. 1999). Considering these within-host interactions might help elucidate relationships among environmental conditions, infection, and heterozygosity.

Since multiple factors are known to contribute to infection in wildlife, we examined the relationship between genome-wide diversity, as well as contributions of specific associated loci, in deer mice (*Peromyscus maniculatus*) and infection with a gastro-intestinal, fecal–oral transmitted coccidian parasite. Coccidia infections are a prevalent, endemic infection in mice and have been shown to affect hosts’ susceptibility to other infections (Pedersen and Antonovics 2013). Moreover, resistance to coccidia is known to have a genetic basis in other...
Deer mice carry multiple zoonotic infections, seven of which have been deemed “high priority” by the National Institute of Allergy and Infectious Disease (Barbour 2017, NIAID 2018). Although they have been proposed as a model organism for studying natural variation and adaption (Bedford and Hoekstra 2015) and have numerous genetic and genomic resources available (Kenney-Hunt et al. 2014; Harris et al. 2015; BCM-HGSC 2016; Long et al. 2019), there are surprisingly only two study of heterozygosity and infection in *Peromyscus*. Specifically, population-level allo-enzyme heterozygosity was negatively correlated with the prevalence of a gastrointestinal nematode (Meagher 1999), and a study of a single MHC-II gene found that rare alleles, but not heterozygosity, was associated with gastrointestinal helminth infection (André et al. 2017). We hypothesize that host genetic variation contributes to likelihood of coccidia infection, infection severity, and infection intensity, and will affect the likelihood of nematode coinfection.

We utilize full genome sequencing to examine per-locus measures of heterozygosity to assess the contribution of neutral genetic diversity to infection susceptibility and severity, as well as association tests to identify potential candidate genes of interest. Additionally, per-individual measures of genome-wide heterozygosity were used to test for relationships among host condition, heterozygosity, and infection. This genome-wide approach is more comprehensive than earlier neutral microsatellite methods, which may not be a good proxy for genome-wide variation and have yielded mixed findings about heterozygosity-infection relationships in wildlife (reviewed in: Blanchong et al. 2016; DeCandia et al. 2018). Genome-wide methods are also increasingly preferred over targeted gene approaches (e.g. MHC or TLR receptors) because they are not focused on a single aspect of defense (Blanchong et al. 2016; DeCandia et al. 2018). For example, genome-wide association tests, such as we use here, can uncover potential functional genes contributing to coccidia infection that may be missed from a targeted gene approach. Our combination of genome-wide per-locus and per-individual heterozygosity approaches, along with genome-wide association tests, provides a way to begin understanding the contributions of genetic factors and potential environmental-dependencies of parasite-mediated inbreeding depression.

**Materials and methods**

**Field sampling methods**

Deer mice (*Peromyscus maniculatus*) were live-trapped at the Bernard Field Station in Claremont, CA from June 27 to August 9, 2019. Eight 40×40 m grids with traps spaced 10 m apart (25 traps/grid) were established in paired locations throughout the sage-scrub habitat. At each grid, Sherman traps were baited with seeds for three nights in a row. Three-night trapping sessions were repeated at each grid every two weeks. Upon their first capture each session, mice were weighed, measured, and a retro-orbital blood sample was collected. Each mouse was also given an individually-numbered ear tag upon its first capture. Females were considered reproductive if they were pregnant or lactating, and males were classified as reproductive based on testes size. Fecal samples from each individual were collected from the traps, weighed, and stored in neutral-buffered formalin. Blood samples were kept on ice, then spun at 2000 g for 10 min to separate plasma. Plasma and pellet fractions were frozen separately at -80°C. Host trait and parasite infection data were used from each individual’s initial capture, while the blood samples sequenced sometimes came from subsequent recaptures. An average of 3.9 ± 0.6 (mean ± se) mice were included from each grid (range: 2–7 mice).
Parasite and condition measurements

To determine body condition, we regressed weight with body length then used the residuals as a measure of condition. Mass-length residuals are the most common condition metric used in condition-infection studies (Sánchez et al. 2018), and this metric has been shown to be highly correlated with body fat in deer mice (Schulte-Hostedde et al. 2005). The number of coccidia oocysts and helminth eggs per g of feces were counted using a modified McMaster egg fecal counting protocol (MAFF 1980). There were 9 uninfected and 22 coccidia-infected individuals. Following the methods of Ezenwa et al. (2021), we used a model-based approach to identify statistically significant clusters in the oocyst intensity data (S3) using the mclust package (Fraley et al. 2021). We then classified infected individuals into two groups based on these clusters [Low (<100 oocysts per gram feces (opg)): n = 11; High (>100 opg): n = 10]. This also corresponded to a natural break in the data since no samples had between 80–135 opg. For two individuals, feces weight data were not collected, so only infection status, not abundance, could be determined. One of these two individuals only had a single coccidia oocyst detected, so it was included in the Low infection severity category.

Lastly, we classified individuals as whether or not they were coinfected with coccidia (Eimeria sp.) and gastrointestinal nematodes (Aspicularis americana (a pinworm) and 3 stryongylid (hookworm) species) (Uninfected: n = 9; Low coccidia (<100 opg): n = 9; High coccidia (>100 opg): n = 7; Coinfected: n = 5; Table S2). Due to their similar biology and the general suite of host immune responses they trigger, grouping all gastrointestinal nematodes together for analysis is common for genetic studies of host resistance (Oppelt et al. 2010; Brambilla et al. 2015; Portanier et al. 2019; Ezenwa et al. 2021; Ahbara et al. 2021) as well as coccidia-nematode coinfection studies (Knowles et al. 2013; Pedersen and Antonovics 2013; Gorsich et al. 2014; Yan et al. 2021).

Sequencing and bioinformatics workflow

DNA was extracted from heparin-treated blood pellet samples with the Qiagen DNEasy Blood and Tissue Extraction Kit. DNA was quantified on a Qubit™ 2.0 fluorometer using a dsDNA High Sensitivity Assay Kit (Invitrogen). Sequencing libraries were prepared by Admera Health with a KAPA Hyper Prep Kit. Individually barcoded libraries were sequenced on an Illumina NovaSeq S4 (paired end, 2 × 150 bp), at a target depth of ~200 million PE reads per sample.

We conducted an initial assessment of read quality using FASTQC v0.11.8 with default values (Andrews 2010). TRIMMOMATIC v0.39 was then used to trim adapter sequences, the leading and trailing three bases, and low quality sequences (leading:3 trailing:3 slidingwindow:4:15 minlen:36; Bolger et al. 2014). Filtered sequences were mapped to the P. maniculatus genome (GCF_000500345.1_Pman_1.0_genomic.fna) using BWA mem v0.7.17 with default values (Li). After sorting (“sort”) and indexing (“index”) BAM files with SAMTOOLS v1.14 (Li et al. 2009), PICARD v 2.26.1 was used to remove duplicates (“MarkDuplicates”) and add read groups ("AddOrReplaceReadGroups"; validation_stringency=lenient; Broad Institute 2019).

We called variants in the parallel mode of FREEBAYES v1.3.5 with 16 threads, resulting in an initial total of 153,221,207 variant sites (Garrison and Marth 2012). To retain only high quality SNPs, we first used vcftools 0.1.17 to filter variant sites with >5% missing
data (–max-missing 0.95), quality scores < 40 (–minQ 40), minimum depth < 8 (–minDP 8), and a minor allele frequency of at least 1.5% (–maf 0.015) (Danecek et al. 2011). We next filtered for allele balance, or the proportion of reads supporting a variant, to reduce paralogs and errors using the vcffilter tool in the vcflib package v1.0.1 (Garrison et al. 2021). We removed sites with an allele balance < 0.25 or > 0.75, with the exception that we retained allele balances close to zero to keep variants where all individuals were homozygous for one or the other allele. Vcftools v0.1.17 was then used to retain only biallelic SNPs with a genotype quality > 20 (–min-alleles 2, –max-alleles 2, –minGQ 20, –remove-indels) and remove any loci that were out of Hardy–Weinberg Equilibrium (P < 0.01; –hwe 0.01) in Uninfected individuals (Danecek et al. 2011). Linkage disequilibrium filtering was performed by pruning out sites with $r^2 > 0.6$ in 1000 site windows with bcftools v1.11 (+ prune -m 0.6 -w 1000; Li 2011; Danecek et al. 2021). After filtering, 573,838 high-quality SNPs were retained for downstream analyses.

**Genetic diversity analyses**

We predicted that individuals with higher levels of genetic diversity would be less likely to be infected and have less severe infections. To this end, we compared genetic per-locus diversity metrics across individuals varying in (a) coccidia infection status (Uninfected/infected); (b) coccidia severity (Uninfected/Low/High); and (c) coinfection status (Uninfected/Low/High/coinfected). Per-locus measures of diversity average heterozygosity metrics across each SNP for individuals within each infection-status group for each site in the genome (n = 573,838 SNPs per group). We used the basicstats function of hierstat v0.5–7 to calculate per-locus observed heterozygosity ($H_O$), expected heterozygosity ($H_E$), and the inbreeding coefficient ($F_{IS}$)(Goudet and Jombart 2020). In addition to the measures of heterozygosity described here, we wanted to estimate the number of segregating alleles per group or “allelic richness”. Because allelic richness can be sensitive to small sample size (Pruett and Winker 2008), we used a rarefaction approach in ADZE v1.0 to calculate per-locus allelic richness ($A_R$) and private allelic richness ($P_{AR}$) (Szpiech et al. 2008). For $A_R$, we standardized sample sizes from two to the number of individuals in the largest group (infection status: 22; severity: 11; coinfection: 9). For $P_{AR}$, we set the standardized sample size to the size of the smallest group (9 for infection status and severity; 5 for coinfection). All other parameters were default values. Because the number of coinfectected individuals is lower than the other categories, we only considered analyses that accounted for differences in sample sizes (i.e., $P_{AR}$ and $A_R$).

Per-individual genetic diversity metrics were calculated in vcfTools v0.1.17 and included $H_O$, $H_E$, and $F_{IS}$. Per-individual measures of diversity average across all loci in the genome for each individual (n = 31 individuals). We assessed statistical differences of genetic diversity means among groups using Welch’s t-test (infection status: Uninfected vs. Infected; sex: male vs. female) or ANOVA (severity: Uninfected vs. Low vs. High; coinfection: Uninfected vs. Low vs. High vs coinfection) in R version 4.0.4 (R Core Team 2021). Tukey’s HSD was applied for post-hoc comparisons where appropriate. For rarefaction metrics ($A_R$, $P_{AR}$), we calculated 95% confidence intervals in ADZE v1.0. VCFs were manipulated using the vcfR v1.12.0 (Knaus and Grünwald 2017) package and figures were made with the ggplot2 v3.3.5 (Wickham 2016) and cowplot v1.1.1 (Wilke 2020) packages in R.

We used a linear modeling approach to determine how per-individual genetic diversity, as well as other factors, influenced the response variables of infection status,
abundance and intensity. For infection status, individuals were coded as Uninfected and Infected (as above). Abundance was measured as the log oocysts per gram (opg) coccidia of all individuals and included Uninfected individuals. Intensity was measured as the log oocysts per gram coccidia in Infected individuals only. Covariates included a metric of individual-level genetic diversity (either $H_O$ or $F_{IS}$), sex (male/female), age (adult/sub-adult), grid (trapping location: HM, NK, RC, FH), and residual body condition. Genetic diversity metrics for each individual were standardized by subtracting the population mean and dividing by the standard deviation. To make sure that potential relatedness of mice within grids was not affecting our results, we used ANOVAs to confirm that grid was not associated with per-individual $H_O$, $H_E$, and $F_{IS}$ ($H_O$: $F = 0.41, p = 0.90$; $H_E$: $F = 0.81, p = 0.60$; $F_{IS}$: $F = 0.43, p = 0.89$). Additionally, all association tests (below) controlled for pairwise genetic relatedness.

For testing the relationship between per-individual heterozygosity and infection status, we built general linearized models fitted to a binomial distribution with a logit link function to determine the significance of fixed effects. For abundance and intensity, we built general linearized models fitted to gaussian distributions. Backward stepwise model selection was performed by comparing AIC values in R with the stepAIC function from the MASS package (Venables and Ripley 2002). Linear models for host condition were parameterized similarly, with covariates including: a metric of individual-level genetic diversity (either standardized $H_O$ or $F_{IS}$), sex (male/female), age (adult/sub-adult), grid (trapping location: HM, NK, RC, FH), and log-transformed coccidia abundance. We also included an interaction between the metric of genetic diversity and coccidia abundance as a measure of tolerance. Backward stepwise model selection was performed by comparing AIC values in R with the stepAIC function. Post-hoc power analyses for regression and binomial analyses were conducted using the pwr and pwr2ppl packages, respectively (Aberson 2019; Champely et al. 2020).

**Association tests**

We used GEMMA v0.98.4 to perform univariate single-marker association tests to identify loci associated with infection status (Zhou and Stephens 2012, 2014). Sex was included as a covariate and a centered pairwise genetic relatedness matrix was calculated to account for population structure (gk -1). Infection status was defined as Uninfected = 1 and Infected = 2. The lrt $P$-values were adjusted using a false discovery rate (Benjamini and Hochberg 1995). We used the relevant.gff file to determine whether significant loci were in genes. When necessary, location IDs were converted to gene names in NCBI gene (https://www.ncbi.nlm.nih.gov/gene). For the set of genes that contained a significantly associated locus, we performed a gene ontology analysis using the gost function from the gprofiler2 package v0.20.0 (Kolberg et al. 2020). Gene functions were ascertained in the Online Mendelian Inheritance in Man (https://www.omim.org/), NCBI gene (https://www.ncbi.nlm.nih.gov/gene), and/or Uniprot (https://www.uniprot.org/) databases. Because the genomic and immunological basis of resistance to coccidia infection has been previously investigated (e.g. Chapman et al. 2013; Boulton et al. 2018), we cross-referenced the functions of genes with associated loci with previously identified potential genes and pathways involved in coccidia resistance.
Results

Sampling and sequencing summary

We performed whole genome sequencing and disease phenotyping on 31 *P. maniculatus* (Table S1). Mice were classified according to infection status (Infected vs. Uninfected), severity (Uninfected, Low, High), and coinfection with nematodes. After trimming adapters and low quality regions, individuals had an average of 350 million reads retained yielding an average coverage of 20.77. Initial variant calling produced 153,221,207 variable sites. After filtering, 573,838 high quality SNPs were retained for downstream analyses.

Infection status

When considering infection status, Uninfected mice had higher levels of per-locus genetic diversity than Infected mice across most metrics (Fig. 1). Per-locus, Uninfected mice had higher levels of $H_O$ (two-sided t-test; Inf = 0.486 ± 0.0004, Uninf = 0.492 ± 0.0004, $t_{1138387} = 11.944$, $p < 0.001$), $H_E$ (two-sided t-test; Inf = 0.357 ± 0.0002, Uninf = 0.358 ± 0.0002, $t_{1136475} = 3.179$, $p = 0.001479$), and $F_{IS}$ (two-sided t-test; Inf = -0.286 ± 0.0004, Uninf = -0.311 ± 0.0005, $t_{1075762} = -38.226$, $p < 0.001$) than Infected mice (Fig. 1a,b,c). Although $H_O$, though not $H_E$, can be sensitive to smaller sample sizes, we note that any bias would predict groups with the lowest number of samples to have lower diversity values (Pruett and Winker 2008). Instead, we find that the smaller group (Uninfected mice, n = 9), has significantly higher levels of all three heterozygosity metrics than the larger group (Infected mice, n = 22). While $A_R$ was not different across all sample sizes tested (overlapping error bars), $P_{AR}$ was lower in Infected (0.0851; 95% CIs: 0.0847–0.0855) than Uninfected mice (0.091; 95% CI: 0.0906 – 0.0914) at the sample size equivalent to the Uninfected group size (n = 9; Fig. 1d,e). We also note that our SNP filtering may have affected our ability to detect differences in allelic richness, as we excluded any SNPs with more than 2 alleles and removed indels.

Infection severity

Mice with higher levels of per-locus genetic diversity had less severe infections across several metrics of genetic diversity (Fig. 3). $H_O$ was highest in Uninfected mice (0.492 ± 0.0004), at intermediate levels in Low-coccidia mice (0.491 ± 0.0004), and lowest in High-coccidia mice (0.482 ± 0.0004; ANOVA: $F_{1715013} = 209.9$, $p < 0.001$; Tukey’s HSD $p < 0.001$ in all pairwise comparisons; Fig. 2a). High-coccidia mice had significantly lower levels of $H_E$ (0.357 ± 0.0002) than Low-coccidia (0.358 ± 0.0002) and Uninfected (0.358 ± 0.0002) mice, though Low-coccidia and Uninfected mice were not significantly different from each other (ANOVA: $F_{1715013} = 20.29$, $p < 0.001$; Tukey’s HSD: High vs Low $p < 0.001$, High vs Uninfected $p < 0.001$, Low vs Uninfected: $p = 0.710$; Fig. 2b). Inbreeding coefficients ($F_{IS}$) were associated with severity in the predicted direction (High = -0.286 ± 0.0005, Low = -0.298 ± 0.0005, Uninfected -0.311 ± 0.0005; ANOVA: $F_{1608741} = 612.7$, $p < 0.001$; Tukey’s HSD $P < 0.001$ in all pairwise comparisons; Fig. 2c). Again, our results are in the opposite direction of any bias imposed by small sample sizes, as we find the smallest group (Uninfected mice) to have the largest $H_O$. However, $H_E$, which is less sensitive to sample size issues, only detected differences in High vs. the other two categories (Pruett and Winker 2008);
Fig. 1 Genome-wide per-locus diversity metrics for mice categorized according to infection status (Uninfected: n = 9; Infected n = 22). Per-locus genomic diversity metrics are significantly lower in Uninfected vs Infected mice for a) observed heterozygosity ($H_O$; $p < 0.001$), b) expected heterozygosity ($H_E$; $p = 0.001479$), c) the inbreeding coefficient ($F_{IS}$; $p < 0.001$) and, d) private allelic richness ($P_{AR}$; non-overlapping 95% CIs). Across all sample sizes, e) allelic richness ($A_R$) does not differ among groups (overlapping confidence intervals). Points are means and error bars are ± one standard error, with the exception of $F_{IS}$ which is shown with ± two standard errors for visualization purposes. For $A_R$ (e), the curve is the smoothed conditional mean for Uninfected (dashed) and Infected (solid) mice, and the shading represents 95% confidence intervals.
Fig. 2 Genome-wide per locus diversity metrics for mice categorized according to infection severity (Uninfected: n = 9; Low: n = 11; High: n = 10). Per-locus genomic diversity metrics include a) observed heterozygosity ($H_O$; all comparisons $p < 0.001$), b) expected heterozygosity ($H_E$; High vs Low/Uninfected $p < 0.001$; Low vs Uninfected ns), c) the inbreeding coefficient ($F_{IS}$; all comparisons $p < 0.001$), d) private allelic richness ($P_{AR}$; Uninfected vs Low/High non-overlapping CIs; Low vs High ns), and e) allelic richness ($A_R$; overlapping CIs). Points are means and error bars are ± two standard errors. For $A_R$ (e), the curve is the smoothed conditional mean for Uninfected (white, dotted), Low (medium gray, long dash), and High mice (dark grey, solid), and the shading represents 95% confidence intervals.
this suggests that highly infected mice have lower neutral diversity, but differences in low vs uninfected mice should be interpreted with caution.

Allelic diversity showed mixed patterns. \(P_{AR}\) was higher in Uninfected mice (0.0343; 95% CI: 0.0341–0.0346) than other groups, though Low (0.0313; 95% CI: 0.0311–0.0316) and High mice (0.0317; 95% CI: 0.0314–0.0319) had similar levels of \(P_{AR}\) (Fig. 3d). \(A_R\) did not differ across groups at all sample sizes tested (overlapping error bars; Fig. 2e), and previous caveats about SNP filtering apply here as well.

### Coinfection

Cinfeected individuals had lower levels of per-locus allelic diversity than other groups. At a sample size of 5 individuals, \(P_{AR}\) was significantly lower in coinfeected mice (0.0231; 95% CI: 0.0229–0.0232) than all other groups (Fig. 3a; Uninfected – 0.0315, 95% CI: 0.0313–0.0317; Low – 0.0287, 95% CI: 0.0285–0.0289; High – 0.0301, 95% CI: 0.0299–0.0303). Indeed, all groups had significantly different levels of \(P_{AR}\), with Uninfected mice having the highest values, followed by High, Low, and coinfeected mice. Coinfeected mice also showed a nonsignificant trend towards lower levels of \(A_R\) across all samples sizes tested (Fig. 3b; low overlap of confidence intervals). While rarefaction analyses such as \(A_R\) are recommended for small samples sizes, we note that our results for coinfection are in the predicted direction of any bias. Here, we find the group with the smallest sample size (coinfeected, \(n=5\)) also shows the lowest values of per-locus allelic diversity. Further work with larger sample sizes should be done to confirm this finding.
Individual-level analyses

In addition to per-locus metrics, we used a linear modeling approached to examine how per-individual levels of heterozygosity (standardized $H_O$ and standardized $F_{IS}$), condition, and other factors influenced infection. Neither standardized $H_O$ or $F_{IS}$ affected the likelihood that an individual was infected with coccidia, though both terms were retained in reduced models (Table 1). In those models, body condition was negatively associated with infection, indicating that individuals with higher body condition were less likely to be infected. Both sex and age were retained in the final models and had marginal p-values, with females and adults perhaps being slightly more likely to be infected (Table 1).

Looking at the quantitative measures of infection, $H_O$ and $F_{IS}$ were the only terms retained in the final linear models for coccidia intensity (Table 1). Their p-values were marginal with a negative trend observed for intensity and $H_O$ ($p = 0.078$), and a positive

Table 1 Results of reduced linear models for coccidia infection status (Infected vs. Uninfected), intensity (log opg), and abundance (log opg), versus residual body condition. The results of the analyses for condition as a function of $H_O$ and $F_{IS}$ are identical (and thus reported only once) because the heterozygosity measures drop from both models. Significant p-values are in bold, marginal values are in italics

|                           | Estimate | Std. Error | t value | p value |
|---------------------------|----------|------------|---------|---------|
| Infection status ~ $H_O$  |          |            |         |         |
| $H_O$                     | $-1.04$  | $0.75$     | $-1.37$ | $0.169$ |
| Sex (M)                   | $-3.62$  | $1.97$     | $-1.84$ | $0.066$ |
| Age (SA)                  | $-4.30$  | $2.49$     | $-1.72$ | $0.085$ |
| Condition                 | $-0.77$  | $0.29$     | $-2.60$ | $0.009$ |
| Infection status ~ $F_{IS}$|          |            |         |         |
| $F_{IS}$                  | $0.99$   | $0.75$     | $1.32$  | $0.186$ |
| Sex (M)                   | $-3.59$  | $1.95$     | $-1.84$ | $0.066$ |
| Age (SA)                  | $-4.29$  | $2.50$     | $-1.72$ | $0.086$ |
| Condition                 | $-0.76$  | $0.29$     | $-2.61$ | $0.009$ |
| Intensity ~ $H_O$         |          |            |         |         |
| $H_O$                     | $-0.27$  | $0.140$    | $-1.87$ | $0.078$ |
| Inensity ~ $F_{IS}$       |          |            |         |         |
| $F_{IS}$                  | $0.27$   | $0.142$    | $1.93$  | $0.069$ |
| Abundance ~ $H_O$         |          |            |         |         |
| $H_O$                     | $-0.33$  | $0.20$     | $-1.61$ | $0.123$ |
| Sex (M)                   | $-1.01$  | $0.50$     | $-2.04$ | $0.056$ |
| Age (SA)                  | $-1.51$  | $0.74$     | $-2.04$ | $0.055$ |
| Condition                 | $-0.21$  | $0.07$     | $-2.90$ | $0.009$ |
| Abundance ~ $F_{IS}$      |          |            |         |         |
| $F_{IS}$                  | $0.33$   | $0.18$     | $1.81$  | $0.083$ |
| Sex (M)                   | $-0.77$  | $0.41$     | $-1.87$ | $0.074$ |
| Age (SA)                  | $-1.30$  | $0.64$     | $-2.03$ | $0.054$ |
| Condition                 | $-0.19$  | $0.06$     | $-2.97$ | $0.007$ |
| Condition ~ $H_O$ or $F_{IS}$|          |            |         |         |
| Grid (HM)                 | $-0.91$  | $1.36$     | $-0.67$ | $0.511$ |
| Grid (NK)                 | $1.97$   | $1.59$     | $1.24$  | $0.228$ |
| Grid (RC)                 | $-1.30$  | $1.35$     | $-0.97$ | $0.344$ |
| Sex (M)                   | $-1.75$  | $1.06$     | $-1.66$ | $0.112$ |
| Age (SA)                  | $-6.09$  | $1.55$     | $-3.93$ | $0.001$ |
| Coccidia Abundance        | $-1.07$  | $0.46$     | $-2.32$ | $0.030$ |
trend for $F_{IS}$ ($p = 0.069$). $H_O$ was not associated with coccidia abundance, but a marginal positive trend was again observed for $F_{IS}$ (Fig. 4a, Table 1, $p = 0.083$). However, these analyses were slightly underpowered (power $\approx 0.71$ for a large effect size). Condition was negatively associated with coccidia abundance (Fig. 4b, Table 1, $p = 0.007$). Similar to the infection status analyses, the abundance models also retained sex and age; they had marginal $p$-values indicating females and adults might have slightly higher infection abundances than males and subadults, respectively (Table 1, $p = 0.054$ (age), $p = 0.074$ (sex)).

Both $H_O$ and $F_{IS}$ were dropped from linear models of condition (Table 1), and visually it is clear that there is no relationship between $F_{IS}$ and condition (Fig. 5c, $p = 0.76$). The interaction between genetic diversity and parasite load, a proxy measure of tolerance, also dropped from the models (Table 1). Coccidia abundance was negatively associated with condition, and subadults had lower condition indices than adults (Table 1, $p = 0.030$ (condition), $p = 0.001$ (age)). We also examined whether individual-level heterozygosity metrics varied between the sexes and found no significant differences ($H_O$: $p = 0.115$, $H_E$: $p = 0.335$, $F_{IS}$: $p = 0.119$).

**Association tests**

We found 73 SNPs associated with coccidia infection status that passed filtering after applying a false discovery rate (Table S3). Sixteen of the associated loci fell within annotated genes (Table 2). For 35 loci, the associated SNP was associated with a reduction in disease prevalence (6 within annotated genes; i.e., negative beta coefficient), while 38 associate loci were associated with an increase in prevalence (10 within annotated genes; i.e., positive beta coefficient). The gene ontology analysis revealed significant enrichment for one molecular function (protein serine/phosphatase activity; GO:0004,722; $p = 0.024$), one cellular component (RNA polymerase II, holoenzyme; GO:0016,591; $p = 0.036$), and no biological processes.

---

**Fig. 4** Partial residual plots from linear regression models show a) the marginally positive correlation between per-individual $F_{IS}$ and coccidia abundance ($p = 0.083$), b) the negative correlation between residual body condition and coccidia abundance ($p = 0.007$), and c) the lack of relationship between per-individual $F_{IS}$ and residual body condition ($p = 0.76$)
Discussion

Identifying the factors that determine host susceptibility to infection can inform the management and evolutionary consequences of wildlife disease. We explored the genetic factors contributing to coccidia infection in *Peromyscus maniculatus*. Our results suggest that both neutral diversity, as well as the presence of certain genetic variants, influence coccidia infection. We find strong evidence that per-locus host genomic diversity is negatively

---

Table 2  Single nucleotide polymorphism (SNP) loci that are significantly associated with disease prevalence. The complete list of associated SNPs is detailed in Table S3. Note: Scaffold number; SNP position; minor allele (A1); major allele (A2); frequency of the minor allele in the full population (f(A1)); frequency of the minor allele in uninfected individuals (f(U)); frequency of the minor allele in infected individuals (f(I)); beta coefficient; standard error of beta; P calculated from a likelihood ratio test; gene name; gene annotation

| Scaffold | SNP | A1 | A2 | f(A1) | f(U) | f(I) | Beta  | SE    | P   | Gene Description |
|----------|-----|----|----|-------|------|------|-------|-------|-----|------------------|
| NW_006501040.1 | 10082353 | T  | G  | 0.08  | 0.00 | 0.28 | -8.46E-01 | 1.71E-01 | 1.03E-06 | PPA2 Pyrophosphatase, inorganic 2 |
| NW_006501050.1 | 3461837 | A  | G  | 0.37  | 0.48 | 0.11 | 7.83E-01 | 1.26E-01 | 3.18E-08 | PSME4 Proteasome activator subunit 4 |
| NW_006501082.1 | 5389931 | G  | A  | 0.42  | 0.50 | 0.22 | 8.35E-01 | 1.60E-01 | 4.96E-07 | KLF3 Kruppel-like factor 3 |
| NW_006501128.1 | 151420 | A  | G  | 0.27  | 0.36 | 0.06 | 5.95E-01 | 1.21E-01 | 3.71E-06 | KCNN2 Potassium channel, calcium-activated, subfamily N, member 2 |
| NW_006501173.1 | 3102328 | T  | C  | 0.18  | 0.05 | 0.50 | -6.38E-01 | 1.01E-01 | 3.02E-06 | MAP7 Microtubule-associated protein 7 |
| NW_006501271.1 | 4971055 | A  | G  | 0.37  | 0.48 | 0.11 | 7.83E-01 | 1.26E-01 | 3.18E-08 | PSME4 Proteasome activator subunit 4 |
| NW_006501309.1 | 12430996 | G  | A  | 0.23  | 0.11 | 0.50 | -6.39E-01 | 1.60E-01 | 4.96E-07 | RPAF2 RNA polymerase II-associated protein 2 |
| NW_006501442.1 | 604457 | C  | G  | 0.24  | 0.34 | 0.00 | 5.67E-01 | 1.16E-01 | 2.39E-06 | TTF2 Transcription termination factor 2 |
| NW_006501554.1 | 508093 | T  | C  | 0.32  | 0.43 | 0.06 | 6.57E-01 | 1.15E-01 | 1.57E-07 | CYP4F14 Leukotriene B4 omega-hydroxylase 3-like |
| NW_006501554.2 | 1588396 | G  | C  | 0.38  | 0.48 | 0.11 | 7.05E-01 | 1.39E-01 | 8.86E-07 | HAVCR1 Hepatitis A virus cellular receptor 1 homolog |
| NW_006501554.3 | 18267 | T  | C  | 0.12  | 0.02 | 0.33 | 7.25E-01 | 1.54E-01 | 5.64E-06 | BIRC3 Baculoviral IAP repeat-containing protein 1e-like; Also NAP |
| NW_006501554.4 | 100354 | C  | T  | 0.39  | 0.48 | 0.17 | 7.25E-01 | 1.54E-01 | 5.64E-06 | SIRPB1 Signal-regulatory protein beta-1-like |
| NW_006501554.5 | 14107 | C  | T  | 0.44  | 0.50 | 0.28 | 8.73E-01 | 1.53E-01 | 1.53E-06 | ODC1 Ornithine decarboxylase 1 |
| NW_006501554.6 | 75024 | G  | A  | 0.25  | 0.36 | 0.00 | 5.93E-01 | 1.23E-01 | 3.75E-06 | CD300LD3 CMRF35-like molecule 3 |
| NW_006501554.7 | 85396 | T  | C  | 0.42  | 0.50 | 0.22 | 8.38E-01 | 1.67E-01 | 1.13E-06 | TAD1A Transcriptional adapter 3 |
| NW_006501554.8 | 11314 | C  | A  | 0.15  | 0.05 | 0.39 | 6.81E-01 | 1.37E-01 | 2.26E-06 | U1AT3E1 Sulfotransferase family 1E, estrogen-prefering, member 1 |
associated with infection status and severity, and weaker evidence that lower genetic diversity increases likelihood of coinfection. We also placed per-individual heterozygosity in context with several other factors that may contribute to infection state, including condition. Condition, along with a trend for individual-level heterozygosity, were negatively correlated with infection abundance and intensity, though the effects of condition and heterozygosity, since they were uncorrelated, are independent. Our results underscore the importance of considering genomic variation, in conjunction with other measures of fitness, to understand the ecology and evolution of host-parasite interactions.

Genomic diversity and infection

Loss of genomic diversity and inbreeding are hypothesized to increase host susceptibility to pathogens through decreasing the adaptive potential of hosts and unmasking deleterious recessive alleles (Spielman et al. 2004; Blanchong et al. 2016). Here, we find that genome-wide neutral variation is an important factor contributing to parasite infection in *P. maniculatus*. Specifically, lower levels of per-locus genomic variation result in a higher likelihood of infection and increased infection severity. These results are consistent whether considering patterns of heterozygosity (*H*$_O$), measures of nucleotide diversity (*H*$_E$), relative inbreeding (*F*$_{IS}$), or allelic variation (*P*$_{AR}$). The main effect of genomic diversity on infection is mediated through differences in patterns of heterozygosity, rather than allelic richness, with the exception that private alleles also play a role. In per-individual analyses factoring in host traits, we observed marginal relationships between standardized heterozygosity metrics and infection intensity and abundance.

Our per-locus analyses show genomic diversity affects host resistance to coccidia, suggesting more heterozygous hosts reduce their parasite load once infected (e.g. lower intensity) and perhaps clear some infections entirely (e.g. lower likelihood of infection). However, our individual-scale analyses, which had lower power due to sample size, found no relationship between heterozygosity and infection status. Previous studies from diverse systems often (Acevedo-Whitehouse et al. 2005, 2006; Townsend et al. 2018; Banks et al. 2020; DeCandia et al. 2021), but not always (Acevedo-Whitehouse et al. 2009; Townsend et al. 2018; Kubacka et al. 2020; Martin et al. 2021), find negative associations between genome-wide variation and infection status (reviewed in (Blanchong et al. 2016)). While there is well-described functional genetic variation for host susceptibility to coccidia infection (Klesius and Hinds 1979; Kim et al. 2006; Boulton et al. 2018; present study) reviewed in (Chapman et al. 2013)), here we also document a role for genome-wide neutral diversity. In *P. maniculatus*, the population-level prevalence of a different fecal–oral parasite (*Capillaria hepatica*) negatively correlates with genetic diversity, after controlling for density (Meagher 1999). Other systems that explore fecal–oral parasites have found similar relationships with genetic diversity and infection status (e.g., Coltman et al. 1999; Acevedo-Whitehouse et al. 2006), though this is not always the case (Schwensow et al. 2007; Townsend et al. 2018).

In addition to associations with infection status, we documented negative correlations between per-locus genomic diversity and parasite load. Recent work on sarcoptic mange in wolves that also used genome-wide SNP data found a similar negative relationship between per-locus measures of genomic variation and disease severity (DeCandia et al. 2021). In other systems, multi-locus heterozygosity negatively correlates with overall fecal–oral parasite loads (Brambilla et al. 2015; Mitchell et al. 2017), though the effect disappeared when examining coccidia load separately in at least one taxa (Mitchell et al. 2017). A reduction
in coccidia load has also been linked to heterozygosity at specific immune loci (Rödel et al. 2020), as well as the presence of specific alleles contributing toward resistance (Kim et al. 2006; Worley et al. 2010; Boulton et al. 2018). The per-locus severity relationship (and trends at the individual level) could be due to higher parasite replication within the host. Because *Eimeria* undergo an asexual reinfective cycle prior to the sexual reproduction that produces the oocysts (Bowman 2009), reduced immunity in relatively inbred individuals could result in higher coccidia oocyst shedding.

Nematode-coccidia coinfected individuals had lower levels of private allelic richness than other groups and overall allelic richness followed a similar trend. Specifically, individuals with less allelic variation were more likely to be coinfected, and coinfected individuals had the fewest number of unique alleles. Both having a smaller number of alleles ($A_R$), as well as the presence of particular unique alleles ($P_{AR}$), could increase an individual’s risk of coinfection. For example, in African buffalo, particular alleles were associated with the likelihood and severity of nematode infection, and a worm-resistant phenotype increased the progression and severity of bovine tuberculosis coinfection (Ezenwa et al. 2021). Notably, our analyses were limited by low sample size due to the low prevalence of nematodes in the study population. Additionally, since we found no individuals infected with only nematodes, we cannot distinguish the possibility that the lower $P_{AR}$ is a consequence of nematode-infection, rather than coinfection. Therefore, further investigations with increased sampling are required to know whether our coinfection results are robust.

A positive relationship between heterozygosity and sequencing depth, as well as mis-mapping, can confound estimates of genome-wide heterozygosity in studies such as the present one. At least five lines of reasoning mitigate these concerns here: i) we applied very stringent filtering to variant sites, including retaining only sites with high read depth, mapping quality, and genotype quality (<0.5% of initial calls were kept for final analyses), ii) only sites that were in Hardy–Weinberg Equilibrium in Uninfected individuals were kept, iii) such sequencing or mapping errors do not predict systematic biases that vary according to infection state, iv) patterns are consistent across all metrics of genomic variation, and v) a preliminary analysis using a different bioinformatic workflow recovered similar relationships between heterozygosity and coccidia infection (PLATYPUS)(WGS500 Consortium et al. 2014). As further evidence that the infection-based patterns are real, we evaluated whether heterozygosity varied according to sex (i.e., another way to categorize individuals) and found no marginal or significant differences.

**Condition, heterozygosity and infection**

Host body condition was negatively correlated with coccidia infection and abundance, but was not associated with either measure of individual-level heterozygosity ($H_0$ and $F_{IS}$). Thus, we found no evidence that general costs of homozygosity throughout the genome were manifest in lower body condition. Two non-exclusive mechanisms could be driving the negative correlation between infection and host condition: 1) since immune responses are energetically costly, individuals in lower condition might be less resistant to infection, and/or 2) lower condition could be a consequence of parasite infection (Beldomenico and Begon 2010; Sánchez et al. 2018). If condition is affecting infection via altered immunity (Fig. 5a), then condition and heterozygosity are having synergistic effects on infection status and intensity. However, we cannot rule out the possibility that the effects of heterozygosity on infection are also mediated via changes in host immunity and resistance. We found no evidence that heterozygosity affected host tolerance. Future analyses focused on
whether immune genes experience elevated diversity relative to non-immune genes could elucidate the mechanism underlying the per-locus infection-heterozygosity correlations as well as examine if the individual-level trends reach significance with larger sample sizes and multiple timepoint infection measures.

Alternatively, the lower condition observed in infected individuals could be a consequence of coccidia infection, and thus an indirect effect of host heterozygosity (Fig. 5b). Individuals with higher parasite loads may not necessarily have decreased infection-imposed fitness. For example, while more inbred sea lions had a higher prevalence of hookworm, anemia was only associated with homozygosity at a single locus (Acevedo-Whitehouse et al. 2006). However, in a similar longitudinal field study of deer mice, the frequency of infection with the coccidian parasite *Eimeria arizonensis* was associated with lower body weight and decreased overwinter survival, but these effects were dependent on sex and reproductive status (Fuller 1996). Specifically, effects of coccidia on mouse weight were found only in lactating females, and effects on survival were found only in male mice (Fuller 1996). In our study population, longitudinal infection and condition data could help determine if condition is a cause or consequence, or both a cause and consequence, of infection (Fig. 5c).

**Associated loci point towards a role for the inflammatory response and echo previous work**

Because coccidia resistance is known to be heritable (Kim et al. 2006; Beraldi et al. 2007b) and specific immune loci have been linked to coccidia resistance (e.g., Kim et al. 2006; Worley et al. 2010; Boulton et al. 2018; Rödel et al. 2020), we performed a genome-wide association test to identify functional candidates associated with coccidia infection in *Peromyscus*. Previous work has shown that the genomic basis of coccidia resistance is polygenic with both many loci of small effect and a few loci of large effect contributing to this complex trait (e.g., Kim et al. 2006, 2010; Hamzić et al. 2015; Boulton et al. 2018). Our study identified numerous SNPs across at least 16 distinct genes that are significantly associated with the coccidia infection status (Table 2). The discovery of specific associated loci complements genome-wide patterns of diversity, suggesting that both the broad benefits of heterozygosity as well as certain genetic variants contribute to disease resistance. This finding mirrors other studies of coccidia resistance in wild populations of terrestrial mammals; both diversity at candidate immune genes and neutral genetic diversity were correlated with gastrointestinal nematode abundance in European mouflon, though surprisingly this study found no relationship between heterozygosity and coccidia abundance (Portanier et al. 2019).

Many of the associated loci found here echo previous work from poultry and lab mice, suggesting that the genetic pathways controlling coccidia infection may be shared across diverse domesticated and wild species (Table 2). A host's ability to control coccidia requires a strong, rapid inflammatory response and contributing genetic factors link to cell-mediated, humoral, and innate components of immunity (Chapman et al. 2013; Boulton et al. 2018). We find numerous associated loci that are important in the inflammatory response. At least five associated SNPs are in genes that interact with the transcription factor NF-κB (e.g., *PSME4, KLF3, CD300LD3, BIRC3, HAVCR1*), a pivotal mediator of inflammatory responses that is at the center of pathways conferring coccidia resistance in other species (Alcala-Canto and Ibarra-Velarde 2008; Psifidi et al. 2016; Boulton et al. 2018; Jin et al. 2020; Knights et al. 2020). One exciting functional candidate is *CD300LD3*
(also known as CLM-3 or CMRF35-like molecule 3), which can activate cytokine production in mast cells and lead to increased production of proinflammatory cytokines (Wu et al. 2011). CMRF35-like molecules also stimulate toll-like receptor 9 pathways and subsequently NF-κB, both of which are implicated in coccidia resistance in chickens (Boulton et al. 2018). Additionally, several candidate genes are known to show enhanced expression in neutrophils (SIRPB1, CD300LD3, MAP7; https://www.proteinatlas.org/) or to modulate inflammatory signaling (e.g., BIRC3). A GWAS performed in chickens found that leukotriene B4 receptor 2 (LTB4R2), a cell surface receptor of the inflammatory leukotriene B4, was significantly associated with resistance to coccidia parasitism (Psifidi et al. 2016). Our study also implicates the leukotriene B4 pathway, as we find a significantly associated SNP in CYP4F14 (leukotriene-B4 omega-hydroxylase 3), a protein involved in leukotriene B4 degradation. Our results also offer novel candidates for contributing to coccidia resistance. For example, a significantly associated SNP is located in the helper-T cell-surface receptor HAVCR1 (Hepatitis A virus cellular receptor 1 homolog), which regulates cytokine production (McIntire et al. 2003). After mice were infected with hepatitis A, certain variants of this gene conferred protective immunity against non-viral diseases (McIntire et al. 2003). Future work can focus on molecular experiments with these functional candidates to investigate and validate their role in coccidia infection. Taken together, the resonance of specific associated loci found here with previously documented associations underscores our ability to detect host genetic factors contributing to the coccidia infection, despite small sample sizes and a single snapshot of a sublethal pathogen.

**Conclusions and future directions**

We find that increased neutral genomic diversity, as well as the presence of specific genetic variants, contribute to decreased coccidia infection in wild populations of deer mice. Our genome-wide association tests identify well-described, as well as novel, functional candidates that can be the focus of future molecular work exploring the effect of particular loci on coccidia resistance. But what is the mechanistic basis of associations between neutral genomic diversity and infection? Our study does not distinguish among whether fitness benefits of heterozygosity are conferred through direct effects of loci, functional effects of linked (likely immune) loci, or general benefits of heterozygosity (Hansson and Westerberg 2002). While it is clear that the immune system plays a role, it is not yet clear how host resistance intersects with genome-wide variation. It is likely that increased diversity at immune genes contributes in some way to our pattern, but the fact that we examined > 550,000 loci with low linkage disequilibrium suggests a genome-wide advantage of heterozygosity that is not mediated through improved condition (e.g., differences in costly anti-parasite behaviors (Luong et al. 2007)). Future analyses on candidate genes, as well as transcriptomics or experimental challenges, can help to disentangle the role of genome-wide heterozygosity vs. single-locus effects in coccidia infections. Further, our work underscores the importance of preserving genetic diversity in wild populations.

Coccidia infections can be dynamic, yet in this study, the parasite infection data came from a single timepoint, so it is surprising that the per-locus patterns were detectable, and perhaps why, in addition to sample size, the individual-level ones were marginal. The classification of individuals into uninfected, low, and high oocyst shedding groups, as was done for the per-locus analyses, might also better reflect individuals’ general infection propensities than the single oocyst count used in the individual-level intensity and abundance analyses. Given that the lifecycle of coccidia includes both intracellular and extracellular stages,
host immunological defenses to coccidia are similarly multi-faceted and include both multiple innate and adaptive components of immunity (Yun et al. 2000). Thus, the relationships between per-locus heterozygosity and resistance to coccidia and the candidate genes associated with some of these SNPs may be informative for other intracellular and extracellular parasites that involve similar immune pathways. In *P. maniculatus*, heterozygosity-coccidia interactions are far easier to study and can provide insight into how host genetic diversity may affect rodent-borne zoonotic diseases like Sin Nombre virus (Barbour 2017). While bats carry some of the most deadly zoonotic viral infections to humans (Brook and Dobson 2015), by far more zoonoses have been identified as coming from rodents (Han et al. 2016). Moreover, it is likely that rodents will also be reservoirs of future zoonoses we have yet to encounter (Han et al. 2015), so it is critical to better understand how both neutral and functional genetic diversity contribute to disease resistance.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10682-022-10175-8.

Acknowledgements This work was performed at the Claremont Colleges’ Robert J. Bernard Biological Field Station. We thank Finley Melnikoff for assistance with the field sampling and fecal egg counts. Erin Alexander, Tamara Mehta, Moira McCarthy, and Binita Pandya also assisted with fecal egg counts. Emma Garval helped with preliminary mixed modeling analyses. Funding was provided by the W.M. Keck Science Department.

Authors contribution S.A. Budischak and F. Finseth developed the project and wrote the manuscript. S.A. Budischak carried out the field sampling, parasite counts, and performed analyses. F. Finseth prepared the samples for sequencing, performed the bioinformatics, and analyzed the data. S. Halvorsen analyzed the data and provided feedback on the manuscript.

Data availability The infection and SNP data generated during this study are available on Dryad (https://doi.org/10.5061/dryad.k0p2ngfb4).

Declarations

Conflict of interest The authors declare that they have no relevant financial or non-conflicts of interest.

Research involving human and animals All animal research protocols were approved by the W.M. Keck Science Department IACUC (Protocol # 18–025) and the California Department of Fish and Wildlife (Scientific Collection Permit SC-182880008).

References

Aberson CL (2019) Power analyses for common designs (Power to the People). Version 0.2.0
Acevedo-Whitehouse K, Petetti L, Duignan P, Castinel A (2009) Hookworm infection, anaemia and genetic variability of the New Zealand sea lion. Proc Biol Sci 276:3523–3529
Acevedo-Whitehouse K, Spraker TR, Lyons E et al (2006) Contrasting effects of heterozygosity on survival and hookworm resistance in California sea lion pups. Mol Ecol 15:1973–1982
Acevedo-Whitehouse K, Vicente J, Gortazar C et al (2005) Genetic resistance to bovine tuberculosis in the Iberian wild boar. Mol Ecol 14:3209–3217. https://doi.org/10.1111/j.1365-294X.2005.02656.x
Ahbara AM, Rouatbi M, Gharbi M et al (2021) Genome-wide insights on gastrointestinal nematode resistance in autochthonous Tunisian sheep. Sci Rep 11:9250. https://doi.org/10.1038/s41598-021-88501-3
Alcala-Canto Y, Ibarra-Velarde F (2008) Cytokine gene expression and NF-κB activation following infection of intestinal epithelial cells with *Eimeria bovis* or *Eimeria alabamensis* in vitro. Parasite Immunol 30:175–179. https://doi.org/10.1111/j.1365-3024.2007.01015.x
André A, Millien V, Galan M et al (2017) Effects of parasite and historic driven selection on the diversity and structure of a MHC-II gene in a small mammal species (*Peromyscus leucopus*) undergoing range expansion. Evol Ecol 31:785–801. https://doi.org/10.1007/s10682-017-9898-z

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. Version 0.11.3. https://www.qubebush.org/resources/fastqc

Banks SC, Scheele BC, Macris A et al (2020) Chytrid fungus infection in alpine tree frogs is associated with individual heterozygosity and population isolation but not population-genetic diversity. Front Biogeogr 12. https://doi.org/10.21425/F5FBG43875

Barbour AG (2017) Infection resistance and tolerance in *Peromyscus spp.*, natural reservoirs of microbes that are virulent for humans. Semin Cell Dev Biol 61:115–122. https://doi.org/10.1016/j.semcdb.2016.07.002

BCM-HGSC (2016) Peromyscus Genome Project. In: Bayl. Coll. Med. Hum. Genome Seq. Cent. https://www.hgsc.bcm.edu/other-mammals/peromyscus-genome-project. Accessed 11 Jan 2021

Bedford NL, Hoekstra HE (2015) *Peromyscus* mice as a model for studying natural variation. eLife 4:e06613. https://doi.org/10.7554/eLife.06613

Beldomenico PM, Begon M (2010) Disease spread, susceptibility and infection intensity: vicious circles? Trends Ecol Evol 25:21–27

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Beraldi D, McRae AF, Gratten J et al (2007) Quantitative trait loci (QTL) mapping of resistance to strongyles and coccidia in the free-living Soay sheep (*Ovis aries*). Int J Parasitol 37:121–129. https://doi.org/10.1016/j.ijpara.2006.09.007

Blanchong JA, Robinson SJ, Samuel MD, Foster JT (2016) Application of genetics and genomics to wildlife epidemiology. J Wildl Manag 80:593–608. https://doi.org/10.1002/jwmg.1064

Bolger AM, Lohse M, Usadel B (2014) Trimmmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boulton K, Nolan MJ, Wu Z et al (2018) Dissecting the genomic architecture of resistance to *Eimeria maxima* parasitism in the chicken. Front Genet 9.https://doi.org/10.3389/fgen.2018.00528

Bowman DD (2009) Georgis’ parasitology for veterinarians, 9th edn. Saunders Elsevier, St. Louis, MO

Brambilla A, Biebach I, Bassano B et al (2015) Direct and indirect causal effects of heterozygosity on fitness-related traits in Alpine ibex. Proc R Soc Lond B Biol Sci 282:20141873. https://doi.org/10.1098/rspb.2014.1873

Broad Institute (2019) Picard toolkit. Version 2.25.5. https://github.com/broadinstitute/picard

Brook CE, Dobson AP (2015) Bats as ‘special’ reservoirs for emerging zoonotic pathogens. Trends Microbiol 23:172–180. https://doi.org/10.1016/j.tim.2014.12.004

Budischak SA, Halvorsen S, Finseth F (2022) Data from: genomic heterozygosity is associated with parasite abundance, but the effects are not mediated by host condition. Dryad. https://doi.org/10.5061/dryad.k0p2ngfb4

Champely S, Ejsstrin C, Dakgaard P et al (2020) Basic functions for power analysis. Version 1.3–0. https://github.com/heliosdrm/pwr

Chapman HD, Barta JR, Blake D et al (2013) A selective review of advances in coccidiosis research. In: Advances in Parasitology. Elsevier, pp 93–171

Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. Evolution 53:1259–1267. https://doi.org/10.2307/2640828

Cunningham AA, Daszak P, Wood JLN (2017) One Health, emerging infectious diseases and wildlife: two decades of progress? Philos Trans R Soc B Biol Sci 372:20160167. https://doi.org/10.1098/rstb.2016.0167

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek P, Bonfield JK, Liddle J et al (2021) Twelve years of SAMtools and BCFtools. GigaScience 10:gia008. https://doi.org/10.11961/gigascience.gia008

Daszak P, Cunningham AA, Hyatt AD (2000) Emerging infectious diseases of wildlife– threats to biodiversity and human health. Science 287:443–449. https://doi.org/10.1126/science.287.5452.443

DeCandia AL, Dobson AP, vonHoldt BM (2018) Toward an integrative molecular approach to wildlife disease. Conserv Biol 32:798–807. https://doi.org/10.1111/cobi.13083

DeCandia AL, Schrom EC, Brandell EE et al (2021) Sarcoptic mange severity is associated with reduced genomic variation and evidence of selection in Yellowstone National Park wolves (*Canis lupus*). Evol Appl 14:429–445. https://doi.org/10.1111/eva.13127
Ezenwa VO, Budischak SA, Buss P et al (2021) Natural resistance to worms exacerbates bovine tuberculosis severity independently of worm coinfection. Proc Natl Acad Sci 118:e2015080118. https://doi.org/10.1073/pnas.2015080118

Fisher MC, Garner TWJ (2020) Chytrid fungi and global amphibian declines. Nat Rev Microbiol 18:332–343. https://doi.org/10.1038/s41579-020-0335-x

Fraley C, Raftery A, Scrucca L et al (2021) Gaussian mixture modelling for model-based clustering, classification, and density estimation. mclust. Version 5.4.9. URL https://mclust-org.github.io/mclust/

French SS, Moore MC, Demas GE (2009) Ecological immunology: The organism in context. Integr Comp Biol 49:246–253. https://doi.org/10.1093/icb/icp032

Frutos R, Serra-Cobo J, Pinault L et al (2021) Emergence of bat-related betacoronaviruses: hazard and risks. Front Microbiol 12:437. https://doi.org/10.3389/fmicb.2021.591535

Fuller CA (1996) Variable levels of immunity to experimental Eimeria arizonensis infections in natural, seminatural, and laboratory populations of deer mice (Peromyscus maniculatus). Can J Zool-Rev Can Zool 74:750–757. https://doi.org/10.1139/z96-085

Garrison E, Kronenberg ZN, Dawson ET et al (2021) Vcflib and tools for processing the VCF variant call format. Version 1.0.1. Cold Spring Harbor Laboratory. URL https://github.com/vcflib/vcflib

Gorsich EE, Ezenwa VO, Jolles AE (2014) Nematode–coccidia parasite co-infections in African buffalo: Epidemiology and associations with host condition and pregnancy. Int J Parasitol Parasites Wildl 3:124–134. https://doi.org/10.1016/j.ippaw.2014.05.003

Goudet J, Jombart T (2020) hierfstat: estimation and tests of hierarchical F-statistics. Version 0.5–7. https://CRAN.R-project.org/package=hierfstat

Hamzić E, Buitenhuis B, Hérault F et al (2015) Genome-wide association study and biological pathway analysis of the Eimeria maxima response in broilers. Genet Sel Evol 47:91. https://doi.org/10.1186/s12711-015-0170-0

Han BA, Kramer AM, Drake JM (2016) Global patterns of zoonotic disease in mammals. Trends Parasitol 32:565–577. https://doi.org/10.1016/j.pt.2016.04.007

Han BA, Schmidt JP, Bowden SE, Drake JM (2015) Rodent reservoirs of future zoonotic diseases. Proc Natl Acad Sci 112:7039–7044. https://doi.org/10.1073/pnas.1501598112

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. Mol Ecol 11:2467–2474. https://doi.org/10.1046/j.1365-294X.2002.01644.x

Harris SE, O’Neill RJ, Munshi-South J (2015) Transcriptome resources for the white-footed mouse (Peromyscus leucopus): new genomic tools for investigating ecologically divergent urban and rural populations. Mol Ecol Resour 15:382–394. https://doi.org/10.1111/1755-0998.12301

Jin H, Haicheng Y, Caiyun Z et al (2020) The expression of NF-kB signaling pathway was inhibited by silencing TGF-b4 in chicken IECs infected with E. tenella. Braz J Poult Sci 22:eRBCA-2020–1338. https://doi.org/10.1590/1806-9061-2020-1338

Jones KE, Patel NG, Levy MA et al (2008) Global trends in emerging infectious diseases. Nature 451:990-U4. https://doi.org/10.1038/nature06536

Kenney-Hunt J, Lewandowski A, Glenn TC et al (2014) A genetic map of Peromyscus with chromosomal assignment of linkage groups (a Peromyscus genetic map). Mamm Genome 25:160–179. https://doi.org/10.1007/s00335-014-9500-8

Kim E-S, Hong YH, Lillehoj HS (2010) Genetic effects analysis of myeloid leukemia factor 2 and T cell receptor-β on resistance to coccidiosis in chickens. Poult Sci 89:20–27. https://doi.org/10.3382/ps.2009-00351

Kim E-S, Hong YH, Min W, Lillehoj HS (2006) Fine-mapping of coccidial-resistant quantitative trait loci in chickens. Poult Sci 85:2028–2030. https://doi.org/10.1093/ps/85.11.2028

Klesius PH, Hinds SE (1979) Strain-dependent differences in murine susceptibility to coccidia. Infect Immun 26:1111–1115. https://doi.org/10.1128/iai.26.3.1111-1115.1979

Knaps BJ, Grünwald NJ (2017) vcf-r: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour 17:44–53. https://doi.org/10.1111/1755-0998.12549

Knights AJ, Yang L, Shah M et al (2020) Krüppel-like factor 3 (KLF3) suppresses NF-κB–driven inflammation in mice. J Biol Chem 295:6080–6091. https://doi.org/10.1074/jbc.RA120.013114

Knowles SCL, Fenton A, Petchey OJ et al (2013) Stability of within-host - parasite communities in a wild mammal system. Proc R Soc B-Biol Sci 280.https://doi.org/10.1098/rspb.2013.0598
Kolberg L, Raudvere U, Kuzmin I et al (2020) gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Research 9:709. https://doi.org/10.12688/f1000research.24956.1

Kubakka J, Podmokla E, Korb J, Dubiec A (2020) Heterozygosity and fitness in a threatened songbird: blood parasite infection is explained by single-locus but not genome-wide effects. J Ornithol 161:803–817. https://doi.org/10.1007/s10336-020-01753-0

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]

Li H, Handsaker B, Wysoker A et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Long AD, Baldwin-Brown J, Tao Y et al (2019) The genome of Peromyscus leucopus, natural host for Lyme disease and other emerging infections. Sci Adv 5:eaaaw6441. https://doi.org/10.1126/sciadv.aaw6441

Luikart G, Pilgrim K, Visty J et al (2008) Candidate gene microsatellite variation is associated with parasitism in wild bighorn sheep. Biol Lett 4:228–231. https://doi.org/10.1098/rsbl.2007.0633

Luong LT, Heath BD, Polak M (2007) Host inbreeding increases susceptibility to ectoparasitism. J Evol Biol 20:79–86. https://doi.org/10.1111/j.1420-9101.2006.01226.x

MAFF (1980) Manual of veterinary parasitological techniques. Ministry of Agriculture, Fisheries and Food, London

Martin AM, Cassirer EF, Waits LP et al (2021) Genomic association with pathogen carriage in bighorn sheep (Ovis canadensis). Ecol Evol 11:2488–2502. https://doi.org/10.1002/ece3.7159

McIntire JJ, Umetsu SE, Macaubas C et al (2003) Hepatitis A virus link to atopic disease. Nature 425:576–576. https://doi.org/10.1038/425576a

Meagher S (1999) Genetic diversity and Capillaria hepatica (Nematoda) prevalence in Michigan deer mouse populations. Evolution 53:1318–1324. https://doi.org/10.1111/j.1558-5646.1999.tb04547.x

Mitchell J, Vitikainen EIK, Wells DA et al (2017) Heterozygosity but not inbreeding coefficient predicts parasite burdens in the banded mongoose. J Zool 302:32–39. https://doi.org/10.1111/jzo.12424

Murray GG, Woolhouse ME, Tapio M et al (2013) Genetic susceptibility to infectious disease in East African Shorthorn Zebu: a genome-wide analysis of the effect of heterozygosity and exotic introgression. BMC Evol Biol 13:246. https://doi.org/10.1186/1471-2148-13-246

Oppelt C, Starkloff A, Rausch P et al (2010) Major histocompatibility complex variation and age-specific endoparasite load in subadult European rabbits. Mol Ecol 19:4155–4167. https://doi.org/10.1111/j.1365-294X.2010.04766.x

Pruett CL, Winker K (2008) The effects of sample size on population genetic diversity estimates in song sparrows Melospiza melodia. J Avian Biol 39:252–256

Psifidi A, Banos G, Matika O et al (2016) Genome-wide association studies of immune, disease and production traits in indigenous chicken ecotypes. Genet Sel Evol 48:74. https://doi.org/10.1186/s12711-016-0252-7

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Råberg L, Sim D, Read AF (2007) Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. Science 318:812–814. https://doi.org/10.1126/science.1148526

Reid JM, Arcese P, Keller LF et al (2007) Inbreeding effects on immune response in free-living song sparrows (Melospiza melodia). Proc Biol Sci 274:697–706

Rödel HG, Oppelt C, Starkloff A et al (2020) Within-litter covariance of allele-specific MHC heterozygosity, cocciidid endoparasite load and growth is modulated by sibling differences in starting mass. Oecologia 194:345–357. https://doi.org/10.1007/s00442-020-04764-z

Ruiz-López MJ, Monello RJ, Gompper ME, Eggert LS (2012) The effect and relative importance of neutral genetic diversity for predicting parasitism varies across parasite taxa. PLoS ONE 7:e45404. https://doi.org/10.1371/journal.pone.0045404

Sánchez CA, Becker DJ, Teitelbaum CS et al (2018) On the relationship between body condition and parasite infection in wildlife: a review and meta-analysis. Ecol Lett 21:1869–1884. https://doi.org/10.1111/ele.13160
