

RESEARCH

Open Access



Host immunogenetic variation and gut microbiome functionality in a wild vertebrate population

Chuen Zhang Lee^{1,2*}, Sarah F. Worsley¹, Charli S. Davies¹, Jan Komdeur³, Falk Hildebrand^{4,5}, Hannah L. Dugdale³ and David S. Richardson^{1,6*}

Abstract

Background The gut microbiome (GM) –important for host health and survival– is partially shaped by host immunogenetics. However, to date, no study has investigated the influence of host Major Histocompatibility Complex (MHC) genes on gut microbiome functionality in a wild population. Here we use a natural population of the Seychelles warbler (*Acrocephalus sechellensis*) to assess the effects of MHC genes on GM taxonomy and functionality using shotgun metagenomics.

Results Our results show that taxonomic GM composition was associated with MHC-II diversity and the presence of one specific MHC-I allele (*Ase-ua 7*). Specifically, MHC-II diversity was associated with decreased *Lactococcus lactis* and increased *Staphylococcus lloydi* abundance, while *Ase-ua 7* was linked to reduced *Enterococcus casseliflavus* and *Gordonia* sp OPL2 but increased *Escherichia coli* and *Vulcaniibacterium thermophilum*. These taxonomic changes may reflect differences in MHC-mediated microbial recognition. In contrast, functional GM composition was significantly associated with increasing individual MHC-I diversity but not MHC-II diversity. In particular, increasing MHC-I diversity was associated with an increased prevalence of microbial defence genes but a reduced prevalence of microbial metabolism genes. Analysis also revealed that functional GM networks were more fragmented in high compared to low MHC-I diversity hosts.

Conclusion These results suggest that MHC variation (particularly at MHC-I) plays an important role in shaping both the taxonomy and function of the GM in wild vertebrates. In the Seychelles warbler, this results in trade-offs whereby there is an increase in microbial defence and a reduction in GM metabolic potential in individuals with higher MHC-I diversity. Thus, this work sheds light on the possible costs and benefits of maintaining a healthy microbiome, which is essential for understanding how the GM and immune system co-evolve.

Keywords *Acrocephalus sechellensis*, Metagenomics, Gut microbiome, Major histocompatibility complex, Wild population

*Correspondence:

Chuen Zhang Lee
chuen.lee@uea.ac.uk
David S. Richardson
david.richardson@uea.ac.uk

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

The vertebrate gut microbiome (GM), a complex ecosystem of microorganisms inhabiting the gastrointestinal tract, is increasingly recognised as a critical determinant of host health and fitness [118]. However, the composition and function of the GM exhibit extensive variability across individuals, particularly in natural populations [26, 84, 109]. This variation has been attributed to a range of factors, such as diet, age, sex, location and host genetic variation [21, 111, 119].

A growing body of evidence links GM characteristics to host immunogenetic variation [23, 117]. Immune genes influence the immune system's ability to recognise, tolerate, or eliminate microbial populations [20, 65]. Therefore, the immune system must maintain a balance – tolerating beneficial microbes while combating pathogens – to optimise host health [31, 97]. Furthermore, the GM also appears to play a role in the immune defences of the host, with GM dysbiosis (an imbalance in the composition of microbes) resulting in a reduction of host immune function, emphasising the interconnected nature of immune health and GM stability [54, 83].

The major histocompatibility complex (MHC) is a family of immune genes, forming part of the vertebrate acquired immune system [77]. These genes encode cell-surface glycoprotein receptor molecules that bind to antigens before presenting them to T lymphocytes and B cell receptors, which trigger an immune or tolerogenic response [5, 90]. The MHC has two main classes, MHC-I and MHC-II, based on the encoded receptors presenting intracellular or extracellular antigens, respectively [89, 90]. The role of the MHC in combating pathogens has been well-studied [49, 73], with the extraordinarily high polymorphism of MHC genes observed in natural populations thought to be driven by pathogen-mediated selection mechanisms and sexual selection [94]. Individual MHC variation determines the range of microbial antigens recognised by the immune system [6], and is associated with variation in commensal gut microbial communities [21, 53, 92, 99]. Thus, different MHC genotypes could shape individual GM variation by initiating immune responses to potentially pathogenic microbes while maintaining beneficial species [91, 92].

Previous studies examining the impact of MHC variation on the GM in wild animals using 16S metabarcoding, have reported mixed findings. Several have found that increased MHC diversity is associated with decreased microbiome diversity [6, 58, 101] but others associated it with increased [41] or unchanged GM diversity [21, 68]. Similarly, some studies have observed shifts in taxonomic composition with MHC diversity [6, 41, 68], while others have not [21, 29, 30, 101]. Additionally, the presence/absence of specific MHC alleles (rather than the overall

diversity of alleles) has been found to be correlated with GM taxonomic composition [6, 21].

The functional composition of the GM – represented through microbial genes – could provide a more direct representation of host-microbe interactions [112]. However, the consequences of MHC variation for GM functionality have remained underexplored so far [31]. Many microbes share genes and, consequently, have similar functional roles [64]. Therefore, changes in microbial taxa do not always result in altered GM function – i.e. there is functional redundancy [64, 112]. Functional redundancy refers to the ecological concept that multiple species within an ecosystem can perform similar roles, encoding the same gene and/or different genes with the same function, buffering against species loss [64, 112]. Studying functionality is important for understanding if and how host genetic variation interacts with the GM to influence host fitness and evolutionary trajectories [112]. Most studies on MHC and microbial functionality rely on 16S metabarcoding markers and infer function based on known microbial taxa-function association [32, 33, 96, 103]. However, in less studied systems, such as wild animals, functional inferences from 16S metabarcoding markers may lead to misassignments due to the lack of representation of the specific microbes observed in existing databases [96, 100].

In humans and transgenic captive mice (*Mus musculus*), MHC haplotype is associated with GM functional composition [3, 8]. However, captive/domesticated populations often harbour greatly reduced genetic variation and microbial diversity compared to natural populations [106], thus, these results may not be transferable to wild systems. To our knowledge, the only pioneering study of host MHC and GM function in a wild animal so far used 16S functional inferences [68], which are likely to lead to inaccuracies [96, 100]. Shotgun metagenomics or transcriptomic approaches are needed to accurately determine gut microbiome function in response to host MHC variation in wild animal populations.

Here, we investigate the relationship between MHC and GM variation in a population of Seychelles warblers (*Acrocephalus sechellensis*). Despite reduced neutral genetic variation due to past population bottlenecks [95], the Seychelles warbler has maintained variation (albeit reduced) at MHC-I and MHC-II loci [21, 39, 88]. Furthermore, one specific MHC allele (*Ase-ua 4*) and MHC-I diversity overall have been positively correlated with survival and reproductive success [12, 87]. A previous 16S-based analysis of this population has demonstrated that MHC alleles are associated with changes in bacterial GM taxonomic diversity and composition [21]. An analysis of the fungal mycobiome also reported changes in species diversity and composition associated with

MHC alleles and MHC-I diversity, respectively [110]. Since these studies, we have greatly expanded our sample size, identified key environmental control variables, and conducted shotgun sequencing for metagenomic analysis [59].

We leverage a powerful combination of both 16S rRNA metabarcoding (larger sample size) and shotgun metagenomics to provide a comprehensive and high-resolution assessment of the association between host MHC variation and both taxonomic and, importantly, functional components of the bacterial GM in adult Seychelles warblers. First, we test if GM taxonomic diversity and composition correlate with MHC-I and MHC-II diversity or alleles. Next, we test if GM functional diversity and functional composition are associated with this MHC variation. Finally, we assess the role of functional redundancy in preserving the functionality of the GM despite changes in host MHC variation.

Methods

Study system

The population of insectivorous Seychelles warblers on Cousin Island (29 ha; 04° 20' S, 55° 40' E) has been extensively monitored since 1985 [13, 51]. Two field seasons are undertaken annually from ca. January to March (minor) and June to September (major). Each season, as many individuals as possible are caught in the nest (chicks) or using mist nets and sampled (see below). New individuals are marked with a British Trust for Ornithology (BTO) metal ring and a unique combination of three colour rings, allowing them to be monitored throughout their lives. Almost every bird (>96%) on Cousin has been marked this way since 1997 [81, 85]. Age is calculated based on fledge or hatch dates, or eye colour at first catch [50]. This population includes ca. 320 individuals in approximately 115 territories [38, 52].

Sample collection

Faecal sample collection, storage, DNA extraction, library preparation and sequencing were conducted between 2017 and 2023, as part of (and described in full in) previous studies using 16S rRNA metabarcoding [21, 108, 112] and shotgun metagenomics [59]. In brief, caught birds were placed in a flat-bottom paper bag with a sterilised weigh boat under a metal grate, allowing faeces to drop to the weigh boat, while minimising contact with the birds' surface. Faecal matter was transferred into a sterile microcentrifuge tube containing 1 mL of absolute ethanol, stored at 4 °C during fieldwork and then at −80 °C for long-term storage at the University of East Anglia (UEA). The time-of-day (minutes after sunrise; 06:00 AM) of sampling was recorded. Each season, control samples were also taken from the hands of fieldworkers using a

sterile swab and stored in the same manner. A small (ca 25 µL) blood sample was also collected from each bird via brachial venepuncture and stored in 0.7 ml of absolute ethanol at 4 °C. Samples may be collected from the same individual in different field seasons; thus, the identity of each individual sampled (Bird ID) is recorded and used to control for these repeated measures in statistical analyses. In total, 149 unique bird IDs were sampled, of which 72 had repeated samples.

Molecular genotyping

Total genomic DNA was extracted from blood samples using the DNeasy Blood and Tissue kit (Qiagen, Crawley, UK) according to the manufacturer's protocol. All Individuals were genotyped using up to 30 polymorphic microsatellite loci and 3 sexing markers (following [37, 86, 93]) as part of the ongoing determination of parentage and pedigree within this population [82]. Individual genome-wide heterozygosity (Hs) at these neutral loci was calculated using *genhet* 3.1 in R 4.33 [19, 80] as per [113].

Sequencing of amplified MHC-I exon 3 and MHC-II exon 2 variants using Illumina MiSeq technology had already been undertaken for 314 warblers [21]. All confirmed variants (20 MHC-I and 14 MHC-II, hereafter termed alleles) at each of these MHC regions (which contain 4 replicated loci) [46, 88] were used to calculate individual MHC diversity. However, due to statistical power limitations, only alleles present in >5% and <95% of individuals were included in presence/absence analyses. Two MHC-I alleles (*Ase-ua1* and *Ase-ua10* were co-occurring; thus, only one of them, *Ase-ua1*, was retained for downstream analyses. Therefore, nine MHC-I alleles and three MHC-II alleles were used in the presence/absence statistical analyses. Each allele in the presence/absence analysis translates to a unique amino acid sequence with different antigen-binding properties [21]. Pairwise p-distance of MHC alleles was also calculated based on the percentage of different amino acids. The average pairwise p-distance was then calculated per individual to represent functional diversity of the MHC (i.e. MHC divergence).

Gut microbiome screening

Microbial DNA from faecal samples was extracted using the DNeasy PowerSoil Kit (Qiagen, Crawley, UK) and a modified version of the manufacturer's protocol (described in detail [21]). Samples were randomised across extractions to minimise batch effects.

Faecal DNA samples were submitted for 16S rRNA amplicon sequencing. Amplicon sequencing libraries were generated using the V4 primers 515F (5'TGCCAG CMGCCGCGGTAA3') and 806R (5'GGACTACHVGGG TWTCTAAT3'). Libraries were sequenced across seven

batches using 2×250bp, paired-end sequencing on an Illumina MiSeq Platform (see [21, 108, 112]). Control samples were also extracted, library prepped and sequenced the same way ($n=21$ hand controls, 15 negative controls, and 10 positive ZymoBIOMICS Microbial Community Standard (D6300) controls).

Faecal DNA samples underwent library preparation using the LITE protocol [75] and were sequenced using 2×150 bp, paired-end shotgun metagenomic sequencing in two runs on an Illumina NovaSeq X platform (see [59]). Hand controls ($n=6$) and positive controls ($n=3$, two ZymoBIOMICS Microbial Community Standard (D6300), and one ZymoBIOMICS Faecal Reference with TruMatrix™ Technology (D6323)) were also prepped and sequenced as part of the metagenomic samples sequencing.

Bioinformatics

Read processing for 16S metabarcoding was performed as previously described [108, 112]. Briefly, 16S rRNA reads were processed using QIIME2 2019.10, reads were truncated, filtered, and classified into amplicon sequencing variants (ASV) using DADA2 [15]. ASVs were then taxonomically assigned using the naïve-Bayes classifier on the SILVA 132 reference database for 16S rRNA gene sequences [7]. These ASVs were then imported into R 4.3.3 using *phyloseq* 1.46.0 [15, 66], then filtered to remove non-bacterial sequences, reads unassigned to phylum level, and potential contaminants (based on hand controls). Rarefaction curves were constructed with *iNEXT* version 3.0.1 with the default 50 bootstrap replications [18], reaching an asymptote at 8000 reads, indicating sample completeness (Figure S1A). In addition, 27 faecal samples with < 8000 reads were removed and ASVs with < 50 reads across all samples were also removed.

Shotgun metagenomic sequence processing was performed using *MATAFILER* [43] as previously described [59]. Host reads were removed by mapping reads with *Kraken2* 2.1.3 to the Seychelles warbler genome (unpublished; complete BUSCO=96.0% with a total length=1,081,018,985 bp), followed by read quality filtering using *sdm* 2.14 beta; minimum sequence length of 50, minimum average quality of 27 [44, 107], an average of 21% (± 0.07 SE) reads were removed. After trimming, two samples and five hand controls were removed because they did not have enough reads for metagenome assembly. An average of 20,481,040 (SD=13,718,305) paired-end reads per sample were retained for de novo metagenome assembly using *MEGAHIT* 1.2.9 with default parameters and kmer-list of 25,43,67,87,111,131 [61]. Using the resulting assemblies, genes were predicted using Prodigal 2.6.3 [47] and clustered into gene catalogues (95% identity). Genes were functionally annotated

using *eggNOGmapper* 2.1.12 with default parameters and the eggNOG database 4 [16, 79]. Functional categories were also assigned to each functional annotation based on the cluster of gene orthologs (COG) database [98]. Metaphlan4 assignments were used to taxonomically assign shotgun sequencing reads. Rarefaction curves were constructed for metagenomics taxonomy and functional reads with *iNEXT* version 3.0.1 with the default 50 bootstrap replications [18] and showed an asymptote and sample completeness at 5,500 and 100,000 reads, respectively (Figure S1B-C).

Statistical analysis

Adult warblers with both microbiome and MHC data were analysed. For 16S analysis, 253 samples from 149 individuals were included in this study. Of these, 99 samples from 57 adult individuals also had GM shotgun metagenomic data. Individuals carried a mean of 5.13 (SE: 0.088, range 2–7) MHC-I alleles and 2.88 (SE: 0.060, range 1–5) MHC-II alleles. Due to the low number of samples for which we had shotgun metagenomic data, we had to limit the number of predictor variables (i.e. < 9) in each model to avoid overfitting and unreliable estimates. Thus, for MHC diversity and MHC divergence models, all control variables (see below) were included, but we first used the 16S metabarcoding dataset to shortlist which genetic metrics (including specific MHC alleles) should be included in the shotgun metagenomic models. Unless stated otherwise, all statistical analyses were conducted in R 4.3.3 in R Studio 2024.12.0 + 467 [78, 80] and linear mixed effect (LMMs) and generalised linear mixed effect models (GLMMs) were constructed using *lme4* 1.1–35.5 [1].

GM diversity

16S rRNA metabarcoding diversity

Reads were rarefied to 8,000 reads with the *rarefy_even_depth* function in *vegan* 2.6.6 (Oksanen [71]) – the point at which the number of ASVs identified reached an asymptote in rarefaction curves (Figure S1A)—before calculation of alpha diversity metrics. Both ASV richness and Shannon diversity were calculated for each sample using *phyloseq* 1.46.0 [66].

A GLMM with a negative binomial distribution was constructed with ASV richness as a response variable. An LMM with a Gaussian distribution was used to model Shannon diversity. Hereafter, all 16S models were tested with the same set of variables (described below) with either MHC alleles or MHC diversity as the response unless stated otherwise. MHC-I and MHC-II diversity (i.e. the number of alleles per individual) were included as predictors, along with genome-wide heterozygosity, age, season, sample year, sex, sample days at 4°C, and

time of day sampled, as fixed-term control variables and bird ID as a random effect. Quadratic effects of MHC-I diversity and MHC-II diversity were included to test if an intermediate number of alleles influenced GM characteristics, but were dropped sequentially if not significant (least significant first) to allow interpretation of the main terms. Standardised effect sizes of each fixed effect were determined using partial R^2 with *r2beta()* function in *r2glmm* version 0.1.3 [48].

To determine if specific MHC alleles were shaping the GM, a second model was constructed using the presence/absence of MHC-I (*Ase-ua 1*, *Ase-ua 3*, *Ase-ua 4*, *Ase-ua 5*, *Ase-ua 6*, *Ase-ua 7*, *Ase-ua 8*, *Ase-ua 9*, *Ase-ua 11*) and MHC-II alleles (*Ase-dab 3*, *Ase-dab 4*, *Ase-dab 5*) as predictors in place of MHC diversity.

Metagenomic taxonomic diversity

Metaphlan4 assignments were rarefied to 5,500 reads (Figure S1B) – prior to alpha diversity analysis. A GLMM with a negative binomial distribution was then used to model species richness, and an LMM was used to model Shannon diversity. All metagenomics analyses were performed with the same structure (described below, i.e. MHC diversity models included all terms, while MHC presence/absence models only included genetic variables that were identified as significant in the corresponding 16S analysis).

A second model with the presence/absence of specific MHC alleles (identified as significant in the 16S metabarcoding model above) was constructed.

Metagenomic functional diversity

Functional gene annotations (determined using eggNOG mapper described above) were rarefied to 100,000 reads (Figure S1C) before functional alpha diversity analysis. Scaled exponentially transformed functional gene richness and exponentially transformed functional Shannon diversity were modelled separately with LMMs, with either the MHC diversity or the presence/absence of MHC alleles, along with genome-wide heterozygosity and environmental control variables (as described for 16S analyses above). An additional model including the presence/absence of all MHC alleles was constructed, as functional GM characteristics may respond differently to MHC variation than GM taxonomy. However, because the metagenomic dataset ($N=99$ from 57 individuals) contains fewer samples than the 16S metabarcoding ($N=253$ from 149 individuals), incorporating all alleles risks over-parameterising the model. Therefore, a reduced model was also constructed that only included the specific MHC alleles identified as significant in the 16S metabarcoding analysis to evaluate the robustness of the results.

GM composition

16S rRNA metabarcoding composition

Unrarefied reads were used. Rare ASVs (<5% prevalence) were removed prior to analysis, and a centred log ratio (CLR) transformation was applied to the remaining ASV abundances using *microbiome* 1.24.0 [60]. Pairwise Aitchison distances (i.e. composition differences) among GM samples were then modelled via a PERMANOVA using the *adonis2()* function with *by*="margin" in *vegan* 2.6.6 with 9999 permutations. A blocking effect of bird ID was included to account for repeated sampling (Oksanen [71]). The first PERMANOVA model included MHC diversity, genome-wide heterozygosity, age, season, sample year, sex, days at 4°C and time of day as predictors. The second PERMANOVA model had the presence/absence of individual MHC alleles instead of MHC diversity. Heterogeneity of group dispersions among significant MHC alleles, seasons, and sampling years was tested using *betadisper()* and *permutest()* functions in *vegan*. Both these and all subsequent GM composition models were set up in the same way, tested for heterogeneity of dispersion, and visualised with a PCA generated in *phyloseq* 1.46.0 [66] unless stated otherwise.

Metagenomic taxonomic composition

Rare species were removed (<5% prevalence), the remaining unrarefied reads were CLR transformed and used in a PERMANOVA to identify differences in taxonomic composition associated with MHC variation (as described for 16S analysis above). For the MHC alleles model, only genetic predictors identified as significant in the 16S rRNA metabarcoding composition analysis were included, along with all control variables (as described for 16S analyses above).

Metagenomic functional composition

Rare functional genes (<5% prevalence) were removed, and the remaining unrarefied reads were CLR transformed and used in a PERMANOVA to test for differences in functional composition linked to MHC variation (as described for metagenomic taxonomic composition analyses above). For the MHC alleles model, all alleles were included, but that could overparameterize the models; thus, an additional model was created with only genetic predictors identified as significant in the 16S rRNA metabarcoding composition analysis.

Differential abundance analyses

Differential abundance of metagenomic taxonomic species

Differential abundance tests were carried out using *ALDEx2* 1.34.0 [28]. Only common species (>10% prevalence and >0.001% abundance resulting in 49 metagenomic identified species) were included. Abundances

were CLR transformed as part of the *ALDEx2* method [28]. Genome-wide heterozygosity, MHC-I and MHC-II diversity, as well as significant variables identified in the metagenomic taxonomic composition analysis were included as predictors.

Differential abundance of metagenomic functional genes

Abundances of common functional genes (>50% prevalence, >0.1% abundance resulting in 94 eggNOG members) were CLR-transformed using *ALDEx2* 1.34.0 [28] and included in this analysis. Predictors were included as above but based on significant variables in metagenomic functional composition analysis. KEGG orthologs and pathways were assigned manually to significant functional genes by referring to the KEGG website (<https://www.kegg.jp/kegg/kegg2.html>). The relative abundance (determined using CLR transformation) of each significant functional category was tested with a GLMM to directly test for trade-offs between categories. Genome-wide heterozygosity, MHC-I diversity, MHC-II diversity, as well as significant variables identified in the metagenomic functional composition analysis were included as predictors. Bird ID was included as a random effect.

Network analysis

Network analysis of metagenomic taxonomic species

Networks of metagenomic taxonomic species were constructed with SParse Inverse Covariance Estimation for Ecological Association Inference (SPIEC-EASI) version 1.0.7 [55]. The samples were split into two categories based on being higher or lower than the average MHC diversity in the population (see above): low (<6) and high (≥ 6) MHC-I diversity, or low (<3) and high (≥ 3) MHC-II diversity. The raw counts of common bacterial species (>10% prevalence and >0.1% abundance) were used as inputs, with SPIEC-EASI applying a CLR transformation. Common species were used to capture relevant, stable GM species and minimise the influence of rare taxa [25]. The number of nodes (species), the number of edges (interaction between species), the average number of connections per node, modularity and negative-to-positive ratios were calculated. The difference in edge density, modularity, and positive edge percentage between low and high MHC diversity were tested with the *net-Compare()* function in NetCoMi 1.2.0 [76]. The networks were then plotted with the *ggnet2* function in ggnet 0.1.0 [11]. Nodes were coloured by Phylum, and size was based on mean abundance per species.

Network analysis of metagenomic functional genes

Networks were constructed as described above but using common metagenomic functional genes (eggNOG genes)

(>50% prevalence, >0.1% abundance) instead of metagenomic species.

MHC divergence vs. MHC diversity

In response to reviewer's comments, we also tested whether using an alternate measure of MHC diversity, i.e. MHC divergence [22, 40], affected our interpretation of how the GM was associated with MHC variation. To do this, we added complementary models that replace MHC diversity with MHC divergence. This MHC comparison is useful because it captures different biological aspects of the MHC: MHC diversity reflects the number of MHC alleles that are present, whereas MHC divergence quantifies how functionally different those alleles are from one another [22]. In each case, model performance was compared using Akaike Information Criterion (AIC) values calculated with the *AIC()* function. An AIC difference >7 between two models is considered evidence for improved model fit [14]. Given that the results were qualitatively very similar between models using the different metrics (see results), we have added the detailed methods and results for this additional analysis in the supplementary and report the results of MHC diversity models in the main text.

Results

GM diversity

16S rRNA metabarcoding diversity

GM alpha diversity (Shannon diversity or richness) was not significantly associated with MHC-I or MHC-II diversity (Table 1A, Table S1A & S2A & S3A). The presence of the MHC-I allele *Ase-ua 11*—but no other MHC allele—was significantly positively associated with 16S richness (Table 1B, Table S1B & S2B & S3B). No alleles were associated with Shannon diversity.

Metagenomic taxonomic diversity

Taxonomic alpha diversity (Shannon diversity or richness) calculated using shotgun metagenomics data was not associated with MHC-I or MHC-II diversity (Table 1A, Table S1A & S2A & S3A), nor with the presence/absence of *Ase-ua 11* (the MHC variant identified in the 16S analysis above (Table 1B, Table S1B & S2B & S3B)).

Metagenomic functional diversity

Functional alpha diversity (Shannon diversity or richness) of gene annotations derived from shotgun metagenomics data was not associated with MHC-I or MHC-II diversity (Table 1A, Table S1A & S2A & S3A), nor with MHC alleles (Table 1B, Table S1B & S2B & S3B, Table S4).

Table 1 The relationship between gut microbiome alpha diversity (richness) and variation in host (A) Major histocompatibility complex (MHC) diversity and (B) the presence/absence of specific MHC alleles in adult Seychelles warblers. Generalised linear mixed models with a negative binomial distribution were used for 16S ASV diversity ($N=253$ samples from 149 individuals) and metagenomics taxonomy diversity ($N=99$ samples, 57 individuals), and linear mixed models were used for metagenomics functional diversity ($N=99$ samples, 57 individuals). Reference categories for categorical variables were as follows: Female (sex), winter (season), 2017 (Sample year), and absent (in all MHC alleles). Significant ($P < 0.05$) variables are shown in bold. Shannon diversity results are similar and shown in Supplementary Table S1

Predictors	16S ASV diversity				Metagenomics taxonomic diversity				Metagenomics functional diversity			
	Est	SE	z	P	Est	SE	z	P	Est	SE	t	P
A) MHC Diversity												
(Intercept)	5.37	0.32	16.82	<0.001	3.23	0.70	4.62	<0.001	1.06	0.45	2.39	0.02
Heterozygosity	−0.19	0.21	−0.93	0.35	0.43	0.43	1.02	0.31	−0.13	0.28	−0.46	0.65
MHC-I Diversity	0.04	0.03	1.26	0.21	−0.01	0.06	−0.18	0.85	0.07	0.04	1.60	0.12
MHC-II Diversity	−0.01	0.04	−0.23	0.82	−0.14	0.08	−1.83	0.07	−0.02	0.06	−0.26	0.79
Age	−0.02	0.02	−1.28	0.20	−0.04	0.03	−1.12	0.26	−0.04	0.02	−2.07	0.04
Season (summer)	0.04	0.12	0.30	0.76	0.19	0.22	0.90	0.37	−0.04	0.14	−0.26	0.79
Sex (male)	−0.22	0.09	−2.51	0.01	0.15	0.17	0.90	0.37	−0.04	0.12	−0.31	0.76
Days at 4°C	−0.01	0.09	−0.05	0.96	0.02	0.19	0.13	0.90	−0.07	0.10	−0.63	0.53
Time of day	0.02	0.08	0.19	0.85	0.26	0.18	1.46	0.15	−0.06	0.10	−0.60	0.55
Sample Year (2018)	−0.03	0.13	−0.27	0.79	0.01	0.27	0.05	0.96	0.13	0.16	0.84	0.40
Sample Year (2019)	0.14	0.16	0.87	0.39	−0.29	0.37	−0.79	0.43	−0.03	0.21	−0.13	0.90
Sample Year (2020)	0.48	0.21	2.29	0.02	−0.02	0.45	−0.05	0.96	0.08	0.26	0.31	0.76
Sample Year (2021)	0.21	0.16	1.32	0.19	0.00	0.37	−0.01	0.99	0.07	0.20	0.33	0.75
Sample Year (2022)	0.14	0.16	0.88	0.38	0.45	0.31	1.44	0.15	0.31	0.18	1.70	0.09
Sample Year (2023)					0.08	0.36	0.22	0.83	−0.05	0.21	−0.25	0.81
B) Presence/absence of MHC alleles												
(Intercept)	5.66	0.33	17.37	<0.001	3.24	0.29	11.15	<0.001	1.48	0.49	3.02	0.004
Heterozygosity	−0.09	0.21	−0.42	0.68					−0.09	0.32	−0.29	0.773
Ase-dab3	0.29	0.15	1.95	0.05					0.37	0.23	1.62	0.116
Ase-dab4	−0.27	0.15	−1.76	0.08					−0.22	0.24	−0.89	0.378
Ase-dab5	0.18	0.16	1.15	0.25					0.00	0.27	−0.01	0.995
Ase-ua1	0.12	0.18	0.64	0.52					−0.05	0.26	−0.20	0.844
Ase-ua3	−0.11	0.19	−0.60	0.55					−0.02	0.29	−0.08	0.935
Ase-ua4	−0.19	0.14	−1.37	0.17					−0.16	0.25	−0.65	0.520
Ase-ua5	−0.09	0.18	−0.52	0.60					0.02	0.29	0.08	0.938
Ase-ua6	−0.18	0.17	−1.04	0.30					−0.16	0.27	−0.60	0.555
Ase-ua7	−0.16	0.20	−0.80	0.43					−0.15	0.27	−0.54	0.593
Ase-ua8	−0.04	0.14	−0.29	0.78					0.16	0.23	0.70	0.491
Ase-ua9	−0.07	0.17	−0.39	0.70					−0.29	0.29	−1.00	0.324
Ase-ua11	0.38	0.18	2.07	0.04	0.00	0.16	0.01	0.99	0.45	0.28	1.59	0.121
Age	−0.03	0.02	−1.63	0.10	−0.06	0.03	−1.67	0.10	−0.05	0.03	−1.70	0.096
Season (summer)	0.04	0.12	0.31	0.75	0.09	0.22	0.42	0.68	0.04	0.14	0.25	0.803
Sex (male)	−0.25	0.09	−2.88	<0.001	0.12	0.17	0.75	0.45	−0.09	0.13	−0.66	0.512
Days at 4°C	−0.03	0.09	−0.33	0.75	−0.03	0.19	−0.16	0.88	−0.08	0.11	−0.72	0.476
Time of day	0.08	0.08	0.92	0.36	0.27	0.18	1.48	0.14	−0.06	0.11	−0.53	0.600
Sample Year (2018)	−0.06	0.12	−0.45	0.65	0.08	0.28	0.30	0.76	0.14	0.17	0.81	0.420
Sample Year (2019)	0.09	0.16	0.56	0.58	−0.13	0.37	−0.35	0.73	−0.13	0.22	−0.60	0.551
Sample Year (2020)	0.42	0.21	2.02	0.04	−0.08	0.46	−0.18	0.86	0.09	0.27	0.35	0.729
Sample Year (2021)	0.21	0.16	1.30	0.19	0.14	0.37	0.38	0.71	0.09	0.21	0.42	0.678
Sample Year (2022)	0.12	0.15	0.79	0.43	0.57	0.32	1.79	0.07	0.28	0.20	1.35	0.181
Sample Year (2023)					0.29	0.37	0.78	0.43	−0.15	0.24	−0.64	0.526

GM composition

16S rRNA metabarcoding composition

16S GM composition was associated with both a quadratic function of MHC-I diversity and of MHC-II diversity (Table 2A, Fig. 1A-B). It was also associated with season, sample year, days at 4°C, and time of day (Table 2A) but not with genome-wide heterozygosity, age, and sex.

16S GM composition was associated with the MHC-I variants, *Ase-ua 5*, *Ase-ua 7*, and *Ase-ua 9* (Table 2B, Fig. 1C-E), and also season, sample year, days at 4°C, and time of day (Table 2B), but not genome-wide heterozygosity, age, and sex. Despite these significant associations between GM composition and MHC variation, the overall effect sizes were small across all MHC variables ($R^2 < 0.4\%$, Table 2, Fig. 1). Sample dispersion

Table 2 PERMANOVA analyses of gut microbiome composition in relation to individual major histocompatibility complex (MHC) characteristics in adult Seychelles warblers. Performed using Euclidean distance matrices of CLR-transformed abundances of (I) 16S amplicon sequencing variants (ASV) composition, (II) metagenomic taxonomic composition, (III) metagenomic functional gene composition categories. Separate models included (A) MHC diversity and (B) the presence/absence of MHC alleles. Significant predictors ($p < 0.05$) are in bold. $N = 253$ samples from 149 individuals were included in the 16S metabarcoding analyses. $N = 99$ samples from 57 individuals were used for analyses of metagenomic taxonomic and functional composition. Bird ID was included as a blocking factor

Predictor	(I) 16S ASV composition				(II) Metagenomics taxonomic composition				(III) Metagenomics functional gene composition			
	df	R ²	F	p	df	R ²	F	p	df	R ²	F	p
A) MHC Diversity												
Heterozygosity	1	0.003	0.719	0.067	1	0.008	0.797	0.030	1	0.015	1.485	0.052
MHC-I Diversity	1	0.003	0.919	0.007	1	0.010	1.034	0.626	1	0.014	1.423	0.045
MHC-I Diversity^2	1	0.004	0.951	0.006	-	-	-	-	-	-	-	-
MHC-II Diversity	1	0.003	0.839	0.036	1	0.009	0.925	0.034	1	0.011	1.069	0.642
MHC-II Diversity^2	1	0.003	0.913	0.028	-	-	-	-	-	-	-	-
Age	1	0.003	0.918	0.930	1	0.015	1.625	0.528	-	0.010	0.976	0.893
Season	1	0.007	1.955	<0.001	1	0.019	1.995	0.001	1	0.014	1.368	0.155
Sample Year	5	0.038	2.023	<0.001	6	0.085	1.492	<0.001	6	0.063	1.044	0.202
Sex	1	0.003	0.914	0.870	1	0.012	1.228	<0.001	1	0.012	1.206	0.185
Days at 4°C	1	0.010	2.618	0.008	1	0.011	1.116	0.494	1	0.014	1.394	0.015
Time of day	1	0.010	2.525	0.001	1	0.017	1.773	<0.001	1	0.013	1.325	0.168
B) Presence/absence of MHC alleles												
Heterozygosity	1	0.003	0.679	0.246	-	-	-	-	-	-	-	-
<i>Ase-dab3</i>	1	0.004	1.004	0.999	-	-	-	-	1	0.009	0.853	0.406
<i>Ase-dab4</i>	1	0.003	0.894	0.371	-	-	-	-	1	0.013	1.336	0.787
<i>Ase-dab5</i>	1	0.004	0.991	0.740	-	-	-	-	1	0.007	0.744	0.634
<i>Ase-ua1</i>	1	0.004	1.107	0.084	-	-	-	-	1	0.012	1.151	0.125
<i>Ase-ua3</i>	1	0.004	0.934	1.000	-	-	-	-	1	0.009	0.899	0.334
<i>Ase-ua4</i>	1	0.005	1.273	0.876	-	-	-	-	1	0.014	1.377	0.439
<i>Ase-ua5</i>	1	0.004	0.986	0.048	1	0.006	0.584	0.584	1	0.009	0.919	0.770
<i>Ase-ua6</i>	1	0.003	0.873	0.185	-	-	-	-	1	0.010	0.959	0.782
<i>Ase-ua7</i>	1	0.004	1.150	0.010	1	0.009	0.995	<0.001	1	0.009	0.882	0.522
<i>Ase-ua8</i>	1	0.006	1.514	0.605	-	-	-	-	1	0.011	1.099	0.119
<i>Ase-ua9</i>	1	0.003	0.899	0.011	1	0.008	0.834	0.118	1	0.010	1.040	0.672
<i>Ase-ua11</i>	1	0.007	1.743	0.976	-	-	-	-	1	0.014	1.376	0.304
Age	1	0.004	0.937	0.971	1	0.015	1.532	0.518	1	0.014	1.353	0.916
Season	1	0.007	1.966	0.001	1	0.020	2.145	<0.001	1	0.014	1.357	0.145
Sample Year	5	0.038	1.986	<0.001	6	0.083	1.455	<0.001	6	0.062	1.026	0.313
Sex	1	0.003	0.889	0.862	1	0.014	1.481	<0.001	1	0.009	0.855	0.148
Days at 4°C	1	0.010	2.543	0.009	1	0.010	1.083	0.485	1	0.013	1.322	0.012
Time of day	1	0.009	2.401	<0.001	1	0.017	1.794	<0.001	1	0.014	1.428	0.123

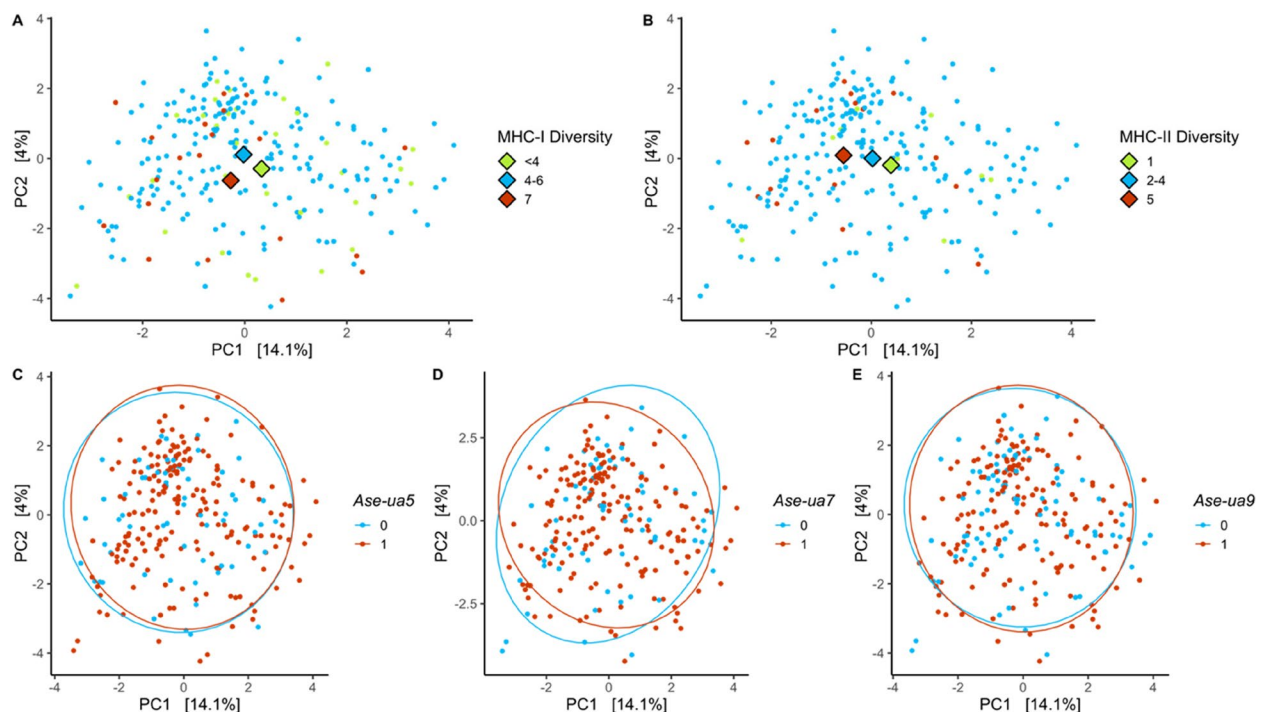


Fig. 1 Principal Component Analyses (PCA) of gut microbiome compositional variation determined using 16S rRNA metabarcoding of adult Seychelles warbler faecal samples in relation to MHC-I diversity (**A**), MHC-II diversity (**B**), and the presence/absence (1/0) of MHC-I allele *Ase-ua 5* (**C**), MHC-I allele *Ase-ua 7* (**D**), MHC-I allele *Ase-ua 9* (**E**). $N = 253$ from 149 birds. Large diamonds represent the group centroids. For clarity, samples were grouped into discrete categories for plotting. In plots **A–B**, the coloured points represent low (green), medium (blue), and high (red) MHC diversity. In plots **C–E**, blue = absence, red = presence of the allele. Ellipses of 95% confidence intervals of each group are drawn around the points. Principal components 1 and 2 explained 14.1% and 4% of the variation in gut microbiome structure, respectively

did not significantly differ with MHC-I variants, *Ase-ua 5*, *Ase-ua 7*, and *Ase-ua 9*, season, and sample year (Table S5).

Metagenomic taxonomic composition

Variation in GM metagenomic taxonomic composition was associated with genome-wide heterozygosity and MHC-II diversity, but not MHC-I diversity (Table 2A, Fig. 2A–B). Of the control variables, sex, sample year, and time of day were associated with metagenomics taxonomic composition (Table 2A), but age, season, and days at 4°C were not.

When assessing MHC variants identified in the 16S analysis, GM metagenomic taxonomic composition was associated with the presence of MHC-I *Ase-ua 7* (Table 2B, Fig. 2C) but not *Ase-ua 5* and *Ase-ua 9*. Sex, season, sample year, and time of day were also associated with metagenomic taxonomic composition (Table 2B), but age and days at 4°C were not. Despite significant differences in GM composition, the overall effect sizes were small across all MHC variables ($R^2 < 0.9\%$, Table 2, Fig. 2). Sample dispersion did not significantly differ with MHC-I variant, *Ase-ua 7*, season, and sample year (Table S5).

Metagenomics functional composition

Functional GM composition was significantly associated with increasing individual MHC-I diversity (Table 2A, Fig. 3) and with days at 4°C (Table 2A). However, genome-wide heterozygosity, MHC-II diversity and all other control variables (age, sex, season, sample year, and time of day) were not (Table 2A).

Functional GM composition was not significantly associated with any MHC alleles (Table 2B, Table S6). Functional GM composition was associated with days stored at 4°C (Table 2B), but not with any other control variables (age, sex, season, sample year, time of day) (Table 2B). Despite significant differences in GM composition, the overall effect sizes were small across all MHC variables ($R^2 < 1.4\%$, Table 2, Fig. 3), but the MHC-I diversity effect size is the largest among GM compositional analyses. Sample dispersion did not significantly differ with season and sample year (Table S5).

Differential abundance analysis

Differential abundance of metagenomic taxonomic species

The abundance of some individual bacterial species (identified using metagenomics) varied in relation to MHC characteristics (Fig. 4AB); the abundance of

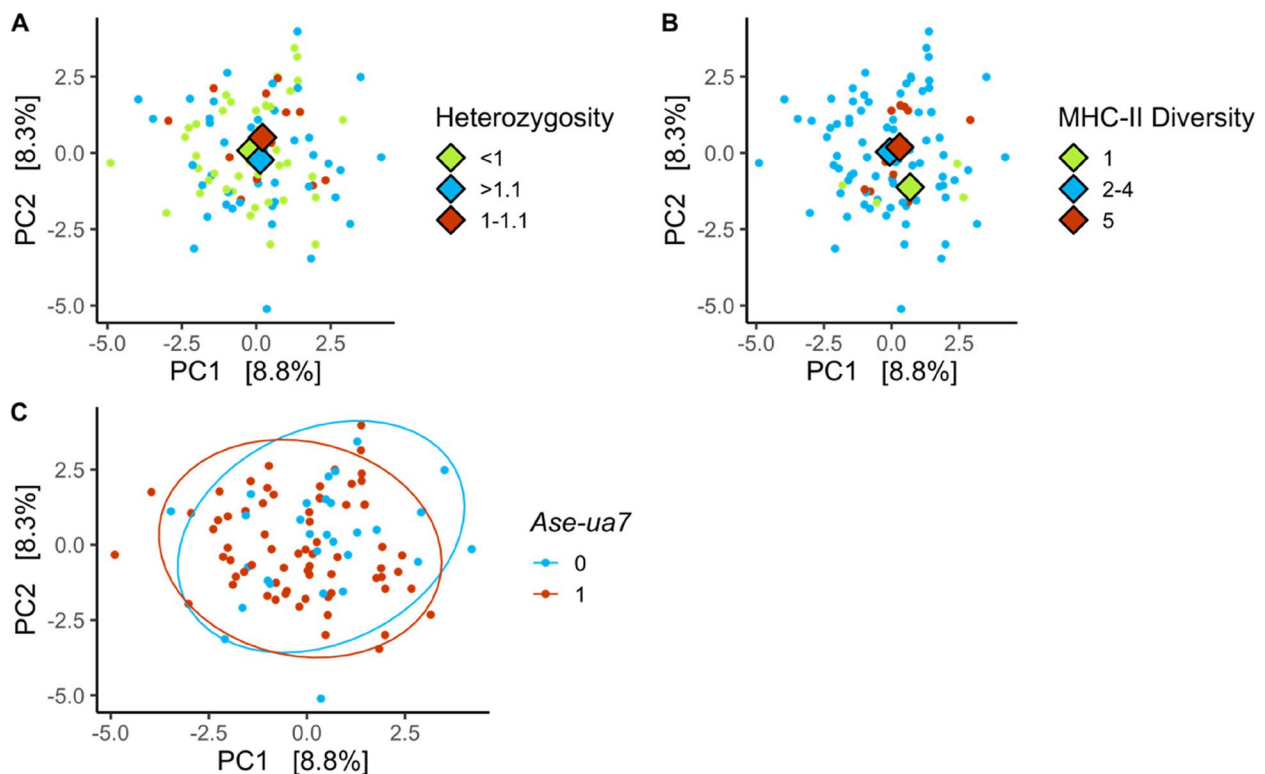


Fig. 2 Principal Component Analyses (PCA) of gut microbiome metagenomic taxonomic compositional variation of Seychelles warbler faecal samples in relation to genome-wide heterozygosity (**A**), MHC-II diversity (**B**), and MHC-I allele *Ase-ua 7* (**C**). $N=99$ from 57 birds. Large diamonds represent the group centroids. For clarity, samples were grouped into discrete categories for plotting. In plots **A** and **B**, the coloured points indicate low (green), middle (blue), and high (red) genome-wide heterozygosity (Heterozygosity) and MHC-II diversity, respectively. In plot **C**, blue = absence (0) and red = presence (1) of *Ase-ua 7*. Ellipses of each group are drawn around the points

Enterococcus casseliflavus decreased, and *Microbacterium enclense* increased with increasing MHC-I diversity (Fig. 4A). The abundance of *Lactococcus lactis* decreased, and the abundance of *Staphylococcus lloydii* increased with increasing MHC-II diversity (Fig. 4B).

The abundances of four bacterial species were significantly associated with the presence/absence of the MHC-I allele *Ase-ua 7* (identified as associated with GM taxonomy composition), i.e. there was decreased prevalence of *Enterococcus casseliflavus* and *Gordonia* sp OPL2, and an increased prevalence of *Escherichia coli* and *Vulcaniibacterium thermophilum*, when *Ase-ua 7* was present (Fig. 4C). The abundance of bacterial species was not significantly related to host genome-wide heterozygosity (Fig. 4D).

Differential abundance of metagenomic functional genes

Abundances of 24 GM functional gene annotations differed significantly in relation to individual host MHC-I diversity (Table S7, Fig. 5A). In total, 9 GM genes increased in abundance and 15 genes decreased in abundance with increasing MHC-I diversity.

The 24 gene annotations were derived from 13 functional gene categories (defined by Cluster of Gene Orthologs (COG)). With increasing MHC-I diversity, two COG categories only increased in prevalence, four COG categories increased and decreased in prevalence, and seven COG categories only decreased in prevalence (Fig. 5B). Five of the seven COG categories that only decreased in prevalence are involved in bacterial metabolism. In addition, one GM functional gene annotation (COG1216) increased in prevalence with increasing genome-wide heterozygosity.

The KEGG orthologs and pathways of MHC-I diversity associated genes (Table S7) further support the findings, as core microbial functions decreased in prevalence; Carbohydrate metabolism (K01885, K01886, K01652, K06131, K01115), lipid metabolism (K00666), amino acid metabolism (K00852, K00847, K00874), transcription (K03043, K02335, K04799), translation (K08832, K15409), and replication and repair (K03581, K04043, K03283) (Table S7). In addition, MHC-I diversity was positively associated with KEGG pathways

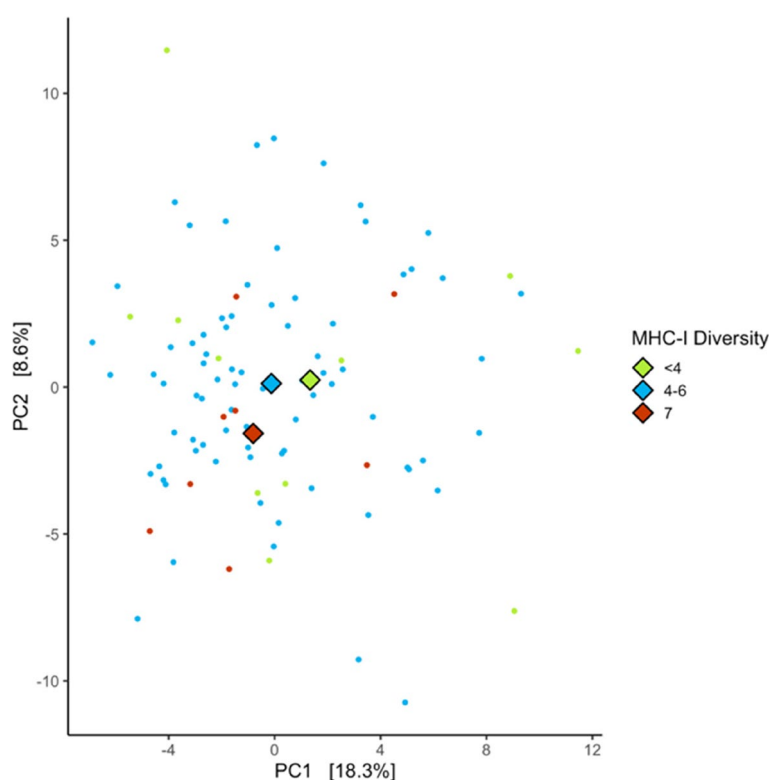


Fig. 3 Principal Component Analyses (PCA) of gut microbiome compositional variation determined using metagenomics function with MHC-I diversity in the gut microbiome of Seychelles warblers ($n=99$ from 57 birds). Large diamonds represent the group centroids. For clarity, samples were grouped into discrete categories for plotting. The coloured points represent the count < 4 (green), 4–6 (blue), and 7 (red) of MHC-I diversity. Principal components 1 and 2 explained 18.3% and 8.6% of gut microbiome structure, respectively

involved in environmental defence or stress adaptation (K02004, K18138, K09800, K04763; Table S7).

A direct test of trade-offs between functional categories of microbial defence and metabolism shows that as MHC-I diversity increases, the relative abundance of microbial defence functional category increases (Estimate = 0.32, $p=0.017$, Table S8, Figure S2), whereas the relative abundance of microbial metabolism functional categories decreases (Estimate = -0.26 , $p=0.001$, Table S8, Figure S2).

Three GM functional gene annotations were differentially abundant with increasing MHC-II diversity: decreased in COG0318 (K00666 – fatty-acyl-CoA synthase), and increased in COG1609 (K02529 – LacI family transcriptional regulator, galactose operon repressor) and COG1653 (K02027 – multiple sugar transport system substrate-binding protein).

Network analysis

Network analysis of metagenomic taxonomic species

The metagenomic taxonomic species network had a higher number of connected nodes and edges in low MHC-I diversity (19 nodes, 13 edges) compared to high

(2 nodes, 1 edge) MHC-I diversity individuals (Fig. 6AB). The average number of edges per connected node was also higher in low (mean = 1.4 edges per node) than in high MHC-I diversity (mean = 1.0 edges per node). Modularity was 0.82 in low MHC-I diversity, but zero in high MHC-I diversity due to only having one edge. The ratio of negative to positive edges in low MHC-I diversity was 0.3; there was only a single positive edge and no negative edges in MHC-I diversity. However, permutation tests show that edge density, modularity, and positive edge percentage of the metagenomic taxonomic species network did not significantly differ between low and high MHC-I diversity (Table S9).

In relation to MHC-II diversity, the metagenomic taxonomic species network had a higher number of connected nodes in low (14 nodes) than in high (10 nodes) MHC-II diversity individuals (Fig. 6CD). However, the number of edges did not differ between low (8 edges) and high (8 edges) MHC-II diversity (Fig. 6CD). The average number of edges per connected node was lower in low (mean = 1.1 edges per node) than in high (mean = 1.6 edges per node) MHC-II diversity.

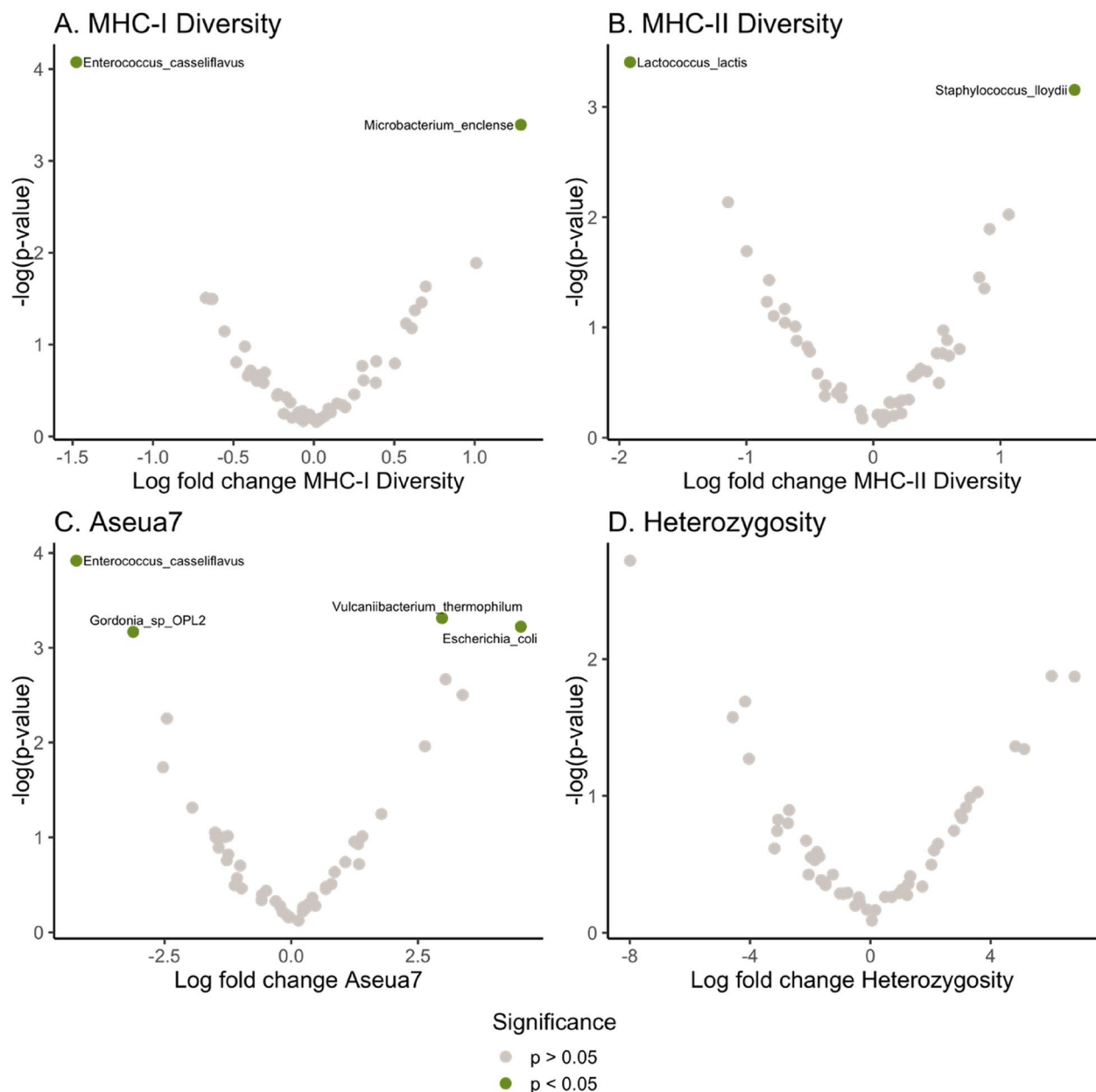


Fig. 4 Differential abundance of metagenomically identified bacterial species in adult Seychelles warblers according to host genome-wide heterozygosity (A), MHC-I diversity (B), MHC-II diversity (C), and presence of MHC-I *Aseua* 7 ($n = 99$ from 57 birds) (D). Points represent bacterial species and are coloured according to significance; green points (with species-level taxonomic annotations) are significantly differentially abundant ($p < 0.05$), and grey points are not

Modularity was also higher in low (0.81) than in high (0.59) MHC-II diversity. The ratio of negative to positive edges was 0.33 in low MHC-II diversity and was 0.14 in high MHC-II diversity. Permutation tests show that edge density, modularity, and positive edge percentage of the metagenomic taxonomic species

network did not significantly differ between low and high MHC-II diversity (Table S9).

Network analysis of metagenomic functional genes

The metagenomic functional genes network had a higher number of connected nodes and edges in individuals with low MHC-I (76 nodes, 87 edges) than high MHC-I

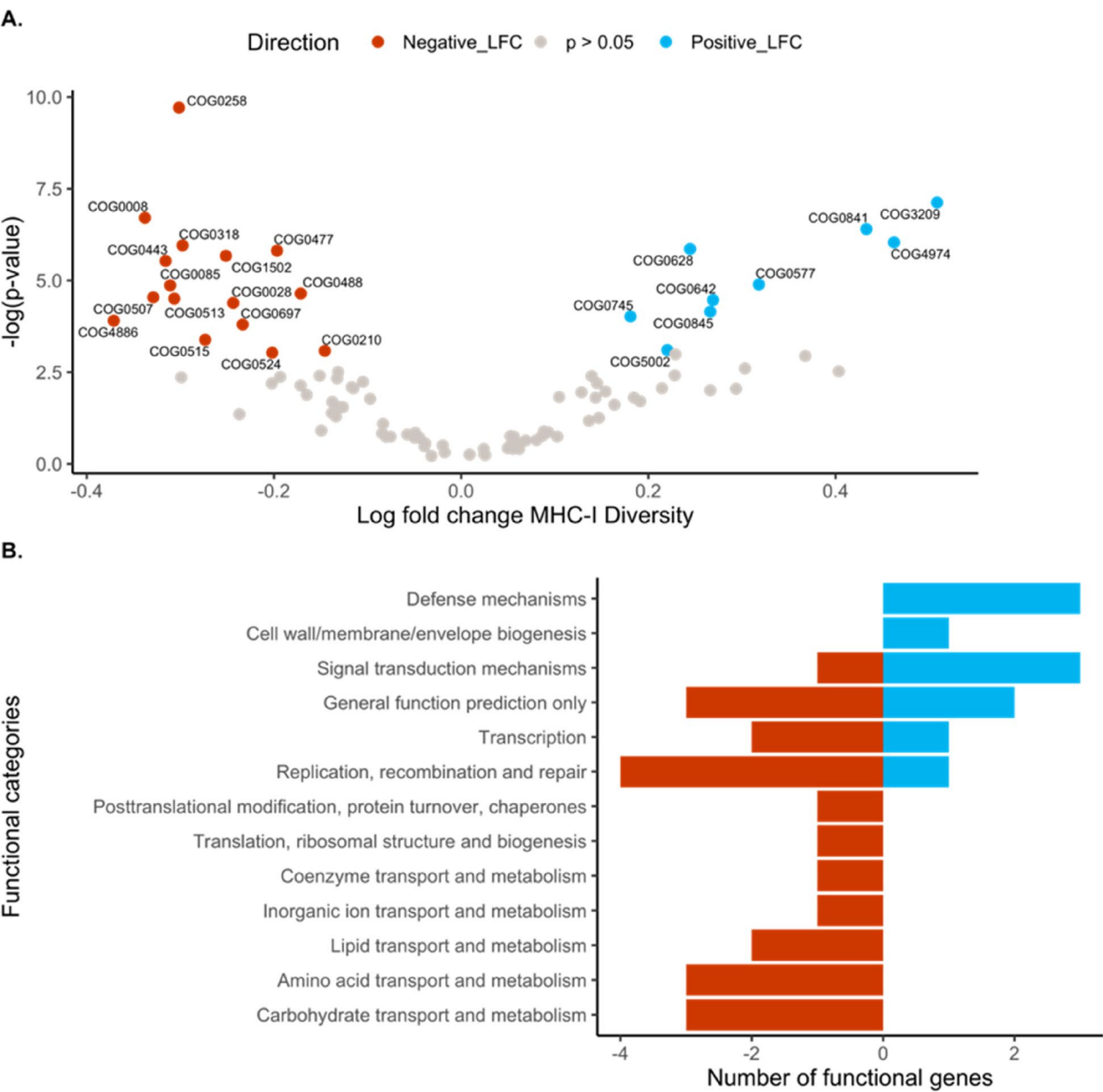


Fig. 5 Variation in the abundance of bacterial functional genes (determined using eggNOG) in the gut microbiome of adult Seychelles warblers in relation to individual MHC-I diversity ($n = 99$ from 57 birds). **A** The results of an ALDEx2 differential abundance test showing the log fold change in abundance of each eggNOG gene annotation with increasing MHC-I diversity. Blue points are significantly ($p < 0.05$) more abundant, red points are significantly ($p < 0.05$) less abundant, and grey points do not differ significantly with MHC-I diversity. Labels on significant points represent eggNOG members. **B** Counts of functional genes per eggNOG category that demonstrated a significant positive (blue) or negative (red) log fold change with increasing MHC-I diversity, respectively

(56 nodes, 44 edges) diversity (Fig. 6EF). The average number of edges per connected node was also higher in low MHC-I (mean=2.3 edges per node) than in high MHC-I diversity (mean=1.6 edges per node). Modularity was lower in low MHC-I (0.72) than in high (0.85) MHC-I diversity. The ratio of negative to positive edges was 0.45 in low MHC-I diversity and 0.19 in high MHC-I

diversity. Permutation tests show that the edge density of the metagenomic functional genes network was significantly higher in low compared to high MHC-I diversity ($p = 0.04$, Table S9). This shows that the metagenomically derived functional genes network is more fragmented in high MHC-I diversity individuals. Modularity and positive edge percentage of the metagenomic functional

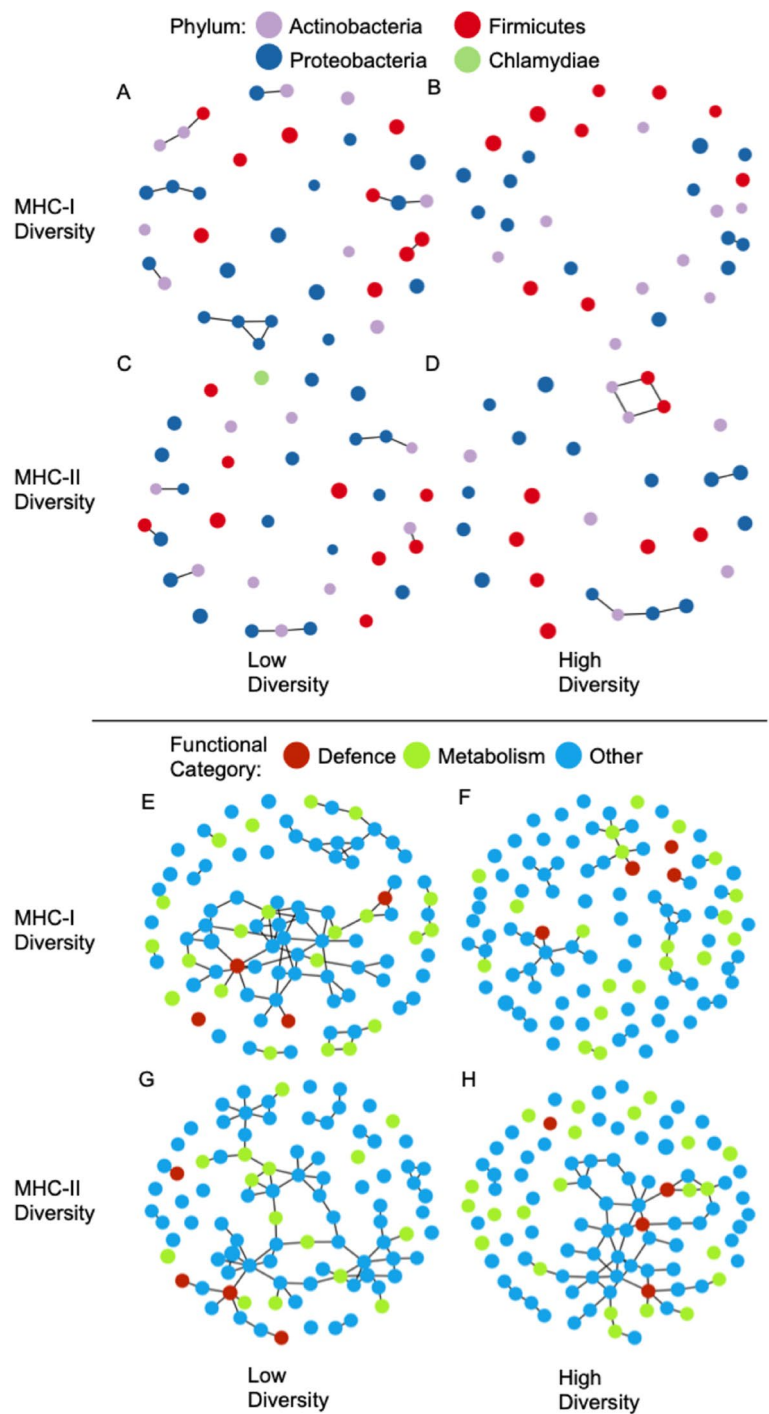


Fig. 6 Network analysis between MHC diversity and the gut microbiome in adult Seychelles warblers ($n = 99$ from 57 birds). Each node represents (A–D) metagenomic taxonomic species or (E–H) metagenomic functional genes. MHC-I diversity (ABEF), MHC-II diversity (CDGH), low MHC diversity (ACEG), high MHC diversity (BDFH). Nodes (species/genes) are coloured by Phylum (A–D) and functional category (E–H). The size of the nodes is proportional to the mean abundance. Lines are edges (interaction between species/genes), connecting nodes that are linked

genes network did not significantly differ between low and high MHC-I diversity. (Table S9).

In relation to MHC-II diversity, the metagenomic functional genes network had a higher number of connected nodes and edges in low (74 nodes, 83 edges) than in high MHC-II diversity individuals (55 nodes, 62 edges) (Fig. 6GH). The average number of edges per connected node was very similar between low (mean=2.2 edges per node) and high (mean=2.3 edges per node) MHC-II diversity individuals. Modularity was also higher in low (0.76) than in high (0.65) MHC-II diversity individuals. The ratio of negative to positive edges was 0.26 in low MHC-II diversity and was 0.44 in high MHC-II diversity. However, permutation tests show that edge density, modularity, and positive edge percentage of the metagenomic functional genes network did not significantly differ between low and high MHC-II diversity (Table S9).

MHC divergence vs. MHC diversity

In GM alpha diversity models, using MHC divergence resulted in a similar model fit (assessed via AIC) to models using the original MHC diversity measures (see supplementary; Table 11, Table S10). However, even in models where the fit was slightly improved by using MHC divergence (16S GM richness, metagenomic taxonomic richness, metagenomic functional richness, and metagenomic functional Shannon diversity), using MHC divergence had little impact on which specific variables were found to be significantly associated with GM diversity (i.e. they showed qualitatively the same results). Of the six alpha diversity models, only one model showed a significant change in variable significance; MHC-I divergence was significantly associated with metagenomic taxonomic richness, but not MHC-I diversity (Table S12).

In GM composition models, using MHC divergence did not improve model fit (AIC) and did not improve variance explained (R^2) of any model compared to using the MHC diversity measure (Table 2, Table S11). Of the three GM composition models, only one model showed a significant change in variable significance; MHC-I divergence was not significantly associated with 16S GM composition, but MHC-I diversity was (Table S12).

Discussion

Our study shows that in adult Seychelles warblers, there was no association between GM diversity (characterised by either 16S ASV or metagenomics) and MHC diversity, though one specific MHC variant (MHC-I allele *Ase-ua 11*) was associated with 16S GM richness but not metagenomic GM diversity. However, GM composition (16S and metagenomic-derived) was associated with MHC diversity. The 16S GM composition was associated with MHC-I and MHC-II diversity in a

non-linear (quadratic function) manner, and with the MHC-I alleles *Ase-ua 5*, *Ase-ua 7*, and *Ase-ua 9*. The metagenomic taxonomic composition was associated with MHC-II diversity and the MHC-I allele *Ase-ua 7*. Furthermore, the overall functional composition of the GM (metagenomically derived) was associated with variation in MHC-I diversity. Two metagenomic bacterial species were differentially abundant with increasing MHC-I and MHC-II diversity, and four bacterial species differed in relation to the presence/absence of MHC-I allele *Ase-ua 7*. Furthermore, 24 functional genes differed in abundance with increasing MHC-I diversity (this corresponded to increases in microbial defence mechanisms and decreases in microbial metabolism) and three with increasing MHC-II diversity. Network analysis showed that high (compared to low) MHC-I diversity was also associated with greater fragmentation in functional GM structure.

Consistent with previous findings on the Seychelles warbler, we found that MHC diversity (either class-I or class-II) was not significantly associated with GM alpha diversity [21, 110]. This contradicts results found in other wild populations [6, 41, 58, 101]. For example, in giant salamanders (*Cryptobranchus alleganiensis bishopi* and *C. a. alleganiensis*) and sticklebacks (*Gasterosteus aculeatus*), increasing MHC allelic diversity is correlated with increases and decreases in overall skin microbiome diversity, respectively [6, 41]. Given that the Seychelles warblers had a recent population bottleneck (<50 individuals in the 1960s, [95]), the MHC is less diverse than in most similar species [88], and may therefore have a reduced impact on GM diversity. However, even so, considerable MHC variation remains in the Cousin Island warbler population [88] and the most functionally divergent MHC alleles were likely retained, as such alleles are more likely to differ in antigen-binding properties and thus confer broader immunological protection [72, 88]. This may also explain why replacing MHC diversity with MHC divergence had little effect on our overall results as there is little difference between these metrics in this species. Thus, we conclude that MHC and GM diversity associations may be host-species dependent and functionally dependent regardless of overall MHC diversity/divergence levels.

In contrast to the GM diversity results, shifts in GM composition (both 16S and metagenomic) were associated with MHC characteristics. An association between MHC-II diversity and GM taxonomic composition, as seen in our study, has been shown multiple times in captive and wild systems (see review [90]). In our study, only two bacterial species differed in abundance among individuals with different levels of MHC diversity (increased *Staphylococcus loydii* and decreased *Lactococcus lactis*).

There is evidence of *Lactococcus lactis* causing detrimental infections in avian species [36], but it has also been used as a probiotic for broiler chickens (*Gallus gallus domesticus*) [9, 69]. *Staphylococcus loydii* hasn't been observed to cause infections in any species. These results may suggest that in the Seychelles warblers, increasing MHC-II diversity may suppress pathogenic species. In addition, the MHC-I allele *Ase-ua 7* was significantly correlated with differences in GM composition, consistent with previous work on this population [21]. *Ase-ua 7* was associated with decreases in *Enterococcus casseliflavus* and *Gordonia* sp OPL2, (gram-positive bacteria), but increased prevalence of the gram-negative bacteria, *Escherichia coli* and *Vulcaniibacterium thermophilum*, [57, 62, 70, 116]. The World Health Organization (WHO) priority list of pathogens primarily consists of Gram-negative bacteria [10]. Therefore, we speculate that the effect of *Ase-ua 7* on GM composition could be based on microbial cell wall structure and could be important in controlling pathogens. Indeed, during inflammation, MHC expression was upregulated in the small intestinal stem cells of humans and mice (*Mus musculus*) [42].

In our study, the effects of MHC-II diversity and the MHC-I allele *Ase-ua 7* on taxonomic GM are relatively small (two and four, out of 49, core species were differentially abundant, respectively). Such limited effects have also been found in the few previous similar studies, with the MHC variations influencing only a very small number of bacterial species rather than the overall composition [6, 68, 110]. This small taxonomic effect could explain why both MHC-II diversity and MHC-I allele *Ase-ua 7* were not significantly associated with metagenomic function (see below). As only a few microbial species are influenced, other microbes may replace their function through functional redundancy [64, 112]. Moreover, network analyses of both taxonomic and functional GM revealed that MHC-II diversity was not associated with greater fragmentation, modularity, and a negative-to-positive interaction ratio; these differences were more pronounced for MHC-I diversity. This further supports the conclusion that MHC-II diversity has a relatively small effect on the gut microbiome.

In contrast, the diversity of MHC-I alleles (but not class-II) did appear to influence GM functional differences (determined metagenomically), with higher MHC-I diversity leading to the presence of an increased number of microbial defence genes, whilst decreasing the number of metabolism-related genes. This is further shown in network analysis, where functional microbial networks appear more fragmented in individuals with high MHC-I diversity, indicating reduced microbial interactions. These patterns may reflect increased microbial competition, fewer metabolic interactions among microbes, or

immune-mediated disruption under high MHC-I diversity. Trade-offs between defence and growth are common (although not ubiquitous) in microbial species [27, 63]. Our results suggest that these microbial trade-offs can also amount to costs for the host, whereby control of the microbiome (via the immune system) can result in a reduction in the GM's metabolic potential. Presumably, these costs may be outweighed by the benefits of eliminating pathogens and maintaining a healthy microbiome [34, 67]. Quantifying the relative costs and benefits of maintaining a dynamic microbiome remains largely unexplored but is essential for understanding how host-microbiome interactions and host control mechanisms (including the immune system) evolve [34, 67, 105].

Overall, our study indicates that MHC-I, not MHC-II, plays a greater role in shaping the GM, both taxonomically and functionally. This is consistent with previous findings in the Seychelles warbler [21, 110], as well as in the reddish-gray mouse lemur (*Microcebus griseorufus*) [68]. MHC-I molecules encode receptors that typically act intracellularly, while MHC-II receptors interact extracellularly. Thus, intuitively the class-II receptors should have greater interaction with and influence over gut microbes [89]. However, the mechanisms by which MHC-I alleles and diversity influence GM function remain unclear [90]. MHC-I receptors can be triggered by bacteriophages [2], which play an important role in shaping bacterial composition [45]. Future work on host MHC and the gut virome in the Seychelles warbler could help to elucidate the mechanisms by which MHC-I influences GM composition.

It is also possible that MHC-I indirectly impacts the gut microbiome via effects on host health. In the Seychelles warblers, MHC-I diversity has been positively correlated with both survival and reproductive success [12, 87, 88]. In the current study, we also found that MHC-I diversity was associated with GM metagenomic function. It is plausible that MHC-I diversity affects GM functionality by controlling mutualistic bacteria, indirectly influencing host survival. However, MHC-I variation could affect other components of the host's overall health, indirectly leading to differences in GM function. Consequently, multiple mechanisms may be involved in the relationship between MHC-I diversity and individual differences in the GM.

An individual's sex and genome-wide heterozygosity, independent of MHC variation, were also associated with shifts in GM metagenomic species. However, these changes were not detectable in terms of 16S ASVs and metagenomics function, which indicates that the changes are species/strain specific and do not influence microbial function. This result is reinforced by the fact that no common species were differentially abundant with

increasing levels of genome-wide heterozygosity. Various other variables including sample year, season, and time of day, were also predictors of 16S ASVs and metagenomic species, but not metagenomic function. The fact that so many factors affect the species present in the GM, but not the overall GM functionality is likely attributed to functional redundancy, as the types of environmental microbes change with time, the overall GM function is replaced by other microbes (potentially due to changes in diet), thereby preserving the overall GM function [64, 112].

Interestingly, the number of days stored at 4°C was associated with shifts in metagenomics function but not 16S ASVs and metagenomic species. Larger DNA fragments degrade quicker in storage, and these larger DNA fragments are required for accurate gene recognition and gene annotation, as these steps require a start codon followed by an open reading frame [47]. Whereas smaller DNA fragments may still be sufficient for accurate taxonomic assignment of 16S ASVs and metagenomic species, as specific marker genes are typically used for taxonomic annotations [4, 74].

One limitation of our study is that it is purely correlative, and we are unable to validate our findings experimentally. Nonetheless, our findings in both datasets (16S and metagenomic methods) are comparable despite the smaller sample size in the metagenomic work. Future experimental work in more amenable study systems could investigate potential pathways by which host MHC-I diversity and alleles may affect the GM, for example by introducing immune-triggering bacteriophages and measuring their impact on the gut bacteriophage and bacteria community. [35, 104]. However, such experiments are likely lab-based and the GM would be radically altered [102], hence the generalisation to natural populations would be limited. A further limitation of our study is that it lacks gene expression data for the MHC and the GM. Therefore, the functional relevance is only on the DNA level, reflecting potential rather than actual function of the MHC and GM. However, in the bottlenecked Seychelles warbler population, the few remaining MHC alleles are highly divergent, preventing further reduction into functional supertypes [21, 88, 113]. Future research could use transcriptomics and proteomics to quantify both MHC and GM function. The lack of expression data could plausibly explain the modest effect sizes in this study as a locus could be present but not expressing. In addition, although multiple studies have found that MHC is associated with the microbiome [6, 41, 58, 101], the host regulation of the GM may be more directly shaped by innate immunity, such as by toll-like receptors, NOD-like receptors or defensins [105]. However, a previous study in the Seychelles warbler found no associations

between toll-like receptor 3 (variation at which influences host survival in this species) and the GM alpha diversity or composition [21]. Therefore, while the MHC may only explain a limited amount of variation in the GM (as shown in this study and [21, 110]), this variation may still be important.

Despite reporting several associations between MHC characteristics and the GM, our study is limited to only assessing genome-wide heterozygosity and the specific MHC candidate alleles we had already screened for. A genome-wide association study on the host and its GM could reveal more loci, and potentially more nuanced or polygenic effects, by accounting for multiple genes concurrently [114]. However, this requires whole-genome data on many individual birds, in combination with metagenomic sequencing of their GM, which would be very costly and labour-intensive [56, 115]. Leveraging a sequencing approach that targets specific individuals and utilises recent technological advances, such as the imputation of low-coverage samples, could be a feasible way to conduct such studies [115]. In the same vein, while heterozygosity measured using 30 neutral microsatellite loci does reflect genome-wide heterozygosity/inbreeding in this species [95], whole-genome sequencing of individuals would be a more accurate measure. It would enable us to determine runs of homozygosity, and thus provide the resolution needed for a powerful investigation of the effect of host inbreeding on the GM [17].

In conclusion, our study suggests that both MHC class I and II influence an individual's GM in the Seychelles warbler, but that despite being an intracellular receptor, MHC-I has a greater influence on GM composition, than MHC-II. We also found that MHC-I allele *Ase-ua 7* changes the GM taxonomic composition, while MHC-I diversity alters GM function. These results may explain previous findings that MHC-I diversity is positively correlated with fitness in this population [12, 87], potentially as a result of inducing changes in the GM functionality. However, this could be an indirect effect, rather than the GM actively contributing to increased fitness. Various pathways are involved in regulating the immune system, underscoring the need for host and gut transcriptomics and metabolomic data to enable mechanistic investigation of immunogenetics and the GM [24, 90].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-026-02341-9>.

Additional file 1: Supplementary materials.

Acknowledgements

We thank Nature Seychelles for facilitating fieldwork on Cousin Island and the Seychelles Bureau of Standards and the Ministry of Agriculture, Climate

Change & Environment for providing permission to conduct fieldwork and sample collection. This study would not have been possible without the contribution of exceptional fieldworkers, laboratory technicians and database managers associated with the Seychelles Warbler Project. The research presented in this paper was carried out on the High-Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia.

Code availability

All code required to recreate analyses, output plots and tables can be found on https://github.com/Chuen-Lee/SW_MHC_GM.git.

Authors' contributions

CZL and DSR conceived the study. CZL performed the research and analysed the data with input from DSR. DSR and HLD acquired funding. DSR, CSD, SFW, JK, FH, HLD supervised the research. CZL wrote the paper with the help of DSR and input from all co-authors. All authors contributed to the review and editing of the paper.

Funding

CZL was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) Norwich Research Park Biosciences Doctoral Training Partnership (Grant number BB/T008717/1). DSR and HLD were funded by a Natural Environment Research Council (NERC) grant (NE/S010939/1). SFW was funded by a Leverhulme Trust Early Career Fellowship (ECF-2023–433). CSD was funded by the Natural Environment Research Council and EnvEast DTP (NE/L002582/1) and UK Natural Environment Research Council (NERC) Biomolecular Analysis Facility at the University of Sheffield (NBAF1150). FH was supported by the European Research Council H2020 StG (erc-stg-948219, EPYC), BBSRC Institute Strategic Programme Food Microbiome and Health (BB/X011054/1, BBS/E/F/000PR13631), Earlham Institute ISP Decoding Biodiversity (BBX011089/1, BBS/E/ER/230002A and BBS/E/ER/230002B). JK and DSR were funded by Dutch Science Council grant (ALW NWO Grant No. ALWOP531), JK was funded by NWO TOP grant 854.11.003 and NWO VICI 823.01.014 also from Dutch Science Council. HLD was funded by a Rosalind Franklin Fellowship from the University of Groningen.

Data availability

All 16S sequencing data used in this study can be accessed from European Nucleotide Archive (ENA) database under the study accession numbers PRJEB45408 (samples taken in 2017 and 2018) and PRJEB47095 (samples taken in 2019 and 2020) and PRJEB67634 (samples taken in 2021 and 2022). All shotgun metagenomics data used in this study can be accessed from ENA database under the study accession number PRJEB81709.

Declarations

Ethics approval and consent to participate

Fieldwork was carried out in accordance with local ethical regulations and agreements (UEA ethics approval ID ETH2223-0665). The Seychelles Department of Environment and the Seychelles Bureau of Standards approved the fieldwork (permit number A0157).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹School of Biological Sciences, University of East Anglia, Norwich, Norfolk, UK. ²Centre for Microbial Interactions, Norwich Research Park, Norwich, Norfolk, UK. ³Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, The Netherlands. ⁴Quadram Institute Bioscience, Norwich Research Park, Norwich, Norfolk, UK. ⁵Earlham Institute, Norwich Research Park, Norwich, Norfolk, UK. ⁶Nature Seychelles, Roche Caiman, Mahé, Republic of Seychelles.

Received: 31 May 2025 Accepted: 6 January 2026

Published online: 12 March 2026

References

- Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1). <https://doi.org/10.18637/jss.v067.i01>.
- Bazan J, Calkosiński I, Gamian A. Phage display—a powerful technique for immunotherapy. *Hum Vaccin Immunother*. 2012;8(12):1817–28. <https://doi.org/10.4161/hv.21703>.
- Berryman MA, Ilonen J, Triplett EW, Ludvigsson J. Functional metagenomic analysis reveals potential inflammatory triggers associated with genetic risk for autoimmune disease. *J Autoimmun*. 2024;148:103290. <https://doi.org/10.1016/j.jaut.2024.103290>.
- Blanco-Míguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol*. 2023;41(11):1633–44. <https://doi.org/10.1038/s41587-023-01688-w>.
- Blum JS, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annu Rev Immunol*. 2013;31(1):443–73. <https://doi.org/10.1146/annurev-immunol-032712-095910>.
- Bolnick DI, Snowberg LK, Caporaso JG, Lauber C, Knight R, Stutz WE. Major histocompatibility complex class IIb polymorphism influences gut microbiota composition and diversity. *Mol Ecol*. 2014;23(19):4831–45. <https://doi.org/10.1111/mec.12846>.
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852–7. <https://doi.org/10.1038/s41587-019-0209-9>.
- Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, et al. The effect of host genetics on the gut microbiome. *Nat Genet*. 2016;48(11):1407–12. <https://doi.org/10.1038/ng.3663>.
- Boodhoo N, Shojadoost B, Alizadeh M, Astill J, Behboudi S, Sharif S. Effect of treatment with *Lactococcus lactis* NZ9000 on intestinal microbiota and mucosal immune responses against *Clostridium perfringens* in broiler chickens. *Front Microbiol*. 2023;14. <https://doi.org/10.3389/fmicb.2023.1257819>.
- Breijyeh Z, Jubeh B, Karaman R. Resistance of Gram-negative bacteria to current antibacterial agents and approaches to resolve it. *Molecules*. 2020;25(6):1340. <https://doi.org/10.3390/molecules25061340>.
- Briatte F. _ggnet: Functions to plot networks with ggplot2_. (0.1.0). R package. 2025.
- Brouwer L, Barr I, Van De POL M, Burke T, Komdeur J, Richardson DS. MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Mol Ecol*. 2010;19(16):3444–55. <https://doi.org/10.1111/j.1365-294X.2010.04750.x>.
- Brown TJ, Dugdale HL, Hammers M, Komdeur J, Richardson DS. Seychelles warblers with silver spoons: Juvenile body mass is a lifelong predictor of annual survival, but not annual reproduction or senescence. *Ecol Evol*. 2022;12(7):e9049. <https://doi.org/10.1002/ece3.9049>.
- Burnham KP, Anderson DR, Huyvaert KP. AIC model selection and multi-model inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol*. 2011;65(1):23–35. <https://doi.org/10.1007/s00265-010-1029-6>.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581–3. <https://doi.org/10.1038/nmeth.3869>.
- Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;38(12):5825–5829. <https://doi.org/10.1093/molbev/msab293>.
- Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of homozygosity: a comparison of SNP array and whole genome sequence low coverage data. *BMC Genomics*. 2018;19(1):106. <https://doi.org/10.1186/s12864-018-4489-0>.
- Chao A, Gotelli NJ, Hsieh TC, Sander EL, Ma KH, Colwell RK, et al. Rarefaction and extrapolation with hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol Monogr*. 2014;84(1):45–67. <https://doi.org/10.1890/13-0133.1>.
- Coulon A. <scpg>genhet</scpg> : an easy-to-use R function to estimate individual heterozygosity. *Mol Ecol Resour*. 2010;10(1):167–169. <https://doi.org/10.1111/j.1755-0998.2009.02731.x>.
- Criscitelli MF, de Figueiredo P. Fifty shades of immune defense. *PLoS Pathogens*. 2013;9(2). <https://doi.org/10.1371/journal.ppat.1003110>.

21. Davies CS, Worsley SF, Maher KH, Komdeur J, Burke T, Dugdale HL, et al. Immunogenetic variation shapes the gut microbiome in a natural vertebrate population. *Microbiome*. 2022;10(1):41. <https://doi.org/10.1186/s40168-022-01233-y>.
22. Dearborn DC, Gager AB, McArthur AG, Gilmour ME, Mandzhukova E, Mauck RA. Gene duplication and divergence produce divergent MHC genotypes without disassortative mating. *Mol Ecol*. 2016;25(17):4355–67. <https://doi.org/10.1111/mec.13747>.
23. Dzierozynski L, Queen J, Sears CL. Subtle, persistent shaping of the gut microbiome by host genes: a critical determinant of host biology. *Cell Host Microbe*. 2023;31(10):1569–73. <https://doi.org/10.1016/j.chom.2023.09.007>.
24. Eshleman EM, Shao TY, Woo V, Rice T, Engleman L, Didriksen BJ, Whitt J, Haslam DB, Way SS, Alenghat T. Intestinal epithelial HDAC3 and MHC class II coordinate microbiota-specific immunity. *J Clin Invest*. 2023;133(4). <https://doi.org/10.1172/JCI162190>.
25. Fabbri M, Scicchitano D, Candela M, Turroni S, Rampelli S. Connect the dots: sketching out microbiome interactions through networking approaches. *Microbiome Res Rep*. 2023;2(4). <https://doi.org/10.20517/mrr.2023.25>.
26. Fenn J, Taylor C, Goertz S, Wanlik KM, Paterson S, Begon M, et al. Discrete patterns of microbiome variability across timescales in a wild rodent population. *BMC Microbiol*. 2023;23(1):87. <https://doi.org/10.1186/s12866-023-02824-x>.
27. Ferenci T. Trade-off mechanisms shaping the diversity of bacteria. *Trends Microbiol*. 2016;24(3):209–23. <https://doi.org/10.1016/j.tim.2015.11.009>.
28. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. Anova-like differential expression (ALDEx) analysis for mixed population RNA-seq. *PLoS ONE*. 2013;8(7):e67019. <https://doi.org/10.1371/journal.pone.0067019>.
29. Fleischer R, Risely A, Hoeck PEA, Keller LF, Sommer S. Mechanisms governing avian phyllosymbiosis: Genetic dissimilarity based on neutral and MHC regions exhibits little relationship with gut microbiome distributions of Galápagos mockingbirds. *Ecol Evol*. 2020;10(23):13345–54. <https://doi.org/10.1002/ecs3.6934>.
30. Fleischer R, Schmid DW, Wasimuddin Brändel SD, Rasche A, Corman VM, Drosten C, Tschapka M, Sommer S. Interaction between MHC diversity and constitution, gut microbiota and Astrovirus infections in a neotropical bat. *Mol Ecol*. 2022;31(12):3342–3359. <https://doi.org/10.1111/mec.16491>.
31. Fuess LE, den Haan S, Ling F, Weber JN, Steinel NC, Bolnick DI. Immune gene expression covaries with gut microbiome composition in stickleback. *MBio*. 2021;12(3). <https://doi.org/10.1128/mBio.00145-21>.
32. Gill T, Asquith M, Brooks SR, Rosenbaum JT, Colbert RA. Effects of HLA-B*27 on gut microbiota in experimental spondyloarthritis implicate an ecological model of dysbiosis. *Arthritis Rheumatol*. 2018;70(4):555–65. <https://doi.org/10.1002/art.40405>.
33. Gill T, Brooks SR, Rosenbaum JT, Asquith M, Colbert RA. Novel inter-omic analysis reveals relationships between diverse gut microbiota and host immune dysregulation in HLA-B*27-induced experimental spondyloarthritis. *Arthritis Rheumatol*. 2019;71(11):1849–57. <https://doi.org/10.1002/art.41018>.
34. Gillingham MA, Prüter H, Montero BK, Kempnaers B. The costs and benefits of a dynamic host microbiome. *Trends Ecol Evol*. 2025;40(3):255–72. <https://doi.org/10.1016/j.tree.2024.11.008>.
35. González-Mora A, Hernández-Pérez J, Iqbal HMN, Rito-Palomares M, Benavides J. Bacteriophage-based vaccines: a potent approach for antigen delivery. *Vaccines*. 2020;8(3):504. <https://doi.org/10.3390/vacc8030504>.
36. Goyache J, Vela AI, Gibello A, Blanco MM, Briones V, González S, et al. *Lactococcus lactis* subsp. *lactis* infection in waterfowl: first confirmation in animals. *Emerg Infect Dis*. 2001;7(5):884–6. <https://doi.org/10.3201/eid0705.017519>.
37. Hadfield JD, Richardson DS, Burke T. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Mol Ecol*. 2006;15(12):3715–30. <https://doi.org/10.1111/j.1365-294X.2006.03050.x>.
38. Hammers M, Kingma SA, Spurgin LG, Bebbington K, Dugdale HL, Burke T, et al. Breeders that receive help age more slowly in a cooperatively breeding bird. *Nat Commun*. 2019;10(1):1301. <https://doi.org/10.1038/s41467-019-09229-3>.
39. Hansson B, Richardson DS. Genetic variation in two endangered *Acrocephalus* species compared to a widespread congener: estimates based on functional and random loci. *Anim Conserv*. 2005;8(1):83–90. <https://doi.org/10.1017/S1367943004001878>.
40. Henikoff S. Scores for sequence searches and alignments. *Curr Opin Struct Biol*. 1996;6(3):353–60. [https://doi.org/10.1016/S0959-440X\(96\)80055-8](https://doi.org/10.1016/S0959-440X(96)80055-8).
41. Hernández-Gómez O, Briggler JT, Williams RN. Influence of immunogenetics, sex and body condition on the cutaneous microbial communities of two giant salamanders. *Mol Ecol*. 2018;27(8):1915–29. <https://doi.org/10.1111/mec.14500>.
42. Heuberger C, Pott J, Maloy KJ. Why do intestinal epithelial cells express MHC class II? *Immunology*. 2021;162(4):357–67. <https://doi.org/10.1111/imm.13270>.
43. Hildebrand F, Gossmann TI, Frioux C, Özkurt E, Myers PN, Ferretti P, et al. Dispersal strategies shape persistence and evolution of human gut bacteria. *Cell Host Microbe*. 2021;29(7):1167–1176.e9. <https://doi.org/10.1016/j.chom.2021.05.008>.
44. Hildebrand F, Tadeo R, Voigt AY, Bork P, Raes J. LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome*. 2014;2(1). <https://doi.org/10.1186/2049-2618-2-30>.
45. Hughes AL, Yeager M. Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet*. 1998;32(1):415–35. <https://doi.org/10.1146/annurev.genet.32.1.415>.
46. Hutchings K. Parasite-mediated selection in an island endemic, the Seychelles warbler (*Acrocephalus Sechellensis*). University of East Anglia. School of Biological Sciences. 2009. https://scholar.google.com/scholar_lookup?hl=en&publication_year=2009&author=K.+Hutchings&title=Parasite%20mediated+selection+in+an+island+endemic%20C+the+Seychelles+warbler+%28Acrocephalus+s.
47. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. 2010. <http://www.biomedcentral.com/1471-2105/11/119>.
48. Jaeger BC, Edwards LJ, Das K, Sen PK. An R^2 statistic for fixed effects in the generalized linear mixed model. *J Appl Stat*. 2017;44(6):1086–105. <https://doi.org/10.1080/02664763.2016.1193725>.
49. Janeway CJ, Travers P, Walport M, et al. Immunobiology: the immune system in health and disease. 5th edition. (5th ed.). Garland Science. 2001.
50. Komdeur J. Cooperative breeding in the Seychelles warbler. University of Cambridge. 1991.
51. Komdeur J. Importance of habitat saturation and territory quality for evolution of cooperative breeding in the Seychelles warbler. *Nature*. 1992;358(6386). <https://doi.org/10.1038/358493a0>.
52. Komdeur J, Pels MD. Rescue of the Seychelles warbler on Cousin Island, Seychelles: the role of habitat restoration. *Biol Conserv*. 2005;124(1):15–26. <https://doi.org/10.1016/j.biocon.2004.12.009>.
53. Kubinak JL, Stephens WZ, Soto R, Petersen C, Chiaro T, Gogokhia L, et al. MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection. *Nat Commun*. 2015;6(1):8642. <https://doi.org/10.1038/ncomms9642>.
54. Kuhn KA, Stappenbeck TS. Peripheral education of the immune system by the colonic microbiota. *Semin Immunol*. 2013;25(5):364–9. <https://doi.org/10.1016/j.smim.2013.10.002>.
55. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol*. 2015;11(5):e1004226. <https://doi.org/10.1371/journal.pcbi.1004226>.
56. La Reau AJ, Strom NB, Filvaroff E, Mavrommatis K, Ward TL, Knights D. Shallow shotgun sequencing reduces technical variation in microbiome analysis. *Sci Rep*. 2023;13(1):7668. <https://doi.org/10.1038/s41598-023-33489-1>.
57. Larke-Mejia NL, Carrión O, Crombie AT, McGenity TJ, Murrell JC. *Sphingopyxis* sp. strain OPL5, an isoprene-degrading bacterium from the Sphingomonadaceae family isolated from oil palm leaves. *Microorganisms*. 2020;8(10):1557. <https://doi.org/10.3390/microorganisms8101557>.
58. Leclaire S, Strandh M, Dell'Arcidia G, Gabirot M, Westerdahl H, Bonadonna F. Plumage microbiota covaries with the major

- histocompatibility complex in blue petrels. *Mol Ecol*. 2019;28(4):833–46. <https://doi.org/10.1111/mec.14993>.
59. Lee C, Worsley SF, Davies C, Silan E, Burke T, Komdeur J, Hildebrand F, Dugdale H, Richardson D. Gut microbiome composition and function – including transposase gene abundance - varies with age, but not senescence, in a wild vertebrate. 2024. <https://doi.org/10.32942/X2PP8R>.
 60. Lahti L, Shetty S. microbiome R package. 2019.
 61. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674–6. <https://doi.org/10.1093/bioinformatics/btv033>.
 62. Lim JY, Yoon J, Hovde CJ. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. *J Microbiol Biotechnol*. 2010;20(1):5–14.
 63. Liu ZL, Liu J, Niu DK. Bacterial defenses and their trade-off with growth are not ubiquitous but depend on ecological contexts. 2024. <https://doi.org/10.1101/2024.03.24.586452>.
 64. Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, et al. Function and functional redundancy in microbial systems. *Nat Ecol Evol*. 2018;2(6):936–43. <https://doi.org/10.1038/s41559-018-0519-1>.
 65. McConnell SC, Hernandez KM, Andrade J, de Jong JLO. Immune gene variation associated with chromosome-scale differences among individual zebrafish genomes. *Sci Rep*. 2023;13(1). <https://doi.org/10.1038/s41598-023-34467-3>.
 66. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*. 2013;8(4). <https://doi.org/10.1371/journal.pone.0061217>.
 67. Metcalf CJE, Koskella B. Protective microbiomes can limit the evolution of host pathogen defense. *Evol Lett*. 2019;3(5):534–43. <https://doi.org/10.1002/evl3.140>.
 68. Montero BK, Uddin W, Schwensow N, Gillingham MAF, Ratovonamana YR, Rakotonandranary SJ, Corman V, Drosten C, Ganzhorn JU, Sommer S. Evidence of MHC class I and II influencing viral and helminth infection via the microbiome in a non-human primate. *PLoS Pathogens*. 2021;17(11). <https://doi.org/10.1371/journal.ppat.1009675>.
 69. Navale VD, Yadav R, Khilari A, Dharne M, Shanmugam D, Vamkudoth KR. Dietary supplementation of *Lactococcus lactis* subsp. *lactis* BIONCL17752 on growth performance, and gut microbiota of broiler chickens. *Probiotics and Antimicrobial Proteins*. 2024. <https://doi.org/10.1007/s12602-024-10313-4>.
 70. Niu XK, Narsing Rao MP, Dong ZY, Kan Y, Li QR, Huang J, et al. *Vulcanibacterium gelatinicum* sp. nov., a moderately thermophilic bacterium isolated from a hot spring. *Int J Syst Evol Microbiol*. 2020;70(3):1571–7. <https://doi.org/10.1099/ijsem.0.003934>.
 71. Oksanen J, Simpson Gavin L, Guillaume Blanchet F, Friendly M, KR, Legendre P, McGlinn D, Minchin PR, ORB, Simpson GL, Solymos P, Stevenes MHH, Wagner HH, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, et al. vegan: community ecology package. 2024.
 72. Oliver MK, Piertney SB. Selection maintains MHC diversity through a natural population bottleneck. *Mol Biol Evol*. 2012;29(7):1713–20. <https://doi.org/10.1093/molbev/mss063>.
 73. Özer O, Lenz TL. Unique pathogen peptidomes facilitate pathogen-specific selection and specialization of MHC alleles. *Mol Biol Evol*. 2021;38(10):4376–87. <https://doi.org/10.1093/molbev/msab176>.
 74. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res*. 2022;50(D1):D785–94. <https://doi.org/10.1093/nar/gkab776>.
 75. Perez-Sepulveda BM, Heavens D, Pulford CV, Predeus AV, Low R, Webster H, Dykes GF, Schudoma C, Rowe W, Lipscombe J, Watkins C, Kumwenda B, Shearer N, Costigan K, Baker KS, Feasey NA, Hinton JCD, Hall N, Perez-Sepulveda BM, et al. An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. *Genome Biol*. 2021;22(1). <https://doi.org/10.1186/s13059-021-02536-3>.
 76. Peschel S, Müller CL, von Mutius E, Boulesteix AL, Depner M. NetCoMi: network construction and comparison for microbiome data in R. *Brief Bioinform*. 2021;22(4). <https://doi.org/10.1093/bib/bbaa290>.
 77. Piertney SB, Oliver MK. The evolutionary ecology of the major histocompatibility complex. *Heredity*. 2006;96(1):7–21. <https://doi.org/10.1038/sj.hdy.6800724>.
 78. Posit team. RStudio: integrated development environment for R. (2023.12.1+402). Posit Software, PBC. 2024.
 79. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei T, Creevey C, Kuhn M, Jensen LJ, Von Mering C, Bork P. EggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res*. 2014;42(D1). <https://doi.org/10.1093/nar/gkt1253>.
 80. R Core Team. R: a language and environment for statistical computing (R version 4.3.3). R Foundation for Statistical Computing. 2024.
 81. Raj Pant S, Komdeur J, Burke TA, Dugdale HL, Richardson DS. Socio-ecological conditions and female infidelity in the Seychelles warbler. *Behav Ecol*. 2019;30(5):1254–64. <https://doi.org/10.1093/beheco/azr072>.
 82. Raj Pant S, Versteegh MA, Hammers M, Burke T, Dugdale HL, Richardson DS, et al. The contribution of extra-pair paternity to the variation in lifetime and age-specific male reproductive success in a socially monogamous species. *Evolution*. 2022;76(5):915–30. <https://doi.org/10.1111/evo.14473>.
 83. Reikvam DH, Erofeev A, Sandvik A, Grcic V, Jahnsen FL, Gaustad P, et al. Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. *PLoS ONE*. 2011;6(3):e17996. <https://doi.org/10.1371/journal.pone.0017996>.
 84. Ren T, Grieneisen LE, Alberts SC, Archie EA, Wu M. Development, diet and dynamism: longitudinal and cross-sectional predictors of gut microbial communities in wild baboons. *Environ Microbiol*. 2016;18(5):1312–25. <https://doi.org/10.1111/1462-2920.12852>.
 85. Richardson DS, Burke T, Komdeur J. Grandparent helpers: the adaptive significance of older, postdominant helpers in the Seychelles Warbler. *Evolution*. 2007;61(12):2790–800. <https://doi.org/10.1111/j.1558-5646.2007.00222.x>.
 86. Richardson DS, Jury FL, Blaakmeer K, Komdeur J, Burke T. Parentage assignment and extra-group paternity in a cooperative breeder: the Seychelles warbler (*Acrocephalus sechellensis*). *Mol Ecol*. 2001;10(9):2263–73. <https://doi.org/10.1046/j.0962-1083.2001.01355.x>.
 87. Richardson DS, Komdeur J, Burke T, von Schantz T. MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. *Proc Biol Sci*. 2005;272(1564):759–67. <https://doi.org/10.1098/rspb.2004.3028>.
 88. Richardson DAVIDS, Westerdahl H. MHC diversity in two *Acrocephalus* species: the outbred Great reed warbler and the inbred Seychelles warbler. *Mol Ecol*. 2003;12(12):3523–9. <https://doi.org/10.1046/j.1365-294X.2003.02005.x>.
 89. Rock KL, Reits E, Neefjes J. Present yourself! by MHC class I and MHC class II molecules. *Trends Immunol*. 2016;37(11):724–37. <https://doi.org/10.1016/j.it.2016.08.010>.
 90. Roland MM, Mohammed AD, Kubinak JL. How MHCII signaling promotes benign host-microbiota interactions. *PLoS Pathog*. 2020;16(6):e1008558. <https://doi.org/10.1371/journal.ppat.1008558>.
 91. Russell JT, Roesch LFW, Ördberg M, Ilonen J, Atkinson MA, Schatz DA, et al. Genetic risk for autoimmunity is associated with distinct changes in the human gut microbiome. *Nat Commun*. 2019;10(1):3621. <https://doi.org/10.1038/s41467-019-11460-x>.
 92. Silverman M, Kua L, Tanca A, Pala M, Palomba A, Tanes C, et al. Protective major histocompatibility complex allele prevents type 1 diabetes by shaping the intestinal microbiota early in ontogeny. *Proc Natl Acad Sci U S A*. 2017;114(36):9671–6. <https://doi.org/10.1073/pnas.1712280114>.
 93. Sparks AM, Spurgin LG, van der Velde M, Fairfield EA, Komdeur J, Burke T, et al. Telomere heritability and parental age at conception effects in a wild avian population. *Mol Ecol*. 2022;31(23):6324–38. <https://doi.org/10.1111/mec.15804>.
 94. Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc Biol Sci*. 2010;277(1684):979–88. <https://doi.org/10.1098/rspb.2009.2084>.
 95. Spurgin LG, Wright DJ, van der Velde M, Collar NJ, Komdeur J, Burke T, et al. Museum <scp>DNA</scp> reveals the demographic history of the endangered Seychelles warbler. *Evol Appl*. 2014;7(9):1134–43. <https://doi.org/10.1111/eva.12191>.
 96. Sun S, Jones RB, Fodor AA. Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. *Microbiome*. 2020;8(1). <https://doi.org/10.1186/s40168-020-00815-y>.

97. Tanoue T, Umesaki Y, Honda K. Immune responses to gut microbiota-commensals and pathogens. *Gut Microbes*. 2010;1(4):224–33. <https://doi.org/10.4161/gmic.1.4.12613>.
98. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*. 2000;28(1). <http://www.ncbi.nlm.nih.gov/COG>.
99. Toivanen P, Vaahtovuori J, Eerola E. Influence of major histocompatibility complex on bacterial composition of fecal flora. *Infect Immun*. 2001;69(4):2372–7. <https://doi.org/10.1128/IAI.69.4.2372-2377.2001>.
100. Toole DR, Zhao J, Martens-Habbena W, Strauss SL. Bacterial functional prediction tools detect but underestimate metabolic diversity compared to shotgun metagenomics in southwest Florida soils. *Appl Soil Ecol*. 2021;168. <https://doi.org/10.1016/j.apsoil.2021.10412>.
101. Uren Webster TM, Consuegra S, Hitchings M, Garcia de Leaniz C. Interpopulation variation in the Atlantic salmon microbiome reflects environmental and genetic diversity. 2018. <https://doi.org/10.1128/AEM>.
102. van Leeuwen P, Mykytczuk N, Mastromonaco GF, Schulte-Hostedde AI. Effects of captivity, diet, and relocation on the gut bacterial communities of white-footed mice. *Ecol Evol*. 2020;10(11):4677–90. <https://doi.org/10.1002/ece3.6221>.
103. Wadud Khan MA, Zac Stephens W, Mohammed AD, Round JL, Kubinak JL. Does MHC heterozygosity influence microbiota form and function? *PLoS ONE*. 2019;14(5). <https://doi.org/10.1371/journal.pone.0215946>.
104. Wan Y, Wu Y, Bian J, Wang X, Zhou W, Jia Z, et al. Induction of hepatitis B virus-specific cytotoxic T lymphocytes response in vivo by filamentous phage display vaccine. *Vaccine*. 2001;19(20–22):2918–23. [https://doi.org/10.1016/S0264-410X\(00\)00561-2](https://doi.org/10.1016/S0264-410X(00)00561-2).
105. Wilde J, Slack E, Foster KR. Host control of the microbiome: Mechanisms, evolution, and disease. *Science*. 2024;385(6706). <https://doi.org/10.1126/science.adj3338>.
106. Williams CE, Hammer TJ, Williams CL. Diversity alone does not reliably indicate the healthiness of an animal microbiome. *ISME J*. 2024;18(1). <https://doi.org/10.1093/ismej/wrae133>.
107. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20(1). <https://doi.org/10.1186/s13059-019-1891-0>.
108. Worsley SF, Davies CS, Lee CZ, Mannarelli M, Burke T, Komdeur J, Dugdale HL, Richardson DS. Longitudinal gut microbiome dynamics in relation to age and senescence in a wild animal population. *Mol Ecol*. 2024. <https://doi.org/10.1111/mec.17477>.
109. Worsley SF, Davies CS, Mannarelli ME, Hutchings MI, Komdeur J, Burke T, et al. Gut microbiome composition, not alpha diversity, is associated with survival in a natural vertebrate population. *Anim Microbiome*. 2021;3(1):84. <https://doi.org/10.1186/s42523-021-00149-6>.
110. Worsley SF, Davies CS, Mannarelli M-E, Komdeur J, Dugdale HL, Richardson DS. Assessing the causes and consequences of gut microbiome variation in a wild population of the Seychelles warbler. *Microbiome*. 2022;10(1):242. <https://doi.org/10.1186/s40168-022-01432-7>.
111. Worsley SF, Lee CZ, Versteegh MA, Burke T, Komdeur J, Dugdale HL, Richardson DS. Gut microbiome communities demonstrate fine-scale spatial variation in a closed, island bird population. *ISME Communications*. 2025. <https://doi.org/10.1093/ismeco/ycaf138>.
112. Worsley SF, Mazel F, Videvall E, Harrison XA, Björk JR, Wanelik KM. Probing the functional significance of wild animal microbiomes using omics data. 2024.
113. Wright DJ, Brouwer L, Mannarelli M-E, Burke T, Komdeur J, Richardson DS. Social pairing of Seychelles warblers under reduced constraints: MHC, neutral heterozygosity, and age. *Behav Ecol*. 2016;27(1):295–303. <https://doi.org/10.1093/beheco/arv150>.
114. Xiang R, Kelemen M, Xu Y, Harris LW, Parkinson H, Inouye M, et al. Recent advances in polygenic scores: translation, equitability, methods and FAIR tools. *Genome Med*. 2024;16(1):33. <https://doi.org/10.1186/s13073-024-01304-9>.
115. Yoshida GM, Lhorente JP, Correa K, Soto J, Salas D, Yáñez JM. Genome-wide association study and cost-efficient genomic predictions for growth and fillet yield in Nile tilapia (*Oreochromis niloticus*). *G3 Genes|Genomes|Genetics*. 2019;9(8):2597–607. <https://doi.org/10.1534/g3.119.400116>.
116. Yoshino Y. *Enterococcus casseliflavus* infection: a review of clinical features and treatment. *Infect Drug Resist*. 2023;16:363–8. <https://doi.org/10.2147/IDR.S398739>.
117. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. *Cell Res*. 2020;30(6):492–506. <https://doi.org/10.1038/s41422-020-0332-7>.
118. Zhu L, Wang J, Bahrndorff S. Editorial: The wildlife gut microbiome and its implication for conservation biology. *Front Microbiol*. 2021;12. <https://doi.org/10.3389/fmicb.2021.697499>.
119. Zoelzer F, Burger AL, Dierkes PW. Unraveling differences in fecal microbiota stability in mammals: from high variable carnivores and consistently stable herbivores. *Anim Microbiome*. 2021;3(1):77. <https://doi.org/10.1186/s42523-021-00141-0>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.