Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Immunogenetic Variation Shapes the Gut Microbiome in a Natural Vertebrate Population

**Charli S. Davies** ( charli.davies@yahoo.co.uk )
University of East Anglia https://orcid.org/0000-0001-9030-7820
**Sarah F. Worsley**
University of East Anglia
**Kathryn H. Maher**
The University of Sheffield
**Jan Komdeur**
RUG: Rijksuniversiteit Groningen
**Terry Burke**
The University of Sheffield
**Hannah L. Dugdale**
RUG: Rijksuniversiteit Groningen
**David S. Richardson**
University of East Anglia

**Research**

**Keywords:** Major Histocompatibility Complex, gut microbiome, microbial diversity, Acrocephalus sechellensis, genetic variation, life history

**Posted Date:** July 14th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-703361/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Version of Record:** A version of this preprint was published at Microbiome on March 8th, 2022. See the published version at https://doi.org/10.1186/s40168-022-01233-y.

# Abstract

**Background:** The gut microbiome (GM) can influence many biological processes in the host, impacting its health and survival, but the GM can also be influenced by the host's traits. In vertebrates, Major Histocompatibility Complex (MHC) genes play a pivotal role in combatting pathogens and are thought to shape the host's GM. However, despite this - and the documented importance of both GM and MHC variation to individual fitness - few studies have investigated the association between the GM and MHC in the wild.

**Results:** We characterised MHC class I (MHC-I), MHC class II (MHC-II), and GM variation in individuals within a natural population of the Seychelles warbler, *Acrocephalus sechellensis*. We determined how the diversity and composition of the GM varied with MHC characteristics, in addition to environmental factors and other host intrinsic traits. Our results show that the presence of specific MHC alleles, but not MHC diversity, influences both the diversity and composition of the GM in this population. We found that MHC-I alleles, rather than MHC-II alleles, had the greatest impact on the GM. GM diversity was negatively associated with the presence of three MHC-I alleles (*Ase-ua3, Ase-ua4, Ase-ua5*), and one MHC-II allele (*Ase-dab4*), while changes in GM composition were associated with the presence of four different MHC-I alleles (*Ase-ua1, Ase-ua7, Ase-ua10, Ase-ua11*). GM diversity was also positively correlated with genome-wide heterozygosity and varied with host age and field period.

**Conclusions**: These results suggest that components of the host's immune system play a role in shaping the GM of wild animals. Host genotype – specifically, adaptive immune receptor MHC-I alleles – can modulate the GM, although whether this occurs directly, or indirectly through effects on host health is unclear. To our knowledge this is the first-time individual MHC-I variation has been shown to be associated with differences in GM characteristics in a natural population. Importantly, it suggests that host–microbiome coevolution may play a role in maintaining functional immunogenetic variation within natural vertebrate populations.


# Background

Most animals harbour a complex microbial community including bacteria, archaea, viruses, and microbial eukaryotes, collectively known as the host's microbiome. Members of this diverse community have coevolved with their hosts over evolutionary time and, as a result, play a fundamental role in their host's biology and function [1]. Consequently, it is important to understand both the ecological and evolutionary processes that shape host-associated microbial communities [2]. This is particularly true of the vertebrate gut, where the gut microbiome (GM) contributes to many key biological processes in the host, from enabling nutrient uptake [3] to pathogen defence [4]. As such, studies on domestic or laboratory animals have demonstrated that disrupting the GM can have significant negative consequences for host health and survival [5]. However, the factors that govern the diversity and composition of the GM often remain unclear, particularly in natural populations where organisms are exposed to far greater levels of environmental and microbial complexity compared to domestic or laboratory animals [6].

Studies on wild vertebrates are now attempting to unravel the importance of extrinsic factors, such as the environment, season, and diet in contributing to GM variation among species, populations, and individuals [e.g., 7–9]. Within a population, individual variation in the GM has also been linked to a multitude of host factors including sex [10], age [11], body condition [12], cognition [13], sociality [14], and hormones [15, 16]. Host genetics may also play an important role in determining individual GM variation within a population [17]. For example, increased genetic relatedness predicts decreased GM dissimilarity [10, 18] and composition changes [19, 20].

Immunogenetic variation may play a particularly important role in driving differences in the GM across individuals (reviewed in [21]). Microbial complexity presents a unique challenge to the host immune system, which has evolved to prevent pathogenic microbes from proliferating, whilst still allowing beneficial microbes (i.e., those that form mutualistic and commensal interactions) to remain [22]. In humans, inappropriate immune responses can lead to detrimental compositional changes to the GM via a loss of diversity and stability, as well as an increase in the proliferation of pathogenic bacteria [23]. In turn, such changes in GM composition have been linked to serious health outcomes [5, 24], so accurate detection of pathogenic microbes by the host is essential. Variation in GM composition is associated with variation in immune receptor genes (which detect pathogens, inducing an immune response) [25], making them key candidate genes that might influence the relationship between host genotype and the GM. In particular, pattern recognition receptor genes involved in the innate immune response, including toll-like receptors (TLR), nucleotide-binding oligomerization-like receptors, RIG-I-like receptors, and C-type lectin receptors, play an important role in recognising and regulating the GM in humans and captive animals (reviewed in [25, 26]).

In vertebrates, Major Histocompatibility Complex (MHC) genes, which play a key role in pathogen detection in the adaptive immune response [27], may also shape GM variation. MHC genes code for receptor molecules that bind to specific non-self antigens, initiating an adaptive immune response. MHC class I (MHC-I) molecules, which are expressed on the surface of virtually all somatic cells, mainly recognise intracellular microbes, while MHC class II (MHC-II) molecules are present on antigen-presenting cells and mainly recognise extracellular microbes, including bacterial GM species [28]. Functional variants of MHC genes can confer differential pathogen recognition and directly affect fitness [29, 30], with pathogen-mediated selection thought to drive the extraordinary within- and among-population variation observed at the MHC [31, 32]. Individuals with different MHC genotypes are, therefore, likely to recognise different microbial species, which could contribute to the inter-individual GM variation seen in natural vertebrate systems. Likewise, comparing functional immunogenetic variation with the diversity and composition of the GM could provide further understanding of the selective pressures acting on the maintenance of host genetic variation, with the GM being one possible factor driving pathogen-mediated selection at MHC genes.

Various studies on captive animals have found links between individual MHC variation and GM composition [33–35]. However, captivity can profoundly alter an organism's microbiome [36] and the complexities associated with natural populations cannot be captured in such studies [6]. Few studies

have investigated associations between the microbiome and individual MHC variation in wild animals. Those that have, have all focused on MHC-II variation, as this is central to humoral immunity against extracellular microbes [37]. For example, in a population of threespine sticklebacks (*Gasterosteus aculeatus*), GM diversity and MHC-II diversity were negatively correlated, and changes in GM composition and diversity were associated with the presence of specific MHC-II alleles [38]. In contrast, in two species of giant salamander, Ozark hellbender (*Cryptobranchus alleganiensis bishopi*) and eastern hellbender (*C. a. alleganiensis*), although the composition of the cutaneous microbiome was associated with the presence of specific MHC-II alleles, cutaneous microbiome diversity and composition were positively correlated with individual MHC-II divergence [39]. Lastly, two studies on seabirds (Leach's storm petrel (*Oceanodroma leucorhoa*), and the blue petrel (*Halobaena caerulea*), found that variation in body surface (feather and glandular) bacterial communities that are responsible for the production of volatile cues involved in mate choice were associated with MHC-II variation [40, 41]. To our knowledge, no studies have tested for an association between MHC-I genes (or any other host immune genes) and the microbiome in wild systems. This is despite the fact that MHC-I variation could impact the GM indirectly, for example via differences in individual susceptibility to intracellular infections (such as viruses) that could then impact the health and/or GM of the host [42]. In this study we investigate the extent to which MHC-1 genes, MHC-II genes, *TLR3* genotypes, and genome wide heterozygosity are associated with GM characteristics in one population of Seychelles warblers (*Acrocephalus sechellensis*), alongside other host and environmental factors.

The natural, isolated population of Seychelles warbler on Cousin Island has been intensively monitored since 1985 [43, 44]. The age, sex, status and territory of nearly every individual is known, and corresponding DNA samples have been collected since 1990 [45]. This population harbours reduced genetic variation as a result of a past genetic bottleneck [46]. However, variation still exists at neutral loci across the genome [47] and, importantly, at both MHC-I [48] and MHC-II [49] loci, as well as at some TLR genes [50]. Differences in individual fitness have been linked with this genetic variation. For example, individual condition [51] and reproductive success [52] are negatively correlated with genome-wide homozygosity. Likewise, differential survival and reproductive success are associated with variation in the viral-sensing *TLR3* gene [45]. Lastly, survival is positively associated with a specific MHC-I allele, *Ase-ua4*, and MHC-I diversity overall [53], and the occurrence of extra-pair paternity is negatively related to MHC-I diversity of the social male [54]. But whether there is functional variation at MHC-II in the Seychelles warbler, and if this is related to variation in fitness remains unresolved.

Here, we aim to test whether immunogenetic variation is associated with individual GM variation in the Seychelles warbler. Specifically, we test if bacterial diversity and community composition are associated with individual MHC-I and MHC-II gene diversity, or the presence of specific alleles at MHC-I, MHC-II or *TLR3* loci. It is difficult to make clear predictions about such associations. Individual GM diversity might be negatively associated with immunogenetic variation [38] if this genetic diversity enables hosts to recognise and eliminate more bacterial species, meaning that a reduced diversity of mutualistic strains, or a few highly competitive strains, can persist in the gut. Alternatively, GM diversity might be positively associated with greater, or optimal, immunogenetic diversity. This could occur via two different pathways.

First, directly, with greater immunogenetic diversity helping to eliminate highly competitive (potentially pathogenic) bacterial taxa, while still tolerating a network of commensal and mutualistic bacterial species. Second, indirectly, with greater, or optimal, immunogenetic diversity conferring increased fitness or host health, which in turn has often been associated with greater GM diversity in humans [23]. We also predict that GM composition and diversity will differ in relation to the presence or absence of specific immune alleles, via differences in immunity and tolerance. This is expected to be most marked for MHC-II alleles, as these are expressed extracellularly on antigen-presenting cells that can extend into the gut lumen, and are therefore important in the recognition of gut microbes [22].

## Methods

# Study species and sample collection

The Seychelles warbler is a small insectivorous passerine, endemic to the Seychelles archipelago. They are facultative cooperative breeders, defending strict year-round territories [43]. The population on Cousin Island (4°20’S,55°40’E; 0.29km$^2$) has been extensively monitored since 1985, increasing in intensity from 1997 onwards [43, 44]. The Cousin Island population is small, with a carrying capacity of around 320 adult individuals, existing in 115 territories [43, 55, 56], and there is virtually no migration to or from Cousin [57]. Individuals can live for a maximum of 19 years, with a median post-fledging lifespan of 5.5 years [58]. A comprehensive population census is conducted bi-yearly during the major breeding season (June–September) in the south-east monsoon, and the minor breeding season (January–March) in the north-west monsoon [59]. Territory quality varies quantifiably within and between years [60]; thus, it is possible to separate the influence of environmental factors from host-intrinsic factors when investigating GM variation at the individual level.

Individuals are either caught as chicks in the nest, or by mist net. At first catch, each bird is given a metal British Trust for Ornithology (BTO) ring and a unique combination of three colour rings, allowing it to be individually identifiable. Birds are aged based on hatch date, behaviour or eye colour; grey eyes indicate an age < 5 months, light brown eyes are characteristic of sub-adults (5–12 months), and adults (> 12 months) have red-brown eyes [43, 61]. Blood samples (25 µl) are collected by brachial venipuncture and stored in 0.8 ml of absolute ethanol at either room temperature or 4°C.

Captured birds were placed into a clean bag immediately following capture. In the first major breeding season of our study (2017) this was a laundered cotton bag, however, for all following seasons, birds were placed into a single use plastic-lined, flat-bottomed paper bag containing a plastic tray covered by a metal grate, according to an established protocol [62]. The metal grate and tray were sterilised with a 10% bleach solution between use. Individuals were removed from the bag once they had defaecated, or after a maximum of 30 minutes. A sterile flocked swab was used to transfer faecal material into a sterile microcentrifuge tube containing 1 ml of absolute ethanol. If the bird defaecated outside of the bag or tray then a sample was collected. Control samples from possible sources of contamination such as the bag, grate, tray, and fieldworkers’ hands, were taken throughout each sampling season, using sterile flocked

swabs. Faecal samples were stored in the field at 4°C, before being transported to the lab, where they were stored at -80°C prior to extraction.

# Molecular methods

Genomic DNA (gDNA) was extracted from blood using the DNeasy blood and tissue kit (Qiagen) Individuals were genotyped at 30 polymorphic microsatellite loci [47, 56] and standardised individual microsatellite heterozygosity ($H_S$) was calculated using the R 3.6.1 function Genhet 3.1 [63]. Sex was determined via PCR [56, 64]. Variation at one non-synonymous SNP within the leucine-rich repeat domain of TLR3 exon 4 was determined following [45].

# *MHC sequencing and bioinformatics*

In total, 314 samples were MHC sequenced, including 229 samples from individuals that had microbiome data and 31 samples from individuals previously MHC screened at either MHC-I [65] or MHC-II [49] using older techniques. The remainder included 30 repeated samples, 23 negative controls (making up at least 5% of each plate), and four samples (one per plate) from one great reed warbler (*Acrocephalus arundinaceus*) individual to serve as a positive control.

MHC-I exon 3 and MHC-II exon 2 were amplified and sequenced using previously validated primer sets [48, 49] (Additional file 1: Table S1), with the addition of Illumina index sites. Additionally, six random hexamers (N) were added to the first round PCR (PCR1) primers to increase diversity and improve cluster separation [66]. The two primer pairs amplifying MHC-I each included a motif-specific primer situated within exon 3, and one general primer situated in intron 3, and so amplified 262 bp of the full exon (274 bp). These primers had been designed to preferentially amplify functional variants, while avoiding known pseudogenes [48, 67]. The primers for MHC-II, situated within the flanking introns 1 and 2 of exon 2, amplify a 291 bp fragment. These sequences were then edited to the 270 bp MHC-II exon 2 [49]. The term 'allele' is used to describe the different variants amplified for each class of MHC, consistent with other publications investigating MHC diversity; however, alleles cannot be assigned to specific (duplicated) loci within each MHC class. Previous work suggests that, in the Seychelles warbler, there are a minimum of four duplicated MHC-I loci and six MHC-II loci [48, 49]. Sequencing of the MHC-I and MHC-II exons was carried out using 2x 250-bp paired-end sequencing on an Illumina MiSeq platform (Illumina, San Diego) (see Additional file 1: Supplementary methods for details).

Processing and MHC genotyping of raw Illumina MiSeq data was conducted using the Amplicon Sequencing Assignment (AmpliSAS) tool [68]. First, FastQC was used to check read quality, before merging pair-ended reads together using AmpliMERGE (10,257,832 merged sequences). AmpliCLEAN was then used to remove low-quality reads (Phred score of <30) and any that were missing either primers or barcodes (e.g., from residual PhiX). MHC-I and MHC-II sequences were separated at this stage, resulting in 3,044,897 raw MHC-I sequences, and 6,144,575 raw MHC-II sequences. Following this step, all

downstream bioinformatics and analysis were conducted separately for MHC-I and MHC-II. Cutadapt 1.6 [69] was used to remove MHC-specific primers, the six random hexamers and short reads (<100 bp). For MHC-II sequences, remaining intron regions were also removed, leaving a 270-bp fragment spanning the full exon. AmpliCHECK was used for preliminary data exploration, using Illumina-based default settings. Finally, AmpliSAS was used for demultiplexing, clustering, and filtering reads. First, a subset of 30 duplicated samples were used to optimise parameters for MHC-I and MHC-II, testing both minimum dominant frequency settings for the clustering step, and minimum amplicon depth for the filtering step, as recommended in [68]. Based on these results (Additional file 1: Table S2, S3) Illumina default clustering settings were used (1% substitution errors, 0.001% indel errors, 25% minimum dominant frequency) for both MHC classes. For the filtering step, chimaeras and sequences that only appeared in one sample were removed, and the minimum amplicon frequency was set as 1.6% for MHC-I, and 1.8% for MHC-II. This resulted in 1,267,410 raw MHC-I sequences, and 1,385,049 raw MHC-II sequences. Due to computational limitations, the MHC-II dataset was split into two halves and analysed using AmpliSAS separately, before being combined using AmpliCOMBINE in the web version of AmpliSAS.

For MHC-I, the majority of putative alleles were 262 bp, although three sequences that were <262 bp were present in >80% individuals – these were not homologous to any known MHC gene when checked using blastn. These shorter fragments were removed from downstream analysis. The majority of MHC-II putative alleles were the full 270 bp length, although there were also sequences between 267–269 bp, which were similar to MHC genes (see results). All MHC-II sequences <267 bp in length were not similar to any known MHC genes and, as putative sequencing artifacts, were removed from downstream analysis. MHC-I and MHC-II putative alleles were first checked against all known Seychelles warbler alleles. Any unknown putative alleles were then checked against the GenBank (NCBI) nucleotide database (accessed on 25$^{th}$ June 2020) to assess similarity to known MHC alleles from other related species. Additionally, samples of insufficient read depth based on rarefaction curves, which equated to a minimum read depth of 150 per amplicon for MHC-I, and 100 per amplicon for MHC-II, were removed. For 30 individuals sequenced twice to confirm repeatability, the sample with the greatest read depth was retained. After processing, the total number of reads assigned to an allele was 1,071,525 for MHC-I (mean ± SEM = 4391.5 ± 149.3 per sample) and 1,123,211 for MHC-II (mean ± SEM = 4603.3 ± 888.2 per sample) in the Seychelles warbler.

## *Microbial extraction, sequencing, and bioinformatics*

In total, 400 faecal samples were sequenced across three sequencing runs (two plates per sequencing run). This included 343 unique faecal samples (from 235 individuals) and 14 control samples, the latter of which included six extraction negative controls, four positive controls (using a microbial community standard), and four sampling controls. Additionally, 43 faecal samples were sequenced twice (20 within the same run and 23 across different runs) to determine sequencing repeatability.

Faecal samples were centrifuged for 10 minutes at 10,000 rpm and the supernatant was removed. To remove any residual ethanol, the resulting pellet was washed with 100 µl RNase/DNA-free molecular grade water by centrifuging at 10,000 rpm for 10 minutes. The supernatant was then removed, and the washing step repeated a further two times. Microbial DNA was then extracted from 0.05–0.1 g of each sample using the DNeasy PowerSoil Kit (Qiagen), according to an optimised version of the manufacturer's instructions. Modifications consisted of a heat block step (65°C for 10 minutes) prior to bead-beating, and elution of DNA in a final volume of 60 µl elution buffer. A ZymoBIOMICS microbial community standard (D6300) was extracted as a positive control using a ZymoBIOMICS DNA miniprep kit (Zymo Research), according to the manufacturer's instructions.

Extracted gDNA was quantified using a Qubit 4.0 Fluorometer (Invitrogen) with a Qubit dsDNA HS assay kit (Invitrogen). Aliquots of gDNA were shipped on dry ice to the Centre for Genomic Research, University of Liverpool for library preparation, pooling and sequencing. Bacterial barcoding was performed with a 2-step amplification process using the primers 515F (5'TGCCAGCMGCCGCGGTAA3') and 806R (5'GGACTACHVGGGTWTCTAAT3') [70], which amplify the V4 region of the 16S rRNA gene (see Additional file 1: Supplementary methods for details). Paired-end sequencing was carried out using 2x 250-bp paired-end sequencing on an Illumina MiSeq platform (Illumina, San Diego).

For each sequencing run, raw reads were first trimmed using Cutadapt 1.2.1 [69] to remove Illumina adapter sequences. Reads were further trimmed using Sickle 1.200 with a minimum window quality score of 20, resulting in totals of 12,308,047, 9,397,303 and 9,831,508 demultiplexed reads for the three runs, respectively (mean ± SEM per sample: 102,567.1 ± 10,454.8, 67,123.6 ± 6,633.1, 70,225.1 ± 5,423.5). Sequences were then analysed using QIIME2 2019.10 [71]. Based on overall quality scores, the first 10 bases of each read were trimmed, and sequences truncated to 210 bp for both forward and reverse reads. The DADA2 plugin 2019.10.0 was used to join paired-end reads, denoise, remove chimeras and residual PhiX reads, dereplicate and call amplicon sequence variants (ASVs) [72, 73]. Following this, results from the three separate runs were merged, resulting in a total of 22,997,693 reads (mean ± SEM per sample: 57,494.3 ± 3424.8) with 36,182 ASVs. A mid-point rooted phylogeny was then constructed using the masked alignment MAFFT [74] and the Fast Tree approach [75]. Taxonomic assignment of ASVs was performed by training a naïve-Bayes classifier on the SILVA 132 16S dataset using 99% sequence similarity [76, 77]. Plastid-like and archaeal sequences were removed, as well as singletons which likely represent sequencing errors. Additionally, two ASVs – one from the genus *Defltia* (relative abundance of 90.5% in a negative extraction control from the first run) and one from the genus *Limnobacter* (relative abundance of 99.9% in a negative extraction control from the third run) were removed as probable contaminants. This resulted in a total of 21,904,965 reads (mean ± SEM per sample: 54,899.7 ± 3,429.5) with 35,428 ASVs. The final sample metadata, ASV and taxonomy tables were all exported from QIIME2 into R 3.6.1 where they were processed using phyloseq 1.28.0 [78]. Sample completeness and rarefaction curves were generated using iNEXT 2.0.20 [79]; completeness plateaued at approximately 10,000 reads and 34 samples (including all six negative extraction controls) with fewer reads were excluded from downstream analyses. Overall, 320 unique samples (93%) were retained from 224 individuals.

The repeatability of GM sequencing was tested by comparing the 37 samples that were sequenced multiply within and across sequencing runs. Euclidean dissimilarity between pairs of samples was compared using one metric of alpha diversity (the Shannon diversity index), and two metrics of beta diversity (unweighted and weighted UniFrac of between and within duplicated samples) using Kruskal–Wallis tests.

# Statistical analyses

Unless otherwise stated, all analyses were conducted in R 3.6.1.

To characterise the Seychelles warbler GM, samples sequenced twice for repeatability analysis (sample duplicates) were filtered such that only the sample with the greatest read-depth was retained for downstream analysis. Samples collected from the same bird during the same catch session (catch duplicates) were also filtered to retain the single sample with the smallest potential exposure to external contamination, i.e., samples collected from cleaned trays were prioritised over those collected from other substrates, then the sample with the highest read depth was prioritised. The removal of sample and catch duplicates resulted in 281 samples (from 224 individuals). For all alpha diversity, beta diversity and differential abundance analyses, microbiome samples taken from chicks were excluded due to a small sample size ($n$ = 11). Individuals with incomplete MHC genotype data ($n$ = 25) were also excluded. Lastly, to prevent pseudo-replication, where an individual had multiple samples taken at different capture events, a single sample was selected at random, giving an overall sample size of 195 samples from 195 individuals from Cousin Island. This resulted in a total of 10,998,587 reads (mean ± SEM per sample: 56,403 ± 4181.0) with 35,428 ASVs in the non-rarefied dataset.

## *Alpha diversity*

All 195 samples were rarefied to a depth of 10,000 reads, based on sample completeness curves, leaving a total of 1,950,000 reads and 27,861 ASVs across samples in the rarefied dataset. Analyses were run using both rarefied and non-rarefied data, however, as results were comparable between datasets and library size was highly variable across samples, only the outcome of models using the rarefied dataset are presented here. Three metrics of alpha diversity were calculated: Chao1 [80] (a measure of microbial species richness), Shannon diversity index [81] (a measure of species richness, taking into account sample evenness), and Faith’s phylogenetic diversity index [82] (a measure of the phylogenetic diversity of a sample). Chao1 and Shannon diversity indices were calculated using phyloseq 1.28.0 [78], and Faith’s phylogenetic diversity was calculated using btools 0.0.1 [83]. Both Chao1 and Faith’s phylogenetic diversity were log-transformed in models to improve residual fit.

Linear models with a Gaussian distribution were constructed using glmmTMB 0.2.3 [84] to determine whether the alpha diversity of the GM differed with: (1) the presence or absence of individual immune genes, and (2) immune gene diversity. The first set of models included the presence or absence of all

MHC-I and MHC-II alleles that were present in at least 15% of sampled individuals, and that were the correct length (see above). This included the following alleles: *Ase-dab3, Ase-dab4, Ase-dab5, Ase-ua1/10, Ase-ua3, Ase-ua4, Ase-ua5, Ase-ua6, Ase-ua7, Ase-ua8, Ase-ua9* and *Ase-ua11 (Ase-ua1* and *Ase-ua10* were perfectly correlated, so only *Ase-ua1* was included in analyses). The second set of models contained MHC-I diversity, MHC-II diversity and the squares of each of these terms, since optimal, rather than a greater diversity, of MHC alleles could be more beneficial [85]. Both sets of models also included individual heterozygosity ($H_s$), and *TLR3* genotype ($TLR3^{AA}$, $TLR3^{AC}$ or $TLR3^{CC}$). The field period sampled (major 2017, major 2018 or minor 2018), sex (male, female), and age (fledgling, old fledging, sub-adult or adult) were also included in models, as these factors have been shown to influence GM variation in other studies. Alpha diversity (Shannon, log Chao1, or log Faith) was entered as the response variable. In all models, continuous factors were standardised (scaled and centred) using arm 1.10-1 [86]. All biologically relevant interactions were initially included in models but were removed prior to model averaging to enable interpretation of first-order effects, as all were non-significant ($P < 0.1$). Field period and territory quality were correlated (linear model; $F_{2,185} = 117.2$, $P < 0.001$), so only field period was included in the models. Collinearity between independent variables was tested using variance inflation factors ensuring an upper limit of three. Collinearity between the presence and absence of immune alleles was further assessed using GGally 2.0.0 [87]. DHARMA 0.2.4 [88] was used to confirm that there was no over or under dispersion, or residual spatial or temporal autocorrelation in the models. Model averaging – an information-theoretic approach using the dredge function in MUMIn 1.43.6 [89] – was used to select plausible models. All models within seven AICc of the top model were included in the averaged model, to obtain the final conditional model [90].

## Beta diversity

The unrarefied dataset was further filtered to remove ASVs that appeared in fewer than five samples, and that had a total read count of <50 across samples, based on an assessment of overall ASV prevalence and abundance (Additional file 1: Fig S1). Overall, 1,944 out of 35,428 ASVs were retained following filtering. To account for uneven sequencing depth across samples, reads were normalised using the cumulative sum scaling function [91] in metagenomeSeq 1.26.3 [92]. Two beta diversity metrics that incorporate phylogenetic distance were then calculated using phyloseq 1.28.0 [93]; these were unweighted UniFrac distance (based on the presence/absence of microbial taxa) and weighted UniFrac distance (a quantitative measure which also accounts for differences in the abundances of microbial taxa) [94]. Marginal Permutational Analysis of Variance tests (PERMANOVAs) were used to assess whether GM community composition differed with immune gene characteristics, using the adonis2 function in Vegan 2.5.6 [95] with 10,000 permutations. As with alpha diversity models, two sets of PERMANOVA tests were constructed for each beta diversity metric, with the first set of models including the presence/absence of MHC alleles, and the second set of models including MHC diversity; other variables were included as in alpha diversity models. To clarify whether significant differences detected in PERMANOVA tests were caused by differences in mean values, rather than variation in dispersion across

groups [96], homogeneity of group dispersions was tested using the betadisper function in Vegan 2.5.6 [95]. Principle coordinate analysis (PCoA), based on weighted and unweighted UniFrac distances, was used to visualise the differences in composition between groups.

### *Differential abundance analysis*

To assess whether particular ASVs were differentially abundant between groups of individuals with different immune gene characteristics, DESeq2 1.24.0 [97] was used. For this analysis, unrarefied reads were filtered but untransformed, as DeSeq2 uses its own variance stabilising transformation to account for variation in library size across samples. DeSeq2 estimates the log2-fold change in microbial abundance between sample groups using a negative binomial distribution to model ASV counts. Only variables that were associated with significant compositional shifts in PERMANOVA tests (see above) were included in this analysis to avoid over-parametrization (Additional file 1: Table S4). To account for the large number of zero counts for individual ASVs, the “poscounts” estimator was included when estimating size factors. Differential ASV abundance was assessed using negative binomial Wald tests and $P$ values were adjusted using the Benjamini and Hochberg false-discovery rate correction, with a significance cut-off of $P < 0.01$. Two ASVs did not converge due to a high number of zero counts across samples; these were removed from the analysis.

## Results

### *Seychelles warbler GM profile*

The overall bacterial phyla and class profile of the Seychelles warbler GM was similar to other passerine bird species [98]. We identified 40 bacterial phyla across the 281 samples combined; however, of these, *Proteobacteria* (42% of total reads), *Firmicutes* (22%) and *Actinobacteria* (17%) dominated, with all other phyla being present at lower relative abundances (summing to <5% of the total read count). The dominant bacterial classes were *Gammaproteobacteria* (25%), *Alphaproteobacteria* (16%), *Actinobacteria* (16%), *Bacilli* (16%) and *Clostridia* (6%) (Fig 1).

The core microbiome was further characterised at the family level by extracting bacterial families that appeared in at least 50% of samples with a minimum relative abundance of 0.1% (Additional file 1: Table S5). This resulted in the detection of 28 core families, with ASVs from these families making up 74% of all reads. Of the core families, eight were present in at least 80% of samples, and four accounted for >5% of all reads; this latter group consisted of *Enterobacteriaceae* (23% of total reads), *Streptococcaceae* (10%), *Rhizobiaceae* (6%) and *Enterococcaceae* (5%). Of the assigned genera, 20 were present in at least 50% of samples and ASVs from these genera made up a total of 28% of all reads. However, of these, only two genera (*Microbacterium* and *Enterococcus*) were present in more than 80% of samples.

Despite the presence of a core microbiome, the abundance of individual bacterial taxa was highly variable across individuals (Fig 1). Additionally, there was considerable individual variation in alpha

diversity when measured as Chao1 (mean = 323.2 ± 14.63 SEM), Shannon (mean = 4.0 ± 0.07), and Faith's phylogenetic diversity (mean = 18.8 ± 0.62).

Repeatability of GM sequencing for the same sample was tested using three metrics of diversity (Shannon, unweighted, and weighted UniFrac). As expected, pairwise distances between samples were significantly greater (or more dissimilar) than within-sample comparisons, i.e., pairwise distances when gDNA from the same sample was sequenced twice ($n$ = 37, P < 0.001) (Additional file 1: Fig S2).

## *MHC characteristics*

244 individuals were successfully genotyped at MHC-I exon 3 and MHC-II exon 2 genes. The repeatability of MHC-I genotypes was 95.0% and repeatability of MHC-II was 90.1%, based on 26 and 24 duplicate samples, respectively (Additional file 1: Table S2, S3). The great reed warbler positive control sample had four MHC-I and four MHC-II alleles – all of which mapped with 100% similarity to previously sequenced great reed warbler MHC alleles. Barring one MHC-I sample, which was adjacent to a positive control sample during sequencing and was subsequently removed from the analysis, no other Seychelles warbler samples contained these alleles.

On average, individuals had 5.0 MHC-I alleles: 2–7 alleles per individual. Of these, 10 MHC-I alleles were present in >5% but <95% individuals, and another 10 were present in ≤5% of individuals (Fig 2). Comparing these 20 alleles to previously sequenced data [48], nine of the ten common alleles matched previously sequenced alleles, with an average of 98% sequence similarity. One allele, *Ase-ua2*, was not present in the current sequencing cohort. When comparing 29 individuals also genotyped using reference strand-mediated conformation analysis [65], and excluding *Ase-ua2*, there was 95% similarity between genotyping methods.

Including all MHC-II alleles, individuals had on average 5.8 alleles (range 3–11) out of a total of 24 alleles (Fig 2a). However, of these 24 MHC-II alleles, only 14 were of the full exon length (270 bp), six alleles had a 1-bp deletion (269 bp in length), two alleles had a 2-bp deletion (268 bp) and two alleles had a 3-bp deletion (267 bp). Of the 10 alleles which contained indels, three of these also contained stop codons, and all were missing the Cys74 residue, which along with Cys10 residue creates a disulfide bridge, which is important for conformation of the mature MHC protein; therefore, these alleles were removed as putative pseudo- or non-functional alleles. Concentrating on putatively functional MHC-II alleles, there were 2.9 alleles on average per individual (range 1–5 alleles per individual). Of these, only three were present in >5% but < 95% of individuals (Fig 2b). Of the other putatively functional alleles, two were present in virtually all individuals, while nine alleles had a frequency <5%.

For downstream analysis, only alleles of the full, correct length (i.e., MHC-I: 262 bp, MHC-II: 270 bp) were included when calculating diversity or for presence/absence. Again, for the presence/absence of alleles, only alleles present in >5% but <95% of individuals were included. Of those alleles included in the final

presence/absence analyses, all 10 of the MHC-I and three of the MHC-II alleles translated into unique amino acid sequences.

## *The effect of MHC and other host variables on GM alpha diversity*

The presence of four MHC alleles was associated with reduced diversity and richness of the GM (Fig 3). Individuals with the *Ase-ua5* allele had significantly lower alpha diversity for all calculated metrics, compared to individuals without a copy (Additional file 1: Table S6, Fig 3), indicating that *Ase-ua5* negatively influences the richness, evenness and the phylogenetic diversity of the GM. Presence of the *Ase-ua3* allele was also associated with a decrease in both Shannon diversity and Chao1 richness (Additional file 1: Table S6a, Fig 3), but there was no significant difference in the phylogenetic diversity of the GM between individuals with or without the *Ase-ua3* allele (Additional file 1: Table S6a, Fig 3). Presence of the *Ase-ua4* allele was associated with reduced GM richness (Additional file 1: Table S6a, Fig 3) and this effect approached significance when taking phylogenetic diversity into account (log Faith's phylogenetic diversity: $P = 0.059$, Additional file 1: Table S6a, Fig 3). None of the remaining MHC-I alleles or the *TLR3* genotype was associated with alpha diversity (Additional file 1: Table S6a, Fig 3). Likewise, most MHC-II alleles were not associated with changes in GM diversity. However, the presence of one allele, *Ase-dab4*, was associated with a reduction in Shannon diversity (Additional file 1: Table S6a, Fig 3), but not Chao1 richness or Faith's phylogenetic diversity (Additional file 1: Table S6a, Fig 3). There was no significant effect of MHC-I or MHC-II diversity, or diversity$^2$ on alpha diversity (Additional file 1: Table S6b). By contrast, individual heterozygosity was positively associated with Shannon diversity (Additional file 1: Table S6a, Fig 3), but not Chao1 or Faith's phylogenetic diversity, though these both showed the same pattern.

Regarding other factors, males had a borderline tendency to have reduced richness compared to females (logChao1 $P = 0.062$, log Faith $P = 0.066$, Additional file 1: Table S6, Fig 3). There was also an association between GM diversity and age, with old fledglings having reduced Shannon diversity and Chao1 richness compared to all other age classes (Additional file 1: Table S6, Fig 3), although the phylogenetic diversity of the GM did not change between age classes (Additional file 1: Table S6, Fig 3). There was no association between GM diversity and field period (Additional file 1: Table S6a, Fig 3), suggesting that environmental variation across field periods had little influence on the observed variation in alpha diversity values across individuals.

## *The effect of host variables on GM composition*

In addition to effects on alpha diversity, compositional differences in the GMs of individuals with, or without, specific MHC-I alleles were identified, although the alleles that showed an effect were not the same as those associated with shifts in GM alpha diversity. PERMANOVA tests showed that the overall

composition of the GM was significantly different for individuals with the *Ase-ua11* allele versus those without it (Additional file 1: Table S4, Fig 4a). GM composition was also significantly different for individuals possessing either the *Ase-ua7* allele (Additional file 1: Table S4, Fig 4b) or for *Ase-ua1* (and *Ase-ua10* as these alleles were co-occurring) (Additional file 1: Table S4, Fig 4c) compared to individuals without these alleles, but only when weighted UniFrac distances were used as a beta diversity metric. None of the differences detected in PERMANOVA tests were due to differential dispersion (all betadisper tests: $P > 0.05$), indicating that the results reflected differences in mean values across groups. However, although statistically significant, the presence and absence of specific alleles only explained a minimum of 1.4% and a maximum of 1.7% (per allele) of the variation in GM composition, suggesting that each allele only had a small individual influence on overall GM composition (Additional file 1: Table S4). The remaining MHC-I and MHC-II alleles, as well as MHC-I and MHC-II diversity (or diversity$^2$) had no effect on GM composition (Additional file 1: Table S4). Additionally, *TLR3* genotype and $H_s$ were not associated with any of the beta diversity metrics (Additional file 1: Table S4).

Looking at other host factors, age class was associated with a compositional shift in the GM in PERMANOVAs based on unweighted UniFrac distances (Additional file 1: Table S4) and explained 1.9% of the variance in GM composition. Based on the Principal Coordinate analysis plots, this difference was due to old fledglings being slightly more differentiated compared to other age classes (Additional file 1: Fig S3). However, this effect was absent in models based on weighted UniFrac, which takes the abundance of ASVs into account (Additional file 1: Table S4). This suggests that the changes in composition with age class may be due to differences in the presence/absence of different bacterial taxa in the GM, rather than differing abundances of the same taxa. There were no differences in beta diversity between males and females (Additional file 1: Table S4). Focusing on the influence of extrinsic factors on the GM, there were significant compositional differences in the GM between field periods, which overall explained either 1.7% or 2.0% variance for unweighted and weighted UniFrac distance, respectively (Additional file 1: Table S4).

## *The influence of host variables on the abundance of specific ASVs*

The co-occurring *Ase-ua1/10* alleles were associated with the greatest change in ASV abundance, with 32 ASVs (across 15 bacterial orders) being significantly more abundant when the alleles were absent and 70 ASVs (across 33 orders) being more abundant when they were present (Fig 5c). Fewer taxa were differentially abundant between groups of individuals with/without *Ase-ua11*. In this instance, 32 ASVs (across 17 orders) were significantly more abundant when the allele was absent, and 12 ASVs (across 7 orders) were more abundant when the allele was present (Fig 5a). Overall, 29 ASVs were significantly more abundant when the allele *Ase-ua7* was present and 22 ASVs were more abundant when the allele was absent (Fig 5b).

Many ASVs were significantly more abundant in old fledglings compared to other age groups (old fledglings compared to fledglings: 149 vs 26; old-fledgling compared to sub-adults: 133 vs 57; old-fledglings compared to adults: 169 vs 19; Additional file 1: Fig S4). In comparison, the numbers of differentially abundant taxa between other age groups were more even (fledglings compared to sub-adults: 55 vs 48; fledglings compared to adults: 22 vs 35; sub-adults compared to adults: 104 vs 22; Additional file 1: Fig S4).

Concentrating on extrinsic associations with GM, 229, 225, and 146 ASVs were significantly differentially abundant between the three field periods, respectively (Additional file 1: Fig S5). The majority of ASVs were overrepresented in the minor 2018 season compared to either major season (166 in the minor 2018 vs 59 in the major 2017 season, and 192 in the minor 2018 season vs 37 in the major 2018 season). Of these, 150 ASVs from the minor season differed in abundance across analyses.

## Discussion

In this study, we screened both GM variation and the MHC class I and II characteristics of individuals in a natural population of the Seychelles warbler. This enabled us to assess how the diversity and composition of the bacterial component of the GM is associated with individual immunogenetic variation, i.e., MHC and *TLR3* genotypes. Our results indicate that differences in GM diversity are associated with the presence of certain MHC alleles (specifically, lower diversity is associated with four of the 13 tested MHC alleles). Furthermore, differences in GM composition are associated with the presence/absence of four (different) MHC-I alleles, including the differential abundance of certain bacterial taxa. While we found no effect of MHC diversity or *TLR3* genotype on GM diversity or composition, we did find a positive association between bacterial GM diversity and individual genome-wide heterozygosity. Lastly, GM characteristics were also associated with several other extrinsic and host specific variables, namely sex, age and sampling period.

# MHC variation in the Seychelles warbler

Although previous studies have already characterised MHC variation in the Seychelles warbler [48, 49], here we screened variation at the MHC-I exon 3 and MHC-II exon 2 genes using next-generation sequencing for the first time. We found reduced functional allelic diversity at MHC-II compared to MHC-I, consistent with what has been found in other passerines [99]. Previous studies on the Seychelles warbler have provided evidence that balancing selection has maintained variation at both the MHC-I [48] and MHC-II [49]. However, the latter study did not fully resolve individual MHC-II characteristics because of difficulties with the cloning and reference strand-mediated conformation analysis techniques used. In the present study, we were able to confirm (class-I) and fully characterise (class-II) individual MHC variation. Our results, showing that variation has been maintained at both sites in this species despite reduced genome-wide variation [46], support the idea that balancing selection is maintaining variation at both MHC classes. Given the MHC's role in pathogen detection, this is likely to be pathogen-mediated selection.

## MHC variation shapes GM variation

We found that GM diversity was negatively associated with the presence of three (out of 10) MHC-I alleles (*Ase-ua3, Ase-ua4, Ase-ua5*), and one (out of three) MHC-II alleles (*Ase-dab4*). All three MHC-I alleles were consistently associated with a reduction in GM bacterial richness and *Ase-ua5* was additionally associated with reduced phylogenetic diversity in the GM. This suggests that these alleles may lead to the selective elimination of certain bacterial taxa from the gut, resulting in a reduced community of species with a narrower phylogenetic range. This is similar to the findings of another study that identified MHC-II motifs associated with reduced GM diversity in natural threespine sticklebacks [38]. It is likely that MHC-II alleles – such as *Ase-dab4* in the Seychelles warbler – directly impact GM diversity, since MHC-II molecules are produced in antigen-presenting cells, which are abundant in the lamina propria behind the gut epithelium [22]. Such dendritic cells can extend between epithelial cells where they phagocytose particles, including microbes, from the gut lumen [100]. Antigens from these microbes are then exported to the cell surface by MHC-II molecules, so that they can be presented to B- and T-cell populations [101] and instigate an immune response.

Our study expands on previous work by also investigating how variation at MHC-I genes impact GM variation. To our knowledge, this is the first-time differences in GM characteristics have been associated with individual MHC-I variation in a natural population. MHC-I molecules typically respond to intracellular pathogens, rather than extracellular microbes, and play a central role in anti-viral and anti-tumour immunity [28]. As such, we would not necessarily expect these molecules to recognise bacterial antigens, although cross reactivity via the presentation of exogenously derived antigens can occur [102]. Instead, MHC-I variation may indirectly affect GM characteristics by impacting other aspects of the host's health and physiology. In the Seychelles warbler, we know that survival is associated with a specific MHC-I allele (*Ase-ua4*) [53] and, although we do not know what drives this effect (i.e. we have not identified the host–pathogen interaction responsible), any such interaction could also shape changes in the GM. In our study, the *Ase-ua4* allele was associated with a reduction in GM richness but did not significantly change the composition of the GM. Numerous studies have linked MHC-I variation with susceptibility to malarial infection in passerines [103, 104] and other taxa [29]. Malaria infections alter the GM, via disruption to immune homeostasis [105] and could play a role in the Seychelles warbler in which a single strain of the malarial parasite *Haemoproteus nucleocondensus* has been identified [106]. Alternatively, MHC-I might alter an individual's susceptibility to another, as yet unidentified pathogen, such as a virus, that could also drive differences in the diversity of the GM [42, 107].

Given the negative, but lack of positive, associations identified between GM alpha diversity and MHC alleles, it is surprising that there was no significant effect of overall MHC diversity on the GM. One might expect the cumulative negative effect of each allele (Fig. 3) to cause at least a weak negative correlation between GM diversity and MHC diversity. However, given the multitude of factors involved in determining both the host's GM and immune response, this lack of association between GM and MHC diversity could simply be due to limited power to detect weak effects, as is often the case when examining associations between immunocompetence and MHC variation [108]. Assessing how the functional divergence of MHC

alleles within an individual – which provides information about the range of antigens that can be detected [109] – rather than just the number of MHC genes has provided additional resolution in other MHC studies, [e.g.,110]. In future, this approach could provide extra clarity, particularly when considering the diversity of bacterial taxa that are present within the GM.

We also observed compositional differences in the GM associated with four MHC-I alleles (*Ase-ua7*, *Ase-ua11*, and the co-occurring *Ase-ua1/10* alleles) but no MHC-II alleles. Alleles associated with compositional differences were different to those that were negatively associated with GM alpha diversity. The biggest compositional shift was associated with *Ase-ua1/10*, whose presence/absence caused the greatest number of differentially abundant ASVs. This is perhaps not surprising, as individuals with both alleles would be able to recognise a larger number of antigens, thus providing a broader immune response compared to a single allele. However, the presence/absence of *Ase-ua1/10* only explained 0.5–1.4% (depending on the metric of beta diversity) of the variance in GM composition suggesting that, separately, each allele has a relatively small impact on the GM overall.

Several ASVs were not assigned beyond the level of bacterial family and many bacterial taxa have not been fully characterised, making it difficult to draw conclusions about the functional significance of compositional changes in the GM for the host. However, there were several potentially interesting, shared candidate taxa that were differentially abundant between individuals with/without the *Ase-ua1/10* and *Ase-ua11* alleles. For example, individuals with these alleles had a greater abundance of ASVs from the order *Lactobacillales*, a lactic acid-producing bacterial taxon, generally thought to be a beneficial member of the GM. Indeed, members of this order are used as probiotics in poultry farms to boost the immune response of chickens [111]. In contrast, individuals with the *Ase-ua1/10* and *Ase-ua11* alleles had a reduced abundance of ASVs from *Bacteroidales*, an order commonly associated with chronic intestinal inflammation [112]. Two of these ASVs were from the genus *Bacteroides*; while species from this genus can be mutualistic, opportunistic pathogenic infections can occur in humans and other animals [113]. Likewise, a third ASV from the genus *Alistipes*, has a pathogenic role in various human and animal diseases [114]. The patterns of change associated with *Ase-ua7* were different to those arising from the presence/absence of *Ase-ua1/10* and *Ase-ua11*, with fewer ASVs from the orders *Lactobacillales* and *Bacteriodales* being differentially abundant. Instead, several ASVs in the order *Clostridiales* were significantly more abundant when the *Ase-ua7* allele was present (and less abundant when the *Ase-ua11* or *Ase-ua1/10* alleles were present), suggesting that this order could have been selectively tolerated, or that ASVs in this order proliferated when other competing taxa were removed.

Cumulatively, the variance in composition explained by overall MHC allele presence or absence was low (6.3% or 9.0% for unweighted or weighted UniFrac, respectively). However, this is not unusual when investigating the factors that influence GM composition across individuals within a single population; for example, environmental and host factors explained between 0.4–10% variance in a population of North American red squirrels (*Tamiasciurus hudsonicus*) [115]. Even sampling period, the most significant predictor of beta diversity in our study, explained only 2% of variation in the GM. One explanation for the low level of explained variance could be the greater presence of transitionary microbiota in the avian gut

[116]. Adaptations for flight have placed constraints on avian morphology, leading to shorter intestinal lengths, and consequently, shorter food retention times [117]; this may reduce the potential for bacterial species to adapt to the avian gut and to variation in host ecology. Secondly, if many bacterial taxa carry out the same function in the host gut, there could be a high turnover of species without any consequences for the host [118]. This can give rise to high inter-individual variation in the GM and may explain why the variables analysed here (or indeed in many within-population studies of the GM) explain a low proportion of the overall variance. Indeed, functional GM diversity may be more important than species diversity for host fitness [119]. To address this, future work incorporating metagenomic analysis would allow greater resolution of bacterial species and a more accurate assessment of the functional composition of the GM [120].

Apart from the acquired immune response, the GM may also be affected by the innate immune response (underpinned by genes such as TLRs) [26]. However, we detected no effect of *TLR3* genotype (one of the few TLRs to have functional variation in this system [50]) on the GM; this is despite the fact that survival and reproductive success are significantly associated with *TLR3* variation in the Seychelles warbler [45]. This is perhaps not surprising, given *TLR3*'s role in recognising viral dsRNA [121], rather than any bacterial conserved structures.

Individual heterozygosity was positively correlated with GM bacterial richness in the Seychelles warbler, though this was not associated with differences in GM phylogenetic diversity or composition. A decrease in individual heterozygosity (or increase in homozygosity) may reflect increased inbreeding. In the Seychelles warbler, increased inbreeding is associated with poorer individual condition (via reduced telomere length [51]) and reproductive success, with maternal homozygosity negatively predicting offspring survival in years with high mortality [52, 122]. Thus, we may be detecting an indirect effect of increased inbreeding, whereby the decreased fitness or health of individuals in turn negatively impacts GM diversity [123]. However, it is also possible that this association could be driven by reduced heterozygosity of currently unknown, functional loci directly reducing GM diversity. In future studies, it could be informative to use either quantitative trait locus mapping [124], or genome-wide association studies to identify candidate genes associated with GM variation, [e.g., 19, 125]. The Seychelles warbler could be particularly useful for this, as it underwent a bottleneck in the 1960s, resulting in a 25% reduction in genome-wide variation [46], thus making it a more tractable study system in which to disentangle the associations between host genetic variation and the GM.

It is difficult to assess the impact of the identified relationships between GM variation and MHC alleles on host fitness. Typically, greater GM alpha diversity is thought to be beneficial, as it correlates with increased health and survival in humans [23]. However, other studies have shown that high alpha diversity can indicate dis-regulation and GM instability [126]. Similarly, the function of many bacteria in the GM of wild animals is unknown. Given the key role that MHC genes play in pathogen resistance, it is possible that the observed negative correlation between GM alpha diversity and presence of specific MHC alleles is beneficial to the host. For example, Seychelles warblers with a copy of the *Ase-ua4* allele had reduced GM alpha diversity, this same allele conferred a survival advantage in individuals with a copy

[53]. To fully unpick the consequences of these GM/MHC relationships in the Seychelles warbler, longitudinal data and analyses accounting for within- and between-individual differences are needed to fully test whether there are fitness differences between individuals with different MHC alleles and GM characteristics. This is no small undertaking, and will require extensive and powerful datasets, which are not yet available.

# The GM may drive the evolution of immune genes

Variation in the GM can affect traits important to the host's own fitness [127], including host immune function [128], the severity of diseases [129], and ultimately, survival [130], providing the potential for host evolutionary adaptation [131]. If immune genes can regulate fitness through modulation of the microbiome, then it is plausible that the microbiome can also influence selection on immune genes. Thus, the microbiome can act as a selective pressure, shaping host phenotypes (reviewed in [132, 133]), ultimately resulting in in host-microbiome co-evolution, adaptation, or speciation [9, 134]. Coevolution between host immune genes and their microbiota can occur via direct pathways (by MHC genes differentially recognising bacterial species), or indirect pathways (by protection from pathogens through competition by commensal bacteria [22, 135]). Pathogen-mediated selection is thought to be central in maintaining diversity at MHC genes [136–138]. If components of the GM act as a driver of MHC variation, this could explain how variation at MHC genes is maintained in the previously bottlenecked Seychelles warbler [46, 48], despite the very limited macroparasite fauna in this population [49]. We detected no effect of MHC diversity (or optimality) on the GM, and therefore no evidence of MHC heterozygote advantage in relation to the GM [139]. However, specific alleles were associated with differences in GM composition; this is consistent with either rare allele [140] or fluctuating selection [29] mechanisms, although differentiating between the two is extremely difficult [32]. Identifying the function of GM taxa that are associated with MHC alleles, whether they be pathogenic, beneficial, or commensal, could help infer the significance and direction of these associations.

# Effects of age, sex, and field period on the GM

In addition to genetic factors, several other key variables influenced the GM in our analysis. Our results indicated a relationship between GM composition and age class in the Seychelles warbler. Specifically, old fledglings had reduced GM alpha diversity and compositional differences compared to all other age group comparisons (which did not differ from one another). In the Seychelles warbler, old fledglings are newly independent and start to forage for themselves and so may be eating different – perhaps lower quality – food items compared to older birds. This may explain the greater number of differentially abundant taxa present in this age group, including a greater abundance of ASVs in the order *Planctomycetes*, which are typically transient colonisers of the gut (but see [141]). Alternatively, exposure to stress via glucocorticoids alters host GM in other species [16]. Thus, increased stress in young individuals as they encounter new situations and pathogens could contribute to differences between age groups. Indeed, mortality is greatest during the first year of life in the Seychelles warbler [52].

While sex is an important determinant of individual variation in natural populations, its importance as a driver of GM variation varies across vertebrate species [10, 14, 115, 142, 143]. Sex was only associated with a minor difference in the GM of the Seychelles warbler, with males having marginally reduced diversity, but no difference in composition compared to females. It is, perhaps, not surprising that the effect of sex on the GM was so limited, given that Seychelles warblers of both sexes have the same diet, and exhibit limited differences in morphology and behaviour. In threespine sticklebacks GM–MHC associations were sex dependent [38]; however, we found no evidence of this in the Seychelles warbler.

Within a species, seasonal changes in diet can be an important factor driving GM variation [12, 115, 144]. In the Seychelles warbler, field period explains 1.7–2.0% of the variance in GM composition. Although the temperature on Cousin Island is relatively stable, there are measurable differences between seasons and years [145], which could lead to variation in the type and abundance of insect prey species. This may explain the observed difference in GM composition, but not diversity, between field periods. For example, mean island-wide territory quality increased by 80% in the major 2017 field period and 75% in the minor 2018 field period, compared to the later major 2018 field period. Alternatively, increases in food availability between seasons could also act indirectly on the GM by buffering individuals against stress or susceptibility to pathogens.

## Conclusions

Our results show that variation has been maintained at MHC-I and MHC-II genes in the Seychelles warbler, and that the presence of specific alleles, but not MHC diversity, was associated with differences in GM diversity and composition. It is possible that such GM–MHC interactions might explain previous results in this population showing that specific MHC alleles are associated with higher survival. However, further longitudinal data are needed to establish whether these associations equate to fitness differences between individuals and to better understand host immunogenetic–GM coevolution.

## Declarations

### *Ethics approval and consent to participate*

Fieldwork was carried out in accordance with local ethical regulations and agreements. The Seychelles Department of Environment and the Seychelles Bureau of Standards approved the fieldwork.

### *Consent for publication*

Not applicable.

### *Availability of data and material*

All 16S rRNA gene amplicon sequences have been submitted to the European Nucleotide Archive (ENA) database under the study accession number PRJEB45408.

The 44 MHC alleles have been deposited at GenBank, accession numbers for MHC

class I alleles are MZ509455-74, and for MHC class II alleles are MZ509475-98.

All metadata, along with R scripts used to run analyses, are available in the Github Repository, https://github.com/Seychelles-Warbler-Project/Davies_2021_Microbiome.

# *Competing interests*

Not applicable.

# *Funding*

CSD was funded by the Natural Environment Research Council and EnvEast DTP (NE/L002582/1). The MHC laboratory work was supported by the UK Natural Environment Research Council (NERC) Biomolecular Analysis Facility at the University of Sheffield (NBAF1150). The microbiome sequencing was funded by a NERC NBAF Pilot Scheme Grant (NBAF1092) and by a NERC grant (NE/S010939/1) to DSR.

# *Authors' contributions*

The study was conceived by CSD and DSR. CSD, DSR, and TB performed fieldwork. CSD conducted the MHC lab work while affiliated with NBAF Sheffield under the guidance of TB. CSD conducted the microbiome bioinformatics with input from SFW. CSD conducted the MHC genotyping with input from KHM. CSD performed the microbiome extractions, analyses and drafted the manuscript with supervision from DSR. DSR, HLD, JK and TB managed the Seychelles warbler project. All authors contributed critically to the work and approved the final manuscript for publication.

# *Acknowledgements*

We thank the Seychelles Bureau of Standards and the Department of Environment for permission for fieldwork overall, and Nature Seychelles for facilitating fieldwork on Cousin Island. This study would not have been possible without the contribution of exceptional fieldworkers and technicians. We particularly thank Gavin Horsburgh for assistance in MHC sequencing, Maria-Elena Mannarelli for assistance in extracting the faecal samples, and Marco van der Velde for microsatellite genotyping. Bacterial data generation and analysis were carried out by the Centre for Genomic Research, University of Liverpool. The research presented in this paper was carried out on the High Performance Computing Clusters supported

respectively by the Research and Specialist Computing Support service at the University of East Anglia and IT Services at the University of Sheffield.

# References


1. Hird SM. Evolutionary Biology Needs Wild Microbiomes. Front Microbiol. 2017; 8.
2. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al: Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci USA. 2013; 110.
3. Hooper LV, Midtvedt T, Gordon JI. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu Rev Nutr. 2002; 22.
4. Pickard JM, Zeng MY, Caruso R, Núñez G. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. Immunol Rev. 2017; 279.
5. Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol. 2009; 9.
6. Amato KR. Co-evolution in context: The importance of studying gut microbiomes in wild animals. Microbiome Sci Med. 2013; 1.
7. Maurice CF, Knowles S, Ladau J, Pollard KS, Fenton A, Pedersen AB, et al: Marked seasonal variation in the wild mouse gut microbiota. ISME J. 2015; 9.
8. Hird SM, Sánchez C, Carstens BC, Brumfield RT. Comparative gut microbiota of 59 neotropical bird species. Front Microbiol. 2015; 6.
9. Greene LK, Williams CV, Junge RE, Mahefarisoa KL, Rajaonarivelo T, Rakotondrainibe H, et al: A role for gut microbiota in host niche differentiation. ISME J. 2020; 14.
10. Stoffel MA, Acevedo-Whitehouse K, Morales-Durán N, Grosser S, Chakarov N, Krüger O, et al: Early sexual dimorphism in the developing gut microbiome of northern elephant seals. Mol Ecol. 2020; 29.
11. Videvall E, Song SJ, Bensch HM, Strandh M, Engelbrecht A, Serfontein N, et al: The development of gut microbiota in ostriches and its association with juvenile growth. bioRxiv. 2018.
12. Bolnick DI, Snowberg LK, Hirsch PE, Lauber CL, Knight R, Caporaso JG, et al: Individuals' diet diversity influences gut microbial diversity in two freshwater fish (threespine stickleback and Eurasian perch). Ecol Lett. 2014; 17.
13. Davidson GL, Cooke AC, Johnson CN, Quinn JL. The gut microbiome as a driver of individual variation in cognition and functional behaviour. Philos Trans R Soc Lon B Biol Sci. 2018; 373.
14. Tung J, Barreiro LB, Burns MB, Grenier J-C, Lynch J, Grieneisen LE, et al: Social networks predict gut microbiome composition in wild baboons. eLife. 2015; 4.
15. Mallott EK, Borries C, Koenig A, Amato KR, Lu A. Reproductive hormones mediate changes in the gut microbiome during pregnancy and lactation in Phayre’s leaf monkeys. Sci Rep. 2020; 10.
16. Noguera JC, Aira M, Pérez-Losada M, Domínguez J, Velando A. Glucocorticoids modulate gastrointestinal microbiome in a wild bird. Royal Soc Open Sci. 2018; 5.

17. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol. 2011; 9.
18. Perofsky AC, Lewis RJ, Abondano LA, Fiore AD, Meyers LA. Hierarchical social networks shape gut microbial composition in wild Verreaux's sifaka. Proc Royal Soc B. 2017; 284.
19. Suzuki TA, Phifer-Rixey M, Mack KL, Sheehan MJ, Lin D, Bi K, et al: Host genetic determinants of the gut microbiota of wild mice. Mol Ecol. 2019; 28.
20. Couch CE, Arnold HK, Crowhurst RS, Jolles AE, Sharpton TJ, Witczak MF, et al: Bighorn sheep gut microbiomes associate with genetic and spatial structure across a metapopulation. Sci Rep. 2020; 10.
21. Marietta E, Rishi A, Taneja V. Immunogenetic control of the intestinal microbiota. Immunology. 2015; 145.
22. Hooper LV, Littman DR, Macpherson AJ. Interactions Between the Microbiota and the Immune System. Science. 2012; 336.
23. Claesson MJ, Jeffery IB, Conde S, Power SE, O/'Connor EM, Cusack S, et al: Gut microbiota composition correlates with diet and health in the elderly. Nature. 2012; 488.
24. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012; 13.
25. Kurilshikov A, Wijmenga C, Fu J, Zhernakova A. Host genetics and gut microbiome: Challenges and perspectives. Trends Immunol. 2017; 38.
26. Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. Nature. 2016; 535.
27. Piertney SB, Oliver MK. The evolutionary ecology of the major histocompatibility complex. Heredity. 2005; 96.
28. Hughes AL, Yeager M. Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet. 1998; 32.
29. Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, et al: Common West African HLA antigens are associated with protection from severe malaria. Nature. 1991; 352.
30. Loiseau C, Zoorob R, Garnier S, Birard J, Federici P, Julliard R, et al: Antagonistic effects of a Mhc class I allele on malaria-infected house sparrows. Ecol Lett. 2008; 11.
31. Hedrick PW. Evolutionary genetics of the major histocompatibility complex. Am Nat. 1994; 143.
32. Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc B. 2010; 277.
33. Toivanen P, Vaahtovuo J, Eerola E. Influence of major histocompatibility complex on bacterial composition of fecal flora. Infect Immun. 2001; 69.
34. Khan MAW, Stephens WZ, Mohammed AD, Round JL, Kubinak JL. Does MHC heterozygosity influence microbiota form and function. PLoS One. 2019; 14.
35. Lin P, Bach M, Asquith M, Lee AY, Akileswaran L, Stauffer P, et al: HLA-B27 and human β2-Microglobulin affect the gut microbiota of transgenic rats. PLoS One. 2014; 9.

36. Clayton JB, Vangay P, Huang H, Ward T, Hillmann BM, Al-Ghalith GA, et al: Captivity humanizes the primate microbiome. Proc Natl Acad Sci USA. 2016; 113.
37. Jensen PE. Recent advances in antigen processing and presentation. Nat Immunol. 2007; 8.
38. 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.
39. 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.
40. Leclaire S, Strandh M, Dell'Ariccia G, Gabirot M, Westerdahl H, Bonadonna F. Plumage microbiota covaries with the major histocompatibility complex in blue petrels. Mol Ecol. 2019; 28.
41. Pearce DS, Hoover BA, Jennings S, Nevitt GA, Docherty KM. Morphological and genetic factors shape the microbiome of a seabird species (Oceanodroma leucorhoa) more than environmental and social factors. Microbiome. 2017; 5.
42. Li N, Ma W-T, Pang M, Fan Q-L, Hua J-L. The Commensal microbiota and viral infection: A comprehensive review. Front Immunol. 2019; 10.
43. Komdeur J. Importance of habitat saturation and territory quality for evolution of cooperative breeding in the Seychelles warbler. Nature. 1992; 358.
44. 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. Nature Com. 2019; 10.
45. Davies CS, Taylor MI, Hammers M, Burke T, Komdeur J, Dugdale HL, et al: Contemporary evolution of the innate immune receptor gene TLR3 in an isolated vertebrate population. Mol Ecol. 2021; 30.
46. Spurgin LG, Wright DJ, van der Velde M, Collar NJ, Komdeur J, Burke T, et al: Museum DNA reveals the demographic history of the endangered Seychelles warbler. Evol Appl. 2014; 7.
47. Richardson D,S, Jury F, 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.
48. Richardson DS, Westerdahl H. MHC diversity in two Acrocephalus species: the outbred Great reed warbler and the inbred Seychelles warbler. Mol Ecol. 2003; 12.
49. Hutchings K. Parasite-mediated selection in an island endemic, the Seychelles warbler (Acrocephalus sechellensis). University of East Anglia; 2009.
50. Gilroy D, van Oosterhout C, Komdeur J, Richardson DS. Toll-like receptor variation in the bottlenecked population of the endangered Seychelles warbler. Anim Conserv. 2017; 20.
51. Bebbington K, Spurgin LG, Fairfield EA, Dugdale HL, Komdeur J, Burke T, et al: Telomere length reveals cumulative individual and transgenerational inbreeding effects in a passerine bird. Mol Ecol. 2016; 25.
52. Brouwer L, Komdeur J, Richardson DS. Heterozygosity–fitness correlations in a bottlenecked island species: a case study on the Seychelles warbler. Mol Ecol. 2007; 16.

53. 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.
54. Richardson DS, Komdeur J, Burke T, von Schantz T. MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. Proc R Soc B. 2005; 272.
55. Brouwer L, Tinbergen JM, Both C, Bristol R, Richardson DS, Komdeur J. Experimental evidence for density-dependent reproduction in a cooperatively breeding passerine. Ecology. 2009; 90.
56. 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. 2021.
57. Komdeur J, Piersma T, Kraaijeveld K, Kraaijeveld-Smit F, Richardson DS: Why Seychelles Warblers fail to recolonize nearby islands: unwilling or unable to fly there? Ibis. 2004; 146.
58. Hammers M, Brouwer L. Rescue behaviour in a social bird: removal of sticky 'bird-catcher tree'seeds by group members. Behaviour. 2017; 154.
59. Komdeur J, Daan S. Breeding in the monsoon: semi-annual reproduction in the Seychelles warbler (Acrocephalus sechellensis). J Ornithol. 2005; 146.
60. Komdeur J, Burke T, Dugdale HL, Richardson DS: Seychelles warblers: Complexities of the helping paradox. Cooperative breeding in vertebrates: Studies of ecology, evolution, and behavior. 2016.
61. Wright DJ. Evolutionary and conservation genetics of the Seychelles warbler (Acrocephalus sechellensis). University of East Anglia; 2014.
62. Knutie SA, Gotanda KM. A non-invasive method to collect fecal samples from wild birds for microbiome studies. Microb. 2018; 76.
63. Coulon A. genhet: an easy-to-use R function to estimate individual heterozygosity. Mol Ecol Resour. 2010; 10.
64. Griffiths R, Double MC, Orr K, Dawson RJG. A DNA test to sex most birds. Mol Ecol. 1998; 7.
65. Wright DJ, Spurgin LG, Collar NJ, Komdeur J, Burke T, Richardson DS. The impact of translocations on neutral and functional genetic diversity within and among populations of the Seychelles warbler. Mol Ecol. 2014; 23.
66. Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, et al: MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. R Soc Open Sci. 2015; 2.
67. Westerdahl H, Wittzell H, Schantz Tv, Bensch S. MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. Heredity. 2004; 92.
68. Sebastian A, Herdegen M, Migalska M, Radwan J. Amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. Mol Ecol Resour. 2016; 16.
69. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011; 17.

70. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al: Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA. 2011; 108.
71. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al: QIIME allows analysis of high-throughput community sequencing data. Nat methods. 2010; 7.
72. 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.
73. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017; 11.
74. Katoh K, Misawa K, Kuma Ki, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30.
75. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mole Biol Evol. 2009; 26.
76. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al: The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2012; 41.
77. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al: The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Res. 2013; 42.
78. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS One. 2013; 8.
79. Hsieh T, Ma K, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (H ill numbers). Methods Ecol Evol. 2016; 7.
80. Chao A. Nonparametric estimation of the number of classes in a population. Scand Stat Theory Appl. 1984.
81. Shannon CE. A mathematical theory of communication. Bell Syst Techn J. 1948; 27.
82. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992; 61.
83. Battaglia T: btools: A Suite of R Function for All Types of Microbial Diversity Analyses. 0.0.1 edition; 2018.
84. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al: glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R journal. 2017; 9.
85. Nowak MA, Tarczy-Hornoch K, Austyn JM. The optimal number of major histocompatibility complex molecules in an individual. Proc Natl Acad Sci USA. 1992; 89.
86. Gelman A, Su Y, Masanao Y, Zheng T, Dorie V: arm: Data Analysis Using Regression and Multilevel/Hierarchical Models, version 1.10-1. 2018.
87. Schloerke B, Crowley J, Cook D, Hofmann H, Wickham H, Briatte F, et al: Ggally: Extension to ggplot2. 2011.

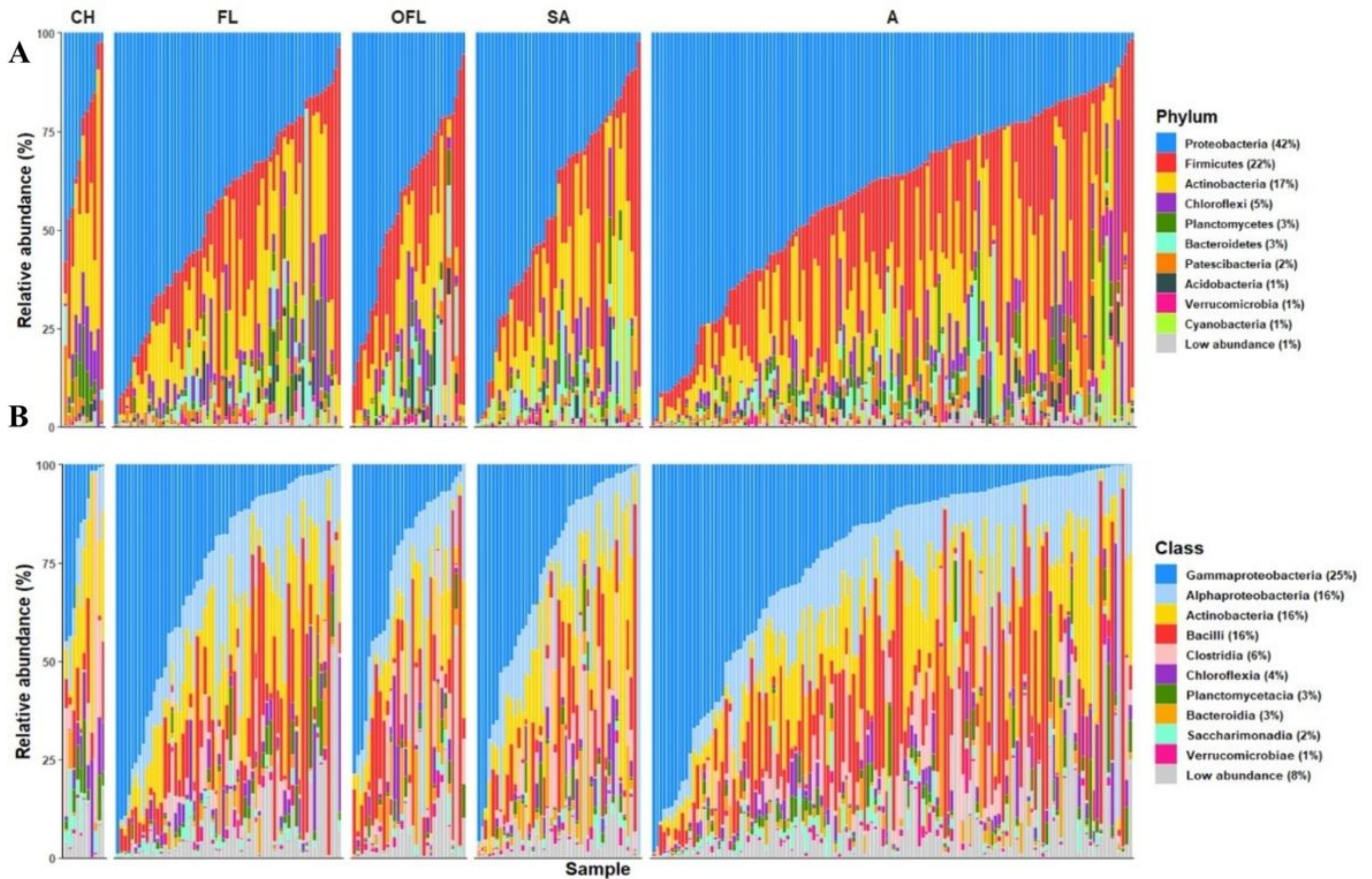
88. Hartig F: DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 024. 2017; 5.
89. Barton K, Barton MK: Package 'MuMIn'. R package version 1.43.6 edition; 2019.
90. Burnham KP, Anderson DR, Huyvaert KP. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav Ecol Sociobiol. 2011; 65.
91. Paulson JN, Stine O, Bravo H, Pop M. Robust methods for differential abundance analysis in marker gene surveys. Nat Methods. 2013; 10.
92. Paulson JN, Pop M, Bravo HC. metagenomeSeq: Statistical analysis for sparse high-throughput sequencing. Bioconductor package. 2013; 1.
93. Lozupone C, Hamady M, Knight R. UniFrac–an online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinform. 2006; 7.
94. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial Ccommunities. Appl Environ. 2007; 73.
95. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, et al: The vegan package. Community ecology package. 2007; 10.
96. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001; 26.
97. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15.
98. Kropáčková L, Těšický M, Albrecht T, Kubovčiak J, Čížková D, Tomášek O, et al: Codiversification of gastrointestinal microbiota and phylogeny in passerines is not explained by ecological divergence. Mol Ecol. 2017.
99. Minias P, Pikus E, Whittingham LA, Dunn PO: A global analysis of selection at the avian MHC. Evolution. 2018; 72.
100. Niess JH, Brand S, Gu X, Landsman L, Jung S, McCormick BA, et al: CX3CR1-Mediated dendritic cell access to the intestinal lumen and bacterial clearance. Science. 2005; 307.
101. Hooper LV, Macpherson AJ. Immune adaptations that maintain homeostasis with the intestinal microbiota. Nat Rev Immunol. 2010; 10.
102. Ackerman AL, Cresswell P. Cellular mechanisms governing cross-presentation of exogenous antigens. Nat Immunol. 2004; 5.
103. Biedrzycka A, Bielański W, Ćmiel A, Solarz W, Zając T, Migalska M, et al: Blood parasites shape extreme major histocompatibility complex diversity in a migratory passerine. Mol Ecol. 2018; 27.
104. Westerdahl H, Waldenström J, Hansson B, Hasselquist D, von Schantz T, Bensch S. Associations between malaria and MHC genes in a migratory songbird. Proc R Soc B. 2005; 272.
105. Mukherjee D, Chora ÂF, Mota MM. Microbiota, a third player in the host plasmodium affair. Trends Parasitol. 2020; 36.

106. Fairfield EA, Hutchings K, Gilroy DL, Kingma SA, Burke T, Komdeur J, et al: The impact of conservation-driven translocations on blood parasite prevalence in the Seychelles warbler. Sci Reports. 2016; 6.
107. Rosshart SP, Vassallo BG, Angeletti D, Hutchinson DS, Morgan AP, Takeda K, et al: Wild mouse gut microbiota promotes host fitness and improves disease resistance. Cell. 2017; 171.
108. Gaigher A, Burri R, San-Jose LM, Roulin A, Fumagalli L. Lack of statistical power as a major limitation in understanding MHC-mediated immunocompetence in wild vertebrate populations. Mol Ecol. 2019; 28.
109. Wakeland EK, Boehme S, She JX, Lu C-C, McIndoe RA, Cheng I, et al: Ancestral polymorphisms of MHC class II genes: Divergent allele advantage. J Immunol Res. 1990; 9.
110. Pineaux M, Merkling T, Danchin E, Hatch S, Duneau D, Blanchard P, et al: Sex and hatching order modulate the association between MHC-II diversity and fitness in early-life stages of a wild seabird. Mol Ecol. 2020; 29.
111. Brisbin JT, Gong J, Orouji S, Esufali J, Mallick AI, Parvizi P, et al: Oral treatment of chickens with lactobacilli influences elicitation of immune responses. Clin Vaccine Immunol. 2011; 18.
112. Zitomersky NL, Atkinson BJ, Franklin SW, Mitchell PD, Snapper SB, Comstock LE, et al: Characterization of adherent bacteroidales from intestinal biopsies of children and young adults with inflammatory bowel disease. PLoS ONE. 2013; 8.
113. Wexler HM. Bacteroides: the good, the bad, and the nitty-gritty. Clin. 2007; 20.
114. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. The genus Alistipes: Gut bacteria with emerging implications to inflammation, cancer, and mental health. Front Immunol. 2020; 11.
115. Ren T, Boutin S, Humphries MM, Dantzer B, Gorrell JC, Coltman DW, et al: Seasonal, spatial, and maternal effects on gut microbiome in wild red squirrels. Microbiome. 2017; 5.
116. Song SJ, Sanders JG, Delsuc F, Metcalf J, Amato K, Taylor MW, et al: Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. mBio. 2020; 11.
117. Caviedes-Vidal E, McWhorter TJ, Lavin SR, Chediack JG, Tracy CR, Karasov WH. The digestive adaptation of flying vertebrates: High intestinal paracellular absorption compensates for smaller guts. Proc Natl Acad Sci USA. 2007; 104.
118. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al: Structure, function and diversity of the healthy human microbiome. Nature. 2012; 486.
119. 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.
120. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun. 2016; 469.
121. Barton GM. Viral recognition by Toll-like receptors. Semin Immunol. 2007; 19.

122. Richardson DS, Komdeur J, Burke T. Inbreeding in the Seychelles warbler: environment-dependent maternal effects. Evolution. 2004; 58.
123. Grosser S, Sauer J, Paijmans AJ, Caspers BA, Forcada J, Wolf JBW, et al: Fur seal microbiota are shaped by the social and physical environment, show mother–offspring similarities and are associated with host genetic quality. Mol Ecol. 2019; 28.
124. Snijders AM, Langley SA, Kim Y-M, Brislawn CJ, Noecker C, Zink EM, et al: Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. Nat Microbiol. 2016; 2.
125. 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.
126. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: Networks, competition, and stability. Science. 2015; 350.
127. Suzuki TA. Links between natural variation in the microbiome and host fitness in wild mammals. Integr. 2017; 57.
128. van Veelen HPJ, Falcão Salles J, Matson KD, van der Velde M, Tieleman BI. Microbial environment shapes immune function and cloacal microbiota dynamics in zebra finches Taeniopygia guttata. Animal Microbiome. 2020; 2.
129. Villarino NF, LeCleir GR, Denny JE, Dearth SP, Harding CL, Sloan SS, et al: Composition of the gut microbiota modulates the severity of malaria. Proc Natl Acad Sci USA. 2016; 113.
130. Benskin CMH, Rhodes G, Pickup RW, Mainwaring MC, Wilson K, Hartley IR. Life history correlates of fecal bacterial species richness in a wild population of the blue tit Cyanistes caeruleus. Ecol Evol. 2015; 5.
131. Kolodny O, Callahan BJ, Douglas AE. The role of the microbiome in host evolution. Philos Trans R Soc Lond B Bio Sci. 2020; 375.
132. Koskella B, Bergelson J. The study of host-microbiome (co)evolution across levels of selection. Philos Trans R Soc Lond B Bio Sci. 2020; 375.
133. Rowe M, Veerus L, Trosvik P, Buckling A, Pizzari T. The reproductive microbiome: An emerging driver of sexual selection, sexual conflict, mating systems, and reproductive isolation. Trends Ecol Evol. 2020; 35.
134. Moeller AH, Sanders JG. Roles of the gut microbiota in the adaptive evolution of mammalian species. Philos Trans R Soc Lond B Bio Sci. 2020; 375.
135. 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 Comms. 2015; 6.
136. Eizaguirre C, Lenz TL, Kalbe M, Milinski M. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nat Comms. 2012; 3.
137. Hedrick PW. Pathogen resistance and genetic variation at MHC loci. Evolution. 2002; 56.

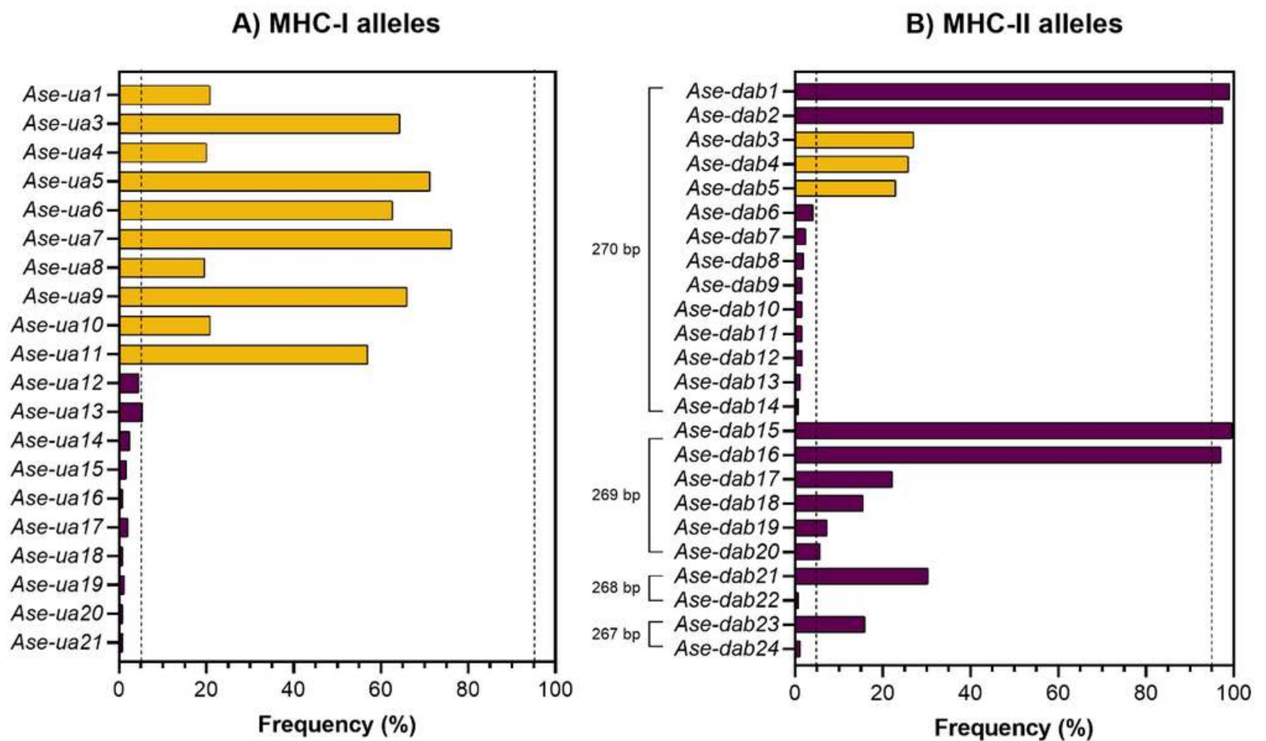
138. Bernatchez L, Landry C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol. 2003; 16.
139. Doherty PC, Zinkernagel RM. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature. 1975; 256.
140. Slade R, McCallum H. Overdominant vs. frequency-dependent selection at MHC loci. Genetics. 1992; 132.
141. Cayrou C, Sambe B, Armougom F, Raoult D, Drancourt M. Molecular diversity of the Planctomycetes in the human gut microbiota in France and Senegal. APMIS. 2013; 121.
142. Teyssier A, Rouffaer LO, Saleh Hudin N, Strubbe D, Matthysen E, Lens L, et al: Inside the guts of the city: Urban-induced alterations of the gut microbiota in a wild passerine. Sci Total Environ. 2018; 612.
143. Nelson TM, Rogers TL, Carlini AR, Brown MV. Diet and phylogeny shape the gut microbiota of Antarctic seals: a comparison of wild and captive animals. Environ. 2013; 15.
144. Michel AJ, Ward LM, Goffredi SK, Dawson KS, Baldassarre DT, Brenner A, et al: The gut of the finch: uniqueness of the gut microbiome of the Galápagos vampire finch. Microbiome. 2018; 6.
145. van de Crommenacker J, Komdeur J, Burke T, Richardson DS. Spatio-temporal variation in territory quality and oxidative status: a natural experiment in the Seychelles warbler (Acrocephalus sechellensis). J Anim Ecol. 2011; 80.

# Figures



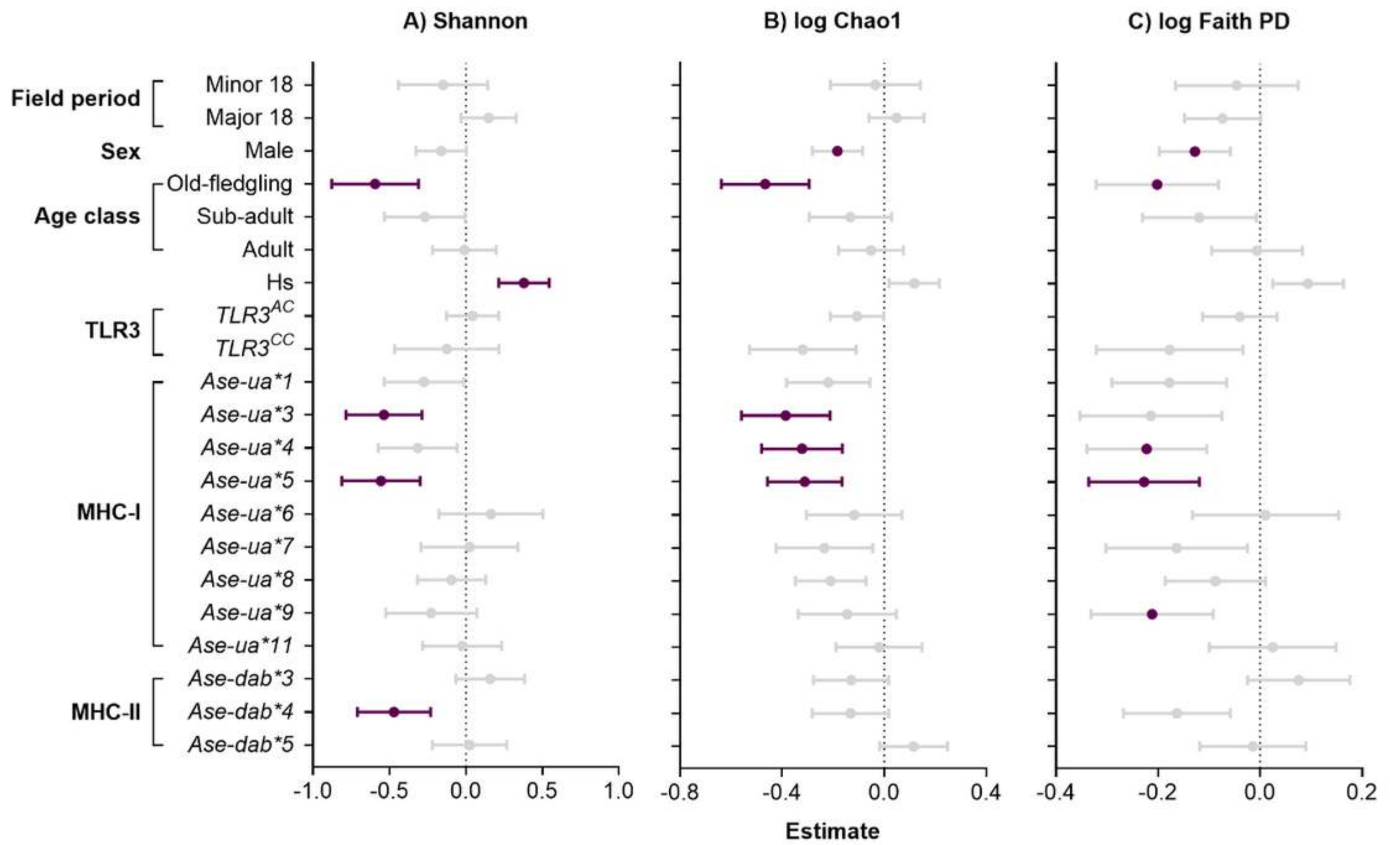

## Figure 1

The relative abundance of bacterial (A) phyla and (B) classes in Seychelles warbler faecal samples. Each column represents a single faecal sample (281 faecal samples, collected from 224 Seychelles warblers). Samples are separated by age class: CH = chick, FL = fledgling, OFL = old fledgling, SA = sub-adult, A = adult. Y-axis shows the relative abundance (%) of the 10 most abundant bacterial taxa. All other taxa are collapsed into the low abundance category.



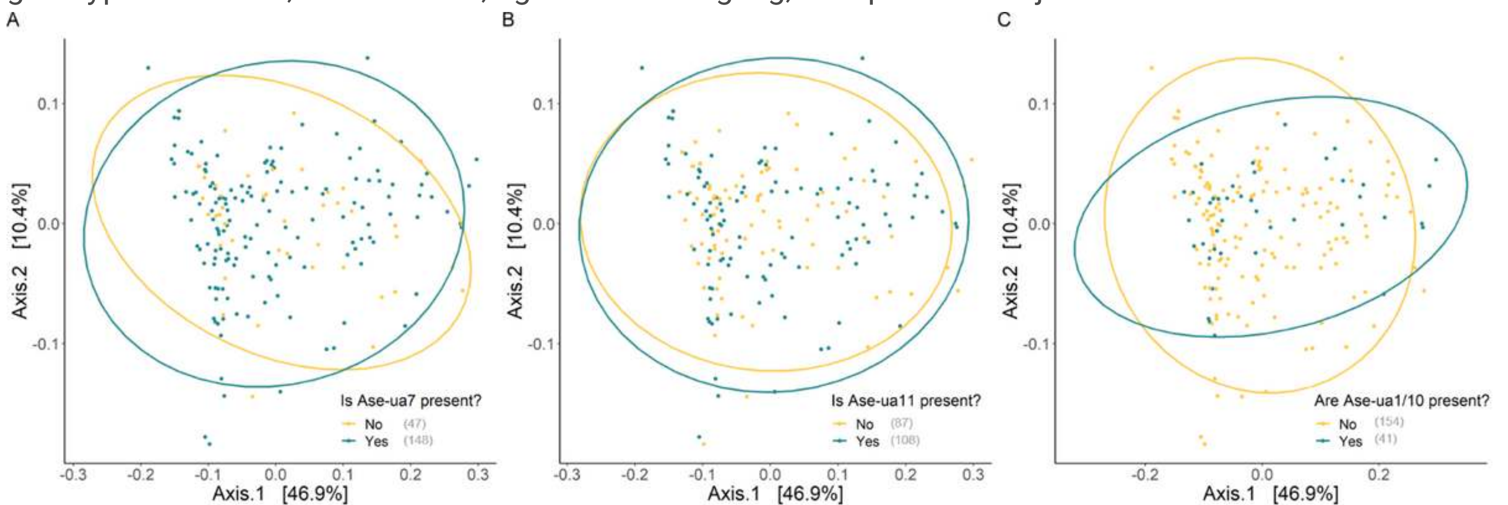

## Figure 2

Variation in MHC-I exon 3 and MHC-II exon 2 in 244 Seychelles warblers. Each bar represents the frequency (%) of each (A) MHC-I allele and (B) MHC-II allele. Bars represent MHC alleles included (yellow) or not included (purple) in presence/absence analysis. Dashed lines indicate 5% or 95% frequency cut offs.



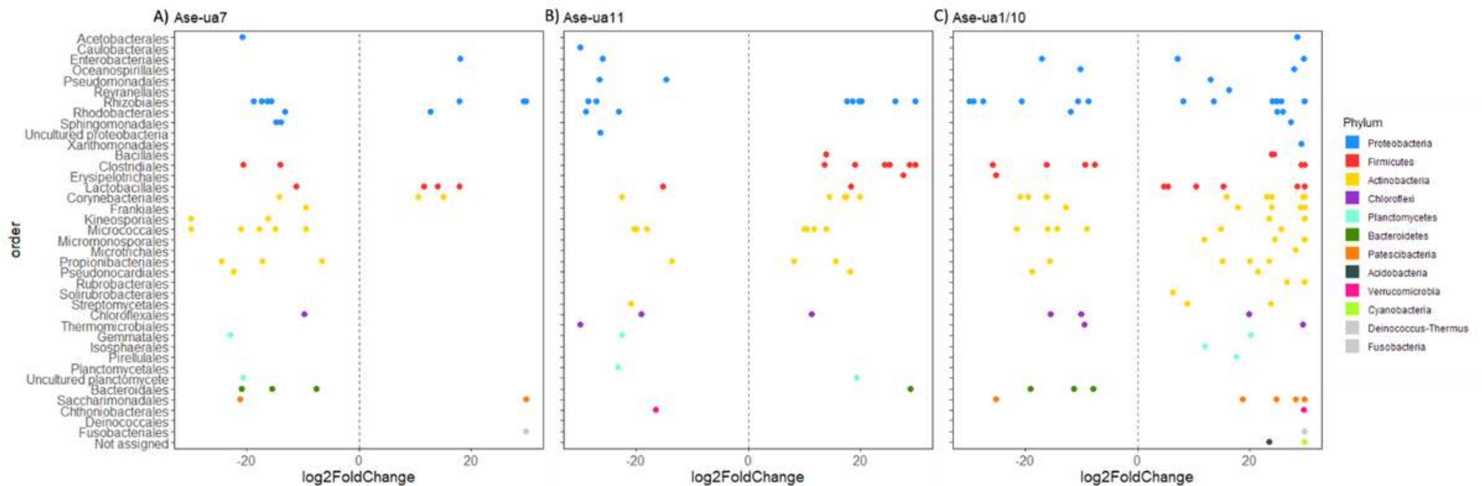

## Figure 3

Effects MHC alleles, TLR3 genotype and genome-wide heterozygosity and host variables on alpha diversity in 195 Seychelles warblers. Alpha diversity metrics are (A) Shannon diversity, (B) Chao1, and (C) Faith's phylogenetic diversity (PD). Estimates and standard errors are based on linear conditional model-averaged estimates. An estimate >0 indicates increased alpha diversity, while <0 indicates decreased alpha diversity. Significant terms (P < 0.05) are highlighted in purple, and terms approaching significance (P < 0.1) are indicated with a purple point. Estimates are in reference to MHC allele = absent, TLR3 genotype = TLR3AA, sex = female, age class = fledgling, field period = major 2017.



### Figure 4

Beta diversity of Seychelles warbler gut microbiome composition depending on presence of three MHC-I alleles. The principal coordinate plots are based on weighted UniFrac distances according to the presence or absence of the MHC-I (A) Ase-ua7 allele, (B) Ase-ua11 allele, or (C) Ase-ua1/10 allele. Points represent a single faecal sample from a single individual (n = 195). Sample sizes are specified in brackets in the legend, and colours indicate the presence (blue) or absence (yellow) of the MHC-I allele. Ellipses represent a 95% confidence interval around the cluster centroids.




### Figure 5

Differentially abundant ASVs (Padj < 0.01) in the gut microbiomes of Seychelles warblers, according to the presence/absence of the MHC-I alleles A) Ase-ua7 B) Ase-ua11 or C) Ase-ua1/10. ASVs are grouped at the level of bacterial order and coloured according to bacterial phylum. ASVs shown with a log2-fold change greater than zero are significantly more abundant in individuals with this allele and ASVs with a log2 fold change smaller than zero are significantly more abundant in individuals without a copy of this allele.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1Microbiome09062021.docx