Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity

DANIEL I. BOLNICK,* LISA K. SNOWBERG,† J. GREGORY CAPORASO,‡ CHRIS LAUBER,§ ROB KNIGHT¶ and WILLIAM E. STUTZ†**

*Howard Hughes Medical Institute and Section of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA, †Section of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA, ‡Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA, §Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309-0216, USA, ¶Howard Hughes Medical Institute and Department of Chemistry and Biochemistry and BioFrontiers Institute, University of Colorado, Boulder, CO 80309-0215, USA, **Department of Ecology and Evolution, University of Colorado, Boulder, CO 80309-0215, USA

Abstract

Animals harbour diverse communities of symbiotic bacteria, which differ dramatically among host individuals. This heterogeneity poses an immunological challenge: distinguishing between mutualistic and pathogenic members of diverse and host-specific microbial communities. We propose that Major Histocompatibility class II (MHC) genotypes contribute to recognition and regulation of gut microbes, and thus, MHC polymorphism contributes to microbial variation among hosts. Here, we show that MHC IIb polymorphism is associated with among-individual variation in gut microbiota within a single wild vertebrate population of a small fish, the threespine stickleback. We sampled stickleback from Cedar Lake, on Vancouver Island, and used next-generation sequencing to genotype the sticklebacks’ gut microbiota (16S sequencing) and their MHC class IIb exon 2 sequences. The presence of certain MHC motifs was associated with altered relative abundance (increase or decrease) of some microbial Families. The effect sizes are modest and entail a minority of microbial taxa, but these results represent the first indication that MHC genotype may affect gut microbiota composition in natural populations (MHC-microbe associations have also been found in a few studies of lab mice). Surprisingly, these MHC effects were frequently sex-dependent. Finally, hosts with more diverse MHC motifs had less diverse gut microbiota. One implication is that MHC might influence the efficacy of therapeutic strategies to treat dysbiosis-associated disease, including the outcome of microbial transplants between healthy and diseased patients. We also speculate that macroparasite-driven selection on MHC has the potential to indirectly alter the host gut microbiota, and vice versa.

Keywords: Gasterosteus aculeatus, genetic diversity, Major Histocompatibility Complex, microbiota, mucosal immunity, threespine stickleback

Received 10 April 2014; revision received 10 June 2014; accepted 11 June 2014

Introduction

Animals harbour diverse communities of symbiotic microbes on the skin and on mucosal surfaces such as the gut lumen (Ley et al. 2008), comprising thousands of microbial species from diverse Phyla. This diversity
poses a substantial immunological challenge to the host, which must distinguish between beneficial mutualists (Balcazar et al. 2007; Cheesman et al. 2011; Lathrop et al. 2011), vs. pathogens (Cerf-Bensussan & Gaboriau-Routhiau 2010). Improper immune responses induce deleterious changes in microbial community composition (dysbiosis) resulting either in loss of mutualist benefits, or pathogen infection. Dysbiosis can lead to a variety of metabolic and inflammatory diseases (Turnbaugh & Gordon 2009; Jostins et al. 2012; Koeth et al. 2013; Markle et al. 2013; Proal et al. 2013; Yoshimoto et al. 2013).

Regulating diverse gut microbiota is made more difficult by substantial among-individual variation in microbial composition (Ley et al. 2008; Wu et al. 2011). Within a host species, individuals living in the same location can have radically different microbiota (e.g. Fig. S1, Supporting information), due to a combination of genetic, environmental and dietary differences among hosts (Muegge et al. 2011; Spor et al. 2011). Consequently, co-occurring individuals must regulate different combinations of microbial taxa. A major challenge in ecological immunology is to understand how the host genome encodes immunological tools that interact with such a diverse community that also differs among related host individuals.

The mucosal immune system relies on both innate and adaptive immune pathways that use diverse protein receptors to detect pathogen-associated molecular patterns (Hooper et al. 2012; Robertson & Girardin 2013) and induce appropriate immunological responses leading to tolerance or attack. The vertebrate adaptive immune system is particularly promising for regulating the gut microbiota for four reasons. First, mis-regulation of adaptive immunity (e.g. deficits of Treg cells) can alter microbial composition and can lead to inflammatory diseases (Atarashi et al. 2013). Second, the adaptive immune response can be fine-tuned to each individual’s prior exposure to microbial antigens, via positive selection of T- and B-cell receptors that match previously encountered microbes. Third, adaptive immunity leads to lasting immunological memory that could help explain to the resilience of individuals’ unique microbial communities in the face of perturbations such as antibiotic treatment (Dethlefsen & Relman 2011). Finally, the adaptive immune system includes some of the most polymorphic genes in the vertebrate genome, such as germ-line variation in the Major Histocompatibility Complex (MHC) and somatic variation in T- and B-cell receptor loci. This diversity in immune receptor structure, both within- and among-individuals, might contribute to microbial diversity variation within and among hosts.

The MHC class II gene family, among the most polymorphic genes known (Eizaguirre & Lenz 2010; Spurgin & Richardson 2010), is a logical candidate for regulating diverse and host-specific microbiota. MHC class II is produced in antigen-presenting cells such as macrophages and dendritic cells (Inaba et al. 1998; Hepworth et al. 2013). These cells phagocytize extracellular particles including, but not limited to, bacteria, viruses and antigens shed from multicellular parasites including helminths. Dendritic cells are abundant in the lamina propria just behind the gut epithelium, where they extend dendrites between epithelial cells to phagocytize particles (including gut microbes) from the inner mucus layer of the gut lumen (Niess et al. 2005; Hooper et al. 2012). These particles are sheared into short peptide chains, which bind to the antigen-binding groove of the MHC II protein. This complex is then exported to the cell surface to present the antigen to T cells (Mortha et al. 2014). We focus here on MHC II because it predominantly presents phagocytized antigens (as would be the case for gut lumen microbes), unlike MHC class I that presents intracellular antigens.

Allelic variants at the MHC Ila and MHC IIb genes give rise to MHC class II proteins that have different peptide-binding region (PBR) sequences and consequently bind to different antigen peptide chains, allowing co-dominant recognition of diverse foreign antigens (Doherty & Zinkernagel 1975; McClelland et al. 2003). Only antigens whose protein sequence binds to the PBR are exported to the cell surface and presented to T cells (Hunt et al. 1992). A T cell will adhere to the dendritic cell if its T-cell receptor (TCR) binds to the MHC/antigen complex, leading to T-cell activation (Hooper et al. 2012). Depending on other signals of infection, the MHC/T-cell interaction can lead to (i) apoptosis of the T cell to minimize auto-immunity (Kappler et al. 1988), (ii) maturation of the T cell into effector T\textsubscript{H1}, T\textsubscript{H2} or T\textsubscript{H17} cells that initiate a variety of inflammatory or anti-inflammatory immune reactions against pathogens (Teh et al. 1988), or (iii) maturation into Foxp3\textsuperscript{+} T\textsubscript{reg} cells with immunosuppressive activity (Steinman et al. 2003; Sun et al. 2007). The latter point is particularly significant because altered microbiota can lead to a paucity of T\textsubscript{reg} cells, resulting in autoimmune disorders (Lee & Mazmanian 2010; Kranich et al. 2011). Finally, effector T-cell activation regulates gut microbes through a variety of mechanisms, which include activating B-cells to secrete protective antibodies directly into the gut mucosa and lumen (Fagarasan & Honjo 2003; Cebula et al. 2013; Farace et al. 2013; Mortha et al. 2014).

Due to the extreme polymorphism in MHC, we hypothesized that individuals with different MHC II genotypes will recognize different combinations of microbes, leading to among-individual differences in immunity or tolerance. Consequently, gut microbial composition should covary with MHC genotype.
Surprisingly, despite the well-known role of dendritic cells in sampling proteins in the gut and training T cells (Hooper et al. 2012), there is little evidence that natural polymorphism in MHC explains variation in gut microbiota composition. Congenic laboratory mice with different MHC genotypes have different faecal microbes (Toivanen et al. 2001), but this study has been criticized on the grounds of not properly controlling for family structure (Spor et al. 2011). Another study of 20 human newborns showed that HLA-DQ genotype (a subset of MHC II genes) affects the development of the gut microbiota, as individuals with Coeliac-associated HLA alleles developed a distinctive microbiota (De Palma et al. 2010). However, very few other studies have examined MHC-microbiota associations, and no studies in natural populations.

We examined the composition of the gut microbiota and MHC sequences from a sample of individuals from a single natural population of threespine stickleback (Gasterosteus aculeatus). This small north temperate fish species is a major model organism in evolutionary genetics and evolutionary ecology that has widely been studied to understand the role of parasite communities and nonrandom mating in sustaining MHC polymorphism (Aeschlimann et al. 2003; Wegner et al. 2003; Kurtz et al. 2004; Kalbe et al. 2009). We show that there are numerous associations between MHC genotype and the abundance of particular members of the gut microbial community, and MHC diversity influences gut microbial diversity.

**Methods**

**Sample acquisition**

In June 2008, we sampled 400 threespine stickleback from Cedar Lake, British Columbia (50°12′09″ N, 125°33′58″ W), which contains a typical single population of stickleback (Berner et al. 2010). Individuals were captured in unbaited minnow traps set along a short (~300 m) segment of shoreline on a single day. Specimens were frozen at −80 °C and later thawed and dissected to determine sex and remove the intestinal tract to extract microbial DNA from a subset of 150 individuals.

**Microbiota genotyping**

We extracted DNA from the entire intestine, using the Powersoil Kit (MO BIO Laboratories, Carlsbad CA, USA). We PCR-amplified the V4 hypervariable region of the bacterial 16S ribosomal RNA gene (sites 515–806 following E. coli positions) as described in Bolnick et al. (2014a). 16S rRNA gene was PCR-amplified using standard Earth Microbiome Project PCR conditions (Caporaso et al. 2011, 2012) with primers and individual-sample barcodes described in Bates et al. (2011) and Bergmann et al. (2011). Barcoded amplicons were pooled and sequenced on an Illumina HiSeq 2000 at the University of Colorado, as described in Caporaso et al. (2012). Sequences were demultiplexed and filtered using QIIME default parameters for sequence quality.

Operational Taxonomic Units (OTUs, the equivalent of species) were picked using a closed-reference protocol against the Greengenes database (12_10 release) at 97% similarity. Reads that did not match any reference sequence with at least 97% identity were discarded. Nearly all discarded reads were host mitochondrial 16S rRNA gene sequence, or spiked-in Phi-X. On average, we retained 56,010 sequence reads per individual (median = 35,652). OTUs were assigned taxonomic identities based on their best match in Greengenes. Using closed-reference OTU-picking discards roughly 30% of the bacterial 16S rRNA gene sequences (the exact percentage varies among individuals). Using open-reference OTU-picking instead yields higher overall microbial diversity, due to the inclusion of OTUs with uncertain taxonomic affiliation. The OTUs discarded during closed-reference OTU picking tend to be rather rare within and among hosts, and many of them cannot be clearly assigned to known microbial Families and so would not have been included in our core analyses anyway. Finally, alpha and beta diversity measures (e.g. PD, species richness, PCoA scores) are highly correlated between open- and closed-reference OTU-picking results (r > 0.96 for any of these diversity measures). Consequently, the use of closed-reference OTU-picking is unlikely to appreciably influence our results.

We calculated alpha diversity (chaol, phylogenetic diversity, and OTU richness) within each host (Lozupone & Knight 2008). To account for dependence of diversity metrics on sampling effort, all samples were rarified to 3448 sequence reads before calculating diversity measures. This value was chosen to avoid discarding samples with modest coverage, and higher rarefaction (e.g. 10,000 reads) yielded highly correlated diversity measures (r > 0.98). To quantify the major axes of phylogenetic microbiota variation among individuals, we calculated weighted and unweighted UniFrac distances among hosts using the rarified data (Lozupone et al. 2006) and performed Principal Coordinate Analysis (PCoA) analysis on the resulting distance matrices based on OTU presence/absence (unweighted PCoAs) or relative abundance (weighted PCoAs).

Two caveats are worth noting. First, we examine a snapshot of each individual’s gut microbiota at one point in time, and it is not known how stable individual sticklebacks’ gut microbiota are through time. However, studies in adult humans suggest that individuals’ gut
microbiota are very stable through time (Lozupone et al. 2012). Second, it is not presently known which microbes in the stickleback gut are pathogens, food-acquired transients, or commensal residents. Consequently, we treat all microbes equally and do not make assumptions about their ecology, although we excluded chloroplast 16S sequences (typically <10% of the microbiota) that are almost certainly from ingested algae, and eukaryotic mitochondrial 16S rRNA gene.

**MHC genotyping**

We purified genomic DNA from a fin clip from each stickleback, using Promega Wizard Genomic DNA Purification Kits. We used barcoded forward and reverse general primers to PCR amplify the highly variable exon 2 from all MHC Iib loci simultaneously (forward primer 5’-TGTCTTTAACTCCACGGAGC3’; reverse primer 5’-CTCTGACTCACCAGACTTAG3’). This exon is the most variable portion of MHC II because it encompasses the PBR that binds antigens. Our primers, similar to ones independently developed by Lenz et al. (2009), amplify a 213-bp amplicon (excluding primers) from all MHC Iib paralogues. The published genome indicates there are at least five functional MHC copies (Dijkstra et al. 2013), but estimates have ranged from 4 to 6 (Sato et al. 1998; Reusch & Langefors 2005); discrepancies may reflect copy number variation.

We used 50 μL PCRs containing 25 ng of genomic template DNA, 10 μL of 10× PCR buffer, 300 50 μM MgCl₂, 20 μM dNTPs, 20 μM of each primer (uniquely barcoded for each sample), and 1 unit Platinum Taq DNA polymerase. PCR was carried out with 120 s of 94 °C initialization followed by 25 cycles in which denaturation at 94 °C (30 s) was followed by 57 °C annealing (30 s), extension at 72 °C (60 s) and elongation at 72 °C for 240 s. PCR products were purified with Agencourt AMPure XP beads, quantified with Picogreen, and equimolar aliquots were pooled across samples to generate a multiplexed library that was sequenced on a Roche/454 FLX sequencer.

We obtained 39 692 sequences, with between 1 and 4886 reads per fish (median = 99), and retained individuals with >50 reads. With this cut-off, we have a 94.9% probability of observing all alleles in an individual with 10 unique alleles (e.g. heterozygous at each of five paralogues; most individuals will have fewer alleles). Discarding samples with insufficient sequencing depth, we retained 123 fish (55 females, 68 males) with a median of 127 sequence reads (mean of 227).

Sequencing by 454 is somewhat error-prone; as a result, when sequencing a single animal’s MHC a large number of sequence reads are likely to deviate slightly from the true allele sequences. We inferred the true sequences by Stepwise Threshold Clustering (STC; described fully in Stutz & Bolnick In press), which clusters the erroneous sequences with the true allele sequences. Briefly, STC uses a quasi-Dirichlet process to iteratively group similar reads into clusters of similar sequences for each individual. A cluster inferred to represent one true allele will contain the correct allelic sequence at relatively high frequency and erroneous sequences at lower frequency. Size and similarity criteria are applied to each cluster to determine whether each cluster represents one or more than one true allele. In the latter case, the reads are reclustered at more stringent similarity thresholds until each cluster represents a single allele. This approach combines strengths of multiple existing methods for genotyping diverse gene families with NGS data (Babik et al. 2009; Galan et al. 2010; Prosperi et al. 2010; Sommer et al. 2013) with the added advantage of not requiring sequencing of duplicate libraries of every sample. Using the STC output, we measured individual MHC diversity in two ways: the number of unique sequence motifs (richness), and the average pairwise sequence differences among motifs (‘divergence’).

**Data analysis**

We calculated the relative abundance of each microbial taxon in each stickleback (# reads observed out of total # reads per host). We then used quasibinomial General Linear Models (GLMs) to test for associations between each MHC motif and the relative abundance of each microbial taxon (focusing on Family-level taxa). We focus on pairwise associations between common microbes and common alleles (each found in at least five hosts), to avoid inflating the problem of multiple comparisons with low-power analyses. GLMs allow us to account for variable sequencing depth across hosts and corresponding uncertainty in taxon relative abundance, without having to discard data (e.g., no need to rarify to a lowest common sequence depth). Binomial GLMs explicitly account for uncertainty in the relative abundance of a given microbe taxon in a host individual. When microbe taxon abundances are overdispersed (excessive variance and skew), binomial GLMs can yield excessive false positives, whereas quasibinomial GLMs are more robust. Note that these associations represent correlation, not causation, and even causal associations may entail indirect effects of MHC on microbes (e.g. via MHC effects on host parasite load, immune activity, condition).

We used false discovery rate (FDR) analysis to acquire q-values to determine when particular MHC/Family associations are robust to multiple comparisons. We also use Fisher’s combined probability to test the
null hypothesis that an entire set of tests includes only false positives. However, both FDR and Fisher’s probability are flawed because they assume multiple tests are independent, whereas multiple motifs or microbes can show correlated effects because they are genetically similar, or (for MHC) motifs may be in linkage disequilibrium (Fig. S2, Supporting information). To adjust for potential non-independence, we examined the distribution of significant results in the matrix of P-values from GLMs comparing motifs (rows) and microbial Families (columns). If significant results are entirely due to false positives, they should be distributed randomly across this matrix. Conversely, if certain motifs have real effects on the microbiota, then the corresponding row of P-values should be enriched for significant effects. Similarly, if a certain microbial Family is particularly sensitive to MHC, its column should be enriched for significant P-values. We therefore calculated the row and column marginal frequencies of significant and non-significant results and used chi-squared tests to check for disproportionate concentration of significant effects in certain motifs, or on certain Families.

We analysed several types of relationship between genetic and microbial diversity. First, we used quasibinomial GLMs to test whether the relative abundance of each microbial Family depends on MHC diversity (richness or divergence). Conversely, do particular MHC motifs coincide with higher or lower microbial diversity (PD, richness, or chao1, tested with ANOVAs)? Finally, we tested for correlations between measures of MHC diversity and measures of microbial diversity.

**Results**

**Among-individual variation in microbiota and MHC**

Illumina sequencing of bacterial and archaeal 16S rRNA genes identified a total of 2426 microbial OTUs present with at least 10 sequence reads, 1303 of which were found in at least two fish (959 in at least three fish). Most OTUs confined to a single host were exceedingly rare [on average only 0.11% (maximum 5%) of the sequences from the single host individual where they were found]. Excluding these OTUs found in only one host (following Caporaso et al. 2011), we found an average of 115 OTUs per host (ranging from 10 to 349, median of 98, after rarefaction to 3448 reads per individual; Table S1, Supporting information). Even for microbes with high prevalence (found in all hosts), relative abundance varied dramatically among hosts (e.g. Fig. S3, Supporting information).

We identified 87 unique MHC Iib DNA sequences and 82 unique amino acid sequences (hereafter referred to as sequence motifs, the protein analogue to alleles), after STC pooled individually rare sequence errors with the most similar common sequence. We excluded five presumptive pseudogenes that had premature stop codons but were found in multiple fish. We focus on amino acid sequences (motifs), rather than DNA sequences (alleles) because the former interact directly with microbial antigens. Below, we focus on 14 common MHC motifs (found in ≥5 individuals; Fig. S4, Supporting information), but we used all motifs in calculating MHC diversity measures. Individuals carried between 1 and 9 distinct motifs (mean = 5.08, SD = 1.66; Fig S4, Supporting information), consistent with the presence of at least five MHC paralogues and similar to polymorphism seen in benthic species of stickleback from Paxton Lake on nearby Texada Island (Matthews et al. 2010).

**Associations between MHC motifs and microbial taxa**

We found a small number of significant pairwise associations between MHC motifs and microbial Families. At the α = 0.05 level, there was a negligible excess of significant results above false-positive expectations (81 of 1456; P = 0.35), but at α = 0.01 there was a significant excess of statistically significant associations (35/1456; P < 0.0001; Fig. S5, Supporting information). Selected examples of significant MHC/microbe associations are shown in detail in Fig. 1; all effects are shown in Fig. 2. Applying false discovery rate (FDR) corrections to all 1456 comparisons, seven MHC/Family associations remain significant at q < 0.05. However, FDR assumes independent statistical tests, which is clearly violated due to pairwise comparisons and ancestral structure among alleles and among microbes. Indeed, some MHC motifs exhibit a disproportionate number of significant microbial associations, rather than having significant effects randomly distributed among motifs (χ² = 63, d.f. = 13, P = 2.04×10⁻⁸). Six of the 14 common MHC motifs were unambiguously associated with changes in the relative abundance of at least some microbial Families (motifs P123, P119, P121, P88, P8 and P12; Fisher’s combined P < 0.05 and q value <0.05) and significantly affect microbial PCoA scores in MANOVAs (e.g. Figs S6 & S7, Supporting information). One striking feature of the distribution of MHC effects is that common MHC motifs typically have negative effects on the abundance of microbial Families, whereas rare MHC motifs typically have positive effects averaged across microbe Families (Fig. S8, Supporting information).

Likewise, a few microbial Families exhibit a disproportionate concentration of significant MHC/Family associations (χ² = 157, d.f. = 103, P = 0.0004). Fifteen of 104 Families exhibited overall significant responses.
across MHC motifs (Fisher’s combined probability <0.05; Fig. 2). Examples in Fig. 1 illustrate associations between MHC motifs and microbes (Fisher’s P < 0.05 & q < 0.05). Families exhibiting significant MHC effects are taxonomically clustered (Fig. 2). For instance, motif P77 affects the abundance of three Families in the Order Solirubrobacterales (Phylum Actinobacteria); Families in the order Rhodospirillales (Class Alphaproteobacteria) are associated with P121 and P12; and most Families in the Phylum Planctomycetes are associated with multiple MHC motifs. Note that the concentration of motif-associations in a few motifs and a few Families is not observed in null models where we permute individuals’ MHC genotype (Fig. S9, Supporting information).

To summarize, although we found significant effects of MHC on gut microbiota composition, this effect was generally weak, involving a minority of gut microbial Families. To put this into perspective, we used linear models to simultaneously estimate the effect size (% variance explained) of stickleback MHC genotype (six motifs with Fisher’s P < 0.05), sex, size, diet, and ecomorphology (three PC axes) on microbial PCoA axes (see Fig. S10, Supporting information for detailed methods). We found that MHC explained a cumulative 9.99% of microbial PCoA variation, compared with 1.3% for diet, 0.5% for sex, 0.4% for size, and 3.5% for ecomorphology (Fig. S10, Supporting information).

Immune function is frequently reported to be sexually dimorphic, both due to sex differences in nutrition, parasite exposure, and immunosuppressive effects of male sex hormones (Kurtz et al. 2007). When we re-ran GLMs for each motif-Family pair, including sex and...
sex motif interactions, we found two microbial Families (Enterobacteriaceae and Anaplasmataceae) that were particularly prone to sex-specific MHC allele effects (e.g. Fig. S11, Supporting information). More generally, effects of MHC motifs on microbiome composition in males were only very weakly correlated with the comparable motif-Family effects in females (Fig. S12, Supporting information).

Effects of MHC diversity

Major Histocompatibility class II sequence divergence (mean pairwise amino acid sequence differences within an individual) influenced the relative abundance of 17 microbial Families (Fig. 3; specific examples are illustrated in Fig. S13, Supporting information). Some Families were more common in individuals with high MHC divergence, others Families were correspondingly less common. Notably, many Families responded differently to MHC diversity in males vs. females: 24 microbial Families covaried with MHC divergence in at least one sex, but only one Family showed significant parallel effects of diversity in both sexes (Fig. 3).

Several MHC motifs are associated with changes in microbial diversity (measured either as PD, chao1, or richness). The presence of motifs P24, P12, and P123 coincided with significant increases, of between 23% and 33%, in microbial diversity (Fig. S14, Supporting information). In contrast, the presence of motif P121 reduced microbial diversity by 26%.

Finally, individuals with more divergent MHC motifs exhibited less diverse microbiota (significant for PD, chao1, and richness; Fig. 4A). This relationship did not hold when counting MHC motif richness without considering amino acid diversity (Fig. 4B). We infer that protein structure disparity among MHC motifs within a host reduces the diversity of gut microbial communities. This effect appears to be mediated by a high frequency of Firmicutes in low MHC diversity individuals (Fig. S15, Supporting information).

Discussion

In recent years, biologists have devoted considerable attention to understanding immune regulation of the gut microbiota (Hooper et al. 2012). Much of this research has focused on innate immune mechanisms such as Toll-like and Nod-like receptors (Round et al. 2011; Robertson & Girardin 2013), which recognize conserved motifs on microbial cell surfaces and then initiate immunological responses. However, given the limited polymorphism in these genes, it is not clear how effective these pattern recognition proteins are at making fine distinctions among diverse gut microbes, nor how these relatively conserved protein families could work to interact with complex microbiota that are also unique to each individual host. Adaptive cellular immunity, on the other hand, has the capacity to mount flexible responses to host-specific microbial communities, resulting in inflammatory immune attacks or tolerance via anti-inflammatory pathways that may be essential in regulating the gut microbiota (Lee & Mazmanian 2010; Hand et al. 2012; Belkaid et al. 2013; Cebula et al. 2013; Mortha et al. 2014).

Given the role of dendritic cells in sampling gut microbes and using MHC II to present antigens to activate T cells (Steinman et al. 2003; Niess et al. 2005; Sun et al. 2007; Farache et al. 2013; Mortha et al. 2014), we hypothesized that MHC class II genotype may influence which microbes a host can recognize. Consequently, polymorphism in MHC should generate among-host differences in gut microbiome composition. Our correlative data from a wild population of stickleback supports this hypothesis. The presence of some MHC motifs coincides with modest but significant changes in gut microbiota. To our knowledge, this represents the first evidence that MHC affects the composition of the gut microbiota in natural populations, consistent with previous results from inbred laboratory mice (Toivanen et al. 2001). The effect of MHC on the gut microbiota of the wild-caught fish studied here, while modest, exceeds the effects of sex, size, diet and ecomorphology (Fig. S10, Supporting information).

The stickleback gut microbiome is sexually dimorphic, and MHC genotype include some clades that include pathogens. One of the strongest effects is seen in the family Anaplasmataceae, which includes the often-pathogenic Wolbachia, Rickettsia and Neorickettsia. It is not clear at present whether the Anaplasmataceae are (i) pathogens infecting stickleback, (ii) pathogens infecting helminth parasites such as nematodes, or (iii) probiotic members of the fish gut microbiota that protect against helminth infections. The family Legionellaceae is also strongly associated with MHC in this sample and includes some notable pathogens. Other families strongly associated with MHC include Acetobacteraceae, Beijerinckiaceae, Gemmataceae, Pirellulaceae, Streptomyctaceae and Xanthobacteraceae; their function roles in the stickleback gut remain unknown.

Curiously, most MHC motif effects were significant in only one sex and therefore exhibited a significant sex MHC interaction. This is consistent with findings, from the same sample of stickleback, that stickleback sex and diet interactively modify the gut microbiota (Bolnick et al. 2014b). The mechanism by which sex alters MHC or diet effects on microbes remains obscure, though there is evidence from mammals that sex hormones influence gut microbiota composition (Koren et al. 2012; Markle et al. 2013).
A few MHC motifs disproportionately affected microbial Family abundances, whereas others were associated with few if any microbial changes. The strongest effects involved MHC motif P123 (Fig. 2), a member of a small clade of motifs that show atypically large sequence divergence from other motifs and fewer variants (Fig. S2, Supporting information). Alleles from this clade appear to segregate like a single locus (nearly all individuals carry 1 or 2 alleles from this clade). It appears this distinct MHC-like locus disproportionately affects gut microbes, given the high concentration of microbial effects. This may reflect specialization to bind microbial antigens, and/or preferential expression of these alleles in gut mucosal dendritic cells. The concentration of MHC effects on a few microbial Families is equally intriguing. The most likely explanation is that MHC PBR has evolved to recognize antigens from particular families that impose the strongest selection, either because they are severe pathogens or important mutualists.

Major Histocompatibility class II effects on microbiota relative abundances include both positive and negative changes in abundance, depending on the motif and Family, suggesting that MHC is involved in activating both tolerance and attack mechanisms. There is prior evidence for dendritic cell–mediated microbial tolerance (Steinman et al. 2003; Mortha et al. 2014) and clearance (Niess et al. 2005), so both mechanisms are plausible. A caveat is that we examined changes in relative rather than absolute microbial abundance. Consequently, we must be cautious in ascribing any given motif-microbe association as direct suppression or tolerance. For example, a MHC motif might confer susceptibility to a pathogen which reaches high absolute abundance, thereby reducing the relative abundance of other taxa whose absolute abundance may in fact be unchanged.

Curiously, the most common MHC alleles tended to have generally suppressive effects on many Families’ relative abundances. This could arise if alleles that reduce pathogen abundance are positively selected and become common. Alternatively, pathogens may evolve to escape recognition by common alleles (more than rare alleles); the resulting proliferation of one or a few pathogen taxa could indirectly reduce the relative abundance of all other taxa, generating the negative correlation shown in Fig. S8 (Supporting information).

Because individual MHC motifs tend to alter the abundance of particular microbes, the diversity of MHC alleles an individual carries also affects microbe taxon abundance, and overall diversity. We found that particular motifs altered microbial diversity, and correspondingly, the diversity of MHC was negatively correlated with overall microbial diversity. In general, MHC protein sequence divergence has a greater impact on the microbiota than does the number of unique motifs. Based on this result, we anticipate that reductions in MHC diversity within populations [e.g. in response to population bottlenecks or directional selection by parasites (Sutton et al. 2011)] might have correlated effects on gut microbial diversity. Altered microbial diversity has, in turn, been linked to health impacts such as increased risk of obesity and autoimmunity in mammals (Turnbaugh et al. 2008; Markle et al. 2013). Whether this is true in fish as well remains to be seen: body condition does covary with gut microbial diversity in both stickleback and perch (Bolnick et al. 2014). More generally, it is clear that particular microbes provide health benefits in fish, in the form of both improved growth and protection against pathogens (Balcazar et al. 2007; Vendrell et al. 2008; Merrifield et al. 2010).

The associations between MHC diversity and microbial diversity most likely reflect an emergent property of the motif by taxon associations described above. The presence of a distinctive MHC motif could simultaneously inflate MHC diversity and, by inducing tolerance of a highly competitive microbe, reduce microbial diversity. For example, Anaplasmataceae (which include pathogenic Rickettsiales) are highly abundant in some stickleback (97% of microbes), and completely absent in other stickleback. High abundance of this one taxon reduces microbial diversity (regression of PD on Anaplasmataceae abundance, $t = -4.28$, $P < 0.0001$). Alleles that apparently promote tolerance of Anaplasmataceae (P24, P71, P117; Fig. 2), thereby reduce microbial diversity. Because individuals with more diverse MHC are more likely to carry these motifs, MHC diversity can reduce microbial diversity. Conversely, individuals with low MHC diversity carry on...
average more Firmicutes (Fig. S15, Supporting information), which elevates microbial phylogenetic diversity.

One important caveat regarding our results is that we have taken a strictly correlative approach. Because MHC of congenic laboratory mice affects gut microbes (Toivanen et al. 2001), we expect our results reflect a causal effect of MHC. However, we cannot rule out the possibility that this correlation arises through mutual association with an unmeasured variable. In particular, MHC allele variants might be in linkage disequilibrium with another innate or adaptive immune gene that regulates microbiota, or a gene that indirectly affects microbiota via host diet, physiology or endocrine levels. We consider this unlikely because we focused exclusively on variation within a single panmictic population where interbreeding and high effective population size maintains small areas of linkage disequilibrium around MHC. For comparison, in other lakes (Farewell and Roberts Lakes) in the same watershed, genotyping ~20 000 SNPs (via ddRADseq) shows that LD effectively decays in <10 kb (D. I. Bolnick, J. Weber, G. Bradburd & W. E. Stutz, unpublished). However, our results would ultimately benefit from experimental validation using transgenic or congenic approaches that vary MHC genotype against a controlled background, as has been done in mice (Toivanen et al. 2001). However, such laboratory studies typically entail a less-than-natural microbiota (Chandler et al. 2011; Bolnick et al. 2014a) and may not happen to sample the few critical MHC motifs (or microbe Families), so there is some benefit gained from correlative studies in completely natural settings.

A second caveat is that MHC may indirectly affect the microbiota. Numerous studies have documented effects of MHC II on resistance (or susceptibility) to helminth parasites (Hedrick 2002; Wedekind 2005; Pierney & Oliver 2006; Spurgin & Richardson 2010; Eizaguirre et al. 2012; Yasukochi & Satta 2013). These helminth parasites could plausibly alter the gut microbiota, leading to an indirect effect of MHC on gut microbes. This indirect effect can arise in several ways. First, parasite infections may introduce microbial helminth symbionts or pathogens. As we do not have detailed helminth infection data for these individual fish, we cannot at present rule out this possibility. Helminths can also alter the environment in the gut lumen, competing for

![Fig. 3](image-url) Positive (green) and negative (blue) effects of MHC divergence on microbial Family relative abundance within each sex and overall. MHC divergence is measured as the mean pairwise amino acid divergence among motifs within each individual. Significant ($P < 0.05$) positive and negative effects are indicated by coloured cells, using GLMs within each sex separately (columns for males and females), or in a combined model with effects of MHC divergence (‘overall’), sex, and a sex*MHC interaction. To the right of the heatmap, we indicate which microbial Families showed significant sex effects (M when the Family is more abundant in males, F when more abundant in females). An asterisk indicates microbial Families with a significant sex by MHC divergence interaction. Microbial Families are ordered taxonomically to match Fig. 2.

![Fig. 4](image-url) Relationship between MHC diversity and microbial diversity (chao1). MHC diversity can be measured either as (A) sequence diversity (mean pairwise distance of co-occurring motifs), or (B) the number of distinct motifs (regardless of their protein sequence divergence). Correlation coefficients ($r$) and $P$-values are from Spearman rank correlation tests due to the high-leverage outlier in panel A. The regression trend line is added for the significant relationship to highlight the trend. A second (dashed) significant trend line is plotted for a test removing the three lowest outliers. Similar effects are found using PD or microbial OTU richness; for these metrics, the trend in panel A is again negative ($P = 0.059$ and 0.039, respectively), and the trend in panel B is nonsignificant.
nutrients, or changing host immune state. For example, helminths induce proliferation of anti-inflammatory T_{H2} cells that have ancillary effects on microbiota (Moser & Murphy 2000). These possibilities raise the fascinating idea that helminth-driven evolution of stickleback MHC might have correlated effects on microbiota (and vice versa). Such multiway co-evolution will only occur, however, if the same MHC alleles affect both microbes and macroparasites. The alternative is that MHC paralogues are functionally specialized to recognize different taxonomic groups, or be expressed in different tissues, which could decouple MHC-microbe and MHC-helminth interactions.

Conclusions

The substantial microbial variation among individual hosts, found in many species, is a deep puzzle. Our data indicate that major histocompatibility (MHC) genes can contribute to this among-individual variation, although effect sizes are weak (cumulatively, MHC explains roughly 10% of microbial variation, Fig. S10, Supporting information). The MHC itself is highly polymorphic, thereby allowing individual hosts to recognize a diversity of antigens and generating appreciable among-individual variation. Also, MHC is essential to presenting microbial antigens to T cells, thereby inducing T-cell activation and specialization. The resulting adaptive immunity can fine-tune the hosts’ tolerance or resistance to diverse and individual-specific microbial communities in ways that fixed pattern recognition proteins cannot. Interestingly, adaptive immune regulation of the microbiota could explain why hosts whose microbiota are cleared by antibiotics nevertheless are recolonized by a microbiota that closely resembles their pre-treatment state (Dethlefsen & Relman 2011) despite exposure to different microbes. That is, the resilience of gut microbiota within a host might be a result of MHC-dependent rejection of other microbes.

Our results raise the question of whether MHC might also modify human gut microbiota, and whether this association might underlie some diseases. For example, MHC class I has been linked to type 1 diabetes (Nejentsev et al. 2007), which is also associated with dysbiosis (Markle et al. 2013). Also, MHC-associated celiac disease (Marsh 1992) coincides with changes in microbiota composition (Sanz et al. 2011). In both cases, it is unclear at present whether the microbiota change is an incidential side effect of things like altered patient diets, or whether MHC-induced changes in microbiota are a causal link in the disease. Our results highlight the value of further research into MHC/microbiota associations in humans and other species, and therapeutic implications of these associations (De Palma et al. 2010). For instance, might microbial transplants from healthy to diseased patients be more effective between MHC-matched donors and recipients? The significance of MHC/microbe connections for health will depend on exactly which microbes MHC regulates in the human gut, how strong this regulation is, and whether those microbes are probiotic or pathogenic.

To conclude, we show that among-individual differences in MHC genotype and MHC diversity are associated with variation in gut microbiota composition. Our findings raise a variety of interesting questions. What mechanism explains the sex-specific effect of MHC on gut microbes? Do the same MHC alleles control gut microbes and macroparasites? If so, does host-macroparasite co-evolution (Spurgin & Richardson 2010) have ancillary effects on the gut microbiota? Finally, a growing literature suggests that MHC influences host olfactory cues, mating success (Milinski 2003), and speciation (Eizaguirre et al. 2009); might microbiota play a key role in generating those MHC-dependent olfactory cues?

Acknowledgements

This research was supported by the Howard Hughes Medical Institute (DB and RK), David and Lucille Packard Foundation (DIB).

References

Aeschlimann PB, Haberi MA, Reusch TBH, Boehm T, Milinski M (2003) Female sticklebacks Gasterosteus aculeatus use self-reference to optimize MHC allele number during mate selection. Behavioral Ecology and Sociobiology, 54, 119–126.

Atarashi K, Tanoue T, Oshima K et al. (2013) Tr1 cells terminate colitis by corrective production of IL-10. Nature, 500, 232–236.

Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009) New genera of the Verrucomicrobia from the gut of nematode herbivores in the marine environment. Molecular Ecology Resources, 9, 713–719.

Balcazar JL, de Blas I, Ruiz-Zarzuelo I et al. (2007) Enhancement of the immune response and protection induced by probiotic lactic acid bacteria against furunculosis in rainbow trout (Oncorhynchus mykiss). FEMS Immunology and Medical Microbiology, 51, 185–193.

Bates ST, Cropsey GW, Caporaso JG, Knight R, Fierer N (2011) Bacterial communities associated with the lichen symbiosis. Applied Environmental Microbiology, 77, 1309–1314.

Belkaid Y, Bouladoux N, Hand TW (2013) Effectors and memory T cell responses to commensal bacteria. Trends in Immunology, 34, 299–306.

Bergmann GT, Bates ST, Eilers KG et al. (2011) The under-recognized dominance of Verruromicrobia in soil bacterial communities. Soil Biology and Biochemistry, 43, 1450–1455.

Berner D, Stutz WE, Bolnick DI (2010) Foraging trait (co)variances in stickleback evolve deterministically and do not predict trajectories of adaptive diversification. Evolution, 64, 2265–2277.
AIDS, De Palma G, Capilla A, Nadal I. Chandler JA, Lang JM, Bhatnagar S, Eisen JA, Kopp A (2011) Incomplete recovery and indi-

© 2014 The Authors. Molecular Ecology Published by John Wiley & Sons Ltd.
ies in the model organism three-spined stickleback Gasterosteus aculeatus. BMC Evolutionary Biology, 9, 57.
Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI (2008) Worlds within worlds: evolution of the vertebrate gut microbiota. Nature Reviews Microbiology, 6, 776–788.
Lozupone CA, Knight R (2008) Species divergence and the measurement of microbial diversity. FEMS Microbiology Reviews, 32, 557–578.
Lozupone CA, Hamady M, Knight R (2006) UniFrac - an online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics, 7, 371.
Lozupone CA, Stormbaugh JJ, Gordon JI, Jansson JK, Knight R (2012) Diversity, stability, and resilience of the human gut microbiota. Nature, 489, 220–230.
Markle JGM, Frank DN, Morton-Toth S et al. (2013) Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. Science, 339, 1084–1088.
Marsh MN (1992) Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity (‘celiac sprue’). Gastroenterology, 102, 330–354.
Matthews B, Harmon LJ, M’Conigle L, Marchinko KB, Schasch H (2010) Sympatric and allopatric divergence of MHC genes in threespine stickleback. PLoS One, 5, e10948.
McClelland EE, Penn DJ, Potts WK (2003) Major histocompatibility complex heterozygote superiority during coinfection. Infection and Immunity, 71, 2079–2086.
Merrifield DL, Dimitroglou A, Bradley G et al. (2010) Probiotic applications for rainbow trout (Oncorhynchus mykiss Walbaum) I. Effects on growth performance, feed utilization, intestinal microbiota and related health criteria. Aquaculture Nutrition, 16, 504–510.
Miliński M (2003) The function of mate choice in sticklebacks: optimizing Mhc genetics. Journal of Fish Biology, 63(Suppl A), 1–16.
Mortha A, Chudnovskiy A, Hashimoto D et al. (2014) Microbiota-dependent crosstalk between macrophages and ILC3 promotes intestinal homeostasis. Science, 343, 1249288.
Moser M, Murphy KM (2000) Dendritic cell regulation of TH1-TH2 development. Nature Immunology, 1, 199–205.
Muegge BD, Kuczynski J, Knights D et al. (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science, 332, 970–974.
Nejenstev S, Howson JM, Walker NM et al. (2007) Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature, 450, 887–892.
Niess JH, Brand S, Gu X et al. (2007) CX3CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. Science, 307, 254–258.
Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. Heredity, 96, 7–21.
Proal AD, Albert PJ, Marshall TG (2013) The human microbiome and autoimmunity. Current Opinion in Rheumatology, 25, 234–240.
Prosperi MCF, De Luca A, Di Giambenedetto S et al. (2010) The threshold bootstrap clustering: a new approach to find families or transmission clusters within molecular quasispecies. PLoS One, 5, e13619.
Reusch TBH, Langefor A (2005) Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback Gasterosteus aculeatus. Journal of Molecular Evolution, 61, 531–541.
Robertson SJ, Girardin SE (2013) Nod-like receptors in intestinal host defense: controlling pathogens, the microbiota, or both? Current Opinion in Gastroenterology, 29, 15–22.
Round JL, Lee SM, Li J et al. (2011) The toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. Science, 332, 974–977.
Sanz Y, De Palma G, Laparra M (2011) Unraveling the ties between celiac disease and intestinal microbiota. International Reviews of Immunology, 30, 207–218.
Sato A, Figueroa F, O’Hughin C, Steck N, Klein J (1998) Cloning of major histocompatibility complex (Mhc) genes from three-spine stickleback, Gasterosteus aculeatus. Molecular Marine Biology and Biotechnology, 7, 221–231.
Sommer S, Courtiol A, Mazzoni CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. BMC Genomics, 14, 542.
Spor A, Koren O, Ley R (2011) Unravelling the effects of the environment and host genotype on the gut microbiome. Nature Reviews Microbiology, 9, 279–290.
Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proceedings of the Royal Society B: Biological Sciences, 277, 979–988.
Steinman RM, Hawiger D, Nussenzwerg MC (2003) Tolerogenic dendritic cells. Annual Review of Immunology, 21, 685–711.
Stutz WE, Bolnick DI (In press) A Stepwise Threshold Clustering (STC) method to infer genotypes from error-prone next-generation sequencing of multi-allele genes such as the Major Histocompatibility Complex (MHC). PLoS One.
Sun C-M, Hall JA, Blank RB et al. (2007) Small intestine lamina propria dendritic cells promote de novo generation of Foxp3 Treg cells via retinoic acid. Journal of Experimental Medicine, 204, 1775–1785.
Sutton JT, Nakagawa S, Robertson BC, Jamieson IG (2011) Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immune genes. Molecular Ecology, 20, 4408–4420.
Teh HS, Kisielow P, Scott B et al. (1988) ThyMIC major histocompatibility complex antigens and the ab T-cell receptor pathway establishes colonization by a commensal of the three-spined stickleback. Current Opinion in Rheumatology, 25, 233–237.
Turnbaugh PJ, Gordon JI (2009) The core gut microbiome, energy balance and obesity. Journal of Physiology, 587, 4153–4158.
Turnbaugh PJ, Backhed F, Fulton L, Gordon JI (2008) Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host & Microbe, 3, 213–223.
Vendrell D, Balcazar JL, de Blas I et al. (2008) Protection of rainbow trout (Oncorhynchus mykiss) from lactococcosis by probiotic bacteria. Comparative Immunology, Microbiology, and Infectious Diseases, 31, 337–345.
Wedekind C (2005) The course of malaria in mice: major histocompatibility complex (MHC) effects, but no general MHC
heterozygote advantage in single-strain infections. Genetics, 170, 1427–1430.

Wegner KM, Reusch TBH, Kalbe M (2003) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. Journal of Evolutionary Biology, 16, 224–232.

Wu GD, Chen J, Hoffmann C et al. (2011) Linking long-term dietary patterns with gut microbial enterotypes. Science, 334, 105–108.

Yasukochi Y, Satta Y (2013) Current perspectives on the intensity of natural selection of MHC loci. Immunogenetics, 65, 479–483.

Yoshimoto S, Loo TM, Atarashi K et al. (2013) Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature, 499, 97–101.

DIB conceived of the study, carried out the statistical analyses, and wrote the paper. LKS did the laboratory work to genotype gut microbiota. WES did the laboratory work and bioinformatics to genotype host MHC. RK and GC did the bioinformatic analyses to obtain OTU identifications and diversity metrics of gut microbiota from sequence data. All authors worked on editing the manuscript. The authors have no competing financial interests.

Data accessibility

The raw sequence data for gut microbiota sequencing, and host metadata, can be accessed via accession nos. ERP006029 and ERP006030 at the European Bioinformatics Institute, http://www.ebi.ac.uk. Summary data of MHC genotypes, and OTU calls and alpha and beta diversity measures, and fish metadata (sex, morphology, diet) have been deposited in Dryad (doi:10.5061/dryad.2s07s). To recreate analyses for this study, we provide:

1 A table of microbial OTUs. Each microbe OTU is a row, columns include OTU taxonomic assignment, and read counts in each individual stickleback (one column per host).

2 Tables of metadata with individual fish as rows, columns include sex, morphology, presence/absence of each MHC allele, MHC diversity metrics, microbial diversity metrics, and microbial beta diversity scores (weighted and unweighted PCoA axes).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Pairwise microbial similarity among individual humans, among stickleback, and among Eurasian Perch.

Fig. S2 Non-independence among MHC motifs.

Fig. S3 Among-individual variation in gut microbial composition.

Fig. S4 Genetic variation in MHC from Cedar Lake stickleback.

Fig. S5 Effect of MHC at various microbial taxonomic ranks.

Fig. S6 Effects of an MHC motif on weighted PCoA axes 1–2.

Fig. S7 Effects of an MHC motif on unweighted PCoA axes 1–3.

Fig. S8 Correlation between mean MHC effect and motif frequency.

Fig. S9 Null distribution of two MHC-microbe associations.

Fig. S10 Comparison of MHC, sex, diet, morphology effect sizes.

Fig. S11 Sex-specific effects of an MHC motif.

Fig. S12 MHC motif effect sizes in males vs. females.

Fig. S13 Effects of MHC diversity on abundance of three microbes.

Fig. S14 Effects of individual MHC motifs on microbial diversity.

Fig. S15 Barchart of Phylum relative abundance by MHC diversity.

Table S1 Taxonomic composition of stickleback gut microbiota from Cedar Lake, Vancouver Island, British Columbia.