Journal of Autoimmunity 148 (2024) 103290

Contents lists available at ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

Functional metagenomic analysis reveals potential inflammatory triggers associated with genetic risk for autoimmune disease

Meghan A. Berryman [a], Jorma Ilonen [b], Eric W. Triplett [a,*], Johnny Ludvigsson [c]

[a] Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA
[b] Immunogenetics Laboratory, Institute of Biomedicine, University of Turku, Turku, Finland
[c] Crown Princess Victoria's Children's Hospital and Division of Pediatrics, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

ARTICLE INFO

Handling editor: C Selmi

Keywords:
ABIS
Type 1 diabetes
HLA
DR3-DQ2
DR4-DQ8
Inflammation
Autoimmune disease

ABSTRACT

To assess functional differences between the microbiomes of individuals with autoimmune risk-associated human leukocyte antigen (HLA) genetics and autoimmune protection-associated HLA, we performed a metagenomic analysis of stool samples from 72 infants in the All Babies in Southeast Sweden general-population cohort and assessed haplotype-peptide binding affinities. Infants with risk-associated HLA DR3-DQ2.5 and DR4-DQ8 had a higher abundance of known pathogen-associated molecular patterns and virulence related genes than infants with protection-associated HLA DR15-DQ6.2. However, there was limited overlap in the type of inflammatory trigger between risk groups. Supported by a high Firmicutes/Bacteroides ratio and differentially abundant flagellated species, genes related to the synthesis of flagella were prominent in those with HLA DR3-DQ2.5. However, this haplotype had a significantly lower likelihood of binding affinity to flagellin peptides. O-antigen biosynthesis genes were significantly correlated with the risk genotypes and absent from protective genotype association, supported by the differential abundance of gram-negative bacteria seen in the risk-associated groups. Genes related to vitamin B biosynthesis stood out in higher abundance in infants with HLA DR3-DQ2.5/DR4-DQ8 heterozygosity compared to those with autoimmune-protective genetics. *Prevotella* species and genus were significantly abundant in all infant groups with high risk for autoimmune disease. The potential inflammatory triggers associated with genetic risk for autoimmunity have significant implications. These results suggest that certain HLA haplotypes may be creating the opportunity for dysbiosis and subsequent inflammation early in life by clearing beneficial microbes or not clearing proinflammatory microbes. This HLA gatekeeping may prevent genetically at-risk individuals from benefiting from probiotic therapies by restricting the colonization of those beneficial bacteria.

1. Introduction

Environmental and genetic factors contribute in concert to the onset of autoimmune diseases like type 1 diabetes [1–5]. The genetic region with by far the greatest association with autoimmunity is the major histocompatibility complex (MHC) region [6,7]. Broadly categorized into three classes, human leukocyte antigen (HLA) alleles play an integral role in the initiation of immune response to foreign entities. Predominantly expressed on the surface of antigen presenting cells, class II HLA proteins, like HLA DR and HLA DQ, present externally found antigens, such as bacterial or foreign peptides, to $CD4^+$ T cells. Specificity in antigen recognition from both the HLA protein and T cell receptors is foundational to an effective immune reaction. Class II HLA DR and DQ loci are also highly polymorphic with elevated linkage disequilibrium, which results in characteristically differential HLA haplotype proteins [8]. Despite differing etiologies, many autoimmune diseases share the same risk-associated genetics often resulting in comorbidity. For example, HLA haplotypes DR3-DQ2.5 and DR4-DQ8 are risk-associated for future type 1 diabetes, celiac disease, autoimmune thyroiditis, and autoimmune arthritis [5].

The role of the gut microbiome has become increasingly evident in defining the pathogenesis of these autoimmune diseases [9–14]. There is a distinct need for research into autoimmune-associated genetic influence on immune response to commensal bacteria in humans. The association between genetic polymorphisms within the MHC region and gut microbiome composition at the intestinal epithelium has been detected

* Corresponding author.
E-mail address: ewt@ufl.edu (E.W. Triplett).

https://doi.org/10.1016/j.jaut.2024.103290
Received 24 October 2023; Received in revised form 28 April 2024; Accepted 14 July 2024
Available online 20 July 2024

in mice and the development of diabetes in NOD mouse model has been found to be dependent on microbial exposure [15–18]. Individuals with autoimmune diseases often experience intestinal upset and altered microbial composition and patterns emerge when comparing HLA-associated and autoimmune-associated microbiomes [5]. HLA genotypes may be predisposing an individual to systemic inflammation originating from the gut microbiome by clearing beneficial microbes and creating dysbiosis early in life.

The All Babies in Southeast Sweden (ABIS) general-population pediatric cohort provides a unique repository from which to compare genetic communities during development and establishment of the gut microbiome prior to disease onset. ABIS-based research shows significant correlations between autoimmune-associated HLA haplotypes and potentially dysbiotic microbial composition in children years before autoimmune disease onset [19–22]. Stool samples from ABIS children at approximately one year of age showed that HLA haplotypes have significant influence over whole gut microbiome composition [19], as well as total abundance and relative abundance of *Bifidobacterium* and *Lactobacillus* strains [20,22]. Additionally, through ABIS, correlations between gut microbiome composition and future type 1 diabetes diagnosis are seen over a decade before disease occurs [21].

The objective of this study was to elevate microbiome analysis beyond 16S rRNA-based taxonomic composition into defining the metabolic potential of commensal microbiota through metagenomic long-read sequencing. Metagenomic analysis was performed using stool samples from infants with either high-risk or protective HLA alleles for type 1 diabetes or celiac disease. We hypothesized that the differences observed could answer questions about how microbial composition leads to inflammation and autoimmunity.

2. Material and methods

2.1. Sample collection and institutional review board approval

This analysis is a portion of a broader study called All Babies In Southeast Sweden (ABIS). ABIS is a prospective, general-population, cohort of 17,055 individuals from southeast Sweden born between October 1, 1997 and October 1, 1999. Oral, written, and video information about the study informed parental consent. Following parental consent, participating parents completed questionnaires and diaries from infant birth through the first year of life [23]. Collected information included but was not limited to birth type, infection history, antibiotic use, duration of breastfeeding, dietary introductions, living conditions, and additional environmental factors. Stool samples were collected at an average of 1 year of age from 1733 and processed as previously reported [19]. Approval was obtained by the Research Ethics Committees of the Faculty of Health Science at Linköping University, Sweden, Ref. 1997/96287 and 2003/03–092 and the Medical Faculty of Lund University, Sweden (Dnr 99227, Dnr 99321) as described previously [19]. Microbial analysis performed at the University of Florida was approved by the University of Florida's Institutional Review Board as an exempt study IRB201800903.

2.2. HLA genotyping

HLA DR-DQ haplotypes were defined using typing for DRB1, DQB1, and DQA1 alleles for common European DR-DQ haplotypes. Typing of serum was performed using hybridization with sequence-specific lanthanide-labelled oligonucleotide probes [24]. The identified HLA DR-DQ haplotypes and corresponding confirmed alleles are as follows: DR3-DQ2.5 (*DRB1*03-DQA1*05:01-DQB1*02:01*); DR4-DQ8 (*DRB1*04-DQA1*03-DQB1*03:02*); DR15-DQ6.2 (*DRB1*15-DQA1*01-DQB1*06:02*). Those at the highest genetic risk for developing autoimmunity are homozygous for DR3-DQ2.5 or DR4-DQ8 or a heterozygous combination of the two. The most protective haplotype against developing autoimmunity is DR15-DQ6.2.

2.3. DNA Extraction, Sequencing and processing

DNA was extracted from 72 ABIS stool samples using the E.Z.N.A Stool DNA Kit according to manufacturer's protocol (Omega Bio-tek, Inc., Norcross, GA). Preparation of sequencing library was performed using the Ligation Sequencing Kit XL V14 for long fragments (Oxford Nanopore Technologies, Oxford, UK) according to the manufacturer's protocol. Sequencing was performed using the GridIONx5 (Oxford Nanopore Technologies, Oxford, UK) with one sample per GridION Flow Cell R10 (Oxford Nanopore Technologies, Oxford, UK). An average of 7.68 Gb per sample and 5.16 M reads per sample were generated. Raw fast5 data was basecalled using Guppy v6.4.2. Following basecalling, fastq files were assembled in sequential mode via metaFlye, contigs under 200 base pairs were removed using prinseq, RNAs were predicted using Barrnap, 16S rRNA sequences were taxonomically classified with the RDP classifier, open reading frames were predicted with Prodigal, contigs were mapped with minimap2, and function was annotated using Diamond with the KEGG database through SqueezeMeta v1.6.0 [25]. SqueezeMeta output files were extracted, converted into phyloseq objects via the phyloseq R package, and further analyzed in R Studio [26].

2.4. Statistical analysis

All packages were run on RStudio v2023.03.0 + 386 [27]. Variables impacting binomial beta diversity were tested for using the permutational multivariate analysis of variance (PERMANOVA) test through the adonis function in the package vegan (version 2.4–6) [28]. Core taxonomic and functional profiles were determined through the R package PIME v0.1.0 [29]. The associated principle coordinate analysis (PCoA) plot was created through the plot_ordination function. Chi square analysis was performed with R function chisq.test. Differential analysis was performed with R packages, DESeq2 and edgeR [30,31]. Wilcoxon rank sum test using the R function wilcox.test. Logistic regression was calculated using R function glm. Odds ratio was calculated with MedCalc Software [32]. P-values were corrected for false discovery rate (FDR) using the Benjamini-Hochberg method using R function p.adjust. Volcano plots were created in R package EnhancedVolcano [33]. Heatmaps were designed using the R package pheatmap (version 1.0.12) [34].

2.5. Binding affinity

Binding affinity calculations were performed with NetMHCIIpan version 4.1 [35]. The amino acid sequences for flagellin (Uniprot: P04949) from *Escherichia coli* (strain K12) and dTDP-4-amino-4,6-dideoxygalactose transaminase (Uniprot: F4V8B3) from *Escherichia coli* (strain TA280) were acquired from Uniprot [36]. Each amino acid sequence was separated into 15 peptide length sequences and run against DRB1*0301, DRB1*0401, DRB1*1501, DQA1*0501-DQB1*0201, DQA1*0301-DQB1*0302, and DQA1*0102-DQB1*0602. For flagellin, a total of 484 peptide sequences were created. For dTDP-4-amino-4,6-dideoxygalactose transaminase, a total of 362 peptide sequences were created. Threshold for strong binder was set at 1 % rank and weak binder was set at 5 % rank.

3. Results

3.1. Cohort description and study design

ABIS is a prospective, general-population cohort of 17,055 individuals born between October 1, 1997 and October 1, 1999 throughout southeast Sweden [23]. A biobank of stool, urine, blood, and hair was collected from birth to 20 years. Metadata was attained in the form of HLA genotyping, questionnaires, and first-year diary including delivery mode, duration of breastfeeding, infection history, antibiotic use, introduction to and frequency of certain foods, living conditions,

and additional environmental factors. All infants were free of any autoimmune disease diagnoses at the time of stool sampling.

This study is a metagenomic analysis of stool samples from 14 homozygous HLA DR3-DQ2.5 (DR3), 20 homozygous HLA DR4-DQ8 (DR4), and 14 heterozygous HLA DR3-DQ2.5/DR4-DQ8 (DR3/DR4) infants at a single time point, each one-to-one matched on region of residence and age of stool collection to a homozygous HLA DR15-DQ6.2 (DR15) infant, 24 DR15 infants in total (Supplementary Table 1). Given their positive associations with type 1 diabetes, celiac disease, autoimmune thyroiditis, and autoimmune arthritis, DR3, DR4, and DR3/DR4 were considered to be risk-associated [5]. DR15 was considered to be protection-associated for the purpose of this analysis because it is protective against type 1 diabetes, although it is known to be associated with multiple sclerosis [37,38]. All samples analyzed here had greater than 100,000 genes per million reads. To reduce noise, genes with under 1000 counts were removed (mean: 55,505 counts, max: 1,471,284 counts). To normalize data, following matching, each cohort was rarefied to 10,000 genes per sample, which allowed for the inclusion of the greatest number of samples and genes. The DR3 vs. DR15 matched cohort (3v15) had 985 unique genes. The DR4 vs. DR15 matched cohort (4v15) had 989 unique genes. The DR3/DR4 vs. DR15 matched cohort (3/4v15) had 985 unique genes. The average age of stool collection for 3v15 was 11.9 months (±1.8 months). The average age of stool collection for 4v15 was 12.6 months (±1.4 months). The average age of stool collection for 3/4v15 was 13.0 months (±2.7 months). Neither alpha diversity nor beta diversity between risk groups and matched protective controls were significantly different. None of the metadata provided had a significant impact on beta diversity of any of the cohorts. According to the Chi-square test there was no significant difference between the proportion of each metadata variable between the risk and protective haplotype groups in each cohort, except there were a significantly greater number of DR3/DR4 infants that went on to get an autoimmune disease compared to the matched DR15 infants ($p = 0.008$). None of the infants in this study were described by their parent as having a diet that included breast milk at the time of stool collection.

A risk-comparison cohort comprised of 13 DR3 children and 13 DR4 children (3v4) were matched on region and age of stool collection with an average age of stool collection of 12.3 months (±1.1 months). There was no difference in alpha diversity or beta diversity between DR3 and DR4.

Genomes were assembled, annotated, and assigned function through the SqueezeMeta pipeline [25]. Taxonomic assignment were classified using RDP [39]. Functional assignments of genes were designated from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database using the best average hit based on average bitscore [40]. All gene outputs were presented in the form of a KEGG ID and corresponding KEGG pathway orthology.

3.2. Taxonomic abundance analysis of whole microbiome

Taxonomic analysis of each cohort was performed. Relative abundance comparisons and differential abundance analysis helped define the significant phyla, genera, and species within each group. Taxonomic differences at the phylum level highlighted a doubled Firmicutes/Bacteroides ratio in the DR3 group compared to the other three groups (Fig. 1A). Firmicutes made up the largest percent of the microbiomes from each group. While Bacteroides made up the second largest percentage of the DR15, DR4, and DR3/DR4. Proteobacteria was the second

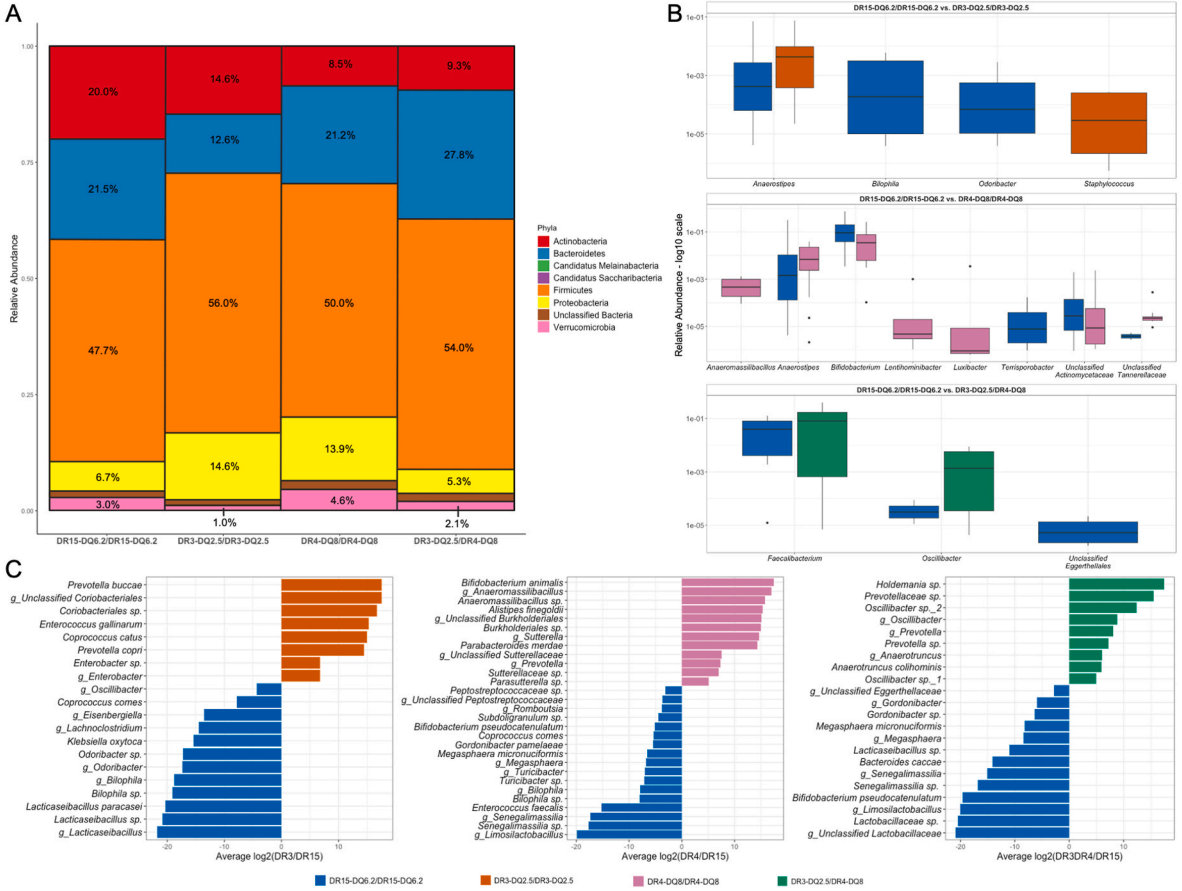

Fig. 1. Taxonomic comparison between cohort groups. A. Percent relative abundances of phyla in each group. **B.** Statistically significant relative abundance of genera within each cohort. $p < 0.05$. **C.** Average differential abundance of genera and species within each cohort. Species name is given if determined. LFC $> |2|$, $p < 0.05$.

most relatively abundant bacteria for the DR3 group. Also, percent Actinobacteria was over double in the DR15 group compared to the DR4 and DR3/DR4 groups. Wilcoxon rank sum test did not identify any phylum as significantly associated with one group over another.

Bacteria at the genus level were analyzed for significant difference in relative abundance according to Wilcoxon rank sum test (Fig. 1B). DR3 had a higher relative abundance of *Anaerostipes* and *Staphylococcus* compared to DR15. *Bilophila* and *Odoribacter* were more relatively abundant in DR15. Bacteria at both the genus level and species level were analyzed for differential abundance between risk and protective group in each cohort (Fig. 1C). At the genus level, *Enterobacter* and an unclassified Coriobacteriales genus were associated with DR3; while *Bilophila*, *Eisenbergiella*, *Lachnoclostridium*, *Lacticaseibacillus*, *Odoribacter*, and *Oscillibacter* were DR15 associated. At the species level, *Coriobaceriales* sp., *Coprococcus catus, Enterobacter* sp., *Enterococcus gallinarum, Prevotella buccae,* and *Prevotella copri* were associated with DR3 compared to the protective group. *Bilophila* sp., *Coprococcus comes, Klebsiella oxytoca, Lacticaseibacillus* sp., *Lacticaseibacillus paracasei,* and *Odoribacter* sp. were associated with DR15 compared to the risk group. Two *Prevotella* species were associated with DR3 and two *Lacticaseibacillus* species were associated with DR15. *Lactocaseibacillus* sp. was also differentially abundant in DR15 compared to DR3/DR4. *Bilophila* and *Ordibacter* stood out as both significant in relative abundance and differential abundance for DR15 compared to the risk group.

The DR4 group had a higher relative abundance of *Anaeromassilibacillus*, *Anaerostipes, Lentihominibacter, Luxibacter,* and an unclassified Tannerellaaceae genus compared to the DR15 group (Fig. 1B). *Bifidobacterium, Terrisporobacter,* and an unclassified Actinomycetaceae genus were more relatively abundant in DR15. The DR4 differentially abundant genera were *Anaeromassilibacillus*, an unclassified Burkholderiales genus, *Prevotella*, *Sutterella,* and an unclassified Sutterellaceae compared to the protective group (Fig. 1C). The genus *Prevotella* was also differentially abundant in DR3/DR4 compared to the protective group. Seven genera were associated with DR15 compared to DR4—*Bilophila, Limosilactobacillus, Megasphaera,* an unclassified Peptostreptococcaceae genus, *Romboutsia, Senegalimassilia,* and *Turicibacter.* At the species level, *Allistipes finegoldii, Anaeromassilibacillus* sp., *Bifidobacterium animalis, Burkholderiales* sp., *Parabacteroides merdae, Parasutterella* sp., and *Sutterellaceae* sp. were differentially abundant in DR4. *Bifidobacterium pseudocatenulatum, Bilophila* sp., *Coprococcus comes, Enterococcus faecalis, Gordonibacter pamelaeae, Megasphaera micronuciformis,* a Peptostreptococcaceae species, *Senegalimassilia* sp., *Subdoligranulum* sp., and *Turicibacter* sp. were associated with DR15 compared to the risk group. *Bilophila* sp. and *C. comes* were also differentially abundant in DR15 compared to DR3. *M. micronuciformis, Senegalimassilia* sp., and *B. pseudocatenulatum* were also differentially abundant in DR15 compared to DR3/DR4. *Anaerostipes* stood out as both significant in relative abundance and differential abundance for DR4 compared to the protective group.

The DR3/DR4 group had higher relatively abundance of *Oscillibacter* and *Faecalibacterium* compared to the protective group (Fig. 1B). Only an unclassified Eggerthellales genus was more relatively abundant in the DR15 group. Three genera were differentially abundant in the DR3/DR4 group compared to the protective group—*Anaerotruncus, Oscillibacter,* and *Prevotella* (Fig. 1C). Six genera were associated with DR15 compared to DR3/DR4—an unclassified Eggerthellaceae genera, *Gordonibacter,* an unclassified Lactobacillaceae genus, *Limosilactobacillus, Megasphaera,* and *Senegalimassilia*. At the species level, *Anaerotruncus colihominis, Holdemania* sp., two *Oscillibacter* sp., an unclassified Prevotellaceae species, and a *Prevotella* sp. were differentially abundant in DR3/DR4. *Bacteroides caccae, Bifidobacterium pseudocatenulatum, Gordonibacter* sp., an unclassified Lactobacillaceae species, *Lacticaseibacillus* sp., *Megasphaera micronuciformis,* and *Senegalimassilia* sp. were associated with DR15 compared to the risk group. *Oscillibacter* stood out as both significant in relative abundance and differential abundance for DR3/DR4 compared to the protective group.

3.3. Functional profile of whole metagenome for genetically at-risk population

Analysis of the genes that define the whole metagenome of the autoimmune-risk population was performed. Wilcoxon rank sum test, differential abundance analysis, and logistic regression modeling helped define the significant differences between the functional profile of the risk-associated gut and the protective-associated gut in each cohort. A heatmap of the genes determined to be significantly associated with the risk genotype in each cohort through Wilcoxon rank sum test was created (Fig. 2A). Significance symbols represent within cohort significant differences in gene abundance between the risk group and the protective group. Color and intensity show gene z-score across cohorts. Euclidean distance grouped genes into dendrogram clades. Each clade was predominantly represented by whether a gene is equally likely to be in all risk genotypes, more likely to be in two of the risk genotypes, or more likely to be in just one of the risk genotypes. Across cohort z-score analysis identified which genes were prominent in each genotype. Differential abundance analysis further indicated genes that were risk associated within each cohort (Fig. 3). Also, logistic regression modeling identified genes that had the probability of being more abundant in risk verses a protective group in each cohort (Table 1).

3.3.1. Virulence related genes dominate the HLA DR3-DQ2.5-associated metagenome

Twenty-eight genes were significantly more abundant in the DR3 group than the protective group in the 3v15 cohort. Over 46 % of the DR3-associated genes were virulence related genes (n = 13), which has an odds ratio of 3.46 (p = 0.06) compared to the number of virulence related genes in the protective DR15 group (n = 4). The DR3-associated virulence related genes were three flagellar assembly genes, *flgM* (K02398), *fliA* (K02405), *motA* (K02556), three *Caulobacter* cell cycle stalked phase genes, *lon* (K01338), *pleD* (K02488), *murG* (K02563), two biofilm formation genes, *glgP* (K00688), *hfq* (K03666), one O-antigen biosynthesis gene, *wecE* (K02805), one pilus gene, *cpA* (K02278), one type IV secretion system gene, *virB11* (K03196), one SOS response factor, *lexA* (K01356), one outer membrane vesicle gene (K06027) (Supplemental Table 2). Of the 28 significantly DR3-associated genes, nine were dominant in the DR3 group compared to the other risk groups across cohorts according to z-score (Fig. 2A). Of note, all the risk-associated flagella assembly and *Caulobacter* cell cycle genes were found in high abundance in the DR3 group. Logistic regression identified seven genes with a significantly increased probability of abundance in the DR3 group compared to the DR15 group (Table 1). Six genes of those genes were significant also in Wilcoxon rank sum test—two flagellar assembly genes, *fliA* (K02405), *motA* (K02556), one *Caulobacter* cell cycle gene, *lon* (K01338), one O-antigen biosynthesis gene, *wecE* (K02805), PTS transport gene, *crr/ptsG* (K02749), tRNA modification gene, *selD* (K01008). Differential abundance analysis highlighted four genes associated with DR3—two flagella assembly genes, *fliA* (K02405), *motA* (K02556), O-antigen biosynthesis gene, *wecE* (K02805), and PTS gene, *crr/ptsG* (K02749) (Fig. 3). All four genes were also significantly associated with DR3 by Wilcoxon rank sum test and logistic regression. Of the four genes significantly associated with the DR3 group in all three tests, *fliA* (K02405), *motA* (K02556), and *wecE* (K02805) were also risk-associated in the core metagenome (Supplemental Table 3). Flagella-related and O-antigen related genes persisted as significantly associated with the DR3 group metagenome across all statistical tests.

3.3.2. Genes for outer membrane proteins abundant in DR4-DQ8-associated metagenome

Twenty-nine genes were significantly more abundant in the DR4 group than the protective group in the 4v15 cohort. Over 27 % of the significantly DR4-associated genes (n = 8) were virulence related—three O-antigen biosynthesis genes, *wecE* (K02805), *perE/rfbE* (K13010), *wbpO* (K02474), antimicrobial resistance gene, *ddl* (K01921),

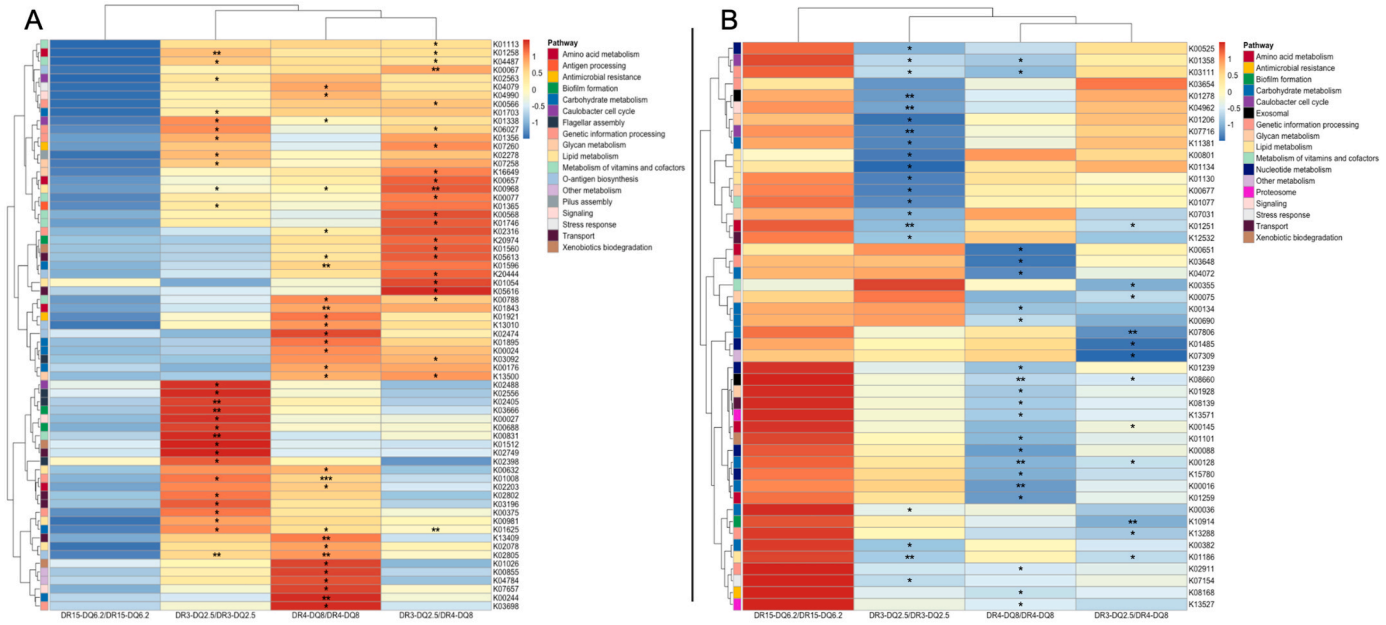

Fig. 2. Gene abundance comparison heatmap. A. Genes significantly more abundant in the risk-associated group compared to the protective-associated group in each cohort. **B.** Genes significantly more abundant in the protective-associated group compared to the risk-associated group in each cohort. Within cohort significant differences between risk and protective groups was determined by Wilcoxon rank sum test indicated by asterisk. Across cohort by row z-score designated by color. Dendrogram by Euclidian distance. $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***.

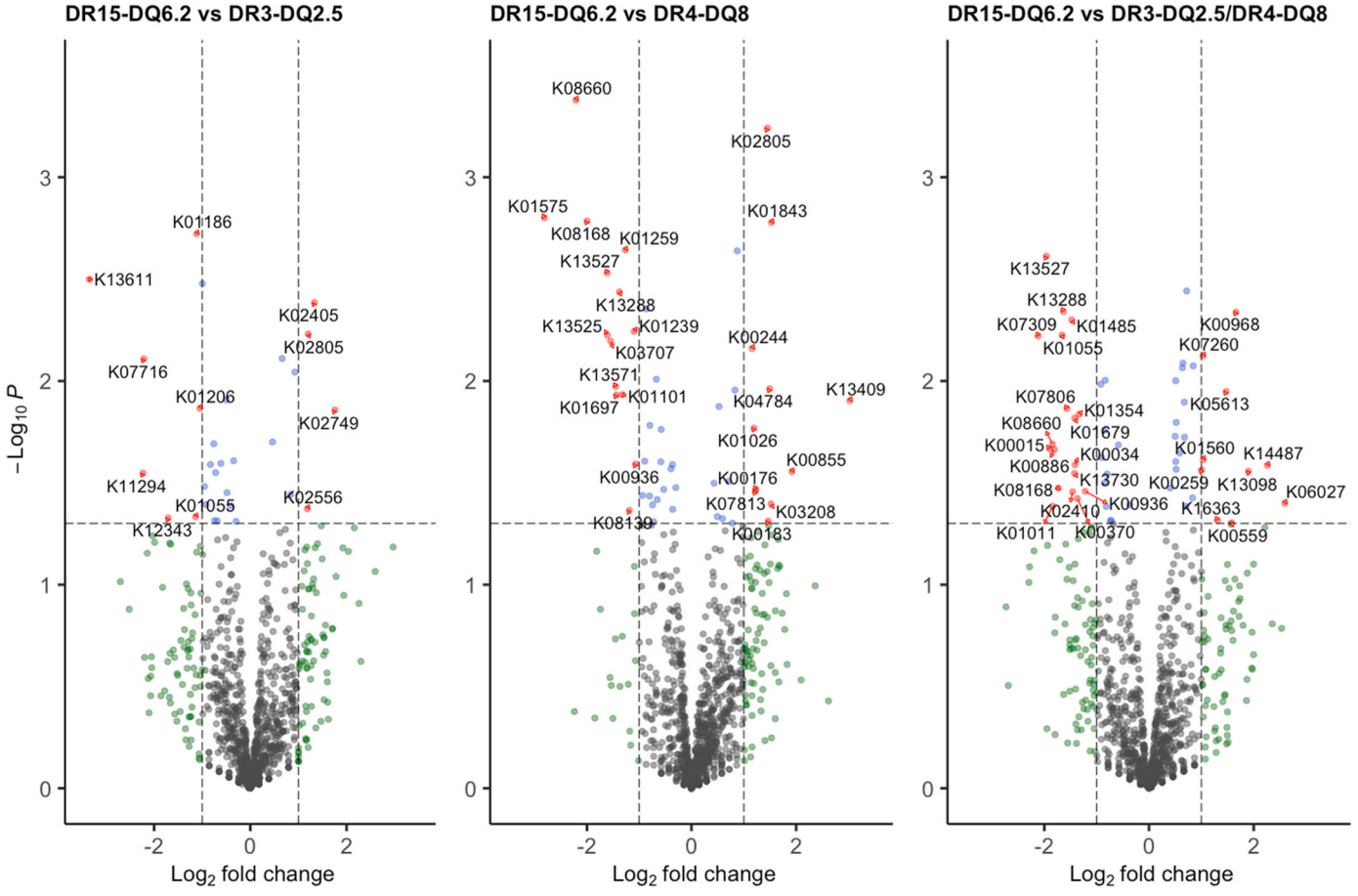

Fig. 3. Differential abundance of genes within each cohort. Protective group: Log_2 fold change (LFC) < 0. Risk group: LFC > 0. Significant genes considered to be LFC ≥ 1 and $p < 0.05$. Red: significant differential abundance. Blue: $p < 0.05$, LFC < 1. Green: $p \geq 0.05$, LFC ≥ 1.

Table 1
Significant genes from logistic regression modeling in each cohort. Only $p < 0.05$ included. Negative T value is protective associated. Positive T value is risk associated.

3v15				4v15				3/4v15			
KEGG	P value	T value	Error	KEGG	P value	T value	Error	KEGG	P value	T value	Error
K00812	0.006	−2.98	3.82	**K08660**	0.005	−2.99	1.49	**K07806**	0.006	−2.96	0.77
K01077	0.007	−2.91	3.55	**K01259**	0.008	−2.81	1.81	**K01485**	0.007	−2.92	1.81
K01186	0.008	−2.90	4.50	**K01101**	0.013	−2.61	1.80	**K07309**	0.008	−2.88	1.14
K01251	0.01	−2.77	1.84	**K13527**	0.019	−2.45	2.47	**K01679**	0.011	−2.72	1.05
K03654	0.013	−2.67	4.52	**K13571**	0.022	−2.39	1.69	**K13730**	0.016	−2.57	1.03
K01278	0.016	−2.59	5.18	**K01239**	0.025	−2.34	4.44	**K08660**	0.025	−2.38	1.95
K07716	0.027	−2.34	0.98	**K01575**	0.03	−2.26	1.31	**K13527**	0.03	−2.29	3.17
K00677	0.03	−2.30	2.01	**K08168**	0.03	−2.25	1.65	**K01354**	0.036	−2.22	4.06
K01191	0.03	−2.30	6.63	**K00936**	0.033	−2.22	1.62	**K01055**	0.036	−2.21	2.14
K01206	0.032	−2.27	9.01	**K01697**	0.04	−2.12	1.77	**K00370**	0.047	−2.08	3.02
K01358	0.032	−2.27	3.66	**K03707**	0.042	−2.11	1.31	**K14487**	0.05	2.06	1.08
K13611	0.041	−2.15	1.58	**K13288**	0.046	−2.06	3.04	**K05613**	0.034	2.24	2.01
K00970	0.043	−2.14	3.06	**K00244**	0.036	2.18	10.91	**K07260**	0.031	2.29	4.19
K02556	0.04	2.17	1.57	**K00855**	0.035	2.19	1.00	**K06027**	0.027	2.34	0.49
K07260	0.035	2.22	3.47	**K13409**	0.029	2.27	0.75	**K00259**	0.014	2.65	1.16
K01338	0.027	2.35	7.03	**K04784**	0.02	2.43	4.80	**K01560**	0.012	2.69	1.29
K01008	0.021	2.47	2.43	**K01026**	0.019	2.45	2.16	**K00968**	0.003	3.27	1.40
K02749	0.017	2.57	0.99	**K00176**	0.018	2.47	0.87				
K02405	0.005	3.12	1.42	**K02805**	0.003	3.20	1.84				
K02805	0.003	3.25	1.29	**K01843**	0.003	3.23	1.78				

Table 2
Total number of flagellin peptides with binding affinity for each haplotype. Total of 484 fliC peptide sequences were run against each HLA. Total of 362 wecE were run against each HLA. Shared binding peptides refers to peptide sequences that share weak or strong predicted binding affinity with another haplotype.

Flagellin (fliC)									
HLA	Total Strong	Total Weak	Total Binding	Shared Binding Peptides					
				DR3	DQ2.5	DR4	DQ8	DR15	DQ6.2
DR3	13	18	31	–	4	5	3	0	2
DQ2.5	9	25	34	4	–	17	28	0	12
DR4	24	45	69	5	17	–	24	6	15
DQ8	11	38	49	3	28	24	–	2	15
DR15	2	15	17	0	0	6	2	–	0
DQ6.2	21	109	130	2	12	15	15	0	–
dTDP-4-amino-4,6-dideoxygalactose transaminase (wecE)									
HLA	**Total Strong**	**Total Weak**	**Total Binding**	**Shared Binding Peptides**					
				DR3	**DQ2.5**	**DR4**	**DQ8**	**DR15**	**DQ6.2**
DR3	8	22	30	–	7	12	5	1	6
DQ2.5	4	18	22	7	–	5	10	1	5
DR4	8	29	37	12	5	–	5	2	8
DQ8	7	12	19	5	10	5	–	1	5
DR15	4	10	14	1	1	2	1	–	1
DQ6.2	3	16	19	6	5	8	5	1	–

Caulobacter cell cycle gene, *lon* (K01338), yersiniabactin gene, *irp2* (K04784), stress response gene, *htpG* (K04079), and type I secretion system gene, *raxB* (K13409) (Supplemental Table 2). Over 17 % of the significantly DR4-associated genes (n = 5) coded for the synthesis of outer membrane proteins—the three O-antigen biosynthesis genes mentioned above, the type I secretion system gene, and a OmpR regulator gene, *phoB* (K07657) (Fig. 2A). None of the protective-associated genes in this cohort were outer membrane related. Differential abundance analysis highlighted 11 genes significantly more abundant in DR4 group (Fig. 3). Eight of differentially abundant genes were significant also in Wilcoxon rank sum test and logistic regression—three carbon fixation genes, *korD/oorD* (K00176), *frdA* (K00244) and *prkB* (K00855), one amino acid metabolism geen, *kamA,* (K01843), one yersiniabactin synthesis gene, *irp2* (K04784), one O-antigen biosynthesis gene, *wecE* (K02805), one type I secretion system gene, *cvaB/mchF/raxB* (K13409), and one styrene degradation gene, *pct* (K01026) (Table 1). Of the eight genes significant across all tests, *wecE* (K02805) and *frdA* (K00244) were also significantly associated with the DR4 group in core metagenome analysis.

3.3.3. Metabolism of vitamins and cofactors pathway genes are predominant portion of heterozygote risk group associated genes

Twenty-three genes were significantly more abundant in the DR3/DR4 group than the protective group in the 3/4v15 cohort (Fig. 2A). Five of the risk group-associated genes were virulence related genes and 18 were housekeeping genes. Six of the genes code for metabolism of vitamins and cofactor proteins (26.1 %). The six vitamin and cofactor metabolism genes were from vitamin B1 (K04487, K00788), vitamin B5 (K00077), vitamin B9 (K01113), one carbon pool by folate (K01746), and ubiquinone metabolism pathways (K00568) (Supplemental Table 2). This high proportion of vitamin and cofactor metabolism genes in significant compared to the number of vitamin and cofactor metabolism genes in the DR3 (n = 2) and DR4 (n = 1) high abundant genes (OR = 6.35, p = 0.02). Only two of the 23 genes were dominant in the DR3/DR4 group compared to risk groups across cohorts. The two DR3/DR4 dominant genes were from transport, *slc1a5* (K05616), and lipid metabolism, *mgll* (K01054). The DR3/DR4 group shared 40 significantly abundant genes in common with at least one other risk group according to z-score, 19 were shared with only the DR4 group and 10 were shared

with only DR3 (OR = 2.7, p = 0.04). Differential abundance analysis highlighted ten genes associated with DR3/DR4 (Fig. 3). Five of the ten differentially abundant genes were also significantly associated with DR3/DR4 in Wilcoxon rank sum test and logistic regression modeling—one vancomycin resistance gene, *vanY* (K07260), one outer membrane vesicle trafficking protein (K06027), one glycerophospholipid metabolism gene, *tagD* (K00968), one glutamate transport gene, *slc1a3* (K05613), and one chloroalkene degradation gene (K01560) (Table 1). Of the three genes significantly associate with DR3/DR4 in the whole metagenome analysis, *tagD* (K00968), *slc1a3* (K05613), and xenobiotic degradation gene (K01560) were also risk associated in the core metagenome analysis of 3/4v15 (Supplemental Table 3).

3.3.4. Limited overlap between all three risk groups

Of the 80 significant genes associated with risk groups, only two genes were significant in all three risk groups in their respective cohorts according to Wilcoxon rank sum test—lipid metabolism gene, *tagD* (K00968), and carbohydrate metabolism gene, *eda* (K01625) (Fig. 2A). Three genes were significant in DR3 and DR4 alone—tRNA modification gene, *selD* (K01008), *Caulobacter* cell cycle gene, *lon* (K01338), and O-antigen biosynthesis gene, *wecE* (K02805). Three genes were significant in DR3 and DR3/DR4—amino acid metabolism gene, *pepT* (K01258), thiamine metabolism gene, *iscS* (K04487), and membrane trafficking gene (K06027). Three genes were also significant in both DR4 and DR3/DR4—thiamine metabolism gene, *thiE* (K00788), glutamate transporter gene, *SLC1A2* (K05613), glycan biosynthesis gene, *fcbD* (K13500). One gene stood out as significant according to differential abundance analysis, logistic regression modelling, and Wilcoxon rank sum test in both DR3 and DR4 risk groups compared to protective in their cohorts—O-antigen biosynthesis gene, *wecE* (K02805) (Supplemental Table 2). The gene, *wecE*, was also significantly associated with DR3 and DR4 when analyzing the core metagenomes (Supplemental Table 3). No genes were significant according to all three tests in all three risk groups. Eleven genes were equally likely to be in all risk genotypes across cohorts according to z-score.

Risk groups DR3 and DR4 were compared to identify genes that could define the differences between the groups. Fourteen genes were differentially abundant and significantly different between DR3 (n = 4 genes) and DR4 (n = 10 genes) (Supplemental Table 4). All significantly different genes were housekeeping genes. Three of the 10 DR4-associated genes were also significant in logistic regression—lipid metabolism gene, *atsA/arsA* (K01134), LPS biosynthesis gene, *lpxA* (K00677), and lipid biosynthesis gene, *pksJ/baeJ* (K13611).

3.4. Functional profile of whole metagenome for genetically protected population

Analysis of the genes that define the whole metagenome of the autoimmune-protected population was performed. Wilcoxon rank sum test, differential abundance analysis, and logistic regression modeling helped define the protective-associated gut in each cohort. A heatmap of the genes determined to be significantly associated with the protective genotype in each cohort through Wilcoxon rank sum test was created (Fig. 2B). Significance symbols were used to represent within cohort significant differences in gene abundance between the risk group and the protective group. Color and intensity showed gene z-score across cohorts. Euclidean distance was used to group genes into dendrogram clades.

3.4.1. Metabolism genes predominantly associated with protective group in DR3-DQ2.5 vs. DR15-DQ6.2 cohort comparison

Twenty genes were significantly more abundant in the DR15 group than the risk group in the 3v15 cohort according to Wilcoxon rank sum test (Fig. 2B). Sixteen were housekeeping-associated genes and four virulence-associated genes. The odds ratio of the protective group housekeeping genes was significant compared to the risk group in the 3v15 cohort (OR = 4.62, p = 0.02). Three of the twenty genes were shared across cohorts. *Caulobacter* cell cycle gene, *clpP* (K01358), was also protective correlated in the 4v15 cohort. Amino acid metabolism gene, *ahcY* (K01251), and lipid metabolism gene, *nanB* (K01186), were also protective correlated in the 3/4v15 cohort. Thirteen genes had a higher probability of abundance in the DR15 group according to logistic regression modeling (Table 1). Seven genes were differentially abundant in DR15 in the 3v15 cohort (Fig. 3). Four genes were both differentially abundant and significantly probable, but only three genes were significant in all three statistical tests—*Caulobacter* cell cycle gene, *pleC* (K07716), glycan biosynthesis gene, *fucA1* (K01206), and lipid metabolism gene, *nanB* (K01186). Of the three genes significantly associate with DR15 in the whole metagenome analysis, *nanB* (K01186) was also protective associated in the core metagenome analysis of 3v15. Unlike the DR3 group, there were no flagellar assembly genes associated with DR15 in the 3v15 cohort.

3.4.2. Proteosome related genes significantly associated with DR15-DQ6.2 group in DR4-DQ8 vs. DR15-DQ6.2 cohort comparison

Twenty-one genes were significantly more abundant in the DR15 group than the risk group in the 4v15 cohort (Supplemental Table 5). Eighteen are housekeeping genes, four of which are from the glycolysis pathway. Unlike the DR4 group, there are no outer membrane associated genes significantly correlated with DR15 in the 4v15 cohort. One gene was associated with DR15 in both the 3v15 cohort and the 4v15 cohort—*Caulobacter* cell cycle gene, *clpP* (K01358) (Fig. 2B). Logistic regression identified twelve genes with a significantly increased probability of abundance in the DR15 group compared to the DR4 group (Table 1). Fourteen genes were differentially associated with the DR15 group (Fig. 3). Seven genes were significant by all three statistical tests—one was amino acid metabolism gene, *pip* (K01259); one was antimicrobial resistance gene, *tetB* (K08168); one coded for an exosomal protein of breastmilk, *CNDP2* (K08660); one was nucleotide metabolism gene, *iunH/rihA* (K01239); two were proteosomes, *mpa* (K13527) and *pafA* (K13571); and one was xenobiotic degradation gene (K01101). The gene for exosomal protein in breastmilk was significantly correlated with DR15 in both 4v15 and the 3/4v15 cohort (Fig. 2B). Of the seven genes significantly abundant in DR15 from the whole metagenome analysis, *pip* (K01259) and *iunH/rihA* (K01239) were also protective associated in the core metagenome analysis of 4v15.

3.4.3. Housekeeping related genes associated with DR15-DQ6.2 group in DR3-DQ2.5/DR4-DQ8 vs. DR15-DQ6.2 cohort comparison

Twelve genes were significantly more abundant in the DR15 group than the risk group in the 3/4v15 cohort, ten of which are housekeeping-associated genes (Supplemental Table 5). Two genes are significantly correlated with the protective group in each of the other cohorts (Fig. 2B). Logistic regression identified ten genes with a significantly increased probability of abundance in the DR15 group compared to the DR3/DR4 group (Table 1). Eighteen genes were differentially associated with the DR15 group (Fig. 3). Four genes were significant by all three statistical tests—carbohydrate metabolism gene, *arnB/pmrH* (K07806), exosomal gene, *CNDP2* (K08660), selenocompound metabolism gene, *ynfE* (K07309), and nucleotide metabolism gene, *codA* (K01485). Of the four genes significantly associate with DR15 in the whole metagenome analysis, *arnB/pmrH* (K07806) and *codA* (K01485) were also protective associated in the core metagenome analysis of 3/4v15.

3.5. Distinct differences in core metagenomes

To delve deeper into the genes that were the most prominent in each group, analysis of the core microbiomes at 90 % prevalence was performed. The intent was to identify genes that were in almost every individual of the group. The statistical analysis tool, prevalence interval for microbiome evaluation (PIME), was used to determine the core

metagenomes for each cohort [29]. With PIME, the cohorts were split by HLA genotype. Genes under 90 % prevalent within the subset were filtered out. Random forest was then used for group prediction, compute variable importance, and mean decrease accuracy (Supplemental Table 6). At 90 % prevalence, the predicted core metagenomes are distinctly different (Fig. 4A). The core metagenome of 3v15 included 415 genes, 44 were unique to DR3 and 40 were unique to DR15 (Table 3). The core metagenome of 4v15 included 378 genes, 77 were unique to DR4 and 7 were unique to DR15. The core metagenome of 3/4v15 included 419 genes, 58 were unique to DR3/DR4 and 21 were unique to DR15. Wilcoxon rank sum test was performed on the core metagenomes of each cohort to determine which gene abundances were statistically different between the risk genotype and the protective (3v15: 88 genes, 46 risk-associated, 42 protective-associated; 4v15: 87 genes, 78 risk-associated, 9 protective-associated; 3/4v15: 81 genes, 59 risk-associated, 22 protective-associated). To further determine functional differences between the risk and protective gut metagenomes, each gene was categorized according to KEGG pathway orthology (Supplemental Table 3). The number of genes that were from the same pathway were compared between risk and protective (Fig. 4).

3.5.1. *Flagellar assembly genes make up significant portion of DR3-DQ2.5-associated core metagenome*

Flagellar assembly (15.22 %, n = 7 genes) and vitamin and cofactor metabolism (13.04 %, n = 6 genes) genes made up the largest portion of risk-associated genes in the 3v15 core metagenome (Fig. 4B). The significant flagellar assembly genes were *flgM* (K02398), *flhA* (K02400), *fliA* (K02405), *fliC* (K02406), *fliG* (K02410), *motA* (K02556), and *motB* (K02557) (Supplemental Table 3). The odds ratio of a flagellar assembly gene being in the risk-associated core metagenome of this cohort was 16.14 (p = 0.06). The vitamin and cofactor metabolism genes included one vitamin B6 metabolism gene (K00831), one vitamin B7 metabolism gene (K01012), one vitamin B3 metabolism gene (K01081), one vitamin B2 and B9 metabolism gene (K14652), one carbon pool by folate gene (K01746), and one porphyrin metabolism gene (K01698). The odds ratio of specifically vitamin B related gene being in the risk-associated core metagenome of this cohort was 11.27 (p = 0.10). Antimicrobial resistance genes comprised the largest portion of protective-associated core genes (11.90 %, n = 5). However, five antimicrobial resistance genes were also risk-associated. There was no odds ratio significance.

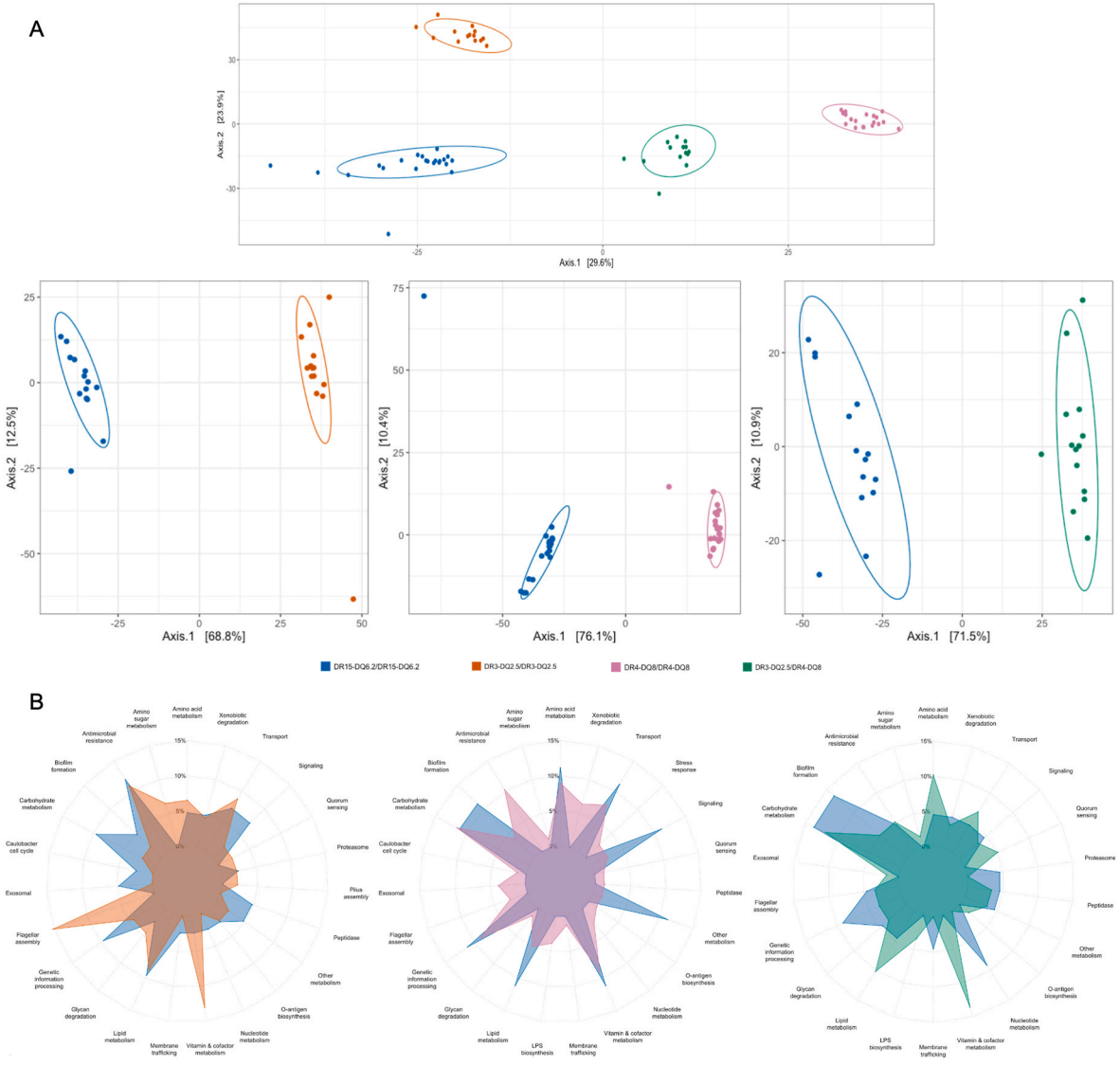

Fig. 4. Differences in the core metagenomes of each cohort and group. A. Principal coordinates analysis plot of core metagenomes at 90 % gene prevalence. Top: Across cohort comparison. Bottom: Within cohort comparison. **B.** Percent of genes categorized into pathway for each within cohort core metagenome. Only significantly different genes compared. Significance determined by Wilcoxon rank sum test. 3v15: 88 genes; 4v15: 87 genes; 3/4v15: 81 genes. $p < 0.05$.

Table 3
Number of genes that make up the core microbiomes of the groups in each cohort. PIME was used to determined core metagenome at 90 % prevalence. Wilcoxon rank sum test was used to determine significant abundance. $p < 0.05$.

Cohort	Group	Total Core Genes	Core Genes Unique to Group	Significantly Abundant Core Genes
3v15	DR15	375	40	42
	DR3	371	44	46
4v15	DR15	301	7	9
	DR4	371	77	78
3/4v15	DR15	361	21	22
	DR3/DR4	398	58	59

3.5.2. Beta-lactam and vancomycin resistance prominent in DR4-DQ8-associated core metagenome

Only nine of the 87 significantly different 4v15 genes were protective associated, none of which categorized together (Fig. 4B). All the protective-associated core genes were housekeeping related genes. The most predominant risk-associated pathways were carbohydrate metabolism (11.54 %, n = 9), antimicrobial resistance (10.26 %, n = 8), amino acid metabolism (8.97 %, n = 7), and vitamin and cofactor metabolism (8.97 %, n = 7). The carbohydrate metabolism genes were from butanoate metabolism, citrate cycle, glycolysis, pentose phosphate pathway, pyruvate metabolism, and starch and sugar metabolism. Within the antimicrobial resistance pathway, four were vancomycin resistance genes (K01929, K07260, K18345, K18346) and four were beta-lactam resistance genes (K02171, K02172, K08218, K07644) (Supplemental Table 3). There was no odds ratio significance due to the low number of protective-associated genes.

3.5.3. Vitamin B metabolism prominent in DR3-DQ2.5/DR4-DQ8-associated core metagenome

Vitamin and cofactor metabolism (13.56 %, n = 8) and carbohydrate metabolism (11.86 %, n = 7) genes comprised the largest portion of significantly risk-associated genes in the 3/4v15 core metagenome (Fig. 4B). The significant vitamin and cofactor metabolism genes were from two vitamin B2 metabolism (K09474, K14652), two one carbon pool by folate (K01746, K03465), one vitamin B3 metabolism (K01081), one vitamin B5 metabolism (K01579), one vitamin B7 metabolism (K01012), and one porphyrin metabolism pathways (K01698) (Supplemental Table 3). The largest portion of protective-associated genes in the 3/4v15 core metagenome were from biofilm formation (13.64 %, n = 3 genes) and carbohydrate metabolism (13.64 %, n = 3 genes) pathways made up the highest percentage of protective-associated genes; however, biofilm formation and carbohydrate metabolism genes were also risk-associated.

3.6. In silico predicted HLA binding affinity for flagellin differs for HLA DR3-DQ2.5

The prominence of flagellar assembly genes associated with the DR3 group compared to the protective group and other risk groups in this study led to exploration into the binding affinity of each of the HLA haplotypes studied here to flagellin peptides. NetMHCIIpan version 4.1 was used to predict binding affinity [35]. The flagellin protein from *Escherichia coli* (strain K12) was selected to test. The 498 amino acid flagellin sequence was separated into 484 peptides each with a 15 amino acid length. Each sequence was run against DR3, DR4, DR15, DQ2.5, DQ8, and DQ6.2 separately for predicted binding affinity. Affinity was categorized as strong binding or weak binding (Supplemental Table 7). The DR3-DQ2.5 haplotype had significantly fewer flagellin peptides (n = 61 unique peptides) that evoked possible binding affinity compared to the DR15-DQ6.2 haplotype (n = 147 unique peptides, OR = 0.36, $p < 0.0001$) and compared to the DR4-DQ8 haplotype (n = 94 unique peptides, OR = 0.60, p = 0.0041) (Table 2). DR3 shared four predicted flagellin binding peptides with DR4 and zero predicted binding peptides with DR15. Three of the predicted binding peptides shared between DR3 and DR4 had strong binding affinity for both HLA proteins, and one peptide had weak binding for both HLA (Supplemental Table 7). Whereas DR4 shared six predicted binding peptides with DR15, and all had weak binding with DR15. There were a greater number of shared predicted binding peptides between the DQ proteins. DQ2.5 and DQ8 shared 28 predicted flagellin binding peptides, with over 80 % of all flagellin peptides with binding affinity for DQ2.5.

Protein dTDP-4-amino-4,6-dideoxygalactose transaminase, wecE (K02805), was chosen for comparison to represent a cytoplasmic protein that would not readily be an accessible antigen for HLA binding. The *wecE* gene was significantly associated with both DR3 and DR4 across all statistical tests. The wecE protein from *Escherichia coli* (TA280) was selected to test. The 376 amino acid wecE sequence was separated into 362 peptides each with a 15 amino acid length. Each sequence was run against DR3, DR4, DR15, DQ2.5, DQ8, and DQ6.2 separately for predicted binding affinity. Affinity was categorized as strong binding or weak binding (Supplemental Table 8). There were no significant differences between total predicted wecE binding peptides for DR3-DQ2.5 (n = 45 unique peptides) and DR4-DQ8 (n = 51 unique peptides). There was no significant difference between the number of binding affinity predictions for DR3-DQ2.5 and DR15-DQ6.2 (n = 31 unique peptides). There were higher odds of DR4-DQ8 having binding affinity for wecE than DR15-DQ6.2 (OR = 1.69, p = 0.03). The number of predicted peptides with strong binding affinity was lower for all HLA except DR15 compared to the predicted binding to flagellin. The total number of predicted binding peptides were lower for wecE than flagellin in all HLA.

4. Discussion

Infants with autoimmune-risk HLA exhibited an increased potential for gut inflammation compared to children of the same age and from the same parts of Sweden who did not have risk-associated HLA. Genes coding for the synthesis of known pathogen-associated molecular patterns (PAMPs) and other virulence factors were significantly more abundant in the gut microbiomes of at-risk infants compared to protected infants. However, there was limited overlap in the type of inflammatory trigger between risk groups. Flagella assembly was notably the most prominent pathway associated with HLA DR3-DQ2.5. Whereas genes coding for outer membrane proteins were in high abundance in HLA DR4-DQ8 homozygous infants. Meanwhile housekeeping genes, essential genes for cellular maintenance like nucleotide and amino acid metabolism, were associated with the HLA DR15-DQ6.2 regardless of comparison group. This difference is of interest due to the protective nature of HLA DR15-DQ6.2 against type 1 diabetes versus the autoimmune risk association of HLA DR3-DQ2.5 and DR4-DQ8 [2,5,38]. Many studies group HLA haplotypes by autoimmune risk, thereby losing the behavioral differences between DR3-DQ2.5 and DR4-DQ8 [5,19]. Our observation that these risk haplotypes are promoting inflammation in separate ways is in accordance with the differential association of HLA risk haplotype and age of autoantibody detection as well as the preferential association of HLA DR4-DQ8 with insulin autoantibodies as the first autoantibody detected and HLA DR3-DQ2.5 with glutamic acid decarboxylase antibodies as the first autoantibody detected [41]. Differences between major T1D phenotypes have led to a proposed endotype concept to address the heterogeneity within T1D [42].

Flagellar assembly genes were a prominent feature of the DR3 group metagenome. Flagella gene abundance was buoyed by the observed high Firmicutes/Bacteroides ratio and differential abundance of flagellated bacteria like *Enterococcus gallinarum* [43,44]. Flagella is a recognized PAMP. It is detected by host toll-like receptor 5 (TLR-5), which leads to cytokine induction and inflammatory response [45]. High levels of flagellin have been previously associated with breakdown of the mucosal barrier and subsequent inflammation [44]. The prominence of flagellar assembly genes associated with the DR3 group compared to the protective and other risk groups led to exploration into the *in silico*, predicted, binding affinity of each of the HLA haplotypes studied here to flagellin peptides. Significantly fewer potential binding peptides were observed for DR3-DQ2.5 compared to DR15-DQ6.2 and DR4-DQ8, which suggested that DR3-DQ2.5 had an overall lower predicted binding affinity for flagellin. This same differential affinity was not observed when comparing affinity for a cytoplasmic protein that would not readily be an accessible antigen for HLA binding. When taking into consideration how the higher predicted binding affinity of DQ2.5 for gliadin results in a T cell mediated immune response to gluten [46], it is reasonable to consider that immune systems with DR3-DQ2.5 HLA may not be clearing flagellar bacteria at the same rate as non-DR3-DQ2.5 adaptive immune responses would. If this is the case, flagella could continue to trigger the innate immune response through TLR-5 on dendritic cells, macrophages, intestinal epithelial cells. This idea is supported by the observed higher abundance of flagellar assembly genes in the DR3 group and the association of DR3-DQ2.5 with possible dysbiosis [5].

Genes from the O-antigen biosynthesis pathway were seen in high abundance in all risk groups and never associated with the protective group across the HLA haplotypes. However, genes that coded for outer membrane proteins in general were more prominent in the DR4 group than the other risk groups. O-antigen is a highly variable and highly immunogenic carbohydrate antigen located on the outer membrane of gram-negative bacteria [47]. Specifically, O-antigen biosynthesis gene, *wecE,* was significantly higher across multiple tests for both DR3 and DR4 compared to DR15. The gene, *wecE*, codes for an aminotransferase that is required for the successful biosynthesis of enterobacterial common antigen and modification of lipid A, known PAMPs [48,49]. The consistency of *wecE* risk-associated significance suggests higher levels of Enterobacterales across both homozygous genotypes. This is supported by evidence that higher abundance of family Enterobacteriaceae is associated with HLA DR3-DQ2 and DR4-DQ8 and autoimmune disease [5]. Importantly, previous studies show that an inflamed gut is conducive to Enterobacteriaceae over growth [50].

Genes related to vitamin B biosynthesis stood out in higher abundance in the DR3/DR4 risk group compared to the DR15 protective group. Humans cannot produce their own vitamin B and must rely on the acquisition of these essential micronutrients through diet or microbial synthesis. However, not all bacteria produce vitamin B—among the vitamin B producing bacteria are *Bacteroides fragilis, Escherichia coli, Prevotella copri, Clostridium difficlie*, *Lactobacillus* sp., and *Bifidobacterium* sp. [51,52]. Non-producing bacteria must also rely on their neighboring vitamin B-producing bacteria to provide this important nutrient for them. One hypothesis is that due to this competition for nutrients, the gut microbiome is not capable to providing the host with sufficient B vitamins to fulfill daily recommendations and acquisition through diet is required [52]. Thus, an increased supply of vitamin B could be due to an increased demand by commensal bacteria or by the human host. B vitamins are necessary for the proper functioning of the immune system [51]. There were multiple trials in the 1990s and early 2000s that assessed the supplementation of nicotinamide for the prevention or delay of type 1 diabetes because of the observed prevention of autoimmune diabetes by vitamin B3 in animal models [53–55]. However, the studies found no significance between human treatment groups.

There may be a connection between bacterial vitamin B production and rheumatoid arthritis as *Prevotella copri,* associated with HLA DR4-DQ8, has also been shown to be higher in the gut of rheumatoid arthritis subjects compared to healthy controls [5,51,56]. In this study, *Prevotella* species and genus were significantly associated with all risk groups—DR3, DR4, and DR3/DR4. Further research is needed to assess how the interplay of microbial production versus acquisition of vitamin B affects immune system function.

Many large cohort studies of this kind focus exclusively on those who are at risk for autoimmune diseases. The general population nature of ABIS allows for comparison of risk-associated microbiomes to those without risk genetics. This work suggests that beneficial strains capable of colonizing the gut of high genetic risk individuals be cultured and tested in clinical trials for their efficacy in reducing inflammatory marker in the gut.

More studies could benefit from general population analysis to help determine aberrant features. Future ABIS research will involve and benefit from the inclusion of different time points along with increasing the number of children studied. A larger prospective analysis of the ABIS cohort are expected to help further define the potential indicators of dysbiosis and inflammation in the early-life gut.

The objective of this study was to elevate microbiome analysis beyond 16S rRNA sequencing and taxonomic composition into defining the possible functions of which commensal microbiota are capable. Metagenomic analysis was performed to determine the metabolic potential of microbiomes from opposing autoimmune-associated HLA groups. The differences observed between risk groups suggest how HLA-regulated microbiome composition may contribute to inflammation which, in turn, triggers autoimmunity. HLA haplotypes that increase the risk of type 1 diabetes appear to prevent the colonization of beneficial microbes and/or encourage the colonization of potentially harmful ones at a very early stage in life, year prior to autoimmune diagnosis [19–21]. The results presented here suggest that early screening for HLA haplotypes be considered at birth. Those infants found to be at high genetic risk can then be monitored over time and given beneficial microbiomes tailored for their HLA genotype when probiotic intervention is considered age appropriate. Alternatively, prescription of medications that reduce inflammation may be appropriate at an early age.

More research is needed to determine whether HLA-associated dysbiosis and subsequent inflammation is a common denominator to autoimmune diseases. This HLA gatekeeping of gut bacterial colonization may prevent genetically at-risk individuals from benefiting from probiotic therapies by restricting the colonization of those beneficial bacteria [19,20].

Funding

ABIS was supported by Barndiabetesfonden (Swedish Child Diabetes Foundation); Swedish Council for Working Life and Social Research, Grant/Award Numbers: FAS2004-1775, FAS2004–1775; Swedish Research Council, Grant/Award Numbers: K2005-72X-11242-11A and K2008-69X-20826-01-4, K2008-69X-20826-01-4; Östgöta Brandstodsbolag; Medical Research Council of Southeast Sweden (FORSS); JDRF Wallenberg Foundation, Grant/Award Number: K 98-99D-12813-01A; ALF-and LFoU grants from Region Östergötland and Linköping university, Sweden; Joanna Cocozza Foundation. This research was funded by JDRF grant no. 1-INO-2018-637-A-N.

CRediT authorship contribution statement

Meghan A. Berryman: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Jorma Ilonen:** Methodology. **Eric W. Triplett:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. **Johnny Ludvigsson:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

Data will be made available on request.

Acknowledgments

We are grateful to all ABIS participants and parents, to Ingela Johansson and Gosia Smolinska for handling biological samples, and to Åshild Faresjö and Andrea Lebena for taking care of the ABIS register data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jaut.2024.103290.

References

[1] O. Vaarala, M.A. Atkinson, J. Neu, The "perfect storm" for type 1 diabetes, Diabetes 57 (2008) 2555–2562, https://doi.org/10.2337/db08-0331.
[2] C. Wasserfall, K. Nead, C. Mathews, M.A. Atkinson, The threshold hypothesis: solving the equation of nurture vs nature in type 1 diabetes, Diabetologia 54 (2011) 2232–2236, https://doi.org/10.1007/s00125-011-2244-z.
[3] F. Pociot, Å. Lernmark, Genetic risk factors for type 1 diabetes, Lancet 387 (2016) 2331–2339, https://doi.org/10.1016/S0140-6736(16)30582-7.
[4] M. Rewers, J. Ludvigsson, Environmental risk factors for type 1 diabetes, Lancet 387 (2016) 2340–2348, https://doi.org/10.1016/S0140-6736(16)30507-4.
[5] M.A. Berryman, J. Ilonen, E.W. Triplett, J. Ludvigsson, Important denominator between autoimmune comorbidities: a review of class II HLA, autoimmune disease, and the gut, Front. Immunol. 14 (2023) 1270488.
[6] J. Trowsdale, J.C. Knight, Major histocompatibility complex genomics and human disease, Annu Rev Genomics Hum Genet 14 (2013) 301–323, https://doi.org/10.1146/annurev-genom-091212-153455.
[7] C.A. Dendrou, J. Petersen, J. Rossjohn, L. Fugger, HLA variation and disease, Nat. Rev. Immunol. 18 (2018) 325–339, https://doi.org/10.1038/nri.2017.143.
[8] I. Djilali-Saiah, V. Benini, S. Daniel, R. Assan, J.F. Bach, S. Caillat-Zucman, Linkage disequilibrium between HLA class II (DR, DQ, DP) and antigen processing (LMP, TAP, DM) genes of the major histocompatibility complex, Tissue Antigens 48 (1996) 87–92, https://doi.org/10.1111/j.1399-0039.1996.tb02612.x.
[9] C. Heidt, U. Kämmerer, M. Fobker, A. Rüffer, T. Marquardt, M. Reuss-Borst, Assessment of intestinal permeability and inflammation bio-markers in patients with rheumatoid arthritis, Nutrients 15 (2023) 2386, https://doi.org/10.3390/nu15102386.
[10] A. Piccioni, F. Rosa, S. Mannucci, F. Manca, G. Merra, S. Chiloiro, M. Candelli, M. Covino, A. Gasbarrini, F. Franceschi, Gut microbiota, LADA, and type 1 diabetes mellitus: an evolving relationship, Biomedicines 11 (2023) 707, https://doi.org/10.3390/biomedicines11030707.
[11] M. Monticolo, K. Mucha, B. Foroncewicz, Lupus nephritis and dysbiosis, Biomedicines 11 (2023) 1165, https://doi.org/10.3390/biomedicines11041165.
[12] M. Knip, H. Siljander, The role of the intestinal microbiota in type 1 diabetes mellitus, Nat. Rev. Endocrinol. 12 (2016) 154–167, https://doi.org/10.1038/nrendo.2015.218.
[13] R.E. Rossi, G. Dispinzieri, A. Elvevi, S. Massironi, Interaction between gut microbiota and celiac disease: from pathogenesis to treatment, Cells 12 (2023) 823, https://doi.org/10.3390/cells12060823.
[14] Q. Xu, J.-J. Ni, B.-X. Han, S.-S. Yan, X.-T. Wei, G.-J. Feng, H. Zhang, L. Zhang, B. Li, Y.-F. Pei, Causal relationship between gut microbiota and autoimmune diseases: a two-sample mendelian randomization study, Front. Immunol. 12 (2022) 746998, https://doi.org/10.3389/fimmu.2021.746998.
[15] J.L. Kubinak, W.Z. Stephens, R. Soto, C. Petersen, T. Chiaro, L. Gogokhia, R. Bell, N.J. Ajami, J.F. Petrosino, L. Morrison, W.K. Potts, P.E. Jensen, R.M. O'Connell, J. L. Round, MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection, Nat. Commun. 6 (2015), https://doi.org/10.1038/ncomms9642.
[16] M. Silverman, L. Kua, A. Tanca, M. Pala, A. Palomba, C. Tanes, K. Bittinger, S. Uzzau, C. Benoist, D. Mathis, Protective major histocompatibility complex allele prevents type 1 diabetes by shaping the intestinal microbiota early in ontogeny, Proc Natl Acad Sci U S A 114 (2017) 9671–9676, https://doi.org/10.1073/pnas.1712280114.
[17] L. Wen, R.E. Ley, P.Y. Volchkov, P.B. Stranges, L. Avanesyan, A.C. Stonebraker, C. Hu, F.S. Wong, G.L. Szot, J.A. Bluestone, J.I. Gordon, A.V. Chervonsky, Innate immunity and intestinal microbiota in the development of Type 1 diabetes, Nature 455 (2008) 1109–1113, https://doi.org/10.1038/nature07336.
[18] P. Toivanen, J. Vaahtovuo, E. Eerola, Influence of major histocompatibility complex on bacterial composition of fecal flora, Infect. Immun. 69 (2001) 2372–2377, https://doi.org/10.1128/IAI.69.4.2372-2377.2001.
[19] J.T. Russell, L.F.W. Roesch, M. Ördberg, J. Ilonen, M.A. Atkinson, D.A. Schatz, E. W. Triplett, J. Ludvigsson, Genetic risk for autoimmunity is associated with distinct changes in the human gut microbiome, Nat. Commun. 10 (2019) 3621, https://doi.org/10.1038/s41467-019-11460-x.
[20] M.A. Berryman, P.L. Milletich, J.R. Petrone, L.F.W. Roesch, J. Ilonen, E.W. Triplett, J. Ludvigsson, Autoimmune-associated genetics impact probiotic colonization of the infant gut, J. Autoimmun. 133 (2022) 102943, https://doi.org/10.1016/j.jaut.2022.102943.
[21] M. Bélteky, P.L. Milletich, A.P. Ahrens, E.W. Triplett, J. Ludvigsson, Infant gut microbiome composition correlated with type 1 diabetes acquisition in the general population: the ABIS study, Diabetologia 66 (2023) 1116–1128, https://doi.org/10.1007/s00125-023-05895-7.
[22] M.A. Berryman, E.W. Triplett, J. Ludvigsson, Human leukocyte antigen-dependent colonization of Lactobacillus in the early-life gut, Frontiers in Microbiomes 2 (2023) frontiersin.org/articles/10.3389/frmbi.2023.1192773 (accessed October 5, 2023).
[23] J. Ludvigsson, M. Ludvigsson, A. Sepa, Screening for prediabetes in the general child population: maternal attitude to participation, Pediatr. Diabetes 2 (2001) 170–174, https://doi.org/10.1034/j.1399-5448.2001.20405.x.
[24] J. Ilonen, M. Kiviniemi, J. Lempainen, O. Simell, J. Toppari, R. Veijola, M. Knip, Genetic susceptibility to type 1 diabetes in childhood – estimation of HLA class II associated disease risk and class II effect in various phases of islet autoimmunity, Pediatr. Diabetes 17 (2016) 8–16, https://doi.org/10.1111/pedi.12327.
[25] J. Tamames, F. Puente-Sánchez, SqueezeMeta, A highly portable, fully automatic metagenomic analysis pipeline, Front. Microbiol. 9 (2019) 3349, https://doi.org/10.3389/fmicb.2018.03349.
[26] P.J. McMurdie, S. Holmes, Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data, PLoS One 8 (2013) e61217, https://doi.org/10.1371/journal.pone.0061217.
[27] Posit team, RStudio: integrated development environment for R. http://www.posit.co/, 2023.
[28] J. Oksanen, F.G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, E. Szoecs, H. Wagner, Vegan: community ecology package. https://CRAN.R-project.org/package=vegan, 2020. (Accessed 6 April 2022).
[29] L.F.W. Roesch, P.T. Dobbler, V.S. Pylro, B. Kolaczkowski, J.C. Drew, E.W. Triplett, pime: a package for discovery of novel differences among microbial communities, Molecular Ecology Resources 20 (2020) 415–428, https://doi.org/10.1111/1755-0998.13116.
[30] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.
[31] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140, https://doi.org/10.1093/bioinformatics/btp616.
[32] MedCalc Software Ltd, Odds ratio calculator, MedCalc (n.d.). https://www.medcalc.org/calc/odds_ratio.php (accessed December 5, 2022).
[33] K. Blighe, EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. https://github.com/kevinblighe/EnhancedVolcano, 2023. (Accessed 26 June 2023).
[34] R. Kolde, Pheatmap: Pretty Heatmaps, 2019. https://CRAN.R-project.org/package=pheatmap. (Accessed 6 July 2022).
[35] B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data, Nucleic Acids Res. 48 (2020) W449–W454, https://doi.org/10.1093/nar/gkaa379.
[36] E. Coudert, S. Gehant, E. de Castro, M. Pozzato, D. Baratin, T. Neto, C.J.A. Sigrist, N. Redaschi, A. Bridge, The UniProt consortium, annotation of biologically relevant ligands in UniProtKB using ChEBI, Bioinformatics 39 (2023) btac793, https://doi.org/10.1093/bioinformatics/btac793.
[37] K.H. Stürner, I. Siembab, G. Schön, J.-P. Stellmann, N. Heidari, B. Fehse, C. Heesen, T.H. Eiermann, R. Martin, T.M. Binder, Is multiple sclerosis progression associated with the HLA-DR15 haplotype? Mult Scler J Exp Transl Clin 5 (2019) 2055217319894615 https://doi.org/10.1177/2055217319894615.
[38] N.J. Thomas, J.M. Dennis, S.A. Sharp, A. Kaur, S. Misra, H.C. Walkey, D. G. Johnston, N.S. Oliver, W.A. Hagopian, M.N. Weedon, K.A. Patel, R.A. Oram, DR15-DQ6 remains dominantly protective against type 1 diabetes throughout the first five decades of life, Diabetologia 64 (2021) 2258–2265, https://doi.org/10.1007/s00125-021-05513-4.
[39] Q. Wang, G.M. Garrity, J.M. Tiedje, J.R. Cole, Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, Appl. Environ. Microbiol. 73 (2007) 5261–5267, https://doi.org/10.1128/AEM.00062-07.
[40] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30, https://doi.org/10.1093/nar/28.1.27.
[41] J.P. Krischer, K.F. Lynch, D.A. Schatz, J. Ilonen, Å. Lernmark, W.A. Hagopian, M. J. Rewers, J.-X. She, O.G. Simell, J. Toppari, A.-G. Ziegler, B. Akolkar, E. Bonifacio, TEDDY Study Group, the 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study, Diabetologia 58 (2015) 980–987, https://doi.org/10.1007/s00125-015-3514-y.
[42] M. Battaglia, S. Ahmed, M.S. Anderson, M.A. Atkinson, D. Becker, P.J. Bingley, E. Bosi, T.M. Brusko, L.A. DiMeglio, C. Evans-Molina, S.E. Gitelman, C. J. Greenbaum, P.A. Gottlieb, K.C. Herold, M.J. Hessner, M. Knip, L. Jacobsen, J. P. Krischer, S.A. Long, M. Lundgren, E.F. McKinney, N.G. Morgan, R.A. Oram,

T. Pastinen, M.C. Peters, A. Petrelli, X. Qian, M.J. Redondo, B.O. Roep, D. Schatz, D. Skibinski, M. Peakman, Introducing the endotype concept to address the challenge of disease heterogeneity in type 1 diabetes, Diabetes Care 43 (2020) 5–12, https://doi.org/10.2337/dc19-0880.

[43] D.L. Lauté-Caly, E.J. Raftis, P. Cowie, E. Hennessy, A. Holt, D.A. Panzica, C. Sparre, B. Minter, E. Stroobach, I.E. Mulder, The flagellin of candidate live biotherapeutic Enterococcus gallinarum MRx0518 is a potent immunostimulant, Sci. Rep. 9 (2019) 801, https://doi.org/10.1038/s41598-018-36926-8.

[44] T.C. Cullender, B. Chassaing, A. Janzon, K. Kumar, C. Muller, J.J. Werner, L. T. Angenent, M.E. Bell, A.G. Hay, D.A. Peterson, J. Walter, M. Vijay-Kumar, A. T. Gewirtz, R.E. Ley, Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut, Cell Host Microbe 14 (2013) 571–581, https://doi.org/10.1016/j.chom.2013.10.009.

[45] T. Kawai, S. Akira, Toll-like receptors and their crosstalk with other innate receptors in infection and immunity, Immunity 34 (2011) 637–650, https://doi.org/10.1016/j.immuni.2011.05.006.

[46] S. Tollefsen, H. Arentz-Hansen, B. Fleckenstein, Ø. Molberg, M. Ráki, W.W. Kwok, G. Jung, K.E.A. Lundin, L.M. Sollid, HLA-DQ2 and -DQ8 signatures of gluten T cell epitopes in celiac disease, J. Clin. Invest. 116 (2006) 2226–2236, https://doi.org/10.1172/JCI27620.

[47] A.K. Rai, A.M. Mitchell, Enterobacterial common antigen: synthesis and function of an enigmatic molecule, mBio 11 (2020) 01914, https://doi.org/10.1128/mBio.01914-20, 20.

[48] K.L. Aoyagi, B.D. Brooks, S.W. Bearden, J.A. Montenieri, K.L. Gage, M.A. Fisher, LPS modification promotes maintenance of Yersinia pestis in fleas, Microbiology (Read.) 161 (2015) 628–638, https://doi.org/10.1099/mic.0.000018.

[49] M.A. Jorgenson, S. Kannan, M.E. Laubacher, K.D. Young, Dead-end intermediates in the enterobacterial common antigen pathway induce morphological defects in Escherichia coli by competing for undecaprenyl phosphate, Mol. Microbiol. 100 (2016) 1–14, https://doi.org/10.1111/mmi.13284.

[50] M. Zeng, N. Inohara, G. Nuñez, Mechanisms of inflammation-driven bacterial dysbiosis in the gut, Mucosal Immunol. 10 (2017) 18–26, https://doi.org/10.1038/mi.2016.75.

[51] K. Yoshii, K. Hosomi, K. Sawane, J. Kunisawa, Metabolism of dietary and microbial vitamin B family in the regulation of host immunity, Front. Nutr. 6 (2019) 434971, https://doi.org/10.3389/fnut.2019.00048.

[52] S. Magnúsdóttir, D. Ravcheev, V. de Crécy-Lagard, I. Thiele, Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes, Front. Genet. 6 (2015). https://www.frontiersin.org/articles/10.3389/fgene.2015.00148. (Accessed 29 September 2023).

[53] N. Visalli, M.G. Cavallo, A. Signore, M.G. Baroni, R. Buzzetti, E. Fioriti, C. Mesturino, R. Fiori, L. Lucentini, M.C. Matteoli, A. Crinò, S. Corbi, S. Spera, C. Teodonio, F. Paci, R. Amoretti, L. Pisano, C. Suraci, G. Multari, N. Sulli, M. Cervoni, G. De Mattia, M.R.C. Faldetta, B. Boscherini, M.L.M. Bitti, G. Marietti, F. Ferrazzoli, C. Bizzarri, D. Pitocco, G. Ghirlanda, P. Pozzilli, A multi-centre randomized trial of two different doses of nicotinamide in patients with recent-onset Type 1 diabetes (the IMDIAB VI), Diabetes Metabol. Res. Rev. 15 (1999) 181–185, https://doi.org/10.1002/(SICI)1520-7560(199905/06)15:3<181::AID-DMRR31>3.0.CO;2-H.

[54] P. Chase, J. Dupre, J. Mahon, R. Ehrlich, E. Gale, H. Kolb, E. Lampeter, J. Nerup, Nicotinamide and prevention of diabetes, Lancet 339 (1992) 1051–1052, https://doi.org/10.1016/0140-6736(92)90570-s.

[55] E.a.M. Gale, P.J. Bingley, C.L. Emmett, T. Collier, European nicotinamide diabetes intervention trial (ENDIT) group, European nicotinamide diabetes intervention trial (ENDIT): a randomised controlled trial of intervention before the onset of type 1 diabetes, Lancet 363 (2004) 925–931, https://doi.org/10.1016/S0140-6736(04)15786-3.

[56] K.N. Thompson, K.S. Bonham, N.E. Ilott, G.J. Britton, P. Colmenero, S.J. Bullers, L. J. McIver, S. Ma, L.H. Nguyen, A. Filer, I. Brough, C. Pearson, C. Moussa, V. Kumar, L.H. Lam, M.A. Jackson, A. Pawluk, S. Kiriakidis, P.C. Taylor, L.R. Wedderburn, B. Marsden, S.P. Young, D.R. Littman, J.J. Faith, A.G. Pratt, P. Bowness, K. Raza, F. Powrie, C. Huttenhower, Alterations in the gut microbiome implicate key taxa and metabolic pathways across inflammatory arthritis phenotypes, Sci. Transl. Med. 15 (2023) eabn4722, https://doi.org/10.1126/scitranslmed.abn4722.