# Supplementary Methods

**Ovalbumin (OVA)-driven CD4$^+$ T cell cultures and intracellular cytokine staining**

Mesenteric lymph nodes cells from reconstituted and OVA-sensitized GF mice were labeled with the Violet CellTrace proliferative dye (Invitrogen; Grand Island, NY) and cultured with 200µg/ml OVA and 250pg/ml IL-2 for 72 hours. During the last 4 hours, cultured cells were stimulated with PdBU (500 ng/ml; Sigma-Aldrich, St. Louis MO) and Ionomycin (500 ng/ml; Sigma-Aldrich) in the presence of Brefeldin A (1µg/ml; BD Biosciences – San Jose, CA). Cells were stained with the following conjugated antibodies: CD3 (145-2C11), CD4 (RM4-5), IL-4 (11B11) and IFN-γ (XMG1.2) (eBioscience, San Diego, CA). Intracellular cytokines were detected in Violet CellTrace$^+$ proliferating CD3$^+$CD4$^+$ T cells by using Cytofix/Cytoperm (BD Biosciences) buffers, according to the manufacturer's instructions. Stained cells were analyzed on a LSRII Fortessa cytometer (BD Biosciences) and data processed using Flowjo (Tree Star; Ashland, OR).

**PhyloChip™ data analysis**

*Pre-processing and Data Reduction.* Fluorescent images were captured with the GeneChip Scanner 3000 7G (Affymetrix, Santa Clara, CA). An individual array feature occupied approximately 8x8 pixels in the image file corresponding to a single probe 25mer on the surface. To calculate the summary intensity for each feature on each array, the central 9 pixels of individual features were ranked by intensity and the 75% percentile was used. Probe intensities were background-subtracted and scaled to the PhyloChip™ Control Mix. Array fluorescence intensity was collected as integer values ranging from 0 to 65,536 ($2^{16}$). Fluorescence intensities for sets of probes complementing an operational taxonomic unit (OUT) were averaged after discarding the highest and lowest and the mean was $\log_2$ transformed into numbers ranging from 0 to 16. For compatibility with some statistical operations, the scores were multiplied by 1000 then rounded, allowing a range of integers from 0 to 16,000. These values are referred to as the hybridization score (HybScore). For the complete distribution see Hazen et al, Supplemental Information [1]. The data was reduced to consider the bacterial taxa deemed present as described in Hazen *et al*. [1]. Taxa were filtered to those present in the majority of samples of at least one of the experimental groups and rank-normalized such that taxa in each are represented by their ranked HybScore within that sample only (rank 1 represents the lowest HybScore in that sample).

*Sample-to-Sample Distance Function*. All profiles were inter-compared in a pair-wise fashion to determine a dissimilarity score and results were stored as a distance matrix. The Weighted Unifrac distance measure was chosen because it utilizes the phylogenetic distance between OTUs as well as the abundance of those OTUs to compute a community-wide dissimilarity between any pair of profiles [2, 3]. Similar biological samples produce small Weighted Unifrac dissimilarity scores. When comparing the presence or absence of taxa between profiles, the Unweighted Unifrac distance measure was utilized.

*Statistical Analysis, Ordination, Clustering, and Classification Methods.* The differences between the microbial communities (the entire number of OTUs detected in any one comparison group versus another) was determined by the Adonis test, which is a permutation test based on a dissimilarity matrix, in this case measured by weighted UniFrac. Because the Adonis test considers the multidimensional structure of the data (e.g., compares entire microbial communities), it does not involve multiple hypotheses testing for each microbial taxon found within those communities.

Taxa found increased in their ranked HybScore in one category compared to the alternate categories were identified using the Kruskal-Wallis (KW) test. The aim of a KW filter in the context of this analysis was to reduce the dimensionality of the dataset, and demonstrate that this reduced set of OTUs could still effectively discriminate between samples in terms of their microbial community structures by the ordination and clustering methods listed below.

Two-dimensional ordinations and hierarchical clustering maps of the samples in the form of dendrograms were created to graphically summarize the inter-sample relationships. To create dendrograms, the samples from the distance matrix are clustered hierarchically using the average-neighbor (HC-AN) method [4]. Non-Metric Multidimensional Scaling (NMDS) was employed to visualize relationships between samples by two-dimensional ordination plotting [5]. Ordination points are colored by highlighted groupings. Lists of significant taxa whose abundance characterizes each class is performed using Prediction Analysis for Microarrays (PAM), a classifier (supervised machine learning) based method that utilizes a nearest shrunken centroid method [6].

*Phylogenetic Tree Visualization*. Bacterial families with OTUs found by the KW test to be differentially abundant between two comparison groups (*e.g.* allergen sensitized WT versus *Il4raF709* mice) were identified, and the one OTU with the greatest difference between the two group means from each family was selected. For those families containing OTUs with both higher and lower abundance scores between the two comparison groups, two OTUs were selected. A phylogenetic tree was constructed using FastTree, which was built using one representative 16S ribosomal DNA (rDNA) gene sequence from each of the OTUs selected from the Greengenes multiple sequence alignment [7, 8]. The Tree was displayed with iTOL software [9].

## 16S rDNA sequencing methods and data analysis

*Summary of methodology*. The microbial community structure in each stool sample was assessed by 16S amplicon sequencing on the Roche 454 platform. Sequencing data was processed through a bioinformatics pipeline to obtain distributions of OTUs for each sample. We tested differences in overall microbial community structure between stool samples from different groups using the Dirichlet Multinomial model and a likelihood ratio test [10, 11]. We used hierarchical clustering with the Bray-Curtis (BC) dissimilarity measure to visualize the differences between the distributions of OTUs in samples [12]. The BC measure quantifies the difference between a pair of ecosystems based on the species or OTU composition of samples. A BC value of zero indicates identical OTU distributions; a BC value of one indicates no overlap in the OTUs present in the pair of samples. We used a bootstrapping procedure to estimate 95% confidence intervals on BC measures, and thus evaluate the reproducibility of sample clusterings. Results were visualized with a dendrogram constructed using the bootstrapped values. We also used the UniFrac measure with Principal Coordinates Analysis (PCoA) to visualize differences between microbial communities in samples; this measure takes into account phylogenetic relationships among sequences and does not require clustering sequences into OTUs [2, 3]. Individual OTUs that discriminate between different groups were determined using a Random Forests supervised machine learning approach [13, 14].

*16S rDNA Amplicon sequencing.* DNA pyrosequencing was performed by the Human Genome Sequencing Center at Baylor College of Medicine following protocols benchmarked for the Human Microbiome Project. The V3-V5 hypervariable regions of the 16S rRNA gene were amplified using primer 357F (5'-CCTACGGGAGGCAGCAG-3') modified with the addition of the 454 FLX-titanium adaptor "B" sequence (5'CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3') and primer 926R (5'-CCGTCAATTCMTTTRAGT-3') modified with the addition of unique 6-8 nucleotide barcode sequences and the 454 FLX-titanium adaptor "A" sequence (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3'). Barcode and adaptor sequences are found at http://www.hmpdacc.org/doc/HMP_MDG_454_16S_Protocol_V4_2_102109.pdf. PCR amplification was performed on 2 uL of DNA template in a total volume of 25 uL containing 1x AccuPrime Buffer II (Invitrogen Corp., Carlsbad, CA), 320 uM of each primer, and 0.03 U/uL AccuPrime High Fidelity *Taq* polymerase. Reactions were heated at 95°C for 2 min followed by 30 cycles of 95°C for 20 sec, 50°C for 30 sec, and 72°C for 5 min. The concentration of amplicons in each reaction was determined in triplicate using the PicoGreen fluorescent assay (Invitrogen Corp.) and amplicons were pooled before being sequenced via a multiplexed 454-FLX-titanium pyrosequencing run according to manufacturer's specifications.

*Bioinformatics for 16S data.* Sequences were pre-processed using custom scripts and the packages mothur, CloVR, and QIIME [15-17]. Filtering criteria were: no ambiguous bases, maximum homopolymer length of 8, 1 base difference allowed for barcode matches, and 2 base differences allowed for primer matches. Each sample had approximately 3000 reads after filtering. Sequences were trimmed based on a minimum average quality score of 35 over a window of length 50 nt, and clustered into OTUs with a similarity threshold of 95%.

*Testing for differences in OTU distributions between groups.* OTU relative abundances were assumed to follow the Dirichlet Multinomial (DM) distribution[1,2]. To test for differences in overall community structure between two groups, denoted A and B, we used a likelihood ratio test:

$$S = -2\ln\{\mathrm{P}(X_A, X_B \mid M_{A+B}) / [\mathrm{P}(X_A \mid M_A)\,\mathrm{P}(X_B \mid M_B)]\}$$

Here, $X_A$ and $X_B$ represent the set of vectors of OTU counts for groups A and B respectively. $M_{A+B}$ represents the DM model estimated from the combined groups, and $M_A$ and $M_B$ the corresponding DM models estimated from the separate groups. DM parameters were estimated using the Maximum Likelihood method. The $S$ statistic asymptotically follows a $\chi^2$ distribution with degrees of freedom equal to the number of OTUs in the samples.

*Clustering and visualizing samples.* Bootstrapping was performed to standardize the effects of differing numbers of sequencing reads between samples, and to obtain estimates of the variability of dissimilarity measures between samples. For each pair of samples $i$ and $j$, $m$ reads were drawn independently and with replacement, and the Bray-Curtis Dissimilarity measure[3] was calculated between the bootstrapped reads. We set $m$ equal to the median number of sequencing reads over all samples, and repeated the bootstrapping procedure on each pair of samples 10,000 times. The 95% confidence interval for each sample pair was then estimated from the empirical distribution of values. An average linkage dendrogram was constructed using the 95th centile values between nodes.

*Finding OTUs that discriminate between groups.* To find OTUs discriminating between groups, we used Random Forests[5] (RF) with a wrapper feature based method as implemented in the Boruta[6] package. Briefly, RF is an ensemble based classification method that uses multiple weak classifier decision trees. An importance measure is calculated for each feature (OTU) based on the loss of accuracy in classification. The statistical significance of the importance measure is determined using a permutation based method.

**16S rDNA Pyrosequencing Analysis: Results**

*Comparisons between sensitized and sham sensitized Il4raF709 mice.* We assessed the difference in overall microbial community structure among stool samples from *Il4raF709* homozygous mutant mice sensitized with OVA or sham sensitized with PBS. The distributions of OTUs differed significantly between the groups (Dirichlet Multinomial model, $p$-value = $< 10^{-20}$). BC dissimilarity dendrograms and UniFrac PCoA plots visualizing differences in overall microbial community structure between groups showed overall separation between the two groups, although two mutant PBS samples were close to outlying mutant OVA samples (**Figure E1A, B**). Several bacterial families and genera were found to discriminate between the groups using a supervised machine learning based method, including OTUs classifying to the genera Clostridium, Bacteroides, Alistipes and Streptococcus (**Table E3**).

*Comparisons between unsensitized WT versus Il4raF709 mutant mice***.** We assessed the difference in overall microbial community structure between stool samples from unsensitized WT and *Il4raF709* homozygous mutant littermate mice, and found that the distributions of OTUs differed significantly between the two groups (Dirichlet Multinomial model *P* value = $7x10^{-11}$). We visualized differences in overall microbial community structure between the two groups using a dendrogram with the BC measure (**Figure E4A**) and a UniFrac PCoA plot (**Figure E4B**). Consistent across both techniques, the samples from the *Il4raF709* homozygous mutant mice overall clustered together, although few WT samples clustered with outlying *Il4raF709* samples. These findings suggest that differences between the two groups were relatively subtle and not well-visualized using a dimensionality reduction method. To explore differences in individual OTUs, we used a supervised machine learning based method, and found differences in several OTUs, including those classifying to the genera Helicobacter, Clostridium, Lactobacillus and Odoribacter (data not shown).

*Comparisons between WT and Il4raF709 mutant sensitized mice*. We assessed the difference in overall microbial community structure among stool samples from *Il4raF709* homozygous mutant mice and WT controls sensitized with OVA. The distributions of OTUs differed significantly between each group (Dirichlet Multinomial model, *P* value = $< 10^{-20}$). BC dissimilarity dendrograms and UniFrac PCoA plots visualizing differences in overall microbial community structure between groups showed clear separation between the WT and mutant OVA groups (**Figure E5A, B**). Several bacterial families and genera were found to optimally discriminate between the groups using a supervised machine learning-based method, including Alistipes, Clostridium, Anaeroplasma, Lachnobacterium, and Bacteroides (**Table E9**).

*Assessing recolonization of WT GF mice by flora of OVA-sensitized WT versus mutant mice.* We assessed the difference in overall microbial community structure between stool samples from the two groups of recipient mice collected 8 weeks after colonization (at the end of the OVA sensitization period). The distributions of OTUs in stool samples from the group of mice receiving donor microbiota from allergen-sensitized mutant mice differed significantly from those of the group receiving donor microbiota from WT mice (Dirichlet Multinomial model[1,2] $P$ value < $10^{-20}$). We visualized differences between samples using a dendrogram with the BC measure (**Figure 8A**).

The samples from the group of mice that received donor microbiota from WT mice all clustered tightly, and clustered with the respective donor sample. The samples from the group of mice that received donor microbiota from allergen-sensitized *Il4raF709* mutant mice also clustered closely with one another, but were distinct from those of WT flora recipients. The donor sample from allergen-sensitized mutant mice essentially clustered separately, but was closer to its respective recipient samples in aggregate than it was to the other samples.

## Supplementary Figure Legends

**Figure E1.** The microbial signature and dysbiosis associated with the allergen sensitization of *Il4raF709* mice is reproduced by 16S rDNA pyrosequencing. **A.** Agglomerative clustering of fecal samples from OVA- and sham PBS sensitized *Il4raF709* mice based on 16S OTUs Samples were clustered based on bootstrapped BC dissimilarity values computed on the relative abundances of OTUs in each sample. BC = 0 indicates identical microbial communities; BC = 1 indicates communities with no overlapping OTUs. Heights of lines on the dendrogram indicate bootstrapped BC values

at the 95th percentile. **B.** Principal Coordinate Analysis (PCoA) plot with the UniFrac measure for stool samples from OVA and sham PBS sensitized *Il4raF709* mice. The unweighted UniFrac measure was used, which emphasizes qualitative phylogenetic differences between samples. *n*=5 for the *Il4raF709* OVA group versus 9 mice for the PBS sham sensitized *Il4raF709* group. Dirichlet Multinomial model, *p*-value = < $10^{-20}$.

**Figure E2.** $T_R$ cell-treatment resets the microbiota of allergen-sensitized *Il4raF709* mice into a new baseline distinct from that of sham sensitized and treated control *Il4raF709* mice. **A**. NMDS based on Weighted Unifrac distance between samples of PBS-sensitized mice (*n*=4) versus those of $T_R$-cell treated (*n*=5), OVA-sensitized mice, based on the 786 taxa whose abundance was significantly different between groups using the Kruskal-Wallis (KW) test. **B**. Hierarchical Clustering based on Weighted Unifrac distance between samples. **C**. Nearest shrunken centroid analysis of OTUs that best characterize the difference between allergen-sensitized versus tolerant groups. **D**. Representation of the abundance of the OTUs identified by the nearest shrunken centroid analysis using the PAM method.

**Figure E3**. The microbiota of sham-sensitized *Il4raF709* and WT mice are distinct. **A**. NMDS based on Weighted Unifrac distance between samples of PBS-sensitized *Il4raF709* mice (*n*=4) versus those of WT, PBS-sensitized mice (*n*=4), based on the 813 taxa whose abundance was significantly different between groups using the KW test. **B**. Hierarchical Clustering based on Weighted Unifrac distance between samples. **C**. Nearest shrunken centroid analysis of OTUs that best characterize the difference between allergen-sensitized versus tolerant groups. **D**. Representation of the abundance of the OTUs identified by the nearest shrunken centroid analysis using the PAM method.

**Figure E4**. *Il4raF709* mice conserve their specific microbiota signature when cohoused with WT littermates. **A.** Agglomerative clustering of stool samples from *Il4raF709* (*n*=12) and WT (*n*=5) mice based on 16S OTUs. Samples were clustered based on bootstrapped BC dissimilarity values computed on the relative abundances of OTUs in each sample. BC = 0 indicates identical microbial communities; BC = 1 indicates communities with no overlapping OTUs. Heights of lines on the dendrogram indicate bootstrapped BC values at the 95th percentile. **B.** PCoA plot with the UniFrac measure for stool samples from *Il4raF709* and WT littermate mice. The unweighted UniFrac measure was used, which emphasizes qualitative phylogenetic differences between samples. Dirichlet Multinomial model *P* value = 7x$10^{-11}$.

**Figure E5.** The microbiota signatures of allergen-sensitized *Il4raF709* and WT mice are distinct by 16S rDNA pyrosequencing. **A.** Agglomerative clustering of stool samples from ovalbumin sensitized *Il4raF709* mutant (n=5) and WT (n=4) mice based on 16S OTUs. Samples were clustered based on bootstrapped BC dissimilarity values computed on the relative abundances of OTUs in each sample. BC = 0 indicates identical microbial communities; BC = 1 indicates communities with no overlapping OTUs. Heights of lines on the dendrogram indicate bootstrapped BC values at the 95th percentile. **B.** PCoA plot with the UniFrac measure for stool samples from ovalbumin sensitized mutant and WT mice. The unweighted UniFrac measure was used, which emphasizes qualitative phylogenetic differences between samples. Dirichlet Multinomial model, *p*-value = $< 10^{-20}$.

**Figure E6.** The microbiota signatures of allergen- (OVA +OVA/SEB) sensitized *Il4raF709* and WT mice are distinct. **A**. NMDS based on Weighted Unifrac distance between samples of OVA/SEB-sensitized WT (*n*=6) versus OVA+OVA/SEB sensitized *Il4raF709* mice (*n*=9), based on the 430 taxa whose abundance was significantly different between groups using the KW test. **B**. Hierarchical Clustering based on Weighted Unifrac distance between samples. **C**. Nearest shrunken centroid analysis of OTUs that best characterize the difference between the groups. **D**. Representation of the abundance of the OTUs identified by the nearest shrunken centroid analysis using the PAM method. **E**. Venn diagram showing the abundance levels of different OTUs in relation to the sensitization state of WT and *Il4raF709* mice. The labels define the abundance states of sets of OTUs in relation to specific comparison groups, *e.g.* F709 OVA+OVA/SEB<F709 PBS identifies those OTUs that are less abundant in allergen sensitized (with OVA or with OVA/SEB) *Il4raF709* mice as compared to sham (PBS) sensitized mice. The number of OTUs thus identified is indicated in parentheses. Spheres indicate intersections between two sets, while the colored webs show which intersection of sets form the spheres. **F**. Contingency table representation of the results shown in the Venn diagram. P<0.0001 by the $X^2$ test (excluding the 2993 OTUs that did not change upon sensitization in both WT and *Il4raF709* mice).

**Table E1**. Annotations of Prediction Analysis for Microarrays (PAM)-selected bacterial taxa that discriminate between sham and allergen (OVA and OVA/SEB-sensitized) *Il4raF709* mice (see **Figure 3C, D**).

**Table E2.** Annotations of bacterial taxa showing significantly different abundances between sham and allergen-sensitized *Il4raF709* mice, as shown in the phylogenetic tree in **Figure 4**.

**Table E3.** Annotations of bacterial genera that optimally discriminate between sham and allergen-sensitized *Il4raF709* mice, as revealed by 16S rDNA pyrosequencing (**Figure E1)** and determined using the Random Forest machine learning method.

**Table E4**. Annotations of PAM-selected bacterial taxa that discriminate between allergen (OVA- and OVA/SEB)-sensitized versus $T_R$-cell treated and OVA-sensitized mice. The taxa selected correspond to those graphically presented in **Figure 5C, D**.

**Table E5.** Annotations of bacterial taxa showing significantly different abundances between sham and allergen-sensitized *Il4raF709* mice, as shown in the phylogenetic tree in **Figure 6**.

**Table E6**. Annotations of PAM-selected bacterial taxa that discriminate between $T_R$ cell-treated, allergen (OVA)- and sham-sensitized *Il4raF709* mice. The taxa selected correspond to those graphically presented in **Figure E2C, D**.

**Table E7**. Annotations of PAM-selected bacterial taxa that discriminate between sham- and allergen (OVA/SEB)-sensitized WT mice. The taxa selected correspond to those graphically presented in **Figure E3C, D**.

**Table E8**. Annotations of PAM-selected bacterial taxa that discriminate between allergen (OVA/SEB)-sensitized WT *and Il4raF709* mice. The taxa selected correspond to those graphically presented in **Figure 7C, D**.

**Table E9**. Annotations of bacterial genera that optimally discriminate between allergen-sensitized WT versus *Il4raF709* mice, as revealed by 16S rDNA pyrosequencing (**Figure E5)** and determined using the Random Forest machine learning method.

## Supplementary References

1. Hazen TC, Dubinsky EA, DeSantis TZ, Andersen GL, Piceno YM, Singh N, et al. Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science 2010; 330:204-8.
2. Lozupone C, Hamady M, Knight R. UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 2006; 7:371.
3. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005; 71:8228-35.
4. Legendre P, Legendre L. Numerical ecology. 2nd English ed. Amsterdam ; New York: Elsevier; 1998.
5. Shepard RN. Multidimensional scaling, tree-fitting, and clustering. Science 1980; 210:390-8.
6. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002; 99:6567-72.
7. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 2010; 5:e9490.

8. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006; 72:5069-72.

9. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 2007; 23:127-8.

10. Mosimann JE. On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. Biometrika 1962; 49:65-82.

11. Tvedebrink T. Overdispersion in allelic counts and theta-correction in forensic genetics. Theor Popul Biol 2010; 78:200-10.

12. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs 1957; 27:325-49.

13. Breiman L. Random Forests. Learning Machines 2001; 45:5-32.

14. Kursa MB, Rudnicki WR. Feature selection with the boruta package. j. Stat. Soft. 2010; 36:1-13.

15. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010; 7:335-6.

16. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009; 75:7537-41.

17. Angiuoli SV, Matalka M, Gussman A, Galens K, Vangala M, Riley DR, et al. CloVR: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing. BMC Bioinformatics 2011; 12:356.