SUPPLEMENTAL MATERIAL

METHODS:

Subjects

Subjects ages 1.1 to 19.7 years old undergoing upper endoscopy (EGD) as part of clinical care for diagnostic purposes or management of eosinophilic esophagitis (EoE) were enrolled in the study. All subjects had been on proton pump inhibitor (PPI) therapy for at least 4 weeks prior to endoscopy, had no history of antibiotic or systemic steroid use, and did not carry a diagnosis of any other gastrointestinal (GI) tract inflammatory disease. Informed consent, and when appropriate, assent was obtained from all subjects.

Because we began this study prior to the AGREE (A Working Group on PPI-REE) guidelines [1], EoE diagnosis was based on 2011 clinical guidelines of esophageal symptoms and ≥15 eosinophils per high power field (eos/hpf) after a PPI trial [2]. Subjects not meeting EoE criteria and those without any histological esophageal inflammation in the proximal GI tract after PPI therapy were considered non-EoE controls. The active EoE group was further stratified as active steroid-naïve (≥15 eos/hpf after PPI trial and/or dietary intervention; no topical swallowed steroid (TSS)) or steroid-non-responder (≥15 eos/hpf; currently on TSS). The inactive EoE group was divided into inactive steroid-naïve (<15 eos/hpf after dietary intervention) and steroid-responder (<15 eos/hpf currently on TSS) (Figure 1). A subgroup of 9 subjects who responded to TSS were studied longitudinally (Figure 2). Patients with endoscopic signs of esophageal candidiasis whose samples resulted in positive *Candida* cultures were excluded.

Esophageal biopsies were collected during endoscopy using sterile forceps, placed in dry ice and then stored at -80 degrees Celsius until processing, as previously described [3].

Sample processing and library preparation

DNA was extracted from esophageal biopsies using the MoBio PowerSoil® DNA Isolation Kit (Catalog # 12888-50). To profile bacterial communities, the V1-V2 variable regions of the bacterial 16S rRNA gene were amplified using barcoded primer sequences 27F (AGAGTTTGATCCTGGCTCAG) and 338R (TGCTGCCTCCCGTAGGAGT), as described previously. Purified PCR products were pooled in equal amounts and sequenced on the Illumina MiSeq, yielding 250bp paired-end sequence reads. Four negative control samples were included along with the experimental samples, as well as four synthetic positive control samples [4].

To profile fungal communities, the Internal Transcribed Spacer (ITS) region was amplified from the extracted DNA using barcoded primer sequences ITS1F (CTTGGTCATTTAGAGGAAGTAA) [5] and ITS2 (GCTGCGTTCTTCATCGATGC) [6]. Purified PCR products were quantified by the PicoGreen dsDNA quantitation assay (ThermoFisher, Waltham, MA USA). Paired-end sequencing was carried out on the Illumina MiSeq as for the 16S rRNA gene products. As a positive control, two samples from *Cryptococcus diffluens* culture stocks were included. A set of four negative control samples was also included on the ITS sequencing run, to enable identification and removal of contaminant sequences [7].

Bioinformatics analysis

Read pairs from 16S rRNA marker gene sequencing were processed using QIIME version 1.9 [8]. Reads were joined to form a complete amplicon sequence for the V1-V2 region, with a minimum overlap of 35 base pairs and a maximum overlap difference of 15%. Sequences were filtered to remove low quality reads, with a minimum quality threshold of Q20. Operational Taxonomic Units (OTUs) were generated at 97% sequence similarity using UCLUST v. 1.2.22 [9]. Taxonomic assignments were generated with the default method in QIIME, using the Greengenes reference database v. 13_8 [10]. The counts of OTUs that were assigned to the same taxon at the genus level or the most specific level when they could not be identified at a genus level were summed to obtain taxon level abundances. These counts were divided by the total number of assignments in a sample to calculate taxon level relative abundance. These were then used for differential abundance analysis. A phylogenetic tree was inferred from the OTU data using FastTree2 [11]. Similarity between samples was assessed by weighted and unweighted UniFrac distance [12, 13].

Data from fungal ITS amplicon sequencing were also analyzed using QIIME. Due to the variable length of the ITS region, only the forward reads were used. Sequence reads were quality filtered as above. To improve taxonomic resolution for fungal sequences, OTUs were generated at 100% similarity using UCLUST. The representative sequences of the OTUs were aligned to the NCBI nucleotide database using BLAST, and the consensus taxonomic assignments were generated using the BROCC software [14]. OTUs appearing in *Cryptococcus diffluens* positive control samples were removed from the analysis. OTUs classified in the *Chordata* phylum were also removed from the analysis to eliminate contamination from non-fungal sources. The remaining abundances were normalized using PicoGreen DNA quantification data as described previously [7].

To account for total fungal abundance and potential admixture of environmental sources, we multiplied the fungal OTU relative abundances by the post-PCR PicoGreen concentration for each sample, to obtain a PicoGreen-corrected OTU abundance [7].

Statistical analysis

Data files from QIIME were analyzed in the R environment. Unweighted and weighted UniFrac distances between samples were visualized using Principal Coordinate Analysis (PCoA), and differences between groups were assessed with the PERMANOVA test [15]. For taxa present in more than 80% of the samples, relative abundance was analyzed using generalized linear models. Fisher's exact test was used to assess if taxa were differentially present or absent between groups. Where multiple taxa were tested in the same analysis, the p-values were corrected for false discovery rate using the method of Benjamini and Hochberg [16].

In the longitudinal sub-study, we evaluated the microbial community shift across timepoints by using PERMANOVA on UniFrac distances. We calculated the p values by permuting the time points only within each subject. To identify the bacteria that changed across time points, we used generalized linear mixed effects models and added a random intercept term for each subject.

REFERENCES

1. Dellon, E.S., et al., Updated International Consensus Diagnostic Criteria for Eosinophilic Esophagitis: Proceedings of the AGREE Conference. Gastroenterology, 2018. **155**(4): p. 1022-1033 e10.
2. Liacouras, C.A., et al., Eosinophilic esophagitis: updated consensus recommendations for children and adults. J Allergy Clin Immunol, 2011. **128**(1): p. 3-20 e6; quiz 21-2.
3. Benitez, A.J., et al., Inflammation-associated microbiota in pediatric eosinophilic esophagitis. Microbiome, 2015. **3**: p. 23.
4. Kim, D., et al., Optimizing methods and dodging pitfalls in microbiome research. Microbiome, 2017. **5**(1): p. 52.
5. Gardes, M. and T.D. Bruns, ITS primers with enhanced specificity for basidiomycetes--application to the identification of mycorrhizae and rusts. Mol Ecol, 1993. **2**(2): p. 113-8.
6. White, T.J., T. Bruns, S. Lee & J. Taylor, Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, in PCR Protocols: A Guide to Methods and Applications, N. Innis, D. Gelfand, J. Sninsky & T. White, Eds., Editor. 1990, Academic Press. p. 315-322.
7. Bittinger, K., et al., Improved characterization of medically relevant fungi in the human respiratory tract using next-generation sequencing. Genome Biol, 2014. **15**(10): p. 487.
8. Caporaso, J.G., et al., QIIME allows analysis of high-throughput community sequencing data. Nat Methods, 2010. **7**(5): p. 335-6.
9. Edgar, R.C., Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 2010. **26**(19): p. 2460-1.
10. McDonald, D., et al., An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J, 2012. **6**(3): p. 610-8.
11. Price, M.N., P.S. Dehal, and A.P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One, 2010. **5**(3): p. e9490.
12. Lozupone, C. and R. Knight, UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol, 2005. **71**(12): p. 8228-35.

13. Lozupone, C.A., et al., Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol, 2007. **73**(5): p. 1576-85.
14. Dollive, S., et al., A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. Genome Biol, 2012. **13**(7): p. R60.
15. Anderson, M.J., A new method for non-parametric multivariate analysis of variance. Austral Ecology, 2008.
16. Benjamini, Y.H., Yosef, Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B, 1995. **57**(1): p. 289-300.

SUPPLEMENTAL TABLE AND FIGURES

Supplemental Table 1. Patients' characteristics

	Controls (n=10)	**EoE (n=69)**	
		Active (n=33)	Inactive (n=36)
Demographics			
Age in years – median (IQR)	8.05 (1.6-13.3)	12.10 (6.9-15.8)	9.5 (7.0-13.0)
Gender-n (%)			
Male	6 (60%)	25 (76%)	33 (92%)
Female	4, (40%)	8 (23%)	3 (8%)
Ethnicity - n (%)			
Caucasian	7 (70%)	19 (58%)	31 (86%)
African American	2 (20%)	8 (24%)	3 (8%)
Hispanic	1 (10%)	2 (6%)	1 (3%)
Other	1 (13%)	4 (12%)	1 (3%)
Indication - n (%)			
Abdominal pain	5 (50%)	1 (3%)	2 (6%)
Failure to thrive	2 (20%)	1 (3%)	0 (0%)
Dysphagia	0 (0%)	9 (27%)	1 (3%)
EoE surveillance	0 (0%)	16 (48%)	35 (97%)
Symptomatology - n (%)			
Heartburn	0 (0%)	1 (3%)	1 (3%)
Dysphagia	0 (0%)	12 (36%)	5 (14%)
Regurgitation	0 (0%)	4 (12%)	4 (11%)
Abdominal pain	4 (40%)	3 (9%)	6 (17%)
Food impaction	0 (0%)	9 (27%)	6 (17%)
Allergy history - n (%)			
Asthma	2 (20%)	19 (58%)	25 (69%)
Eczema	0 (0%)	17 (52%)	15 (42%)
Allergic Rhinitis	2 (20%)	11 (33%)	21 (58%)
Medications - n (%)			
PPI	10 (100%)	33 (100%)	36 (100%)
Topical swallowed steroids	0 (0%)	16 (48%)	18 (50%)
Inhaled steroids	0 (0%)	7 (21%)	17 (47%)
Nasal steroids	2 (20%)	9 (27%)	19 (53%)
EGD Findings - n (%)			
Furrows	0 (0%)	19 (58%)	8 (22%)
Exudates	0 (0%)	6 (18%)	2 (6%)
Rings	0 (0%)	3 (9%)	0 (0%)

Histologic Findings			
Eos/hpf - median (IQR)	0	40 (30-50)	0 (0-1)
EoE - eosinophilic esophagitis, IQR - Interquartile Range, Eos/hpf - eosinophils per high power field, PPI - proton pump inhibitor, EGD – esophagogastroduodenoscopy			

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure 1: Schematic of study design. Proton pump inhibitors (PPI). Eosinophils per high power field (eos/hpf). Topical swallowed steroids (TSS).

Supplemental Figure 2: Total PicoGreen-corrected fungal abundance for subjects not on topical swallowed steroid (TSS) therapy (noTSS) and subjects on topical swallowed steroid (TSS) therapy and its association with the number of eosinophils. Figure includes both subjects with active and inactive eosinophilic esophagitis (EoE).