Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome

Liying Yang,[1, 5] Xiaohua Lu,[2, 5] Carlos W. Nossa,[2] Fritz Francois,[2, 3] Richard M. Peek,[4] Zhiheng Pei[1,2, 3]

[1]*Department of Pathology, New York University School of Medicine, New York, NY 10016.*
[2]*Department of Medicine, New York University School of Medicine, New York, NY 10016.*
[3]*Department of Veterans Affairs New York Harbor Healthcare System, New York, NY 10010.*
[4]*Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN 37212.*
[5]*These authors contributed equally to this work.*

SUPPLEMENTARY APPENDIX

This appendix has been provided by the authors to give readers additional information about their work

ACKNOWLEDGEMENTS

We thank the staff of the Division of Gastroenterology, Department of Medicine, Veterans Affairs New York Harbor Healthcare System for patient recruitment and endoscopic biopsies, Dr. Qinghu Ren and Mr. Meisheng Zhou for their assistance in data analyses, Jordan Rupprecht for his assistance in mathematic calculation, Dr. Andrew Martin, Dr. Benli Chai, and Mr. Todd DeSantis for their advice for selection of analytic methods.

ACCESSION NUMBERS

The 16S rRNA gene sequences have been deposited in Genbank at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/) with accession numbers DQ537536-DQ537935 and DQ632752-DQ639751.

SUPPLEMENTARY METHODS

Subjects and histological phenotyping of esophageal biopsies

Patients were recruited from the Department of Veterans Affairs Medical Centers in New York, NY and in Nashville, TN during the years 2003-2006, as previously described.[1, 2] The use of the subjects in this study was approved by the Institutional Review Boards at New York University and Vanderbilt University. Esophagogastroduodenoscopy was performed and all biopsies were taken before the biopsy forceps entered further into the stomach, which differs from the retrograde method used in previous studies.[1, 2] The biopsy location was about 2 cm above the squamocolumnar junction. Each biopsy was examined microscopically, and patients were classified into one of three phenotypes: normal, esophagitis, and Barrett's esophagus (BE) groups based on these histological findings. Patients with no inflammatory infiltrate or only less than 10 lymphocytes per high power field (HPF, 400x magnification) in the squamous epithelium were assigned to the normal group; those with lymphocytes equal or more than 10 cells/HPF or with any numbers of eosinophils or polymorphonuclear leukocytes were assigned to the esophagitis group. Patients whose squamous epithelium was replaced by intestinal type epithelial cells and goblet cells were included as BE.

Sequence quality control

We sequenced more than two plates for each samples. We performed a complex control on the quality of sequences because we were concerned by unexpected high number of novel species reported by many other studies using cultivation-independent techniques. First we manually inspected all sequence reads and eliminated all sequences that failed. We then removed chimeric sequences by using the methods referred to in our previous paper published in PNAS.[1] Both the failed and chimeric sequences accounted for about 30% of the sequences. We performed taxonomic analysis on the remaining sequences. Any sequences unclassifiable at the species level were sequenced on another strand to obtain full length 16S rDNA sequence sequences were reconfirmed at the overlap region. This way, we avoided over-reporting novel species due to poor sequence quality. To be comparable among samples, we included first 200 usable sequences in our analyses. This gave us a final sequence number of 6800. By strictly excluding antilogous and artificial sequences, 308 of 6800 sequences in our dataset were unclassifiable at the species level, but classifiable at the genus level.

Classification of samples of esophageal microbiome

To classify samples of the esophageal microbiome, the 6,800 sequences (200 per sample) from the distal esophageal microbiome were aligned using the NAST aligner,[3] and the alignment was manually curated. The Lane mask[4] was used to restrict calculations to 1,287 conserved columns of aligned characters. Genetic distance between each pair of sequences was calculated from the aligned sequences using the DNAML option of DNADIST in the PHYLIP package[5] hosted at http://greengenes.lbl.gov/cgi-bin/nph-distance_matrix.cgi. Genetic distance between two samples was calculated as the mean of distances between all pairs of sequences from the two samples, by the following equation:

$$D_{i,j} = \frac{1}{n}\sum_{\substack{i=1 \\ j=1}}^{n} d_{i,j} \qquad (1)$$

where $D_{i,j}$ is the average distance between two samples *i* and *j*, $d_{i,j}$ is the distance between a sequence in sample *i* and a sequence in sample *j*, and n is the number of sequence pairs between the two samples. Distance for a sample from itself was defined as zero. Samples were classified by unsupervised hierarchical clustering analysis using the distance matrix calculated for all possible pairs of samples. The dendrogram was constructed using the average linkage algorithm and cosine measure of the distance matrix, by SPSS 13.0 (SPSS Inc., Chicago, IL). Cluster analysis with cosine distance allows better visualization of the between cluster difference because it minimize the within-group differences as compared with other distances.

Defining the normal reference range for esophageal microbiome

The values of between-sample distance calculated by Equation (1) were used in this analysis. To establish a normal reference range (NRR), the mean distance, $\overline{X}_{i,j}$, between normal sample *i* and each of *n* normal samples were calculated by:

$$\overline{X}_{i,j} = \frac{1}{n}\sum_{j=1}^{n} x_{i,j} \qquad (2)$$

Where $x_{i,j}$ is the distance between normal sample *i* and normal sample *j*. A sample was considered an outlier if its mean distance from the *n* normal samples was > 1.5 x IQR (interquartile range)

above or below the third or first quartiles,[6] respectively. The 95% NRR was determined after exclusion of outliers by the following equations:

$$NRR = \overline{X} \pm 1.96 s.d. \quad (3)$$

Where $\overline{X}$ is the average of the mean genetic distance for each normal sample calculated by:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} \overline{X}_{i,j} \quad (4)$$

Where $\overline{X}_{i,j}$ is the mean genetic distance for normal sample *i* as calculated by Equation (2); and *s.d.* is the sample standard deviation. The mean genetic distance for sample *k* in a disease group and the *n* normal samples (excluding any outliers), $\overline{Y}_{k,j}$, was calculated by:

$$\overline{Y}_{k,j} = \frac{1}{n}\sum_{j=1}^{n} y_{k,j} \quad (5)$$

Where $y_{k,j}$ is the genetic distance between sample *k* and normal sample *i*. The assignment of a sample to a normal or abnormal microbiome types was determined by whether its distance to the normal samples fell inside or outside the NRR.

Double principal coordinate analysis

Double principal coordinate analysis (DPCoA)[7] within the R statistical package (www.rproject.org) was used to reduce the complexity of the overall distances between samples to a principal coordinate system that best preserved the integrity of the genetic distance-based microbiome typing scheme. Distances between each pair of sequences and between each pair of samples were calculated by Rao diversity analysis within the R statistical package. The first two principal coordinates were obtained for each sample based on the between-sample distance and plotted to show the distribution of samples in a two-dimensional space, independent of their phenotypes. The assignment of a sample into a microbiome type was based on its relative location on the plot and that of related samples on the first or second principal coordinate (PC1 or PC2), or a combination of both. The border between the two microbiome types was determined by the rule of maximal separation of samples between the two types. To validate, the types assigned by the reduced genetic data by DPCoA was compared with the two microbiome types.

UniFrac test and F_{ST} analysis

UniFrac test was performed to compare difference between microbiome types, among and between phenotypes as well as among samples.[8] Unifrac was performed using online software, UniFrac (http://bmf.colorado.edu/unifrac), with 1000 permuations. For data entry, all 6,800 sequences were used to construct rooted phylogenetic trees using the unweighted pair group method with arithmetic mean (UPGMA) algorithm.[9] F_{ST} analysis was performed on the abundance-weighted the datasets using Arlequin 3.1[10] to examine whether there is a population-wise difference between the two types of microbiome.[11] The Jukes-Cantor algorithm was used to calculate the distances between sequences.

Lineage through time curve analysis

Lineage through time curves were constructed using the number of taxa at various identity levels (ID) calculated by DOTUR.[12] This analysis measures changes in the richness of taxa over the entire taxonomic hierarchy of domain bacteria. The rates of change and their turning points were analyzed by linear regression.

Correlation between microbiome types and relative abundance of taxonomic groups

Microbiome-abundance correlation (MAC) analysis is a type of linear regression analysis that we have designed to further link a microbiome type to a specific group of bacteria. This test determines whether the relative abundance of a group of bacteria significantly correlates with the criteria used in establishing the microbiome types, such as the first principal coordinates in the DPCoA-based reduced typing scheme, by the following equation:

$$y = ax + b \qquad (6)$$

where x is the principal coordinate of a sample calculated by DPCoA and y is the relative abundance of a specific bacterial group in the sample. For a bacterial group that was significantly correlated with the principal coordinate, we calculated the abundance-based 95% NRR for the bacterial group using Equation 3, where $\overline{Y}$ is the mean of relative abundances of the bacterial group in normal samples (excluding outliers). Through the use of NRR, each of the 34 samples was assigned to a normal or abnormal taxonomic type. For validation, the resulting taxonomy-based assignment was compared with the microbiome type assignment based on phylogenetic distance. Microbiome types also were compared using LIBRARY COMPARE at RDP II (release 9.39). Sequences belonging to the same microbiome type were pooled. Each sequence was assigned to a taxonomic unit at the genus rank (or lowest classifiable taxonomic rank above genus) by the RDP naïve Bayesian classifier using an 80% confidence threshold. The probability of the observed difference between the abundance of a given taxon was calculated for each microbiome type using the "digital Northern analysis",[13,14] with $P < 0.05$ to define significance.

AUTHOR CONTRIBUTIONS

L.Y. and X.L. performed all experiments and initial data analyses. C.N. contributed to data analysis. F.F. and R.M.P. contributed to subject recruitment and performed endoscopic biopsy. Z.P. designed the study, performed histological examination, supervised experimentation, and coordinated the project. All authors were involved in manuscript preparation.

SUPPLEMENTARY REFERENCES

1. Pei Z, Bini EJ, Yang L, Zhou M, Francois F, Blaser MJ. Bacterial biota in the human distal esophagus. Proc Natl Acad Sci U S A. 2004;101:4250-5.
2. Pei Z, Yang L, Peek RM, Jr Levine SM, Pride DT, Blaser MJ. Bacterial biota in reflux esophagitis and Barrett's esophagus. World J Gastroenterol. 2005;11:7277-83.
3. DeSantis TZ, Jr., Hugenholtz P, Keller K, et al. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res. 2006;34:W394-9
4. Lane DJ. 16S/23S rRNA sequencing. Nucleic Acid Techniques in Bacterial Systematics. New York, New York , USA: John Wiley and Sons, Inc, 1991:115-175.

5. Felsenstein J. PHYLIP -- Phylogeny Inference Package (Version 3.2). Cladistics. 1989;5:164-6.
6. Tukey JW. Exploratory Data Analysis. Reading, MA: Addison-Wesley, 1977.
7. Pavoine S, Dufour AB, Chessel D. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. J Theor Biol. 2004;228:523-537.
8. Lozupone C, Hamady M, Knight R. UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics. 2006;7:371.
9. Michener CD, Sokal RR. A quantitative approach to a problem in classification. Evolution, 1957;11:130-162.
10. Excoffier LG, Laval G, Schneider S. Arlequin, Version 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 2005;1:47-50.
11. Martin AP. Phylogenetic approaches for describing and comparing the diversity of microbial communities. Appl Environ Microbiol 2002;68(8):3673-82.
12. Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol. 2005;71:1501-6.
13. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261-5267.
14. Audic S, Claverie JM. The significance of digital gene expression profiles. Genome Res. 1997;7:986-995.
15. Lau WF, Wong J, Lam KH, Ong GB. Oesophageal microbial flora in carcinoma of the oesophagus. Aust N Z J Surg. 1981;51:52-5.
16. Finlay IG, Wright PA, Menzies T, McArdle CS. Microbial flora in carcinoma of oesophagus. Thorax. 1982;37:181-4.
17. Mannell A, Plant M. The management of malignant oesophago-airway fistulae. Aust N Z J Surg. 1983;53:31-6.
18. Gagliardi D, Makihara S, Corsi PR, et al. Microbial flora of the normal esophagus. Dis Esophagus. 1998;11:248-50.
19. Pajecki D, Zilberstein B, dos Santos MA, et al. Megaesophagus microbiome: a qualitative and quantitative analysis. J Gastrointest Surg. 2002;6:723-9.
20. Macfarlane S, Furrie E, Macfarlane GT, Dillon JF. Microbial Colonization of the Upper Gastrointestinal Tract in Patients with Barrett's Esophagus. Clinical Infectious Diseases. 2007;45:29-38.
21. Narikiyo M, Tanabe C, Yamada Y, et al. Frequent and preferential infection of Treponema denticola, Streptococcus mitis, and Streptococcus anginosus in esophageal cancers. Cancer Sci. 2004;95:569-74.
22. Armstrong D, Bennett JR, Blum AL, et al. The endoscopic assessment of esophagitis: A progress report on observer agreement Gastroenterology. 1996;111:85-92.
23. Locke GRr, Talley NJ, Fett SL, Zinsmeister AR, 3rd LJM. Risk factors associated with symptoms of gastroesophageal reflux. The American Journal of Medicine. 1999;106:642-9.
24. Mohammed I, Cherkas LF, Riley SA, Spector TD, Trudgill NJ. Genetic influences in irritable bowel syndrome: a twin study. The American Journal of Gastroenterology. 2005;100:1340-4.
25. Kotzan J, Wade W, Yu HH. Assessing NSAID Prescription Use as a Predisposing Factor for Gastroesophageal Reflux Disease in a Medicaid Population. Pharmaceutical Research. 2001;18:1367-72.
26. Conio M, Filiberti R, Blanchi S, et al. Risk factors for Barrett's esophagus: a case-control study. International Journal of Cancer. 2001;97:225-9.
27. Voutilainen M, Sipponen P, Mecklin J-P, Juhola M, Färkkilä M. Gastroesophageal Reflux Disease: Prevalence, Clinical, Endoscopic and Histopathological Findings in 1,128 Consecutive

Patients Referred for Endoscopy due to Dyspeptic and Reflux Symptoms. Digestion. 2000;61:6-13
28. Ruigomez A, Rodriguez LAG, Wallander MA, Johansson S, Graffner H, Dent J. . Natural history of gastro-oesophageal reflux disease diagnosed in general practice. Alimentary Pharmacology & Therapeutics. 2004;20:751-60.
29. Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucl Acids Res. 2007;35:D169-172
30. Chao A. Nonparametric estimation of the number of classes in a population. Scand J Statist. 1984;11:265-70.

Supplementary Table 1. Summary of culture-based and non-culture-based studies on microbiome of the esophagus									
Category	Culture-based 1981-2007						Non-culture-based 2004-2007		
Study	Lau 1981	Finlay 1982	Mannell 1983	Gagliardi 1998	Pajecki 2002	Macfarlane 2007	Narikiyo 2004	Pei 2004	This study 2007
Disease	Cancer	Cancer	Normal Cancer	Normal	Chagas' disease	Normal Barrett's	Cancer	Normal	Normal Esophagitis Barrett's
Specimen	Aspirate	Resection	Aspirate	Aspirate	Aspirate	Biopsy Aspirate	Biopsy	Biopsy	Biopsy
Method	Culture	Culture	Culture	Culture	Culture	Culture	PCR	PCR	PCR
No. cases	79	12	101	30	15	14	20	4	34
No. isolates/clones	61	85	377	30	ND	ND	100	900	6800
No. species	14	15	32	11	ND	46	7	95	166
Mean species/case	1	6	4	1	ND	ND	≤6	43	25
% cases positive for bacteria	64	100	100	67	93	71	ND	100	100
% cases positive for *Streptococcus*	10	92	ND	ND	93	50	87	100	100
% cases positive for *Bacteroides*	39	92	ND	ND	0	ND	ND	100	97
Reference	15	16	17	18	19	20	21	1	This study

ND: not described.

Supplementary Table 2. Patient information

Group	Case#	Age	Sex	Race[a]	Indication for endoscopy	Endoscopic findings	Histological phenotype	Biota type
Normal	E10	51	M	U	Fecal occult blood	HH[b]	Normal	I
	E11	80	M	W	Fecal occult blood	Normal	Normal	I
	E15[c]	75	M	W	Iron deficiency	HH	Normal	II
	E18	54	M	B	Heartburn	Normal	Normal	I
	E30	83	F	B	Fecal occult blood	Normal	Normal	I
	E37	38	M	W	BE follow-up[d]	Normal	Normal	I
	E40	44	F	B	Heartburn	Normal	Normal	I
	E42	65	M	B	Heartburn	HH	Normal	I
	E47	62	M	W	Heartburn	Normal	Normal	I
	E49*	82	M	W	Fecal occult blood	Normal	Normal	I
	E62*	57	M	B	Nausea	LA Grade A esophagitis[e]	Normal	I
	E67	59	M	W	Dysphasia	BE and tumour	Normal[f]	I
Esophagitis	E12	54	M	U	Iron deficiency	HH	Lc, Eo	II
	E19	70	M	B	Heartburn	HH	Lc	II
	E22	68	M	W	Heartburn	Normal	Lc	II
	E38	76	M	W	Heartburn	Schatzki ring, HH	Lc, Eo	I
	E39	75	M	B	Hoarseness	Normal	Lc	I
	E41	63	M	B	BE follow-up	HH	Lc	I
	E44	61	M	W	Chronic chocking	Normal	Lc	II
	E45*	77	M	W	Chest pain	Normal	Lc	I
	E51	54	M	U	Colon polyps	Normal	Lc	II
	E53*	75	M	B	Epigastric pain	Normal	Lc, Eo, PMN[g]	II
	E64*	66	M	U	Fecal occult blood	Normal	Lc, Eo	I
	E66	72	M	W	Fecal occult blood	Normal	Lc, Eo	II
BE	E25	57	M	W	Heartburn	BE, tumour, HH	BE[f]	II
	E27	66	M	B	BE follow-up	BE, HH	BE	II
	E50	70	M	W	Heartburn	BE, HH	BE	I
	E71	57	M	U	Fecal occult blood	BE, HH	BE	II
	E76	77	M	W	BE follow-up	BE	BE	I
	E78	61	M	W	BE follow-up	LA grade B esophagitis, HH	BE	II
	B288	73	M	W	BE follow-up	BE, HH	BE	I
	B322	92	M	W	BE follow-up	BE	BE	II
	B330	59	M	W	Heartburn	BE, HH	BE	I
	B350	61	M	W	BE follow-up	BE, HH	BE	II

[a] B: black; W: white; U: unknown.

[b] HH: hiatus hernia.

[c] Outlier, not included in analyses of normal reference ranges, P test, F_{ST} test, and estimation of species richness.

[d] BE: Barrett's esophagus.

[e] Los Angeles classification of esophagitis.[21]

[f] These biopsies were taken from mucosa near a tumour.

[g] Lc: lymphocytes, Eo: eosinophils, PMN: polymorphonuclear leukocytes.

* Gastric biopsies from these cases are positive for *Helicobacter pylori* on histological examination.

Supplementary Table 3. Summary of hypothesis testing*				
Groups compared			UniFrac	
			Unweighted	Weighted
All 34 samples			0.001	N/A
Normal	Esophagitis	Barrett's	0.001	N/A
Normal	Esophagitis		0.003	0.003
Normal	Barrett's		0.003	0.003
Esophagitis	Barrett's		0.003	0.003
Type I	Type II		0.001	0.001

**P* values corrected by Bonferroni method.

Supplementary Table 4. Comparisons of histological phenotypes in relative abundance of *Streptococcus*						
Omnibus test[a]						
Groups compared	Phenotype					
	Normal[b] (n=11)		Esophagitis (n=12)		BE (n=10)	
P value	0.043					
Follow-up tests[c]						
Groups compared	Phenotype					
	Normal	Esophagitis	Normal	BE	Esophagitis	BE
P value	0.016*		0.029*		0.773	

[a] The Omnibus test was performed using one-way ANOVA.

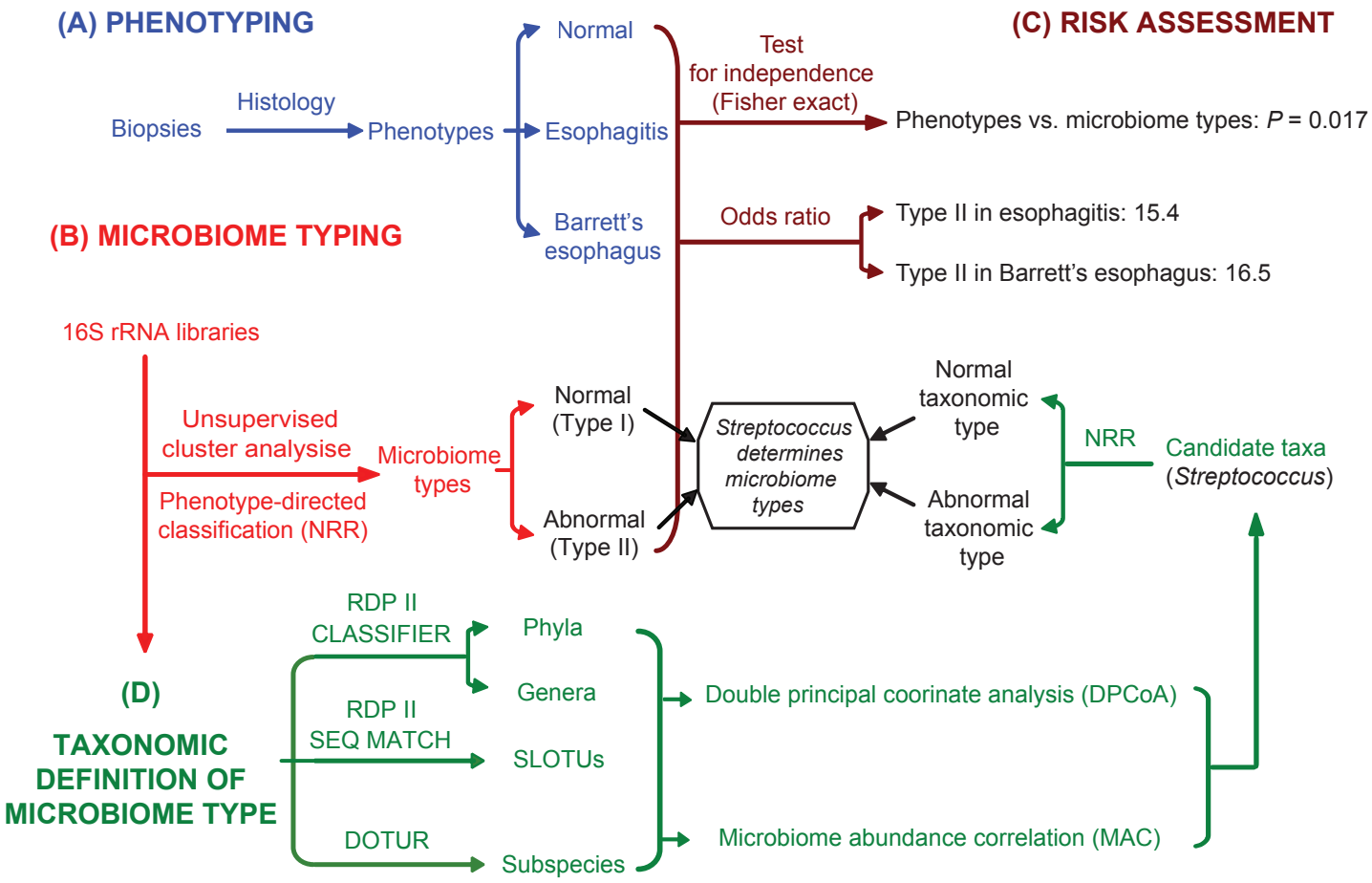
[b] An outlier (E15) in the normal group was not included in comparisons between esophageal phenotypic groups.

[c] The follow-up tests were performed with two-tailed independent *t*-test. Tests that are statistically different at the false discovery rate < 5% are marked by *.

Supplementary Table 5. Comparison of type II microbiome with known risk factors in gastroesophageal reflux disorders

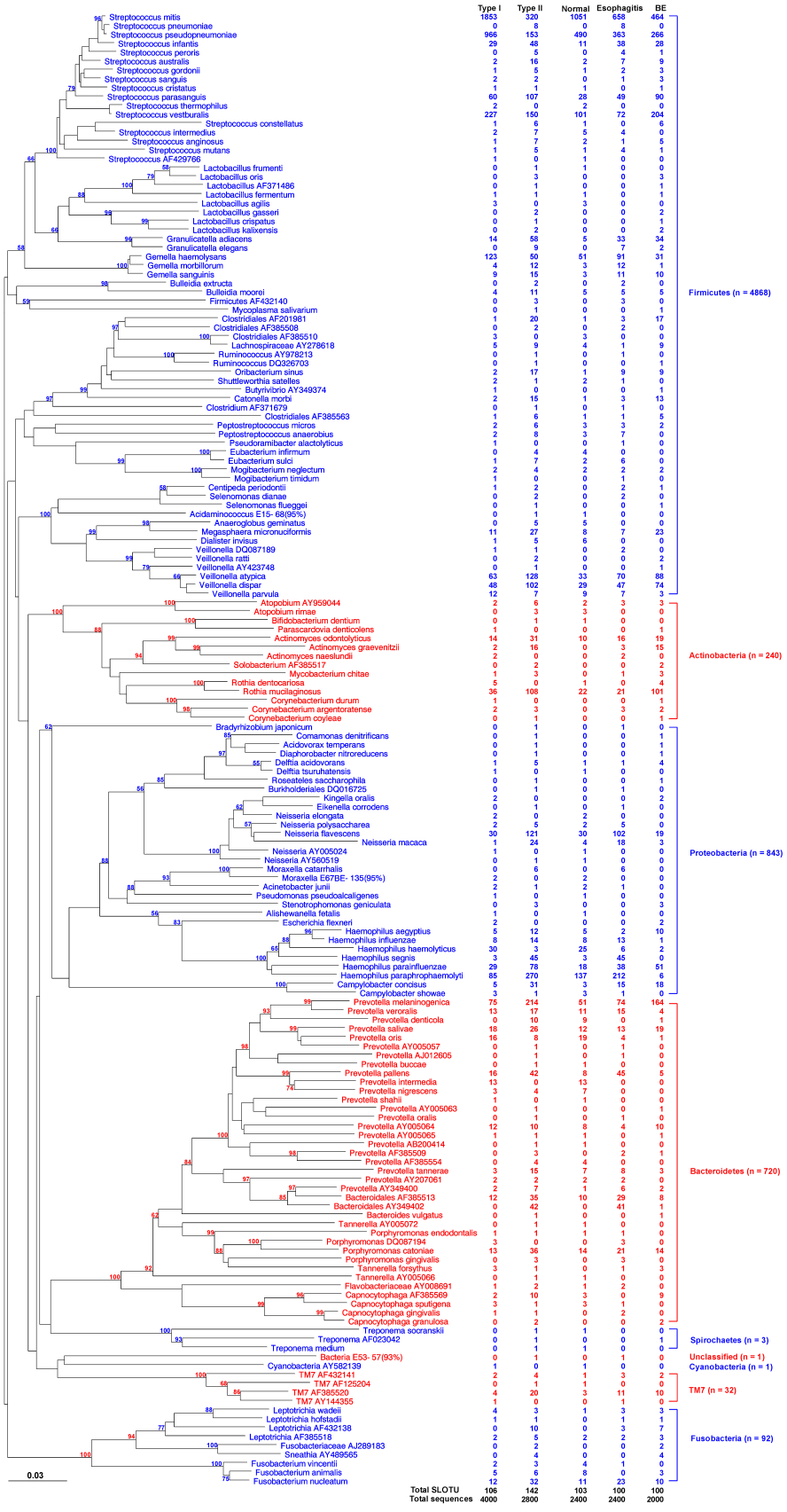
Category	Subcategory	Predictive factor	Predicted outcome (defining method)	Sample size	Odds ratio	95% C.I.	Reference
Host	Genetic	Immediate relatives	Heartburn (questionnaire)	1,524	2.6	1.8-3.7	23
		Parental family history		3,920	1.5	1.2-1.7	24
	Aging	Increasing age	GERD (ICD-9 code)	163,085	1.1	1.0-1.1	25
	Structural	Hiatus hernia			4.2	2.8-6.3	
			BE (histology)	457	3.9	2.5-6.0	26
			Esophagitis (endoscopy)	451	2.4	1.5-4.0	
		Papillae elongation	Heartburn (medical record)	1,128	2.2	1.5-3.2	27
	Symptomatic	Heartburn/regurgitation	Esophagitis (endoscopy)	451	9.4	6.1-14.4	28
			BE (histology)	457	5.8	4.0-8.4	
	Comorbid	Gall bladder disease	GERD (ICD-8 codes)	7,451	3.7	2.1-6.7	28
		Asthma	GERD (ICD-9 code)	163,085	3.2	2.6-4.0	25
		Angina	GERD (ICD-8 codes)	7,451	3.2	2.1-4.9	28
		Obesity	GERD (ICD-9 code)	163,085	2.8	2.1-3.6	25
		Peptic ulcer disease	GERD (ICD-8 codes)	7,451	2.5	1.7-3.6	28
		Chest pain			2.3	1.8-2.8	
		Cough			1.7	1.4-2.1	
		Irritable bowel syndrome			1.6	1.2-2.1	
Environment	Behavioral	Tobacco	GERD (ICD-9 code)	163,085	2.6	1.9-3.5	25
		Alcohol			1.8	1.4-2.4	
	Medical	NSAID			1.8	1.6-2.1	
		Anticholinergic drug	Heartburn (questionnaire)	3,920	1.5	1.1-2.1	24
		Nitrates	GERD (ICD-8 codes)	7,451	1.5	1.1-2.0	28
		Oral steroids			1.3	1.1-1.6	
	Bacterial	Type II microbiome	Esophagitis (histology)	24	15.4	1.5-161.0	This study
			BE (histology)	22	16.5	1.5-183.1	

NSAID: Nonsteroidal anti-inflammatory drug.
GERD: gastroesophageal reflux disease.
BE: Barrett's esophagus.
ICD: international classification of diseases.
ICD-8 codes for GERD: gastroesophageal reflux, esophagitis, esophageal inflammation, or heartburn.
ICD-9 code for GERD: not specified in detail.



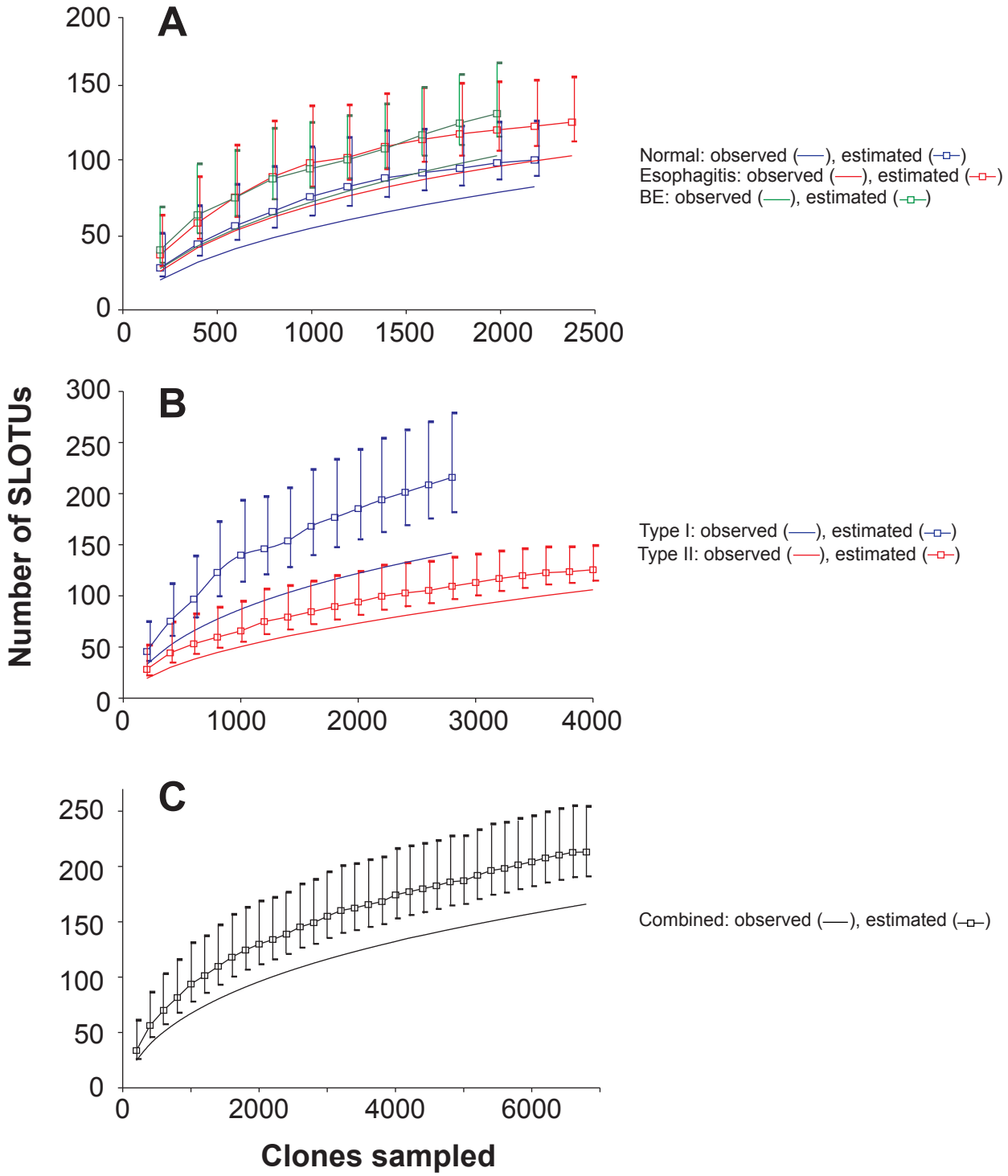
Supplemental Figure 1

Schematic approach to the classifications and risk assessment of microbiome in the distal esophagus. (A) Each biopsy was assigned to a phenotype as normal, esophagitis, or Barrett's esophagus based on histological examination (Supplemental Table 2). (B) Each biopsy also was assigned to a microbiome type based on the collective uniqueness of bacterial 16S rRNA gene sequences recovered from the biopsy. This was accomplished by unsupervised cluster analysis followed by phenotype directed classification using the 95% normal reference range (NRR) calculated by genetic distances between phenotypically normal samples (Figure 1A and 1B). Samples were classified as either normal (type I) or abnormal microbiome (type II). (C) The risk of association of a microbiome type with an abnormal phenotype was assessed by Fisher exact test and calculation of odds ratio (Table 1). (D) To define the microbiome types in easily comprehensible, taxonomic terms, the 6,800 bacterial 16S rRNA gene sequences cloned from the biopsies were binned into taxonomic groups at the phylum and genus levels using the CLASSIFIER at Ribosomal Database Project II (RDP II) (29) (Figure 3) and at the species level (SLOTU: species level operational taxonomic unit) by SEQUENCE MATCH (RDP II) (29) (Supplemental Figure 2). To test whether the complex difference between the two types of microbiomes could be caused by a principal subpopulation, double principal coordinate analysis (DPCoA)(7) was performed (Figure 1C). Microbiome-abundance correlation (MAC) analyses were performed to scan for any taxonomic groups whose relative abundance in the 34 samples significantly correlated with the criteria used in the microbiome typing (Figure 4A). For a bacterial group that was significantly correlated with the typing criteria, the abundance based 95% NRR for the bacterial group was calculated (Figure 4B). Through the use of NRR, each of the 34 samples was assigned to a normal or abnormal taxonomic type. For validation, the resulting taxonomy-based assignment was compared with the assigned microbiome types to determine whether the microbiome types can be translated into the taxonomic types. Main findings are shown by text in black font.



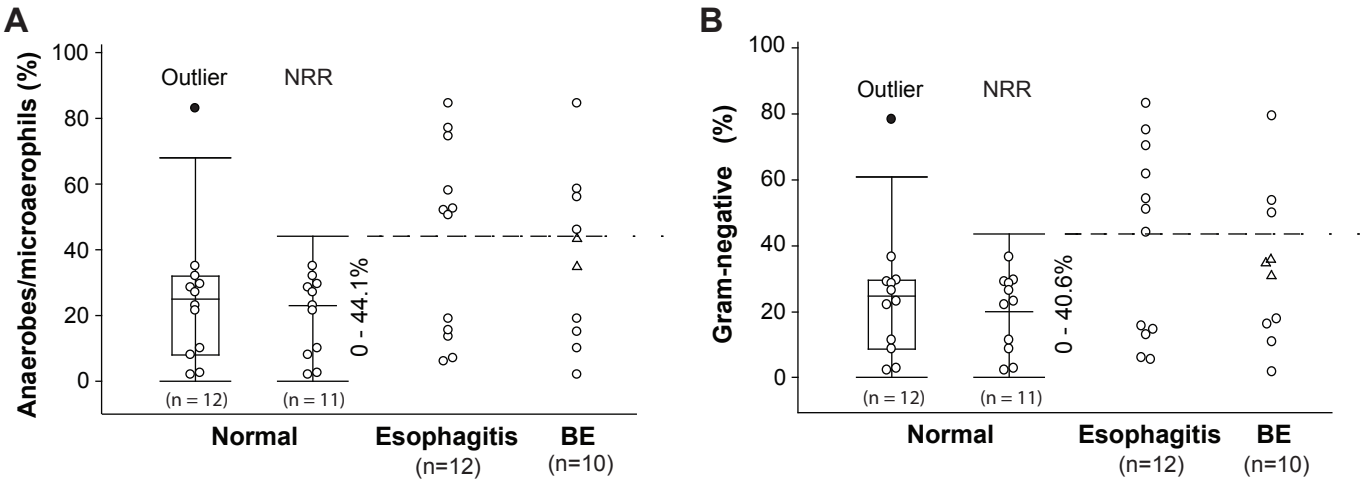
Supplementary Figure 2

Representation of bacterial species in the distal esophagus by phenotypes and microbiome types. Sequences representing unique species were used to construct the phylogram, as described in Experimental Procedures. Bootstrap values (based on 1500 replicates) are indicated at each node when >70%, and the branch length index is represented below the phylogram. Names of species-level operational taxonomic units (SLOTU) are located at the termination of each branch. SLOTUs represented by 16S rRNA gene clones are those that had ≥97% homology with PCR-amplified 16S sequences but no cultivation-defined species. They are named by the closest genus, followed by the Genbank accession number of the best-matched sequence. Unknown SLOTUs are represented by the closest taxon followed by the serial number of the clone used in this study, as well as the percent sequence identity (in parentheses). The abundance of a species and its sources were indicated on the right. The 166 SLOTUs belong to nine phyla, contrasted by alternating red and blue print.



Supplemental Figure 3

Estimation of species richness in the esophagus. The total number of species-level operational taxonomic units (SLOTU) that may be present in the human distal esophagus and their associated 95% confidence interval (C.I.) were calculated using a nonparametric richness estimator, Chao1 (30), hosted at http://purl.oclc.org/estimates. (A) Estimation of SLOTU richness in normal, esophagitis, and Barrett's esophagus (BE). Observed SLOTU richness for normal (blue), esophagitis (red), and BE (green) is represented by solid lines. Predicted SLOTU richness is shown by solid lines with square and 95% C.I. Based on the prediction, the present study has identified 82.5% (observed/predicted: 80/97), 82.0% (100/122), and 78.7% (100/127) SLOTUs in normal, esophagitis, and BE group, respectively (E15, as an outlier, was excluded from analysis of the species richness for the normal group). (B) Estimation of SLOTU richness in type I (blue) and type II (red) microbiome. Based on the prediction, the present study has identified 82.2% (observed/predicted: 106/129) and 68.6% (142/207) SLOTUs in type I and II microbiome, respectively. (C) Estimation of combined SLOTU richness in the distal esophagus based on 6,800 rDNA clones from 34 subjects. Chao I estimation indicates that the human distal esophagus may harbour ~213 (95% C.I. 191-254) SLOTUs and suggests that 77.9% of the SLOTUs (166/213) have been identified in this study.



Supplemental Figure 4

Classification of samples by the relative abundance of anaerobes/microaerophils (A) and Gram-negative bacteria (B). An outlier (solid circle) was excluded using a box plot in which the upper whisker length is 1.5*IQR. The 95% normal reference ranges (NRR) (mean ± 1.96 S.D.) were calculated by the relative abundance of anaerobes/microaerophils or Gram-negative bacteria in the normal samples after excluding the outlier. The dotted line is the upper limit of the 95% NRR, which separates the 34 samples into normal (inside the NRR) and abnormal taxonomic types (outside the NRR). Microbiome type II samples that were classified as the normal taxonomic type are shown by triangles.