Gut Microbiota in T1DM-Onset Pediatric Patients: Machine-Learning Algorithms to Classify Microorganisms as Disease Linked

Roberto Biassoni,[1,*] Eddi Di Marco,[1,*] Margherita Squillario,[2] Annalisa Barla,[2] Gianluca Piccolo,[3] Elisabetta Ugolotti,[1] Cinzia Gatti,[1] Nicola Minuto,[3] Giuseppa Patti,[4,5] Mohamad Maghnie,[3,4,5] and Giuseppe d'Annunzio[3]

[1]Molecular Diagnostics, Analysis Laboratory, IRCCS Istituto Giannina Gaslini, 16147 Genoa, Italy; [2]DIBRIS University of Genoa, 16145, Genoa, Italy; [3]Pediatric Clinic Regional Center for Pediatric Diabetes, IRCCS Istituto Giannina Gaslini, Genoa, 16147, Italy; [4]Department of Pediatrics, IRCCS Istituto Giannina Gaslini, University of Genoa, 16147 Genoa, Italy; and [5]Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University Of Genoa, 16147 Genoa, Italy

ORCiD numbers: 0000-0002-4365-8166 (R. Biassoni); 0000-0002-6217-1919 (G. Patti); 0000-0002-1856-2277 (G. d'Annunzio).

Aims: The purpose of this work is to find the gut microbial fingerprinting of pediatric patients with type 1 diabetes.

Methods: The microbiome of 31 children with type 1 diabetes at onset and of 25 healthy children was determined using multiple polymorphic regions of the 16S ribosomal RNA. We performed machine-learning analyses and metagenome functional analysis to identify significant taxa and their metabolic pathways content.

Results: Compared with healthy controls, patients showed a significantly higher relative abundance of the following most important taxa: *Bacteroides stercoris, Bacteroides fragilis, Bacteroides intestinalis, Bifidobacterium bifidum, Gammaproteobacteria* and its descendants, *Holdemania*, and *Synergistetes* and its descendants. On the contrary, the relative abundance of *Bacteroides vulgatus, Deltaproteobacteria* and its descendants, *Parasutterella* and the *Lactobacillus, Turicibacter* genera were significantly lower in patients with respect to healthy controls. The predicted metabolic pathway more associated with type 1 diabetes patients concerns "carbon metabolism," sugar and iron metabolisms in particular. Among the clinical variables considered, standardized body mass index, anti-insulin autoantibodies, glycemia, hemoglobin A_{1c}, Tanner stage, and age at onset emerged as most significant positively or negatively correlated with specific clusters of taxa.

Conclusions: The relative abundance and supervised analyses confirmed the importance of *B stercoris* in type 1 diabetes patients at onset and showed a relevant role of *Synergistetes* and its descendants in patients with respect to healthy controls. In general the robustness and coherence of the showed results underline the relevance of studying the microbioma using multiple polymorphic regions, different types of analysis, and different approaches within each analysis. **(*J Clin Endocrinol Metab* 105: e3114–e3126, 2020)**

Freeform/Key Words: adolescent, autoimmunity, children, type 1 diabetes mellitus, gut, machine learning algorithms, microbiota

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in USA

Received 14 November 2019. Accepted 1 July 2020.
First Published Online 21 July 2020.
Corrected and Typeset 24 July 2020.

*These authors equally contributed to the manuscript.

Abbreviations: GADA, glutamic acid decarboxylase antibodies; HbA_{1c}, hemoglobin A1c; HD, healthy donor; IA2A, antithyrosine phosphatase antibodies; IAA, anti-insulin autoantibody; KO, KEGG orthology; LDA, linear discriminant analysis; LEfSe, linear discriminant analysis effect size; OTU, operational taxonomic unit; ZnT8, anti-Zn transporter protein 8 antibodies.

Type 1 diabetes mellitus develops in genetically susceptible individuals by means of environmental factors that trigger an autoimmune inflammatory process within the pancreatic islets leading to β-cell loss (1, 2). Up to now the causative mechanisms have not been completely defined, and the identification of risk factors represents a challenge with practical, diagnostic, and therapeutic implications (3-5).

The increasing incidence of diabetes among genetically stable populations, the different incidence among neighboring regions, and the shift to a younger age at diabetes clinical onset strengthen the role of environmental factors in its pathogenesis (6). The main topics are perinatal and socioeconomic factors (7), the hygiene hypothesis (8), dietary components both in mother and in children (7), infectious agents (9), obesity and the "accelerator hypothesis" (10), epigenetic factors (11), and gut permeability (12).

Recently, it has been postulated that the gut immune system and gut-microbiota composition play a key role in human health and in the development of autoimmunity (13). The microbial flora consists of a dynamic ecosystem, with 10^{13} to 10^{14} microorganisms belonging up to thousands different species, whose genome codes for a number of genes up to 150 times human genes (14). Gut microbiota acts as an endocrine organ that translates nutritional factors into hormone-like signals; its composition is influenced by dietary habits and geographical locations. Gut microbiota is involved in several activities: complex carbohydrates digestion, vitamin synthesis, immune and inflammatory responses, and hormone and neurotransmitter synthesis, all influencing host physiology and disease susceptibility (13). In particular gut microbiota interacts with the mucosal environment, modulating intestinal permeability, local and systemic inflammatory activity, and cross-talking with the gut immune system (15). The immunological link between the gut and pancreas by shared lymphocyte homing receptors contributes to the development of diabetes together with gut immunological modification and dysbiosis (16). Moreover, the gut-brain axis and gut-hypothalamus axis are influenced by microorganisms to regulate food intake and energy expenditure (13).

Studies on animal models, particularly in BB rats and the NOD mouse, have reported conflicting results (14, 17). An increased presence of *Bacteroidetes*, *Eubacterium*, and *Ruminococcus* in BB-diabetes prone rats was reported, whereas a decrease of *Firmicutes* and an increase of *Bacteroides* in NOD mice has been described (14). On the other hand, certain strains of *Bacteroidetes*, for example, *Prevotella*, seem to be protective against disease development (17). These different results might be due to design of the experimental model, environmental factors, storage and processing of stool samples, and regional variations in gut microbiota.

Animal studies translated to humans confirmed the involvement of gut microbiota in diabetes pathogenesis (18-20). Several differences in gut microbiota have been reported between healthy children and those with diabetes. Healthy children had a higher percentage of butyrate-producing and mucin-degrading bacteria, as compared to peers with diabetes, who otherwise showed a decrease in *Firmicutes/Bacteroidetes* ratio (18) and a higher abundance of *Bifidobacterium pseudocatenulatum*, *Roseburia hominis*, and *Alistipes shahii* (19). Children with active β-cell autoimmunity and prediabetes showed an increased ratio of *Bacteroidetes* vs *Firmicutes* (17). Recently, a disease-specific duodenal mucosa characterized by a distinctive inflammatory profile and microbiota composition was reported (20). These observations support the role of the gastrointestinal ecosystem in the development of autoimmunity (21) and target intestinal flora as a potential future new strategy for diabetes prevention and treatment (22, 23).

The primary aim of our study was to define gut-microbiota composition in a cohort of newly diagnosed children and adolescents with type 1 diabetes. Our secondary aim was to analyze specific bacteria taxa with clinical and metabolic parameters of the disease.

Methods

We evaluated 31 children and adolescents (20 males, 11 females) with newly diagnosed type 1 diabetes (mean age, 10.3 years with an SD of 4.1, median 9.5 years) between January 1, 2016 and May 31, 2018. All patients were admitted to the emergency section of the hospital and then sent to the in-ward section of the Pediatric Clinic, the Regional Center for Pediatric Diabetes, Giannina Gaslini Institute, Genoa. No patient needed admission to the intensive care unit. Clinical characteristics are reported in Table 1. In particular mean glucose levels at first detection were 462 ± 215 mg/dL (M ± SD), mean hemoglobin A_{1c} (HbA_{1c}) levels were 10.75 ± 1.92% (94.00 ± 21.09 mmol/mol), mean C peptide levels were 0.48 ± 0.33 ng/mL. Mean serum pH values were 7.31 ± 0.13. Inclusion criteria were individuals living in Northern Italy, born from Caucasian parents, singleton birth, diagnosis of type 1 diabetes according to International Society for Pediatric and Adolescent Diabetes guidelines (23), personal history negative for acute or chronic gastrointestinal diseases, and/or antibiotic or probiotics administration in the previous month. Exclusion criteria were forms other than type 1 diabetes, coexistence or concomitant diagnosis of celiac disease, personal history positive for gastrointestinal acute and chronic illnesses, and antibiotics or probiotics administration in the previous month. Data regarding sex, age at diabetes onset, presence of ketoacidosis, gestational age, mode of delivery, duration of breastfeeding, and age at weaning were also recorded. Screening for

Table 1. Newly diagnosed patients with type 1 diabetes and clinical variables

Sample	Sex	Age at onset, y	pH	Glycemia, mg/dL	HbA_{1c}, %	HbA_{1c}, mmol/mol	Tanner	BMI SDS	GADA	IAA	ZnT8	IA2A	C peptide ng/mL	Delivery type	Milk, d	Weaning, mo	Gluten, mo
1D	M	13.86	7.36	275	6.21	44	III	2.45	1	1	1	1	1.5	1	425	6	8
2D	M	7.76	7.35	703	10.42	90	I	2.68	1	2	1	1	0.3	1	240	5	7
3D	M	10.51	7.36	227	11.75	105	II	0.22	1	2			0.5	2	0	6	8
8D	F	4.59	7.34	451	8.7	71	I	0.1	1	2	1	1	0.4	2	500	4	6
10D	F	7.86	7.35	422	8.74	72	I	-0.91	1	2	1	2	0.4	1	360	8	9
12D	M	7.87	7.33	484	9.15	77	I	–6.7	1	1	1	1	0.3	2	0	6	9-12
18D	F	6.05	7.41	476	9.1	76	I	0.67	2	2	NA	NA	0.3	1[a]	365 + f	6.5	6.5
19D	F	9.45	7.4	287	10.8	95	I	–1.98	1	2	NA	NA	0.2	1	120	6	10
21D	M	5.43	7.36	305	12.7	115	I	–0.87	1	2	2	2	0.3	2	300	7	7
22D	M	14.30	7.33	420	14.3	133	IV	–1.76	1	2	1	2	0.4	1[b]	180 + f	5	6[c]-7[c]
25D	M	15.36	7.38	453	11.3	100	V	0.83	1	2	1	1	0.6	1	90	5	5
26D	F	8.01	7.4	348	12.07	108	I	–2.66	1	2	1	1	0.3	1 (36[c] wk)	10	7	5
28D	M	12.80	7.38	400	10.5	91	II	–1.62	1	1	1	1	0.25	1[c]	Formula	6	8
29D	F	4.06	7.11	1340	10.3	89	I	–1.3	1	1	1	1	0.36	1	Breast	NA	NA
30D	F	8.99	7.34	414	9.19	77	I	0.62	1	2	1	2	0.3	1[d]	Formula	4	10
31D	F	16.10	7.34	806	11.28	100	V	–0.41	1	1	2	2	0.61	2	900	6	6
32D	F	8.13	7.32	509	13.23	121	I	–1.19	1	2	1	2	0.79	1	Formula	4-5	6-7
33D	M	8.54	7.39	652	11	97	I	0.2	1	1	1	2	0.09	1	300	6	10
35D	M	14.33	6.96	493	10.47	91	II	–1.43	1	2	2	2	0.21	1	NA	NA	NA
36D	M	8.92	7.33	230	8.8	73	I	2.13	1	2	1	1	0.23	2	Formula	5	7
37D	M	13.32	7.36	438	12.54	114	IV	–0.75	1	2	1	1	0.35	1	900	NA	NA
42D	F	4.16	7.39	404	10.24	88	I	NA	1	2	1	1	0.75	1[a]	1	Formula	5
47D	M	17.41	6.9	420	12.27	111	V	NA	1	2	1	2	1.11	1	120	5	5
48D	M	10.08	7.15	412	14.04	130	I	1.48	1	2	1	1	0.28	1	90[d]	6	6
49D	F	17.18	7.35	298	11.6	103	V	–0.45	1	2	1	1	1.26	1	Formula	6	6
50D	M	15.24	7.35	567	9.03	75	IV	–0.32	1	2	2	2	0.54	1	180	6	9
54D	M	12.75	7.38	657	7.81	62	II	0.14	1	1	1	1	0.7	1	90	6	6
57D	M	11.00	7.3	369	13.19	121	III	–0.87	1	2	1	1	NA	2	450	15	15
58D	M	5.50	7.47	180	9.95	85	I	–1.24	2	2	1	2	0.17	1	60	6	6
60D	M	14.50	7.37	455	13.28	122	IV	–1.19	1	2	NA	NA	NA	1	900	10	10
61D	M	7.20	7.16	432	9.32	78	I	1.95	1	2	NA	NA	0.3	1	60[e]	6	6

Autoantibodies are indicated as 1, present; 2, absent. Type of delivery: 1, natural; 2, cesarean ([a]dystocial; [b]inductive suction cup; [c]dystocial twin; 1[d]dystocial for gestosis). Breastfeeding days or use of formula milk are indicated (f: formula integrated from day 0; f1: formula integrated from day 20; formula: cow milk from 300 days; [e]mixed from beginning.

Abbreviations: BMI SDS, standardized body mass index; F, female; GADA, glutamic acid decarboxylase antibodies; HbA_{1c}, hemoglobinA_{1c}, IA2A, antithyrosine phosphatase antibodies; IAAs, anti-insulin autoantibodies; M, male; ZnT8, anti-Zn transporter protein 8 antibodies.

the most frequent autoimmune diseases associated with type 1 diabetes (ie, celiac and thyroid diseases) was performed in all patients.

As expression of endogenous insulin secretion serum C peptide measurement by electrochemiluminescence assay was performed in all patients. Reference values were 1 to 3 ng/mL.

Immunological markers of type 1 diabetes were detected in all patients. Glutamic acid decarboxylase (GADA), antithyrosine phosphatase (IA2A), and anti-Zn transporter protein 8 (ZnT8) autoantibodies were detected by enzyme-linked immunosorbent assay, and anti-insulin autoantibodies (IAAs) by radioimmunoassay.

As controls, 25 sex-matched healthy donors (HDs) (age 10.3 ± 4.1 years, median 10.0 years) were analyzed using the same approach. Informed consent was obtained by patients and caregivers, and the study was approved by the local ethics committee.

Fecal microbioma analysis

Fecal samples were collected from HDs and from newly diagnosed patients within 1 week after recovery from metabolic decompensation, if present, and during multi-injective daily insulin therapy. All patients were in good clinical conditions at the time of collection and showed near-normal blood glucose levels. Stool samples were transferred on collection at –20 C and stored in the same conditions until DNA extraction. DNA extraction from fecal samples was performed resuspending a tiny quantity of feces into 1 mL of the ASL Stool lysis buffer (Qiagen GmbH) by vigorous pipetting. DNA was extracted from 200 µL of this suspension with a MagDEA DNA 200GC extraction kit and PSS Magtration System 12GC automated platform, according to the manufacturer's instructions (Precision System Science PSS Co, Ltd). DNA was eluted into 100 µL of 10 mM Tris pH 8.0/1 mM EDTA (TE) buffer, and quality and quantity were evaluated by spectrophotometric and Qubit fluorimetric quantitation assays, respectively (Thermo Fisher Scientific). A total of 3 ng of DNA were used for each 16S amplification reaction performed with an Ion 16S Metagenomics Kit (Thermo Fisher Scientific). It allowed the polymerase chain reaction amplification of 7 out of 9 informative 16S polymorphic regions (V2, V4, V8, V3, V6-7, and V9) according to the manufacturer's protocols.

The IonPlus-Library kit for AB library builder (Thermo Fisher Scientific) was used for library synthesis. Differently bar-coded libraries were automatically handled into a Ion 520 chip by the Ion Chef System and sequenced by GeneStudio S5 system (Thermo Fisher Scientific). Data analysis was performed with Ion Reporter suite software (v 5.10) using both the curated Greengenes (v13.5) and the premium curated MicroSEQ ID 16S ribosomal RNA reference library (v2013.1) databases with standard parameters. In detail, the analysis was performed using 2 machine-learning algorithms (Random Forest and Elastic Net l_1l_2), together with complete compositional and biodiversity α and β index analyses. In addition, comparative analysis and predictions of metabolic potentials were studied. In detail, comparative analysis used multiple data filtering and normalization techniques coupled with differential analysis algorithms like linear discriminant analysis (LDA) effect size (LEfSe), metagenomeSeq (based on zero-inflated Gaussian fit or Fitfeature statistical models), EdgeR, and DEseq2. All statistical analyses were performed using *P* values of less than .05 adjusted using the false discovery rate method, and all parameters were always analyzed at phylum, class, order, family, genus, and species levels.

Bioinformatics and statistics

Statistical, visual, and meta-analyses of microbiome data were performed with the MicrobiomeAnalyst tool (24). Data filtering for low abundance and low variance operational taxonomic units (OTUs, based on the prevalence in 20% of samples and interquartile range set at 10%) was applied for all relative abundance comparisons using the algorithms metagenomeSeq, EdgeR, DESeq2, linear discriminant analysis (LDA), and effect size-LEfSe.

Finally, 2 different machine-learning algorithms, Random Forest (24) and l_1l_2 (25), for biomarker classification and identification were applied.

Data preprocessing

The R package microbiomeSeq (available at http://userweb.eng.gla.ac.uk/umer.ijaz/projects/microbiomeSeq_Tutorial.html#content) was used to normalize separately the phylum, class, order, family, genus, and species data matrices. This package is suggested in a recent paper that is focused on normalization methods for microbioma data (26). For both l_1l_2 and Random Forest, we analyzed separately phylum, class, order, family, genus, and species, whose numerosity for the 56 samples was 12, 25, 53, 106, 139, 318, respectively.

Supervised analysis: *l1l2*

For multivariate variable selection, we chose l_1l_2, an embedded regularization method that combines 2 penalty terms, 1 enhancing sparcity (l_1 norm) and the other retaining correlated variables (l_2 norm). The algorithm can be tuned to give sets of discriminative variables of different sizes (25). In particular this method was used inside PALLADIO, a machine-learning framework based on regularization methods. It provides an estimate of its reliability by means of a nonparametric, 2-sample Kolmogorov-Smirnov test, along with an assessment of prediction performance.

Unsupervised analysis: weighted correlation network analysis

To verify if the abundance of particular groups of taxa was significantly associated to relevant clinical features characterizing diabetic patients, we performed a multivariate analysis using weighted correlation network analysis (WGCNA) clustering methodology (27). The R package implementation of this method was used (28). In particular we leveraged on the selections of classes, orders, families, genus, and species identified by l_1l_2 and considered the following clinical features: sex, age at onset, glycemia, glycated HbA_{1c}, and C peptide levels, Tanner stage, BMI-SDS, GADA IA2A, ZnT8, and IAA positivity, type of delivery (natural or cesarean) and timing of gluten introduction.

Inferring functional (metabolic) pathways using PICRUSt

The metagenome functional content was predicted using PICRUSt (29) from biom files, thus we obtained the KEGG orthology (KO) terms table. These data were analyzed with

MicrobiomeAnalyst to identify a list of the most significant KO able to discriminate patients from controls.

Results

Microbiome analysis

The study of the fecal microbiota was performed using next-generation sequencing, enabling the analysis of the complete bacterial contents of the samples. It is known that the hypervariable regions of the 16S ribosomal gene exhibit different degrees of sequence diversity, allowing us to assign the microbial taxonomy of mixed bacterial populations. It is also true that the analysis of a single hypervariable region is not able to distinguish among all bacteria. In our approach we used the sequencing of 7 polymorphic regions to avoid as much as possible potential false-negative taxonomic classifications and any bias in the identification of microorganisms in a complex microbial population.

The 63 analyzed microbiome samples resulted in the identification of 1606 OTUs in patients and 1552 OTUs in HDs. A comparative abundance profile across experimental groups at the phylum, family, and genus taxonomic levels are shown in our supplemental materials (30).

α and β Diversity analysis

To allow comparison of patients and HDs, these analysis were performed rarefying data to the minimum library size. The α diversity profiling indexes, which estimate community richness (observed_species, Chao1-index, and abundance-based coverage estimators or ACEs), increased in HDs compared to patients (as indicated by Mann-Whitney statistical analysis). Indeed, the number of observed species and the Chao1 estimator increased both at the level of class, order, and species ($P \leq .05$), and at family and genus ($P \leq .01$), whereas ACEs reached statistical significance only at the family, genus ($P \leq .01$), and species ($P \leq .05$) levels (data not shown). The use of indexes estimating community richness and evenness (Shannon, Simpson, and Fisher) gave more variegated results. The Shannon and the Simpson indexes both showed no significant differences with the exception of the Shannon index evaluated at the phylum level ($P \leq .05$). The Fisher diversity index showed a statistically significant increasing trend in patient samples at the phylum, class, order ($P \leq .05$), family, genus ($P \leq .01$), and species ($P \leq .05$) levels (data not shown).

The β diversity (unweighted UniFrac and Bray-Curtis index) was performed by principal coordinate analysis as the ordination method and was statistically analyzed with a permutational multivariate analysis of variance test. In detail, unweighted UniFrac allowed for the presence or absence of different taxa ($P < .044$), whereas the Bray-Curtis index gave information about the abundance of data ($P < .008$) (30). Finally, α and β diversity analyses showed a more complex organization of the gut-bacterial flora in patients.

Taxonomy-based differential abundance analysis

The differential abundance of taxa among samples belonging to either HDs or newly diagnosed type 1 diabetes patients was evaluated with different algorithms (metagenomeSeq, EdgeR, DESeq2, LDA, and LEfSe). All *P* values shown have been adjusted using the false discovery rate approach and are indicated as "q=."

Different *Bacteroidetes* species like *B stercoris* (q = 1.473E-4), *B intestinalis* (q = 0.010), and *B fragilis* (q = 0.0452) were statistically significant in patients compared to HDs (Table 2).

We observed a higher relative abundance of *Synergistetes* (q = 0.001), *Synergistia* (q = 4.798E-5), *Synergistales* (q = 2.64E-4), and *Synergistaceae* (q = 0.0017) in patients than in the HD group (Table 2). Patients showed the previously cited taxa about 20 times more than HDs, as indicated by log2 fold changes.

Among *Proteobacteria* the class *Gammaproteobacteria* was more abundant in patients than in HDs (q = 0.0226) as well as its descendant order *Enterobacteriales* (q = 0.0226) and family *Enterobacteriaceae* (q = 0.0379), which showed a 3-times higher abundance in patients. In contrast, the class *Deltaproteobacteria* (q = 0.002) and *Betaproteobacteria* (q = 0.031) were around 3 times less abundant in patient samples than in HDs (Table 2). Among *Deltaproteobacteria* descendants, the *Desulfovibrionales* (q = 0.0016), *Desulfovibrionaceae* (q = 0.016), and *Bilophila* genera (q = 0.036) showed lower relative abundance in patients as compared to HDs (0.30, 0.33, or 0.22, respectively). Within *Betaproteobacteria* descendants, *Parasutterella* (q = 0.018) and *Sutterella* (q = 0.0013) showed the opposite behavior, their abundance in patients vs HD being 0.13 and 7.67, respectively.

Oscillospiraceae (q = 0.027) and unclassified *Clostridiales*, both descendants of *Firmicutes*, showed a 3- and 10-fold relative abundance increase in patients over HDs, respectively. *Eubacterium hallii* and *Eubacterium siraeum* behaved the opposite way, with the relative abundance in patients compared to HD for the former was 0.16 and for the latter 8.86, respectively. Finally, *Bifidobacterium bifidum*, a descendant of *Actinobacteria*, was 8 times more abundant in patients than in HDs (q = 0.0169).

Table 2. Relative abundance analysis obtained with 4 different algorithms in newly diagnosed patients with type 1 diabetes

Taxa	Zero-inflated Gaussian fit	EdgeR	EdgeR	DESeq2	DESeq2	LDA-LEfSe	LDA-LEfSe
Actinobacteria	FDR q	log2FC	FDR q	log2FC	FDR q	LDA score	FDR q
Bifidobacterium bifidum	0.0473	2.9550	0.0169	–	–	–	–
Bacteroidetes	FDR q	log2FC	FDR q	log2FC	FDR q	LDA score	FDR q
Bacteroides stercoris	1.4730E-4	–	–	–	–	–	–
Bacteroides intestinalis	0.0102	–	–	–	–	–	–
Bacteroides fragilis	–	2.8052	0.0452	–	–	–	–
Prevotella	0.0125	–	–	–	–	–	–
Prevotella copri	0.0133	–	–	–	–	–	–
Alistipes	0.0447	–	–	–	–	–	–
Alistipes indistinctus	0.0331	–	–	–	–	–	–
Firmicutes	FDR q	log2FC	FDR q	log2FC	FDR q	LDA score	FDR q
Bacillales	0.0255	–	–	–	–	–	–
Lactobacillus	0.0332	–2.9873	0.0044	–	–	–	–
Holdemania	0.0093	–	–	–	–	–	–
Holdemania filiformis	0.0331	–	–	–	–	–	–
Turicibacter	0.0050	–	–	–	–	–	–
Turicibacter sanguinis	0.0497	–	–	–	–	–	–
Coprococcus catus	0.0331	–	–	–	–	–	–
Oscillospiraceae	0.0271	–	–	–	–	2.86	0.0198
Pseudoflavonifractor	0.0125	–	–	–	–	–	–
Unclassified Clostridiales	–	3.4351	0.0014	–	–	–	–
Eubacterium hallii	–	–2.6969	9.5187E-4	–	–	–	–
Eubacterium siraeum	–	3.1480	0.0299	–	–	–	–
Proteobacteria	FDR q	log2FC	FDR q	log2FC	FDR q	LDA score	FDR q
Gammaproteobacteria	–	1.3140	0.0226	1.7177	0.0026	–	–
Enterobacteriales	–	–	–	1.6479	0.0226	–	–
Enterobacteriaceae	–	–	–	1.6692	0.0379	–	–
Betaproteobacteria	–	–1.4771	0.0312	–	–	–	–
Parasutterella	0.0177	–2.9632	0.0318	–	–	–	–
Parasutterella excrementihominis	0.0331	–	–	–	–	–	–
Sutterella	0.0013	2.9407	0.0452	–	–	–	–
Deltaproteobacteria	–	–1.5005	0.0018	–	–	–	–
Desulfovibrionales	–	–1.7498	0.0016	–1.1091	0.0377	–	–
Desulfovibrionaceae	–	–1.6110	0.0162	–1.3542	0.0379	–	–
Bilophila	–	–2.1806	0.0356	–	–	–	–
Synergistetes	FDR q	log2FC	FDR q	log2FC	FDR q	LDA score	FDR q
Synergistetes	0.0014	4.3392	5.0399E-5	4.4556	0.0427	2.59	0.0050
Synergistia	4.7975E-5	4.6856	4.7975E-5	4.6863	0.0272	2.59	0.0101
Synergistales	2.6399E-4	4.4596	7.1403E-5	4.5482	0.0377	2.59	0.0126
Synergistaceae	0.0017	4.2063	1.8713E-4	–	–	2.59	0.0198
Verrucomicrobia	FDR q	log2FC	FDR q	log2FC	FDR q	LDA score	FDR q
Verrucomicrobiales	0.0200	–	–	–	–	–	–
Akkermansia muciniphila	0.0497	–	–	–	–	–	–

Abbreviations: FC, fold change; FDR, false discovery rate; LDA, linear discriminant analysis; LEfSe, linear discriminant analysis effect size.

Supervised analysis

We applied Random Forest and l_1l_2 algorithms to identify taxable to discriminate between patients and HD with good classification performances. We analyzed separately phylum, class, order, family, genus, and species considering both machine-learning methods. Interestingly, both methods produced overlapping results (Figure 1). Among the identified species, *B stercoris* was the most significant and robust species found both by Random Forest and l_1l_2, with a higher abundance in patients than in HDs, as shown in Fig. 1. *B vulgatus* was another species classified with the highest scores by both methods. Interestingly, although both species are descendants of the *Bacteroidetes* phylum, the *B vulgatus* behaved in the opposite way if compared to *B stercoris* (Fig. 1), since the *B vulgatus* relative abundance was higher in HDs, as already reported (31). Again, *Synergistetes, Synergistia*, *Synergistales*, *and Synergistaceae* were all identified among the features best describing patient samples. These bacteria are suggested to be opportunistic pathogens because they are known to be implicated in periodontal disease, gastrointestinal infections, and soft-tissue infections but they can also be found in HDs (Fig. 1).

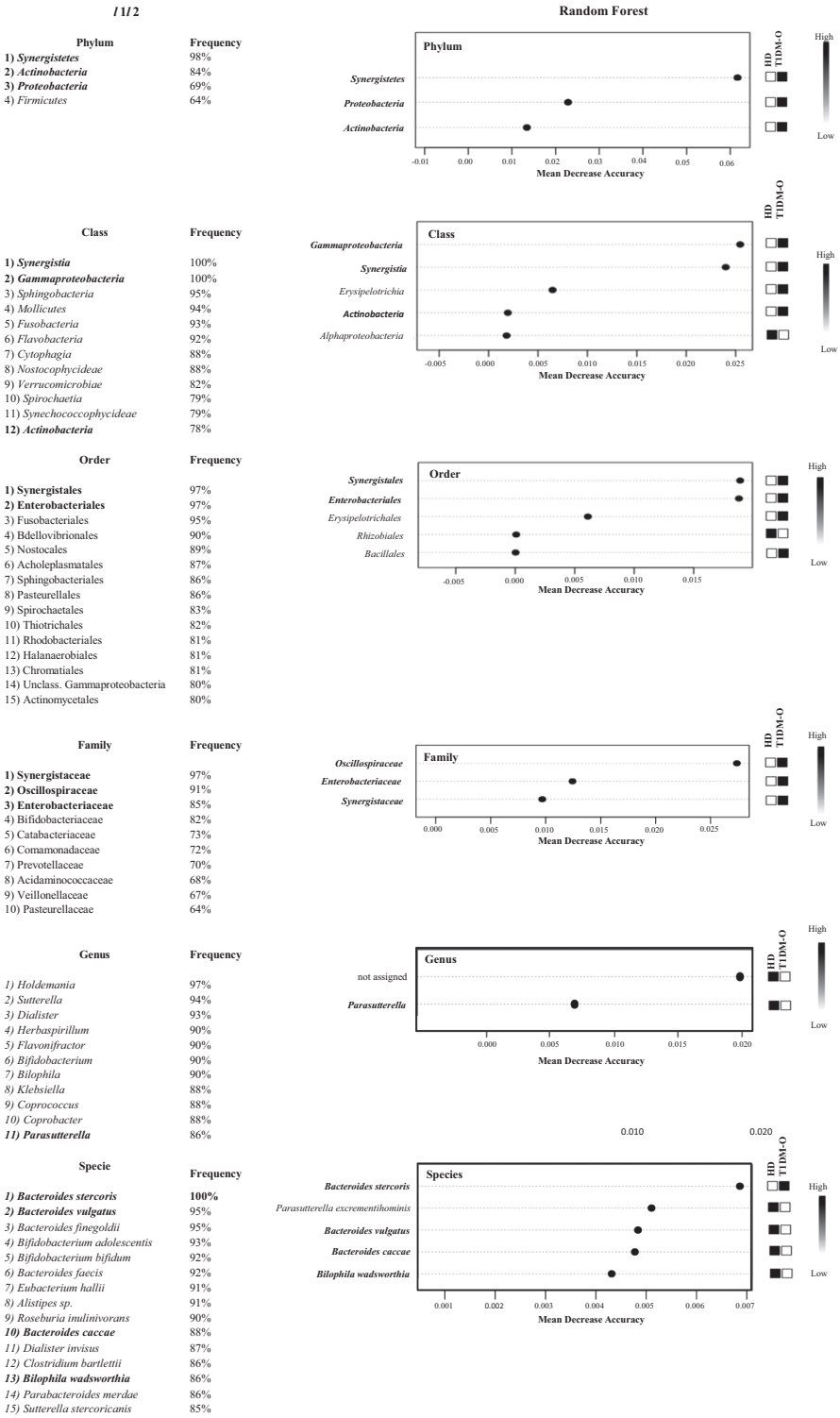

Figure 1. Classification of microbial taxonomy using 2 different machine-learning algorithms. At left are shown the features that are relevant at the different taxonomic levels using the l_1l_2 algorithm. They are classified so the best have the highest percentage (mean balanced accuracy and Matthew correlation coefficient: 0.578, 0.166). Bold and gray-filled taxonomic features are the ones present in the Random Forest classification. At right we show the results of Random Forest based on 5000 decision trees, with the most important variables in the model being highest in the plot and having the largest mean decrease accuracy (mean value of out of bucket error, sensitivity and specificity of 0.408, 0.54, 0.62). The plots indicate only relevant variables. The white- or black-filled boxes on the part of the diagrams indicate the high or low relative abundances both in patients at onset and healthy donor individuals.

The *Proteobacteria* and their descendants *Gammaproteobacteria*, *Enterobacteriales*, and *Enterobacteriace* were among the features best identified by both algorithms. The same was true for *Bilophila and B wadsworthia*, both descendants of *Deltaproteobacteria* (Fig. 1). *Sphingobacteriia* and *Erysipelotrichia* had relevant scores with only 1 of the 2 algorithms used (l_1l_2 or Random Forest). *Sphingobacteriia* is known to produce

sphingolipids, which recently have been shown to be important mediators in the signaling cascades involved in stress responses and inflammation (32). *Erysipelotrichia* and its descendants *Holdemania and Turicibacter* genera were among the differentially abundant features reported in Table 2. *Oscillospiraceae*, *Actinobacteria*, and its *Bifidobacterium* descendant genus have a higher abundance in patients than in HDs (Fig. 1).

Microbiome-associated metabolic pathways analysis

Members of certain taxonomic groups share metabolic pathways and produce metabolites that may be associated or participate in mechanisms triggering pathologies. By the gene content of the microbiota, we obtained functional profiling (metabolic pathways) that characterize these microorganisms (29). Patients showed a group of 2696 KO identifiers based on molecular functions sorted according to LDA score ($P \leq .05$) (30, see Table 1). Fig. 2 shows the best 20 KO-identifiers both for patients and HDs based on their LDA score. In diabetic individuals we had 712 "metabolic pathways," 107 of them associated with "carbon metabolism" and in particular with glucose metabolism. In detail, these pathways in the following order were linked to glycolysis and gluconeogenesis, fructose and mannose metabolism, starch and sucrose metabolism, pentose phosphate pathway and galactose metabolism (45, 42, 39, 27, and 25 pathways, respectively). It is of note that among the best KO identifiers we found some

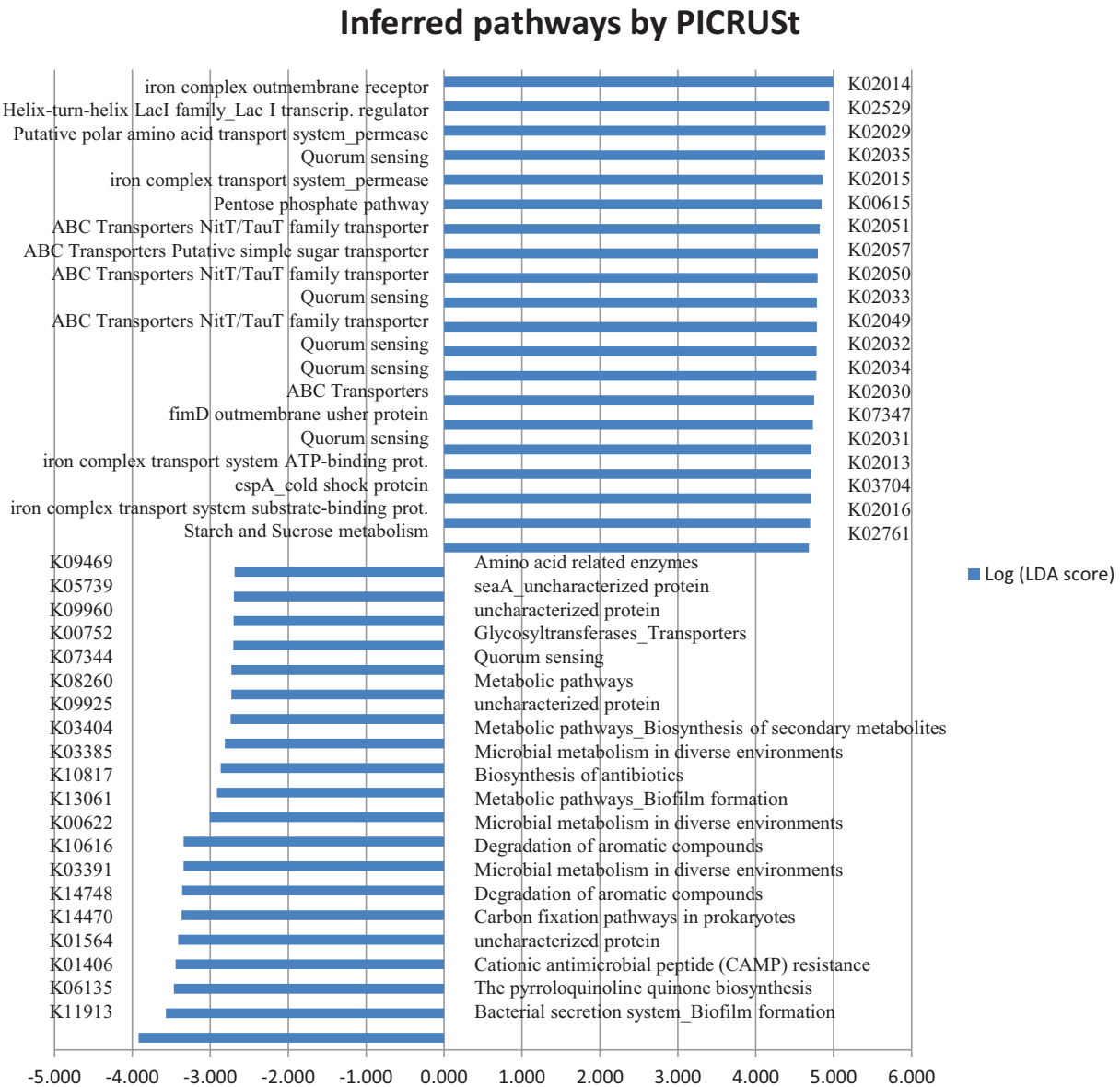

Figure 2. The best 20 KO (Kegg orthology) identifiers both for diabetics (on top) and healthy donors (Characterized by negative LDA Scores). The analysis was based on linear discriminant analysis (LDA) score and *P* less than or equal to .05. Each identifier is indicated on the right for the patients group and on the left for healthy donor participants; the relative pathways, or the name of the proteins, are also shown for each entries.

pathways were associated with the iron complex transport/permease protein K02015 and iron complex outer-membrane receptor protein transporter linked to pore ion-channels, K02014. Higher levels of iron in tissues were described as a risk factor for insulin resistance in type 2 diabetes (33).

Unsupervised analysis

To verify whether the abundance of particular groups of taxa was significantly associated with relevant clinical features that characterize patients with diabetes, we performed a clustering analysis using the WGCNA method (see "Methods").

In particular we leveraged the selections of classes, orders, families, genera, and species and all-taxa identified by l_1l_2, and we considered the list of clinical features specified in "Methods."

Table 3 shows that among the clinical features considered, age at onset, sex, Tanner pubertal stage, BMI SDS, glycemia, glycated HbA_{1c}, and IAA levels and type of delivery were significantly positively or negatively related to the abundance of specific clusters of taxa.

It is of note that in the unsupervised analysis, IAA positivity and BMI SDS were the clinical variables with the highest number of significantly correlated taxa. IAA presence showed the major numbers of correlation with clusters of taxa among the autoantibodies. In detail, a negative correlation with *Flavobacteriia*, *Actinobacteria, Gammaproteobacteria*, *Synergistia, Synergistales, Enterobacteriales*, *Oscillispiraceae*, *Synergistaceae*, *Enterobacteriaceae*, *Catabacteriaceae Bifidobacteriaceae*, and *Clostridium bartletti* was observed both in relative abundance and supervised analyses (see Table 2 and Fig. 1). Again, BMI SDS was negatively correlated with *Flavobacteriia*, *Synergistia*, *Actinobacteria*, *Gammaproteobacteria, Synergistales*, *Catabacteriaceae*, and *Oscillospiraceae*, and positively correlated with *Bifidobacterium adolescentis* and *Bifidobacterium bifidum* (Pearson correlation coefficient [PCC] = 0.43 and *P* value = .02). The *Bilophila*, *Sutterella*, *Flavonifractor*, and *Holdemania* genera and the *Bilophila wadsworthia* and *Sutterella stercoricanis* species showed a negative correlation with HbA_{1c} levels (PCC = –0.48 and *P* value = .006) and were significant in the supervised analysis (see Fig 1). The *Bilophila*, *Sutterella*, and *Holdemania* (with the exception of *Flavonifractor genus)* were relevant in the relative abundance analysis. Whereas for *Bilophila* and *Sutterella* the relative abundance analysis reached a statistical significance level at the genus level, only the supervised and the unsupervised methods were able to reach statistical significance at the species level (*B wadsworthia* and *S stercoricanis*) with the Tanner scale (PCC = –0.44 and *P* value = .01). In addition, *Clostridium bartlettii* and *Bacteroides finegoldii* were statistically significant both in the supervised and the unsupervised methods. It is remarkable that *B wadsworthia* and *S stercoricanis* clustered together and were negatively correlated with age at diabetes onset (PCC = –0.45 and *P* value = .01). Differently from the relative abundance analysis, both l_1l_2 and WGCNA found *Klebsiella* and *Coprobacter* genera being statistically significant, in the former to discriminate between patients and HDs, and in the latter to strongly correlate with glycemia (PCC = 0.58 and *P* value = 7E-04) in a positive direction, while both of them correlated negatively with age at onset. Moreover, analyzing all taxa selected by the l_1l_2 algorithm (ie, all-taxa comprising phylum, class, order, family, genus, and species together) we found a strikingly positive correlation between blood pH at diabetes diagnosis and the cluster *Herbaspirillum*, *Coprococcus, Coprobacter*, *Eubacterium hallii*, *Bacteroides faecis*, *Bacteroides stercoris*, *Bacteroides vulgatus, and Parabacteroides merdae*, *Actinobacteria*, and its descendants (*Bifidobacteriaceae* and *Bifidobacterium*) (PCC = 0.57 and *P* value = 9E-04), and a significant negative correlation between HbA_{1c} levels and the cluster *Bilophila* and *B wadsworthia* (PCC = –0.5, *P*-value = .004) (data not shown).

Discussion

In our study we observed a significant shift in the microbiota composition in newly diagnosed children and adolescents with type 1 diabetes. We used metagenomics to study gut microbiota in patients with type 1 diabetes at clinical onset compared to fecal microbial flora in HDs.

Our data showed that controls had higher α diversity than cases. Others like Cinek did not find any differences, whereas Kostic found a "drop" in case alpha diversity between seroconversion and type 1 diabetes diagnosis (34, 35). Kemppainen showed that bacterial diversity differed by geographical location and the Shannon index differed significantly at each site (36). In addition, our analysis takes into consideration 7 polymorphic regions of the 16S gene, whereas the majority of published articles used a single or 2 different regions instead. Different approaches used might be reflected in α diversity evaluation. Thus, it is possible that different conditions, either biological and/or technical, may influence the α diversity of cases rather than controls, such as time of seroconversion and diagnosis.

Although *Bacteroidetes* do not represent the major phylum in patients (37), *Bacteroides* genus and even

doi:10.1210/clinem/dgaa407

Table 3. Unsupervised analysis: significant results of weighted correlation network analysis in which the correlation between "taxa clusters" and clinical features among the 31 diabetic type 1 diabetes mellitus–onset patients are calculated. Taxa are those found statistically discriminant by *l1l2*. Arrow direction indicates positive (up) and negative (down) correlation. The Tanner scale is complete of its 5 stages

	Synergistetes	Actinobacteria	Proteobacteria	Firmicutes	Bacteroidetes
Age at onset		↑*Bifidobacterium*	↑*Herbaspirillum* ↓ (*Klebsiella*, *Bilophila wadsworthia*, *Sutterella stercoricanis*)	↑*Coprococcus* ↓*Clostridium bartlettii*	↓ (*Coprobacter, Bacteroides finegoldii*)
Sex (male)			↓*Klebsiella*		↓ *Coprobacter*
Tanner			↓ (*Bilophila wadsworthia*, *Sutterella stercoricanis*)	↓*Clostridium bartlettii*	↓ *Bacteroides finegoldii*
Glycemia		↑*Actinomycetales* ↓*Bifidobacteriaceae*	↓ (*Thiotrichales*, *Rhodobacterales*) ↑ *Klebsiella*		↑ *Coprobacter*
Glycated HbA_{1c}			↓ (*Bilophila* *Bilophila wadsworthia Sutterella Sutterella stercoricanis*)	↓ (*Flavonifractor*, *Holdemania*)	
BMI SDS	↓ (*Synergistia*, *Synergistales*)	↓*Actinobacteria* ↑ (*B adolescentis, B bifidum*)	↓*Gammaproteobacteria*	↓(Catabacteriacea, Oscillospiraceae)	↓*Flavobacteriia*
IAA (presence)	↓ (*Synergistia*, *Synergistales*, *Synergistaceae*)	↓ (Actinobacteria, *Bifidobacteriaceae*)	↓ (*Gammaproteobacteria*, *Enterobacteriales*, *Enterobacteriaceae*)	↓ (*Catabacteriaceae*, *Oscillospiraceae*, *Clostridium bartlettii*)	↓*Flavobacteriia*
GADA (presence)		↓ (*Actinobacteria*, *Bifidobacteriaceae*, *Bifidobacterium*)		↓*Eubacterium hallii*	↓(*Coprobacter*, *Alistipes sp*, *Bacteroides caccae, Bacteroides faecis* *Bacteroides stercoris, Bacteroides vulgatus, Parabacteroides merdae*)
IA2A (presence)			↓ (*Bilophila wadsworthia*, *Sutterella stercoricanis*)	↓*Clostridium bartletti*	↓*Bacteroides finegoldii*
Delivery (natural)	↑(Synergistia, Synergistales)		↓ Comamonadaceae	↑ (Catabacteriaceae, Oscillospiraceae, *Dialister invisus*, *Roseburia inulinivorans*) ↓ Acidaminococcaceae	↑*Bacteroides faecis* ↓ Prevotellaceae
pH		↓ (*Actinobacteria*, *Bifidobacteriaceae*, *Bifidobacterium*)	↓ *Herbaspirillum*	↓ (*Coprococcus*, *Eubacterium hallii*)	↓ *(Coprobacter, Bacteroides faecis, Bacteroides stercoris, Bacteroides vulgatus, Parabacteroides merdae*)

Abbreviations: BMI SDS, standardized body mass index; GADA, glutamic acid decarboxylase antibodies; HbA_{1c}, hemoglobinA_{1c}, IA2A, antithyrosine phosphatase antibodies; IAA, anti-insulin autoantibodies.

different descendant species (*B stercoris*, *B intestinalis*, *B cellulosilyticus*, and *B. fragilis*) are among the taxonomic features more abundant in patients. On the contrary, *B vulgatus* showed a higher relative abundance in HD subjects. In addition, we found higher relative abundance of the class *Gammaproteobacteria* and the order *Enterobacteriales* in patients. This might be responsible for intestinal permeability increase (37, 38) and inflammation state; the same microorganisms might impair fasting glucose control in type 2 diabetes patients (39). Interestingly, *B stercoris* was one of the more relevant species observed using 2 machine-learning algorithms (Random Forest and $l_1 l_2$ Elastic Net). A recent report showed that human leukocyte antigen (HLA)-A2–restricted $CD8^+$ T cells autoreactive for pancreatic islets (recognizing the zinc-transporter $8_{186\text{-}194}$) were increased in the pancreas of patients.

$CD8^+$ lymphocytes are activated when short peptide (8-10 amino acids) are presented at the surface of antigen presenting cells (APCs) in the context of HLA class I molecules. The peptide-HLA complex is recognized by the T-cell receptor. Rarely, epitope cross-reactivity, or lack of fine specificity, happens because of cross-conservation of T-cell epitopes between pathogens and autoantigens, as suggested by several examples in multiple sclerosis, Guillain-Barré syndrome, and other pathologies, but at present it is still unclear whether the composition of the intestinal microbiota affects local APC function (40). Recently, $ZnT8_{186\text{-}194}$–reactive $CD8^+$ T cells, preferentially enriched in the pancreas, but not in the blood, of type 1 diabetes donors, were shown to cross-prime an unrelated homologous epitopes (mimotope) of *B stercoris* (41). It is intriguing that these lymphocytes showed cross-reactivity with an epitope from the commensal *B stercoris* strain, a species that we found with higher relative abundance in patients 40-42). As reported, $CD8^+$ T-cell clones displaying the *B stercoris* mimotope had a stronger agonist effect than the native $ZnT8_{186\text{-}194}$ peptide, thus indicating that they can cross-recognize a bacterial mimotope (41). Our findings, together with the presence of bacterial species responsible for an increase of intestinal permeability (see previously), suggest a role for epitope redundancy that might also cause immune response activation and contribute to the development of type 1 diabetes mellitus. On the other hand, in our case series, no significant association between ZnT8 autoantibody positivity and specific microbiota assessment was found. Regardless, these data are not surprising because we evaluated only serum autoantibodies produced by B lymphocytes, whereas Culina et al analyzed CD8+ cytotoxic T cells specific for $ZnT8_{186\text{-}194}$ peptide that resulted in homing into the pancreas (41).

Among *Firmicutes*, *Holdemania* and *H filiformis* were reported as an indicator of impaired lipids and glucose metabolism (39). *Turicibacter* and *T sanguinis* produce lactate, a key regulator of metabolism that has been associated with chronic metabolic diseases and shown to identify in type 2 obese and diabetic mice models (43). Interestingly, in our diabetes patients we found a higher relative abundance of *H filiformis* and, on the contrary, a lower abundance of *T sanguinis* that correlate with impaired metabolic glucose disorders (data not shown).

On the contrary, the genus *Parasutterella*, belonging to the *Betaproteobacteria* phylum, negatively correlated with patients ($P = .004$); among them the *P excrementi hominis* species reached statistical significance ($P = .03$). It has been reported that a high-fat diet regimen was associated with a reduction of *Parasutterella* and *P excrementihominis* (44). *Lactobacillus* genus and *Eubacterium hallii* (both belonging to *Firmicutes*) showed a decrease of relative abundances in patients compared to HDs. The presence of *Lactobacillus* has been shown to attenuate fasting blood glucose, postprandial blood glucose, glucose intolerance, and insulin resistance in type 2 diabetes, thus the decrease of abundance of this genus in patients at onset may indicate a more general protective role for this genus (44, 45).

It is well known that diabetic patients are at higher risk of developing periodontal diseases than controls (46). *Synergistetes* were isolated from subgingival plaque in periodontitis, in root canals from patients suffering endodontic infections, and in cases of dental caries (47). In our report we found the phylum *Synergistetes*, class *Synergistia*, order *Synergistales*, and family *Synergistaceae* all showed a statistically significant higher relative abundance in patients (around 20 times more) compared to HDs. The supervised results also confirmed the importance of this microorganism in defining the patient group. Furthermore, in some patients we defined at the species level the microorganisms belonging to Synergistaceae family and precisely *Pyramidobacter piscolens* (reads identified in the polymorphic regions V2, V3, V4, and V8), *Cloacibacillus porcorum* (V2, V3, V4, V6-7, and V8), and *Cloacibacillus evryensis* (V8 only), stressing the importance of using multiple 16S polymorphic regions for the correct identification of prokaryotic species. Altogether, these data demonstrate the presence of orally derived periodontopathic microorganisms in the fecal microbiota, thus linking periodontitis and systemic disease (48).

Our analysis showed that different pathways associated with glucose metabolism and iron complex levels are impaired in patients with newly diagnosed type 1

diabetes. In addition, histidine metabolism and insulin signaling pathways appear in the list of those associated with patients as well as in a list associated with type 1 diabetes mellitus (49). Moreover, 8 among the 13 pathways that were associated with newly diagnosed patients and that had already been reported in a paper using a direct metaproteomics method were found in the list of the altered pathways observed in our predictive approach (50).

Among the clinical variables, BMI SDS and IAAs were the most significantly correlated across the highest number of phyla. In general we found significant negative correlations among these 2 variables and specific levels of classification of *Synergistetes* (ie, *Synergistia*, *Synegistaceae*), *Actinobacteria* (ie, the phylum and *Bifidobacteriaceae*) and *Proteobacteria* (ie, *Gammaproteobacteria* and *Enterobacteriaceae*) phyla. The most significant correlations we found among others are (i) the positive correlation between glycemia and the cluster *Klebsiella-Coprobacter*, (ii) the negative correlation between HbA_{1c} levels and the cluster *Bilophila-B wadsworthia*, and (iii) the positive correlation between serum pH and the cluster *Bifidobacteriaceae-B bifidum*.

Major strengths of our study include the microbiota detection technique and the same clinical characteristics and dietary habits of the enrolled patients. We are aware that the small sample of cases represents a point of weakness, but on the other hand this is a monocentric study with strict selection criteria. In particular, we evaluated only Caucasian patients with a detailed clinical phenotype, including similar dietary habits and age older than 3 years to avoid different food ingestion and younger age as confounding factors for microbiota composition.

At present it is not possible to clearly state whether gut-microbiota diversity represents a cause or a consequence of autoimmunity in type 1 diabetes.

Further longitudinal studies and increased number of cases will help in obtaining a better knowledge of the role of gut microbiota and autoimmunity and to consider gut-microbiota modulation as a future therapeutic opportunity.

Acknowledgments

We are grateful to Prof Emanuele Bosi, Dr Massimo Locatelli, and Dr. Stefania Del Rosso from the Laboratory Medicine Department, San Raffaele Hospital, Milan, for autoantibodies measurement. We thank Dr Flavia Napoli for English revisions. We are grateful for the Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), University of Genoa, a Department Of Excellence, for the support in the several steps leading to the approval and publishing the Study.

Financial Support: This work was supported by the Fondi Ricerca Corrente. from Italian Ministry of Health granted to Roberto Biassoni.

Author Contributions: R.B. supervised all microbiome characterization, participated in writing the paper, performed all analyses aimed at taxomonic characterization, α and β diversity, as well as the relative abundance and Random Forest analyses. E.D.M. extracted DNA, prepared libraries and sequenced all samples, and participated in writing the paper. M.S. prepared the Biom files, performed data normalization, supervised the $l_1 l_2$ machine learning, performed unsupervised WGCNA clustering and microbiome-associated metabolic pathways (PICRUSt) analysis, and participated in writing the paper. A.B. conducted the supervised and unsupervised analyses, collected clinical and metabolic data. G.P. collected clinical data, followed patients, and helped with manuscript preparation. E.U. helped in sample preparation and sequencing. C.G. selected healthy patient samples. N.M. collected metabolic data and followed patients. G.P. collected clinical data. M.M. revised the manuscript. G.d'A. selected patients, collected clinical and metabolic data, participated in writing the paper, and revised the manuscript.

Additional Information

Correspondence and Reprint Requests: Giuseppe d'Annunzio, MD, Pediatric Clinic, Regional Center for Pediatric Diabetes, IRCCS Istituto Giannina Gaslini, Via Gaslini 5, 16147 Genoa, Italy. E-mail: giuseppedannunzio@gaslini.org.

Disclosure Summary: The authors have nothing to disclose. The authors declare no conflict of interest.

Data Availability: All data generated or analyzed during this study are included in this published article or in the data repositories listed in "References." Raw 16S rRNA gene reads were deposited at the "short read archive" database (BioProject ID SRA: PRJNA525585).

References

1. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. 2014;**383**(9911):69-82.
2. Atkinson MA, Gianani R. The pancreas in human type 1 diabetes: providing new answers to age-old questions. *Curr Opin Endocrinol Diabetes Obes*. 2009;**16**(4):279-285.
3. Eisenbarth GS. Type 1 diabetes: molecular, cellular and clinical immunology. *Adv Exp Med Biol*. 2004;**552**:306-310.
4. Katsarou A, Gudbjörnsdottir S, Rawshani A, et al. Type 1 diabetes mellitus. *Nat Rev Dis Primers*. 2017;**3**:17016.
5. Mayer-Davis EJ, Lawrence JM, Dabelea D, et al; SEARCH for Diabetes in Youth Study. Incidence trends of type 1 and type 2 diabetes among youths, 2002-2012. *N Engl J Med*. 2017;**376**(15):1419-1429.
6. Butalia S, Kaplan GG, Khokhar B, Rabi DM. Environmental risk factors and type 1 diabetes: past, present, and future. *Can J Diabetes*. 2016;**40**(6):586-593.
7. Fronczak CM, Barón AE, Chase HP, et al. In utero dietary exposures and risk of islet autoimmunity in children. *Diabetes Care*. 2003;**26**(12):3237-3242.

8. Versini M, Jeandel PY, Bashi T, Bizzaro G, Blank M, Shoenfeld Y. Unraveling the hygiene hypothesis of helminthes and autoimmunity: origins, pathophysiology, and clinical applications. *BMC Med*. 2015;**13**:81.
9. Barbeau W. What is the key environmental trigger in type 1 diabetes—is it viruses, or wheat, gluten, or both? *Autoimm Rev*. 2012;**12**(2):295-299.
10. Meah FA, DiMeglio LA, Greenbaum CJ, et al; Type 1 Diabetes TrialNet Study Group. The relationship between BMI and insulin resistance and progression from single to multiple autoantibody positivity and type 1 diabetes among TrialNet Pathway to Prevention participants. *Diabetologia*. 2016;**59**(6):1186-1195.
11. Jerram ST, Dang MN, Leslie RD. The role of epigenetics in type 1 diabetes. *Curr Diab Rep*. 2017;**17**(10):89.
12. Hu Y, Wong FS, Wen L. Antibiotics, gut microbiota, environment in early life and type 1 diabetes. *Pharmacol Res*. 2017;**119**:219-226.
13. Sohail MU, Althani A, Anwar H, Rizzi R, Marei HE. Role of the gastrointestinal tract microbiome in the pathophysiology of diabetes mellitus. *J Diabetes Res*. 2017;**2017**:9631435.
14. Qin J, Li R, Raes J, et al; MetaHIT Consortium. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;**464**(7285):59-65.
15. Krych Ł, Nielsen DS, Hansen AK, Hansen CH. Gut microbial markers are associated with diabetes onset, regulatory imbalance, and IFN-γ level in NOD mice. *Gut Microbes*. 2015;**6**(2):101-109.
16. Paun A, Yau C, Danska JS. The influence of the microbiome on type 1 diabetes. *J Immunol*. 2017;**198**(2):590-595.
17. Bibbò S, Dore MP, Pes GM, Delitala G, Delitala AP. Is there a role for gut microbiota in type 1 diabetes pathogenesis? *Ann Med*. 2017;**49**(1):11-22.
18. Vaarala O. Gut microbiota and type 1 diabetes. *Rev Diabet Stud*. 2012;**9**(4):251-259.
19. Vatanen T, Franzosa EA, Schwager R, et al. The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature*. 2018;**562**(7728):589-594.
20. Pellegrini S, Sordi V, Bolla AM, et al. Duodenal mucosa of patients with type 1 diabetes shows distinctive inflammatory profile and microbiota. *J Clin Endocrinol Metab*. 2017;**102**(5):1468-1477.
21. Daft JG, Lorenz RG. Role of the gastrointestinal ecosystem in the development of type 1 diabetes. *Pediatr Diabetes*. 2015;**16**(6):407-418.
22. He C, Shan Y, Song W. Targeting gut microbiota as a possible therapy for diabetes. *Nutr Res*. 2015;**35**(5):361-367.
23. Knip M, Honkanen J. Modulation of type 1 diabetes risk by the intestinal microbiome. *Curr Diab Rep*. 2017;**17**(11):105.
24. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res*. 2017;**45**(W1):W180-W188.
25. De Mol C, Mosci S, Traskine M, Verri A. A regularized method for selecting nested groups of relevant genes from microarray data. *J Comput Biol*. 2009;**16**:1-15.
26. Weiss S, Xu ZZ, Peddada S, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*. 2017;**5**(1):27.
27. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;**4**:17.
28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;**9**:559.
29. Douglas GM, Beiko RG, Langille MGI. Predicting the functional potential of the microbiome from marker genes using PICRUSt. *Methods Mol Biol*. 2018;**1849**:169-177.
30. Biassoni R, Di Marco E, Squillario M, et al. Data from: Gut microbiota in new-onset pediatric patients with type 1 diabetes: machine learning algorithms to classify microorganisms disease-linked. *Figshare*. 2019. Deposited October 16, 2019. Doi: 10.6084/m9.figshare.9989216. https://figshare.com/articles/modified_Online-Only_Supplemental_Material_doc/9989216
31. Sane F, Scuotto A, Pierrat V, Kacet N, Hober D, Romond MB. Diabetes progression and alterations in gut bacterial translocation: prevention by diet supplementation with human milk in NOD mice. *J Nutr Biochem*. 2018;**62**:108-122.
32. Dees SB, Moss CW, Hollis DG, Weaver RE. Chemical characterization of *Flavobacterium odoratum*, *Flavobacterium breve*, and *Flavobacterium*-like groups IIe, IIh, and IIf. *J Clin Microbiol*. 1986;**23**(2):267-273.
33. Lee HJ, Choi JS, Lee HJ, Kim WH, Park SI, Song J. Effect of excess iron on oxidative stress and gluconeogenesis through hepcidin during mitochondrial dysfunction. *J Nutr Biochem*. 2015;**26**(12):1414-1423.
34. Cinek O, Kramna L, Mazankova K, et al. The bacteriome at the onset of type 1 diabetes: a study from four geographically distant African and Asian countries. *Diabetes Res Clin Pract*. 2018;**144**:51-62.
35. Kostic AD, Gevers D, Siljander H, et al; DIABIMMUNE Study Group. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe*. 2015;**17**(2):260-273.
36. Kemppainen KM, Ardissone AN, Davis-Richardson AG, et al; TEDDY Study Group. Early childhood gut microbiomes show strong geographic differences among subjects at high risk for type 1 diabetes. *Diabetes Care*. 2015;**38**(2):329-332.
37. de Goffau MC, Luopajärvi K, Knip M, et al. Fecal microbiota composition differs between children with β-cell autoimmunity and those without. *Diabetes*. 2013;**62**(4):1238-1244.
38. Pedersen C, Ijaz UZ, Gallagher E, et al. Fecal enterobacteriales enrichment is associated with increased in vivo intestinal permeability in humans. *Physiol Rep*. 2018;**6**(7):e13649.
39. Lippert K, Kedenko L, Antonielli L, et al. Gut microbiota dysbiosis associated with glucose metabolism disorders and the metabolic syndrome in older adults. *Benef Microbes*. 2017;**8**(4):545-556.
40. Moise L, Beseme S, Tassone R, et al. T cell epitope redundancy: cross-conservation of the TCR face between pathogens and self and its implications for vaccines and autoimmunity. *Expert Rev Vaccines*. 2016;**15**(5):607-617.
41. Culina S, Lalanne AI, Afonso G, et al. Islet-reactive CD8(+) T cell frequencies in the pancreas but not in blood, distinguish type 1 diabetic patients from healthy donors. *Sci Immunol*. 2018;**3**:eaao4013.
42. Ehlers MR. Who let the dogs out? The ever-present threat of autoreactive T cells. *Sci Immunol*. 2018;**3**(20):eaar6602.
43. Horie M, Miura T, Hirakata S, et al. Comparative analysis of the intestinal flora in type 2 diabetes and nondiabetic mice. *Exp Anim*. 2017;**66**(4):405-416.
44. Kreutzer C, Peters S, Schulte DM, et al. Hypothalamic inflammation in human obesity is mediated by environmental and genetic factors. *Diabetes*. 2017;**66**(9):2407-2415.
45. Wang G, Li X, Zhao J, Zhang H, Chen W. *Lactobacillus casei* CCFM419 attenuates type 2 diabetes via a gut microbiota dependent mechanism. *Food Funct*. 2017;**8**(9):3155-3164.
46. Sun KT, Chen SC, Lin CL, et al. The association between type 1 diabetes mellitus and periodontal diseases. *J Formos Med Assoc*. 2019;**118**(6):1047-1054.
47. Vartoukian SR, Palmer RM, Wade WG. Diversity and morphology of members of the phylum "synergistetes" in periodontal health and disease. *Appl Environ Microbiol*. 2009;**75**(11):3777-3786.
48. Nakajima M, Arimatsu K, Kato T, et al. Oral administration of *P. gingivalis* induces dysbiosis of gut microbiota and impaired barrier function leading to dissemination of enterobacteria to the liver. *PloS One*. 2015;**10**(7):e0134234.
49. Luan M, Shang Z, Teng Y, et al. The shared and specific mechanism of four autoimmune diseases. *Oncotarget*. 2017;**8**(65):108355-108374.
50. Gavin PG, Mullaney JA, Loo D, et al. Intestinal metaproteomics reveals host-microbiota interactions in subjects at risk for type 1 diabetes. *Diabetes Care*. 2018;**41**(10):2178-2186.