Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Special Issue: *Annals Reports*
ORIGINAL ARTICLE

Microbial and metabolic multi-omic correlations in systemic sclerosis patients

Chiara Bellocchi,[1] Álvaro Fernández-Ochoa,[2] Gaia Montanelli,[1] Barbara Vigone,[1] Alessandro Santaniello,[1] Christian Milani,[3] Rosa Quirantes-Piné,[2] Isabel Borrás-Linares,[2] Marco Ventura,[3] Antonio Segura-Carrettero,[2] Marta Eugenia Alarcón-Riquelme,[4,5,a] and Lorenzo Beretta [1,a]

[1]Scleroderma Unit, Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico di Milano, Milan, Italy. [2]Department of Analytical Chemistry, University of Granada, Granada, Spain. [3]Laboratory of Probiogenomics, Department of Life Sciences, University of Parma, Parma, Italy. [4]Centre for Genomics and Oncological Research (GENYO), Pfizer-University of Granada-Andalusian Regional Government, Granada, Spain. [5]Institute for Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

Address for correspondence: Lorenzo Beretta, Scleroderma Unit, Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico di Milano, Via Pace 9, 20122 Milan, Italy.
lorberimm@hotmail.com

Intestinal microbiota has been associated with systemic autoimmune diseases, yet the functional consequences of these associations are elusive. We characterized the fecal microbiota (16S rRNA gene amplification and sequencing) and the plasma metabolome (high-performance liquid chromatography coupled to mass spectrometry) in 59 patients with systemic sclerosis (SSc) and 28 healthy controls (HCs). Microbial and metabolic data were cross-correlated to find meaningful associations after extensive data mining analysis and internal validation. Our data show that a reduced model of nine bacteria is capable of differentiating HCs from SSc patients. SSc gut microbiota is characterized by a reduction in protective butyrate-producing bacteria and by an increase in proinflammatory noxious genera, especially *Desulfovibrio*. From the metabolic point of view, a multivariate model with 17 metabolite intermediates well distinguished cases from controls. The most interesting peaks we found were identified as glycerophospholipid metabolites and benzene derivatives. The microbial and metabolic data showed significant interactions between *Desulfovibrio* and alpha-*N*-phenylacetyl-L-glutamine and 2,4-dinitrobenzenesulfonic acid. Our data suggest that in SSc, intestinal microbiota is characterized by proinflammatory alterations subtly entwined with the metabolic state. *Desulfovibrio* is a relevant actor in gut dysbiosis that may promote intestinal damage and influence amino acid metabolism.

Keywords: systemic sclerosis; microbiota; metabolome; gastrointestinal

Introduction

Systemic sclerosis (SSc) is an autoimmune disease characterized by vascular damage, widespread fibrosis, and immune system activation. The gastrointestinal tract (GIT) is affected in nearly 90% of patients[1] as a consequence of motor disturbances[2] secondary to myenteric neuropathy, muscle atrophy, and fibrosis.[3] GIT involvement presents as a variety of signs and symptoms such as bloating, diarrhea, and constipation. Intestinal bacterial overgrowth and dysbiosis may occur in up to 75% of patients, and this may be related to dysmotility.[4] Dysbiosis is a modification of microbiota with relevant immunological and metabolic consequences. Gut microbiota interacts with the intestinal immune system playing a fundamental role in maintaining immune tolerance,[5–7] exerts antimicrobial functions through signaling with pattern recognition receptors contributing to the development of an intact barrier

[a]Both the authors contributed equally.

doi: 10.1111/nyas.13736

function,[8] and regulates T lymphocytes (T_{reg} cells)[9] and plasma cell production of IgA.[10]

Several studies focusing on microbiota in autoimmune diseases have found distinct profiles in patients with systemic lupus erythematosus (SLE),[11,12] inflammatory arthritis, and spondyloarthritis.[13,14] A few studies conducted in relatively small case-series have studied microbiota in SSc.[15–17] A first study found a distinct colonic microbial signature on cecum and sigmoid mucosal lavage samples in 17 SSc patients compared with healthy controls (HCs).[16] Here, microbial genera and species elsewhere related to inflammatory bowel disease (IBD) were found to be associated with the severity of GIT symptoms. Similar results were found in stool sample analysis when comparing the same patients with 17 patients from Norway.[15] Recent work on 18 Italian SSc patients showed specific microbial differences between SSc patients with or without gastroenteric involvement.[17] Despite the encouraging results described by these reports, a detailed description of intestinal microbiota in larger case-series is lacking, thus challenging the overall generalizability of previous findings. Moreover, the functional consequences of microbial alterations in SSc are still poorly understood.

Next-generation omics approaches have revolutionized the study of microbiota, making feasible the study of the whole intestinal microbial ecology in a high-resolution and culture-independent manner.[18] Sequence-based taxonomic profiling of the microbiome can be integrated with other multi-omics data as for instance, transcriptomics, proteomics, and metabolomics, to fully understand the functional activity of microbial community.[18] Nonetheless, the integration of multi-omics data has mostly been dampened by bioinformatic challenges. Metabolomic is of great interest for functionally characterizing intestinal microbiota: metabolic pathways and shared metabolites are fundamental in the crosstalk among microorganisms, or between microorganisms and the host. Moreover, the metabolome is partially the result of interaction between the body and the microbiota;[19] for example, gut microbiota may alter the concentrations of specific metabolites that lead to the development of cardiac diseases.[20]

The present study was conducted in order to (1) provide a detailed description of intestinal microbiota in the largest case-series of SSc patients analyzed so far (59 cases) in comparison with geographically matched HCs (29 cases); (2) describe the most relevant metabolic alteration occurring in this very same population, and (3) determine to what extent metabolite intermediates are correlated to dysbiosis. Overall, we aim to provide insight into microbial alterations and into the functional consequences of these alterations in SSc.

Patients and methods

Sixty consecutive SSc patients referring to our institution and fulfilling the 2013 ACR/EULAR criteria[21] were included. Exclusion criteria included the presence of hepatitis B or C infection, nephrotic syndrome with proteinuria >3.5 g/day, the presence of overlap syndromes with other systemic autoimmune diseases, and/or the recent (<1 year) use of depletive therapies; all the patients had to be on stable doses of steroids or immunosuppressants for the last 3 months. Patients were categorized into the limited (lc-SSc) or into the diffuse cutaneous (dc-SSc) subset;[22] patients with definite SSc[21] without skin fibrosis yet with puffy fingers were categorized as lc-SSc. Interstitial lung disease (ILD) was defined as the presence of an involvement of lung parenchyma >5% on high-resolution computed tomography accompanied by a reduced forced vital capacity <80% of predicted values and/or a reduced diffusing capacity for carbon monoxide (DLco) < 80% of predicted values.[23]

All SSc patients completed the Italian version of the UCLA GIT 2.0 questionnaire[24,25] to evaluate the severity of self-reported GIT symptoms. The questionnaire provides an overall evaluation of the whole GIT and includes seven domains: reflux, distension/bloating, soilage, diarrhea, social function, emotional well-being, and constipation. The scores from the first six domains were averaged, while constipation was evaluated separately. GIT scores range, from 0 (the least severe involvement) to 3 (the most severe involvement), was deemed appropriate to evaluate the activity and severity of GIT involvement in SSc. To focus on enteric problems, scores from the intestinal domains (distension/bloating, diarrhea, and constipation) were averaged; an arbitrary cutoff point equal to the 50th percentile of the scores distribution was used to discriminate patients with (SSc_{Int}) or without intestinal involvement (SSc_{Ctrl}).

Thirty age- and sex-matched HCs were included as a comparison group. All the HCs came from the

same geographical area as the SSc patients (Milan metropolitan area). HCs were enriched for nonrelatives kindred of the SSc patients.

All patients and controls were asked to complete a 24-item questionnaire to record their dietary habits before the collection of stool samples. The results from the questionnaire were used to determine if short-term dietary modifications may influence gut composition, as hypothesized by some authors,[26] or may have a negligible effect compared to intersubject variability, as suggested by others.[27]

None of the study participants was treated with antibiotics or probiotics the month before stool collection. Fecal samples were frozen at −20 °C until analysis. Within 24 h of stool collection, plasma samples were obtained and frozen at –80 °C until analysis.

The multi-omic analysis is ancillary to the Precisesads project (www.precisesads.eu) that was approved by the local ethic committee (*comitato etico* Area B); written consent was obtained from each participant.

Microbiota identification by 16S rRNA gene amplification and sequencing

DNA was extracted from each stool sample using the QIAamp DNA Stool Mini kit following the manufacturer's instructions (Qiagen).

Partial 16S rRNA gene sequences were amplified from extracted DNA using primer pair Probio_Uni/Probio_Rev, which targets the V3 region of the 16S rRNA gene sequence.[28] 16S rRNA gene amplification and amplicon checks were carried out as previously described.[28] Notably, the primer pair Probio_Uni/Probio_Rev has been specifically developed to maximize coverage and amplification performance of gut bacterial populations.[28]

16S rRNA gene sequencing was performed using a MiSeq Illumina[28] at the DNA sequencing facility of GenProbio srl (www.genprobio.com) according to the protocol previously reported. Following sequencing and demultiplexing, the reads of each sample were filtered to remove low quality and polyclonal sequences and data were exported as .fastq files. The .fastq files were processed using a custom script based on the QIIME software suite.[29] Paired-end reads were assembled to reconstruct the complete Probio_Uni/Probio_Rev amplicons. Quality control (QC) retained those sequences with a 140–400 bp length and mean sequence quality score >20, while sequences with homopolymers >7 bp and mismatched primers were omitted. Chimeric sequences were removed with ChimeraSlayer included in the Qiime software suite. To calculate downstream diversity measures, 16S rRNA operational taxonomic units (OTUs) were defined at ≥97% sequence homology using uclust,[30] and OTUs with <10 sequences were filtered. All reads were classified to the lowest possible taxonomic rank using QIIME[29] and the SILVA database v. 119 clustered at 97% identity as reference dataset.[31]

Metabolomic analysis of plasma samples

Protein content was removed from plasma samples with methanol:ethanol (50:50, v/v).[32] Metabolomics analyses were performed using an Agilent 1260 HPLC instrument coupled to an Agilent 6540 Ultra High Definition (UHD) Accurate Mass Q-TOF equipped with a Jet Stream dual ESI interface. The compounds were separated using a reversed-phase C18 analytical column and a gradient of mobile phases A, water containing 0.1% of formic acid, and B, methanol. Detection was performed in positive-ion mode over a range from 100 to 1700 m/z.

A QC sample, prepared by mixing equal volumes from each sample, was analyzed throughout the sequencing to control for analytical reproducibility. An MS/MS analysis of this sample was performed to facilitate the identification of potential biomarkers.

Once the analysis was performed, features were extracted by means of MassHunter Profinder software (B.06.00, Agilent) with the Recursive Feature Extraction method for small molecules.[33] Peaks were filtered with intensity threshold at 1000 counts. $[M+H]^+$, $[M+Na]^+$, and $[M-H_2O]$ were the considered species. Feature alignment parameters were ± 0.25 min and 40 ppm ± 4 mDa for retention time and mass windows, respectively.

The integrated areas of each feature were normalized by the sum of the total areas of all features from the nearest QC in order to correct the instrumental variability produced throughout the analysis. F features with high variability (CV >30%) were filtered.

Metabolite identification was carried out through the comparison of accurate mass, isotopic distribution, and fragmentation patterns obtained

in MS/MS analysis with online metabolomic databases.

Data analysis

Bioinformatic analysis was restricted to 59 cases and 28 HCs because one SSc sample and two HC samples did not pass the quality test for metabolomic analysis. Microbiome quality filtering results are provided in the Supplementary File S1 (online only).

Canonical analysis of microbiota

Biodiversity of the samples (alpha diversity) was calculated with Chao1 and Shannon indexes using 10 subsampling points for a maximum of 54,890 sequences in order to generate rarefaction curves. Similarities between samples (beta diversity) were calculated by unweighted uniFrac.[34] PCoA representations of beta diversity were performed using QIIME.[29] For the other analyses, bacteria at the genus or species taxonomic rank, with an overall abundance >1%, were considered. To identify differentially represented bacterial populations, the linear discriminant analysis (LDA) effect size (LEfSe) algorithm[35] was used (https://huttenhower.sph.harvard.edu/galaxy/).

Data mining analyses

To model nonlinear multivariate interactions among variables, several inductive data mining algorithms were used. Details about the data mining procedures are reported in Supplementary Figure S1 (online only). In brief, the procedure was carried out to select a reduced subset of variable capable of explaining the differences between groups (i.e., SSc versus HCs; and SSc_{Int} versus SSc_{ctrl}). To this end, the variables were selected to maximize a classification function: the area under receiver operating characteristics curve (AUROC). The AUROC measures the overall discrimination of a classification algorithm, where 1 represents a perfect test and 0.5 a test doing no better than random choice. The study pipeline included a nested cross-validation phase (model selection and data filtering) and an extensive cross-validation phase (to assess the robustness and capability of generalization of the selected model).

Microbiomic and metabolomic cross-correlation

The main aim of the study was to determine whether there were relevant correlations between microbiomic and metabolomic data. The construction of a simple correlation matrix between the two omics would yield too many entries that would not withstand statistical correction. To tackle this issue, the following procedure was devised. Microbiomic and metabolomic models were built taking into account a reduced number of features as described in the data mining section. We refer to these models as "models conditioned to the class" because they maximize the capability of discriminating cases and controls. Selected microbial genera and selected metabolites conditioned to the class were correlated by means of Spearman's ρ. Statistical correction was calculated via a 10,000-fold multivariate permutation test for correlated data.[36] The microbial $\times$ metabolic cross-correlation matrix is visualized via heat maps and plotted via the Orange data mining suite (http://orange.biolab.si/).

Descriptive statistics

Differences between study groups were tested via Student's *t*-test after adjustment (p_c) via a permutation-based step-down *Pmin* procedure[37] to account for multiple testings; cubic root transformation was applied to data before analysis. The scikit-learn algorithm (http://scikit-learn.org/stable/index.html) was used for all the analysis,[38] along with custom python codes implemented by L.B.

Nucleotide sequence accession numbers

The raw 16S rRNA gene amplicon sequences reported in this article have been deposited in the NCBI Short Read Archive (SRA) (PRJNA437384).

Results

Clinical characteristics

Clinical and demographical characteristics of the study participants who passed the QCs (59 SSc and 28 HCs) are reported in Table 1. Overall, the population was mostly composed of females ($n = 52$, 88.1%) and lc-SSc patients ($n = 43$, 72.8%). The average GIT score from SSc patients was 0.48 ± 0.45, while the average score of intestinal domains was 0.475 ± 0.42. Nineteen (32.2%) patients had ILD; the presence of ILD was the most common motivation for the current use of immunosuppressants (10 out of 12 cases, 83.3%), followed by arthritis (remaining two cases, 16.7%). More details about SSc patients are reported in Supplementary Table S1 (online only), including a

Table 1. Clinical characteristics of study groups

Variable	HCs ($n = 28$)	SSc ($n = 59$)	SSc_{Ctrl} ($n = 33$)	SSc_{Int} ($n = 26$)
Age, years	49.3 ± 12.9	56.5 ± 12.7	53.7 ± 13	60 ± 11.9
Females, *n* (%)	22 (78.6)	52 (88.1)	30 (90.9)	22 (84.6)
BMI	25.3 ± 4.2	22.6 ± 3.7	22.07 ± 3.3	23.36 ± 4.04
Disease duration, years	–	12.5 ± 10.4	11.4 ± 10.4	11.8 ± 10.2
Subsets, *n* (%)				
Definite SSc	–	6 (10.1)	0 (0)	6 (22.8)
lc-SSc	–	43 (72.8)	26 (88.8)	17 (65.4)
dc-SSc	–	10 (17.1)	7 (21.2)	3 (11.8)
Autoantibody, *n* (%)				
ANA	–	56 (94.9)	31 (93.9)	25 (96.1)
ACA	–	23 (39)	8 (24.2)	15 (57.7)
Topo I	–	23 (39)	18 (54.6)	5 (19.2)
FVC, % predicted	–	100.1 ± 21.1	98.2 ± 22.6	99.2 ± 23.1
DLco, % predicted	–	73.1 ± 21.8	73.4 ± 22.5	72.7 ± 22.2
ILD, *n* (%)	–	19 (32.2)	13 (39.4)	6 (23.1)
Therapy, *n* (%)				
Prednisone >5 mg/day	–	19 (32.2)	15 (45.5)	4 (15.4)
Immunosuppressants	–	12 (20.3)	7 (21.2)	5 (19.2)
Biologicals	–	9 (15.2)	6 (18.3)	3 (11.4)
GIT total	–	0.48 ± 0.45	0.24 ± 0.32	0.78 ± 0.41

Note: Values expressed as mean ± standard deviation.
HCs, healthy controls; SSc, systemic sclerosis; SSc_{Ctrl}, SSc without intestinal involvement; SSc_{Int}, SSc with intestinal involvement; BMI, body mass index; lc-SSc, limited cutaneous SSc; dc-SSc, diffuse cutaneous SSc, ANAs, antinuclear antibodies; ACAs, anticentromere antibodies; Topo I, antitopoisomerase I antibodies; FVC, forced vital capacity; DLco, diffusing capacity for carbon monoxide; ILDs, interstitial lung diseases; GIT, gastrointestinal tract (UCLA GIT questionnaire).

thorough description of concurrent therapies and GIT subscores. Details about comorbidities in HCs are given in Supplementary Table S2 (online only); overall, only essential hypertension was found in donors ($n = 4$, 14.3%). Comorbidities were those captured in the Precisesads study; no information about other comorbidities is available.

Microbic differences between SSc and HCs

Canonical analysis of the microbiota between SSc and HCs did not provide significant results; the alpha diversity of SSc patients was not different from that of HCs when measured either with the Chao1 index or via the Shannon index. Analysis of beta diversity based on uniFrac distances and PCoA also could not discriminate patients and controls (Supplementary Fig. S2, online only).

The analysis of the microbiota via nonlinear algorithms could sort out different models to explain microbial differences between SSc and HCs. The overall microbial community at the species rank (AUROC = 0.696 ± 0.86) and to a greater extent the genus rank (AUROC = 0.706 ± 0.023) could distinguish SSc patients from HCs. Filtering and data reduction could not sort out a satisfactory model to classify cases from controls at the species level (AUROC = 0.665 ± 0.044). On the contrary, a simplified model comprising just nine microbial genera (out of 88 available) was capable of discriminating SSc from HCs with a fairly good performance and an overall AUROC = 0.711 ± 0.042 after extensive internal validation; this model was thus selected as a reference, and further analyses were conducted at the genus level. The relative abundances of these nine genera in SSc patients and HCs are reported in Supplementary Table S3 (online only). LDA univariate analysis confirmed that a few bacterial species included in the model were over-represented (*Parabacterioides*, unidentified members of the Firmicutes phylum, *Butyricimonas*, and *Desulfovibrio*) or under-represented (*Turicibacter* and unidentified members of the Lachnospiraceae family) in SSc patients (LDA |log score| >2; see Fig. 1).

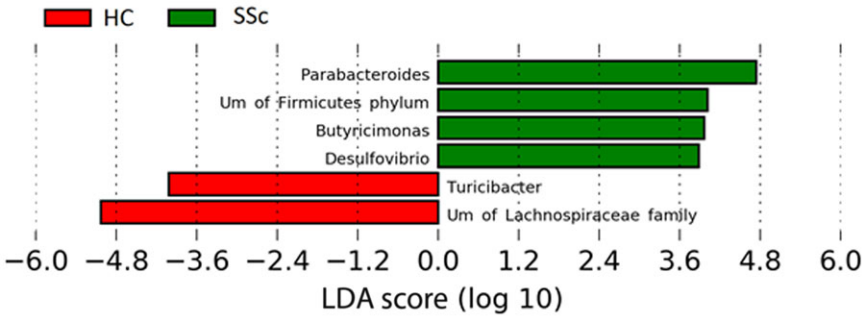

Figure 1. LDA analysis in patients and controls. Linear discriminant analysis (LDA) of bacteria (genus rank) selected in the best classification model. HCs, healthy controls ($n = 28$); SSc, systemic sclerosis ($n = 59$). Significant results with $\log_{10}$ scores >2 are represented.

To rule out the possibility that short-term dietary modification may have accounted for differences between groups, we checked available alimentary diaries from 19 HCs and from 53 SSc participants who returned them correctly compiled. Only the unidentified members of the Erysipelotrichaceae family were related to short-term dietary modifications with an explained variance (R^2) equal to 12.8% ± 6.3%. The abundance of this taxon was, however, equal between HCs and SSc (0.23% ± 0.31% versus 0.43% ± 0.99%; $P = 0.66$). Overall, these results suggest that short-term dietary modification of gut composition was modest compared with intersubject variability, and do not account for HC versus SSc differences (detailed information about 24-h dietary habits in the study groups can be found in Supplementary Table S4, online only).

Plasma metabolic differences between SSc and HCs

Before metabolomic analysis, the desmethyldehydronifedipine, the hydroxydehydro-nifedipine carboxylate, and the omeprazole peaks were excluded from the analysis because the vast majority of our SSc subjects were treated with dihydropyridine calcium channel blockers (84.7%) and proton-pump inhibitors (88.1%); 426 peaks were thus retained for further analysis. A simple model of just 17 metabolites was capable of differentiating SSc patients and HCs, with an AUROC = 0.744 ± 0.029. The normalized peak areas of the 17 metabolites included into the model are reported in Supplementary Table S4 (online only).

Correlations between microbic and metabolic data

The cross-correlation matrix between the nine bacteria and the 17 metabolites that best explained the differences between SSc and HCs is reported in Figure 2. The strongest signals were found between unidentified Firmicutes and alpha-*N*-phenylacetyl-L-glutamine ($\rho = 0.363$, $p_c = 0.083$) and 2,4-dinitrobenzenesulfonic acid ($\rho = 0.393$, $p_c = 0.024$) or between *desulfovibrio* and alpha-*N*-phenylacetyl-L-glutamine ($\rho = 0.389$, $p_c = 0.03$) and 2,4-dinitrobenzenesulfonic acid ($\rho = 0.39$, $p_c = 0.029$). Alpha-*N*-phenylacetyl-L-glutamine and 2,4-dinitrobenzenesulfonic acid were highly correlated ($\rho = 0.79$, $p = 1.1 \times 10^{-19}$). Ps peak areas of both the metabolites were increased in SSc patients as compared to HCs. Alpha-*N*-phenylacetyl-L-glutamine, HCs: 0.357 ± 0.337 versus SSc: 0.837 ± 0.679; $p_c = 7.8 \times 10^{-4}$. 2,4-dinitrobenzenesulfonic acid, HCs: 0.461 ± 0.311 versus SSc: 0.863 ± 0.619; $p_c = 0.003$.

Effect of gut composition on GIT symptoms

Based on the severity of symptoms, as assessed by the intestinal domains of the UCLA GIT 2.0 questionnaire, we were able to classify 26 patients as SSc_{Int} and 33 patients as SSc_{Ctrl}; the threshold to discriminate the two populations was equal to 0.5. The clinical characteristics of the categorized patients are summarized in Table 1. SSc_{Int} patients had an increased prevalence of anticentromere antibodies (SSc_{Int} = 15/26 versus SSc_{Ctrl} = 8/33; $P = 0.015$) and were less likely to use steroids (SSc_{Int} = 4/26 versus SSc_{Ctrl} = 15/33; $P = 0.0237$) compared with SSc_{Ctrl} subjects. SSc_{Int} patients had a reduced

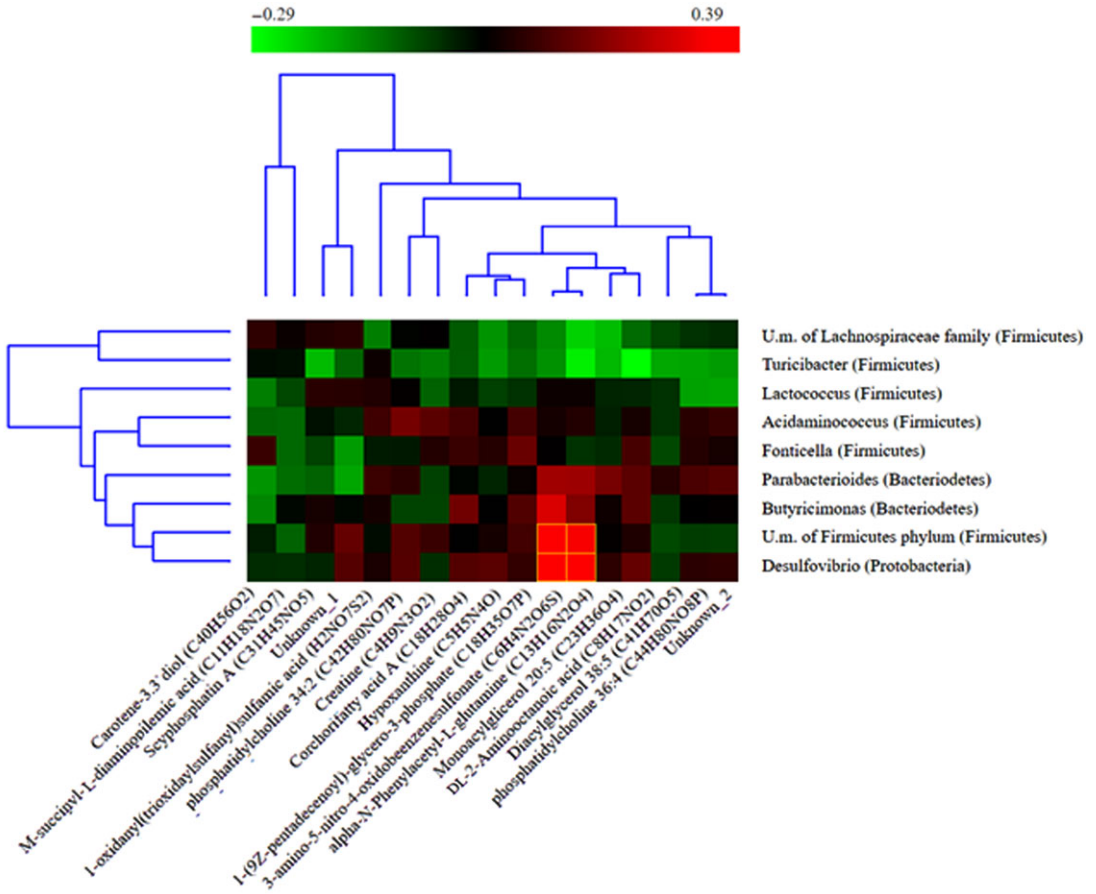

Figure 2. Cross correlation between microbiomic and metabolomic data. Spearman's correlation coefficients between selected bacteria and metabolites. The nine bacteria and the 17 metabolites were selected via bioinformatic analysis and best classified cases versus controls in multivariate models. Individual correlations significant at the 0.05 threshold after 100,000-fold permutation testing for correlated data are highlighted in yellow.

number (richness) and distribution (evenness) of taxa compared to SSc_{Ctrl} patients (Fig. 3).

Overall, 10 genera could be used to model differences among HCs, SSc_{Ctrl}, and SSc_{Int} The multivariate model based on these bacteria had a weighted AUROC $= 0.679 \pm 0.021$. The following pairwise results were observed: HCs versus SSc_{Ctrl}, AUROC $= 0.703 \pm 0.056$; HCs versus SSc_{Int}, AUROC $= 0.64 \pm 0.038$; SSc_{Ctrl} versus SSc_{Int}, AUROC $= 0.689 \pm 0.043$. Within this model, significant LDA scores were observed for *Desulfovibrio* and *Turicibacter* (Fig. 4A). The relative abundance of *Desulfovibrio* in the study groups is illustrated in Figure 4B.

SSc_{Ctrl} and SSc_{Int} patients could not be well differentiated from a metabolic point of view; and no single peak or complex interaction model was associated with either group.

Effect of therapy on gut microbiota

As reported in Table 1, at the time of sample collection, 19 SSc patients (32.2%) were treated with prednisone >5 mg/day, 12 (20.3%) with immunosuppressant, and 9 (15.2%) with biologics. We investigated the effect of any of these therapy on microbiota composition. Beta diversity of patients receiving one or two treatments was different compared to that of patients receiving no treatment as illustrated in Supplementary Figure S3 (online only). Data mining analysis could not sort out adequate model to classify the two subsets of patients.

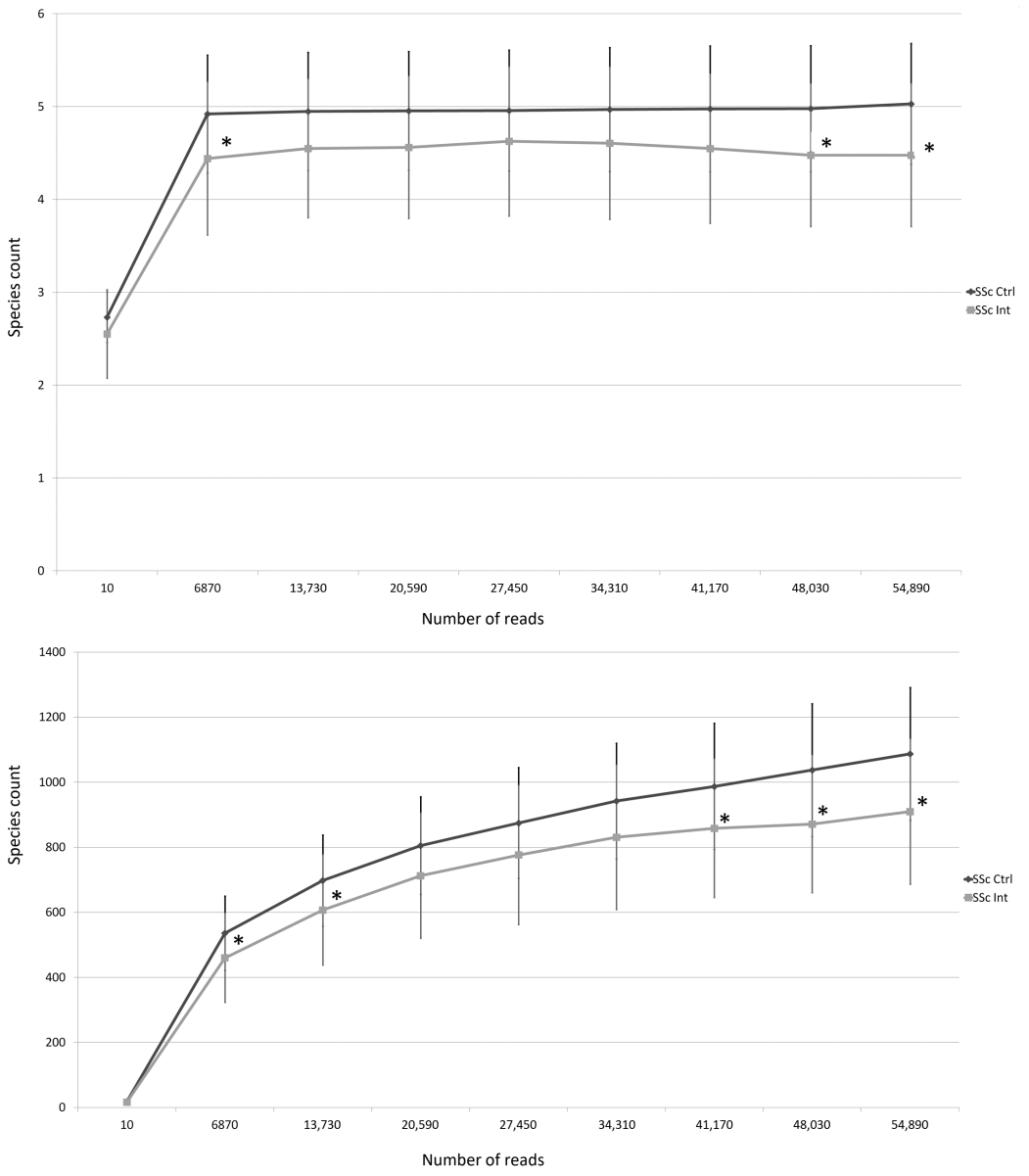

Figure 3. **Alpha diversity in relation to intestinal involvement. Microbial diversity of gut composition in systemic sclerosis (SSc) patients with mild/no self-reported intestinal involvement (SSc Ctrl) or with moderate-to-severe self-reported intestinal symptoms (SSc Int) according to the Shannon index (top panel) or the Chao1 index (bottom panel). Low values indicate a reduced richness and distribution of taxa. *, $P < 0.05$.**

Discussion

We determined the microbiological characteristics and the metabolic profile of a large case-series of SSc patients, demonstrating that (1) Italian SSc patients have a unique mirobial proinflammatory colonic profile; (2) these patients are characterized by peculiar metabolic characteristics; (3) and there are relevant correlations between microbial and metabolomic "fingerprints" conditioned to the disease status.

Among the genera differentially expressed in SSc patients and HCs, nine bacteria were jointly

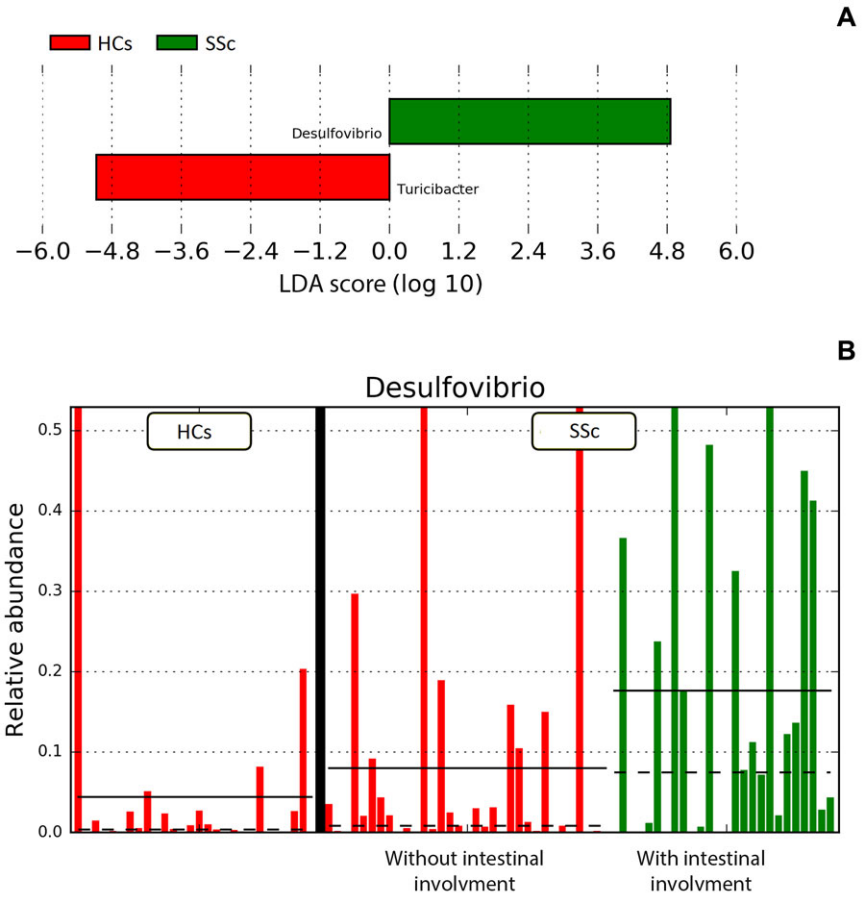

Figure 4. LDA analysis in relation to the intestinal status. Panel (A) shows linear discriminant analysis (LDA) of bacteria (genus rank) selected in the best classification model capable of jointly discriminating healthy controls (HCs) and systemic sclerosis (SSc) patients with or without intestinal involvement; significant results with $\log_{10}$ scores >2 are represented. Panel (B) shows relative abundance of *Desulfovibrio* in the classification model according to the disease status; each bar represents a case; dotted line, median; and solid line, mean.

associated with the disease status. These findings emphasize the multifaceted relationships among bacteria in a complex nonlinear interaction model; however, a number of genera were singularly either over-represented (Fonticella, *Parabacterioides*, unidentified members of the Firmicutes phylum, *Butyricimonas*, and *Desulfovibrio*) or under-represented (*Turicibacter* and unidentified members of the Lachnospiraceae family) in SSc patients.

These findings cannot be directly compared with those presented by Volkmann *et al.*[16] because of the different sample size, the different severity of GIT symptoms, the different clinical picture of enrolled subjects, and the different dietary and lifestyle habits of Italy and U.S. residents. Food intake is a major driver of gut microbiota,[26] whose ecological composition may differ in different geographical regions[39] because of phenotypic variations as well as diet and lifestyle.[40] Differences in gut microbiota were indeed observed in SSc when the same U.S. population described in Ref. 16 was compared to a similar sample of Norwegian patients.[15] Another source of variability with respect to the study by Volkmann *et al.*[16] is the sampling methodology: we analyzed stool samples, while the Volkmann *et al.* analyzed colonic lavage. Each collection method has advantages and drawbacks: some authors claim that colonic lavage is more representative of the actual gut microbial ecology (as determined by biopsy) with respect to stool samples;[41] while others suggest

that bowel cleansing before sample collection may alter the gut microbiota.[42] Despite these considerations, and regardless of the specific findings, our results are in full accordance with other studies (see Refs. 15 and 16), suggesting the presence of a proinflammatory microbial flora in SSc patients, which may have functional consequences on the intestinal integrity.

More specifically, Lachnospiraceae were described as reduced in IBD,[43] in untreated multiple sclerosis subjects,[44] or in SLE patients.[11] Lachnospiraceae is a family of butyrate-producing bacteria, and butyrate exerts anti-inflammatory and antioxidant effects on the colonic mucosa.[45] Similarly, *Turicibacter,* decreased in SSc patients, are protective in models of inflammatory ileitis[46] and in colitis-resistant CD8 T cell–deficient mice,[47] and its abundance decreases after a proinflammatory high-fat diet.[48] In contrast, *Desulfovibrio* promotes ileitis and colonic inflammation in animal models[46,48] and is implicated in IBD.[49,50] Notably, in our samples, the abundance of *Desulfovibrio* was related to the severity of intestinal symptoms and its relative abundance linearly increased from HCs to SSc patients with none/mild intestinal manifestations, to those with overt self-reported intestinal symptoms (Fig. 4). Differences in gut composition between SSc patients with or without intestinal involvement were previously analyzed in a relatively small case-series of Italian patients (18 subjects, of whom half were without enteric involvement) recruited from a different geographical region.[17] As in our study, these authors described a reduced alpha diversity in SSc_{Int} subjects and a different microbial status in relation to intestinal involvement, even if they were able to capture more differences than we found. The difference in findings could be related to sample size and other methodological differences between the studies. In particular, in Ref. 17, controls were defined as the complete absence of self-reported symptoms, while in our work a threshold equal to 0.5 was used to discriminate cases from controls, and indeed this threshold has been validated to distinguish patients with none/mild enteric symptoms from those with moderate-to-severe/very severe complains.[51]

As far as the metabolic findings are concerned, in the multivariate model that characterized the profile of our patients, a few hits were increased and individually associated with SSc: the DL-2-aminooctanoic acid, glycerophospholipid metabolites diacylglycerol 38:5, 1-(9Z-pentadecenoyl)-glycero-3-phosphate, phosphatidylcholine 36:4, and two benzene derivatives, 2,4-dinitrobenzenesulfonic acid and alpha-*N*-phenylacetyl-L-glutamine. The significance of such associations is elusive; it is unknown if they have a causative role in SSc progression or if they reflect the general perturbations underlying SSc pathogenesis. Glycerophospholipids are bioactive molecules that may regulate different cellular pathways, including apoptosis and inflammation;[52] but the source of these compounds in scleroderma patients is unclear.

The correlation of benzene derivatives and *Desulfovibrio* merits further discussion. The association between *Desulfovibrio* and phenylacetylglutamine (alpha-*N*-phenylacetyl-L-glutamine) has not been described before. As discussed previously, *Desulfovibrio* proliferation characterizes a dysbiotic microbiota; in this environment, pathogenic bacteria compete with commensal bacteria.[53] Commensal bacteria possess the capability to catabolize phenyl acetate,[54] and their relative scarcity would promote an excess of substrate and favor phenylacetylglutamine formation. Overall, our results suggest that SSc-related dysbiosis may have metabolic consequences influencing amino-acid metabolism, whose alterations have been already suggested in SSc patients.[55] In our patients, we observed increased amounts of a sulfonate benzene compound (2,4-dinitrobenzenesulfonic acid), whose levels strongly correlate with phenylacetylglutamine, raising the possibility that some of the events that lead to phenylacetylglutamine accumulation might be relevant to 2,4-dinitrobenzenesulfonic acid biosynthesis. Of interest, this sulfone may act as a substrate for *Desulfovibrio* anaerobic respiration (a process called dissimilatory sulfate reduction), further promoting the production of hydrogen sulfide to cause intestinal damage and inflammation.[49]

Some limitations of our study should be acknowledged. First, cross-sectional samples of gut microbiota have an intrinsic variability and do not reflect the dynamic changes of microbial flora in time or in response to exogenous stimuli. Therefore, it is not possible to determine to what extent dysbiosis in SSc is persistent or, rather, if it waxes and wanes in response to therapy or in relation to the severity of GIT symptoms.

Second, no experiment was carried out to correlate microbial alterations with colonic transit time, even if previous work suggests that dysbiosis is the consequence of altered intestinal motility in SSc gut.[4,56] The observation made elsewhere that *Desulfovibrio* abundance positively correlates with colonic transit time[57] and that colonic transit time is increased in subjects with high GIT scores[58] largely supports our findings.

Last, the findings we describe are the results of complex analyses that rely on a local search optimization. This metaheuristic does not guarantee the discovery of an optimal solution, and other potential microbial/metabolic associations in SSc may have been overlooked. In spite of this, we believe that the framework we used could fruitfully be applied in multi-omics research to isolate meaningful signals and to tackle the "curse of dimensionality" problem.

In summary, we have shown that SSc patients have peculiar microbial and metabolic characteristics, and that these are subtly entwined. *Desulfovibrio* is a relevant factor in SSc-related gut dysbiosis in that it may promote intestinal damage and influence amino-acid metabolism. Further studies are needed to characterize these associations in response to therapy and during the disease course of affected individuals.

Acknowledgment

This work was supported by EU/EFPIA/Innovative Medicines Initiative Joint Undertaking PRECISESADS grant No. 115565.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig S1. Pipeline for data analysis.

Fig S2. Beta diversity between SSc patients and HCs.

Fig S3. Beta diversity in relation to therapy.

Table S1. Additional information on systemic sclerosis patients.

Table S2. Comorbidities in healthy controls.

Table S3. Relative abundance of selected taxa.

Table S4. Dietary habits within 24 h from stool collection of the study participants who returned the questionnaires.

Table S5. Selected metabolite concentrations in HCs and SSc.

File S1. Spreadsheet of microbiome quality filtering results.

Competing interests

The authors declare no competing interests.

References

1. Ebert, E.C. 2008. Gastric and enteric involvement in progressive systemic sclerosis. *J. Clin. Gastroenterol.* **42:** 5–12.
2. Henry, M.A., M.C. Harbermann & O.M. Rocha. 1999. Esophageal motor disturbances in progressive systemic sclerosis. *Dis. Esophagus* **12:** 51–53.
3. Sjogren, R.W. 1994. Gastrointestinal motility disorders in scleroderma. *Arthritis Rheum.* **37:** 1265–1282.
4. Andréasson, K., Z. Alrawi, A. Persson, *et al.* 2016. Intestinal dysbiosis is common in systemic sclerosis and associated with gastrointestinal and extraintestinal features of disease. *Arthritis Res. Ther.* **18:** 278.
5. Jandhyala, S.M., R. Talukdar, C. Subramanyam, *et al.* 2015. Role of the normal gut microbiota. *World J. Gastroenterol.* **21:** 8787–8803.
6. Ivanov, I.I. & K. Honda. 2012. Intestinal commensal microbes as immune modulators. *Cell Host Microbe* **12:** 496–508.
7. Palm, N.W., M.R. de Zoete & R.A. Flavell. 2015. Immune–microbiota interactions in health and disease. *Clin. Immunol.* **159:** 122–127.
8. Lundin, A., C.M. Bok, L. Aronsson, *et al.* 2008. Gut flora, Toll-like receptors and nuclear receptors: a tripartite communication that tunes innate immunity in large intestine. *Cell Microbiol.* **10:** 1093–1103.
9. Smith, P.M., M.R. Howitt, N. Panikov, *et al.* 2013. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* **341:** 569–573.
10. Suzuki, K., M. Maruya, S. Kawamoto, *et al.* 2010. The sensing of environmental stimuli by follicular dendritic cells promotes immunoglobulin A generation in the gut. *Immunity* **33:** 71–83.
11. Hevia, A., C. Milani, P. López, *et al.* 2014. Intestinal dysbiosis associated with systemic lupus erythematosus. *MBio* **5:** e01548-14.
12. López, P., B. de Paz, J. Rodríguez-Carrio, *et al.* 2016. Th17 responses and natural IgM antibodies are related to gut microbiota composition in systemic lupus erythematosus patients. *Sci. Rep.* **6:** 24072.
13. Gill, T., M. Asquith, J.T. Rosenbaum, *et al.* 2015. The intestinal microbiome in spondyloarthritis. *Curr. Opin. Rheumatol.* **27:** 319–325.
14. Scher, J.U., D.R. Littman & S.B. Abramson. 2016. Microbiome in inflammatory arthritis and human rheumatic diseases. *Arthritis Rheumatol.* **68:** 35–45.
15. Volkmann, E.R., A.-M. Hoffmann-Vold, Y.-L. Chang, *et al.* 2017. Systemic sclerosis is associated with specific alterations in gastrointestinal microbiota in two independent cohorts. *BMJ Open Gastroenterol.* **4:** e000134.

16. Volkmann, E.R., Y.-L. Chang, N. Barroso, *et al.* 2016. Association of systemic sclerosis with a unique colonic microbial consortium. *Arthritis Rheumatol.* **68:** 1483–1492.
17. Patrone, V., E. Puglisi, M. Cardinali, *et al.* 2017. Gut microbiota profile in systemic sclerosis patients with and without clinical evidence of gastrointestinal involvement. *Sci. Rep.* **7:** 14874.
18. Franzosa, E.A., T. Hsu, A. Sirota-Madi, *et al.* 2015. Sequencing and beyond: integrating molecular "omics" for microbial community profiling. *Nat. Rev. Microbiol.* **13:** 360–372.
19. Nicholson, J.K., E. Holmes, J. Kinross, *et al.* 2012. Host–gut microbiota metabolic interactions. *Science* **336:** 1262–1267.
20. Wang, Z., E. Klipfell, B.J. Bennett, *et al.* 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472:** 57–63.
21. van den Hoogen, F., D. Khanna, J. Fransen, *et al.* 2013. 2013 Classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism Collaborative Initiative. *Arthritis Rheum.* **65:** 2737–2747.
22. LeRoy, E.C., C. Black, R. Fleischmajer, *et al.* 1988. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J. Rheumatol.* **15:** 202–205.
23. Vigone, B., A. Santaniello, M. Marchini, *et al.* 2015. Role of class II human leucocyte antigens in the progression from early to definite systemic sclerosis. *Rheumatology (Oxford)* **54:** 707–711.
24. Khanna, D., R.D. Hays, P. Maranian, *et al.* 2009. Reliability and validity of the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. *Arthritis Rheum.* **61:** 1257–1263.
25. Gualtierotti, R., F. Ingegnoli, R. Two, *et al.* Reliability and validity of the Italian version of the UCLA Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument in patients with systemic sclerosis. *Clin. Exp. Rheumatol.* **33:** S55–S60.
26. David, L.A., C.F. Maurice, R.N. Carmody, *et al.* 2013. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505:** 559–563.
27. Graf, D., R. Di Cagno, F. Fåk, *et al.* 2015. Contribution of diet to the composition of the human gut microbiota. *Microb. Ecol. Health Dis.* **26:** 26164.
28. Milani, C., A. Hevia, E. Foroni, *et al.* 2013. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS One* **8:** e68739.
29. Caporaso, J.G., J. Kuczynski, J. Stombaugh, *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7:** 335–336.
30. Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460–2461.
31. Quast, C., E. Pruesse, P. Yilmaz, *et al.* 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41:** D590–D596.
32. Bruce, S.J., I. Tavazzi, V. Parisod, *et al.* 2009. Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Anal. Chem.* **81:** 3285–3296.
33. Kitagawa, N., S.M. Fischer, J. Roark & M.S.T. Samant. 2013. Novel two-pass feature extraction workflow for the statistical profiling of mass spectrometric data. In *MP379, Proceedings of the 61st ASMS Conference on Mass Spectrometry and Allied Topics*, Minneapolis, Minnesota.
34. Lozupone, C. & R. Knight. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71:** 8228–8235.
35. Segata, N., J. Izard, L. Waldron, *et al.* 2011. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12:** R60.
36. Yoder, P.J., J.U. Blackford, N.G. Waller, *et al.* 2004. Enhancing power while controlling family-wise error: an illustration of the issues using electrocortical studies. *J. Clin. Exp. Neuropsychol.* **26:** 320–331.
37. Westfall, P.H. & S.S. Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment.* John Wiley & Sons, Inc.
38. Pedregosa, F., A. Gramfort, V. Michel, *et al.* 2011. Scikit-learn: machine learning in Python Gaël Varoquaux. *J. Mach. Learn. Res.* **12:** 2825–2830.
39. Suzuki, T.A. & M. Worobey. 2014. Geographical variation of human gut microbial composition. *Biol. Lett.* **10:** 20131037.
40. Conlon, M. & A. Bird. 2014. The impact of diet and lifestyle on gut microbiota and human health. *Nutrients* **7:** 17–44.
41. Watt, E., M.R. Gemmell, S. Berry, *et al.* 2016. Extending colonic mucosal microbiome analysis—assessment of colonic lavage as a proxy for endoscopic colonic biopsies. *Microbiome* **4:** 61.
42. Harrell, L., Y. Wang, D. Antonopoulos, *et al.* 2012. Standard colonic lavage alters the natural state of mucosal-associated microbiota in the human colon. *PLoS One* **7:** e32545.
43. Kang, S., S.E. Denman, M. Morrison, *et al.* 2010. Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm. Bowel Dis.* **16:** 2034–2042.
44. Jangi, S., R. Gandhi, L.M. Cox, *et al.* 2016. Alterations of the human gut microbiome in multiple sclerosis. *Nat. Commun.* **7:** 12015.
45. Hamer, H.M., D. Jonkers, K. Venema, *et al.* 2008. Review article: the role of butyrate on colonic function. *Aliment. Pharmacol. Ther.* **27:** 104–119.
46. Werner, T., S.J. Wagner, I. Martínez, *et al.* 2011. Depletion of luminal iron alters the gut microbiota and prevents Crohn's disease-like ileitis. *Gut* **60:** 325–333.
47. Presley, L.L., B. Wei, J. Braun, *et al.* 2010. Bacteria associated with immunoregulatory cells in mice. *Appl. Environ. Microbiol.* **76:** 936–941.
48. Everard, A., V. Lazarevic, N. Gaïa, *et al.* 2014. Microbiome of prebiotic-treated mice reveals novel targets involved in host response during obesity. *ISME J.* **8:** 2116–2130.
49. Singh, S.B. & H.C. Lin. 2015. Hydrogen sulfide in physiology and diseases of the digestive tract. *Microorganisms* **3:** 866–889.
50. Jia, W., R.N. Whitehead, L. Griffiths, *et al.* 2012. Diversity and distribution of sulphate-reducing bacteria in human faeces from healthy subjects and patients with inflammatory bowel disease. *FEMS Immunol. Med. Microbiol.* **65:** 55–68.

51. Khanna, D., V. Nagaraja, H. Gladue, *et al.* 2013. Measuring response in the gastrointestinal tract in systemic sclerosis. *Curr. Opin. Rheumatol.* **25:** 700–706.
52. Huang, C. & C. Freter. 2015. Lipid metabolism, apoptosis and cancer therapy. *Int. J. Mol. Sci.* **16:** 924–949.
53. Bäumler, A.J. & V. Sperandio. 2016. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535:** 85–93.
54. Teufel, R., V. Mascaraque, W. Ismail, *et al.* 2010. Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc. Natl. Acad. Sci. USA* **107:** 14390–14395.
55. Bengtsson, A.A., J. Trygg, D.M. Wuttge, *et al.* 2016. Metabolic profiling of systemic lupus erythematosus and comparison with primary Sjögren's syndrome and systemic sclerosis. *PLoS One* **11:** e0159384.
56. Basilisco, G., R. Barbera, M. Vanoli, *et al.* 1993. Anorectal dysfunction and delayed colonic transit in patients with progressive systemic sclerosis. *Dig. Dis. Sci.* **38:** 1525–1529.
57. Roager, H.M., L.B.S. Hansen, M.I. Bahl, *et al.* 2016. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat. Microbiol.* **1:** 16093.
58. Sharma, S.K., M.B. Adarsh, S.K. Sinha, *et al.* 2017. The gastrointestinal dysmotility and infections in systemic sclerosis—a real world scenario. *Curr. Rheumatol. Rev.* **13**. https://doi.org/10.2174/1573397113666170425145405.