

ORIGINAL ARTICLE

## Microbiological and bioinformatics analysis of primary Sjogren's syndrome patients with normal salivation<sup>§</sup>

Huma Siddiqui<sup>1</sup>, Tsute Chen<sup>2</sup>, Ardita Aliko<sup>3</sup>, Piotr M Mydel<sup>3,4</sup>, Roland Jonsson<sup>3,5</sup> and Ingar Olsen<sup>1\*</sup>

<sup>1</sup>Department of Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway; <sup>2</sup>Department of Microbiology, Forsyth Institute, Cambridge, MA, USA; <sup>3</sup>Broegelmann Research Laboratory, Department of Clinical Science, University of Bergen, Bergen, Norway; <sup>4</sup>Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology, and Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland; <sup>5</sup>Department of Rheumatology, Haukeland University Hospital, Bergen, Norway

**Background:** Reduced salivation is considered a major clinical feature of most but not all cases of primary Sjögren's syndrome (pSS). Reduced saliva flow may lead to changes in the salivary microbiota. These changes have mainly been studied with culture that typically recovers only 65% of the bacteria present.

**Objective:** This study was to use high throughput sequencing, covering both cultivated and not-yet-cultivated bacteria, to assess the bacterial microbiota of whole saliva in pSS patients with normal salivation.

**Methods:** Bacteria of whole unstimulated saliva from nine pSS patients with normal salivation flow and from nine healthy controls were examined by high throughput sequencing of the hypervariable region V1V2 of 16S rRNA using the 454 GS Junior system. Raw sequence reads were subjected to a species-level, reference-based taxonomy assignment pipeline specially designed for studying the human oral microbial community. Each of the sequence reads was BLASTN-searched against a database consisting of reference sequences representing 1,156 oral and 12,013 non-oral species. Unassigned reads were then screened for high-quality non-chimeras and subjected to *de novo* species-level operational taxonomy unit (OTU) calling for potential novel species. Downstream analyses, including alpha and beta diversities, were analyzed using the Quantitative Insights into Microbial Ecology (QIIME) pipeline. To reveal significant differences between the microbiota of control saliva and Sjögren's saliva, a statistical method introduced in Metastats [www.metastats.cbcb.umd.edu](http://www.metastats.cbcb.umd.edu) was used.

**Results:** Saliva of pSS patients with normal salivation had a significantly higher frequency of Firmicutes compared with controls ( $p = 0.004$ ). Two other major phyla, Synergistetes and Spirochaetes, were significantly depleted in pSS ( $p = 0.001$  for both). In addition, we saw a nearly 17% decrease in the number of genera in pSS (25 vs. 30). While *Prevotella* was almost equally abundant in both groups (25% in pSS and 22% in controls), about a twofold increase in pSS of *Streptococcus* (28% vs. 17%) and *Veillonella* (26% vs. 12%) was detected. *Prevotella melaninogenica* was the major species in controls (13%) while *Veillonella atypica* and the *Veillonella parvula* groups dominated in patient samples (14 and 14%). The scarcity in bacterial species in pSS compared with controls was also demonstrated by alpha and beta diversity analyses, as well as read abundance depicted in a phylogenetic tree.

**Conclusions:** While Firmicutes was significantly higher in pSS patients than in controls, Synergistetes and Spirochaetes were significantly lower. The number of bacterial genera and species was also lower. These data showed that microbial dysbiosis is another key characteristic of pSS whole saliva which can occur independent of hyposalivation.

**Keywords:** primary Sjögren's syndrome; whole saliva; normal salivary flow rate; bacteria; high throughput sequencing; pyrosequencing

Responsible Editor: Ann Progulsk-Fox, University of Florida, United States.

\*Correspondence to: Ingar Olsen, Department of Oral Microbiology, Faculty of Dentistry, University of Oslo, P.O. Box 1052 Blindern, NO-0316 Oslo, Norway, Email: [ingar.olsen@odont.uio.no](mailto:ingar.olsen@odont.uio.no)

To access the supplementary material to this article, please see [Supplementary files](#) under 'Article Tools'.

Received: 25 January 2016; Revised: 20 July 2016; Accepted: 6 September 2016; Published: 20 October 2016

<sup>§</sup>Presented in part at the 13th International Symposium on Sjögren's Syndrome, Bergen, Norway, 20.05.2015 – 22.05.2015.

Primary Sjögren's syndrome (pSS) is an autoimmune, chronic inflammatory disease of unknown etiology characterized by progressive infiltration of mononuclear cells into salivary and lacrimal glands, resulting in impairment of gland function. It is also associated with lymphocytic infiltrates and immunological hyperactivity. The most commonly affected glands are the salivary and lacrimal glands. This causes typical dryness in the mouth and keratoconjunctivitis sicca in the eyes (1). Atkinson et al. (2) found that 88% of Sjögren's syndrome (SS) subjects had reduced salivary flow rate (hyposalivation), and subjective complaints of dry mouth (xerostomia) were reported in 75–92% of such patients (3). pSS occurs when gland inflammation causes both dry mouth and dry eyes and is not associated with another connective tissue disease, while secondary SS is associated with connective tissue diseases such as systemic lupus erythematosus, scleroderma, or rheumatoid arthritis (RA).

Sjögren's syndrome is most prevalent in patients of advanced age, especially in those aged 50 years and older, and women are affected more frequently than men (4). With hypofunction in the salivary glands, the homeostasis of the oral microbiota is changed. This may increase the chances for developing microbial diseases, such as dental caries and oral candidiasis. Parvinen and Larmas (5) found that hyposalivation helped streptococci, actinomyces, and lactobacilli thrive in the oral cavity. In other studies, hyposalivation was reported to be followed by high counts of *Lactobacillus* and *Candida* species (6, 7). In a study of the oral ecology in patients with severe Sjögren's syndrome, *Streptococcus salivarius*, *Neisseria pharyngis*, *Veillonella* species, and *Micrococcus mucilaginosus* were significantly reduced compared with controls while the number of *Staphylococcus aureus* and *Candida* species was significantly increased (8). Thus, previous literature has shown that hyposalivation can affect the composition of the microbiota in the oral cavity, but it is unclear whether a shift in the oral microbiota occurs in pSS patients with normal salivation.

It is noteworthy that previous studies assessed the microbiota with culture, occasionally with selective media and/or with commercial tests for select organisms. It is generally recognized that only 65% of the bacteria in the oral cavity can be recovered by culture.

The aim of the present study was to characterize the bacterial profile in whole saliva of pSS patients with a normal salivary flow rate by high throughput sequencing. This technique recovers both cultivated and not-yet-cultivated bacteria thus giving an in-depth overview of bacteria present.

## Methods

### Sampling and sample processing

Whole unstimulated saliva was collected from nine pSS patients (age 45–79 years) consisting of eight females and one male, and from nine healthy female controls (age 39–68 years). They all had a normal salivation rate of >1.5 ml in 15 min. All patients fulfilled the revised American European Consensus Group criteria for classification of pSS (9). DNA was extracted from the samples (200 µl volume) using the MasterPure™ DNA Purification kit (Epicentre, Illumina Company, Madison, WI) and the final DNA was dissolved in 1 × TE buffer. The 16S rRNA hypervariable region V1V2 was sequenced on a 454 GS Junior system (Roche, Branford, CT) using the primers (9) listed in Table 1. Molecular identifier (MID) tags, 10-mer, were used as sample identifiers and are listed in Supplementary Table 1. Amplification reactions were performed as described by Siddiqui et al. (10), with minor modifications as follows: the cycling program was reduced to 30 cycles and triplicate PCRs were performed for each sample. All PCR products were pooled and purified using Agencourt AMPure PCR purification (Beckman Coulter, Brea, CA). DNA quality and concentration were assessed with Bioanalyzer 2100 (Agilent, Santa Clara, CA) and Nanodrop 3300 Fluorespectrometer (Thermo Scientific, Wilmington, DE).

Table 1. PCR primers used in this study

| Primer                                     | Sequence (5' > 3')   | 16S rDNA region            | Product size        | Reference |
|--|--|----------------------------|---------------------|-----------|
| LibL/A <sup>a</sup> -MID <sup>b</sup> -V1F | CCATCTCATCCCTGCGTGTCTCCGACTCAG -MID-<br>AGAGTTTGATCMTGGCTCAG | V1V2<br>8-361 <sup>c</sup> | 423 bp <sup>d</sup> | 9         |
| LibL/B <sup>b</sup> -V2R                   | CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-<br>CYNACTGCTGCCTCCCGTAG      |                            |                     |           |

<sup>a</sup>LibL/A and LibL/B primers correspond to 454-adaptor sequences as described in protocol for GS Junior. [www.454.com/downloads/my454/documentation/gs-junior/system-wide-documents/GSJunior\\_GuidelinesforAmpliconExperimentalDesign.pdf](http://www.454.com/downloads/my454/documentation/gs-junior/system-wide-documents/GSJunior_GuidelinesforAmpliconExperimentalDesign.pdf), p21.

<sup>b</sup>Molecular identifier tags; MID sequences were adapted from Eurofins Genomics [www.eurofinsgenomics.eu/media/892639/how-to-order-amplicon-sequencing-primers-guide.pdf](http://www.eurofinsgenomics.eu/media/892639/how-to-order-amplicon-sequencing-primers-guide.pdf)

<sup>c</sup>Coordinates are given relative to the 1,542 bp *Escherichia coli* K12 16S rDNA sequence.

<sup>d</sup>Product size includes the primer sequences.

### Bioinformatics analysis of sequence reads

High throughput sequencing was performed following the protocol for unidirectional amplicon sequencing with the GS Junior Titanium emPCR (Lib-L) and Sequencing kit (Roche Diagnostics GmbH Mannheim, Germany), which resulted in 106,614 raw reads. Sequence data generated in this study were submitted to the European Nucleotide Archive with the accession number PRJEB12522 ([www.ebi.ac.uk/ena/data/view/PRJEB12522](http://www.ebi.ac.uk/ena/data/view/PRJEB12522)). Processing of the sequencing data and taxonomy assignment were performed with an algorithm modified, based on the one described by Al-Hebshi et al. (11). To maximize the assignment rate, raw reads were used directly without quality filtering. Reads were first assigned with sample IDs based on the MID sequences and then BLASTN-searched against a combined set of 16S rRNA reference sequences that consist of the HOMDEXTGG set published by Al-Hebshi et al. (11), and the NCBI 16S rRNA reference sequence set (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/16SMicrobial.tar.gz>). These combined, well-curated and near full-length reference sequences represented a total of 1,151 oral and 12,013 non-oral microbial species. The NCBI BLASTN version 2.2.28+ (12) was used with the following parameters: -penalty -5-reward 4-gapopen 5-gapextend 5. Reads with  $\geq 98\%$  sequence identity to the matched reference and at least 90% alignment length (i.e.  $\geq 90\%$  of the read length that was aligned to the reference was used to calculate the sequence percent identity) were classified based on the taxonomy of the reference sequence with the highest sequence identity. If a read matched with reference sequences representing multiple species with equal best scores (i.e. equal percent identity and alignment length), it was subject to chimera checking. Non-chimeric reads with multi-species best hits were considered valid and were assigned as a different species with multiple species names. Unassigned reads (i.e.  $\leq 98\%$  identity or  $\leq 90\%$  alignment length) were pooled together and subject to the *de novo* chimera checking and sequence quality screening using the USEARCH program, version v8.1.1861 (13). The *de novo* chimera checking was done using 98% as the sequence identity cutoff. The sequence quality screening was done with the 'expected error rate'  $\leq 1$  (13) and length  $\geq 200$  bases. Non-chimeric, high-quality unassigned reads were then subject to species-level *de novo* operational taxonomy unit (OTU) calling with 98% as the sequence identity cutoff using USEARCH. Representative reads from each of the OTUs/species were BLASTN-searched against the same reference sequence set again to determine the closest species for these potential novel species. The taxonomy information used in this study is available at [ftp://www.homd.org/publication\\_data/20150519/](ftp://www.homd.org/publication_data/20150519/) for HOM-DEXTGG and <ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy> for the NCBI sequences. All assigned reads were subject to several downstream bioinformatics analyses, including alpha and beta diversity assessments, provided in the

QIIME (Quantitative Insights into Microbial Ecology) software package version 1.9.1 (14). The phylogenetic tree required for constructing the UniFrac-based matrices used in some of the beta diversity analyses was built dynamically from reference sequences with matched reads. The reference sequences were aligned with the software MAFFT version 7.149b (15) prior to tree construction using the QIIME treeing script. Downstream analyses were done for a range of minimal read count (MC) per OTU/species: 1, 2, 5, 10, 50, and 100 separately. All read counts presented in this paper are based on MC 100 analysis.

To reveal significant differences between the microbiota of control saliva and Sjögren's saliva, a statistical method introduced in Metastats [www.metastats.cbc.umd.edu](http://www.metastats.cbc.umd.edu) was used. This method employs a false discovery rate to improve specificity in high complexity environments, and in addition handles sparsely sampled features using Fisher's exact test (16). A *p*-value of  $\leq 0.05$  was considered significant. Only taxa (at different levels) with significant differences in abundance are listed in Table 2. For diversity analysis and variations, QIIME was used (14).

The study was approved by REC West (Regional Committee for Medical and Health Research Ethics), Norway, ref. no. 2009/686 under the title: Translational studies in SS and other rheumatic disorders.

## Results

### Sequence data

Of the 106,614 raw sequence reads obtained from pyrosequencing, 93,458 were successfully assigned with a valid sample ID based on the barcode sequences. Of these, 68,020 matched references with  $\geq 98\%$  identity and  $\geq 90\%$  alignment length, 65,618 matched a single species, and 2,402 multiple species. Of the 2,402 reads that matched multiple species, 1,160 were determined to be chimera and were discarded. The remaining 25,438 (93,458–68,020) unmatched reads were subject to chimera checking and sequence quality screening, and 18,181 (71.48%) were chimeras and 5,758 failed quality screening and were discarded. The final 1,499 good-quality unassigned reads were clustered into 406 species-level OTUs by USEARCH, of which 391 OTUs had BLASTN hits (1,455 reads) and 15 had no hits (44 reads). A summary of this assignment sequencing processing statistics is listed in Table 3. The 68,315 assigned reads belonged to 680 species-level taxonomic units and were subject to a variety of analyses provided in QIIME at various MC cutoffs (1, 2, 5, 10, 50, and 100). The complete QIIME results are available for download at [ftp://www.homd.org/publication\\_data/20160412](ftp://www.homd.org/publication_data/20160412).

### Composition of the salivary microbiota of pSS with normal salivation

Using MC100 as the minimal read count cutoffs, six different phyla were detected in pSS patients, with DNA

**Table 2.** Significant ( $p \leq 0.05$ ) differences in abundance of taxa from control saliva and Sjögren's saliva as estimated by Metastats [www.metastats.cbc.umd.edu](http://www.metastats.cbc.umd.edu)

| Taxon  | Control saliva (n =9) | Sjögren's saliva (n =9) | Metastat p-value |
|--|-----------------------|-------------------------|------------------|
| <b>Phyla</b>   |                       |                         |                  |
| Increased in pSS   |                       |                         |                  |
| Firmicutes   | 0.37366 ± 0.02802     | 0.56124 ± 0.03774       | 0.004            |
| Decreased in pSS   |                       |                         |                  |
| Synergistetes  | 0.00601 ± 0.00525     | 0.00000                 | 0.001            |
| Spirochaetes   | 0.00667 ± 0.00667     | 0.00000                 | 0.001            |
| <b>Genera</b>  |                       |                         |                  |
| Increased in pSS   |                       |                         |                  |
| <i>Streptococcus</i>   | 0.18015 ± 0.03078     | 0.32095 ± 0.05361       | 0.037            |
| Decreased in pSS   |                       |                         |                  |
| <i>Treponema</i>   | 0.00068 ± 0.00068     | 0.00000                 | 0.001            |
| <i>Peptostreptococcaceae</i> _[XIII][G-1]                    | 0.01829 ± 0.01605     | 0.00000                 | 0.001            |
| <i>Bacteroidaceae</i> _[G-1]                                 | 0.00844 ± 0.00775     | 0.00000                 | 0.001            |
| <i>Moryella</i>  | 0.00397 ± 0.00197     | 0.00000                 | 0.001            |
| <i>Fretibacterium</i>  | 0.00601 ± 0.00524     | 0.00000                 | 0.001            |
| <i>Porphyromonas</i>   | 0.08057 ± 0.02712     | 0.01176 ± 0.00430       | 0.012            |
| <i>Tannerella</i>  | 0.04373 ± 0.02111     | 0.00094 ± 0.00046       | 0.020            |
| <i>Catonella</i>   | 0.00475 ± 0.00034     | 0.00027 ± 0.00021       | 0.032            |
| <b>Species</b>   |                       |                         |                  |
| Increased in pSS   |                       |                         |                  |
| <i>Veillonella</i> sp._Oral_Taxon_917                        | 0.00000               | 0.00353 ± 0.00228       | 0.001            |
| Decreased in pSS   |                       |                         |                  |
| <i>Treponema</i> sp._Oral_Taxon_237                          | 0.00667 ± 0.00667     | 0.00000                 | 0.001            |
| <i>Porphyromonas gingivalis</i> _nov_90.93%                  | 0.00280 ± 0.00152     | 0.00000                 | 0.001            |
| <i>Peptostreptococcaceae</i> _[XIII][G-1] sp._Oral_Taxon_113 | 0.01830 ± 0.01605     | 0.00000                 | 0.001            |
| <i>Bacteroidaceae</i> _[G-1] sp._Oral_Taxon_272              | 0.00844 ± 0.00775     | 0.00000                 | 0.001            |
| <i>Moryella</i> sp._Oral_Taxon_373                           | 0.00397 ± 0.00197     | 0.00000                 | 0.001            |
| <i>Fretibacterium multispecies_spp23_2</i>                   | 0.00601 ± 0.00524     | 0.00000                 | 0.001            |
| <i>Prevotella nanceiensis</i>                                | 0.00326 ± 0.00137     | 0.00023 ± 0.00019       | 0.012            |
| <i>Tannerella forsythia</i>                                  | 0.04373 ± 0.02111     | 0.00094 ± 0.00046       | 0.017            |
| <i>Catonella morbi</i>                                       | 0.00475 ± 0.00020     | 0.00027 ± 0.00021       | 0.024            |
| <i>Fusobacterium periodonticum</i>                           | 0.00335 ± 0.01045     | 0.00936 ± 0.00360       | 0.030            |
| <i>Porphyromonas gingivalis</i>                              | 0.05746 ± 0.02815     | 0.00071 ± 0.00046       | 0.038            |
| <i>Prevotella pallens</i>                                    | 0.02479 ± 0.00966     | 0.00375 ± 0.00121       | 0.050            |

Phylum, order and genus levels are listed with mean abundance ± standard error.

sequences predominantly assigned to Firmicutes (57%), Bacteroidetes (26%), Proteobacteria (8%), Actinobacteria (8%), and Fusobacteria (2%). In Fig. 1, a comparative taxonomic view of 16S rDNA sequences is given from bacteria in pSS saliva and control saliva assigned to the phylum level. Saliva from SS displayed a significantly higher frequency of Firmicutes compared with control saliva ( $p = 0.004$ ) while Synergistetes and Spirochaetes were significantly lower than in controls ( $p = 0.001$ , for both, see Table 2).

Twenty-five bacterial genera were detected in SS. Relative abundance of the 18 major genera identified in the sequence pool of nine pSS saliva samples and nine

control saliva samples is shown in Fig. 2. There was a 17% reduction in the numbers of genera identified in SS compared with those of controls (25 genera vs. 30). While *Prevotella* was almost equally abundant in controls and in SS (25 and 22%), an about twofold increase in *Streptococcus* (28% vs. 17%) and *Veillonella* (26% vs. 12%) was observed in pSS patients.

Metastats analysis showed that the genus *Streptococcus* ( $p = 0.037$ ) was significantly higher in pSS patients than in controls (Table 2), while eight other taxa were significantly higher in controls than in pSS patients. Among these were *Peptostreptococcaceae*\_[XIII] [G-1], *Bacteroidaceae*\_[G-1], and *Moryella* ( $p = 0.001$ ).

Table 3. Summary statistics of read assignment

| Category                           | Minimal read count (MC) per species |        |        |        |        |        |
|------------------------------------|-------------------------------------|--------|--------|--------|--------|--------|
|                                    | MC1 (All)                           | MC2    | MC5    | MC10   | MC50   | MC100  |
| Total read counts                  | 93,458                              | 93,458 | 93,458 | 93,458 | 93,458 | 93,458 |
| Reads assigned ( $\geq$ MC)        | 68,315                              | 68,165 | 67,668 | 67,056 | 63,848 | 61,186 |
| Reads unassigned ( $<$ MC)         | 0                                   | 150    | 647    | 1,259  | 4,467  | 7,129  |
| Number of unique species           | 413                                 | 341    | 251    | 203    | 100    | 70     |
| Number of multi-species            | 81                                  | 58     | 40     | 23     | 6      | 2      |
| Number of novel species            | 186                                 | 131    | 59     | 29     | 6      | 2      |
| Reads assigned to unique species   | 65,618                              | 65,546 | 65,306 | 64,994 | 62,607 | 60,507 |
| Reads assigned to multiple species | 1,242                               | 1,219  | 1,164  | 1,052  | 677    | 383    |
| Reads assigned to novel species    | 1,455                               | 1,400  | 1,198  | 1,010  | 564    | 296    |
| Total chimera                      | 19,341                              | 19,341 | 19,341 | 19,341 | 19,341 | 19,341 |
| Reads without BLASTN hits          | 44                                  | 44     | 44     | 44     | 44     | 44     |
| Low-quality reads removed          | 5,758                               | 5,758  | 5,758  | 5,758  | 5,758  | 5,758  |

<sup>a</sup>Total assigned reads is the sum of reads assigned to unique, multiple, and novel species.

In pSS patients, the most abundant species were the *Veillonella parvula* group, *Veillonella atypica*, *Prevotella melaninogenica*, and *Prevotella histicola*. However, *Veillonella* sp.\_Oral\_Taxon\_917 was the only significantly increased species in pSS in contrast to controls ( $p = 0.001$ ). Totally, there was a reduction in the number of species detected in SS compared with controls. *V. atypica* and the *V. parvula* groups were most predominant in the patient samples (14 and 14%). It was also noteworthy that *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema* sp.\_Oral\_Taxon\_237 were significantly more prevalent in

the control group than in the pSS group ( $p = 0.038$ , 0.017, and 0.001, respectively, see Table 2).

A total of 339 species were identified to be absent among the 1,151 known oral species from the Human Oral Microbiome Database, [www.homd.org/](http://www.homd.org/) (May 2016). However only 13 species had a combined read count of  $\geq 100$  across all samples and several of them only differed from the known oral species by naming discrepancy. The two species with most abundant reads were *Streptococcus* sp.\_str\_C300 (1,809 reads) and *Prevotella jejuni* (1,565). *P. jejuni* appeared to be more abundant in the pSS samples. A complete list of the new oral species and novel species is provided in Supplementary Table 2.

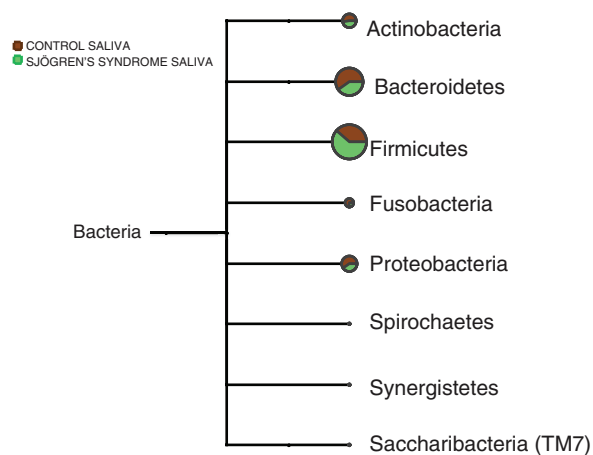
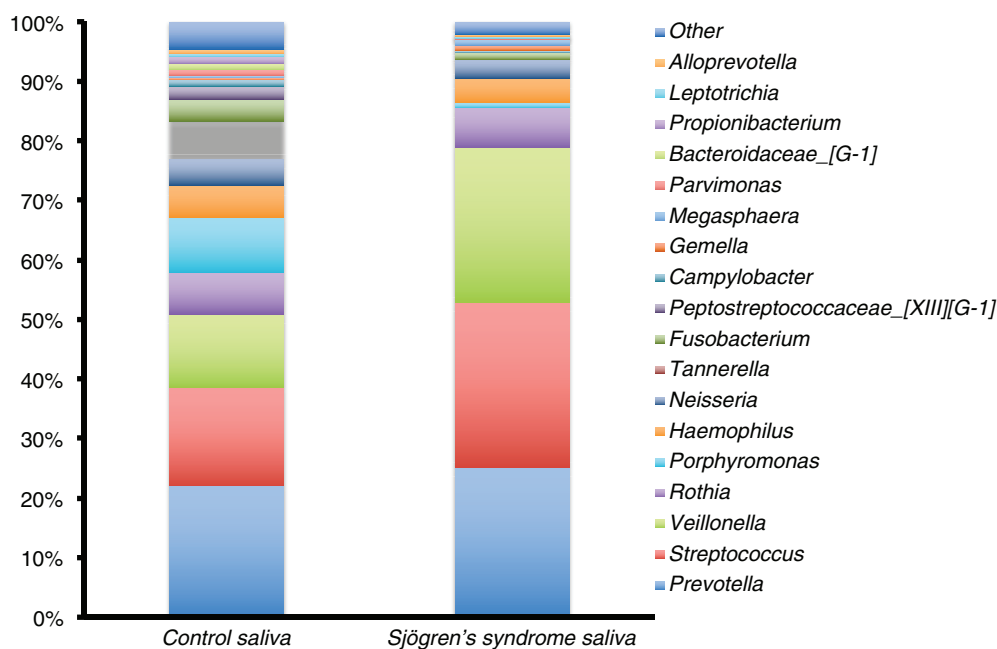


Fig. 1. Summary of the major bacterial phyla detected in pSS saliva ( $n = 9$ ) and control saliva ( $n = 9$ ). MEtaGenome Analyzer (MEGAN) [www-ab.informatik.uni-tuebingen.de/software/megan/welcome.html](http://www-ab.informatik.uni-tuebingen.de/software/megan/welcome.html) was used to visualize taxonomic content. Size of the circles is scaled logarithmically to the total number of reads assigned to the taxon. Fusobacteria, the smallest of the major phyla depicted, constituted 2% of the total.

### Species and richness diversity

The observation that healthy samples had higher species richness and diversity than disease samples are confirmed in Fig. 3. Figure 3A shows that the average species richness is higher in health than in disease, based on the observed species rarefaction curves; Fig. 3B shows the estimated species richness evaluated on the Chao1 matrix and Fig. 3C is the rarefaction curves based on the Shannon index, which evaluated alpha diversity as the evenness of the species of individual samples by taking into account the read count abundance. All these three figures indicated that health samples had higher alpha diversity than the samples from disease. Fig. 3D is the 3D illustration of the PCoA analysis of all samples based on the 'weighed UniFrac' distance matrix and shows that the disease samples can be separated on the second and third principle component, as indication of distinct beta diversities between these two types of samples.

In Fig. 4, a comparison of the phylogenetic distribution of bacterial reads at the species level between the health and the Sjögren samples is given.



**Fig. 2.** Bacterial genera detected in saliva from pSS patients ( $n = 9$ ) and controls ( $n = 9$ ). Groups designated as 'Other' represent minor groups classified. The Y-axis represents relative abundance. An increase in the genera *Streptococcus* and *Veillonella* in pSS saliva relative to control saliva is seen.

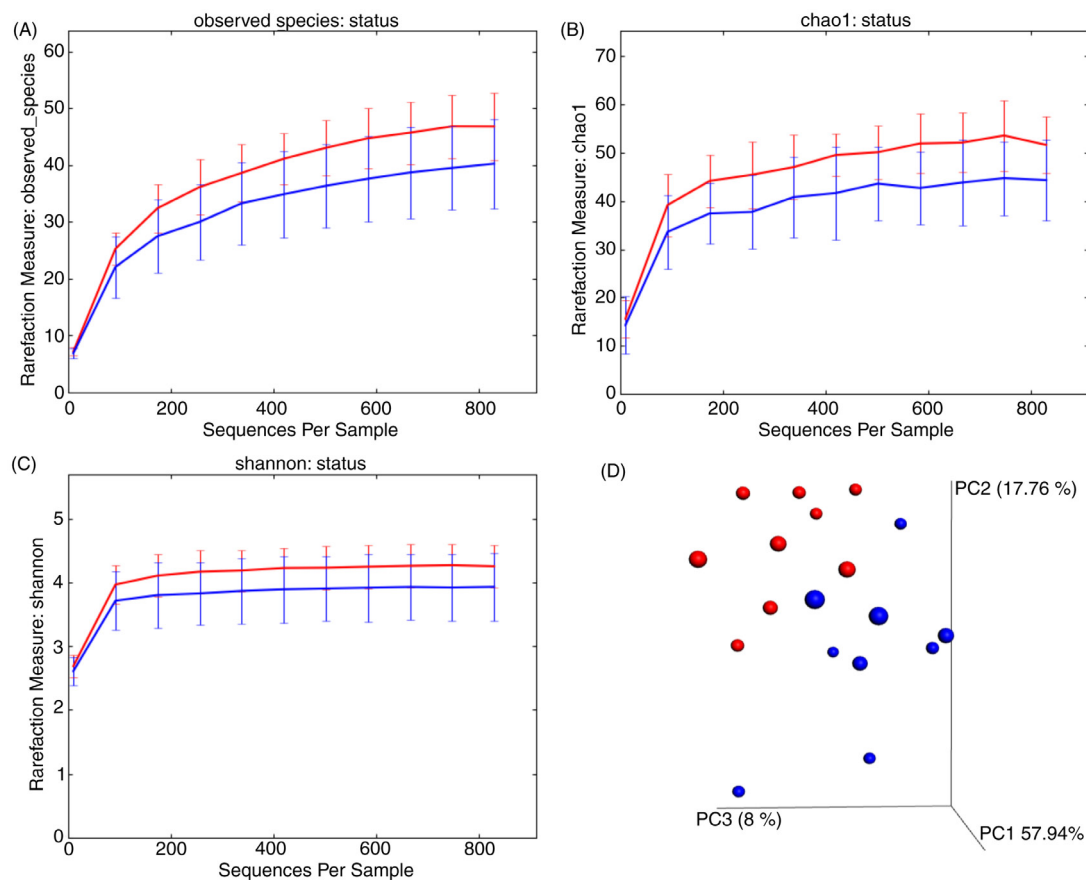
## Discussion

pSS is an autoimmune disorder of unclear etiology. Along with genetic susceptibility, environmental factors are considered important contributors in the development of autoimmunity (17). Especially microorganisms are thought to play a key role in the initiation and development of autoimmune disorders (18), although the mechanisms are still unknown. At present, the majority of studies on the connection between microorganisms and SS focused on the effect of hyposalivation on the composition of the oral microbiota, but it remains unclear whether a shift in the oral microbiota is present independent of hyposalivation in SS. To better demonstrate this, we utilized high throughput sequencing to compare the bacterial composition of the salivary microbiota between pSS patients with normal salivary flow rate and healthy controls.

Indeed, this approach showed alterations in the composition of the salivary microbiota between these two groups. These alterations were characterized by a higher portion of Firmicutes and a reduction in the abundance of four other major phyla, namely Bacteroidetes, Proteobacteria, Fusobacteria, and Actinobacteria. Firmicutes represents a phylum where most bacteria have a Gram-positive cell wall. Among them are oral genera such as *Streptococcus*, *Lactobacillus*, *Selenomonas*, *Clostridium*, and *Eubacterium*. On the contrary, the phyla Bacteroidetes, Proteobacteria, and Fusobacteria are composed of Gram-negative bacteria, while Actinobacteria is a phylum of Gram-positive bacteria. On a genus level, there was an increase in *Streptococcus* and *Veillonella* in pSS patients; however, there was a reduction in the total

number of bacterial genera and species in pSS compared with controls.

These findings indicated that a shift in the oral microbiota can be present in pSS independent of hyposalivation effects. Interestingly, Zhang et al. (19) reported a case-control metagenomic-wide association study of fecal, dental, and salivary samples from patients with RA. In line with our study, they found alterations in the microbiome in all sites, most strikingly an enrichment in Gram-positive bacteria and depletion in Gram-negative bacteria when compared with healthy individuals, which resolved partly after treatment with disease-modifying, anti-rheumatic drugs. *Veillonella* and *Lactobacillus salivarius* were particularly enriched in dental and salivary samples from subjects with RA (19). These findings demonstrate that there is a connection between the oral (and gut) microbiota and autoimmune disease, but it remains unclear whether an altered microbiota causes autoimmunity or the altered microbiota is an effect of the disease. Animal studies have demonstrated that an altered microbiota may indeed have a role in triggering and precipitating autoimmunity. By using animal models of experimental colitis (20) and arthritis (21), it was shown that Gram-negative bacteria, including *Escherichia*, reduced inflammation, possibly through the TLR2/IL-10 axis which results in the repression of the pro-inflammatory pathway (20), whereas Gram-positive bacteria such as *Lactobacillus* contributed to the development of a more severe disease (21). Furthermore, the only study available on the mechanism of action of bacteria in SS development has demonstrated that T-cells from mice

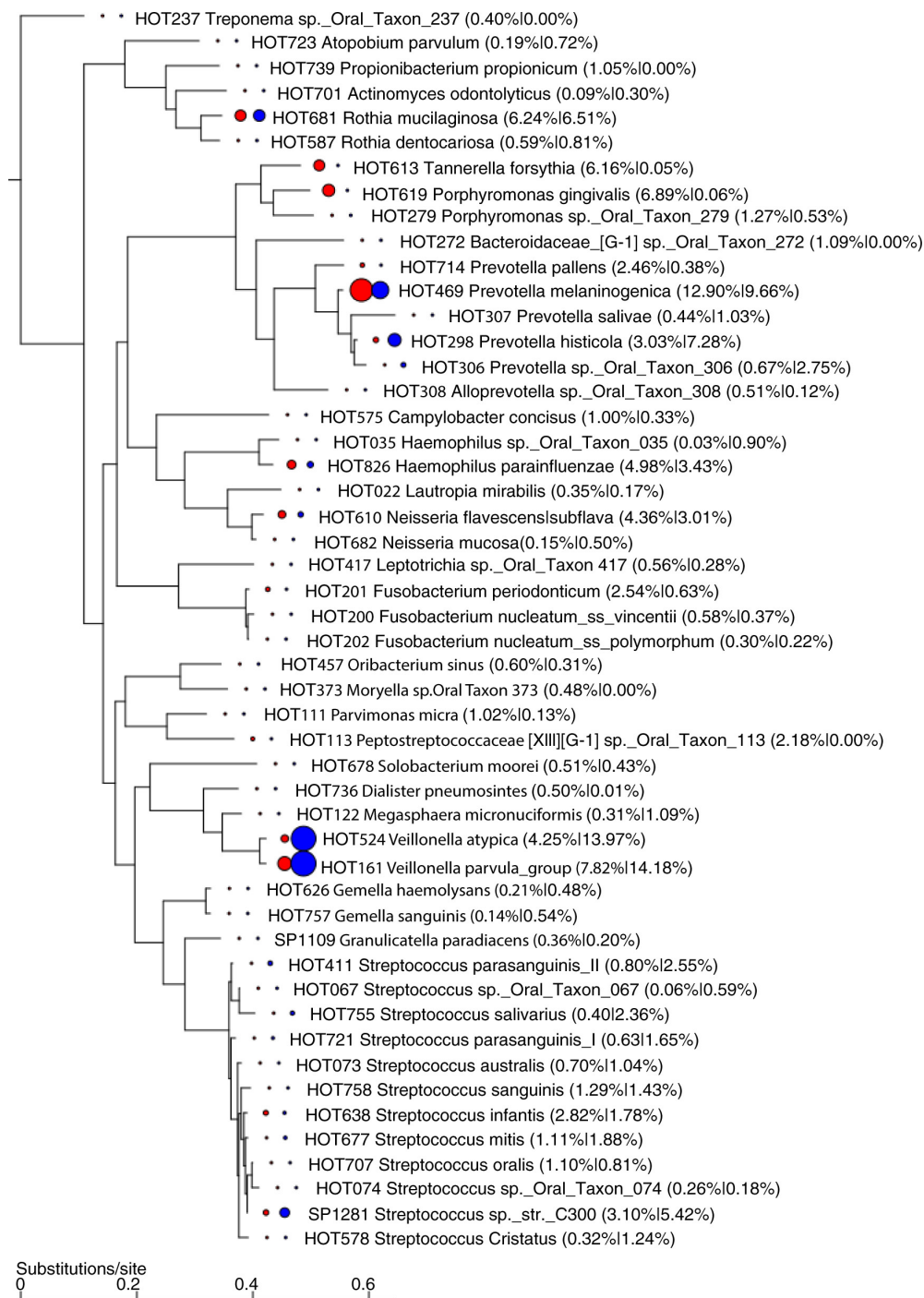


**Fig. 3.** Comparison of microbial diversity between pSS and health samples. A) Average rarefaction curves of pSS and health samples reported as A) observed species, B) Chao1 estimator, and C) Shannon index. D) PCoA 3D plots of all samples with distances calculated using weighed UniFrac matrix. Red curves and dots represent health samples; blue curves represent pSS samples.

immunized with the Ro60 (Sjögren's syndrome antigen A) peptide, derived from one of the major autoantigens in SS, could be activated *in vitro* by peptides from the normal human microbiota, including oral bacteria (22). Among these, a peptide from the von Willebrand factor type A domain protein from the oral microbe *Campylobacter jejuni* was the most potent activator. These intriguing findings support the notion that the microbiome and SS connection might be explained by the molecular mimicry theory. However, no increased abundance of *C. jejuni* in the salivary samples from pSS patients was found in our study.

The increase in *Veillonella* in pSS might be due to the fact that they are known lactate consumers (23). Since there was an increase in *Streptococcus* in pSS patients, it is reasonable to expect that *V. atypica* and the *V. parvula* group had grazed on lactate as a metabolic product from streptococci. Finally, *Porphyromonas gingivalis*, a keystone periodontal pathogen, was more abundant in the control group than in the pSS group, in agreement with other studies that did not find an association between periodontitis or *P. gingivalis* and SS (24, 25).

Another major feature of this study was the comprehensive bioinformatics approach to analyze the next generation 16S rRNA pyrosequencing reads. This provided optimal taxonomy classifications for assessing the microbial communities of clinical samples. Without preliminary sequence quality and chimera checking, the raw reads were directly used for BLASTN-search against the reference sequences. This maximized the taxonomy assignment rate. Our rationale was that, even when some reads have poorer qualities (e.g. lower base quality score, shorter length, or interspecies chimeras), as long as the match length was  $\geq 90\%$  and identity  $\geq 98\%$ , and matched reference sequences represented a single species, one could be confident that the reads were derived from individual species. In the case that a read matched two species under the same criteria, it was possible that the read was a chimera of the two species. Hence, such reads were screened for chimeras at the species level (98% identity) and only non-chimeric reads were retained. For this study, 2,402 reads had multi-species hits and 1,160 were chimeras and were removed from further analyses. Those 1,242 reads that matched multiple species with equal percent



**Fig. 4.** Comparison of phylogenetic distribution of bacterial reads between the health and the Sjögren samples. Percent blue abundance for each species was calculated by assigned reads divided by total reads of all samples of the same clinical state (health or Sjögren). Only those species with a combination of minimum 100 reads were plotted. The size of the colored circles is proportional to the percent abundance, with blue as Sjögren and red as controls.

sequence identity may have represented a new species with the sequence in this region that was equally similar to the hit species. Thus, they were treated as individual species entities for analysis. However, due to the relatively lower read counts in this study, they did not play a significant role that contributed to the characteristic of the samples.

The stringent assignment criteria, that is, the species-level sequence identity (98%) and high alignment length (90%) relative to the reads, prevented the low-quality or chimeric reads being assigned during the BLASTN-search stage. As expected, a high percentage of the unassigned reads (71.48% of 25,438) were determined as chimera. To identify



potential novel species that are not represented in the reference, only high-quality longer reads were used for *de novo* OTUs determination because *de novo* OTU calling is very sensitive to sequence quality noise, and numerous arterial OTUs could be generated if lower quality reads were not excluded.

We used a combined reference sequence dataset that represents a total of 13,164 individual prokaryotic species, of which 1,501 were known species present in human oral cavities. These sequences are well-curated with taxonomy information and were of good quality and length. The inclusion of both GreenGene and NCBI 16S rRNA references enabled the discovery also of non-oral species. Finally, the use of the near full-length 16S rRNA sequences allowed assignment of reads of different hypervariable regions.

For downstream analysis, we used the well-developed QIIME software package, which provided a plethora of visualization tools, as well as various matrices for assessing both alpha and beta diversity. A custom workflow computer script was developed to automatically generate QIIME results based on various MC cutoffs, ranging from 1 to 100. The MC cutoffs are often arbitrarily determined and can have certain impact on the results and the interpretation of biological significance. The use of lower MC will generate higher species richness (alpha diversity). However, regardless of the MC cutoffs and the alpha diversity matrix used, the health samples consistently have higher average alpha diversity than the pSS samples. The human oral cavity is an open environment; thus, it is likely to harbor transient species. However, the quantity is usually very low and including them should not affect species evenness and the best diversity. As expected, although the 2D and 3D best diversity plots varied slightly, the pSS samples in this study were always separable from those of health, as shown in the plots based on the weighed UniFrac matrices.

In conclusion, a noteworthy feature of this study was the dysbiosis demonstrated in the oral microbiota of pSS patients as seen through the marked change in their microbiota of whole saliva compared with controls. This applied to bacterial phyla, genera, and species. Since the salivary flow rate was normal, the present study showed that the shift in the oral microbiota of pSS patients occurred independent of hyposalivation.

## Acknowledgements

HS, PMM, and IO are grateful for funding through the European Commission (FP7-HEALTH-306029 'TRIGGER').

## Conflict of interest and funding

PMM was also supported by a grant from the National Science Center (2014/14/E/NZ6/00162, Poland) and by the Broegelmann Foundation.

## Addendum

A recent paper from Li et al. (2016) has also investigated the oral microbiota in Chinese pSS patients using culture-independent high throughput sequencing technology. Although, the patient group (pSS patients with hyposalivation) and sample type (buccal swab) differs from our study, the paper demonstrates both a shift and a reduction in the diversity of the oral microbiota of Sjögren as reported in our study. These two studies demonstrate that high throughput sequencing analysis of oral microbiota in SS patients can be a powerful tool to better understand this syndrome. <http://dx.doi.org/10.1016/j.archoralbio.2016.06.016>

## References

- Jonsson R, Moen K, Vestrheim D, Szodoray P. Current issues in Sjögren's syndrome. *Oral Dis* 2002; 8: 130–40.
- Atkinson JC, Travis WD, Pillemer SR, Bermudez D, Wolff A, Fox PC. Major salivary gland function in primary Sjögren's syndrome and its relationship to clinical features. *J Rheumatol* 1990; 17: 319–22.
- Boutsi EA, Paikos S, Dafni UG, Moutsopoulos HM, Skopouli FN. Dental and periodontal status of Sjögren's syndrome. *J Clin Periodontol* 2000; 27: 231–5.
- Qin B, Wang J, Yang Z, Yang M, Ma N, Huang F, et al. Epidemiology of primary Sjögren's syndrome: a systematic review and meta-analysis. *Ann Rheum Dis* 2015; 74: 1983–9. doi: <http://dx.doi.org/10.1136/annrheumdis-2014-20537>
- Parvinen T, Larmas M. The relation of stimulated salivary flow rate and pH to *Lactobacillus* and yeast concentrations in saliva. *J Dent Res* 1981; 60: 1929–35.
- Guobis Ž, Kareiviene V, Basevičienė N, Paipalienė P, Niedzelskienė I, Sabalys G, et al. Microflora of the oral cavity in patients with xerostomia. *Medicina (Kaunas)* 2011; 47: 646–51.
- Almståhl A, Wikström M, Stenberg I, Jakobsson A, Fagerberg-Mohlin B. Oral microbiota associated with hyposalivation of different origins. *Oral Microbiol Immunol* 2003; 18: 1–8.
- MacFarlane TW. The oral ecology of patients with severe Sjögren's syndrome. *Microbios* 1984; 41: 99–106.
- Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, et al. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis* 2002; 61: 554–8.
- Siddiqui H, Nederbragt AJ, Lagesen K, Jeansson SL, Jakobsen KS. Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. *BMC Microbiol* 2011; 11: 244. doi: <http://dx.doi.org/10.1186/1471-2180-11-244>
- Al-Hebshi NN, Nasher AT, Idris AM, Chen T. Robust species taxonomy assignment algorithm for 16S rRNA NGS reads: application to oral carcinoma samples. *J Oral Microbiol* 2015; 7: 28934. doi: <http://dx.doi.org/10.3402/jom.v7.28934>
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25: 3389–402.
- Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010; 26: 2460–1. doi: <http://dx.doi.org/10.1093/bioinformatics/btq461>
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010; 7: 335–6. doi: <http://dx.doi.org/10.1038/nmeth.f.303>

15. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; 30: 772–80. doi: <http://dx.doi.org/10.1093/molbev/mst010>.2013
16. White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* 2009; 5: e1000352. doi: <http://dx.doi.org/10.1371/journal.pcbi.1000352>
17. Wahren-Herlenius M, Dörner T. Immunopathogenic mechanisms of systemic autoimmune disease. *Lancet* 2013; 382: 819–31.
18. Van der Meulen TA, Harmsen HJ, Bootsma H, Spijkervet FK, Kroese FG, Vissink A. The microbiome systemic diseases connection. *Oral Dis* 2016. doi: <http://dx.doi.org/10.1111/odi.12472>
19. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med* 2015; 21: 895–905.
20. Oppong GO, Rapsinski GJ, Tursi SA, Biesecker SG, Klein-Szanto AJ, Goulian M, et al. Biofilm-associated bacterial amyloids dampen inflammation in the gut: oral treatment with curli fibres reduces the severity of hapten-induced colitis in mice. *NPJ Biofilms Microbiomes*. 2015; 1. pii: 15019.
21. Kohashi O, Kohashi Y, Takahashi T, Ozawa A, Shigematsu N. Reverse effect of gram-positive bacteria vs. gram-negative bacteria on adjuvant-induced arthritis in germfree rats. *Microbiol Immunol* 1985; 29: 487–97.
22. Szymula A, Rosenthal J, Szczerba BM, Bagavant H, Fu SM, Deshmukh US. T cell epitope mimicry between Sjögren's syndrome Antigen A (SSA)/Ro60 and oral, gut, skin and vaginal bacteria. *Clin Immunol* 2014; 152: 1–9.
23. Periasamy S, Kolenbrander PE. Central role of the early colonizer *Veillonella* sp. in establishing multispecies biofilm communities with initial, middle, and late colonizers of enamel. *J Bacteriol* 2010; 192: 2965–72.
24. Kuru B, McCullough MJ, Yilmaz S, Porter SR. Clinical and microbiological studies of periodontal disease in Sjögren syndrome patients. *J Clin Periodontol* 2002; 29: 92–102.
25. Lugonja B, Yeo L, Milward MR, Smith D, Dietrich T, Chapple IL, et al. Periodontitis prevalence and serum antibody reactivity to periodontal bacteria in primary Sjögren's syndrome: a pilot study. *J Clin Periodontol* 2016; 43: 26–33.