Microbial community interactions: effects of probiotics on oral microcosms
Pham, L.C.

Citation for published version (APA):
Pham, L. C. (2011). Microbial community interactions: effects of probiotics on oral microcosms

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter V - Pyrosequencing analysis of human saliva and saliva-derived microcosms

PHAM LC, BUJJS MJ, TEN CATE JM, CRIELAARD W, ZAURA E. Pyrosequencing analysis of human saliva and saliva-derived microcosms (Manuscript to be submitted to Microbial Ecology).
Microcosms are complex communities derived from natural source such as saliva, and are used to study microbial ecological shifts in controlled in vitro environments. The aim of the current study was to assess the effects of individual inocula on microbiome profiles of the microcosms derived from these individual inocula. Additionally, we aimed to compare the open-ended approach in microbiome profiling by 454 pyrosequencing with the targeted approach by multiplex ligation-dependent probe amplification (MLPA) used for microcosm characterization in our previous study (Pham et al. 2011). To these ends, we amplified hypervariable region V5-V7 of 16S rDNA of individual microcosms and their respective inocula (human saliva screened for high and low counts of mutans streptococci (MS)). The microcosms were obtained under four growth conditions: plain medium, plain medium supplemented with a probiotic strain Lactobacillus rhamnosus GG (LGG), sucrose-supplemented medium and medium supplemented with sucrose and LGG (Pham et al. 2011). Individually tagged amplicon libraries were pooled and sequenced by the GS FLX Titanium system of Roche 454 pyrosequencing technology. Subsequently, the data was processed using the RDP pipeline at Straightforward Novel Webinterface for Microbiome Analysis. In total, 386,869 high quality reads (average length 370 nt) were obtained and clustered in 923 Operational Taxonomic Units (OTUs) at 6% difference level. Microbiome profiles of the six inocula were clearly dichotomized in high and low MS-saliva groups. Two OTUs, classified as Neisseria and Porphyromonas, were significantly more abundant in the low MS-inocula, while two Prevotella OTUs were present in significantly higher proportion in the high MS-salivas. Microcosms differed significantly from the inocula. The representatives of phylum Firmicutes dominated the microcosms. The microbiome profiles of the microcosms clustered according to the type of the inoculum (low or high MS saliva). Both profiling methods, MLPA and pyrosequencing, showed comparable results in discriminating the effects of inocula and growth conditions. We conclude that the choice of the inoculum based on predefined criteria allows modeling of ecological shifts in saliva-derived microcosms. Comparison of targeted profiling by MLPA with high throughput sequencing by 454 pyrosequencing showed that the MLPA is a valid method in
Chapter V

microcosm profiling. The output of the pyrosequencing study will allow further improvement of the MLPA approach (and other targeted methods) by justifying the choice of sequences for specific probe panels.

INTRODUCTION

The oral cavity is a dynamic ecological habitat where an interplay between environmental and host factors determines the composition and function of the microbial communities. Health to disease equilibrium is determined by this interplay, a phenomenon which was introduced to oral microbiology research by Marsh as ‘ecological plaque hypothesis’ (MARSH 1994). According to this hypothesis, repeated episodes of low pH lead to suppression of acid-sensitive species and the outgrowth of microorganisms with an aciduric physiology. Together this leads to a shift towards the disease, dental caries.

Clinical studies provide the only realistic systems to study ecological shifts and their consequences. However, control of environmental parameters in vivo is cumbersome, if possible at all. In vitro model systems allow the control of most, if not all, environmental parameters, but have serious limitations. They either lack the complexity of natural communities, such as by focusing on biofilm consortia of few defined species (GUGGENHEIM et al. 2001; SHU et al. 2000), or are too complex to characterize their composition and to assess the impact of environmental factors, such as in plaque- or saliva-derived communities, called microcosms (MCBAIN et al. 2003; PRATTEN et al. 1998b; SISSONS 1997).

In order to characterize microcosm communities, numerous methods ranging from targeted profiling with selected probes, e.g., by DNA-DNA checkerboard (FILOCHE et al. 2007b; SISSONS 1997) and multiplex ligation-dependent probe amplification (MLPA) (PHAM et al. 2011; TEREFEWORK et al. 2008) to open-ended fingerprinting by denaturing gradient gel electrophoresis (DGGE) (LEDGER et al. 2006; PHAM et al. 2009) have been applied. None of these methods, however, offers both quantifiable and untargeted (open-end) high resolution assessment of microcosm composition. Current developments in sequencing technologies allow the identification by high throughput parallel sequencing (454 pyrosequencing) of short hypervariable
Pyrosequencing analysis of human saliva and microcosms

fragments of small subunit ribosomal DNA of microorganisms at a high resolution (VOELKERDING et al. 2009). Due to relatively low bias and reducing costs, this method is gaining popularity and has been used in several clinical, oral microbiome studies (KEIJSEER et al. 2008; LAZAREVIC et al. 2010; Li et al. 2010; ZAURA et al. 2009). So far, 454 pyrosequencing has been used to confirm the microbial diversity observed by DGGE profiling of a single microcosm sample, derived from saliva pooled of several individuals (TIAN et al. 2010).

In a previous study (PHAM et al. 2011), we obtained microcosms that differed in their inoculum source and in growth conditions. The expected differences in the community profiles were confirmed by the community analyses of targeting 20 selected microorganisms using the MLPA technique. The aim of the current study was to assess the effects of individual inocula on microbiome profiles of microcosms described in Pham et al. (2011) by 454 pyrosequencing and compare with the output of the MLPA profiling.

MATERIALS AND METHODS

In this study we pyrosequenced the hypervariable regions V5-V7 of 16S rDNA from the samples obtained and described in detail in our previous study (PHAM et al. 2011) in Chapter IV. The following is a summary of the experimental procedures and sample processing, followed by a description of DNA amplicon library preparation, pyrosequencing and data analysis.

Generation of microcosm biofilms

Saliva-derived microcosms were grown on bovine enamel-dentin discs using an Active Attachment Biofilm (AAB) model (EXTERKATE et al. 2010). The model consists of a custom-made stainless steel lid with nylon clamps that can accommodate 24 substrata and fit into a 24-well plate. The enamel-dentin discs were prepared and inserted into the nylon clamp in the lid of the AAB model.

The inoculum (human stimulated saliva) was prepared as described previously (PHAM et al. 2009). The use of human saliva was approved by institutional Review Board. We selected six individual salivas based on mutans streptococci (MS) counts.
in saliva: 1) three salivas with more than 10^5 cfu/mL MS in saliva (High MS group) and 2) three salivas with less than 10^5 cfu/mL of salivary MS (Low MS group). The growth medium for microcosms comprised artificial saliva (McBain et al. 2005). The microcosms were grown individually at four experimental conditions: a) in plain medium (Suc-LGG-); b) in plain medium with Lactobacillus rhamnosus GG (LGG) (Suc-LGG+); c) in sucrose-supplemented medium (Suc+LGG-); d) in sucrose-supplemented medium with LGG (Suc+LGG+). The microcosms were harvested after 72 h. In conditions (b) and (d), 10^6 cfu/mL of LGG was added to the medium at the time of the inoculation. Medium was refreshed at 8, 24, 32, 48 and 56 h. Each of the six individual saliva-derived microcosm experiments was performed with triplicate samples and repeated twice.

**Biofilm processing**

After 72 h, the pH of the spent medium was measured and the acidogenicity test was performed: the biofilms were incubated with 0.2% glucose anaerobically for 3 h at 37 °C. After the incubation, the amount of L-lactic acid in the incubation fluid was determined enzymatically using a colorimetric assay (Gutmann and Wahlefeld 1974). After the acid test, the specimens with adherent biofilms were sonicated in PBS and then the biofilm samples were centrifuged. The pellets were stored at -80 °C until DNA extraction and quantification (Zaura et al. 2009).

**Assessment of mineral loss**

The dentin/enamel specimens were used to assess the mineral content of the dentin by transversal microradiography (Lagerweij et al. 1994). Two 200-μm-thick sections were cut from the center of each cylinder perpendicularly to the surface and radiographed (Van Strijp et al. 1995), the images were processed and the integrated mineral loss (IML, vol%×μm) calculated.

**DNA library preparation and pyrosequencing**

For microbiome analyses, a maximum of 102 samples could be individually tagged with sample identification keys and sequenced. Of the 150 samples (6 inocula and
36 samples per each of the four growth conditions), we included all inocula (N=6), and at least two replicates of each of the individual microcosms per growth condition (Suc-LGG-: N=26; Suc-LGG+: N=24; Suc+LGG-: N=22; Suc+LGG+: N=24) for pyrosequencing.

PCR amplicon libraries of the small subunit ribosomal RNA gene V5-V7 hypervariable region were generated for the individual samples. PCR was performed using the forward primer 785F (GGATTAGATACCCBRGTAGTC) and the reverse primer 1175R (ACGTCRTCCCDCTTCCTC). The primers included the 454 Life Sciences (Branford, CT, USA) Adapter A (for forward primers) and B (for reverse primers) fused to the 5' end of the 16S rDNA bacterial primer sequence and a unique 10 nt sample identification key.

The amplification mix contained 2 units of Pfu Ultra II Fusion HS DNA polymerase (Stratagene, CA, USA), 1 unit Buffer Pfu Ultra II [10x], including 2.0 mM MgCl2 (Stratagene), 240 μM dNTP PurePeak DNA polymerase Mix (Pierce Nucleic Acid Technologies, Milwaukee, WI), 0.5 μM of each primer. After denaturation (95 °C; 2 min), 30 cycles were performed consisting of denaturation (94 °C; 30 sec), annealing (50 °C; 40 sec), and extension (72 °C; 80 sec). The amplicons were purified by means of the MinElute kit (Qiagen, Hilden, Germany). The quality and the size of the amplicons were analyzed on the Agilent 2100 Bioanalyzer with the DNA 1000 Chip kit (Agilent Technologies, Santa Clara, CA, USA). The amplicon libraries were pooled in equimolar amounts and sequenced unidirectionally in the reverse direction (B-adaptor) by means of the Genome Sequencer FLX Titanium system (Roche, Basel, Switzerland).

Pyrosequencing data analysis

The sequences were processed using RDP pipeline (COLE et al. 2005) at SNoWMAn - Straightforward Novel Webinterface for Microbiome Analysis (Graz University of Technology, Bioinformatics https://epona.genome.tugraz.at/snowman/). Fasta files with sample sequences, sample identification key files, primer sequences and quality files were uploaded to the webinterface. Two mismatches in each primer sequence were allowed. In RDP preprocessing, the sequences were sorted by RDP Tagsorter
Chapter V

and trimmed by RDP Trimmer by removing primer sequences and low-quality data, sequences that do not have an exact match to the reverse primer, that have an ambiguous base call (N) in the sequence, or that are shorter than 150 nt after trimming. For sequence alignment, the bacteria_2008_12_24_ncbi16S_508_mod5 infernal alignment model was used. NaiveBayesian_rRNA_Classifier_2010_02_01_2.1 model was used for sequence classification. Sequences were clustered in operational taxonomic units (OTUs) within 6% difference level. Taxonomy assignment was accepted only if the assignment probability was 0.8 or higher.

Principal component analysis (PCA), calculation of the Shannon diversity index and the species dominance index were performed in PAST (H\textsc{ammer} et al. 2001). PCA is a commonly used method to reduce the dimensions of multivariate data. Principal components are combined variables that explain a large part of the overall variance. Two-dimensional graphs of the principal components visualize clustering of the individual samples in the space determined by these components. The coefficients of the PCA loadings of each variable (OTU in this case) were used to estimate the impact of each individual OTU on the clustering of the samples in PCA. PCA was performed on normalized and log2 transformed OTU data. For the diversity index \( H' = -\sum p_i \ln p_i \) the proportion of species \( i \) relative to the total number of species (\( p \)) is calculated, and then multiplied by the natural logarithm of this proportion (\( \ln p \)). The resulting product is summed across species, and multiplied by -1. This index accounts for both the abundance and the evenness of the species present. Ecological dominance is the degree to which a particular species is more numerous than its competitors in an ecological community, or makes up more of the biomass. It ranges from 0 (all taxa are equally present) to 1 (one taxon dominates the community completely). The dominance index is calculated as follows: 

\[
D = \sum \left( \frac{n_i}{n} \right)^2
\]

where \( n_i \) is the number of individuals of taxon \( i \) and \( n \) is total number of individuals.

To compare the diversity data (number of taxa, diversity index and dominance data) of the microcosms grown at four different conditions with the diversity data of the inoculum, we performed general linear model repeated measures (GLM-RM) test (SPSS, version 18). Independent samples T-test was used to compare the abundance of those OTUs, which had the highest coefficients of the PCA loadings, between the two different groups (SPSS, version 18). The significance level of all tests was 0.05.
To compare the pyrosequencing results with the MLPA profiles of the same samples (PHAM et al. 2011), the highest taxon (genus or higher taxonomic level) dataset was used. The data of the replicate samples were averaged into a single microbial profile per experiment (N=6 individual microcosm experiments), normalized and log2 transformed before being analyzed by PCA. The relation between the relative abundance of the MLPA probe for genus Prevotella and the final taxon Prevotella in the pyrosequencing dataset was assessed with nonparametric Spearman’s rho correlation (the data was not normally distributed: Kolmogorov-Smirnov normality test p<0.001).

Supplementary files with full list of higher taxa (Supplementary file 1), OTUs at 6% difference (Supplementary file 2) and PCA results (Supplementary file 3) are accessible electronically at:

http://www.onderwijs.acta.nl/data/Supplementary_file_1_taxa_level_CP.zip
http://www.onderwijs.acta.nl/data/Supplementary_file_2_all_OTUs_CP.zip
http://www.onderwijs.acta.nl/data/Supplementary_file_3_PCA_results_without_OTU_94_CP.zip

RESULTS

Overall sequencing and taxonomy output

Sequencing resulted in 456,903 reads that passed the Roche 454 quality control algorithm. Of these, 450,815 reads passed trimming and filtering of the SNoWMAn RDP pipeline. In total, 389,932 reads with average length of 370 nt (355 – 378 nt) passed the orientation check and were used in merging, uniquing and alignment steps of the pipeline. This processing pipeline resulted in 126,347 unique sequences that were further used for clustering in operational taxonomic units (OTUs) at 99%, 97% and 94% similarity. Only the data from 94% similarity (6% OTUs) are described further. We have chosen 6% difference level as this has been accepted previously as a more conservative estimate of species definition than the 3% level (KBISER et al. 2008). Clustering at 6% level resulted in 1,841 OTUs. To further reduce the impact of sequencing errors or potential contaminants, a cut-off of at least 5 reads per OTU was applied to the data. Of all OTUs, 923 OTUs with 386,869 reads passed this cut-off. An individual sample harbored on average 3,863 reads (range 1,177 – 8,261...
The 923 OTUs with 386,869 reads were classified into 64 higher taxa (genus or more inclusive taxon) belonging to seven phyla of the domain Bacteria (Figure 1, supplementary file 1). Firmicutes dominated the data set (92.2% of all reads), while 2.2% of the reads could not be classified to any of the known phyla at the probability of 0.8 or higher (Unclassified bacteria), 2% of reads were classified as Actinobacteria, 1.5% - as Bacteroidetes, 1.2% - as Fusobacteria, 0.9% - as Proteobacteria, 0.1% - as candidate division TM7 and only 9 reads (0.002%) – as candidate division SR1.

**Taxonomical classification and ecological diversity per sample type**

Among the six saliva samples used as inocula for the microcosms, 57% (SD 16.5) reads/sample were classified as Firmicutes, 19.9% (SD 7) - as Bacteroidetes, 10.1% (SD 2.7) – as Actinobacteria, 8.7% (SD 9.1) – as Proteobacteria, 2.3% (SD 1.7) – as Fusobacteria, 0.5% (SD 0.6) – as candidate division TM7, while 1.4% (SD 0.4) of reads per saliva sample could not be classified to any known phylum (Figure 2). Reads classified as candidate division SR1 were found in two out of six inocula (0.1%, SD 0.1). High MS inocula contained statistically significantly lower proportion of phylum Proteobacteria than low MS inocula (p<0.05).
Pyrosequencing analysis of human saliva and microcosms

**Figure 1.** Average relative abundance (%) of reads classified to the five predominant phyla and unclassifiable bacteria in inoculum (N=6) and in microcosm samples grown at four conditions: in plain medium (Suc-LGG-) (N=26), in plain medium with probiotics (Suc-LGG+) (N=24), in sucrose-supplemented medium (Suc+LGG-) (N=22) and in sucrose-supplemented medium with probiotics (Suc+LGG+) (N=24). Error bars – standard deviations.

**Figure 2.** Average relative abundance (%) of the reads classified into the predominant microbial phyla by individual saliva that was used as the inoculum for microcosm experiments. L1, L2, L3 indicate the inocula with low mutans streptococci (MS) counts in the saliva; H1, H2, H3 – high MS inocula (MS >10^5 cfu/mL saliva).
On average, each inoculum contained 123 OTUs (SD 14) or taxa (Figure 3A). The microcosms grown with sucrose supplementation (113 OTUs (SD 30) and 111 OTUs (SD 35) without and with LGG, respectively) harbored similar number of OTUs compared to the inocula, while nearly double amounts of taxa (p<0.05) were found in the samples grown in the absence of sucrose – 208 (SD 43) and 212 (SD 33) OTUs without and with LGG, respectively. The 272 OTUs that were present only in the plain medium (Suc-) groups, were classified as Actinobacteria (8 OTUs), Bacteroidetes (24 OTUs), Firmicutes (132 OTUs), Fusobacteria (2 OTUs), Proteobacteria (22 OTUs) or remained unclassified (84 OTUs). Although each of these 272 OTUs were at a low abundance individually, together they accounted for 2.9% (SD 1.6) of the reads of the respective sample.

Representatives of genus *Veillonella* and *Streptococcus* dominated both the inocula and microcosms (Table 1, supplementary file 2). Four OTUs (#8 and #12 – *Streptococcus*; #7 – *Veillonella*; #18 – Unclassified Firmicutes) were found in all microcosm samples and in all inocula. Additionally, three OTUs (#9 - Lactobacillales; #44 – *Veillonella*; #70 – *Actinomyces*) were found in at least 80% of the samples. Among the six inocula, 26 OTUs were found in all, and 22 OTUs – in all but one inoculum. The predominant and 100%-prevalent inoculum-taxa (OTUs) were classified to genus *Veillonella* (#7), *Streptococcus* (#8, #12), *Actinomyces* (#70), *Prevotella* (#137, #201), *Rothia* (#555), *Porphyromonas* (#336), *Neisseria* (#662), *Gemella* (#1146) and order Lactobacillales (#9).

Apart from the decrease in the number of taxa (OTUs), growth in the presence of sucrose resulted in statistically significantly increased ecological dominance (Figure 3B) and in decreased ecological diversity (Figure 3C) of the community compared to those in the other groups.
Figure 3. Output of the diversity statistics by sample type (inoculum and microcosms obtained from four different growth conditions: plain medium (Suc-LGG-), plain medium with LGG (Suc-LGG+), sucrose-supplemented medium (Suc+LGG-) and sucrose-supplemented medium with LGG (Suc+LGG+)): (A) species richness (number of OTUs or taxa); (B) Dominance index; (C) Shannon H diversity index. The data are average values from replicates of two independent experiments per inoculum. Error bars – standard deviations. * - statistically significantly different from the groups without the asterisk (p<0.05; GLM-RM test).
### Table 1

Relative abundance (%) of top 30 OTUs per sample group. Data are average (standard deviation) values from samples belonging to five groups: inoculum (N=6), microcosms grown in the plain medium (Suc-LGG-) (N=26), in the plain medium with probiotics (Suc-LGG+) (N=24), in sucrose-supplemented medium (Suc+LGG-) (N=22) and in sucrose-supplemented medium with probiotics (Suc+LGG+) (N=24). Full list of OTUs per individual sample is available in the supplementary file 2.

<table>
<thead>
<tr>
<th>OTU</th>
<th>Final Taxon</th>
<th>Inoculum</th>
<th>Suc-LGG-</th>
<th>Suc-LGG+</th>
<th>Suc+LGG-</th>
<th>Suc+LGG+</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Veillonella</td>
<td>17.7 (10.8)</td>
<td>31.6 (13.8)</td>
<td>37.9 (12.3)</td>
<td>55.1 (13.8)</td>
<td>46.3 (13.5)</td>
</tr>
<tr>
<td>12</td>
<td>Streptococcus</td>
<td>10.4 (5.2)</td>
<td>19.6 (5.3)</td>
<td>17.3 (7.4)</td>
<td>6.5 (17.3)</td>
<td>1.3 (1.2)</td>
</tr>
<tr>
<td>8</td>
<td>Lactobacillales</td>
<td>5.4 (2.7)</td>
<td>7.8 (8.2)</td>
<td>8.6 (7.6)</td>
<td>6.8 (2.7)</td>
<td>5.2 (3.5)</td>
</tr>
<tr>
<td>9</td>
<td>Lactobacillaceae</td>
<td>10.5 (11.0)</td>
<td>0.5 (0.5)</td>
<td>0.3 (0.3)</td>
<td>7.5 (3.9)</td>
<td>5.8 (3.9)</td>
</tr>
<tr>
<td>15</td>
<td>Streptococcaceae</td>
<td>1.5 (3.1)</td>
<td>0.9 (1.7)</td>
<td>0.7 (1.2)</td>
<td>4.0 (4.7)</td>
<td>3.5 (4.5)</td>
</tr>
<tr>
<td>70</td>
<td>Actinomyces</td>
<td>4.3 (1.9)</td>
<td>1.5 (0.6)</td>
<td>2.1 (1.0)</td>
<td>0.9 (0.6)</td>
<td>0.6 (0.4)</td>
</tr>
<tr>
<td>46</td>
<td>Veillonella</td>
<td>0.3 (0.2)</td>
<td>3.9 (3.8)</td>
<td>2.5 (2.0)</td>
<td>0.5 (0.4)</td>
<td>0.3 (0.4)</td>
</tr>
<tr>
<td>662</td>
<td>Neisseria</td>
<td>7.0 (7.8)</td>
<td>0.1 (0.2)</td>
<td>0.1 (0.1)</td>
<td>0</td>
<td>0.001 (0.004)</td>
</tr>
<tr>
<td>18</td>
<td>Firmicutes</td>
<td>0.3 (0.2)</td>
<td>2.3 (0.7)</td>
<td>2.4 (0.6)</td>
<td>1.5 (0.7)</td>
<td>0.7 (0.4)</td>
</tr>
<tr>
<td>336</td>
<td>Porphyromonas</td>
<td>6.3 (7.3)</td>
<td>0.2 (0.7)</td>
<td>0.003 (0.07)</td>
<td>0.001 (0.003)</td>
<td>0</td>
</tr>
<tr>
<td>95</td>
<td>Veillonella</td>
<td>0.2 (0.2)</td>
<td>0.5 (0.5)</td>
<td>0.7 (0.4)</td>
<td>2.5 (0.8)</td>
<td>1.5 (0.5)</td>
</tr>
<tr>
<td>42</td>
<td>Fusobacterium</td>
<td>1.5 (1.2)</td>
<td>2.3 (2.6)</td>
<td>1.7 (2.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>137</td>
<td>Prevotella</td>
<td>4.5 (3.2)</td>
<td>0.3 (0.6)</td>
<td>0.5 (1.8)</td>
<td>0.01 (0.03)</td>
<td>0.003 (0.01)</td>
</tr>
<tr>
<td>4</td>
<td>Veillonella</td>
<td>0.2 (0.2)</td>
<td>0.9 (0.5)</td>
<td>1.1 (0.3)</td>
<td>1.6 (0.6)</td>
<td>0.8 (0.4)</td>
</tr>
<tr>
<td>89</td>
<td>Megapora</td>
<td>0.6 (0.6)</td>
<td>2.6 (4.2)</td>
<td>0.6 (1.4)</td>
<td>0.8 (1.3)</td>
<td>0.06 (0.2)</td>
</tr>
<tr>
<td>30</td>
<td>Veillonella</td>
<td>0.1 (0.1)</td>
<td>1.7 (1.0)</td>
<td>1.7 (0.8)</td>
<td>0.6 (0.4)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>36</td>
<td>Firmicutes</td>
<td>0.3 (0.3)</td>
<td>0.1 (0.2)</td>
<td>0.1 (0.2)</td>
<td>1.8 (0.7)</td>
<td>1.1 (0.5)</td>
</tr>
<tr>
<td>555</td>
<td>Rothia</td>
<td>3.5 (2.1)</td>
<td>0.001 (0.01)</td>
<td>0.001 (0.01)</td>
<td>0.001 (0.003)</td>
<td>0</td>
</tr>
<tr>
<td>201</td>
<td>Prevotella</td>
<td>2.5 (1.6)</td>
<td>0.1 (0.2)</td>
<td>0.01 (0.02)</td>
<td>0.01 (0.03)</td>
<td>0.002 (0.01)</td>
</tr>
<tr>
<td>22</td>
<td>Veillonella</td>
<td>0</td>
<td>0.002 (0.01)</td>
<td>0.4 (0.3)</td>
<td>0.001 (0.005)</td>
<td>2.0 (0.6)</td>
</tr>
<tr>
<td>16</td>
<td>Pasteurelaceae</td>
<td>0.6 (0.6)</td>
<td>0.9 (0.9)</td>
<td>0.6 (0.5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1146</td>
<td>Gemella</td>
<td>2.0 (2.7)</td>
<td>0.04 (0.06)</td>
<td>0.03 (0.06)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>194</td>
<td>Solobacterium</td>
<td>0.2 (0.1)</td>
<td>0.7 (0.8)</td>
<td>0.8 (0.8)</td>
<td>0.2 (0.3)</td>
<td>0.1 (0.2)</td>
</tr>
<tr>
<td>186</td>
<td>Firmicutes</td>
<td>0</td>
<td>0.003 (0.01)</td>
<td>0.2 (0.2)</td>
<td>0.001 (0.01)</td>
<td>1.7 (0.4)</td>
</tr>
<tr>
<td>74</td>
<td>Veillonella</td>
<td>0.1 (0.1)</td>
<td>0.5 (0.4)</td>
<td>0.4 (0.3)</td>
<td>0.4 (0.3)</td>
<td>0.2 (0.3)</td>
</tr>
<tr>
<td>44</td>
<td>Veillonella</td>
<td>0.1 (0.1)</td>
<td>0.3 (0.2)</td>
<td>0.3 (0.2)</td>
<td>0.4 (0.2)</td>
<td>0.4 (0.2)</td>
</tr>
<tr>
<td>477</td>
<td>Granulicatella</td>
<td>0.6 (0.3)</td>
<td>0.4 (0.3)</td>
<td>0.4 (0.3)</td>
<td>0.1 (0.1)</td>
<td>0.01 (0.03)</td>
</tr>
<tr>
<td>269</td>
<td>Lactobacillus</td>
<td>0</td>
<td>0</td>
<td>0.1 (0.1)</td>
<td>0.003 (0.01)</td>
<td>1.3 (0.7)</td>
</tr>
<tr>
<td>2</td>
<td>Veillonella</td>
<td>0.1 (0.2)</td>
<td>0.3 (0.1)</td>
<td>0.3 (0.1)</td>
<td>0.4 (0.1)</td>
<td>0.4 (0.2)</td>
</tr>
</tbody>
</table>
Pyrosequencing analysis of human saliva and microcosms

Reads from the microcosm samples (Figure 1) were dominated by a single phylum – Firmicutes, contributing 76–96% of reads per sample (average 88.7%, SD 5.4) if grown in the plain medium (Suc-). Sucrose supplementation to the growth medium (Suc+) increased the proportion of this phylum (93-100% reads/sample; average 98%, SD 1.5) significantly (p<0.05). Irrespective of the presence of sucrose, reads belonging to Actinobacteria, Bacteroidetes, Proteobacteria and candidate division TM7 decreased significantly (p<0.05) in the microcosms compared to the inocula. Fusobacteria remained around 2% of the reads/sample in microcosms grown in the plain medium. Growth in the presence of sucrose nearly eliminated representatives of Bacteroidetes and entirely excluded Fusobacteria and Proteobacteria from the microcosms. Reads classified as representatives of the candidate division TM7 (unculturable and unidentified bacteria) were found in 41% of the microcosms, with an average of 0.05% (range 0–0.3%; SD 0.08) of reads per sample. Higher prevalence of this bacterium was found in the microcosms grown in the plain medium (in 77% and 50% of microcosms, Suc-LGG- and Suc-LGG+, respectively), than in the microcosms grown in sucrose-supplemented medium (in 18% and 13% of the microcosms, Suc+LGG- and Suc+LGG+, respectively). Reads that could not be classified to any of the known phyla (Unclassified Bacteria, Figure 1) increased from 1.4% reads (SD 0.4) per sample in the inocula to 4% of reads (SD 1.3) per Suc- microcosm or decreased to 0.6% reads (SD 0.4) per Suc+ sample.

Firmicutes dominated all microcosm samples (Figure 1). The most abundant group (38–65% of all Firmicutes) was classified into the Veillonellaceae family (Figure 4). Reads from this family were clustered into 115 OTUs, 63 of which were classified as genus Veillonella (found in all groups of samples), 12 – as genus Megasphaera (in all groups of samples), two – as genus Centipeda (in inoculum and Suc- samples), one – genus Selenomonas (inoculum only), one – genus Dialister (in Suc- samples) with remaining 36 OTUs as unclassified Veillonellaceae (in all groups). Among the genus Veillonella OTUs, four OTUs were only found in the inocula, 10 OTUs – only in the Suc- microcosms, and two OTUs – only in the Suc+ samples. Two OTUs (#1784 and #792) were highly prevalent (present in >80% of samples) in Suc- microcosms, but were absent from the Suc+ samples. However, none of the predominant OTUs found in the Suc+ groups were absent in the Suc- samples. Three OTUs (#22, #1636, #1841)
were more prevalent in the microcosms grown in the presence of probiotics (LGG+), irrespective of the presence of sucrose.

**Figure 4.** Distribution of reads (%) within phylum Firmicutes, classified at the Family level by sample types – inoculum and microcosms obtained from four different growth conditions: plain medium (Suc-LGG-), plain medium with LGG (Suc-LGG+), sucrose-supplemented medium (Suc+LGG-) and sucrose-supplemented medium with LGG (Suc+LGG+). * - reads that could not be classified at the family level but at higher levels are presented as one taxonomic group (Higher taxon).

**Clustering of the microbiome profiles by principal component analysis (PCA)**

The dataset with the 923 OTUs was log2 transformed and exposed to data reduction by PCA. First, the PCA was applied to the full dataset including the microcosms and inocula. The two major principal components together explained 50% of the total variance and resulted in clear clustering of microcosms by the growth condition (with or without sucrose and with or without LGG) and separated the inocula from the microcosms (Figure 5).
PCA analysis showed that the principal component 1 (PC1) discriminated between the inocula of the low MS group and the high MS group (Figure 6A) and explained 65% of the total variance. The five main loadings of this component belonged to the OTUs classified as *Porphyromonas*, *Neisseria*, Lactobacillales, *Prevotella* and *Gemella* (Figure 6B). Of these, low MS-inocula contained significantly higher proportion of *Porphyromonas* (OTU #336) and *Neisseria* (OTU #662), and significantly lower proportion of *Prevotella* (OTU #137, #201) than the high MS-inocula (p<0.05).

**Figure 5.** Principal Component Analysis (PCA) plot of the principal component 1 (PC1) and the principal component 2 (PC2) on the complete dataset (inocula and all microcosms) of the normalized and log2 transformed OTU data. Crosses indicate saliva samples used as inocula; open triangles – microcosm samples grown in the plain medium (Suc-LGG-); filled triangles – plain medium with LGG (Suc-LGG+); open circles – sucrose-supplemented medium (Suc+LGG-); filled circles – sucrose-supplemented medium with LGG (Suc+LGG+). PC1 explained 34.9%, PC2 – 15% of the total variance.
Next, we applied PCA to assess the effects of the microcosm source (high or low MS-inoculum) on the microcosms. In the plain medium, microcosms that originated from the low MS-inoculum were clearly separated from the high MS-microcosms by the PC1, explaining 38% of the total variance (Figure 7A). High MS Suc- microcosms harbored significantly higher proportion of the OTUs classified as *Megasphaera* (OTU #89), *Solobacterium* (OTU #194), *Oribacterium* (OTU #178) and *Prevotella* (OTU #72) than the low MS Suc- microcosms, while low MS Suc- microcosms had higher proportion of *Veillonella* (OTU #7) and *Streptococcus* (OTU #46) (Figure 7B).

Among the microcosms grown in the sucrose-supplemented medium, PC1 separated the high and the low MS-microcosms and explained 42% of the variance (Figure 8A). High MS Suc+ microcosms harbored significantly higher proportion of *Lactobacillales* (OTU #9), *Megasphaera* (the same OTU as in Suc- microcosms) and
Actinomyces (OTU #70), while unclassified Firmicutes (OTU #18) were significantly more abundant in low MS Suc+ microcosms than in the high MS samples (Figure 8B).
Figure 8. Effect of the type of the inoculum on the microcosm samples grown in the sucrose-supplemented medium (Suc+LGG-): (A) the PCA plot with the first main components of the PCA. PC1 discriminated between samples derived from the low MS-inocula (open circles) and high MS-inocula (filled circles). (B) Relative abundance of those OTUs which contributed most to the PC1 (values in the brackets are the PC1 loadings of the respective OTUs). * - statistically significant difference between low and high MS samples (p<0.05; Independent samples T-test).

Next, we assessed the effects of the sucrose supplementation in the growth medium on the microbiome profiles of the microcosms (Figure 9). In both, the low and the high MS-inoculum derived microcosms, the first principal component (PC1) discriminated the plain medium grown samples from the sucrose-supplemented samples and explained about 50% of the total variance (Figure 9A, B). Unclassified Lactobacillales (OTU #9), unclassified Firmicutes (OTU #36) and Veillonella (OTU #95) increased significantly with the sucrose supplementation; while Streptococcus (OTU #12) and Fusobacterium (OTU #42) were more abundant in the plain-medium (Suc-) microcosms irrespective of the type of the inoculum (Figure 9C, D). Only in the low MS inoculum group, Streptococcaceae (OTU #15) increased and Streptococcus (OTU #46) decreased significantly in Suc+ compared to Suc- samples (Figure 9C). Specific to the high MS group was the increase of Veillonella (OTU #7) and decrease of Megasphaera (OTU #89), Prevotella (OTU #72) and Oribacterium (OTU #178) in Suc+ samples compared to Suc- group (Figure 9D).
Figure 9. The effects of the sucrose-supplement on microcosms per type of the inoculum (low or high MS in saliva): (A, B) PCA plots with the first two components of the PCA. PC1 discriminated between samples grown in the plain medium (Suc-LGG-) marked as triangles and samples grown in sucrose-supplemented medium (Suc+LGG-) marked as circles. (C, D) Relative abundance of those OTUs which contributed most to the PC1 (values in the brackets are the PC1 loadings of the respective OTUs) in samples inoculated with low MS-saliva (C) and high MS-saliva (D). * - statistically significant difference between low and high MS samples (p<0.05; Independent samples T-test).

Finally, we addressed the effects of the probiotic strain \textit{Lactobacillus rhamnosus} GG (LGG) on the microbiome profiles. To avoid dichotomization of the samples in LGG- and LGG+ groups entirely due to the high abundance of the probiotic LGG in the sample (26% (SD 14) and 57% (SD 15) of \textit{L. rhamnosus} 16S rDNA proportion against total 16S rDNA, in Suc- and Suc+ microcosms, respectively, as reported in Pham \textit{et al}. }
(PHAM et al. 2011), we excluded single predominant OTU that was identified as LGG (OTU #94: 2% of reads in Suc-LGG+ microcosms, 15% of reads in Suc+LGG+ microcosms) from the raw dataset, normalized the data without this OTU, log2 transformed the data and performed the PCA (Figure 10). The effects of the LGG on the plain medium (Suc-) microcosms were minimal (data shown in supplementary file 3) and none of the PCA components discriminated the effect of LGG. In sucrose-supplemented microcosms, however, PC1 of the PCA discriminated clearly between LGG- and LGG+ samples, both in low (37% of variance) and in high (45% of variance) MS-inocula derived groups of samples (Figure 10A, B). Irrespective of the inoculum type, microcosms with LGG supplementation showed significant increase in unclassified Firmicutes (OTUs #186; #50), Lactobacillus (OTU #269), Veillonella (OTUs #22; #124), Lactobacillales (OTUs #255; #52), Bacilli (OTU #197) and decrease in Streptococcus (OTU #12) (Figure 10C, D). Specific to the high MS-group of samples was the decrease of unclassified Streptococcaceae (OTU #15) and genus Megasphaera (OTU #89) with LGG (Figure 10D).
Figure 10. Effects of the probiotic strain *Lactobacillus rhamnosus* GG (LGG) on the microcosms* grown in sucrose-supplemented medium per type of the inoculum (low or high MS in saliva): (A, B) PCA plots with the first two components of the PCA. PC1 discriminated between samples grown without LGG (Suc+LGG-) marked as open circles and samples grown with LGG (Suc+LGG+) marked as filled circles. (C, D) Relative abundance of those OTUs which contributed most to the PC1 (values in the brackets are the PC1 loadings of the respective OTUs) in samples inoculated with low MS-saliva (C) and high MS-saliva (D). * - statistically significant difference between low and high MS samples (p<0.05; Independent samples T-test). ** - Before this PCA analysis, OTU #94 (BLAST hit as *L. rhamnosus* and highly abundant in the samples with LGG supplementation, while absent in LGG- group) was removed from this dataset to prevent that the presence of this OTU splits the microbial profiles LGG-/LGG+ groups. For improved visualization of the graph 10C, the high standard deviation (SD 37) of average relative abundance of the OTU #12 is not shown and is indicated by an arrow.
Comparison between MLPA and pyrosequencing outcome

Of the microcosm samples that were analyzed in the current study with 454 pyrosequencing, 92 samples have also been assessed with multiplex ligation-dependent probe amplification (MLPA) technique (Pham et al. 2011). For the MLPA, we used a probe panel consisting of 20 probes, of which 18 were species-specific, one – genus specific probe targeting genus *Prevotella*, and one probe was specific for yeast species *Candida albicans*. To compare the microbial profiles of the two methods, we used the higher taxon data (genus or more inclusive taxon) that included 64 taxa (supplementary file 1) from the pyrosequencing output. Two main components of the PCA explained about 60% of the total variance in both datasets (Figure 11A, B). The PCA of the pyrosequencing data separated all six inocula of the microcosms from the microcosms groups and clustered the plain-medium grown microcosms (triangles in Figure 11A) from the microcosms grown with sucrose supplement (circles in Figure 11A), while the PCA on the MLPA data showed stronger discrimination between the LGG-containing groups (filled symbols in Figure 11B) and the microcosms without LGG (open symbols) than the sequencing data.

Relative abundance of genus *Prevotella*-specific probe in the MLPA dataset correlated significantly (Spearman’s rho 0.836; p<0.001) with the proportion of reads classified as *Prevotella* in the pyrosequencing dataset (Figure 12). The probability of the pyrosequencing method to detect *Prevotella* in the sample (sensitivity of the method) was 0.72, while the specificity (probability that the sample does not contain any *Prevotella*) of the pyrosequencing was 0.86 against the MLPA (Table 2). All but two of the 14 false negatives of the pyrosequencing method were found at a low proportion (0.1 to 1%) in the MLPA samples. The remaining two samples with no reads classified as *Prevotella* contained 4.8 and 8.3% of *Prevotella* according to the MLPA. False positive samples (N=6) of the pyrosequencing all were found at a very low proportion (0.003 – 0.025% of the reads).
Figure 11. PCA plots of the first two principal components (PC1 and PC2) of (A) the final taxon data (64 taxa at genus or higher level) obtained by pyrosequencing and (B) by MLPA using 20 MLPA probe-set (PHAM et al. 2011). Symbols in the plots: cross – inoculum, open triangle – Suc-LGG--; filled triangle – Suc-LGG--; open circle – Suc+LGG--; filled circle – Suc+LGG++.
Figure 12. The relative abundance of genus *Prevotella* specific MLPA-probe by the relative abundance of genus *Prevotella* as a final taxon in the 454 pyrosequencing dataset per individual sample (N=92). Correlation was statistically significant (Spearman’s rho 0.836; p<0.001).

Table 2. Detection of genus *Prevotella* in the microcosms (N=92) by MLPA and by pyrosequencing.

<table>
<thead>
<tr>
<th>Prevalence of <em>Prevotella</em> in microcosms</th>
<th>Pyrosequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>36</td>
</tr>
<tr>
<td>-</td>
<td>6</td>
</tr>
</tbody>
</table>

MLPA
DISCUSSION

This is, to our knowledge, the first time that effects of inoculum and different growth conditions on saliva-derived microcosms have been assessed by the 454 pyrosequencing approach. This comprehensive parallel sequencing of short fragments of the 16S rDNA gene allowed a non-targeted, open-ended profiling of complex microbial communities of unknown composition.

Pre-screening of the saliva donors for culturable mutans streptococci (MS) allowed us to dichotomize the inocula into low and high MS group. The two groups showed marked differences already at the phylum level. OTU-profile analyses by the principal components confirmed the differences between the two types of inocula. The major differences were in the relative abundance of reads classified as *Neisseria*, *Porphyromonas* (all significantly more abundant in the low MS-inocula) versus *Prevotella*, which was more abundant in the high MS-inocula. Surprisingly, none of the significantly different OTUs could be classified as *Streptococcus mutans* while both culturing on a selective plate and performing a specific qPCR probe counts showed significantly higher mutans streptococci abundance in the high MS group (Pham et al. 2011). It could be due to a relatively low proportion of this organism in saliva in combination with a high sensitivity of targeted methods, or due to the incomplete reference sequences in publicly available 16S rDNA databases.

The microbial profiles of the saliva used as inocula were typical for the salivary microbiome when compared to the salivary profiles obtained in our previous studies using pyrosequencing (Keijser et al. 2008; Zauroa et al. 2009) both by the predominant taxa and by the relative abundance of these taxa. Microcosms, however, differed significantly from the inocula, with representatives of the phylum Firmicutes dominating in all samples. This finding was not surprising because of the selectivity of the growth media employed to simulate saliva (Tian et al. 2010). Tian and colleagues (Tian et al. 2010) performed elaborate series of experiments in search of the “perfect” medium to sustain the original oral communities in vitro. The aim of the experiments in Pham et al (Pham et al. 2011), however, was not to sustain the original community in all its complexity but to address the effects of probiotic strain *Lactobacillus rhamnosus* GG (LGG) on the composition, hence cariogenicity, of the
microcosms. Therefore we used a relatively simple medium (McBain et al. 2005) which supports the growth of the phylum Firmicutes and thereby induces a severe cariogenic challenge mainly by sucrose- and lactate-metabolizing members of the *Streptococcaceae, Lactobacillaceae* and *Veillonellaceae* family of phylum Firmicutes.

Interestingly, the microcosms grown without sucrose supplementation showed a double number of taxa (OTUs) compared to the original inocula or to the microcosms grown with sucrose supplementation. In total, 272 OTUs, though at low abundance, were exclusively found in these plain medium microcosms. The artificial saliva we used must have supported the growth of diverse groups of salivary microorganisms that were below the detection limit in the inoculum, but could establish itself in the 72-h biofilms grown without the caries-inducing challenge.

Another interesting finding was the ability of the representatives of candidate division TM7 to establish in a majority of the plain-medium microcosms and to a lesser degree, in the sucrose-supplemented microcosms. TM7 is a diverse bacterial division (equivalent to a phylum but so far uncharacterized), found in terrestrial, aquatic and clinical habitats, and is known only from environmental 16S rDNA sequence data without any cultivated members (Hugenholtz et al. 2001; Ouverney et al. 2003). It has been associated with mild periodontitis (Brinic et al. 2003) and with active inflammatory bowel disease (Kuehbacher et al. 2008).

Addition of sucrose to the growth medium introduced a severe cariogenic challenge: the pH of the spent medium decreased from pH 7 to pH 5, and the dentin substratum onto which the microcosms were grown showed substantial mineral loss (Pham et al. 2011) irrespective of the inoculum (high or low MS group) for the microcosms. Diversity of the cariogenic microcosms was significantly lower than the diversity of the plain-medium communities, supporting our previous findings (Pham et al. 2009) and clinical reports (Li et al. 2007; Li et al. 2005). The cariogenic communities of this study were dominated by *Veillonellaceae* family members.

*S. mutans* has been associated with dental caries (Loesche 1986; Loesche et al. 1975) and the counts of MS in saliva are used in caries risk assessments (Zhang et al. 2007). In Pham et al. (Pham et al. 2011) we hypothesized that in the absence of *S. mutans* other cariogenic microorganisms must have become abundant under these highly
Pyrosequencing analysis of human saliva and microcosms
cariogenic conditions. The comparison of the microcosms evolving from the low and
the high MS inocula showed that these two groups formed two distinguishable
microbial profiles. The major difference was in the increase of unclassified
Streptococcaceae (OTU #15) and decrease of one representative of genus Streptococcus
(OTU #46) in the low MS communities, and the decrease of Megasphaera (OTU #89),
Prevotella (OTU #72) and Oribacterium (OTU #178) in the high MS microcosms in the
presence of sucrose. Based on these differences, the OTU #15 may have been
involved in cariogenic activity of the low MS microcosms grown with sucrose
supplementation.
The OTUs that increased in proportion in both (low and high MS) microcosm types
if exposed to sucrose, were unclassified Lactobacillales, unclassified Firmicutes and
representatives of genus Veillonella. The presence of sucrose promoted the growth of
Veillonella (OTU #7 in the high MS group and OTU #95 in the both MS groups). The
OTU #7 was highly dominant in the low MS group samples grown without sucrose
(on average, 40% of the reads) and it increased non-significantly in the presence of
sucrose (about 50% of reads), while this particular OTU increased in abundance from
25% (plain medium) to 60% (sucrose-supplemented medium) of the reads in the high
MS group. Veillonellae and streptococci are metabolically linked through
streptococcal fermentation of sugars to lactic acid, which is a carbon source for the
nonsaccharolytic veillonellae with acetate as the main metabolic endproduct
(Chalmers et al. 2008; Delwiche et al. 1985). For this reason veillonellae are
considered non-cariogenic microorganisms. However, both acetate and lactic acid and
their buffers will induce demineralization of enamel and dentin. The rate of
demineralization increases with decreasing pH. Considering buffering in the oral
cavity, demineralization typically occurs at pH values corresponding to the pK
values of the acids present. Acetic acid has a pK value of 4.5, and for lactic acid it is
one pH unit lower. Dentin, as used in this study, has a higher solubility than enamel,
and this tissue will start to dissolve already around pH=6 (Hoppenbrouwers et al.
1987). This pH is reached quickly when acids are formed in the dental plaque
biofilm. In addition, it has been shown that acetate, unlike lactate, penetrates and
thereby disappears into porous dental tissue such as dentin (Geddes et al. 1984).
Since lactate breakdown into acetate would not be inhibited by a build-up of acetate,
Veillonella species would benefit from this process and thus would have an ecological advantage. Recent clinical studies have demonstrated the association of Veillonella species with severe early childhood caries (Becker et al. 2002; Kanasi et al. 2010; Marchant et al. 2001), deep dentinal lesions (Lima et al. 2011), as well as with early and established cavities in children (Aas et al. 2008; Ling et al. 2010). Kanasi et al. (Kanasi et al. 2010) explained the association of Veillonella with severe dental decay by a lactate-rich environment of the carious dental tissue, and proposed for veillonellae a key role in supporting the biofilm of carious infection.

To assess the effects of probiotic strain Lactobacillus rhamnosus GG (LGG) on the microcosms, we removed a single predominant OTU (OTU #94) that we identified as L. rhamnosus prior to principal component analysis (PCA). The PCA showed that addition of the LGG to the inoculum affected the microcosms that were grown in the presence of sucrose, but not in the plain medium. A potential explanation for this difference might be in the low relative abundance of LGG that was found in the plain-medium microcosms compared to the sucrose-supplemented group (Pham et al. 2011). Although LGG does not ferment sucrose, it preferred growth at pH 5 compared to pH 7. In the presence of LGG, several other Firmicutes (Veillonella, Lactobacillus and unclassified members of Lactobacillales, Bacilli and Firmicutes) increased or decreased (Streptococcus, Streptococcaceae, Megasphaera) in their relative abundance. The taxonomic resolution of the current pyrosequencing reads (on average 370 nt, spanning over the hypervariable region V5-V7 of 16S rDNA) and the limitations of publicly available 16S rDNA reference databases (e.g., over-representation of cultured and isolated microorganisms; erroneous sequences due to lack of continuous and updated curation) did not allow us to identify the involved OTUs. One of the OTUs (#15, unclassified Streptococcaceae) that reduced in abundance in high MS microcosms supplemented with sucrose and LGG, was significantly more abundant in the high MS inocula, and increased in abundance in low MS-inocula derived microcosms if grown with sucrose supplementation. We could not confirm that this OTU is Streptococcus mutans (BLAST search resulted in alignment at 96% identity with Streptococcus mutans UA159 sequence), but could not exclude this probability either. To clarify this we may need to sequence numerous
clinical isolates of *S. mutans* and to develop a custom reference database of 16S rDNA sequences of this clinically relevant microorganism.

In our previous study we applied the multiplex ligation-dependent probe amplification (MLPA) technique to profile the same set of experimental and salivary samples with 20 specific probes (PHIAM et al. 2011). The comparison of both techniques indicates that targeting only limited number of microorganisms does not cluster the experimental groups as clearly as the open-ended approach by pyrosequencing. Nevertheless, we could confirm the effects of the individual inocula (high and low MS saliva), plain and sucrose-supplemented medium and the addition of LGG on microcosms using either of the methods. The presence of the genus *Prevotella*-specific probe in the MLPA probe panel allowed us to correlate the relative proportion of this probe in the samples with the proportion of the genus *Prevotella* taxon in the pyrosequencing dataset of the same samples. In 78% of the samples both methods agreed on the presence or absence of *Prevotella*. Sensitivity of the pyrosequencing method is highly dependent on the sequencing depth. In the current study individual samples gave between 1,277 and 8,261 reads (average 3,863). A single read classified as *Prevotella* would contribute to 0.01 - 0.08% (average 0.03%) of the reads of the individual sample. In the samples where *Prevotella* was prevalent by both methods, we found between 0.015% and 12% (average 1.9%) reads/sample classified as *Prevotella*. This clearly shows that by increasing of the number of reads per sample we are able to increase the sensitivity of the pyrosequencing method. In two samples, we found relatively high proportion (4.8% and 8.3%) of *Prevotella* by MLPA but did not find such presence by pyrosequencing. This could be explained by the methodology: MLPA probe of genus *Prevotella* was designed to target exclusively this genus (nt position 579 – 636 of *E. coli* 16S rDNA), while with pyrosequencing all microorganisms were targeted (nt position 785 – 1175 of *E. coli*). For pyrosequencing, the sequenced part of 16S rDNA covers three hypervariable regions (V5-V7) flanked by conserved sequences. In cases where taxonomical classification could not be assigned to the genus level with at least 80% probability, the sequences were assigned to a higher taxon (family, order, class or phylum). This may have resulted in the observed differences at the genus level.
The use of the open-ended microbial profiling approach, 454 pyrosequencing, allowed us to disclose the differences between natural microbial communities (individual salivas) and their *in vitro* evolved counterparts - microcosms. We observed clear differences in the microbial profiles between the two types of inocula (low and high mutans streptococci counts in saliva), supporting the ecological plaque hypothesis. We showed that microcosms derived from these two types of predefined inocula remained distinct at different growth conditions. We conclude that selection of the inoculum for microcosm experiments should be based on defined criteria, which would depend on the aim of the study. The low taxonomic resolution of the pyrosequencing sequences limited the characterization of the microcosms to the OTU level or to higher taxa with maximum resolution at the genus level. This limitation, however, could be improved by the use of a custom-made and curated reference sequence database. Comparison of targeted profiling by MLPA with high throughput sequencing by 454 pyrosequencing showed that the MLPA is a valid method in microcosm profiling. The sequences generated by the open-ended approach will allow relevant probe selection for less costly, targeted community profiling methods.

**ACKNOWLEDGMENTS**

This research was supported by the Dutch Technology Foundation STW, Applied Science Division of NWO and the Technology Program of the Ministry of Economic Affairs (project number 7069). We also thank GABA International AG and MRC-Holland B.V. for financial and technical support. We thank all volunteers participating to this study.