

RESEARCH

Open Access

# Analysis of the salivary microbiome using culture-independent techniques

Vladimir Lazarevic<sup>1\*</sup>, Katrine Whiteson<sup>1</sup>, Nadia Gaïa<sup>1</sup>, Yann Gizard<sup>1</sup>, David Hernandez<sup>1</sup>, Laurent Farinelli<sup>2</sup>, Magne Østerås<sup>2</sup>, Patrice François<sup>1</sup> and Jacques Schrenzel<sup>1</sup>

## Abstract

**Background:** The salivary microbiota is a potential diagnostic indicator of several diseases. Culture-independent techniques are required to study the salivary microbial community since many of its members have not been cultivated.

**Methods:** We explored the bacterial community composition in the saliva sample using metagenomic whole genome shotgun (WGS) sequencing, the extraction of 16S rRNA gene fragments from metagenomic sequences (16S-WGS) and high-throughput sequencing of PCR-amplified bacterial 16S rDNA gene (16S-HTS) regions V1 and V3.

**Results:** The hierarchical clustering of data based on the relative abundance of bacterial genera revealed that distances between 16S-HTS datasets for V1 and V3 regions were greater than those obtained for the same V region with different numbers of PCR cycles. Datasets generated by 16S-HTS and 16S-WGS were even more distant. Finally, comparison of WGS and 16S-based datasets revealed the highest dissimilarity.

The analysis of the 16S-HTS, WGS and 16S-WGS datasets revealed 206, 56 and 39 bacterial genera, respectively, 124 of which have not been previously identified in salivary microbiomes. A large fraction of DNA extracted from saliva corresponded to human DNA. Based on sequence similarity search against completely sequenced genomes, bacterial and viral sequences represented 0.73% and 0.0036% of the salivary metagenome, respectively. Several sequence reads were identified as parts of the human herpesvirus 7.

**Conclusions:** Analysis of the salivary metagenome may have implications in diagnostics e.g. in detection of microorganisms and viruses without designing specific tests for each pathogen.

## Background

The microbiota in the mouth has a significant impact on both the oral and general health. Bacterial species associated with periodontal health and those that are more prevalent in periodontal disease have been identified [1]. The salivary microbiota is a potential diagnostic indicator of several diseases. For instance, a caries-free oral status in children is associated with a significant shift in the relative abundance of *Porphyromonas catoninae* and *Neisseria flavescens* in saliva [2]. Increased salivary counts of *Capnocytophaga gingivalis*, *Prevotella melaninogenica* and *Streptococcus mitis* are associated

with oral cancer [3]. The salivary level of the bacterium *Selenomonas noxia* correlates with obesity in women [4].

The study of the oral microbiota as well as its salivary component requires culture-independent techniques, since about one third of 700 bacterial species identified in the human oral cavity have not been cultivated [5]. These may be based on PCR amplification and high-throughput sequencing of the bacterial 16S rRNA genes (16S-HTS) or the metagenomic whole genome shotgun (WGS) sequencing. The latter approach may include either the analysis of the totality of generated DNA fragments or of the 16S rRNA gene fragments retrieved from the metagenome (16S-WGS) [6]. Both 16S-WGS and 16S-HTS approaches present limitations and advantages over each other [6].

Here we explored the microbial community composition in the saliva sample using WGS, 16S-WGS and 16S-HTS. In addition, to assess putative biases due to

\* Correspondence: vladimir.lazarevic@genomic.ch

<sup>1</sup>Genomic Research Laboratory, Division of Infectious Diseases, Geneva University Hospitals, Rue Gabrielle-Perret-Gentil 4, CH-1211 Geneva 14, Switzerland

Full list of author information is available at the end of the article

PCR amplification, we compared taxonomic composition of 16S-HTS datasets obtained after different number of PCR cycles.

## Methods

### Sampling

The study was conducted according to the current version of Declaration of Helsinki and approved by the Ethics Committee of HUG (09-078). Unstimulated saliva was obtained with informed consent from a 32-year male smoker without obvious signs of oral disease. The sample was collected by spitting in a sterile plastic 50-mL tube at 10:30 a.m., 1.5 hours after eating. Six hundred  $\mu$ L saliva was mixed with the same volume of 2x lysis buffer [Tris 20 mM, EDTA 2 mM (pH 8), Tween 1%] and Proteinase K (Eurobio) 200  $\mu$ g/mL. After a 2.5 hour incubation at 55°C, proteinase K was inactivated by a 10-min heating at 95°C. The saliva lysate was divided in six 200- $\mu$ L aliquots to which RnaseA (Roche) 40  $\mu$ g/mL was added. Samples were incubated for 5 min at room temperature. From that point, the DNeasy Blood & Tissue Kit (QIAGEN) was used following the manufacturer's Spin-Column Protocol for Purification of Total DNA from Animal Blood or Cells (DNeasy Blood & Tissue Handbook 07/2006). DNA was eluted using 110  $\mu$ L of supplied AE Buffer, then the pooled eluate (metagenomic DNA) was concentrated to 80 ng/ $\mu$ L. Total DNA quantity was assessed using a NanoDrop ND-8000 spectrophotometer (NanoDrop Technologies).

### PCR and sequencing

PCR amplification was carried out in a 50- $\mu$ L PrimeStar HS Premix (Takara) containing 8 ng of purified DNA and 0.5  $\mu$ M of each forward and reverse primer. The 16S rDNA V1-3 amplicons generated with primers 5'-GAGTTTGATCMTGGCTCAG (V1 forward) and 5'-CCGCGRCTGCTGGCAC (V3 reverse) corresponded to *E. coli* positions 28 to 514 after exclusion of primers sequences. The samples were run in four replicate PCRs for 20, 25 or 30 cycles using the following parameters: 98°C for 10 s, 60°C for 15 s, and 72°C for 1 min. The four replicate PCRs were then pooled.

Paired-end DNA libraries were prepared according to the manufacturer's (Illumina) instructions. Metagenomic DNA fragments of about 300 bp and 16S rDNA amplicons were barcoded using specific 6-base sequences. The libraries were sequenced from both ends for 100 cycles (excluding barcode sequences) on the Illumina Hi-Seq 2000 using TruSeq SBS v5 kit. A barcoded PhiX reference was spiked in the same channel to estimate the error rate.

### Sequence filtering

Parameters of the initial quality filter were the following: (i) maximum one base below a quality of 5 in the first 70

bases; (ii) a minimum average quality of 10; (iii) no ambiguous base allowed. After filtering, the average Q30 was larger than 75% and the average PhiX error rate was 0.7%. Only pairs were retained in the filtered data, i.e. if one read was filtered out the paired read was removed. Each of the three 16S-HTS datasets (20, 25 and 30 PCR cycles) was reduced by randomly picking 1.2 million sequence read pairs. Then, in the second filtering step, we removed sequences containing incorrect PCR primer sequences or runs of  $\geq 12$  identical nucleotides. The WGS dataset was reduced to one million sequence pairs and was not subject to additional filtering steps. Sequences were deposited in MG-RAST under accession numbers 4477823.3, 4477824.3, 4477839.3, 4477840.3, 4478078.3, 4478079.3, 4478080.3, 4478370.3, 4478371.3, 4479520.3, 4479521.3, 4479522.3, 4479523.3 and 4479524.3.

### Analysis

The 16S rDNA sequences were clustered to operational taxonomic units (OTUs) defined at 95% identity using CD-HIT [7]. The V1 and V3 sequences were assigned the taxonomic identity using the Ribosomal Database Project (RDP) Classifier [8] with a recommended 50% confidence cutoff. Taxonomic assignments of sequences from the WGS dataset were made using BLASTN [9] against NCBI prokaryotic, viral and fungal databases as well as against the human sequences from NCBI and EBI databases. The criteria used were a wordsize of 16,  $\geq 94\%$  identity,  $\geq 90\%$  overlap and  $e\text{-value} \leq 10^{-30}$ . The bacterial 16S rDNA sequences were extracted from the WGS dataset using CAMERA [10] and HMMER search option. They were then filtered using an  $e\text{ value} \leq 10^{-10}$  and assigned to genera using the RDP Classifier.

Group-average clustering of data was performed using a Bray-Curtis similarity matrix in PRIMER-E (Plymouth), based on square-root-transformed genera abundance.

## Results and Discussion

### Illumina sequencing

We explored the microbial community composition in the saliva sample from a male healthy adult using Illumina sequencing. The 100-base paired reads from the whole metagenome fragments as well as 81-base V1 and 84-base V3 reads of 16S rDNA amplicons were analyzed (Table 1). Data from the forward and reverse run for each pool of DNA fragments were analyzed separately since it has been reported that reverse Illumina reads are of lower quality than those from the forward run [11].

### Taxa abundance as a function of the PCR cycle number in the 16S-HTS datasets

Taxa detection and the accuracy of 16S rDNA abundance measurement are affected by the number of PCR cycles used to amplify 16S rDNA from a bacterial community

**Table 1 Description of the 4 sequence datasets and 14 subsets**

	Dataset							
	16S-HTS, 20 cycles		16S-HTS, 25 cycles		16S-HTS, 30 cycles		WGS	
<b>Generated reads</b>	16,345,598		10,732,108		6,150,466		14,724,838	
<b>Randomly chosen read pairs</b>	1,200,000		1,200,000		1,200,000		1,000,000	
<b>16S rDNA region subset</b>	V1	V3	V1	V3	V1	V3		
<b>Filtered forward reads</b>	592,322	566,310	576,701	564,774	546,710	569,459	1,000,000	
<b>Filtered reverse reads</b>	532,945	530,504	534,229	523,548	548,009	495,531	1,000,000	

[12-14]. To investigate how taxonomic composition of the same sample differed depending on the number of PCR cycles, we analyzed forward and reverse reads from V1 and V3 16S-HTS datasets obtained after 20, 25 and 30 cycles. The taxonomic assignment was performed using RDP Classifier (Additional file 1). The changes in proportions were relatively consistent for the different taxa within the same phylum (Figure 1): the relative abundance of sequence reads assigned to taxa belonging to the phyla Actinobacteria and Firmicutes generally decreased with more PCR cycles, whereas the proportion of sequences representing other phyla and their corresponding lower-level taxa generally increased. Instances in which the same direction of change (decrease or increase) occurred across all 8 subsets were found in 63 of the 109 taxa (Figure 1). Moreover, in 95% of cases where a > 25% change in taxa abundance was found in 30- vs 20- cycle-samples, the value obtained after 25 cycles was intermediate relatively to those obtained after 20 and 30 cycles. It seems likely that, for some taxa, more PCR cycles will further increase the bias. We found that the average Firmicutes to Bacteroidetes ratio, which may be an indicator of obesity in intestinal microbiota [15], was on average 3.6 (range 3.4-3.9), 2.3 (range 2.2-2.4) and 2.0 (range 1.9-2.2) in 20-, 25- and 30-cycle 16S-HTS datasets, respectively.

There is a concern that short Illumina reads and sequence errors may compromise the quality of taxonomic assignments [16-18]. To assess the accuracy of taxonomic assignments we extracted 81-base V1 and 84-base V3 sequences from the 16Sr RNA gene for 660 species from the Human Oral Microbiome Database (HOMD) [5] for which the taxonomic information was available at the genus level. These simulated Illumina reads were assigned taxonomy using the RDP Classifier with a recommended 50% bootstrap cutoff. The proportion of V1 and V3 sequences correctly assigned at the genus level reached 68% and 76%, respectively (Additional File 2). For both, V1 and V3 regions, the accuracy of taxonomic assignment at the phylum level was greater than 95%.

We clustered sequence reads generated by Illumina sequencing into OTUs, defined at  $\geq 95\%$  identity, which roughly corresponds to genus-level grouping [19] and may have the effect of absorbing some sequence errors

[20]. Then, we compared the OTU content across datasets obtained after different number of PCR cycles using the phylum-level affiliation of representative OTUs derived from the RDP Classifier. The OTUs that met the criteria described in Additional file 3 were selected for comparisons. This approach confirmed the trend observed when taxonomy was assigned to each sequence read (see above); the relative abundance of the majority of OTUs from the phyla Actinobacteria and Firmicutes was decreased whereas the proportion of most OTUs from the phyla Bacteroidetes, Fusobacteria and Spirochaetes was enhanced by increasing the number of PCR cycles (Additional file 3).

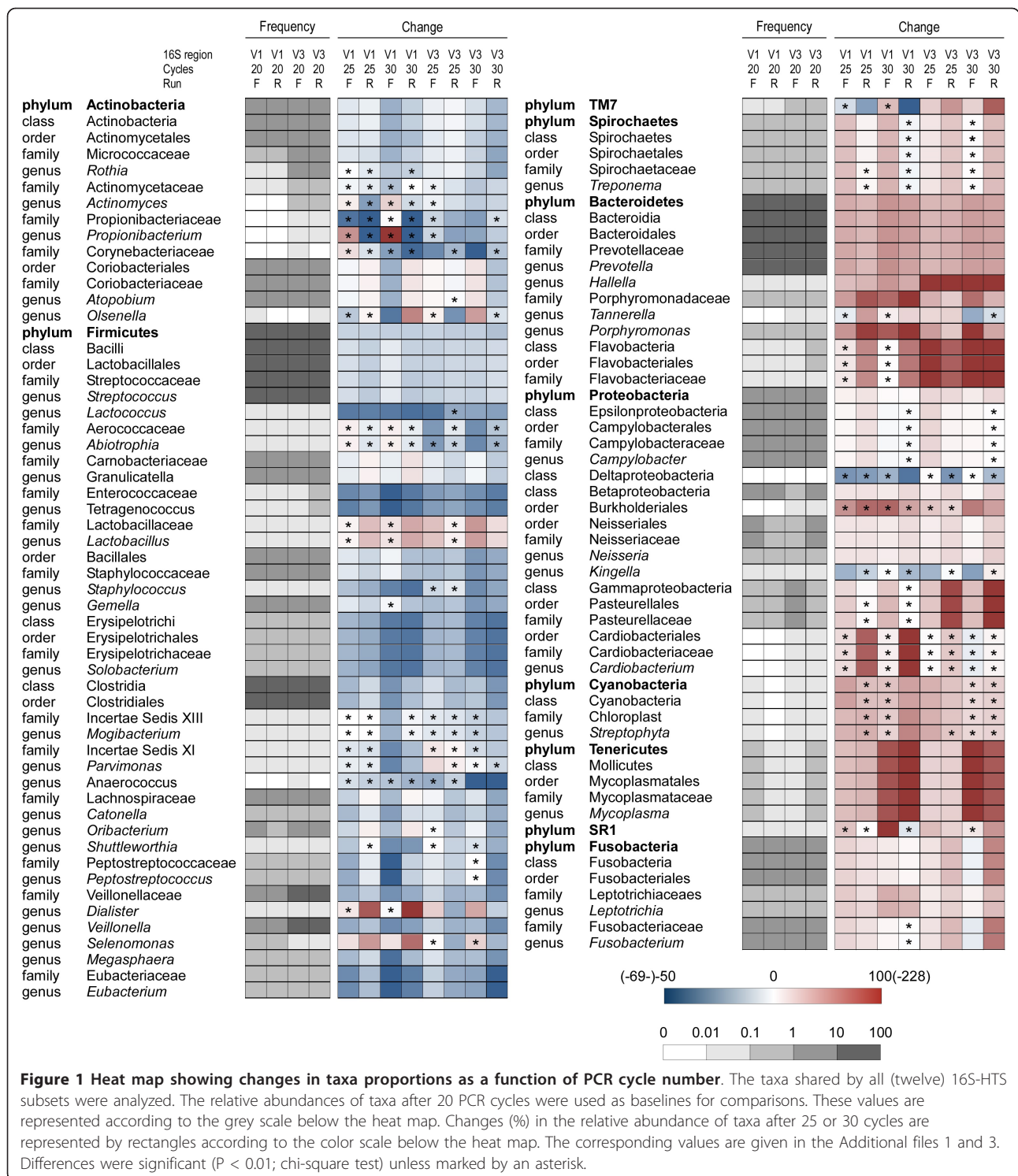
Therefore, performing more PCR cycles, which may be required when little template DNA is available, may introduce amplification biases and increase the distance from samples for which less PCR cycles were performed.

#### Taxonomic assignment in the WGS datasets

In the WGS approach, the taxonomic assignments were inferred from BLASTN searches of individual sequence reads. Sequences were compared to human genomic sequence as well as to databases containing completely sequenced prokaryotic, viral and fungal genomes. Most of the BLASTN hits corresponded to human DNA, whereas bacterial and viral sequences represented 0.73% and 0.0036% respectively (Table 2). Forward reads performed better in terms of assignment yield to these three categories, reflecting a higher sequence quality in comparison to the reverse reads [11]. A total of 369 and 367 16S rRNA gene fragments were extracted from WGS forward and reverse datasets, respectively, using CAMERA.

#### Comparison of the 16S-HTS and WGS datasets

We performed hierarchical clustering of WGS and 16S datasets based on the relative abundance of bacterial genera. Before computing the similarities, we applied a square-root-transformation of the relative abundance data in order to equilibrate the impact of abundant and rare genera. The resulting dendrogram (Figure 2) shows two main clusters that correspond to WGS and 16S-based datasets. The latter is divided into 2 subclusters corresponding to 16S-HTS and 16S-WGS approach. The 16S-HTS subcluster further splits according to the V region



sequenced. Finally, for a given V region, datasets generated by sequencing PCR products obtained after 20 amplification cycles were separated from their 25- and 30-cycle counterparts which clustered together. The highest variation between the forward and reverse datasets was found

for the 16S-WGS approach, which may be due to a rather small number ( $< 400$ ) of sequences analyzed in comparison to the two other approaches (Table 1) and the fact that the forward and reverse reads in the 16S-WGS dataset cover different, randomly distributed segments of 16S



**Table 2 Number of BLASTN hits against human, bacterial and viral databases**

Database	BLASTN hit <sup>a</sup> counts		R/F ratio
	F	R	
Human <sup>b</sup>	801677	642018	80.0
Bacteria <sup>c</sup>	7898	6807	86.7
Viruses <sup>d</sup>	38	34	89.5
Fungi <sup>e</sup>	0	0	-

F, forward run; R, reverse run.

<sup>a</sup> The following BLASTN parameters were used: wordsize = 16, e ≤ 10<sup>-30</sup>, sequence identity ≥ 94%, alignment length ≥ 90 nt.

<sup>b</sup> combined entries from ftp://ftp.ebi.ac.uk/pub/databases/fastafiles/emblrelease/em\_rel\_std\_hum.gz as of 4 March 2011, and ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/human\_genomic.gz as of 12 March 2011.

<sup>c</sup> ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/ as of 16 November 2010.

<sup>d</sup> ftp://ftp.ncbi.nlm.nih.gov/genomes/Viruses/ as of 25 November 2010.

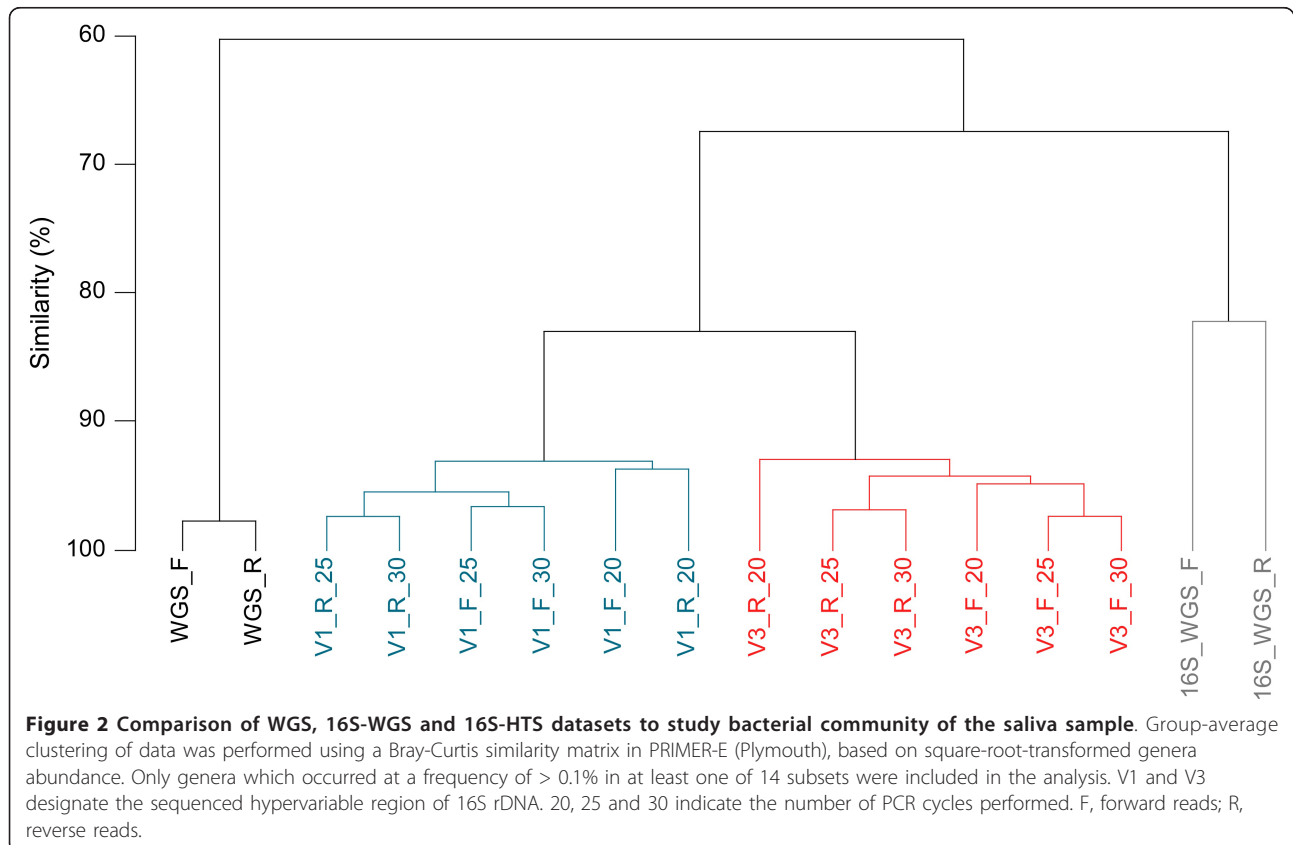
<sup>e</sup> ftp://ftp.ncbi.nlm.nih.gov/genomes/Fungi/ as of 3 March 2011.

rRNA genes which may lead to variation in taxonomic assignments.

In order to confirm that the pattern observed is reproducible across individuals, it would be necessary to analyze a larger number of salivary microbiomes using the same methodologies. As a first step towards addressing this issue, we extracted the 84-base V3 16S rDNA sequences from the 5 salivary microbiomes reported in

a previous study [21]. The relevant data were then included in the construction of the similarity matrix as described above (not presented). We found that the average similarity between the V3 dataset determined in the current study and the five published microbiomes was comparable to the average similarity observed among these five microbiomes (Table 3). Relative to these values, our V3 16S-HTS datasets showed higher degrees of similarity when compared to our V1 16S-HTS and 16S-WGS datasets (Table 3) i.e. the interindividual differences may outweigh to some extent the methodological differences. Still, differences in the genera distribution inferred from 16S-WGS, 16S-HTS and WGS datasets (Table 3) are too great to allow for reliable comparisons of the results generated using different methodologies. However, as the reference microbial genome database will grow, the WGS and 16S-WGS methodologies will likely provide more closely related data.

We identified 206 bacterial genera using 16S-HTS, 108 of which have not been previously found in salivary microbiomes using culture-independent techniques [18,21-24]. This was also the case with 19 out of 56 genera determined by WGS, and 6 out of 39 genera identified by 16S-WGS approach. The majority of the new salivary genera (116/124) were found at a frequency < 0.1%, and only 8 occurred at a frequency between 0.1 and 0.74%



**Table 3 Similarity between the V3 16S and other datasets**

Datasets to which the F and R 30-cycle V3 subsets were compared	Average similarity <sup>a</sup> ± SD (%)	Pearson's correlation coefficient (F, R) <sup>b</sup>
Itself (F vs R)	94.8	0.998
25-cycle V3	97.1 ± 0.3	0.998, 0.999
20-cycle V3	93.2 ± 1.5	0.974, 0.977
30-cycle V1	82.7 ± 0.8	0.987, 0.990
16S-WGS	68.5 ± 3.7	0.933, 0.958
WGS	64.1 ± 0.5	0.650, 0.652
V3 from other individuals <sup>c</sup>	64.8 ± 3.5	0.453-0.871, 0.458-0.867

F, forward run; R, reverse run.

<sup>a</sup> The genera frequency were square-root transformed and used to construct the Bray-Curtis similarity matrix. When applicable, the similarities were calculated separately for the pairs of forward and the pairs of reversed subsets prior to averaging.

<sup>b</sup> Based on untransformed data.

<sup>c</sup> Samples 1.1, 2.1, 3.1, 4.1 and 5.1 from a previous study [21] in which the initial PCR cycles were carried out in the same conditions as those in the present work. Then, 28 cycles were performed using bar-coded primers and the amplicons were sequenced by the Roche/454 platform. The average similarity observed among these five microbiomes was 65.2 ± 7.2%.

(Additional files 1 and 4). This suggests that the most abundant bacterial genera in the saliva of healthy subjects have probably already been identified. However, the inventory and dynamics of low-abundance-genera, whose identification requires a deeper sample coverage, remain largely unknown.

Using WGS sequencing, which, in contrast to the 16S-HTS method applied in this study, does not specifically target bacteria, we did not detect archaea in the saliva sample. This is not surprising since the only archaeon identified so far in the human oral cavity i.e. *Methanobrevibacter oralis*, was found in dental plaques associated with pathological processes [25]. BLASTN similarity search against fungal genomes of the NCBI database did not yield any significant hits. The reasons for this may be: (i) an inefficient disruption of fungal cells by the enzymatic procedure used to release DNA from bacteria; (ii) the presence of fungi in saliva under the detection level; (iii) the absence of the relevant fungal genomes in the database. So far, sequences of only six fungal genera (*Zygosaccharomyces*, *Penicillium*, *Gibberella*, *Saccharomyces*, *Aspergillus*, *Candida*), present in the oral cavity of healthy individuals [26], are available in public databases.

#### Salivary virome

Based on BLASTN comparisons to the NCBI virus database, we identified sequences possibly derived from three different eukaryotic viruses. The most abundant (Table 4) was human herpes virus 7 with 17 and 15 sequence reads in the forward and reverse WGS datasets, respectively. A sequence similar to porcine endogenous retrovirus and the virus of the green alga *Chlorella* were identified as well. Several sequences produced the best BLASTN hits to bacteriophages including *Streptococcus* phage SM1 and two enterobacterial phages, lambda and phiX174.

#### Clinical applications

Metagenomics has the potential to serve as a viral and bacterial infection control strategy in clinical practice because it can discover known as well as new pathogens, and might soon replace many existing typing methods in diagnostics.

HTS of cDNA has already been successfully applied to the detection of new viral pathogens in human serum and liver as well as in the reconstruction of viral genomes [27-29]. Similarly, WGS of a patient's feces samples detected the bacterial pathogen *Campylobacter jejuni* during but not after an acute diarrheal episode [30].

At least six double-stranded DNA human herpes viruses (HHV) i.e. Herpes simplex virus 1, Epstein-Barr virus, cytomegalovirus and human herpesviruses 6, 7 and 8 have been detected in saliva using sensitive PCR assays [31]. These viruses are shed in saliva asymptotically which could facilitate their transmission. Most human adults are infected with HHVs but the prevalence of some HHV is significantly higher in HIV-seropositive

**Table 4 Detection of viral sequences in the WGS dataset using BLASTN**

Virus	Number of hits <sup>a</sup>	
	F	R
Human herpesvirus 7	17	15
Porcine endogenous retrovirus E	1	-
<i>Paramecium bursaria</i> chlorella virus-1 FR483	1	1
<b>Phage</b>		
Enterobacteria phage lambda	12	11
Enterobacteria phage phiX174	6	7
<i>Streptococcus</i> phage SM1	1	-

F, forward run; R, reverse run.

<sup>a</sup> The following BLASTN parameters were used: wordsize = 16, e ≤ 10<sup>-30</sup>, sequence identity ≥ 94%, alignment length ≥ 90 nt.

persons. In subjects with recurrent oral Herpes simplex virus 1 infections, two other HHVs, HHV-6 and HHV-7 were simultaneously present with a frequency of over 93% [31]. HHVs possibly contribute to periodontitis which, in turn, facilitates virus shed into saliva [32]. Recently, using a metagenomic approach Willner et al. [33] identified Epstein-Barr virus in a pool of oropharyngeal swabs from 19 individuals.

In our study WGS sequencing applied on the salivary metagenome allowed identification of sequences showing the best similarity to the human herpesvirus 7 as well as to the putative periodontopathic bacteria *Porphyromonas gingivalis*, *Treponema denticola* and *Aggregatibacter actinomycetemcomitans* [34].

The exact role of bacteria and viruses in periodontitis and other oral diseases is not elucidated. It has been hypothesized that bacteria and viruses cooperate to provoke the disease [35]. The detection of periodontopathic agents is important because periodontitis has been associated with other health problems such as cardiovascular diseases, premature delivery, rheumatoid arthritis and cancer [35].

Although a large fraction of DNA extracted from saliva corresponds to human DNA, we estimate that, at a coverage consisting of a hundred million sequences, which is the current capacity per channel on the Illumina platform, hundreds of thousands of bacterial sequences and thousands of viral and phage sequences may be identified. Therefore, detection of viruses by metagenomic sequencing is possible even without including filtration and concentration steps, although these procedures are effective in enriching the metagenomic samples for viral DNA [33]. In addition, tens of thousands of 16S rDNA sequences, free of amplification anomalies, may be extracted from huge WGS datasets and used to assess taxonomic composition of bacterial communities.

## Conclusions

Analysis of the salivary microbiome is not only of interest from a fundamental perspective, but may have implications in diagnostics e.g. in detection of viruses and microorganisms without including specific tests for each pathogen.

In our study, WGS sequencing compared to 16S-HTS generated a higher fraction of taxonomically unassigned non-human sequences because of the lack of homologs in sequence databases. Using relatively stringent BLASTN parameters about 19% and 35% of sequence reads remained taxonomically unassigned in the forward- and reverse-run WGS subsets, respectively. Nevertheless, the advantage of the WGS approach is that it allows assessment of not only bacterial but also viral (human viruses and phages) and possibly fungal

and archaeal communities which undoubtedly play an important role in oral health or disease. In addition, an in-depth sequencing of a salivary metagenome may provide insights into gene functions and allow for reconstruction of the functional potential of a microbial population [36]. Functional assignments of sequences may be made for instance using CAMERA, CARMA3 [37] or MG-RAST [38], as it was recently shown for the supragingival dental plaque microbiome [39]. The obtained sequences may be assigned to known functions and classified to major categories including, among others, virulence and resistance to antibiotics.

## Additional material

**Additional file 1: Relative abundance of taxa inferred from 16S-HTS and 16S-WGS datasets.** This Excel file lists taxa and their relative abundance. The RDP Classifier with a  $\geq 50\%$  confidence cutoff was used to assign taxonomy.

**Additional file 2: Accuracy of taxonomic assignments for 81-base V1 and 84-base V3 16S rDNA sequences.** This graph shows the accuracy of taxonomic assignments for 660 HOMO species from 118 genera as determined using the RDP Classifier with a 50% bootstrap threshold.

**Additional file 3: Changes in taxa proportions inferred from 16S-HTS subsets as a function of PCR cycle number.** This is a Word file showing changes (%) in the relative abundance of taxa identified after 25 and 30 cycles of PCR. The values obtained after 20 PCR cycles were used as baseline for comparisons. (A) Taxa identified in all (twelve) 16S-HTS subsets were selected for the analysis. (B) 95%-ID OTUs present in six subsets of the specified V region at a frequency  $> 0.1\%$  in at least one 20-cycle-subset are presented. Changes in the relative abundance of taxa after specified number of cycles: red, increase; blue, decrease.

**Additional file 4: Relative abundance of taxa in the WGS dataset.** This is an Excel file with the taxonomic assignments performed using BLASTN and Bergey's taxonomy accessible via the RDP [40]. Positive BLASTN hits values (wordsize of 16,  $\geq 94\%$  identity,  $\geq 90$  overlap and  $e\text{-value} \leq 10^{-30}$ ) were normalized so that they sum to 100%. In cases where two or more top BLASTN hits were identical, the lowest common ancestor was calculated.

## List of abbreviations

HHV: human herpes virus; HTS: high throughput sequencing; OTU: operational taxonomic unit; WGS: whole genome shotgun.

## Author details

<sup>1</sup>Genomic Research Laboratory, Division of Infectious Diseases, Geneva University Hospitals, Rue Gabrielle-Perret-Gentil 4, CH-1211 Geneva 14, Switzerland. <sup>2</sup>Fasteris, Chemin du Pont-du-Centenaire 109, Case postale 28, CH-1228 Plan-les-Ouates, Switzerland.

## Authors' contributions

VL, KW, PF and JS contributed to the experimental design and drafting the manuscript. VL carried out laboratory procedures. MØ and LF performed high-throughput sequencing. VL, NG, YG and DH performed the bioinformatics and taxonomic analyses. All authors approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

Received: 29 August 2011 Accepted: 2 February 2012

Published: 2 February 2012

## References

- Colombo AP, Boches SK, Cotton SL, Goodson JM, Kent R, Haffajee AD, Socransky SS, Hasturk H, Van Dyke TE, Dewhirst F, Paster BJ: **Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray.** *J Periodontol* 2009, **80**:1421-1432.
- Crielaard W, Zaura E, Schuller AA, Huse SM, Montijn RC, Keijsers BJ: **Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health.** *BMC Med Genomics* 2011, **4**:22.
- Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM: **The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects.** *J Transl Med* 2005, **3**:27.
- Goodson JM, Groppo D, Halem S, Carpino E: **Is obesity an oral bacterial disease?** *J Dent Res* 2009, **88**:519-523.
- Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE: **The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information.** *Database (Oxford)* 2010, **2010**:baq013.
- Shah N, Tang H, Doak TG, Ye Y: **Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics.** *Pac Symp Biocomput* 2011, **165**-176.
- Huang Y, Niu B, Gao Y, Fu L, Li W: **CD-HIT Suite: a web server for clustering and comparing biological sequences.** *Bioinformatics* 2010, **26**:680-682.
- Wang Q, Garrity GM, Tiedje JM, Cole JR: **Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.** *Appl Environ Microbiol* 2007, **73**:5261-5267.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
- Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M: **CAMERA: a community resource for metagenomics.** *PLoS Biol* 2007, **5**:e75.
- Claesson MJ, O'Sullivan O, Wang Q, Ninkila J, Marchesi JR, Smidt H, de Vos WM, Ross RP, O'Toole PW: **Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine.** *PLoS One* 2009, **4**:e6669.
- Paliy O, Foy BD: **Mathematical modeling of 16S ribosomal DNA amplification reveals optimal conditions for the interrogation of complex microbial communities with phylogenetic microarrays.** *Bioinformatics* 2011, **27**:2134-2140.
- Sipos R, Szekely AJ, Palatinszky M, Revesz S, Marialigeti K, Nikolausz M: **Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis.** *FEMS Microbiol Ecol* 2007, **60**:341-350.
- Suzuki M, Rappe MS, Giovannoni SJ: **Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity.** *Appl Environ Microbiol* 1998, **64**:4522-4529.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JL: **Microbial ecology: human gut microbes associated with obesity.** *Nature* 2006, **444**:1022-1023.
- Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O'Toole PW: **Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions.** *Nucleic Acids Res* 2010, **38**:e200.
- Degnan PH, Ochman H: **Illumina-based analysis of microbial community diversity.** *ISME J* 2012, **6**:183-194.
- Lazarevic V, Whiteson K, Huse S, Hernandez D, Farinelli L, Østerås M, Schrenzel J, François P: **Metagenomic study of the oral microbiota by Illumina high-throughput sequencing.** *J Microbiol Methods* 2009, **79**:266-271.
- Schloss PD, Handelsman J: **Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness.** *Appl Environ Microbiol* 2005, **71**:1501-1506.
- Kunin V, Engelbrekton A, Ochman H, Hugenholtz P: **Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates.** *Environ Microbiol* 2010, **12**:118-123.
- Lazarevic V, Whiteson K, Hernandez D, François P, Schrenzel J: **Study of inter- and intra-individual variations in the salivary microbiota.** *BMC Genomics* 2010, **11**:523.
- Keijsers BJ, Zaura E, Huse SM, van der Vossen JM, Schuren FH, Montijn RC, ten Cate JM, Crielaard W: **Pyrosequencing analysis of the oral microflora of healthy adults.** *J Dent Res* 2008, **87**:1016-1020.
- Nasidze I, Li J, Quinque D, Tang K, Stoneking M: **Global diversity in the human salivary microbiome.** *Genome Res* 2009, **19**:636-643.
- Zaura E, Keijsers BJ, Huse SM, Crielaard W: **Defining the healthy "core microbiome" of oral microbial communities.** *BMC Microbiol* 2009, **9**:259.
- Lepp PW, Brinig MM, Ouverney CC, Palm K, Armitage GC, Relman DA: **Methanogenic Archaea and human periodontal disease.** *Proc Natl Acad Sci USA* 2004, **101**:6176-6181.
- Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM: **Characterization of the oral fungal microbiome (mycobiome) in healthy individuals.** *PLoS Pathog* 2010, **6**:e1000713.
- Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, Khristova ML, Weyer J, Swanepoel R, Egholm M, Nichol ST, Lipkin WI: **Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa.** *PLoS Pathog* 2009, **5**: e1000455.
- Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan PL, Hui J, Marshall J, Simons JF, Egholm M, Paddock CD, Shieh WJ, Goldsmith CS, Zaki SR, Catton M, Lipkin WI: **A new arenavirus in a cluster of fatal transplant-associated diseases.** *N Engl J Med* 2008, **358**:991-998.
- Towner JS, Sealy TK, Khristova ML, Albarino CG, Conlan S, Reeder SA, Quan PL, Lipkin WI, Downing R, Tappero JW, Okware S, Lutwama J, Bakamutumaho B, Kayiwa J, Comer JA, Rollin PE, Ksiazek TG, Nichol ST: **Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda.** *PLoS Pathog* 2008, **4**:e1000212.
- Nakamura S, Maeda N, Miron IM, Yoh M, Izutsu K, Kataoka C, Honda T, Yasunaga T, Nakaya T, Kawai J, Hayashizaki Y, Horii T, Iida T: **Metagenomic diagnosis of bacterial infections.** *Emerg Infect Dis* 2008, **14**:1784-1786.
- Miller CS, Avdiushko SA, Kryscio RJ, Danaher RJ, Jacob RJ: **Effect of prophylactic valacyclovir on the presence of human herpesvirus DNA in saliva of healthy individuals after dental treatment.** *J Clin Microbiol* 2005, **43**:2173-2180.
- Slots J, Slots H: **Bacterial and viral pathogens in saliva: disease relationship and infectious risk.** *Periodontol* 2000 2011, **55**:48-69.
- Willner D, Furlan M, Schmieder R, Grasis JA, Pride DT, Relman DA, Angly FE, McDole T, Mariella RP, Rohwer F, Haynes M: **Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity.** *Proc Natl Acad Sci USA* 2011, **108**(Suppl 1):4547-4553.
- Umeda M, Chen C, Bakker I, Contreras A, Morrison JL, Slots J: **Risk indicators for harboring periodontal pathogens.** *J Periodontol* 1998, **69**:1111-1118.
- Slots J: **Human viruses in periodontitis.** *Periodontol* 2000 2010, **53**:89-110.
- Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, Rohwer F: **Functional metagenomic profiling of nine biomes.** *Nature* 2008, **452**:629-632.
- Gerlach W, Stoye J: **Taxonomic classification of metagenomic shotgun sequences with CARMA3.** *Nucleic Acids Res* 2011, **39**:e91.
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA: **The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes.** *BMC Bioinformatics* 2008, **9**:386.
- Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simon-Soro A, Pignatelli M, Mira A: **The oral metagenome in health and disease.** *ISME J* 2012, **6**:46-56.
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM: **The Ribosomal Database Project: improved alignments and new tools for rRNA analysis.** *Nucleic Acids Res* 2009, **37**:D141-145.

doi:10.1186/2043-9113-2-4

Cite this article as: Lazarevic et al.: Analysis of the salivary microbiome using culture-independent techniques. *Journal of Clinical Bioinformatics* 2012 **2**:4.