Microbial Community Profiling of Human Saliva Using Shotgun Metagenomic Sequencing

Nur A. Hasan[1,2], Brian A. Young[3], Angela T. Minard-Smith[3], Kelly Saeed[1], Huai Li[1], Esley M. Heizer[3], Nancy J. McMillan[3], Richard Isom[1], Abdul Shakur Abdullah[1], Daniel M. Bornman[3], Seth A. Faith[3], Seon Young Choi[1,2], Michael L. Dickens[3], Thomas A. Cebula[1,4], Rita R. Colwell[1,2,5,6]*

1 CosmosID, College Park, Maryland, United States of America, **2** Maryland Pathogen Research Institute, University of Maryland, College Park, Maryland, United States of America, **3** Battelle, Columbus, Ohio, United States of America, **4** Department of Biology, Johns Hopkins University, Baltimore, Maryland, United States of America, **5** University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland, United States of America, **6** Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States of America

Abstract

Human saliva is clinically informative of both oral and general health. Since next generation shotgun sequencing (NGS) is now widely used to identify and quantify bacteria, we investigated the bacterial flora of saliva microbiomes of two healthy volunteers and five datasets from the Human Microbiome Project, along with a control dataset containing short NGS reads from bacterial species representative of the bacterial flora of human saliva. GENIUS, a system designed to identify and quantify bacterial species using unassembled short NGS reads was used to identify the bacterial species comprising the microbiomes of the saliva samples and datasets. Results, achieved within minutes and at greater than 90% accuracy, showed more than 175 bacterial species comprised the bacterial flora of human saliva, including bacteria known to be commensal human flora but also *Haemophilus influenzae*, *Neisseria meningitidis*, *Streptococcus pneumoniae*, and Gamma proteobacteria. Basic Local Alignment Search Tool (BLASTn) analysis in parallel, reported *ca*. five times more species than those actually comprising the *in silico* sample. Both GENIUSand BLAST analyses of saliva samples identified major genera comprising the bacterial flora of saliva, but GENIUS provided a more precise description of species composition, identifying to strain in most cases and delivered results at least 10,000 times faster. Therefore, GENIUS offers a facile and accurate system for identification and quantification of bacterial species and/or strains in metagenomic samples.

Citation: Hasan NA, Young BA, Minard-Smith AT, Saeed K, Li H, et al. (2014) Microbial Community Profiling of Human Saliva Using Shotgun Metagenomic Sequencing. PLoS ONE 9(5): e97699. doi:10.1371/journal.pone.0097699

Editor: Niyaz Ahmed, University of Hyderabad, , India

Received December 18, 2013; **Accepted** April 22, 2014; **Published** May 20, 2014

Funding: Funding for this study was provided, in part, by NIH Grants 1RO1A139129-01 and 2RO1A1039129-11A2. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Rita R. Colwell is Founder and Chairman of the Board, CosmosID®, and some of the other authors are employees of the company, a bioinformatics company. Rita R. Colwell is also Distinguished University Professor at the University of Maryland, College Park and at Johns Hopkins University Bloomberg School of Public Health. Affiliation with CosmosID and, similarly, with Battelle does not alter the authors' adherence to all PLoS One policies on sharing data and materials online as detailed in the online guide for authors.

* E-mail: rcolwell@umiacs.umd.edu

Introduction

The microbial flora of the human mouth has been extensively studied, providing an understanding of the role of bacterial species, not only in maintaining wellness, but also in dental caries and gingivitis[1,2,3]. The reductionist approach to understanding the microbiology of the human mouth that was followed in the early days of dental microbiology, in some ways inhibited achievement of full understanding of the oral flora as a microbial ecosystem. Recent research on the human microbiome has provided valuable information concerning a variety of processes, including interactions among microbial species in the human mouth. Next generation sequencing (NGS) permits an even more extensive characterization of the microbial ecology of the human body and has triggered an explosion in human microbiome discovery. As a result, the microbiome is now considered by some investigators to represent yet another organ of the human body, dictating health and well-being[4]. From the avalanche of data reported to date, spatio-temporal and host induced variations in microbiomes have been associated with a variety of human conditions, including colorectal carcinoma[5,6], cardiovascular disease[7], inflammatory bowel disease[8], obesity [9], white blood cell cancer [10], and even psychiatric conditions[11].

The oral cavity is a major gateway for bacterial entry to the human body and a natural route for passage to respiratory and digestive tracts and, ultimately, the blood stream. Historically, microorganisms in the oral cavity were found to comprise a diverse and complex community [12], comprised of hundreds of individual bacterial species[13,14]. Recent evidence shows that some bacterial species of the mouth microbiome are linked to oral disease, but are also important in the general health of an individual [12,15,16,17,18,19]. To date, the oral microbiome has been linked to many diseases, namely alveolar osteitis and tonsillitis [12,14,20,21,22,23], bacteremia [24], endocarditis [25], brain and liver abscesses [26,27], stroke [28], diabetes [29,30], pneumonia [31], and premature birth [32]. The mouth can also be considered an important site of genetic exchange among members of the bacterial flora, because of its high bacterial load and richness in species diversity, even where antibiotic resistant bacteria can become established through contact transmission

[33]. The oral microbiome is not homogenous but is made up of subpopulations inhabiting microenvironments within the mouth [15]. A primary example is saliva, which contains a specific bacterial community that helps maintain homeostasis of the mouth ecosystem. Thus, it is not surprising that the oral or salivary microbiome has attracted increased attention as a potential diagnostic tool [23,34,35,36]. Collection of saliva samples is simple and not invasive [15,16,37]. Saliva itself is clinically informative, containing many soluble biomarkers typically found in blood and urine [38,39] that are useful in prognosis of several systemic and oral conditions. In fact, the uniqueness of individual oral commensal flora provides a forensic tool, contributing to the development of the new discipline microbial forensics [40,41]. Because the salivary microbiome has diagnostic, epidemiological, and forensic value, we investigated the bacterial flora of saliva using direct whole genome shotgun (WGS) metagenomic sequencing, an unbiased metagenomic approach to determine the bacterial species and strain composition.

Historical microbiological studies employing conventional culture methods had shown that the human salivary microbiome is comprised of a complex assemblage of bacteria, viruses, fungi, and parasites, with less than half of the bacterial species cultured [12,13,14,42]. 16S rRNA-based identification revealed extensive bacterial diversity, providing results more quickly than traditional culture [23]. However, 16S rRNA is a single gene-centric method, providing less resolution in differentiating closely related species. It also suffers from limitations imposed by non-uniform distribution of sequence dissimilarity among taxa, presence of multiple copies of the 16S rRNA gene [43], failure of target amplification of polymerase chain reaction (PCR) primers [44], and generation of chimeric sequences [45,46].

The accuracy and robustness of any identification methods is dependent on the quality and breadth of the reference database [12]. The popular sequence alignment tool, BLAST [47] relies on the NCBI public database from which even NCBI removes sequences and genomes due to errors [48]. The known limitations of 16S rRNA sequencing for microbial identification has prompted investigators to use WGS sequencing for characterization and resolution of metagenomic communities [12,36], The large amounts of data produced by WGS sequencing, however, present significant challenges in data analysis and interpretation [49]. There are many approaches that have been devised for analysis of WGS data, including alignment, assembly, binning, and gene prediction based methods [50,51]. Read reference alignment or mapping performs reasonably well, but these are computationally very expensive [52]. Compositional binning tools, i.e. MEGAN [53] and MG-RAST [54], are also computationally expensive and do not resolve closely related taxa with the short reads as generated by Illumina and Life Technologies NGS platforms [50].

NGS has progressed today to being able to deliver highly accurate sequences economically and with fast turn-around. As a result, whole genome shotgun (WGS) metagenomic sequencing emerged as a powerful tool for studying the human microbiome [55]. At present, WGS metagenomic data comprise millions to billions of short reads, aiding necessary sequencing depth as needed as well as offering an unprecedented opportunity to identify individual species at or near strain level and determine their relative abundance. In this study, WGS metagenomics was employed, in combination with GENIUS algorithms, to identify and quantitate bacterial species comprising the salivary microbiome.

Materials and Methods

Sample Collection and DNA Isolation

Total DNA was collected from saliva samples provided by two anonymous healthy adult donors following the approved protocol of the Battelle Memorial Institute Internal Review Board. DNA was purified from saliva employing the Oragene-DNA isolation kit (DNA Genoteck, Kanata, ON, Canada), following the manufacturer's recommended protocol.

Next Generation Sequencing and Filtering

DNA samples for metagenomics were prepared for 150 bp and 100 bp single-end sequencing using the Illumina GAIIx and HiSeq 2000 instrument (Illumina, San Diego, CA), respectively. Numerically coded aliquots of approximately 0.5–1 μg DNA per sample were used to create sequencing libraries. First, genomic DNA was fragmented using a Covaris^{TM} S220 Sonicator (Covaris, Inc., Woburn, MA) to approximately 300 base pairs (bp). Fragmented DNA was used to synthesize indexed sequencing libraries using the TruSeq DNA Sample Prep Kit V2 (Illumina, Inc., San Diego, CA), according to manufacturer's recommended protocol. Cluster generation was performed on the cBOT using the TruSeq PE Cluster Kit v3 – cBot – HS (Illumina). Libraries were sequenced with an Illumina HiSeq 2000 at Nationwide Children's Hospital (NCH) Biomedical Genomics Core (Columbus, Ohio) using the TruSeq SBS Kit v3 reagents (Illumina) for paired end sequencing with read lengths of 100 base pairs (bps) (200 cycles) and at CosmosID with an Illumina GAIIx for 150 base pairs (bps) single read using the TruSeq SBS Kit v5 reagents (Illumina). Primary analysis (image analysis and basecalling) were performed using HiSeq Control Software (HCS) version 1.5.15.1 and Real Time Analysis (RTA) version 1.13.48. Secondary Analysis (demultiplexing) was performed using Illumina CASAVA Software v1.6 Post processing of GAIIx reads was performed with RTA/SCS v1.9.35.0 and CASAVA 1.8.0 software. High throughput sequencing reads were quality filtered using the fastq_quality_filter program provided with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) (v. 0.0.13). Only those reads with a quality score ≥17 for at least 80% of the read length (i.e., probability of correct base call ~98%) were retained. Ion Torrent (Life Technologies, NY) sequencing was also performed using amplicons specific to the V4 region of the 16S rRNA gene. Sequence reads are available under NCBI BioProject ID PRJNA231652.

BLASTn Analysis

Sequence data were compared to the NCBI RefSeq database (v. May 19, 2012), but restricted to microbial gis, the NCBI 16S database (v. October 30, 2012), using BLASTn [47] (top hit only) (v. 2.2.25, National Library of Medicine, Bethesda, MD). Table 1 provides details of analyses carried out for each sample with data bases used. Resulting BLASTn hits were filtered to retain only those hits with percent identity ≥97%. An additional filter was applied to the BLASTn hit report to reduce false positives (i.e., reads whose corresponding taxonomic identifier (taxid) appeared ≤0.01% (1:1000)). This was accomplished using a custom script. The Krona (v. 2.2) [56] program, ClassifyBLAST.pl, was also used within a custom script, to obtain a list of organisms identified with read counts associated with each taxon. Krona ImportBLAST.pl program was used to provide interactive visualization of identified bacterial species.

Table 1. Multi-platform BLASTn analysis of two salivary samples against various databases.

Sample	Sequencing Platform		
	Illumina HiSeq 2000	**Illumina GAIIx**	**Ion Torrent (16S, V4)**
	Reference Database used with BLASTn		
VFD10-018	Greengenes 16s[1]	Greengenes 16s[4]	Greengenes 16s[3]
	NCBI 16s	NCBI 16s[4]	NCBI 16s[3]
	NCBI RefSeq (microbial subset)[1]	NCBI RefSeq (microbial subset)[4]	NCBI RefSeq (microbial subset)[3]
VFD12-006	Greengenes 16s[1]	Greengenes 16s[2]	Greengenes 16s[3]
	NCBI 16s[1]	NCBI 16s[2]	NCBI 16s[3]
	NCBI RefSeq (microbial subset)[1]	NCBI RefSeq (microbial subset)[2]	NCBI RefSeq (microbial subset)[3]

[1] Illumina HiSeq2000 sequencing carried out at Nationwide Children's Hospital, Columbus, OH.
[2] Illumina GAIIx sequencing carried out at CosmosID.
[3] Ion Torrent sequencing performed by SeqWright, Inc., Houston, TX.
[4] Illumina GAIIx sequencing performed at The Ohio State University, Columbus, OH.
doi:10.1371/journal.pone.0097699.t001

GENIUS Analysis

Raw unassembled WGS short reads generated by the Illumina GAIIx and HiSeq 2000 platforms were analyzed using GENIUS software package for rapid identification of bacterial species and relative abundance. GENIUS creates sample libraries from unassembled short WGS reads using two algorithms, 5VCE and NmerCE, and utilizes GeneBook® reference libraries derived from curated genomic databases to assign taxonomic membership of sample libraries, employing probabilistic matching. Identification is achieved at species, sub-species, and/or strain level, depending on adequate representation of relevant reference genomes in the GeneBook® libraries

In silico Metagenome Construction

A synthetic metagenome was created comprising a total of 5.5 M reads from ten bacterial species and the human genome (Table 2). The reads, each 100 nucleotides in length, were created using a custom R script from each of the ten bacterial genome and the human chromosome 21 using Illumina sequencing error model.

Results and Discussion

Description of the human microbiome has been made possible by NGS with its significant reduction in cost and improvement in throughput. Metagenomics, as a result, is moving from a16S rRNA gene-centric approach to WGS metagenomic approach. To date, 16S rRNA gene sequencing has been used to identify major taxa and explore the microbial diversity of the human salivary microbiome [23,35,57] linking composition of the microbiome with oral health and/or systemic disease. In this study, WGS metagenomics was used, along with several bioinformatics analysis methods e.g., BLASTn, mapping, and GENIUS, to determine relative performances of taxonomic assignment, and identification of community composition and structure, thereby achieving improved understanding of the human salivary microbiome.

GENIUS 5VCE algorithm was employed to determine the bacterial species composition of a human saliva sample, VFD10-018, sequenced by the Illumina GAIIx (150 bp, $\sim$22 M reads) platform. A total of 26 bacterial genera and 58 species were identified with majority of genera and species previously identified

Table 2. Species composition and simulation statistics of the synthetic metagenomic dataset.

Simulations Statistics				BLAST (RefSeq_genomic)
Species	**Genome Coverage**	**Number of Reads**	**Relative Abundance**	**Number of Reads**
Homo sapiens	0.08	4034394	72.81	3349
Rothia dentocariosa ATCC 17931	47.21	537919	9.71	554
Prevotella melaninogenica ATCC 25845	52.69	430335	7.77	440
Fusobacterium nucleatum subsp. nucleatum ATCC 25586	5.44	53791	0.97	58
Streptococcus oralis Uo5	12.08	107583	1.94	106
Streptococcus mitis B6	11.02	107583	1.94	88
Veillonella parvula DSM 2008	11.1	107583	1.94	83
Peptostreptococcus stomatis DSM 17678	0.99	46261	0.83	NA
Peptostreptococcus anaerobius 653-L	0.94	46261	0.83	NA
Porphyromonas gingivalis W83	0.84	46261	0.83	46
Mycoplasma pneumoniae FH	1.21	23130	0.42	23

The right most column represents number of reads from the sample that BLAST was able to assign to species. NA: Not Assigned due to lack of RefSeq entries.
doi:10.1371/journal.pone.0097699.t002

as members of the human salivary and/or oral microbiome (http://www.homd.org/index.php, Fig. S1). BLASTn (microbial subset) and short read mapping (CLC genomic workbench, using same genome database as GENIUS 5VCE) identified 45 and 102, and 67 and 108, bacterial genera and species in this data set respectively, indicating a much larger microbial community compared to that identified by GENIUS 5VCE (Fig. S1). A global 16S metagenic survey of saliva samples collected from 120 healthy individuals in 12 geographically different locations reported that an individual salivary microbiome typically contains six to 30 bacterial genera [41], an observation in agreement with GENIUS 5VCE identification. Dominant genera identified by GENIUS 5VCE (*Streptococcus,Prevotella, Veillonella, Mycoplasma, Rothia, Haemophilus, Fusobacterium etc.)* were also in agreement with genera identified in the global survey [41]. It is concluded that bacterial taxa identified by BLASTn and short read mapping produces an overestimation of diversity. GENIUS 5VCE performed favorably in both speed, at least 10,000X faster than BLAST, and accuracy in identifying the most likely bacterial community of this dataset with significantly reduced false prediction.

As the actual microbial composition of the saliva samples from the volunteers was unknown, only comparative analysis between orthogonal methods was possible. Therefore, an *in silico* sample (Table 2) containing a composite of 5.5 M reads from ten bacterial species and the human genome was prepared to measure accuracy of the metagenomic analysis. Results show that GENIUS accurately identified bacterial species composition with a negligible computation time (2 minutes for NmerCE and 29 minutes for 5VCE) (Figs. 1 and S2). Despite the fact that the number of sequencing reads for each of the genomes comprising the test set was small, the GENIUS algorithms identified the bacterial species with appropriate strain designation. Identification obtained using GENIUS 5VCE and NmerCE algorithms were in agreement, with one possible false positive identification (*Streptococcus pneumoniae*) by 5VCE. When statistical analysis was carried out, which is a built in function in GENIUS 5VCE algorithm, to provide point estimates for genome coverage and detection confidence limits using random k-mers for each of the identified species, *S. pneumoniae* had the lowest confidence interval (Fig. S2). Furthermore, considering the marginal coverage of strain specific attributes (only 1.6% of unique identifiers), error rate of the sequencing platform, and abundant presence of other *Streptococcus* species in the dataset, this call, at best, would have been tallied as dubious. In contrast, results of BLASTn analysis reported 48 species, even though the sample contained only ten bacterial species, confirming what has been suspected by other investigators as over-estimation of diversity by BLASTn analysis. In contrast, GENIUS was efficient in filtering out false positive signals caused by high genomic similarity among closely related genera and species. Briefly, GENIUS was successful in identifying species of the saliva microbial community, even when only limited sequencing data were available, accomplishing identification that required only minimum computational time. Precise identification was achieved in most cases even with only a small number of reads, i.e., fractional coverage (<1%) of the genome (i.e.,*Peptostreptococcus stomatis, Peptostreptococcus anaerobius*, and *Porphyromonas gingivalis*) were available and also when the target pathogen (i.e., *Mycoplasma pneumoniae)* was present at very low concentration (<1%).

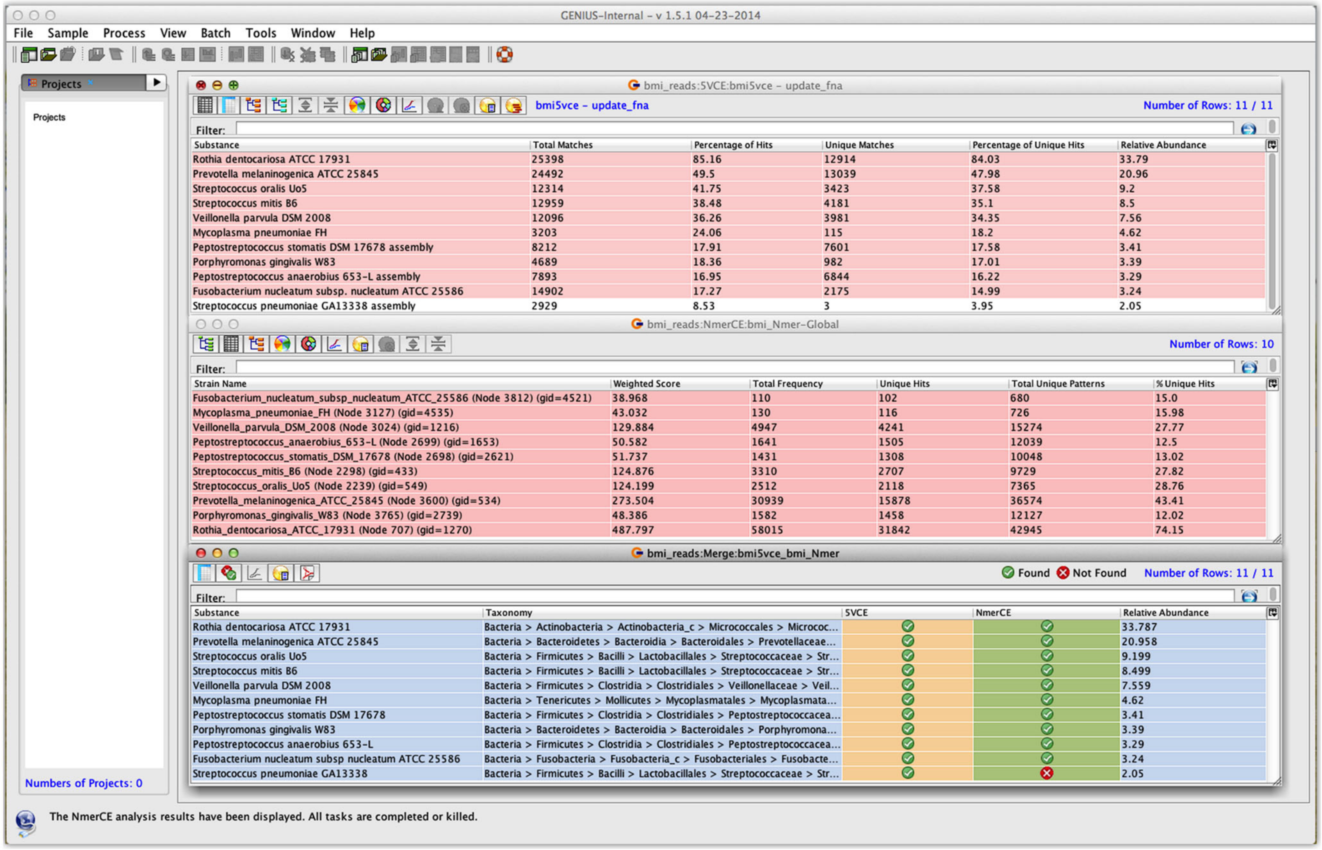
For comparative purposes, 16S rRNA (V4 region)sequencing was carried out using Ion Torrent PGM for salivary samples VFD10 and VFD12 and the sequence data were analyzed by BLAST, using the NCBI 16S ribosomal database (v. 10/30/2012) to enable direct comparison of bacterial communities inferred by both 16S metagenic and WGS metagenomics analysis. Comparison of the accuracy of identification between the two methods was assessed by the number of overlapping genera, since extrapolation of 16S data beyond genus is very limited [43,58,59]. Saliva microbiomes VFD10 and VFD12 showed high concordance (~80%), with respect to genera identified by GENIUS algorithms 5VCE and NmerCE (Fig. 2). Concordance was shared, to a large extent, with 88 genera identified by BLAST-16S (Fig. 2a). GENIUS algorithms 5VCE and NmerCE identified 27 and 26 genera, respectively, in VFD10. Of the 21 genera identified by both algorithms, 17 were also identified by BLAST-16S. The six additional genera identified by5VCE (n = 2) and NmerCE (n = 4) were also identified by BLAST-16S. However, five genera were identified by GENIUS, either by5VCE (n = 4) or NmerCE (n = 1) that were not detected by BLAST-16S. Relative abundance (≥ 1%)of genera determined by GENIUS algorithms and byBLAST-16S (Table S1) were in agreement, with respect to dominant genera, with26 genera comprising 97–99% of the bacterial community.

Results for saliva sample VFD12 are shown in Fig. 2b, with comparison of relative abundance showing 22 genera accounting for 98–99% of the bacterial community (Table S2), providing very little evidence for the large number of genera (n = 82) identified by BLAST-16S. In fact, the global survey of saliva samples [41]reported individual salivary microbiomes contained six to 30 bacterial genera, reinforcing overestimation of diversity by 16S analysis. Based on the literature,16S rRNA sequencing can be biased by unequal amplification of 16S rRNA genes [46] and by taxon-specific biases arising from the primer set used [60]. Generation of chimeric sequences also can skew and inflate diversity estimates significantly [45,46]. Chimera formation is most pronounced when 16S rDNA amplicons are present in low amount, making identification of minor species suspect without supporting data [61].

WGS metagenomics sequencing reads (GAIIx) were compared with the NCBI RefSeq database (microbial subset, v. 05/19/2012), using BLASTn for comparison of species identified by GENIUS algorithms and BLASTn. Fig. 3 shows estimated relative abundance of species at ≥2% for saliva sample VFD10, determined by both GENIUS and BLASTn analyses. A phylogenetic tree of 23 species (depicted as shaded squares) in Fig. 4 shows relative abundance estimated by each method. Most prevalent were *Streptococcus* and *Prevotella*. Similar results were obtained for saliva sample VFD12 (Fig. 4).

A vexing problem in metagenomics is sample diversity, that is, whether sequencing captures the total diversity of a given sample. Libraries generated for saliva sample VFD10 were sequenced using all eight lanes of a GAIIx flow cell, with 18 to 24 million reads (average ~22 million) generated per lane (Fig. S3). GENIUS analysis, with respect to number of species identified, showed good concordance in both identification and relative abundance of bacterial species, for all eight lanes (Fig. S3). Thus, use of an entire flow cell for a single sample library did not influence the diversity estimate.

Illumina HiSeq 2000 allows *ca.*180 million reads per lane, with TruSeq v3 chemistry, compared to *ca.* 40 million reads per lane, with GA v5 chemistry and the GAIIx instrument. When the larger number of reads generated by HiSeq (66–75 million) was analyzed, the number of species identified and percent of each species increased for both saliva samples VFD10 and VFD12, indicating that a larger number of HiSeq reads will contribute breadth and depth in genome coverage (Fig. S4). These results suggest that identification of bacterial species, particularly those present in low number, can be improved with a larger number of


GENIUS-Internal – v 1.5.1 04-23-2014
File Sample Process View Batch Tools Window Help
Projects
Projects
Numbers of Projects: 0

bmi_reads:5VCE:bmi5vce – update_fna

bmi5vce – update_fna — Number of Rows: 11 / 11

Filter:

Substance	Total Matches	Percentage of Hits	Unique Matches	Percentage of Unique Hits	Relative Abundance
Rothia dentocariosa ATCC 17931	25398	85.16	12914	84.03	33.79
Prevotella melaninogenica ATCC 25845	24492	49.5	13039	47.98	20.96
Streptococcus oralis Uo5	12314	41.75	3423	37.58	9.2
Streptococcus mitis B6	12959	38.48	4181	35.1	8.5
Veillonella parvula DSM 2008	12096	36.26	3981	34.35	7.56
Mycoplasma pneumoniae FH	3203	24.06	115	18.2	4.62
Peptostreptococcus stomatis DSM 17678 assembly	8212	17.91	7601	17.58	3.41
Porphyromonas gingivalis W83	4689	18.36	982	17.01	3.39
Peptostreptococcus anaerobius 653-L assembly	7893	16.95	6844	16.22	3.29
Fusobacterium nucleatum subsp. nucleatum ATCC 25586	14902	17.27	2175	14.99	3.24
Streptococcus pneumoniae GA13338 assembly	2929	8.53	3	3.95	2.05

bmi_reads:NmerCE:bmi_Nmer-Global

Number of Rows: 10

Filter:

Strain Name	Weighted Score	Total Frequency	Unique Hits	Total Unique Patterns	% Unique Hits
Fusobacterium_nucleatum_subsp_nucleatum_ATCC_25586 (Node 3812) (gid=4521)	38.968	110	102	680	15.0
Mycoplasma_pneumoniae_FH (Node 3127) (gid=4535)	43.032	130	116	726	15.98
Veillonella_parvula_DSM_2008 (Node 3024) (gid=1216)	129.884	4947	4241	15274	27.77
Peptostreptococcus_anaerobius_653-L (Node 2699) (gid=1653)	50.582	1641	1505	12039	12.5
Peptostreptococcus_stomatis_DSM_17678 (Node 2698) (gid=2621)	51.737	1431	1308	10048	13.02
Streptococcus_mitis_B6 (Node 2298) (gid=433)	124.876	3310	2707	9729	27.82
Streptococcus_oralis_Uo5 (Node 2239) (gid=549)	124.199	2512	2118	7365	28.76
Prevotella_melaninogenica_ATCC_25845 (Node 3600) (gid=534)	273.504	30939	15878	36574	43.41
Porphyromonas_gingivalis_W83 (Node 3765) (gid=2739)	48.386	1582	1458	12127	12.02
Rothia_dentocariosa_ATCC_17931 (Node 707) (gid=1270)	487.797	58015	31842	42945	74.15

bmi_reads:Merge:bmi5vce_bmi_Nmer

✓ Found ☒ Not Found — Number of Rows: 11 / 11

Filter:

Substance	Taxonomy	5VCE	NmerCE	Relative Abundance
Rothia dentocariosa ATCC 17931	Bacteria > Actinobacteria > Actinobacteria_c > Micrococcales > Micrococ...	✓	✓	33.787
Prevotella melaninogenica ATCC 25845	Bacteria > Bacteroidetes > Bacteroidia > Bacteroidales > Prevotellaceae...	✓	✓	20.958
Streptococcus oralis Uo5	Bacteria > Firmicutes > Bacilli > Lactobacillales > Streptococcaceae > Str...	✓	✓	9.199
Streptococcus mitis B6	Bacteria > Firmicutes > Bacilli > Lactobacillales > Streptococcaceae > Str...	✓	✓	8.499
Veillonella parvula DSM 2008	Bacteria > Firmicutes > Clostridia > Clostridiales > Veillonellaceae > Veil...	✓	✓	7.559
Mycoplasma pneumoniae FH	Bacteria > Tenericutes > Mollicutes > Mycoplasmatales > Mycoplasmata...	✓	✓	4.62
Peptostreptococcus stomatis DSM 17678	Bacteria > Firmicutes > Clostridia > Clostridiales > Peptostreptococcacea...	✓	✓	3.41
Porphyromonas gingivalis W83	Bacteria > Bacteroidetes > Bacteroidia > Bacteroidales > Porphyromona...	✓	✓	3.39
Peptostreptococcus anaerobius 653-L	Bacteria > Firmicutes > Clostridia > Clostridiales > Peptostreptococcacea...	✓	✓	3.29
Fusobacterium nucleatum subsp nucleatum ATCC 25586	Bacteria > Fusobacteria > Fusobacteria_c > Fusobacteriales > Fusobacte...	✓	✓	3.24
Streptococcus pneumoniae GA13338	Bacteria > Firmicutes > Bacilli > Lactobacillales > Streptococcaceae > Str...	✓	☒	2.05

The NmerCE analysis results have been displayed. All tasks are completed or killed.

Figure 1. Screenshot of GENIUS client software displaying tabular output of *in silico* metagenomic data. Left panel indicates projects loaded to this graphical user interface. Three table views to the rightmost panel represent output of 5VCE (a)NmerCE (b) and merged output from both 5VCE and NmerCE (c) algorithms, respectively.
doi:10.1371/journal.pone.0097699.g001

reads, especially in the case of samples containing large a mounts of human DNA. Analysis of saliva sample VFD10, using GAIIx sequence data, showed approximately 97% of the sequenced reads was from human DNA. Since background human DNA can range from less than 1% in stool samples to greater than 99% in nasal and vaginal samples [62], the effect of having an increased number of reads to capture species diversity will vary according to host DNA content, as well as complexity of the bacterial population. Improved sampling, extraction, and library construction methods, therefore, should be considered for maximum coverage of species diversity.

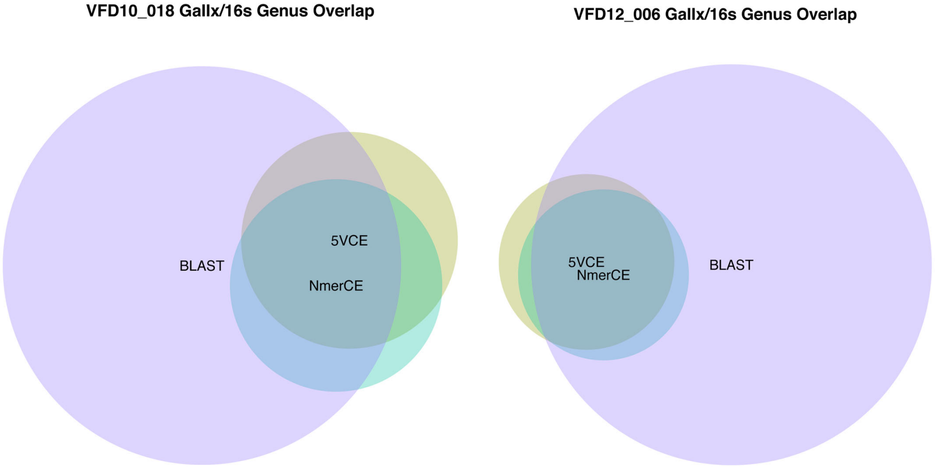


Figure 2. Genus overlap for sample VFD10-018 (a) and VFD12-006 (b) estimated by 16S sequencing/NCBI 16S BLAST and GAIIx sequencing/5VCE-NmerCE.
doi:10.1371/journal.pone.0097699.g002

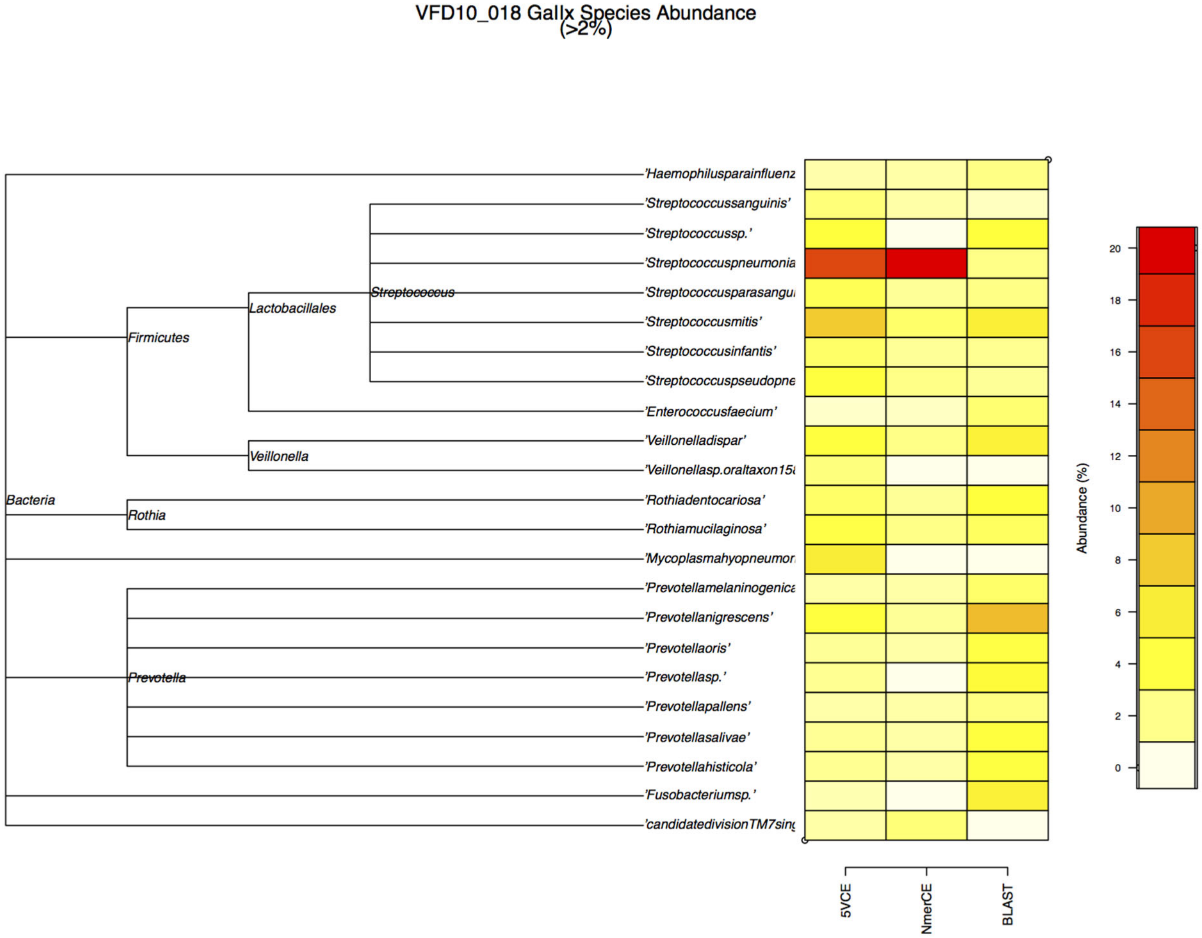


Figure 3. Relative abundance of species in VFD10-018 estimated by GAIIx sequencing and BLAST (microbial reference database), 5VCE, and NmerCE algorithms.
doi:10.1371/journal.pone.0097699.g003

GENIUS was used to analyze salivary datasets from the Human Microbiome Project (HMP) (http://hmpdacc.org/HMASM/). Five human salivary microbiomes were analyzed and nine major phyla were identified by GENIUS 5VCE with Firmicutes, Bacteroidete, Actinobacteria, and Proteobacteria most abundant, Fusobacteria and TM7moderately abundant, and Spirochaetes, Synergistetes, and Tenericutes least abundant. Sixty seven bacterial genera belonging to nine phyla were identified, with eleven genera, *Streptococcus, Prevotella, Veillonella, Neisseria, Haemophilus, Campylobacter, Fusobacterium, Rothia, Mycoplasma,Actinomyces,and Aggregatibacter* comprising ~90% of the bacterial community. Relative abundance estimates of phyla and genera varied (Fig. S5). Overall abundance and distribution of phyla and genera were in agreement with results of studies of human saliva reported by other investigators [34,63,64]. Interestingly, ***Streptococcus*** was observed to be prevalent in most of the HMP datasets, whereas *Prevotella* predominated in the saliva samples, VFD12 and SRS014692, analyzed in this study (Fig. S5). That is, a readily distinguishable abundance profile of ***Prevotella*** species and strains was observed in saliva samples VFD10, VFD12, and SRS014692 (Fig.S6). Greater abundance of ***Prevotella*** in caries-active, **compared to** healthy, individuals has been reported with caries-active individuals often carrying a mixture of *Prevotella* species different from normal healthy individuals [65].

GENIUS identified more than 175 bacterial species, including bacteria commensal to the human salivary microbiome ([65],HOMD, www.homd.org) but also others not usually found in the saliva flora of healthy individuals, including *Haemophilus influenza*e, *Neisseria meningitidis*, *Streptococcus pneumoniae* and Gammaproteobacteria. Several bacterial species, including ***Aggregatibacter actinomycetemcomitans, Porphyromonas gingivalis,*** *Treponema denticola, Fusobacterium nucleatum, Campylobacter rectus, Parvimonas micra, Eikenella corrodens, Prevotella melaninogenica, Prevotella nigrescens, Eubacterium saburreum*, and *Eubacterium yurii*, associated with periodontitis [12,66,67,68], were also identified in varying abundance and distribution (Fig. S7). Although presence of these bacterial species may indicate disease, i.e., periodontitis, it has been shown that periodontitis can be attributed to genetic factors [69]. Therefore, diagnosis of periodontitis cannot yet be made by bacteria present in saliva.

Principal component analysis (PCA) of the five HMP saliva samples and the two saliva samples sequenced in this study showed the bacterial species composition comprised two major clusters (Fig. 5). Four saliva samples, VFD10, VFD12, SRS015055, and SRS014692, clustered separately from SRS09210, SRS013942, and SRS014468. Distinction between clusters, as well as variation within each population, was apparent from a double hierarchical dendrogram showing abundance and distribution of the bacterial species (Fig. 6). Even though two major clusters were observed,

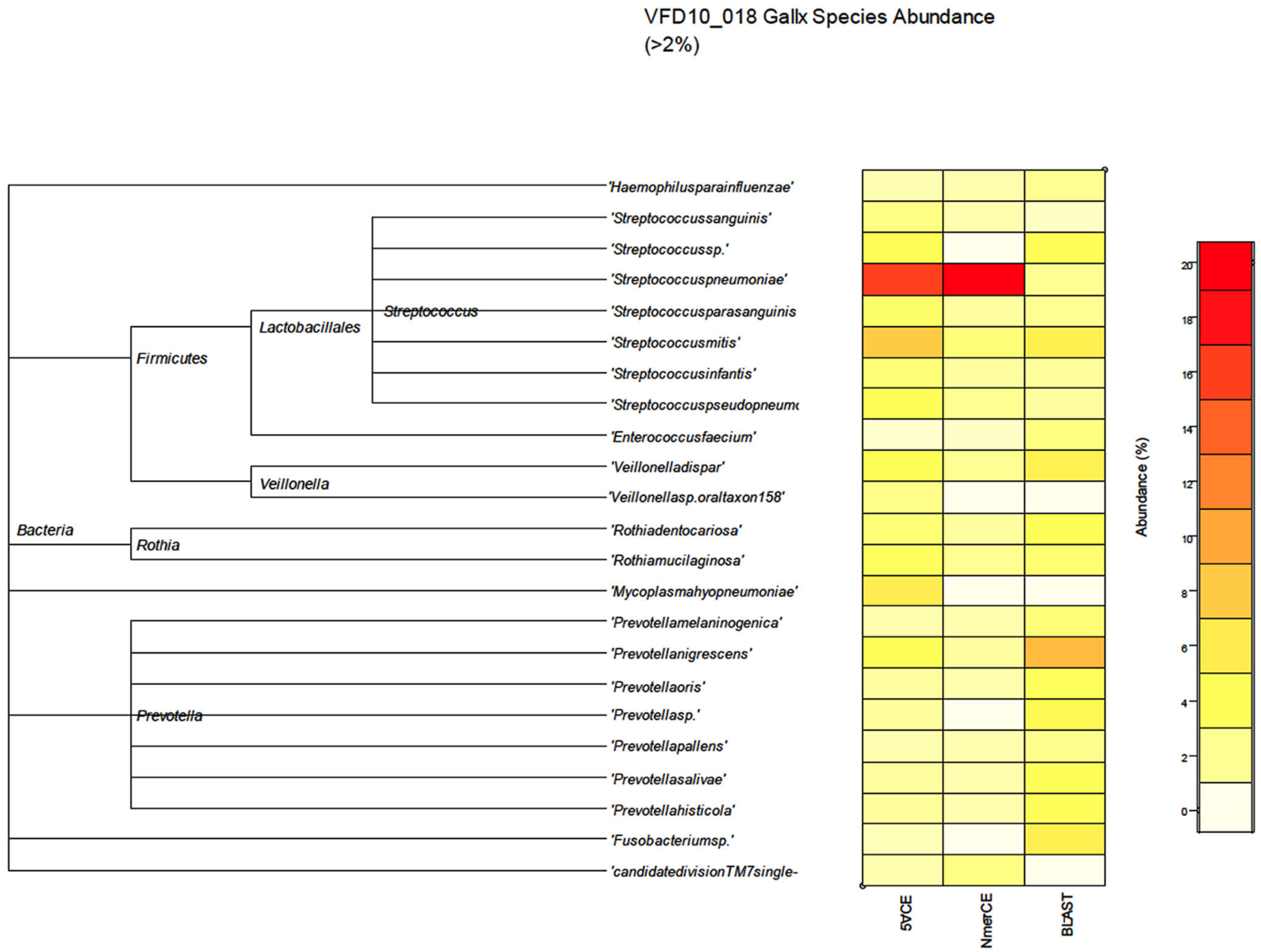


Figure 4. Relative abundance of species in VFD12-006 estimated by GAIIx sequencing and BLAST (microbial reference database), 5VCE, and NmerCE algorithms.
doi:10.1371/journal.pone.0097699.g004

clearly diversity and species assemblage of the saliva samples were not identical. Such differences reflect diet, hygiene, and/or family and culture, all of which influence the oral microbiome. Centroid classification [70] within the two major groupings showed four genera, *Haemophilus*, *Streptococcus*, *Neisseria*, and *Aggregatibacter*, were over-represented in cluster B (Fig. S8). Over representation of *Haemophilus* species, i.e., *H. influenza*e, *H. parainfluenza*e, and *Aggregatibacter aphrophilus*(née, *Haemophilus aphrophilus*)in cluster B is particularly interesting, since these bacterial species have been shown to be associated with *Haemophilus* endocarditis [71,72]. Although *Haemophilus* species can cause adult endocarditis (0.8–1.3%) [73], the presence of a significant number of each of these species represent a skewing from healthy human saliva.

Conclusion

WGS metagenomics applied to the human microbiome has provided useful information applicable to public health and personalized medicine, especially as high-throughput ultra-deep sequencing approaches real time and becomes cost effective. However, post sequencing processing and analysis of data generated by WGS metagenomics are extremely challenging. While traditional BLAST analysis is hindered by factors like time for analysis, low resolution, and large computational requirements, a marker gene approach will speed detection, but sacrifices resolution. Genome mapping and reconstruction ensure precise identification but takes a long 'time to identification' and require powerful computational infrastructures and skilled manpower. In this study, GENIUS algorithms and WGS metagenomic data were used to identify the bacterial community composition of human saliva. Compared to16S metagenic sequencing and analysis, WGS metagenomics provided greater accuracy, both in identification and quantitation of bacterial species and less biased estimate of diversity, when GENIUS algorithms were used. Superior speed, accuracy, and precision in identification were achieved compared to 16S which significantly overestimated diversity. GENIUS algorithms provide high specificity and accurate identification of species, even those present in low abundance and with fractional genome coverage. WGS metagenomics employing GENIUS algorithms is proposed as method of choice for rapid, accurate, and user friendly bacterial identification and metagenomics.

In this study, it has been demonstrated that WGS metagenomics provides a practical approach in answering questions about the human salivary microbiome. Therefore, metagenomic analysis of clinical samples, not only of the salivary microbiome, but also other microbial flora, in general, offers greater power of decision making, precision, and speed compared to traditional methods.

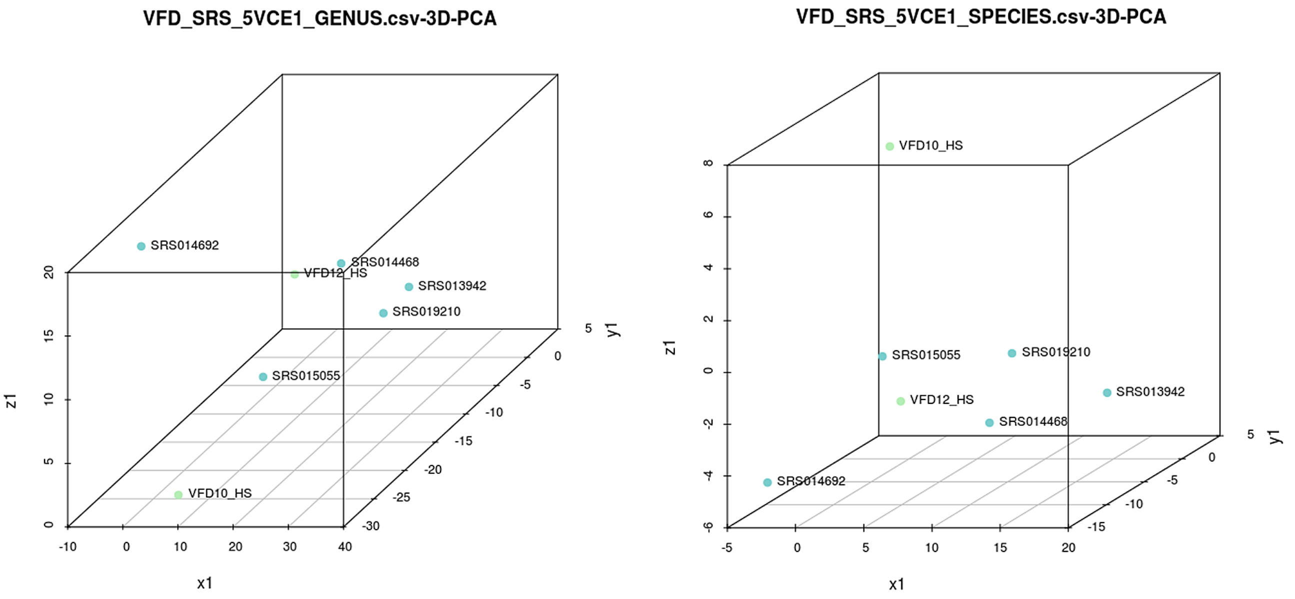


Figure 5. Principal component analysis of data for seven saliva samples analyzed by GENIUS.
doi:10.1371/journal.pone.0097699.g005

The typical "time to answer" for culture-based methods requires weeks for completion, whereas sequencing approaches can reduce the timeline to a few days. For WGS metagenomics for microbial detection and identification, laboratory protocols require a fraction of time compared to culture-based methods, especially since culturing is not required prior to library construction. The time required for a sample being processed and sequenced is approximately two days if MiSeq, 454 GS junior, Ion Torrent PGM, NextSeq or HiSeq X platforms are used for sequencing throughput. However, this timeline could be reduced through automation and will soon be less than a day. The main difference is actually the data analysis, when GENIUS is used, it takes only half an hour or less for analysis of metagenomic data derived from any routine clinical samples and does not require time-consuming alignment or mapping.

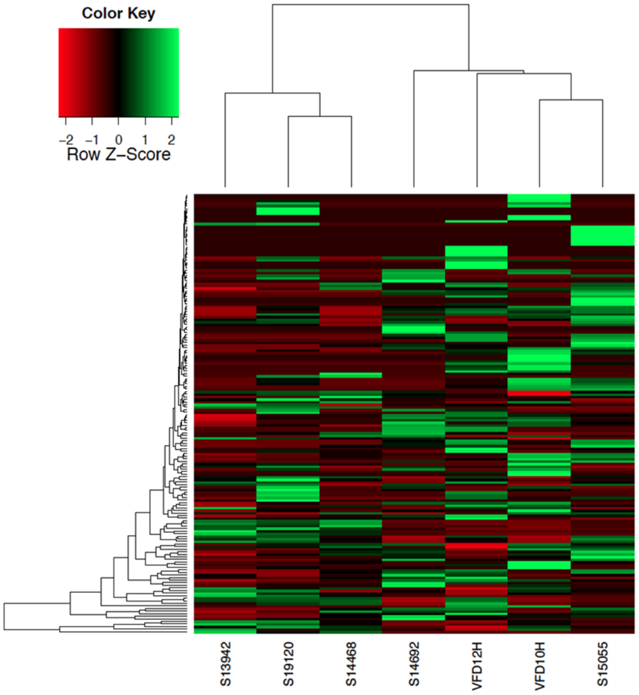


Figure 6. Double hierarchical dendrogram showing bacterial distribution at the species level for seven saliva samples. The relative values for bacterial species are depicted by color intensity, with legend indicated at the top of the figure.
doi:10.1371/journal.pone.0097699.g006

The cost of NGS has reduced dramatically and continues to decrease. It is clear that sequencing is not yet the least expensive method, but considering the amount of information obtained from NGS and the depth of resolution of the analyses, it is proving to be more cost-effective because of the greater breadth and depth of information provided compared to traditional methods involving culture and other bioassays, with battery of tests and reagents required. Metagenomics also provides opportunity to interrogate the same dataset against multiple databases (i.e, GeneBook libraries) for detection of bacteria, viruses and their virulence factors and/or antibiotic resistance in a single assay. It is concluded that application of GENIUS and GeneBook libraries can be utilized effectively for wider application in the clinical laboratory.

Supporting Information

Figure S1 Comparative analysis of human saliva sample VFD10 sequenced by Illumina GAIIx using GENIUS 5VCE, BLAST (NCBI, microbial subset) and short read mapping.
(TIFF)

Figure S2 Statistical analysis of confidence interval by GENIUS and visualization of the metagenomic community using the Krona visualization tool.
(TIFF)

Figure S3 GENIUS 5VCE prediction of species relative abundance in eight lanes of an Illumina GAIIx flowcell. The smaller chart to the upper right corner represents the number of reads generated per lane.
(TIFF)

Figure S4 GENIUS 5VCE prediction of percent of total hits for identified bacterial species in VFD10 and VFD12, sequenced by both GAIIx and HiSeq 2000. The smaller chart to the upper right corner shows number of reads generated for two samples by GAIIx and HiSeq 2000.
(TIFF)

Figure S5 Abundance of bacterial phyla and genera identified in the salivary microbiome by GENIUS.
(TIFF)

Figure S6 Distribution and abundance of different *Prevotella* spp. and strains in salivary microbiomes.
(TIFF)

Figure S7 Occurrence and relative abundance of bacterial species associated with periodontal disease.
(TIFF)

Figure S8 The centroid classification analysis of cluster A and B salivary samples. The top ranking markers can differentiate cluster B (green) and cluster A (red) by the centroid scores. Species and genera with nonzero components in each class are almost mutually exclusive.
(TIFF)

Table S1 Comparison of relative abundance of genera in VFD10-018 using different methods.
(DOCX)

Table S2 Comparison of relative abundance of genera in VFD12-006using different methods.
(DOCX)

Acknowledgments

The authors wish to extend thanks to the Nationwide Children's Hospital Biomedical Genomics Core, Columbus, Ohio for sequencing services.

Author Contributions

Conceived and designed the experiments: TAC RRC NAH BAY DMB SAF. Analyzed the data: NAH HL KS RI ASA SYC DMB ATM-S EMH NJM SAF BAY. Contributed reagents/materials/analysis tools: RRC HL BAY MLD. Wrote the paper: NAH RRC TAC HL RI KS SYC ATM-S NJM EMH BAY MLD. Critical review of the paper for intellectual content: RRC NAH TAC BAY MLD.

References

1. Socransky SS, Haffajee AD, Smith GL, Dzink JL (1987) Difficulties encountered in the search for the etiologic agents of destructive periodontal diseases. J Clin Periodontol 14: 588–593.
2. Haffajee AD, Socransky SS (1994) Microbial etiological agents of destructive periodontal diseases. Periodontol 2000 5: 78–111.
3. Marcotte H, Lavoie MC (1998) Oral microbial ecology and the role of salivary immunoglobulin A. Microbiol Mol Biol Rev 62: 71–109.
4. Weinstock GM (2012) Genomic approaches to studying the human microbiota. Nature 489: 250–256.
5. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, et al. (2012) Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res 22: 299–306.
6. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, et al. (2012) Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res 22: 292–298.
7. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, et al. (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472: 57–63.
8. Garrett WS, Gordon JI, Glimcher LH (2010) Homeostasis and inflammation in the intestine. Cell 140: 859–870.
9. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, et al. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444: 1027–1031.
10. Yamamoto ML, Maier I, Dang AT, Berry D, Liu J, et al. (2013) Intestinal bacteria modify lymphoma incidence and latency by affecting systemic inflammatory state, oxidative stress, and leukocyte genotoxicity. Cancer Res 73: 4222–4232.
11. Gonzalez A, Stombaugh J, Lozupone C, Turnbaugh PJ, Gordon JI, et al. (2011) The mind-body-microbial continuum. Dialogues Clin Neurosci 13: 55–62.
12. Wade WG (2013) The oral microbiome in health and disease. Pharmacol Res 69: 137–143.
13. Paster BJ, Boches SK, Galvin JL, Ericson RE, Lau CN, et al. (2001) Bacterial diversity in human subgingival plaque. J Bacteriol 183: 3770–3783.
14. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, et al. (2010) The human oral microbiome. J Bacteriol 192: 5002–5017.
15. Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, et al. (2005) 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 3: 27.
16. Slots J, Slots H (2011) Bacterial and viral pathogens in saliva: disease relationship and infectious risk. Periodontol 2000 55: 48–69.
17. Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simon-Soro A, et al. (2012) The oral metagenome in health and disease. ISME J 6: 46–56.
18. Watanabe T, Shibata K, Yoshikawa T, Dong L, Hasebe A, et al. (1998) Detection of Mycoplasma salivarium and Mycoplasma fermentans in synovial fluids of temporomandibular joints of patients with disorders in the joints. FEMS Immunol Med Microbiol 22: 241–246.
19. Wu T, Trevisan M, Genco RJ, Dorn JP, Falkner KL, et al. (2000) Periodontal disease and risk of cerebrovascular disease: the first national health and nutrition examination survey and its follow-up study. Arch Intern Med 160: 2749–2755.
20. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL Jr (1998) Microbial complexes in subgingival plaque. J Clin Periodontol 25: 134–144.
21. Faveri M, Mayer MP, Feres M, de Figueiredo LC, Dewhirst FE, et al. (2008) Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. Oral Microbiol Immunol 23: 112–118.
22. Colombo AP, Boches SK, Cotton SL, Goodson JM, Kent R, et al. (2009) Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray. J Periodontol 80: 1421–1432.
23. Crielaard W, Zaura E, Schuller AA, Huse SM, Montijn RC, et al. (2011) Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. BMC Med Genomics 4: 22.
24. Poveda-Roda R, Jimenez Y, Carbonell E, Gavalda C, Margaix-Munoz MM, et al. (2008) Bacteremia originating in the oral cavity. A review. Med Oral Patol Oral Cir Bucal 13: E355–362.
25. Parahitiyawa NB, Jin LJ, Leung WK, Yam WC, Samaranayake LP (2009) Microbiology of odontogenic bacteremia: beyond endocarditis. Clin Microbiol Rev 22: 46–64, Table of Contents.
26. Schiff E, Pick N, Oliven A, Odeh M (2003) Multiple liver abscesses after dental treatment. J Clin Gastroenterol 36: 369–371.
27. Franca AV, Martinelli A, Silva OC Jr (2004) Brain metastasis of hepatocellular carcinoma detected after liver transplantation. Arq Gastroenterol 41: 199–201.
28. Joshipura KJ, Hung HC, Rimm EB, Willett WC, Ascherio A (2003) Periodontal disease, tooth loss, and incidence of ischemic stroke. Stroke 34: 47–52.
29. Genco RJ, Grossi SG, Ho A, Nishimura F, Murayama Y (2005) A proposed model linking inflammation to obesity, diabetes, and periodontal infections. J Periodontol 76: 2075–2084.
30. Saremi A, Nelson RG, Tulloch-Reid M, Hanson RL, Sievers ML, et al. (2005) Periodontal disease and mortality in type 2 diabetes. Diabetes Care 28: 27–32.
31. Awano S, Ansai T, Takata Y, Soh I, Akifusa S, et al. (2008) Oral health and mortality risk from pneumonia in the elderly. J Dent Res 87: 334–339.
32. Buduneli N, Baylas H, Buduneli E, Turkoglu O, Kose T, et al. (2005) Periodontal infections and pre-term low birth weight: a case-control study. J Clin Periodontol 32: 174–181.
33. Palacios G, Druce J, Du L, Tran T, Birch C, et al. (2008) A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med 358: 991–998.
34. Lazarevic V, Whiteson K, François P, Schrenzel J (2010) The salivary microbiome, assessed by a high-throughput and culture-independent approach. Journal of Integrated OMICS 1: 28–35.
35. Cephas KD, Kim J, Mathai RA, Barry KA, Dowd SE, et al. (2011) Comparative analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or primary care givers using pyrosequencing. PLoS One 6: e23503.
36. Lazarevic V, Whiteson K, Gaia N, Gizard Y, Hernandez D, et al. (2012) Analysis of the salivary microbiome using culture-independent techniques. J Clin Bioinforma 2: 4.
37. Hu S, Loo JA, Wong DT (2007) Human saliva proteome analysis and disease biomarker discovery. Expert Rev Proteomics 4: 531–538.
38. Malamud D (2011) Saliva as a diagnostic fluid. Dent Clin North Am 55: 159–178.
39. Refulio Z, Rocafuerte M, de la Rosa M, Mendoza G, Chambrone L (2013) Association among stress, salivary cortisol levels, and chronic periodontitis. J Periodontal Implant Sci 43: 96–100.

40. Young EA, Aggen SH, Prescott CA, Kendler KS (2000) Similarity in saliva cortisol measures in monozygotic twins and the influence of past major depression. Biol Psychiatry 48: 70–74.
41. Nasidze I, Quinque D, Li J, Li M, Tang K, et al. (2009) Comparative analysis of human saliva microbiome diversity by barcoded pyrosequencing and cloning approaches. Anal Biochem 391: 64–68.
42. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, et al. (2010) The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database (Oxford) 2010: baq013.
43. Garrity GM, Thompson LM, Ussery DW, Paskin N, Baker D, et al. (2009) Studies on monitoring and tracking genetic resources: an executive summary. Stand Genomic Sci 1: 78–86.
44. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66–74.
45. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. Nat Methods 6: 639–641.
46. Shah N, Tang H, Doak TG, Ye Y (2011) Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. Pac Symp Biocomput: 165–176.
47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
48. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I (2014) RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res 42: D553–559.
49. Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P (2008) A bioinformatician's guide to metagenomics. Microbiol Mol Biol Rev 72: 557–578, Table of Contents.
50. Thomas T, Gilbert J, Meyer F (2012) Metagenomics - a guide from sampling to data analysis. Microb Inform Exp 2: 3.
51. Mavromatis K, Ivanova N, Barry K, Shapiro H, Goltsman E, et al. (2007) Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. Nat Methods 4: 495–500.
52. Martin J, Sykes S, Young S, Kota K, Sanka R, et al. (2012) Optimizing read mapping to reference genomes to determine composition and species prevalence in microbial communities. PLoS One 7: e36427.
53. Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17: 377–386.
54. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9: 386.
55. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW (2012) Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13: 601–612.
56. Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12: 385.
57. Lazarevic V, Whiteson K, Hernandez D, Francois P, Schrenzel J (2010) Study of inter- and intra-individual variations in the salivary microbiota. BMC Genomics 11: 523.
58. Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, et al. (2010) Direct sequencing of the human microbiome readily reveals community differences. Genome Biol 11: 210.
59. Chakravorty S, Helb D, Burday M, Connell N, Alland D (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods 69: 330–339.
60. Shakya M, Quince C, Campbell JH, Yang ZK, Schadt CW, et al. (2013) Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities. Environ Microbiol 15: 1882–1899.
61. Wang GC, Wang Y (1997) Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. Appl Environ Microbiol 63: 4645–4650.
62. Gevers D, Pop M, Schloss PD, Huttenhower C (2012) Bioinformatics for the Human Microbiome Project. PLoS Comput Biol 8: e1002779.
63. Keijser BJ, Zaura E, Huse SM, van der Vossen JM, Schuren FH, et al. (2008) Pyrosequencing analysis of the oral microflora of healthy adults. J Dent Res 87: 1016–1020.
64. Ling Z, Kong J, Jia P, Wei C, Wang Y, et al. (2010) Analysis of oral microbiota in children with dental caries by PCR-DGGE and barcoded pyrosequencing. Microb Ecol 60: 677–690.
65. Yang F, Zeng X, Ning K, Liu KL, Lo CC, et al. (2012) Saliva microbiomes distinguish caries-active from healthy human populations. ISME J 6: 1–10.
66. Zhou M, Rong R, Munro D, Zhu C, Gao X, et al. (2013) Investigation of the effect of type 2 diabetes mellitus on subgingival plaque microbiota by high-throughput 16S rDNA pyrosequencing. PLoS One 8: e61516.
67. Inagaki S, Onishi S, Kuramitsu HK, Sharma A (2006) Porphyromonas gingivalis vesicles enhance attachment, and the leucine-rich repeat BspA protein is required for invasion of epithelial cells by "Tannerella forsythia". Infect Immun 74: 5023–5028.
68. Kinane DF (2000) Aetiology and pathogenesis of periodontal disease. Ann R Australas Coll Dent Surg 15: 42–50.
69. Michalowicz BS, Diehl SR, Gunsolley JC, Sparks BS, Brooks CN, et al. (2000) Evidence of a substantial genetic basis for risk of adult periodontitis. J Periodontol 71: 1699–1707.
70. Tibshirani R, Hastie T, Narasimhan B, Chu G (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99: 6567–6572.
71. Ratnayake L, Olver WJ, Fardon T (2011) Aggregatibacter aphrophilus in a patient with recurrent empyema: a case report. J Med Case Rep 5: 448.
72. Lynn DJ, Kane JG, Parker RH (1977) Haemophilus parainfluenzae and influenzae endocarditis: a review of forty cases. Medicine (Baltimore) 56: 115–128.
73. Pai RK, Pergam SA, Kedia A, Cadman CS, Osborn LA (2004) Pacemaker lead infection secondary to Haemophilus parainfluenzae. Pacing Clin Electrophysiol 27: 1008–1010.