



OPEN ACCESS

EDITED BY

Stephen J. Bush,
University of Oxford, United Kingdom

REVIEWED BY

Nathan Olson,
National Institute of Standards and
Technology (NIST), United States
Jaiprakash Shewale,
Arcadia Consumer Healthcare, Inc.,
United States
Bradley W. Langhorst,
New England Biolabs, United States

*CORRESPONDENCE

Najim Ameziane,
✉ najimameziane@gmail.com

SPECIALTY SECTION

This article was submitted to Human and
Medical Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 27 October 2022

ACCEPTED 24 January 2023

PUBLISHED 07 February 2023

CITATION

Kumar A, Skrahina V, Atta J, Boettcher V,
Hanig N, Rolfs A, Oprea G and Ameziane N
(2023), Microbial contamination and
composition of oral samples subjected to
clinical whole genome sequencing.
Front. Genet. 14:1081424.
doi: 10.3389/fgene.2023.1081424

COPYRIGHT

© 2023 Kumar, Skrahina, Atta, Boettcher,
Hanig, Rolfs, Oprea and Ameziane. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Microbial contamination and composition of oral samples subjected to clinical whole genome sequencing

Abhishek Kumar¹, Volha Skrahina¹, Joshua Atta¹,
Veronika Boettcher², Nicola Hanig², Arndt Rolfs^{1,3}, Gabriela Oprea¹
and Najim Ameziane^{1*}

¹Arcensus Diagnostics, Rostock, Germany, ²CeGaT GmbH, Tübingen, Germany, ³Medical Faculty University of Rostock, Rostock, Germany

Biological material from the oral cavity is an excellent source of samples for genetic diagnostics. This is because collection is quick, easy-to-access, and non-invasive. We have set-up clinical whole genome sequence testing for patients with suspected hereditary disease. Beside the excellent quality of human DNA that can be isolated from such samples, we observed the presence of non-human DNA sequences at varying percentages. We investigated the proportion of non-human mapped reads (NHMR) sequenced from buccal swabs and saliva, the type of microbial genomes from which they were derived, and impact on molecular classification. Read sequences that did not map to the human reference genome were aligned to complete reference microbial reference sequences from the National Center for Biotechnology Information's (NCBI) RefSeq database using Kraken2. Out of 765 analyzed samples over 80% demonstrated more than 5% NHMRs. The majority of NHMRs were from bacterial genomes (average 69%, buccal swabs and 54% saliva), while the proportion of viruses was low, averaging 0.32% (buccal swabs) and 0.07% (saliva). We identified more than 30 different bacterial families of which *Streptococcus mitis* and *Rothia mucilaginosa* were the most common species. Importantly, the level of contamination did not impact the diagnostic yield.

KEYWORDS

WGS, gentic testing, genetic diagnostics, microbial contamination, molecular diagnostics

Introduction

Saliva and buccal samples are increasingly used for medical research studies, including those using modern “omics” platforms. The oral cavity is an excellent source of biological material for genetic and omics studies (Theda et al., 2018). Oral sampling is fast, easy-to-handle, and non-invasive compared to tissues such as blood (Juarez-Cedillo et al., 2010; Langie et al., 2016). Because of the simplicity and safety of the procedure, subjects can collect the samples themselves and samples can be sent by post.

We established a whole genome sequencing (WGS) testing pipeline for genetic diagnosis. While mapping the genomic data, we noticed that several samples have large amounts of reads not mapping to the human reference assembly hg38. Exploration of these reads indicated the presence of microbial DNA. In this study, we systematically examined the microbial composition of samples obtained from buccal swabs and saliva. In addition, we investigated whether the presence of microbial contamination had an impact on the molecular diagnosis of the subjects. A positive diagnosis is considered achieved if 1) a known pathogenic or likely

pathogenic variant is identified in a gene that can explain the symptoms in the respective subject or 2) a unique or rare variant (MAF < 0.001%) that is expected, based on multiple *in silico* pathogenicity predictions tools, to have a deleterious effect on transcript expression or protein function in a disease associated gene.

Results and discussion

The workflow for the identification of non-human mapped reads (NHMR) and their microbial composition from human whole genomic is performed using publicly available tools SAMtools (Li et al., 2009), Kraken2 (Wood et al., 2019) and Pavian (Breitwieser and Salzberg, 2020) and is schematically represented in Supplementary Figure S1.

Impact of non-human DNA contamination on sequencing statistics

A total of 765 samples were included in this study, of which 653 were from buccal swabs (Pakistan: 537, Albania: 112, Germany: 2, and Middle East: 6) and 112 (Pakistan: 95, Albania: 3, Germany: 8, and Middle East: 2) were from saliva. A median average of 777 million reads were generated per sample that passed Illumina chastity filter (<http://www.Illumina.com>), resulting in a median whole genome sequencing depth of coverage of 29x. Buccal swabs showed less variability in the total number of reads compared to saliva, while there were outliers with a maximum of 2.8 billion reads and a minimum of 479 million reads (Supplementary Figure S2). These differences can be explained due to pooling procedures prior to loading sample libraries on the sequencing flow cells. As expected, a clear correlation is observed between the total number of reads and the average depth of coverage, where the correlation coefficient is drastically enhanced after considering reads mapping to human reference genome only (Supplementary Figure S2). The percentage of NHMR for buccal swabs and saliva was similar.

The subject's gender did not affect the level of NHMR, while a slight difference in NHMRs was observed between samples from different geographical regions. Subjects from Germany demonstrated the lowest percentage of NHMR (Figure 1). However, since only 10 cases from Germany and eight cases from the Middle East were included, the case numbers were too low to establish a solid correlation between region's sample origin and NHMR.

Majority of NHMR is of bacterial origin and a minor fraction is of viral origin

Out of a total of 653 buccal swab and 112 saliva samples, 535 (82%) and 90 (80.4%) cases demonstrated more than 5% of NHMR, respectively (Figure 1). The samples were subsequently classified into low (<5% NHMR), medium (5%–20% NHMR) and high (>20% NHMR) NHMR categories (Figure 1; Supplementary Table S1). The median percentage of NHMRs was 13.6% for buccal swabs and 13.3% for saliva samples. The buccal swabs demonstrated extreme outliers with cases of NHMRs of up to 81%, while the highest NHMRs for saliva reached 48%. Absence of extreme outliers for saliva can be

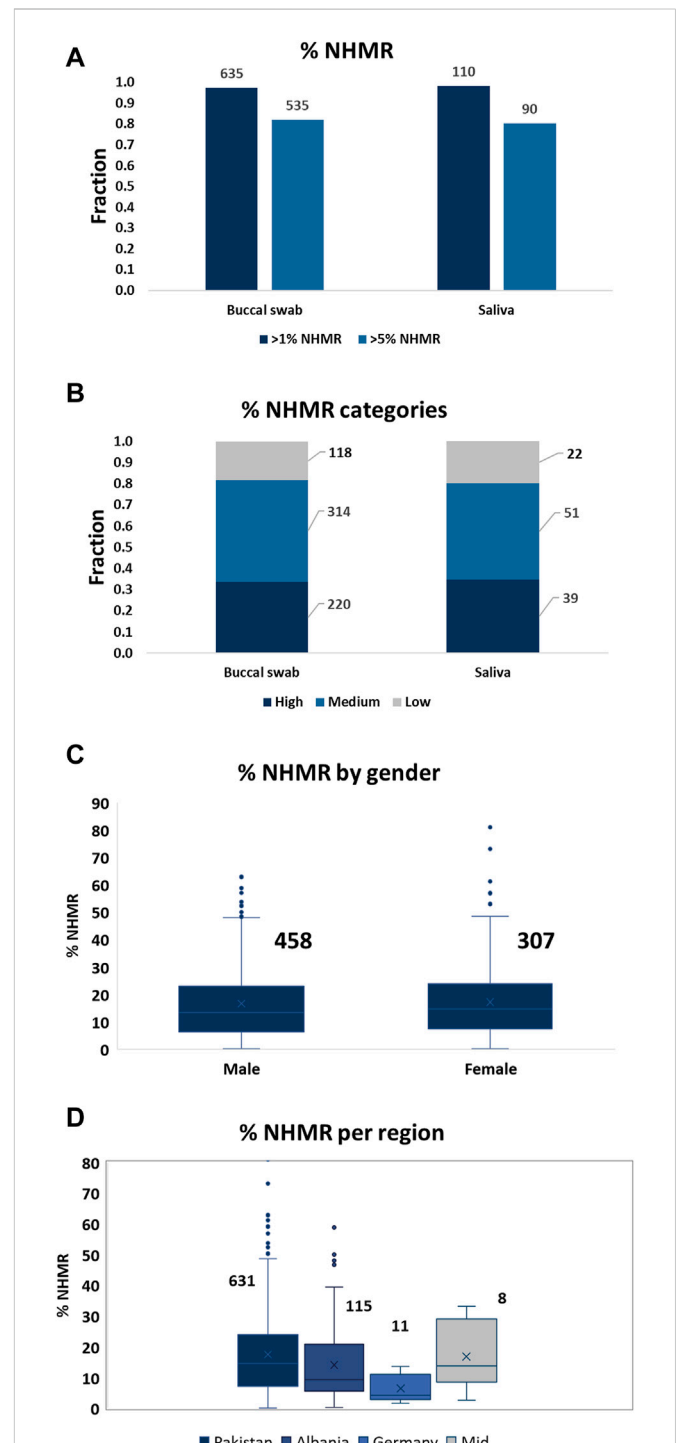


FIGURE 1

The level of NHMR among the samples, per gender, and per geographical region. (A) The distribution of samples with at least 1% or 5% NHMR for buccal swab and saliva samples. The total numbers for the categories are displayed on top of the bars. An unpaired t-test (Mann-Whitney test) demonstrated no significant difference in %NHMR between buccal swab and saliva samples (p -value = 0.8). (B) The distribution of samples into low (<5% NHMR), medium (5%–20% NHMR) and high (>20% NHMR) categories. The total number of cases per category are displayed on the side. (C) The distribution of % NHMR for samples obtained from males and females. The total numbers for the categories are displayed on top of the bars. The difference in %NHMR between males and females is not statistically significant. (D) The distribution of %NHMR for the samples from different geographical regions. Pakistan, 631; Albania, 115; Germany, 11; Middle East, 8.

TABLE 1 The prevalence of bacterial families and species among the samples.

	Buccal swab	Saliva
Bacterial families		
Streptococcaceae	611 (94%)	78 (70%)
Neisseriaceae	287 (44%)	53 (47%)
Pasteurellaceae	179 (27%)	31 (28%)
Actinomycetaceae	160 (25%)	19 (17%)
Micrococcaceae	151 (23%)	33 (29%)
Veillonellaceae	143 (22%)	26 (23%)
Prevotellaceae	106 (16%)	43 (38%)
Flavobacteriaceae	70 (11%)	5 (4%)
Corynebacteriaceae	35 (5%)	5 (4%)
Burkholderiaceae	24 (4%)	3 (3%)
Bacterial species		
<i>Streptococcus mitis</i>	418 (64%)	36 (32%)
<i>Streptococcus pneumoniae</i>	109 (17%)	6 (5%)
<i>Rothia mucilaginosa</i>	96 (15%)	24 (21%)
<i>Veillonella parvula</i>	65 (10%)	6 (5%)
<i>Prevotella melaninogenica</i>	39 (6%)	17 (15%)
<i>Streptococcus oralis</i>	37 (6%)	2 (2%)

explained by the low number of samples compared to the total number of buccal swabs. Indeed, the distribution of the NHMR for buccal swab and saliva samples seem to follow a similar pattern (Supplementary Figure S7).

Most NHMRs were mappable to bacterial genomes with a median average of 69% and 62% of NHMRs for buccal swabs and saliva samples, respectively. Considering at least 5% of NHMR aligning to microbial genomes, most samples demonstrated the presence of more than one bacterial family while in two saliva samples and six buccal swab samples six different bacterial families were detected (Supplementary Table S1). The presence of viral DNA was relatively rare. A median average of 0.05% and 0.03% of NHMRs mapped to viral genomes for buccal swabs and saliva samples, respectively (Supplementary Table S1). Outliers of up to 31.25% and 1.3% NHMRs mapping to viral genomes were observed for buccal swabs and saliva, respectively. Considering at least 0.5% of NHMRs aligning to microbial genomes, which accounts for about 200,000 reads, 23 buccal swab and two saliva samples demonstrated the presence of viral genomes. Expectedly, most of these viral DNA reads originated from bacteriophages (Supplementary Table S1). Exceptionally, some samples have higher viral contents like three buccal swab samples (arc000174 (31.25%), arc000372 (22.5%) and arc000274 (12.73%)) have >10% of NHMRs with viral origins. In all the three cases, the genetic etiology could be identified suggesting that the high viral abundance did not impact the molecular diagnosis.

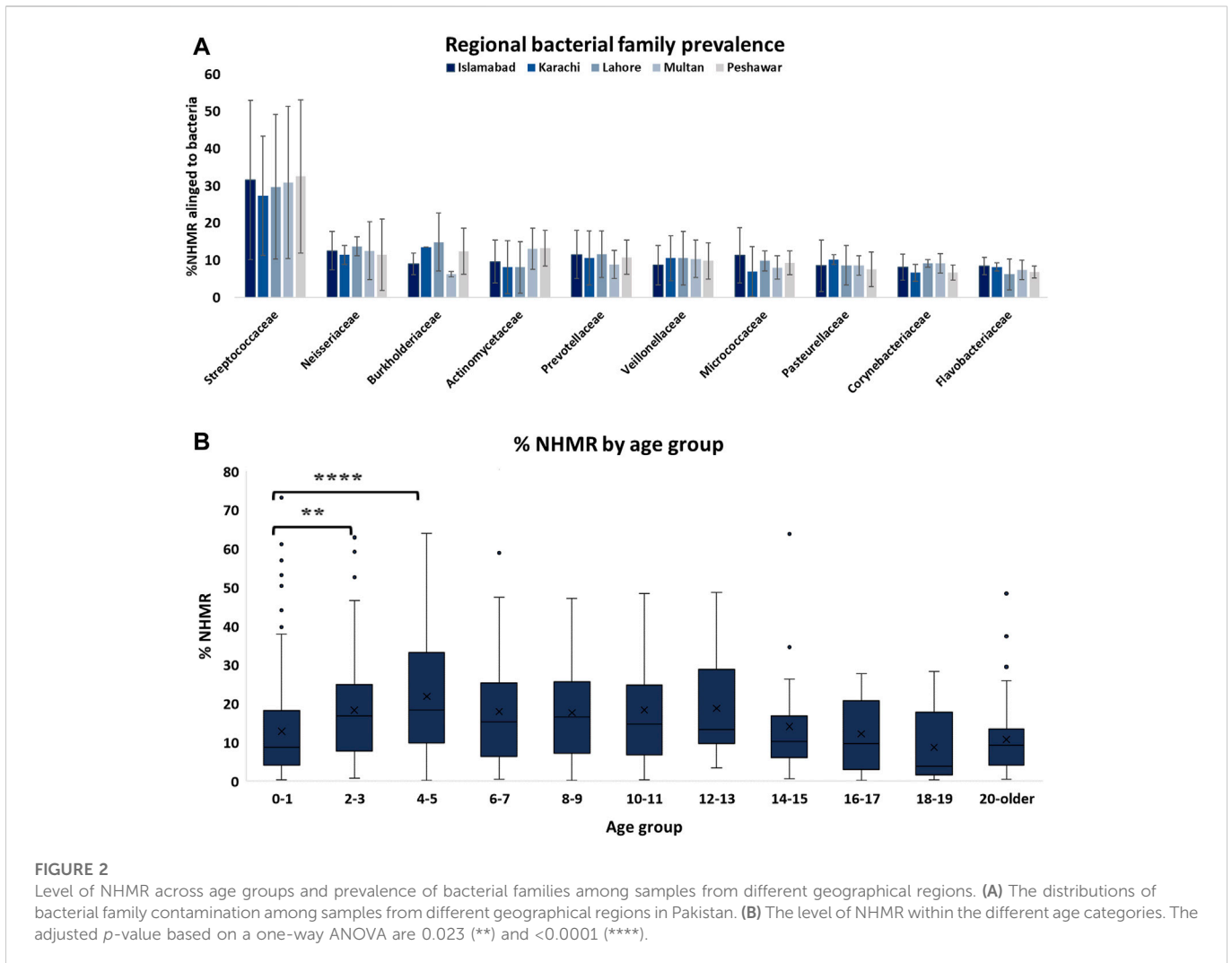
A median average of almost 32% of the NHMR could not be assigned to any of over 83 thousand addressed bacterial or viral genomes. These reads may fall into any or a combination of the following categories: a) bad quality reads due to sequencing errors that do not allow proper alignment to the interrogated genomes; b) the reads are of human origin, but complex genomic alterations in a sample does not allow proper initial alignment to the reference genome; c) the reference genome(s) of the bacterial and/or viral species from which the read originate are not interrogated by Kraken2; d) the reads originate from microorganisms other than bacteria and viruses, such as fungi and protozoa.

Streptococcaceae is the most prevalent microbial family with *Streptococcus mitis* as top bacterial species

Using a cut-off of >5% NHMRs and a minimum of 20 samples for which the same bacterial species is observed, Streptococcaceae was the most abundant bacterial family; identified in 611 buccal swab and 78 saliva samples, accounting for 94% and 70% the cases, respectively (Table 1). Other bacterial major families are Neisseriaceae, Pasteurellaceae, Actinomycetaceae, Micrococcaceae, Veillonellaceae and Prevotellaceae. *Streptococcus mitis* is the most prevalent in these samples, when considering >5% NHMRs and a minimum of five samples (Table 1). In previous studies, it was demonstrated that *Streptococcus* species, *S. mutans* (Psoter et al., 2011; Vieira et al., 2011) and *S. parasanguinis* (Mahfuz et al., 2013), are the major source of NHMRs in human oral cavity. Streptococci are the first inhabitants of the oral cavity with acquisition immediately after birth (Abranches et al., 2018). Hence, these bacteria are major components of the oral microbiota. Additionally, Streptococci are causative agents of dental caries including *S. mitis* (Banas et al., 2016). *S. mitis* serve as one of the major components of the oral streptococci classification systems (Whiley and Beighton, 1998; Richards et al., 2014). Oral Streptococci are potent producers of adhesive molecules, which are instrumental in colonizing these microbes within different oral tissues (Abranches et al., 2018). Taken together, it is evident that oral cavities have a suitable environment for growth of other Streptococci including *S. mitis*.

Microbial composition is variable across samples

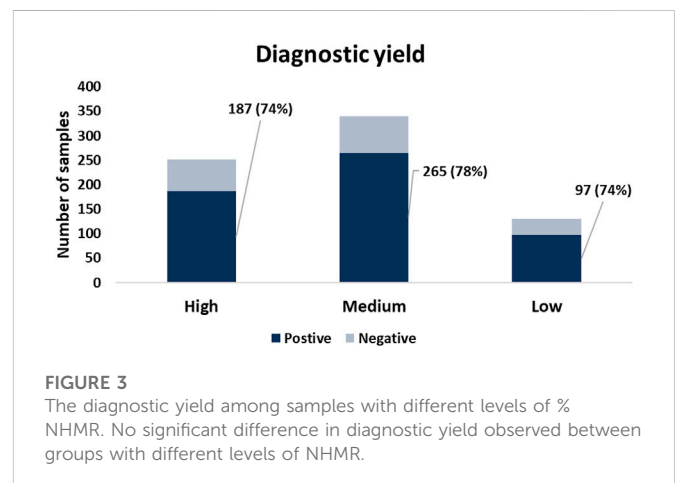
The level and composition of microbial species within samples was variable. When considering the top ten most prevalent bacterial families, no significant difference in contamination level from different bacterial families was observed between the samples from the different geographies (Figure 2). An increased level of NHMR was observed for individuals between the age of two and 13 years of age (Figure 2). A significant difference in % NHMR was identified when comparing the first age group (0–1) to age groups 2–3 and 4–5 but not when compared to older age groups. This may be explained by the absence of teeth in children under the first year of age and better oral hygiene for subjects of the age six and older. A diverse bacterial composition within subjects was observed, as is illustrated but not limited to subjects with the highest NHMRs (Supplementary Figure S4). Sample arc000199, with 19.11% NHMRs, has 65.6 million reads (M) aligning to *Ottowia* sp. oral taxon 894. Other major bacterial components prevalent in the same sample (Supplementary Figure S4), are *Bergeyella cardium* (32.1 M), *Rothia mucilaginosa* (22.3 M). The microbial composition for the different samples is shown in Supplementary Figure S3, with *S. mitis* observed in all samples at quantities and the highest abundance in sample arc000427. Buccal swab sample arc000019 has various bacterial families (Supplementary Figure S4) with top three species: *Neisseria subflava* (18.7 M), *Prevotella melaninogenica* (16.1 M), and *Haemophilus parainfluenzae* (9.1 M). Similarly, buccal swab sample arc000204 has major bacterial fragments



(Supplementary Figure S4) as *P. melaninogenica* (18.3 M), *P. jejuni* (16.3 M), and *Veillonella parvula* (15.9 M). Saliva-based sample arc000701 has several bacterial components (Supplementary Figure S4), the top three of which are *P. melaninogenica* (6.0 M), *N. subflava* (4.3 M), and *P. intermedia* (4.3 M). Saliva sample arc000743 has major bacterial fragments (Supplementary Figure S4) of which the top three are *N. sicca* (9.99 M), *P. melaninogenica* (9.7 M), and *Fusobacterium nucleatum* (7.3 M).

High prevalence of beta herpesvirus 5 in buccal swabs

Three buccal swab samples (samples: arc000174 arc000372 and arc000274) demonstrated more than 10% of NHMRs aligning to viral genomes (Supplementary Table S1) compared to an average of 0.32% for all buccal swab samples. The NHMRs in these samples were predominantly aligning to beta-herpesvirus 5 with a total of 4.92, 2.48, and 2.09 M in arc000174, arc000372 and arc000274, respectively (Supplementary Figure S5). When the threshold of NHMRs aligning to viral genomes was lowered to >0.5%, beta-herpesvirus 5 was identified in 28 samples (2 saliva samples and 26 buccal swab



samples). Three other prominent viral family genomes that were detected in the samples were Myoviridae (5x), Podoviridae (5x) and Siphoviridae (7x). Notably, these three viral families have bacteria as their natural host, which could explain their high prevalence.

No impact of microbial contamination on genetic diagnosis

To assess whether the presence of NHMR had impact on the molecular diagnosis of the subjects, we focussed on subjects with either a positive or a negative primary molecular classification. Cases were categorized in low, medium, and high contamination groups based on NHMRs of less than 5% (131), >5–20% (340), and >20% (252), respectively. Remarkably, the diagnostic yield was highly comparable between the low (74%), medium (78%) and high (74%) contamination groups (Figure 3). The chi-square test demonstrated a *p*-value of 0.49, indicating no significant difference between groups. A genetic diagnosis could be established for cases with NHMRs over 70%. Also, for both buccal swab and saliva samples, the level of NHMR did not demonstrate a significant impact on diagnostic yield (unpaired *t*-test, *p*-values 0.27 and 0.84, respectively. Supplementary Figure S6). The findings for all cases are summarized in Supplementary Table S1. Altogether, although samples collected from the oral cavity for the purpose of clinical molecular diagnostics are often contaminated by bacterial and viral species, the diagnostic yield is not impacted. Due to its safety, simplicity and ease, oral sampling can be considered an excellent DNA collection approach in a clinical diagnostic setting.

In this study, we systematically investigated the origin of non-human mapped reads of samples obtained from the oral cavity that were whole genome sequenced for clinical genetic testing. Over 70% of cases demonstrated the presence of more than 5% microbial contamination. The contamination ranged from a single microbial species to up to six species in a single sample. Nevertheless, the contamination did not negatively impact the genetic diagnosis of the subjects. The reference genomes of the identified highly prevalent microbial species can be added to the human reference genome as decoy in routine clinical diagnostic WGS analysis to improve read alignment and variant calling, especially for samples collected from the oral cavity.

Methods

Sample collection

During hospital visits, patients suspected to suffer from a genetic disease were sampled by physicians who, in parallel, carefully evaluated and documented the phenotypes. Consent to participate in research projects was obtained for all subjects. Buccal swabs and saliva samples were collected using DNA ORAcollect (Genotek, Canada) OCR-100 and OG-500 kits, respectively. The participants or guardians of the participating children signed written informed consent forms. The study was approved by the Ethics Committee of Rostock University (Germany), no. A2022-0072, 25.04 2022.

DNA preparation and WGS

DNA extraction from buccal swabs and saliva was performed using prepIT.L2P kit (DNA Genotek, Canada). DNA libraries for WGS were prepared using TruSeq Nano DNA Library Prep Kit

(Illumina) and sequenced with the 150 bp paired-end protocol on an Illumina platform to yield an average coverage depth of 30x for the nuclear genome and at least 1000x for the mitochondrial genome.

Data analysis

Raw reads were mapped to reference human genome assembly GRCh38.p13 using bwa (version 0.7.17-r1188) software with the mem algorithm (Li and Durbin, 2009). A bam file subset was created containing reads not aligned to human genome (flagged “-f4”) using SAMtools version 1.10⁴ after which these reads were extracted using SAMtools bamtofastq module (Li et al., 2009) to generate paired-end zipped fastq files of non-human mapped reads (NHMR). Next, these fastq files were analyzed with Kraken version 2.0.8-beta (Wood et al., 2019) against the complete reference microbial reference sequences from the National Center for Biotechnology Information’s (NCBI) RefSeq database; minikraken2_v2_8_GB_201904. A complete list of interrogated genomes can be found in Supplementary Table S1. For the visualization of the microbial composition, Kraken2 output files were taken to create Sankey diagrams using Pavian (Breitwieser and Salzberg, 2020). A schematic overview of the procedure is illustrated in Supplementary Figure S1.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: <https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA917836>.

Ethics statement

Studies involving human participants were reviewed and approved by Ethics Committee of Rostock University (Germany), no. A2022-0072, 25.04 2022. Written informed consent to participate in this study was provided by the participants’ legal guardian/next of kin.

Author contributions

Conceived the study: AK, AR, GO, and NA; Designed the study: AK, AR, GO, and NA; Conducted the analysis: AK and NA; Data generation: VS, JA, VB, and NH; Interpreted the results: AK, NA; Drafted the manuscript: AK and NA; Made critical revisions to the manuscript: VS, AR, GO, and NA; All authors read and approved the final manuscript.

Conflict of interest

AK, VS, JA, AR, GO, and NA were employed by the Company Arcensus Diagnostics.

VB and NH were employed by the company CeGaT GmbH.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1081424/full#supplementary-material>

SUPPLEMENTARY FIGURE S1

Schematic overview of the bioinformatic pipeline used to explore microbial presence and composition of samples collected from the oral cavity.

SUPPLEMENTARY FIGURE S2

Sequencing statistics. (A) The distribution of total reads obtained for buccal swabs and saliva samples. (B) The average depth of coverage distribution for the samples. (C) The correlation between total reads and average depth of coverage before and (D) after considering exclusively reads mapped to human reference genome.

SUPPLEMENTARY FIGURE S3

Microbial composition of buccal swab samples. The distribution of NHMR aligning to the different bacterial and viral species for samples with the highest percentage of NHMR. The number of reads aligning to the different species are indicated with colored bars as illustrate the frequency among species. The top panel indicates the percentage of reads aligned to the human reference genome GRCh38.p13, starting with 19.11% to 90.41%.

SUPPLEMENTARY FIGURE S4

Phylogenetic tree of microbial composition for samples arc000199 (A), arc000019 (B), arc000204 (C), arc000701 (D), arc000743 (E).

SUPPLEMENTARY FIGURE S5

Phylogenetic tree of microbial composition of buccal swab samples containing high beta herpesvirus 6 levels. D-Domain, K-Kingdom, P- Phylum, C-Class, O-Order, F-Family, G-Genus, and S-Species. The numbers at the nodes indicate the total NHMR aligned at the respective level. (A) Sample arc000174. (B) Sample arc000372. (C) Sample arc000274.

SUPPLEMENTARY FIGURE S6

The level of NHMR does not affect the diagnostic yield. Box plots showing the distribution of NHMR in percentages (Y-axis) for cases with either a positive or negative molecular diagnostic finding for buccal swab (A) and saliva (B) samples.

SUPPLEMENTARY FIGURE S7

The NHMR distribution of the collected samples. A density plot demonstrating the distribution of the level NHMR for buccal swab and saliva samples. The percentage of NHMR is plotted on the X-axis.

SUPPLEMENTARY TABLE S1

Overview of 765 samples collected either by buccal swab ($n = 653$) or saliva ($n = 112$). TotalReads: total number of sequencing reads. averageDP: average whole genome depth of coverage. Reads aligned %: Percentage of reads aligned to human reference genome. %NHMR: percentage of non-human mapped reads. % Bacteria (of NHMR): percentage of NHMR aligned to bacterial genomes. %Viruses (of NHMR): percentage of NHMR aligned to viral genomes. Material: Sample collection method. Country: Country origin of sample. DOB: Date of birth of the individual from whom the sample was collected. YOB: Year of birth. Gender: sex of the sample. Status: Provided clinical status of proband. Diagnostic_Outcome: Whether genetic etiology for proband was detected. NHMR_Cat: The category of NHMR, high/medium/low. Number of Species: number of detected microbial species, which have a minimum of 200 k reads. Top3Species: The top three detected species ordered from most abundant to less abundant.

SUPPLEMENTARY TABLE S2

Lists of microbial genomes separated in two sheets: A bacterial- and a viral genomes sheet.

References

- Abranches, J., Zeng, L., Kajfasz, J. K., Palmer, S. R., ChakraBorty, B., Wen, Z. T., et al. (2018). Biology of oral streptococci. *Microbiol. Spectr.* 6. doi:10.1128/microbiolspec.GPP3-0042-2018
- Banas, J. A., Zhu, M., Dawson, D. V., Cao, H., and Levy, S. M. (2016). PCR-based identification of oral streptococcal species. *Int. J. Dent.* 2016, 3465163. doi:10.1155/2016/3465163
- Breitwieser, F. P., and Salzberg, S. L. (2020). Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* 36, 1303–1304. doi:10.1093/bioinformatics/btz715
- Juarez-Cedillo, T., Sanchez-Garcia, S., Mould-Quevedo, J. F., GarCia-Pena, C., Gallo, J. J., Wagner, F. A., et al. (2010). Cost-effective analysis of genotyping using oral cells in the geriatric population. *Genet. Mol. Res.* 9, 1886–1895. doi:10.4238/vol9-3gmr939
- Langie, S. A., Szarc Vel Szic, K., Declerck, K., Traen, S., Koppen, G., Van Camp, G., et al. (2016). Whole-genome saliva and blood DNA methylation profiling in individuals with a respiratory allergy. *PLoS One* 11, e0151109. doi:10.1371/journal.pone.0151109
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Mahfuz, I., Cheng, W., and White, S. J. (2013). Identification of *Streptococcus parasanguinis* DNA contamination in human buccal DNA samples. *BMC Res. Notes* 6, 481. doi:10.1186/1756-0500-6-481
- Psoter, W. J., Ge, Y., Russell, S. L., Chen, Z., Katz, R. V., Jean-Charles, G., et al. (2011). PCR detection of *Streptococcus mutans* and *Aggregatibacter actinomycetemcomitans* in dental plaque samples from Haitian adolescents. *Clin. Oral Investig.* 15, 461–469. doi:10.1007/s00784-010-0413-y
- Richards, V. P., Palmer, S. R., Pavinski Bitar, P. D., Qin, X., Weinstock, G. M., Highlander, S. K., et al. (2014). Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. *Genome Biol. Evol.* 6, 741–753. doi:10.1093/gbe/evu048
- Theda, C., Hwang, S. H., Czajko, A., Loke, Y. J., Leong, P., and Craig, J. M. (2018). Quantitation of the cellular content of saliva and buccal swab samples. *Sci. Rep.* 8, 6944. doi:10.1038/s41598-018-25311-0
- Vieira, A. R., Deeley, K. B., Callahan, N. F., Noel, J. B., Anjomshoa, I., Carricato, W. M., et al. (2011). Detection of *Streptococcus mutans* genomic DNA in human DNA samples extracted from saliva and blood. *ISRN Dent.* 2011, 543561. doi:10.5402/2011/543561
- Whiley, R. A., and Beighton, D. (1998). Current classification of the oral streptococci. *Oral Microbiol. Immunol.* 13, 195–216. doi:10.1111/j.1399-302x.1998.tb00698.x
- Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257. doi:10.1186/s13059-019-1891-0