

RESEARCH ARTICLE

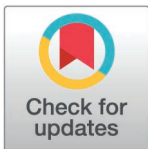
Shotgun metagenomic analysis of the oral microbiomes of children with noma

Michael Olaleye^{1☯}, Angus M. O’Ferrall^{2,3☯}, Richard N. Goodman^{2,3☯}, Deogracia Wa Kabila¹, Miriam Peters¹, Gregoire Falq⁴, Joseph Samuel¹, Donal Doyle⁵, Diana Gomez⁵, Gbemisola Oloruntu¹, Shafi’u Isah⁶, Adeniyi Semiyu Adetunji⁶, Elise Farley¹, Nicholas. J. Evans⁷, Mark Sherlock⁵, Adam P. Roberts^{2,3*}, Mohana Amirtharajah^{5*}, Stuart Ainsworth^{1☯7*}

1 Noma Children’s Hospital, Médecins Sans Frontières, Sokoto, Nigeria, **2** Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom, **3** Centre for Neglected Tropical Diseases, Liverpool School of Tropical Medicine, Liverpool, United Kingdom, **4** Manson Unit, Médecins Sans Frontières, London, United Kingdom, **5** Operational Centre Amsterdam, Médecins Sans Frontières, Amsterdam, The Netherlands, **6** Department of Clinical Services, Noma Children’s Hospital, Sokoto, Nigeria, **7** Department of Infection Biology and Microbiomes, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom

☯ These authors contributed equally

* adam.roberts@lstmed.ac.uk (APR), mohana.amirtharajah@amsterdam.msf.org (MA), stuart.ainsworth@liverpool.ac.uk (SA)



OPEN ACCESS

Citation: Olaleye M, O’Ferrall AM, Goodman RN, Kabila DW, Peters M, Falq G, et al. (2026) Shotgun metagenomic analysis of the oral microbiomes of children with noma. *PLoS Negl Trop Dis* 20(3): e0014118. <https://doi.org/10.1371/journal.pntd.0014118>

Editor: Michael Marks, London School of Hygiene and Tropical Medicine, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

Received: November 11, 2025

Accepted: March 5, 2026

Published: March 20, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pntd.0014118>

Copyright: © 2026 Olaleye et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/),

Abstract

Noma is a rapidly progressive orofacial gangrene that predominantly affects children living in extreme poverty. Despite its documentation since antiquity and its designation as a World Health Organisation Neglected Tropical Disease in 2023, the microbiological cause of noma remains poorly understood, with no specific organisms confidently identified as definitive aetiological agents. Here, we present the first deep shotgun metagenomic profiling of oral saliva microbiomes from 19 Nigerian children with acute noma. Our analyses of this preliminary study reveal marked microbial dysbiosis in noma microbiomes, with machine learning and multivariate statistical analyses indicating significant enrichment of *Treponema*, *Porphyromonas*, and *Bacteroides*, alongside depletion of *Streptococcus* and *Rothia*, as key microbial signatures of noma disease. From the dataset we recovered 40 high-quality *Treponema* metagenome assembled genomes (MAGs) spanning 19 species, 14 of which were novel. Notably, a novel species designated *Treponema* sp. A was detected in 15 of the 19 noma participants and was entirely absent from an internationally representative set of healthy saliva metagenomes. Re-analysis of previously published 16S rRNA datasets from children with noma in Niger also revealed *Treponema* sp. A to be highly prevalent in noma cases but extremely rare in controls. While these findings highlight *Treponema*, particularly *Treponema* sp. A, as an organism of interest and a potential contributor to noma pathogenesis, further comprehensive studies will be required to confirm this association and to clarify whether it reflects a causal role and/or is a genuine marker of noma dysbiosis. Additionally, analysis of antimicrobial

which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Data availability All anonymized data used in this study is available in supplementary materials. Short reads from all samples sequenced in this study have been deposited in the Sequence Read Archive (SRA) under BioProject PRJNA1273078, with accession numbers SRR33858003–SRR33858030. All accessions for MAGs defined in this project can be found in S1 Data. Code availability statement All code to reproduce this analysis is publicly available through GitHub: <https://rngoodman.github.io/noma-metagenomics>.

Funding: This study was funded by Médecins Sans Frontières (MSF). M.O., D.W.K., M.P., G.F., J.S., D.D., G.O., E.F., M.S., and M.A. are employees of MSF and contributed to the study design, data collection, analysis, and manuscript preparation as members of the author team. A.M.O. is supported by the Medical Research Council via the Liverpool School of Tropical Medicine-Lancaster University doctoral training partnership (grant no. MR/W007037/1). R.N.G and A.P.R. are supported by funding from the Medical Research Council (MRC), Biotechnology and Biological Sciences Research Council (BBSRC) and Natural Environmental Research Council (NERC), which are all Councils of UK Research and Innovation (Grant no. MR/W030578/1) under the umbrella of the JPIAMR - Joint Programming Initiative on Antimicrobial Resistance via the STRESST project, and from UKRI through the Strength in Places Fund (grant no. S1PF 36348), as part of the infection innovation Consortium (iiCON). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

resistance determinants detected in noma metagenomes revealed concerning levels of resistance to antibiotics commonly used in noma treatment, particularly β -lactams and metronidazole, especially among *Prevotella* spp. These findings provide the first high-resolution microbial framework for noma and offer a foundation for future research into its pathogenesis and the development of novel diagnostics, therapeutics, and preventive strategies in endemic settings.

Author summary

Noma is a rapidly progressing necrotic disease that destroys the tissues of the mouth and face, mainly affecting children living in abject poverty. Although antibiotic treatment can be lifesaving, the microbes involved in noma are still largely unknown, making prevention and early diagnosis difficult. To help address this, we used metagenomic DNA sequencing to profile bacteria from the mouths of 19 Nigerian children with noma treated at the Noma Children's Hospital in Sokoto, and compared these results with publicly available datasets from healthy individuals. The oral microbiomes of children with noma were remarkably different from healthy controls, with higher levels of certain bacterial species including *Treponema*, *Porphyromonas*, and *Bacteroides* and lower levels of bacteria commonly seen in healthy oral cavities, including *Streptococcus* and *Rothia*. We also identified a previously undescribed *Treponema* species in oral microbiomes of children with noma in this study and found it was also common in earlier noma samples from an independent study in Niger. Finally, we detected genes associated with resistance to antibiotics commonly used to treat noma, including beta-lactams and metronidazole. These findings provide a foundation for larger studies which will hopefully one day lead to diagnostics, targeted therapies, and preventive strategies for noma in endemic settings.

Introduction

Noma (*cancrum oris*) is an orofacial gangrene of unknown etiology that predominantly affects children (most often aged 2–6 years) living in chronic poverty with limited access to healthcare [1]. Once initiated, noma rapidly destroys the soft tissues of the face, including the cheeks, lips, nose and occasionally the eyes [2]. It is estimated that, without treatment, noma case-fatality can be as high as ~90%; however, early access to appropriate care is associated with markedly lower mortality [3]. Survivors are often left with permanent debilitating scars and sequelae often requiring complex surgical intervention [4,5]. The social and economic impact of noma on impoverished communities whilst not thoroughly documented, is thought to be substantial [6] resulting in noma being declared a Neglected Tropical Disease by World Health Organization (WHO) in 2023 [7].

Determining global incidence of noma has been notoriously difficult due to lack of data, with the most widely cited figure suggesting 140,000 children a year may suffer from this severe disease [8], although the veracity of this figure has been questioned [9]. The vast majority of reported cases are from West and East Africa, particularly Nigeria as there is a large research presence there [10–12]. However, noma has a global distribution, appearing in many countries with high levels of chronic malnourishment and poor healthcare, or in immunocompromised individuals [9]. Indeed, historically, noma was reported in Europe and North America during the 19th century [13], was referred to as “trench mouth” during the first world war [14], and was routinely observed among inmates of the Nazi concentration camps of the second world war [15].

Under the current WHO classification of noma progression, introduced in 2016 [8], the disease initiates as acute necrotizing ulcerative gingival lesion (stage 1), before progressing through oedema (stage 2), gangrene (stage 3), scarring (stage 4), and sequalae (stage 5). Prompt intervention in the early stages of noma, including aggressive antibiotic treatment, nutritional supplementation, and surgical debridement, can lead to substantial reduction in mortality, albeit often still with significant sequalae and life-long socio-economic implications.

Although the major risk factors of noma are well understood and include chronic malnourishment and immunosuppression [3,16–18], the mechanism of initiation of infection, the associated microorganisms, and the role of the host immune response, are poorly understood or unknown [19]. No single etiological agent has been identified as the cause of noma, although its susceptibility to aggressive antibiotic treatment certainly suggests it is bacterial in nature. It has long been suspected that noma occurs, similar to other periodontal diseases, due to a dysbiosis in normal healthy functional oral microbiome and dysregulation or dysfunction in immune responses, driven by living in abject poverty [19,20].

Despite no single etiological agent being identified, microbiological surveys have suggested bacterial taxa which may be involved in noma pathology. A recent systematic review of published evidence on noma identified 121 taxa identified as being associated, at some point, with noma [3]. Organisms have primarily been associated with noma through culture or, to a lesser extent, 16S rRNA sequencing [3,16,19–26]. Studies focused on the microbiomes of noma participants reported substantially different microbiomes to controls, frequently identifying *Prevotella intermedia*, various spirochetes and *Fusobacterium nucleatum* as indicators of noma microbiota. However, these studies have noted technical limitations in regard to sampling sites (e.g., sampling dead tissue) and methodological biases (inherent biased amplification 16S rRNA genes, partial sequencing, and discriminatory culture media), which are acknowledged as not being able to provide an unbiased appraisal of a microbiome [19,27,28], thus not enabling definitive descriptions of noma related microbiomes.

Shotgun metagenomics overcomes many of the limitations of previous microbiological survey techniques employed for noma. The technique sequences all DNA present within a sample, thus avoiding any amplification or primer biases, providing an unbiased reflection of a metagenome [29]. Additionally, It can also offer insights into the functional potential of microorganisms in the sample, allowing retrieval of whole genomes and analysis of specific traits, such as antimicrobial resistance [30]. As far as we are aware, shotgun metagenomic sequencing has yet to be utilised in noma research [19].

Here we present a preliminary investigation on the first microbiological profiling of noma using deep shotgun metagenomics, describing the oral metagenomes of 19 participants with acute noma and retrospectively analysing existing data sets to enable us to profile the microbiological taxa we found associated with the disease.

Results

During the recruitment period, 53 Stage 1–4 noma participants aged two to 12 years were admitted at the Noma Children’s Hospital. No stage 0 participants at risk of noma were admitted during the study period. We enrolled 21 of these participants, however one participant was unable to provide a sample after enrolment, so was excluded from the study, leaving us with a total of 20 participants, while another participant’s samples failed extraction (see “Total DNA extraction and metagenome sequencing” below), leaving a total of 19 enrolled participants in the study. Of the 32 participants not enrolled in the study, 18 had recently taken antibiotics, eight had been hospitalized in different health facilities before admission, five were missed due to arriving out of hours, and one had wound dressing before admission at the hospital (S1 Table).

Sociodemographic and disease progression characteristics of enrolled participants

Amongst the 19 enrolled participants ([S1 Table](#)), the majority ($n = 16$, 84%) were aged two to six years, while the remaining participants were between seven and 10 years ($n = 3$, 16%). The participants included 10 males (52.6%) and 9 females (47.4%).

One participant was diagnosed upon admission with Stage 1 noma (acute necrotizing ulcerative gingivitis, 5.3%), 9 with Stage 2 (odema, 47%), seven with Stage 3 (gangrene, 37%), and two with Stage 4 (scarring, 11%) ([S1 Table](#)). All stage 2 and 3 participants were admitted for treatment with antibiotics and wound dressing, while the only stage 1 participant was an outpatient who received treatment but did not have anthropometric measurements recorded. All stage 1–3 participants recovered from their initial infection after receiving treatment. The two stage 4 participants were admitted for physiotherapy.

The majority of participants were from Sokoto State ($n = 9$, 47.4%), with the remainder coming from neighbouring states of Zamfara ($n = 6$, 31.6%) and Kebbi ($n = 2$, 10.5%), whilst a single participant each from more distant Kaduna and Kano states was enrolled ([S2 Table](#)). The Noma Children's Hospital was the first healthcare facility visited by most participants ($n = 15$, 78.9%) ([S3 Table](#)). Of the four who had visited a health facility/ provider before visiting the Noma Children's Hospital, three (75%) had visited a primary health facility and the other a chemist/drugstore. None of these four participants were provided with any treatment at the facilities before they were referred to the Noma Children's Hospital. Amongst the 19 participants, two (10.5%) experienced challenges in seeking healthcare at the Noma Children's Hospital, both citing expensive transportation as an issue ([S3 Table](#)). Two participants (10.5%) had visited an oral health care provider for a check-up in the past year, both at primary health facilities.

Eight caregivers (42.1%) indicated that their child fell sick seven to 14 days before hospitalization at the Noma Children's Hospital, while seven caregivers (36.8%) reported that their infants became sick two to six days prior to admission ([S1 Table](#)). Participants reported several first symptoms including a painful mouth ($n = 18$, 94.7%), foul mouth odour ($n = 14$, 73.7%), red gums ($n = 11$, 57.9%), swollen cheek ($n = 10$, 52.6%) bleeding gums ($n = 8$, 42.1%) or a hole in their cheek ($n = 6$, 31.6%). Six (31.6%) of the participants had visible bleeding from the site of the wound at the time of sampling ([S1 Table](#)).

Most of the participants who were aged two to five years were not acutely malnourished, having MUAC of 12.5 cm and above, while two participants in this age bracket had moderate to severe malnutrition ([S1 Table](#)). The majority of participants ($n = 19$, 95%) had not previously received Haemophilus influenzae b, Human papillomavirus, Meningococcal A and C vaccines, with most ($n = 17$, 85%) also not having received Diphtheria, Measles, or BCG vaccines, either. The Poliovirus vaccine was an exception, with 50% coverage ($n = 10$).

Total DNA extraction and metagenome sequencing

Three saliva samples and nine swab samples failed total DNA extraction, yielding insufficient DNA for sequencing. In total, noma oral metagenomes were obtained for 19 individuals (swab and saliva $n = 9$, saliva only $n = 8$, swab only $n = 2$). The number of raw read pairs per sample ranged from 40.23 million to 80.36 million (mean = 55.30 million), with between 2.76 and 75.56 million (mean = 12.44 million) reads remaining after trimming and host DNA removal ([S4 Table](#)). These reads had a mean Q20 of 98.73% (sd = 0.32%) and a mean Q30 of 95.52% (sd = 0.95%).

Assessing the impact of sample type and categorical variables on noma metagenomic diversity

Initially we wished to determine whether there were any differences between swab and saliva samples collected from the same participants, as well as to assess overall differences between swab and saliva samples across the entire dataset. Comparison of the swab and saliva metagenomes obtained from the same participants ($n = 9$) demonstrates that the relative abundance of genera is similar across the two types of samples ([Fig 1A](#)) and the sample type (i.e., swab vs

saliva) was not a significant factor when using PERMANOVA ($R^2=0.01$, $p=0.946$). We then used the Wilcoxon signed-rank test with Benjamini-Hochberg False Discovery Rate (FDR) correction for multiple comparisons to look at specific genera between swab and saliva samples with adjusted p-values represented as q-values. With the exception of samples from participant N9, which demonstrated a highly anomalous microbiome dominated by *Escherichia*, all other participants with both saliva and swab samples (i.e., N4, N5, N6, N7, N10, N13, N17, N18) had no significant differences between relative abundances of *Prevotella* ($q=0.784$), *Treponema* ($q=0.91$), *Neisseria* ($q=0.784$), *Bacteroides* ($q=0.784$), *Filifactor* ($q=0.68$), *Porphyromonas* ($q=0.808$), *Fusobacterium* ($q=0.797$), *Escherichia* ($q=0.784$), *Selenomonas* ($q=0.688$), *Aggregatibacter* ($q=0.688$) and *Capnocytophaga* ($q=0.6875$).

Relative abundance of top genera in noma and global healthy saliva dataset microbiomes

Twenty publicly available shotgun metagenomic saliva samples from healthy individuals from three countries (USA, Japan and Denmark) [31–33] were used to compare against the noma saliva samples from this study (S6 Table). We found significant differences between countries within the global healthy saliva dataset samples ($p=0.001$) with anosim, mrpp, and permanova. When assessing the differences in specific genera within the global healthy saliva dataset, *Treponema* ($p=0.001$), *Bacteroides* ($p=0.019$), *Porphyromonas* ($p=0.004$), *Fusobacterium* ($p=0.001$), *Streptococcus* ($p=0.001$), and *Rothia* ($p=0.001$) were all significantly different between countries. However, when using differential analysis to compare the normalised counts of genera in the noma dataset against each of the healthy datasets, the noma-associated shifts in *Treponema*, *Rothia*, *Veillonella*, *Gemella*, and *Streptococcus* far exceeded the residual variation among healthy samples (S2 Fig).

Next, we compared the relative abundance of the top 22 genera between the noma and global healthy saliva dataset samples (Fig 1B) using a Wilcoxon rank-sum test, with p-values adjusted for multiple comparisons using the Benjamini-Hochberg FDR method. We also calculated the \log_2 fold change (LFC), relative to noma samples, between the medians of the healthy and noma datasets. Of the 22 genera, 19 were significantly different between noma and global healthy saliva dataset samples (q -values <0.05) with only *Campylobacter*, *Haemophilus*, and *Leptotrichia* not demonstrating significant differences. S3 Fig shows the differences between the top 22 genera; those more abundant in noma samples include *Treponema* (LFC = +5.79, $q=1.38 \times 10^{-9}$), *Bacteroides* (LFC = +2.17, $q=6.45 \times 10^{-9}$), *Filifactor* (LFC = +4.84, $q=2.34 \times 10^{-7}$), *Porphyromonas* (LFC = +3.15, $q=4.79 \times 10^{-7}$), *Fusobacterium* (LFC = +2.13, $q=1.53 \times 10^{-4}$), *Escherichia* (LFC = +1.08, $q=2.24 \times 10^{-4}$), *Selenomonas* (LFC = +2.55, $q=3.25 \times 10^{-4}$), *Aggregatibacter* (LFC = +3.19, $q=1.34 \times 10^{-3}$), *Neisseria* (LFC = +1.87, $q=0.016$), *Prevotella* (LFC = +1.39, $q=0.021$), and *Capnocytophaga* (LFC = +1.97, $q=0.021$). Those more abundant in the global healthy saliva dataset include *Streptococcus* (LFC = -3.41, $q=1.38 \times 10^{-9}$), *Rothia* (LFC = -6.68, $q=4.63 \times 10^{-8}$), *Schaalia* (LFC = -6.59, $q=2.34 \times 10^{-7}$), *Veillonella* (LFC = -3.7, $q=3.53 \times 10^{-6}$), *Actinomyces* (LFC = -4.03, $q=3.98 \times 10^{-6}$) and *Gemella* (LFC = -1.16, $q=6.13 \times 10^{-4}$).

Differential analysis between noma and global healthy saliva samples indicate genera which are significantly different between healthy and diseased states

In order to further determine which genera were significantly different between the global healthy saliva dataset and the noma saliva samples, we ran a differential analysis of normalised count data using the DESeq2 model (Fig 2A) [34]. Given that the dataset contained 1883 genera, many with very low counts, we focused our analysis on the most prominent taxa. We filtered for genera with a mean relative abundance greater than 1% across all noma and healthy samples, resulting in a core set of 16 genera for comparison (S4 Fig).

Of these 16 genera we identified 11 genera that were differently abundant between the noma and healthy groups (q -value <0.05). Four genera were significantly enriched in the noma samples (Fig 2C and 2E) including *Treponema* (LFC = +4.65, $q=4.13 \times 10^{-23}$), *Porphyromonas* (LFC = +2.68, $q=2.41 \times 10^{-10}$), *Bacteroides* (LFC = +2.04, $q=2.21 \times 10^{-7}$) and *Selenomonas* (LFC = +1.02, $q=2.62 \times 10^{-2}$). Seven genera were significantly depleted in the noma samples and

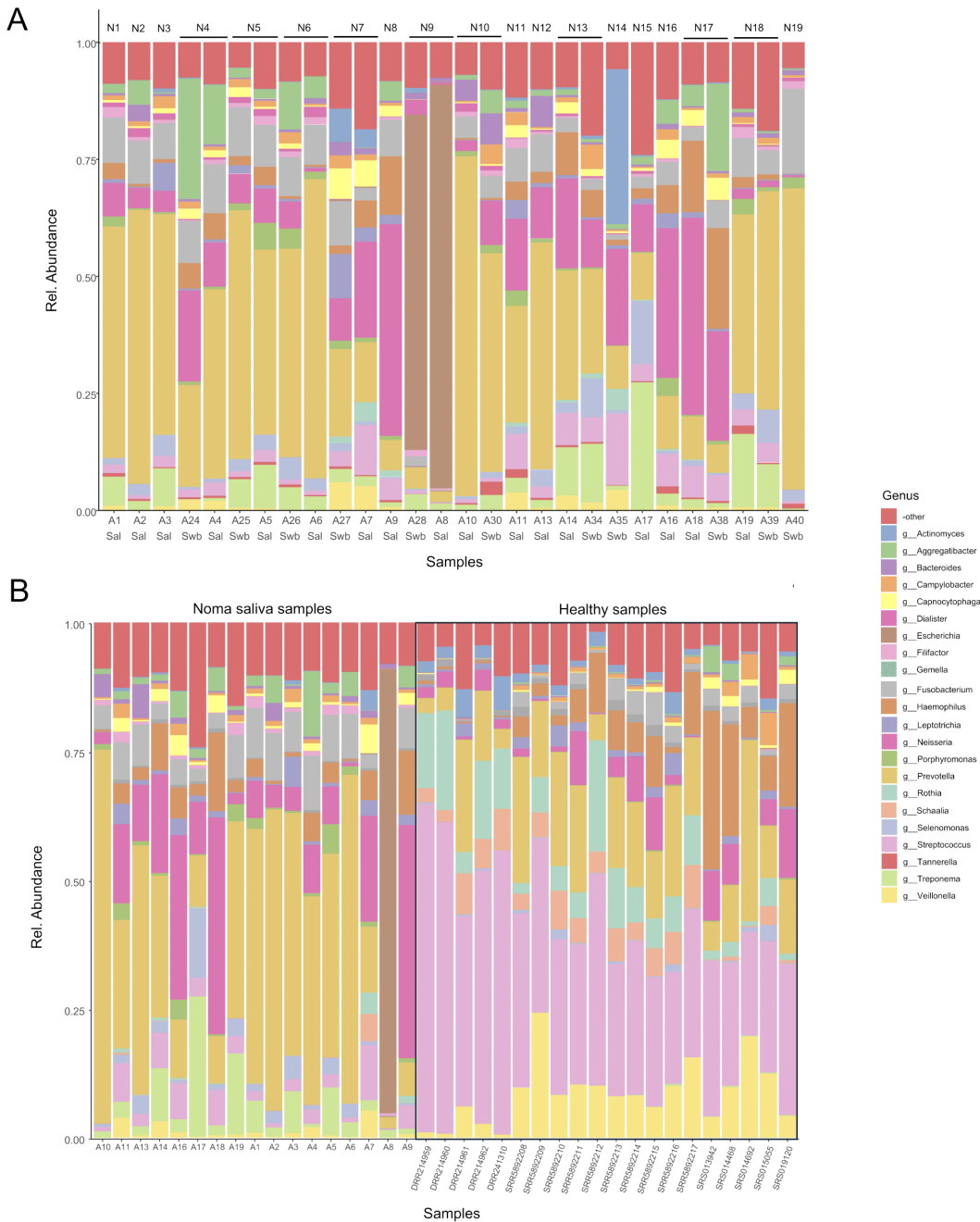


Fig 1. Genus level comparison of relative abundance between noma samples and healthy controls. The relative abundance of the top 22 most abundant genera is shown, with all other genera displayed under “-other”. (A) Saliva and swab samples from noma participants. Participant IDs (N1-N19) are displayed on top and sample names (A1-A40) are displayed on the bottom, along with whether the sample came from saliva (Sal) or swab (Swb). (B) saliva samples from noma participants compared to a global dataset of healthy saliva samples (n = 20).

<https://doi.org/10.1371/journal.pntd.0014118.g001>

show significantly higher median counts in the global healthy saliva dataset (Fig 2B and 2D), these included *Streptococcus* (LFC = -4.55, $q = 2.64 \times 10^{-31}$), *Veillonella* (LFC = -3.90, $q = 2.45 \times 10^{-13}$), *Gemella* (LFC = -2.35, $q = 4.48 \times 10^{-11}$), *Schaalia* (LFC = -4.85, $q = 4.89 \times 10^{-11}$), *Rothia* (LFC = -5.61, $q = 9.61 \times 10^{-11}$), *Actinomyces* (LFC = -3.82, $q = 2.14 \times 10^{-5}$), and

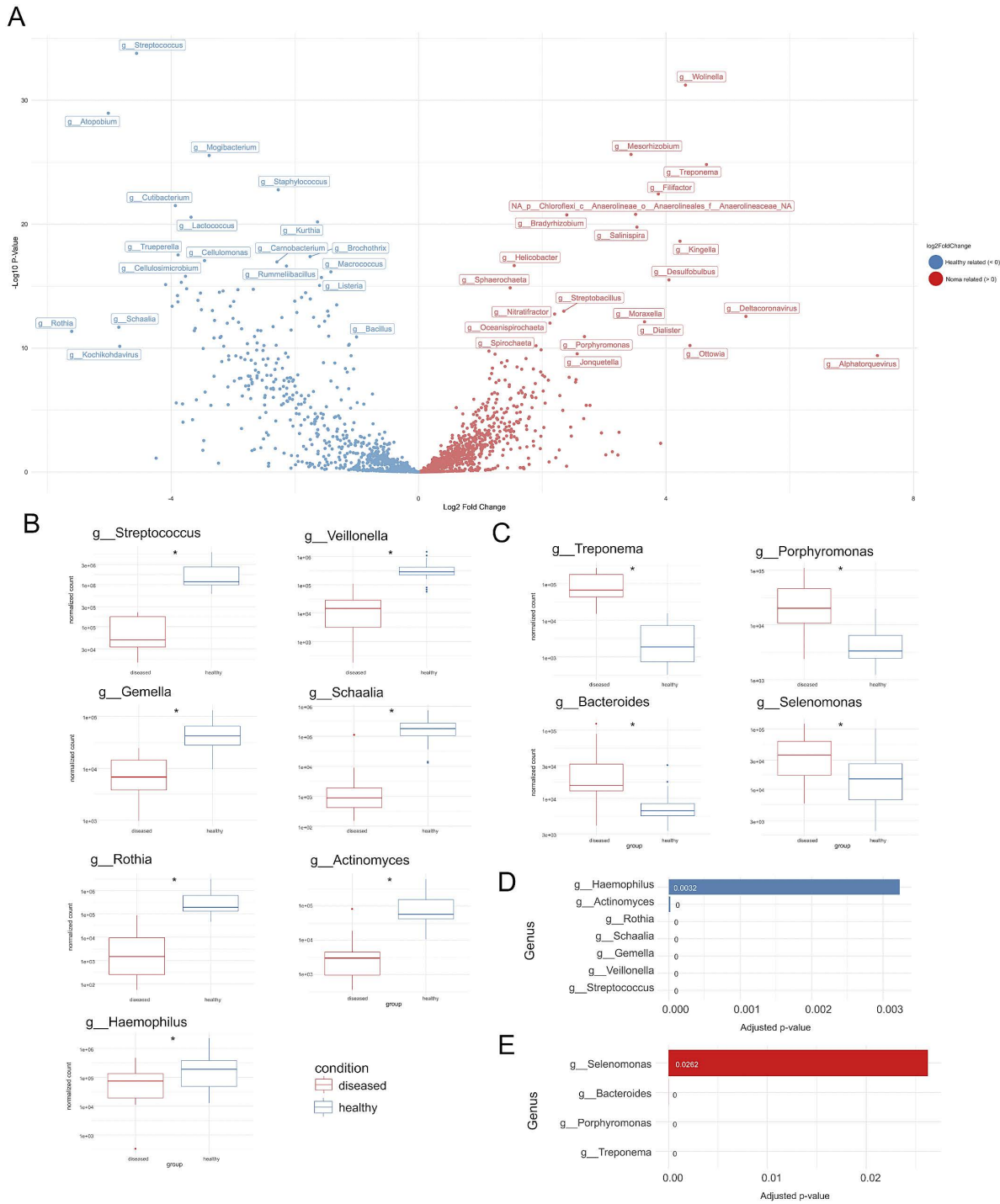


Fig 2. Differential analysis between noma and healthy samples. The Deseq model was used to calculate adjusted p-values and fold change for all genera. (A) A volcano plot with the log₂ fold-change (healthy vs noma) across the x axis and the -log₁₀ p-values across the y-axis, genera which have a positive fold change (> 0) are enriched in noma samples and genera which have a negative fold change (< 0) are enriched in healthy control samples. The normalized counts were plotted as boxplots for the genera with a mean relative abundance of more than 1% (top 1%) that are significantly related to healthy control (B) and noma samples (C). The boxplots summarise the distribution of data points, with the box representing the interquartile range (IQR), the horizontal line within the box representing the median and whiskers extending to 1.5 times the IQR, any outliers beyond this are shown as individual points. The adjusted p-values (q-values) are displayed as barplots for the top genera significantly related to healthy control (D) and noma samples (E).

<https://doi.org/10.1371/journal.pntd.0014118.g002>

Haemophilus (LFC = -1.92, $q = 3.2 \times 10^{-3}$). Some genera such as *Bacteroides* and *Veillonella* show a greater degree of variability as shown by the presence of outliers. The distribution of normalised counts between diseased and healthy groups for the 11 genera are visualized in Fig 2B and 2C. The adjusted p-values (q-values) are visualized in Fig 2D and 2E.

Machine learning and multivariate statistical analyses indicate genera which contribute most to healthy or diseased states

To further determine which genera are most strongly associated with healthy and diseased states, we ran three independent additional analyses; random forests, ordination (PCoA), and hierarchical clustering/CLR transformation (Fig 3). A random forests classifier was used to determine how well categorical metadata variables (noma vs healthy) predict differences in the microbial communities. Each genus was treated as a predictor for the metadata classification. The importance of individual genera in distinguishing between the healthy and diseased states was assessed by those with the highest percent increase in mean squared error (%IncMSE) derived from the random forest model. From the top most abundant genera (S4 Fig), these included typically orally pathogenic *Treponema* (%IncMSE = 3.85), *Bacteroides* (2.43), *Porphyromonas* (2.07) and orally commensal *Streptococcus*, (2.51) and *Rothia* (1.94) (Fig 3A). These genera had the greatest influence on the model prediction and therefore contribute the most to the distinction between noma and global healthy saliva dataset samples.

A cross validation of the dataset (see Methods) was shown to exhibit complete prediction with an area under the receiver operating characteristic (ROC) curve (AUC) of 1.0 (Fig 3B). This shows that there is a distinct separation between the microbiomes of global healthy saliva dataset and noma samples. This near-perfect classification reflects the profound magnitude of dysbiosis in noma, where the microbial community structure is fundamentally distinct from that of healthy individuals (as visualized in the non-overlapping clusters of the PCoA ordination shown in Fig 3D). As an AUC of 1.0 indicates such a clear distinction between healthy and diseased samples, a permutation test ($n = 1000$) was conducted to evaluate the statistical significance of the observed classification accuracy (see Methods). The mean accuracy of the permuted models was significantly lower than the observed accuracy of the RF model (Fig 3C), providing strong evidence that the models performance is not due to random chance or model over-parameterisation. However, this distinct separation should be interpreted in the context of the study limitations; the comparison against a global healthy dataset, rather than geographically matched controls, likely maximises the distinguishability of the two groups by combining disease-specific signals with geographic microbiome variations.

Principal Coordinates Analysis (PCoA) with Bray-Curtis dissimilarity was used to display variation across samples in relation to disease state (Fig 3D). There was a distinct separation between noma samples and global healthy saliva dataset samples. A PERMANOVA analysis on the Bray-Curtis dissimilarity-matrix confirmed this ($R^2 = 0.40$, $p = 0.001$). The relative abundance of bacterial genera was subjected to CLR transformation and z-score standardization, then the top 20 most abundant genera were selected for analysis and visualization. Fig 3E shows the genus-level variation across samples ranging from a CLR z-score of -6 (lower abundance) to +6 (higher abundance), hierarchically clustered by the complete method and Euclidean distances. Notably, certain genera exhibited disease-state-specific enrichment patterns, some genera had a higher mean CLR z-score in noma samples such as *Treponema* (0.90), *Filifactor* (0.86), *Porphyromonas* (0.75), and *Bacteroides* (0.70), which reveals they are more abundant in noma samples. Other genera had a higher mean CLR z-score in global healthy saliva dataset samples such as *Streptococcus* (0.80), *Schaalia* (0.80), *Rothia* (0.75), and *Actinomyces* (0.72) showing they are more abundant in global healthy saliva dataset samples (Fig 3F). Taken together this reveals which genera contribute the most to a healthy oral microbiome or a microbiome associated with noma.

Recovery of novel *Treponema* sp. MAGs from noma participant microbiome samples

We next sought to obtain MAGs from species associated by multivariate and machine learning analysis with disease status in noma metagenomes through a binning process. *Treponema* MAGs made up 25 of 139 high-quality MAGs identified

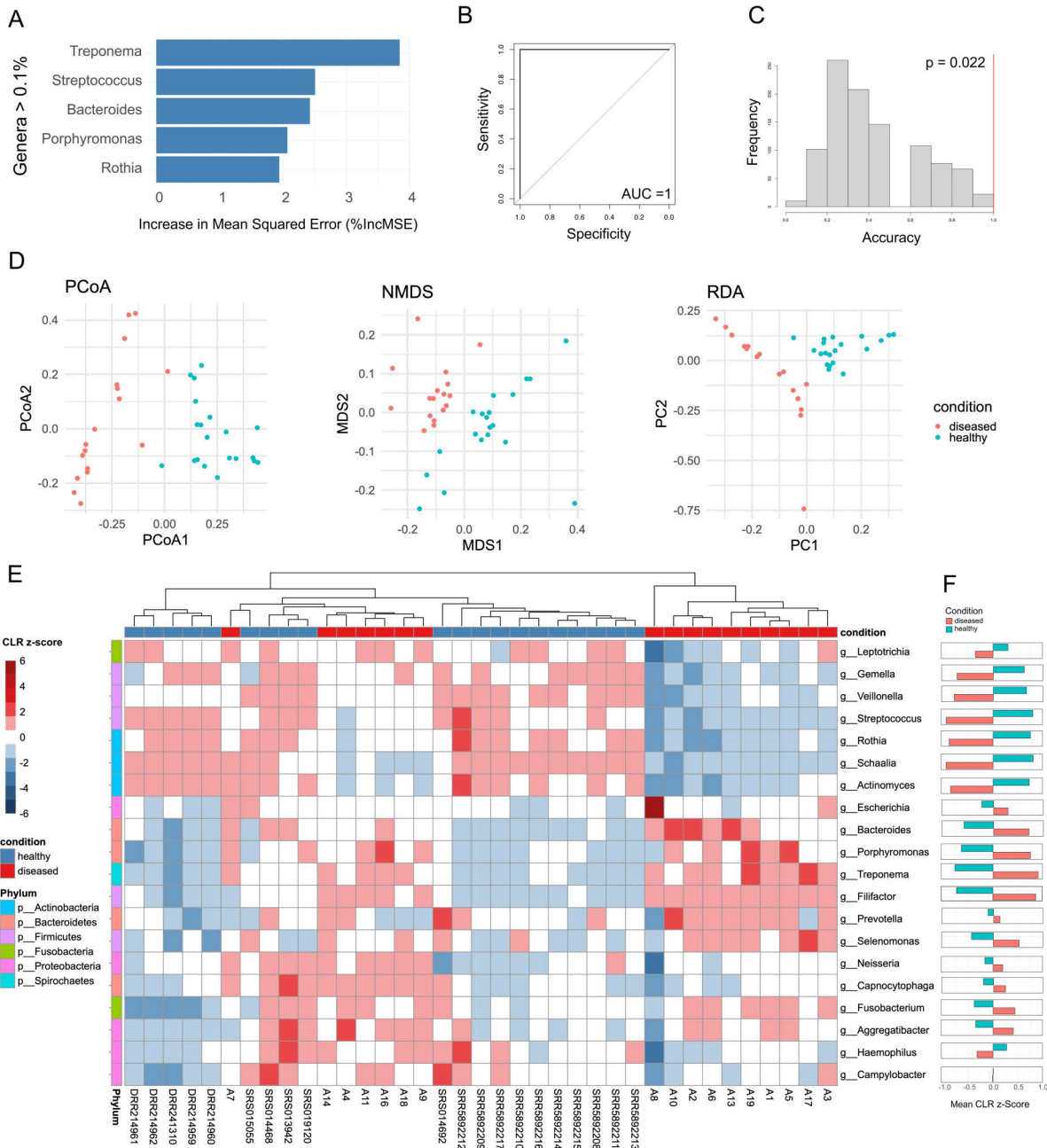


Fig 3. Machine learning and multivariate statistical analyses. Several analyses were applied to determine the differences between healthy and diseased states and which genera contributed the most to this difference. **(A)** The genera with the highest percent increase in mean squared error (%IncMSE) showing they contributed most to the distinction between healthy and diseased in the random forest model **(B)** A receiver operating characteristic (ROC) curve from cross validation of random forest model **(C)** Null distribution of accuracy from permutation testing of the cross validation of the random forest model. This displays the classification accuracy across 1000 permutations, with the observed accuracy of the RF model against our data is shown as a red line. **(D)** Ordination of multivariate data into two dimensions using Principal Coordinate Analysis (PCoA) with Bray-Curtis dissimilarity. Points are coloured by disease state: healthy control (blue) and noma samples (red) **(E)** A heatmap displaying the z-score standardized centred log-ratio (CLR) of the top 20 most abundant genera, rows represent bacteria genera classified by Phylum and columns represent samples classified by disease state. A higher CLR z-score (Darker red) signifies higher abundance, and a lower score (darker blue) signifies lower abundance. The samples are hierarchically clustered by the complete method using Euclidean distances. **(F)** Mean CLR Z-scores by condition shown for each genus.

<https://doi.org/10.1371/journal.pntd.0014118.g003>

in total. Five high-quality *Bacteroides* MAGs were recovered, of which three could be identified at species level (two *Bacteroides heparinolyticus* and one *Bacteroides pyogenes*), while seven high-quality *Porphyromonas* MAGs were recovered (six of which were identified at species level: three *Porphyromonas gingivalis* and three *Porphyromonas endodontalis*).

When including medium-quality MAGs in further analysis targeting the most strongly disease associated genus, we recovered 40 unique medium- or high-quality *Treponema* MAGs (Fig 4). The MAGs belonged to 19 species, of which 14 were novel and could not be assigned to a known species in the genome taxonomy database (Fig 4A). The remaining five species, each of which only appeared in one MAG across all samples, were identifiable at species level against the genome taxonomy database, and were therefore not deemed novel: *Treponema lecithinolyticum*, *Treponema maltophilum*, *Treponema medium*, *Treponema* sp014334325, *Treponema* sp905372235. We positioned all recovered high-quality *Treponema* sp. MAGs in the context of previously characterised oral treponeme reference strains within NCBI RefSeq, as well as the agent of syphilis, *Treponema pallidum* subsp. *pallidum*. A maximum-likelihood phylogenetic tree constructed based on single nucleotide polymorphisms (SNPs) in soft core genes (Fig 4B) showed that novel *Treponema* species, as well as known species, were distributed across diverse clades with varying degrees of relatedness to oral reference strains, emphasizing the previously uncharacterized diversity of *Treponema* in the oral microbiome.

'*Treponema* sp. A' was the most prevalent novel *Treponema* species and was present as a medium- or high-quality MAG in samples from 11/19 participants (57.9%) (S5 Table). We observed marked genetic diversity within *Treponema* sp. A, in which the average nucleotide identity (ANI) between MAGs varied upwards of 96% (Fig 4C), indicating the presence of various strains. Three *Treponema* sp. A MAGs (Avii, Aviii, Axi) recovered from different participants appeared to be very closely related, with 99.8% ANI in each pairwise comparison. Five other novel *Treponema* species were present in samples from more than one noma participant, but none at a prevalence of 30% or more.

One high-quality MAG recovered from *Treponema* sp. A (Aix) contained a 926 bp partial 16S fragment which shared 99.46% identity with a partial 16S rRNA gene from uncultured *Treponema* identified within a subgingival plaque sample from a noma participant published in 2012 (GenBank Accession no. AM420013) [21]. We downloaded the 1505 bp AM420013 sequence to use as a 16S rRNA reference gene for *Treponema* sp. A, to screen across 16S rRNA genes recovered in our samples which could not be binned to a putative MAG. In addition to the eleven participants from which a medium- or high-quality *Treponema* sp. A MAG was recovered from swab or saliva samples, we identified a further four participants with partial *Treponema* sp. A 16S rRNA genes in their microbiomes. Through a combination of MAG and 16S rRNA gene recovery, we therefore detected *Treponema* sp. A in 15/19 noma participants assessed in this study, including 90% (8/9) of Stage 2 participants with oedema, and 100% (7/7) of Stage 3 participants with gangrenous lesions (S5 Table). *Treponema* sp. A was not detected in the participant with Stage 1 disease, nor the two participants with Stage 4 disease (scarring). Neither *Treponema* sp. A MAGs nor 16S rRNA genes were recoverable from the 20 metagenomes in the healthy global saliva dataset.

We examined previous metagenomic data sets for the presence of *Treponema* sp. A. Available 16S rRNA gene fragments (V1–3 region) from a study in Niger, in which samples from 12 healthy children, 12 children with Stage 1 noma (referred to as acute necrotizing ulcerative gingivitis in this study) and 12 children with established noma disease were collected [20], to determine the prevalence of *Treponema* sp. A within the different groups of their study by aligning 16S rRNA gene fragments to the AM420013 16S rRNA reference gene with BLAST. If 99% identity between 16S rRNA gene fragments was considered as a same species hit, *Treponema* sp. A was detected in 2/12 (16.7%), 10/12 (83.3%) and 7/12 (58.3%) of healthy, Stage 1 noma and established noma samples, respectively. Chi-square testing confirmed significant differences between the groups ($p=0.004$). If the same-species threshold was dropped to 98%, all 12 samples from participants with stage 1 noma contained *Treponema* sp. A, while the prevalence within the healthy and established noma groups remained the same (2/12). However, sequences of the variable regions within 16S rRNA genes cannot provide the same level of taxonomic resolution as entire gene sequences, and sequence fragments with 97–99% identity should be treated with caution when inferring same species [35].

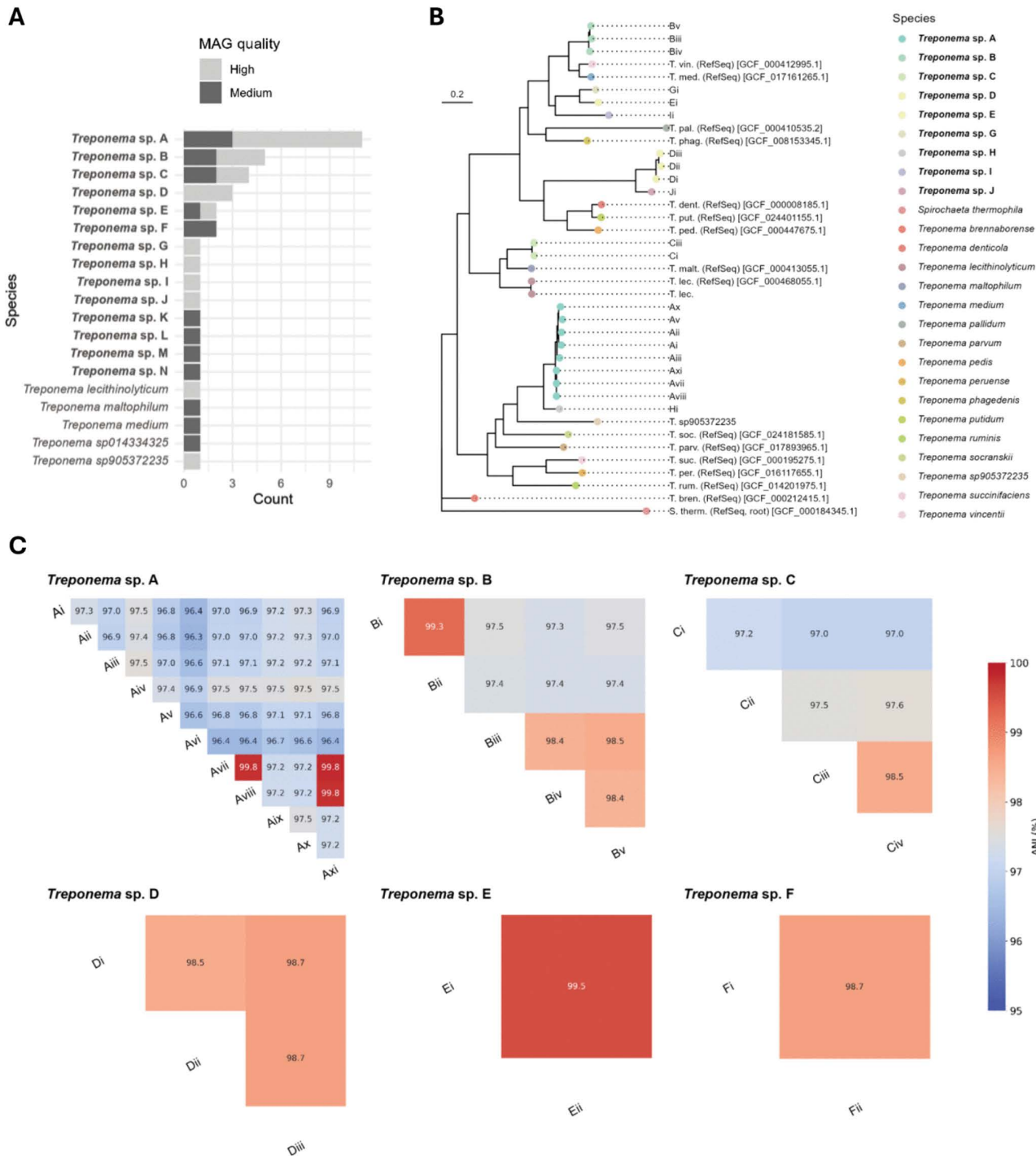


Fig 4. Summary and taxonomic placement of Treponema MAGs recovered from metagenomes. Novel species are in bold. a) frequency of different Treponema species MAGs; b) core-genome SNP maximum-likelihood tree placing high-quality Treponema MAGs within the context of oral Treponema species references; c) matrices showing ANI over aligning regions of MAGs within each Treponema species from which multiple MAGs were recovered.

<https://doi.org/10.1371/journal.pntd.0014118.g004>

Antimicrobial resistance determinants identified in noma participant oral microbiomes

We identified 39 unique AMR determinants in noma oral microbiome samples (Fig 5A). Determinants putatively conferring resistance to beta-lactams and tetracyclines were the most prevalent, namely CfxA3 (17/19 participants) and Tet(Q) (18/19 participants), respectively. This pattern was also observed in the global dataset of saliva metagenomes from healthy individuals (Fig 5A). However, *NimI* metronidazole resistance determinants, which were not detected in healthy saliva metagenomes, were detected in 8/19 noma participant samples. Additionally, aminoglycoside resistance determinants were more commonly detected in noma samples, with *aph(3'')-Ib* being identified in 12/19 (63.2%) of samples, while each aminoglycoside resistance determinant identified in samples from the healthy saliva dataset was not present in more than two samples. In contrast, macrolide resistance determinants *mef(A)* and *msr(D)* were detected in all saliva metagenomes from healthy individuals, but only in samples from 3/19 and 1/19 participants with noma, respectively.

To identify the specific organisms harbouring antimicrobial resistance genes (ARGs) in noma samples, all medium- and high-quality MAGs were screened against the Resfinder database. Of the ARGs that were binned to MAGs, three of the four *cfxA3* genes and all four *nimI* genes were identified in putative *Prevotella* genomes (Fig 5B). These ARGs can confer resistance to two of the treatments routinely given to noma participants at our study hospital (co-amoxiclav and metronidazole). ARGs known to confer resistance to the gentamicin (aminoglycoside), the third antibiotic routinely given, could not be binned to medium- or high-quality MAGs. No known resistance determinants were identified in *Treponema* MAGs.

Discussion

Noma, a devastating disease primarily affecting children in abject poverty, has long been understood to result from an opportunistic bacterial infection, evidenced by its prompt response to antibiotics [36–38]. However, the precise microbiological agents responsible for noma have remained unclear, hampering prevention, diagnosis, and treatment of the disease.

To interpret the shifts observed here, it is useful to note that the oral cavity is a complex ecological environment with multiple specific niches [39]. Generally, a healthy oral microbiome is dominated by typical abundant species, which are thought to be largely stable in humans globally [40–42] (as reflected the global healthy dataset, Fig 1) with complex immunological communication between the flora and the host. Factors such as the shared environment [43], diet, hygiene, and immunological status and infection have all been demonstrated to substantially affect the oral microbiome and in some cases result in periodontal disease [44]. Chronic malnourishment, poor oral hygiene, high rates of co-morbidities, and, in the case of children, immature immune systems, are all common factors applicable to noma participants [16,18]. However, not all children with these factors will develop noma, suggesting additional factors, such as presence or absence of specific microorganisms, amongst other things, are key to disease development.

Previous metagenomic 16S rRNA gene surveys noted a substantial loss of bacterial diversity in oral microbiomes of noma participants [22]. Similarly, our results reveal a drastic shift in microbial populations within the oral cavity of noma participants compared to that of global healthy saliva microbiomes (Fig 1). In particular, differential analysis showed that noma metagenomes were significantly enriched with *Treponema*, *Porphyromonas*, *Bacteroides*, and *Selenomonas*, whereas the global healthy saliva dataset comprised primarily of *Streptococcus*, *Veillonella*, *Gemella*, *Schaalia*, *Rothia*, *Actinomyces*, and *Haemophilus* species (Fig 2). The shift from a predominance of acid tolerant Gram-positive facultative anaerobes in healthy controls to mainly Gram-negative proteolytic obligate anaerobes, commonly associated with inflammation, mucosal invasion, and opportunistic infection, in the oral microbiomes of noma patients is notable. This microbial transition may suggest that oral salivary pH may be mildly alkaline in individuals with noma, favouring the growth of Gram-negative proteolytic anaerobic pathogens at the expense of Gram-positive facultative anaerobes [45]. Examining the oral pH of children with noma warrants inclusion in future research.

The presence and abundance of individual genera in isolation is not adequate to deconvolute complex microbial interactions in a polymicrobial disease [46]. To identify key microbial signatures in the oral microbiomes of noma participants

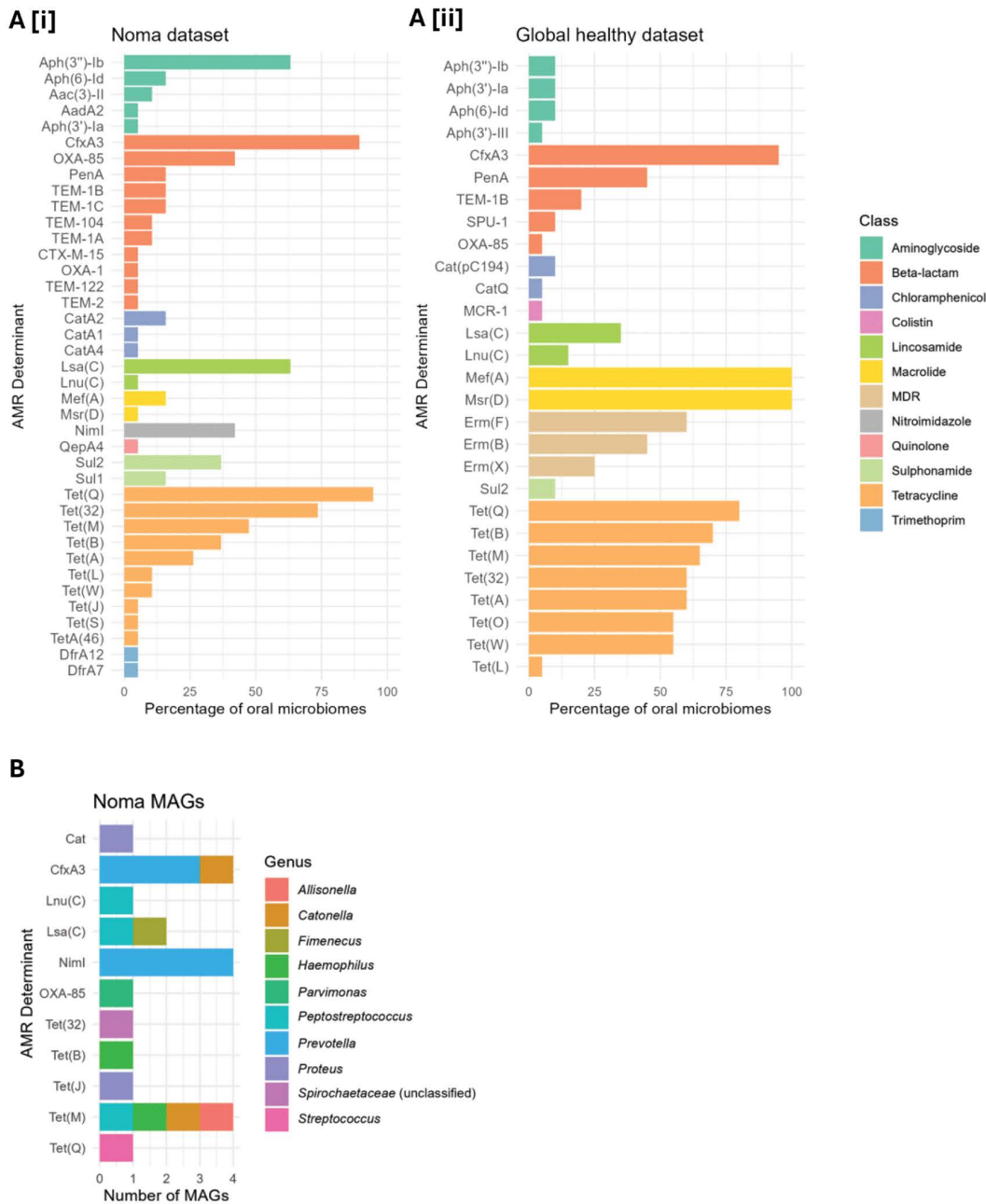


Fig 5. Description of detected AMR determinants in noma metagenomes. AMR determinants identified in: a) metagenome samples from [i] noma participants in this study, and [ii] saliva samples from the healthy global dataset; b) medium- and high-quality MAGs recovered from noma oral microbiome samples. In part (a), AMR determinants are grouped by class, with the most abundant determinants appearing at the top of each class. In part (b), colors indicate the genus of the MAG in which the AMR determinant was identified.

<https://doi.org/10.1371/journal.pntd.0014118.g005>

vs. microbiomes of healthy individuals which may be indicative of noma, we employed machine learning and multivariate statistical analyses, techniques which analyse the presence, abundance, and interactions of various microbes to determine patterns which correlate with specific disease states [47,48]. The key genera identified as signatures of noma

disease by these analyses were *Treponema*, *Porphyromonas* and *Bacteroides* (Fig 3). Previously *Porphyromonas*, a genus which contains well known oral pathogens such as *P. gingivalis*, and *P. endodontalis*, has been associated with noma, but more recent doubt was cast on their role in noma due to higher prevalence being observed in healthy vs. diseased participants in a 16S rRNA study in Nigeria [20]. *Bacteroides* is a genus which to date has not been associated with noma. Studies from the 1960s and 1980s which suggested *Bacteroides* as a potential contributor to noma were referencing *Prevotella* species prior to their reclassification as a distinct genus [38,49]. More recent noma metagenomic 16S rRNA analyses did not find members of the genus *Bacteroides* at greater quantities than healthy controls [20]. *Treponema* are normal inhabitants of the oral cavity, however, specific *Treponema* species are recognised as key oral pathogens [50,51]. Members of the genus are also the causative agents of many diseases, often affecting children in impoverished tropical populations and often causing necrotic lesions, for example, syphilis, pinta, yaws, and bejel, the latter two causing lesions in oral cavities [52,53]. Furthermore several *Treponema* species are associated with other multi-etiologic veterinary diseases with comparable disease progression in defined stages [54], including ovine contagious digital dermatitis [55] and bovine digital dermatitis [56].

Despite links to oral pathology, polymicrobial disease and conditions that disproportionately affect marginalised populations [57], the role of *Treponema* and other spirochaetes in noma has often been overlooked, even though they are frequently reported [14,16,20,21,26]. The exception to this is a study by Whiteson et al, [20] which performed a multivariate analyses on 16S rRNA metagenomes extracted from subgingival fluids collected from distinct oral healthy and lesion sites from noma participants in Niger. Their results demonstrated that *Treponema*, along with *Prevotella*, *Peptostreptococcus*, and *Sharpea*, were indicators of noma. Further studies on the same cohort of children provided mixed results regarding the association of spirochetes with noma. There are several reasons as to why *Treponema* may have been overlooked or proved inconsistent in previous microbiological and metagenomic surveys. *Treponema* are notoriously fastidious, and can often be only cultured on highly complex media in anaerobic conditions [38,58]. Furthermore, limitations of ‘universal’ 16S rRNA primers have been demonstrated to underestimate *Treponema* abundance and diversity in samples [22,59,60], thus their presence being likely substantially underestimated in previous surveys of noma.

Notably, we identified an uncultured, novel *Treponema* species (*Treponema* A) with high confidence from the majority (15/19) of noma patients in this study (Fig 4). Examination of the phylogeny of this novel species reveals it is more closely related to *Treponema socranskii*, a species frequently associated with severe periodontal tissue destruction [61], than it is to other known oral treponemes. Examination of previous 16S rRNA noma and associated ANG metagenomes also revealed the presence of *Treponema* sp. A in the majority of lesions sampled from participants with noma and ANG, while it was detected at very low frequency in the oral cavity of age matched control metagenomes in the same study [20]. The detection of this specific, novel treponeme in the oral metagenomes of noma participants from two independent studies, using different metagenomic techniques, and its apparent very low frequency in healthy oral microbiomes, suggests that *Treponema* A may be associated with the development of noma. Future, more comprehensive microbiome and genomic investigations will be important to confirm the robustness of this association and clarify its potential role in disease pathogenesis.

Whilst noting the overall shifts in the oral microbiomes of noma participants compared to healthy individuals, many previous studies on noma have limited the role of commensal bacteria to those of potential opportunistic pathogens, overlooking the potential protective role commensal bacteria may contribute to preventing noma. In many microbiotas, the maintenance of a commensal, normal microbiota is essential to the health function of the associated organ/tissue, as the commensals prevent pathogen dominance [62,63]. Notably, our analyses, in addition to the identification of the presence of *Treponema*, *Bacteroides* and *Porphyromonas*, also identified the lack of *Streptococcus* and *Rothia* species as a key indicator of noma disease (Fig 3).

Both *Streptococcus* and *Rothia* species are highly associated with healthy oral microbiomes in both adults and children [64,65]. *Streptococcus* spp. are one of the first colonisers of the oral cavity and are key in periodontal development and

maintaining oral health [65]. Similarly, *Rothia* species usually colonise the mouth after the age of one and are persistent colonisers for life [64]. Depletion of these species from oral microbiomes can be due to various factors, including poor oral hygiene, antibiotic use and compromised immune status, either through infection or chronic or acute malnourishment [44], all of which have been associated as risk factors for noma development [16,18]. The association of the depletion of these species as a key indicator of noma suggests that attempts to recolonise the oral cavity with *Streptococcus* and *Rothia*, or targeted prevention of their depletion altogether, may be a potential low cost, protective approach to reduce the incidence of noma. Multiple *Streptococcus* species, including *Streptococcus salivarius* and *Streptococcus mutans* are common food supplements, possess 'generally recognized as safe status', and are marketed as oral probiotics. Multiple studies have demonstrated health benefits of oral probiotics as adjunct therapies for treatment of dental caries and periodontitis [66,67], although overall robust clinical evidence of health benefits in the oral cavity remains limited [68,69]. Despite this, further investigation of oral probiotics as an adjunct strategy for noma prevention in at-risk children may be warranted to explore their potential protective role.

It is notable that *Prevotella*, a genus that has repeatedly linked to noma [16,20,21,25,38], whilst overall more abundant in noma metagenomes in this study (Fig 1), was not found to be significantly different compared to global healthy metagenomes (Figs 2 and S2), nor was *Prevotella* identified as a significant contributor to noma by machine learning or multivariate statistical analyses (Fig 3), suggesting that *Prevotella* may simply be more abundant due to poor oral hygiene or as a result of depletion of other microorganisms, but may not contribute directly to noma disease. Future, more comprehensive microbiome sequencing will be important to clarify whether this observed lack of significance of *Prevotella* relative to previous reports on noma microbiology reflects true biological differences.

The greater number of different ARGs identified in noma samples compared to the healthy global dataset (Fig 5) is perhaps a reflection of the altered taxonomic profiles of noma-associated oral microbiomes, although other local epidemiological factors not accounted for in our study may also contribute. Both beta-lactam and metronidazole resistance determinants, which were common in noma samples, were binned to putative *Prevotella* MAGs. Despite being abundant across noma participant microbiomes, *Prevotella* was not identified in our analysis as being a noma-associated genus (Fig 2). Therefore, the importance of identifying AMR determinants in *Prevotella* MAGs is debatable, as killing these organisms may not be clinically beneficial. While no AMR determinants could be linked to *Treponema* MAGs, the various novel *Treponema* species described here have not previously been cultured. Culture and antimicrobial susceptibility testing are vital next steps that may enable narrowed-spectrum therapy, including the possibility of oral-only antibiotics regimens for the treatment of noma if the novel noma-associated organism (*Treponema* sp. A) is proven to be pathogenic, and if phenotypic susceptibility data match genomic predictions.

The primary limitation of this preliminary study is focused on the use of a global healthy saliva dataset instead of generating healthy oral saliva microbiomes from a similar population to that of the noma participants. Since we found significant differences between countries (USA, Denmark and Japan) within the global healthy saliva dataset, the geography of the samples may be a confounding factor when comparing healthy and diseased samples. However, differential analysis showed the magnitude of difference between the noma and healthy datasets was substantially greater than between healthy datasets in several genera including *Treponema*, *Rothia*, and *Streptococcus* (S2 Fig). Despite these results and general acceptance that healthy oral microbiomes are stable globally [40], it is possible that our global dataset, originating from individuals in high-income countries in North America, Europe and Asia, may not be reflective of typical healthy African population oral microbiomes [70]. Ideally, future studies will include healthy controls from the same geographic location as the noma samples. Furthermore, since we only have 17 saliva samples from noma participants, the study is powered to detect only large effect sizes but may miss smaller but still biologically relevant differences. The metagenomic classification at the genus level, while necessary for specificity and reliability [71], may also miss species or strain-level variations. Despite these limitations, this data provides the first microbiological profiling of noma using deep shotgun metagenomics and highlights key shifts in microbial populations within the oral cavity of noma participants.

The lack of progress in elucidating noma microbiology, despite previous attempts to do so, has led some to question the value of understanding the etiology of the disease. Indeed, the disease was virtually eradicated from Europe and North America long before the discovery of antibiotics [13]. However, lack of detailed understanding of noma microbiology, in addition to the lack of understanding of many facets of noma pathogenesis [3], hinders targeted prevention efforts, hampers therapeutic choice, increases the risk of development of AMR, and limits the potential development of diagnostics, all of which are essential tools and considerations for eradicating noma more rapidly [72]. We are hopeful that the analysis presented here will provide a foundation for targeted microbiological studies of noma the future.

Methods

Ethics statement

The study protocol was reviewed and approved by the following ethical review boards: i) the Usman Danfodiyo University Teaching Hospital Health Research and Ethics Committee (UDUTH/HREC/2022/1152/V2), ii) the Sokoto State Ministry of Health Research Ethics Committee (SKHREC/042/2022), iii) the National Health Research and Ethics Committee in Nigeria (NHREC/01/01/2007-02/12/2022), iv) The Liverpool School of Tropical Medicine Ethical Review Board (22–045) and iv) the MSF Ethical Review Board (2326), which acted as primary sponsor.

Participant recruitment

We conducted a pilot cross-sectional study at the Noma Children's Hospital, Sokoto, Nigeria from 16th October 2023–23rd February 2024, utilizing a questionnaire and collection of saliva and swab samples.

We enrolled pediatric participants admitted to the Noma Children's Hospital with noma through a non-probability convenience sampling. Inclusion criteria consisted of participants who arrived at the Noma Children's Hospital and were: i) diagnosed as having Stage 0 (at risk) to stage 4 noma; ii) aged below 12 years; and iii) whose caregiver provided consent and participant provided assent (where applicable). Exclusion criteria consisted of participants who arrived at the Noma Children's Hospital and were i) diagnosed as having Stage 5 noma or any other condition; ii) aged 12 years and above; and iii) whose caregiver did not provide consent or participant did not provide assent (where applicable). In addition to the pre-specified exclusion criteria in the protocol, the following were applied during recruitment at the hospital and not explicitly included in the original written protocol due to an oversight: iv) a recent history of consuming any antibiotics in the seven days prior to admission and v) having receiving any treatment (including wound debridement and dressing) in another hospital prior to admission to Noma Children's Hospital. These additional criteria were discussed and agreed upon by the research team prior to implementation to ensure that biological samples were obtained before any treatment interventions.

Formal written consent was obtained from the guardian of each child. Participants aged seven to 12 years provided written assent. Information sheets and the informed consent and assent forms were written in English and Hausa.

Clinical information and questionnaire

Information routinely collected at the hospital included participant socio-demographic details, noma stage on admission, Mid Upper Arm Circumference (MUAC), weight, height, and vaccination history (mostly self-reported and subsequently confirmed or by vaccination card). Study specific data collected included caregiver demographics and recent treatment history.

Sample collection and storage

Saliva Samples were collected using DNA/RNA Shield SafeCollect Saliva Collection Kit (Zymo Research, USA) and swabs were collected using a DNA/ RNA Shield Collection Tube with Swab (Zymo Research, USA), following

manufacturer's instructions. Buccal swab samples were collected by rubbing the swab against the inner diseased cheek six times. For saliva samples, participants were asked to dribble saliva into the sample tube until the required volume (1ml) had been obtained.

DNA extraction and sequencing

Total DNA was extracted using the EchoLUTION Tissue DNA Micro Kit (BioEcho, Germany) as per manufacturer instructions. The metagenomic DNA sample was fragmented into short fragments. These DNA fragments were then end-polished, A-tailed, and ligated with full-length adapters for Illumina sequencing before further size selection. PCR amplification was then conducted unless specified as PCRfree. Purification was then conducted through the AMPure XP system (Beverly). The resulting library was assessed on the Agilent Fragment Analyzer System (Agilent) and quantified to 1.5nM through Qubit (Thermo Fisher Scientific) and qPCR. Whole metagenomic sequencing (paired-end 150 bp Illumina reads) was carried out by Novogene. The qualified libraries were pooled and sequenced on Illumina platforms, according to the effective library concentration and data amount required.

Sequence processing

Raw sequence files were quality assessed using FASTQC (<https://github.com/s-andrews/FastQC>). Trimming with Trimmomatic [73] version 0.39 was performed to remove adaptor sequences and low-quality bases with a sliding window quality cutoff of Q20 and a minimum read length of 50 bp. Human (host) DNA was filtered out by aligning reads to the human genome sequence CRCh38.p13 using Bowtie2 (<https://github.com/BenLangmead/bowtie2>) Version 2.2.5 with the “-un-conc” option.

Global healthy saliva sequence collection

Saliva sequences from healthy oral microbiomes were found by searching for metagenomic saliva samples that had been sequenced with Illumina technologies. We identified saliva metagenomes from 20 healthy individuals from three countries across three continents (USA, Japan and Denmark). The samples from Japan and Denmark (SRR and DRS) were downloaded from the ftp site on ENA using wget in December 2024. The samples from the USA (SRS) were part of the Human Microbiome Project [33] and downloaded directly from <https://www.hmpdacc.org/HMASM/> (NCBI BioProject PRJNA43017). The reads were quality-checked with FASTQC.

Taxonomic classification and relative abundance

Read based taxonomic assignment was run against the filtered reads using Kraken2 v2.1.3 [74,75] and Bracken v2.9 [76]. The bracken reports were converted to a metaphlan format and combined using KrakenTools scripts (<https://github.com/jenniferlu717/KrakenTools>) [77]. This combined report was imported into R v4.3.1 as a tree summarized experiment object with the mia v1.10.0 package [78]. An additional assay was constructed for relative abundance and an alternative experiment (altExp) was constructed at the genus level using the SingleCellExperiment v1.24.0 package. The top 20 most abundant genera at each taxonomic level were selected from across the entire dataset, with all others designated “-other”, and visualised with the plotAbundance function of miaViz v1.10.0.

Differential analysis of genera

Differential analysis used the DESeq2 model [34] as part of the DESeq2 v1.42.1 package. The statistical model was defined by the design formula “~ condition”, which tested for the effect of the disease state (‘diseased’ vs. ‘healthy’) on the normalized counts of each genus individually. The adjusted p-values and log2 fold change (LFC) for all genera were plotted together in a volcano plot. A genus was considered significantly differentially abundant if the False Discovery Rate

(FDR) adjusted p-value (padj) was less than 0.05. The most abundant genera (top 1%) were determined across the entire dataset and those which were significantly related to the healthy saliva dataset and noma samples had their normalized counts plotted as boxplots. Subject ID was not included as a variable in the model as the analysis was based on a cross-sectional design comparing two independent groups of subjects.

Multivariate statistical analyses and machine learning approaches

The tree summarized experiment object was converted into a Phyloseq object with the mia package (v1.10.0) for the random forests and ordination analyses. Random forest classifiers were applied using the randomForest package (v4.7-1.2) in R (v4.3.1) based on 500 estimator trees to identify key predictors of microbial community composition and their association with noma or healthy states. This machine learning technique was utilised to rank variable importance, expressed as percent increase in mean squared error (%IncMSE) derived from the random forest model, those variables with the highest %IncMSE contributed the most to the metadata classification. Each genus was treated as a predictor for the metadata classification of disease state (noma vs healthy). A cross validation of the dataset taking 20% of samples as test data and 80% as training data, was used to assess the predictive power of the classifier. To prevent data leakage, the test subset was isolated from the training phase and was utilised exclusively for final model evaluation. Predictive power was assessed as an area under the receiver operating characteristic (ROC) curve (AUC) between 0 and 1 using the pROC package (v1.18.5). A permutation test ($n=1000$) was conducted to evaluate the statistical significance of the observed classification accuracy. This allowed for a null distribution of accuracy to be plotted. The observed accuracy of the random forest model was overlaid on the null distribution.

Ordination techniques including Principal Coordinate Analysis (PCoA) were performed on a Bray-Curtis dissimilarity matrix of the data with the ordinate and distance functions from the phyloseq package (v1.46.0). The first two axes (i.e., PCoA1 and PCoA2) were visualized using the ggplot2 package (v 3.5.1), where samples were represented as points colored by their disease state.

Using the mia package (v1.10.0), a centered log-ratio (CLR) transformation was run on the relative abundance of genera, with a pseudocount to account for compositional effects. A z-score standardization was then applied across genera to normalize variability. The top 20 most abundant genera were selected and a heatmap was generated with the pheatmap package (v 1.0.12) annotated with disease state. Mean CLR z-scores were calculated for each genus in healthy and diseased conditions by aggregating the data by disease state and summarizing with the dplyr package (v1.1.4).

The vegan package (v2.6-8) was used to run PERMANOVA across the entire dataset for age, sex, sample type and disease status, and to run ANOSIM for respondent ID. For PERMANOVA the proportion of variance explained by each variable was reported as the R^2 effect size, for ANOSIM the R statistic was used. All tests used a Bray-Curtis dissimilarity matrix calculated with the phyloseq package (v1.46.0). For the healthy dataset differences in country of origin, individual genera were assessed using the Kruskal-Wallis test. The Wilcoxon rank sum test was used to compare the relative abundance of certain genera within the noma dataset (swab vs saliva) and between the noma and global healthy saliva dataset samples (noma vs healthy). The p-values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg method to control the false discovery rate FDR. An FDR-adjusted p-value, or q-value, of less than 0.05 was considered statistically significant.

Metagenome assembly and MAG recovery

Reads were assembled using metaSPAdes [79] version 3.11.1 at default settings. QUAST [80] version: 5.0.2 was used to check the quality of metagenome assemblies, with assembly statistics generated using contiguous sequences (contigs) of ≥ 500 bp in length. Contigs were indexed using the Burrows Wheeler Aligner (BWA) Version 0.7.17 before the raw reads were mapped against the indexed contigs using “bwa-mem” (<https://github.com/lh3/bwa>) [81] The contigs and resulting sorted BAM files were parsed to the “jgi_summarize_bam_contig_depth” script from MetaBAT2 [82] version 2.17,

before the resulting depth files were used by MetaBAT2 to bin assembled contigs of $\geq 2,500$ bp in length to metagenome assembled genomes (MAGs). CheckM [83] version 1.1.2 was used to assess the completeness and contamination of each MAG. CheckM outputs were compared to standards set by the genome standards consortium to classify MAGs as high-quality ($>90\%$ completeness, $<5\%$ contamination), medium-quality MAGs ($>50\%$ completeness, $<10\%$ contamination) or low-quality ($<50\%$ complete and $<10\%$ contamination, or $>10\%$ contamination) [84]. MAGs classified as medium- or high-quality progressed to the next analysis steps.

MAG taxonomic profiling and placement of *Treponema* MAGs

MAGs underwent taxonomic classification using GTDB-Tk using with the "classify" workflow version 2.1.1 against the Genome Taxonomy Database Version 09-RS220 (24th April 2024) [85]. The JSpeciesWS ANIm tool was used for pairwise comparison of *Treponema* MAGs which could not be assigned to a known species by GTDB-Tk, to identify novel species groups [86,87]. In the case that the same MAG was recovered from swab and saliva samples from one participant, the MAG with the greatest CheckM completeness proceeded to next steps of analysis.

Treponema MAGs in phylogenetic context

Published RefSeq genomes for known *Treponema* species were downloaded, including *Treponema brennaborensis* (GCF_000212415.1), *Treponema denticola* (GCF_000008185.1), *Treponema lecithinolyticum* (GCF_000468055.1), *Treponema maltophilum* (GCF_000413055.1), *Treponema medium* (GCF_017161265.1), *Treponema pallidum* subsp. *pallidum* (GCF_000410535.2), *Treponema parvum* (GCF_017893965.1), *Treponema pedis* (GCF_000447675.1), *Treponema peruense* (GCF_016117655.1), *Treponema phagedenis* (GCF_008153345.1), *Treponema putidum* (GCF_024401155.1), *Treponema ruminis* (GCF_014201975.1), *Treponema socranskii* subsp. *buccale* (GCF_024181585.1), *Treponema succinifaciens* (GCF_000195275.1), and *Treponema vincentii* (GCF_000412995.1). Additionally, the RefSeq *Spirochaeta thermophila* (GCF_000184345.1) genome was downloaded to act as an outgroup reference genome from a different genus within the Spirochaetaceae family. Annotations of our recovered *Treponema* spp. MAGs, as well as RefSeq genomes, were performed with Bakta Version 1.9.3 (full database Version 5.1) [88]. Panaroo [89] version 1.3.2 was used to retrieve soft core gene alignments to infer phylogeny, with clean mode set to moderate, a 70% sequence identity threshold for orthologue detection and a 90% threshold to ensure that genes included were present in at least 90% of genomes. Variant sites containing single nucleotide polymorphisms (SNPs) were then extracted from the alignments using SNP-Sites (<https://github.com/sanger-pathogens/snp-sites>) version 2.5.1. A maximum-likelihood tree was built using 1,000 bootstraps in IQ-TREE [90] version 1.6.1, with the most appropriate model (GTR+F+ASC+R3) determined by ModelFinder [91] within IQ-TREE. The tree was rooted by the *Spirochaeta thermophila* outgroup. The resulting treefile was visualized in the *ggtree* Version 3.12.0 package in R.

16S rRNA gene identification and reconstruction

To enable comparison to 16S microbiome datasets, Barrnap (<https://github.com/tseemann/barrnap>) Version 0.9 was used to quickly identify complete and partial ribosomal RNA genes in MAGs, where present. 16S rRNA genes of interest were queried against the GenBank nucleotide database using the Nucleotide Basic Local Alignment Search Tool (BLAST) webtool (megablast algorithm) to identify similar sequences of interest. However, 16S rRNA genes are often missing from MAGs due to difficulties both with resolving conserved regions in initial assembly algorithms and with handling resulting sequences in binning algorithms [92]. Therefore, after linking any given novel species MAG to a 16S rRNA gene sequence (in the minority situation where a 16S rRNA gene could be binned to the MAG), we were able to use the 16S sequence as a reference for the novel species. MATAM [93] was used alongside the SILVA SSU database Version 138.2 (<https://www.arb-silva.de/documentation/release-1382/>) to re-construct 16S rRNA genes in metagenome samples. We were then able

to confirm the presence of a novel species within samples from which a medium- to high-quality MAG from that species could not be recovered by querying MATAM assemblies against the MAG-linked 16S reference.

Identification of antimicrobial resistance determinants

Metagenome assemblies from both noma microbiome samples and the global healthy saliva dataset, and MAGs, were screened against the ResFinder database using ABRicate (<https://github.com/tseemann/abricate>) Version 1.0.1, to identify putative antimicrobial resistance genes (ARGs). Thresholds for minimum nucleotide identity and minimum query cover were set at 90%.

Supporting information

S1 Table. Symptoms, clinical assessments, and treatment.

(DOCX)

S2 Table. Participant demographic information.

(DOCX)

S3 Table. Health facility visit and treatment.

(DOCX)

S4 Table. Sequencing and assembly statistics.

(DOCX)

S5 Table. Identification of *Treponema sp. A* in noma participant oral microbiome samples in MAGs and 16S rRNA genes.

(DOCX)

S6 Table. Healthy control metadata descriptions.

(DOCX)

S1 Fig. Counts of *Treponema* across noma stages.

(DOCX)

S2 Fig. Counts of twelve genera across the noma dataset and the three separate healthy cohorts.

(DOCX)

S3 Fig. Statistical differences in relative abundance of genera in healthy control and noma samples of top 20 genera.

(DOCX)

S4 Fig. Top 1% of genera across entire dataset.

(DOCX)

S1 Data. Accessions for MAGs defined in this project.

(CSV)

Acknowledgments

We gratefully acknowledge the children and their families who participated in this study. We would like to thank Simon Wagstaff and Andrew Bennett at the LSTM Scientific Computing Unit for providing access to a high-performance computing server for data storage and analysis.

Author contributions

Conceptualization: Michael Olaleye, Miriam Peters, Shafi'u Isah, Adeniyi Semiyu Adetunji, Elise Farley, Mark Sherlock, Adam P. Roberts, Mohana Amirtharajah, Stuart Ainsworth.

Data curation: Michael Olaleye, Angus M. O'Ferrall, Richard N. Goodman.

Formal analysis: Michael Olaleye, Angus M. O'Ferrall, Richard N. Goodman, Gregoire Falq, Nichoas J Evans, Stuart Ainsworth.

Investigation: Angus M. O'Ferrall, Richard N. Goodman.

Methodology: Angus M. O'Ferrall, Richard N. Goodman.

Project administration: Michael Olaleye, Shafi'u Isah, Adeniyi Semiyu Adetunji, Elise Farley, Mark Sherlock, Adam P. Roberts, Mohana Amirtharajah, Stuart Ainsworth.

Resources: Joseph Samuel, Donal Doyle, Diana Gomez, Gbemisola Oloruntuyi, Adam P. Roberts, Mohana Amirtharajah.

Supervision: Adam P. Roberts, Mohana Amirtharajah, Stuart Ainsworth.

Visualization: Angus M. O'Ferrall, Richard N. Goodman.

Writing – original draft: Michael Olaleye, Angus M. O'Ferrall, Richard N. Goodman, Adam P. Roberts, Stuart Ainsworth.

Writing – review & editing: Michael Olaleye, Angus M. O'Ferrall, Richard N. Goodman, Deogracia Wa Kabila, Miriam Peters, Gregoire Falq, Joseph Samuel, Donal Doyle, Diana Gomez, Gbemisola Oloruntuyi, Shafi'u Isah, Adeniyi Semiyu Adetunji, Elise Farley, Nichoas J Evans, Mark Sherlock, Adam P. Roberts, Mohana Amirtharajah, Stuart Ainsworth.

References

1. Leila Srour M, Marck KW, Baratti-Mayer D. Noma: neglected, forgotten and a human rights issue. *Int Health*. 2015;7(3):149–50. <https://doi.org/10.1093/inthealth/ihv001> PMID: [25609756](https://pubmed.ncbi.nlm.nih.gov/25609756/)
2. Gebretsadik HG, de Kiev LC. A retrospective clinical, multi-center cross-sectional study to assess the severity and sequela of Noma/Cancrum oris in Ethiopia. *PLoS Negl Trop Dis*. 2022;16(9):e0010372. <https://doi.org/10.1371/journal.pntd.0010372> PMID: [36099293](https://pubmed.ncbi.nlm.nih.gov/36099293/)
3. Maguire BJ, Shrestha R, Dahal P, Ngu R, Nizigama L, Rashan S, et al. Systematic scoping review of the noma evidence landscape: current knowledge and gaps. *BMJ Glob Health*. 2025;10(7):e018023. <https://doi.org/10.1136/bmjgh-2024-018023> PMID: [40744672](https://pubmed.ncbi.nlm.nih.gov/40744672/)
4. Speiser S, Langridge B, Birkl MM, Kubiena H, Rodgers W. Update on Noma: systematic review on classification, outcomes and follow-up of patients undergoing reconstructive surgery after Noma disease. *BMJ Open*. 2021;11(8):e046303. <https://doi.org/10.1136/bmjopen-2020-046303> PMID: [34353795](https://pubmed.ncbi.nlm.nih.gov/34353795/)
5. Shaye DA, Winters R, Rabbels J, Adetunji AS, Magee A, Vo D. Noma surgery. *Laryngoscope*. 2019;129(1):96–9. <https://doi.org/10.1002/lary.27230> PMID: [30194753](https://pubmed.ncbi.nlm.nih.gov/30194753/)
6. Farley E, Ariti C, Amirtharajah M, Kamu C, Oluyide B, Shoaib M, et al. Noma, a neglected disease: A viewpoint article. *PLoS Negl Trop Dis*. 2021;15(6):e0009437. <https://doi.org/10.1371/journal.pntd.0009437> PMID: [34138861](https://pubmed.ncbi.nlm.nih.gov/34138861/)
7. Ainsworth S. Noma finally recognised as a neglected tropical disease. *PLoS Negl Trop Dis*. 2024;18(5):e0012177. <https://doi.org/10.1371/journal.pntd.0012177> PMID: [38814853](https://pubmed.ncbi.nlm.nih.gov/38814853/)
8. World Health Organization. Noma is a severe disease: it is treatable if detected and managed early. Brazzaville, Republic of Congo: WHO Regional Office for Africa. 2016.
9. Galli A, Brugger C, Fürst T, Monnier N, Winkler MS, Steinmann P. Prevalence, incidence, and reported global distribution of noma: a systematic literature review. *Lancet Infect Dis*. 2022;22(8):e221–30. [https://doi.org/10.1016/S1473-3099\(21\)00698-8](https://doi.org/10.1016/S1473-3099(21)00698-8) PMID: [35303478](https://pubmed.ncbi.nlm.nih.gov/35303478/)
10. Farley E, Oyemakinde MJ, Schuurmans J, Ariti C, Saleh F, Uzoigwe G, et al. The prevalence of noma in northwest Nigeria. *BMJ Glob Health*. 2020;5(4):e002141. <https://doi.org/10.1136/bmjgh-2019-002141> PMID: [32377404](https://pubmed.ncbi.nlm.nih.gov/32377404/)
11. Farley E, Mehta U, Srour ML, Lenglet A. Noma (cancrum oris): A scoping literature review of a neglected disease (1843 to 2021). *PLoS Negl Trop Dis*. 2021;15(12):e0009844. <https://doi.org/10.1371/journal.pntd.0009844> PMID: [34905547](https://pubmed.ncbi.nlm.nih.gov/34905547/)
12. Braimah RO, Adeoye J, Taiwo AO, Bello S, Bala M, Butali A, et al. Estimated incidence and clinical presentation of Noma in Northern Nigeria (1999–2024). *PLoS Negl Trop Dis*. 2025;19(5):e0012818. <https://doi.org/10.1371/journal.pntd.0012818> PMID: [40440286](https://pubmed.ncbi.nlm.nih.gov/40440286/)

13. Marck KW. A history of noma, the "Face of Poverty". *Plast Reconstr Surg.* 2003;111(5):1702–7. <https://doi.org/10.1097/01.PRS.0000055445.84307.3C> PMID: [12655218](https://pubmed.ncbi.nlm.nih.gov/12655218/)
14. Bowman FB. Ulcerative stomatitis and gingivitis (trench mouth). *Can Med Assoc J.* 1940;43:471–2.
15. Nolte SH. From trench mouth to noma: Experiences from Nazi extermination camps. *Clin Dermatol.* 2021;39(6):990–5. <https://doi.org/10.1016/j.clinidmatol.2021.05.019> PMID: [34920836](https://pubmed.ncbi.nlm.nih.gov/34920836/)
16. Baratti-Mayer D, Gayet-Ageron A, Hugonnet S, François P, Pittet-Cuenod B, Huyghe A, et al. Risk factors for noma disease: a 6-year, prospective, matched case-control study in Niger. *Lancet Glob Health.* 2013;1(2):e87–96. [https://doi.org/10.1016/S2214-109X\(13\)70015-9](https://doi.org/10.1016/S2214-109X(13)70015-9) PMID: [25104163](https://pubmed.ncbi.nlm.nih.gov/25104163/)
17. Adeniyi SA, Awosan KJ. Pattern of noma (cancrum oris) and its risk factors in Northwestern Nigeria: A hospital-based retrospective study. *Ann Afr Med.* 2019;18(1):17–22. https://doi.org/10.4103/aam.aam_5_18 PMID: [30729928](https://pubmed.ncbi.nlm.nih.gov/30729928/)
18. Farley E, Lenglet A, Ariti C, Jiya NM, Adetunji AS, van der Kam S, et al. Risk factors for diagnosed noma in northwest Nigeria: A case-control study, 2017. *PLoS Negl Trop Dis.* 2018;12(8):e0006631. <https://doi.org/10.1371/journal.pntd.0006631> PMID: [30138374](https://pubmed.ncbi.nlm.nih.gov/30138374/)
19. Uzochukwu I, Moyes D, Proctor G, Ide M. The key players of dysbiosis in Noma disease; A systematic review of etiological studies. *Front Oral Health.* 2023;4:1095858. <https://doi.org/10.3389/froh.2023.1095858> PMID: [36937503](https://pubmed.ncbi.nlm.nih.gov/36937503/)
20. Whiteson KL, Lazarevic V, Tangomo-Bento M, Girard M, Maughan H, Pittet D, et al. Noma affected children from Niger have distinct oral microbial communities based on high-throughput sequencing of 16S rRNA gene fragments. *PLoS Negl Trop Dis.* 2014;8(12):e3240. <https://doi.org/10.1371/journal.pntd.0003240> PMID: [25474262](https://pubmed.ncbi.nlm.nih.gov/25474262/)
21. Bolivar I, Whiteson K, Stadelmann B, Baratti-Mayer D, Gizard Y, Mombelli A, et al. Bacterial diversity in oral samples of children in niger with acute noma, acute necrotizing gingivitis, and healthy controls. *PLoS Negl Trop Dis.* 2012;6(3):e1556. <https://doi.org/10.1371/journal.pntd.0001556> PMID: [22413030](https://pubmed.ncbi.nlm.nih.gov/22413030/)
22. Huyghe A, François P, Mombelli A, Tangomo M, Girard M, Baratti-Mayer D, et al. Microarray analysis of microbiota of gingival lesions in noma patients. *PLoS Negl Trop Dis.* 2013;7(9):e2453. <https://doi.org/10.1371/journal.pntd.0002453> PMID: [24086784](https://pubmed.ncbi.nlm.nih.gov/24086784/)
23. Chidzonga MM, Mahomva L. Noma (cancrum oris) in human immunodeficiency virus infection and acquired immunodeficiency syndrome (HIV and AIDS): clinical experience in Zimbabwe. *J Oral Maxillofac Surg.* 2008;66(3):475–85. <https://doi.org/10.1016/j.joms.2007.09.024> PMID: [18280380](https://pubmed.ncbi.nlm.nih.gov/18280380/)
24. Falkler WA Jr, Enwonwu CO, Idigbe EO. Isolation of *Fusobacterium necrophorum* from cancrum oris (noma). *Am J Trop Med Hyg.* 1999;60(1):150–6. <https://doi.org/10.4269/ajtmh.1999.60.150> PMID: [9988340](https://pubmed.ncbi.nlm.nih.gov/9988340/)
25. Paster BJ, Falkler WA Jr, Enwonwu CO, Idigbe EO, Savage KO, Levanos VA, et al. Prevalent bacterial species and novel phylotypes in advanced noma lesions. *J Clin Microbiol.* 2002;40(6):2187–91. <https://doi.org/10.1128/JCM.40.6.2187-2191.2002> PMID: [12037085](https://pubmed.ncbi.nlm.nih.gov/12037085/)
26. Phillips RS, Enwonwu CO, Falkler WA. Pro- versus anti-inflammatory cytokine profile in African children with acute oro-facial noma (cancrum oris, noma). *Eur Cytokine Netw.* 2005;16(1):70–7. PMID: [15809209](https://pubmed.ncbi.nlm.nih.gov/15809209/)
27. Weinroth MD, Belk AD, Dean C, Noyes N, Dittoe DK, Rothrock MJ, et al. Considerations and best practices in animal science 16S ribosomal RNA gene sequencing microbiome studies. *J Anim Sci.* 2022;100(2):skab346. <https://doi.org/10.1093/jas/skab346> PMID: [35106579](https://pubmed.ncbi.nlm.nih.gov/35106579/)
28. Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One.* 2014;9(4):e93827. <https://doi.org/10.1371/journal.pone.0093827> PMID: [24714158](https://pubmed.ncbi.nlm.nih.gov/24714158/)
29. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet.* 2019;20(6):341–55. <https://doi.org/10.1038/s41576-019-0113-7> PMID: [30918369](https://pubmed.ncbi.nlm.nih.gov/30918369/)
30. Lee K, Raguideau S, Sirén K, Asnicar F, Cumbo F, Hildebrand F, et al. Population-level impacts of antibiotic usage on the human gut microbiome. *Nat Commun.* 2023;14(1):1191. <https://doi.org/10.1038/s41467-023-36633-7> PMID: [36864029](https://pubmed.ncbi.nlm.nih.gov/36864029/)
31. Yahara K, Suzuki M, Hirabayashi A, Suda W, Hattori M, Suzuki Y, et al. Long-read metagenomics using PromethION uncovers oral bacteriophages and their interaction with host bacteria. *Nat Commun.* 2021;12(1):27. <https://doi.org/10.1038/s41467-020-20199-9> PMID: [33397904](https://pubmed.ncbi.nlm.nih.gov/33397904/)
32. Belstrøm D, Constancias F, Liu Y, Yang L, Drautz-Moses DI, Schuster SC, et al. Metagenomic and metatranscriptomic analysis of saliva reveals disease-associated microbiota in patients with periodontitis and dental caries. *NPJ Biofilms Microbiomes.* 2017;3:23. <https://doi.org/10.1038/s41522-017-0031-4> PMID: [28979798](https://pubmed.ncbi.nlm.nih.gov/28979798/)
33. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature.* 2012;486(7402):207–14. <https://doi.org/10.1038/nature11234> PMID: [22699609](https://pubmed.ncbi.nlm.nih.gov/22699609/)
34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
35. Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun.* 2019;10(1):5029. <https://doi.org/10.1038/s41467-019-13036-1> PMID: [31695033](https://pubmed.ncbi.nlm.nih.gov/31695033/)
36. D'Agostino FJ. A review of noma; report of a case treated with aureomycin. *Oral Surg Oral Med Oral Pathol.* 1951;4(8):1000–6. [https://doi.org/10.1016/0030-4220\(51\)90446-x](https://doi.org/10.1016/0030-4220(51)90446-x) PMID: [14863869](https://pubmed.ncbi.nlm.nih.gov/14863869/)
37. Hickey BB. Cancrum oris. *Br Med J.* 1949;1:367.
38. Falkler WA Jr, Enwonwu CO, Idigbe EO. Microbiological understandings and mysteries of noma (cancrum oris). *Oral Dis.* 1999;5(2):150–5. <https://doi.org/10.1111/j.1601-0825.1999.tb00081.x> PMID: [10522213](https://pubmed.ncbi.nlm.nih.gov/10522213/)
39. Deo PN, Deshmukh R. Oral microbiome: Unveiling the fundamentals. *J Oral Maxillofac Pathol.* 2019;23(1):122–8. https://doi.org/10.4103/jomfp.JOMFP_304_18 PMID: [31110428](https://pubmed.ncbi.nlm.nih.gov/31110428/)

40. Baker JL, Mark Welch JL, Kauffman KM, McLean JS, He X. The oral microbiome: diversity, biogeography and human health. *Nat Rev Microbiol*. 2024;22(2):89–104. <https://doi.org/10.1038/s41579-023-00963-6> PMID: [37700024](https://pubmed.ncbi.nlm.nih.gov/37700024/)
41. Zaura E, Keijsers BJF, Huse SM, Crielaard W. Defining the healthy “core microbiome” of oral microbial communities. *BMC Microbiol*. 2009;9:259. <https://doi.org/10.1186/1471-2180-9-259> PMID: [20003481](https://pubmed.ncbi.nlm.nih.gov/20003481/)
42. Wang J, Feng J, Zhu Y, Li D, Wang J, Chi W. Diversity and Biogeography of Human Oral Saliva Microbial Communities Revealed by the Earth Microbiome Project. *Front Microbiol*. 2022;13:931065. <https://doi.org/10.3389/fmicb.2022.931065> PMID: [35770164](https://pubmed.ncbi.nlm.nih.gov/35770164/)
43. Shaw L, Ribeiro ALR, Levine AP, Pontikos N, Balloux F, Segal AW, et al. The Human Salivary Microbiome Is Shaped by Shared Environment Rather than Genetics: Evidence from a Large Family of Closely Related Individuals. *mBio*. 2017;8(5):e01237-17. <https://doi.org/10.1128/mBio.01237-17> PMID: [28900019](https://pubmed.ncbi.nlm.nih.gov/28900019/)
44. Tjandrawinata RR, Amalia N, Tandii YYP, Athallah AF, Afif Wibowo C, Aditya MR, et al. The forgotten link: how the oral microbiome shapes childhood growth and development. *Front Oral Health*. 2025;6:1547099. <https://doi.org/10.3389/froh.2025.1547099> PMID: [39989601](https://pubmed.ncbi.nlm.nih.gov/39989601/)
45. Pöllänen MT, Paino A, Ihalin R. Environmental stimuli shape biofilm formation and the virulence of periodontal pathogens. *Int J Mol Sci*. 2013;14(8):17221–37. <https://doi.org/10.3390/ijms140817221> PMID: [23965982](https://pubmed.ncbi.nlm.nih.gov/23965982/)
46. Hajishengallis G, Lamont RJ. Polymicrobial communities in periodontal disease: Their quasi-organismal nature and dialogue with the host. *Periodontol 2000*. 2021;86(1):210–30. <https://doi.org/10.1111/prd.12371> PMID: [33690950](https://pubmed.ncbi.nlm.nih.gov/33690950/)
47. Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovic V, Aasmets O, et al. Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. *Front Microbiol*. 2021;12:634511. <https://doi.org/10.3389/fmicb.2021.634511> PMID: [33737920](https://pubmed.ncbi.nlm.nih.gov/33737920/)
48. Asnicar F, Thomas AM, Passerini A, Waldron L, Segata N. Machine learning for microbiologists. *Nat Rev Microbiol*. 2024;22(4):191–205. <https://doi.org/10.1038/s41579-023-00984-1> PMID: [37968359](https://pubmed.ncbi.nlm.nih.gov/37968359/)
49. Adekeye EO, Ord RA. Cancrum oris: principles of management and reconstructive surgery. *J Maxillofac Surg*. 1983;11(4):160–70. [https://doi.org/10.1016/s0301-0503\(83\)80040-x](https://doi.org/10.1016/s0301-0503(83)80040-x) PMID: [6579147](https://pubmed.ncbi.nlm.nih.gov/6579147/)
50. You M, Mo S, Leung WK, Watt RM. Comparative analysis of oral treponemes associated with periodontal health and disease. *BMC Infect Dis*. 2013;13:174. <https://doi.org/10.1186/1471-2334-13-174> PMID: [23578286](https://pubmed.ncbi.nlm.nih.gov/23578286/)
51. Seshadri R, Myers GSA, Tettelin H, Eisen JA, Heidelberg JF, Dodson RJ, et al. Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc Natl Acad Sci U S A*. 2004;101(15):5646–51. <https://doi.org/10.1073/pnas.0307639101> PMID: [15064399](https://pubmed.ncbi.nlm.nih.gov/15064399/)
52. Mitjå O, Šmajš D, Bassat Q. Advances in the diagnosis of endemic treponematoses: yaws, bejel, and pinta. *PLoS Negl Trop Dis*. 2013;7(10):e2283. <https://doi.org/10.1371/journal.pntd.0002283> PMID: [24205410](https://pubmed.ncbi.nlm.nih.gov/24205410/)
53. Marks M, Solomon AW, Mabey DC. Endemic treponemal diseases. *Trans R Soc Trop Med Hyg*. 2014;108(10):601–7. <https://doi.org/10.1093/trstmh/tru128> PMID: [25157125](https://pubmed.ncbi.nlm.nih.gov/25157125/)
54. Holzhauser M, Bartels CJM, Döpfer D, van Schaik G. Clinical course of digital dermatitis lesions in an endemically infected herd without preventive herd strategies. *Vet J*. 2008;177(2):222–30. <https://doi.org/10.1016/j.tvjl.2007.05.004> PMID: [17618149](https://pubmed.ncbi.nlm.nih.gov/17618149/)
55. Angell JW, Crosby-Durrani HE, Duncan JS, Carter SD, Blundell R. Histopathological Characterization of the Lesions of Contagious Ovine Digital Dermatitis and Immunolabelling of *Treponema*-like Organisms. *J Comp Pathol*. 2015;153(4):212–26. <https://doi.org/10.1016/j.jcpa.2015.10.178> PMID: [26597022](https://pubmed.ncbi.nlm.nih.gov/26597022/)
56. Evans NJ, Murray RD, Carter SD. Bovine digital dermatitis: Current concepts from laboratory to farm. *Vet J*. 2016;211:3–13. <https://doi.org/10.1016/j.tvjl.2015.10.028> PMID: [27061657](https://pubmed.ncbi.nlm.nih.gov/27061657/)
57. Baker BJ, Crane-Kramer G, Dee MW, Gregoricka LA, Henneberg M, Lee C, et al. Advancing the understanding of treponemal disease in the past and present. *Am J Phys Anthropol*. 2020;171 Suppl 70:5–41. <https://doi.org/10.1002/ajpa.23988> PMID: [31956996](https://pubmed.ncbi.nlm.nih.gov/31956996/)
58. Enwonwu CO, Falkler WA, Idigbe EO. Oro-facial gangrene (noma/cancrum oris): pathogenetic mechanisms. *Crit Rev Oral Biol Med*. 2000;11(2):159–71. <https://doi.org/10.1177/10454411000110020201> PMID: [12002813](https://pubmed.ncbi.nlm.nih.gov/12002813/)
59. Bekele AZ, Koike S, Kobayashi Y. Phylogenetic diversity and dietary association of rumen *Treponema* revealed using group-specific 16S rRNA gene-based analysis. *FEMS Microbiol Lett*. 2011;316(1):51–60. <https://doi.org/10.1111/j.1574-6968.2010.02191.x> PMID: [21204927](https://pubmed.ncbi.nlm.nih.gov/21204927/)
60. Wang Y, Qian P-Y. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One*. 2009;4(10):e7401. <https://doi.org/10.1371/journal.pone.0007401> PMID: [19816594](https://pubmed.ncbi.nlm.nih.gov/19816594/)
61. Takeuchi Y, Umeda M, Sakamoto M, Benno Y, Huang Y, Ishikawa I. *Treponema socranskii*, *Treponema denticola*, and *Porphyromonas gingivalis* are associated with severity of periodontal tissue destruction. *J Periodontol*. 2001;72(10):1354–63. <https://doi.org/10.1902/jop.2001.72.10.1354> PMID: [11699477](https://pubmed.ncbi.nlm.nih.gov/11699477/)
62. Ivanov II, Littman DR. Modulation of immune homeostasis by commensal bacteria. *Curr Opin Microbiol*. 2011;14(1):106–14. <https://doi.org/10.1016/j.mib.2010.12.003> PMID: [21215684](https://pubmed.ncbi.nlm.nih.gov/21215684/)
63. Artis D. Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nat Rev Immunol*. 2008;8(6):411–20. <https://doi.org/10.1038/nri2316> PMID: [18469830](https://pubmed.ncbi.nlm.nih.gov/18469830/)

64. Dzidic M, Collado MC, Abrahamsson T, Artacho A, Stensson M, Jenmalm MC, et al. Oral microbiome development during childhood: an ecological succession influenced by postnatal factors and associated with tooth decay. *ISME J.* 2018;12(9):2292–306. <https://doi.org/10.1038/s41396-018-0204-z> PMID: 29899505
65. Baty JJ, Stoner SN, Scofield JA. Oral commensal streptococci: gatekeepers of the oral cavity. *J Bacteriol.* 2022;204:e00257-22.
66. Bustamante M, Oomah BD, Mosi-Roa Y, Rubilar M, Burgos-Díaz C. Probiotics as an Adjunct Therapy for the Treatment of Halitosis, Dental Caries and Periodontitis. *Probiotics Antimicrob Proteins.* 2020;12(2):325–34. <https://doi.org/10.1007/s12602-019-9521-4> PMID: 30729452
67. Tagg JR, Harold LK, Jain R, Hale JDF. Beneficial modulation of human health in the oral cavity and beyond using bacteriocin-like inhibitory substance-producing streptococcal probiotics. *Front Microbiol.* 2023;14:1161155. <https://doi.org/10.3389/fmicb.2023.1161155> PMID: 37056747
68. Inchingolo F. The benefits of probiotics on oral health: systematic review of the literature. *Pharmaceuticals.* 2023;16:1313.
69. Teughels W, Loozen G, Quirynen M. Do probiotics offer opportunities to manipulate the periodontal oral microbiota? *J Clin Periodontol.* 2011;38 Suppl 11:159–77. <https://doi.org/10.1111/j.1600-051X.2010.01665.x> PMID: 21323712
70. Willis JR, Gabaldón T. The Human Oral Microbiome in Health and Disease: From Sequences to Ecosystems. *Microorganisms.* 2020;8(2):308. <https://doi.org/10.3390/microorganisms8020308> PMID: 32102216
71. Galloway-Peña J, Hanson B. Tools for Analysis of the Microbiome. *Dig Dis Sci.* 2020;65(3):674–85. <https://doi.org/10.1007/s10620-020-06091-y> PMID: 32002757
72. Galli A, Comparet M, Dagne DA, Baratti-Mayer D, Cao TH, Guérin PJ, et al. Defining the noma research agenda. *PLoS Negl Trop Dis.* 2025;19(4):e0012940. <https://doi.org/10.1371/journal.pntd.0012940> PMID: 40257975
73. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
74. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014;15(3):R46. <https://doi.org/10.1186/gb-2014-15-3-r46> PMID: 24580807
75. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20(1):257. <https://doi.org/10.1186/s13059-019-1891-0> PMID: 31779668
76. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci.* 2017;3:e104. <https://doi.org/10.7717/peerj-cs.104> PMID: 40271438
77. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. *Nat Protoc.* 2022;17(12):2815–39. <https://doi.org/10.1038/s41596-022-00738-y> PMID: 36171387
78. Huang R, Soneson C, Ernst FGM, Rue-Albrecht KC, Yu G, Hicks SC, et al. TreeSummarizedExperiment: a S4 class for data with hierarchical structure. *F1000Res.* 2020;9:1246. <https://doi.org/10.12688/f1000research.26669.2> PMID: 33274053
79. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 2017;27(5):824–34. <https://doi.org/10.1101/gr.213959.116> PMID: 28298430
80. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29(8):1072–5. <https://doi.org/10.1093/bioinformatics/btt086> PMID: 23422339
81. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
82. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* 2019;7:e7359. <https://doi.org/10.7717/peerj.7359> PMID: 31388474
83. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25(7):1043–55. <https://doi.org/10.1101/gr.186072.114> PMID: 25977477
84. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol.* 2017;35(8):725–31. <https://doi.org/10.1038/nbt.3893> PMID: 28787424
85. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2019;36(6):1925–7. <https://doi.org/10.1093/bioinformatics/btz848>
86. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics.* 2016;32(6):929–31. <https://doi.org/10.1093/bioinformatics/btv681> PMID: 26576653
87. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A.* 2009;106(45):19126–31. <https://doi.org/10.1073/pnas.0906412106> PMID: 19855009
88. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom.* 2021;7(11):000685. <https://doi.org/10.1099/mgen.0.000685> PMID: 34739369
89. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 2020;21(1):180. <https://doi.org/10.1186/s13059-020-02090-4> PMID: 32698896
90. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430

91. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–9. <https://doi.org/10.1038/nmeth.4285> PMID: [28481363](https://pubmed.ncbi.nlm.nih.gov/28481363/)
92. Song W, Zhang S, Thomas T. MarkerMAG: linking metagenome-assembled genomes (MAGs) with 16S rRNA marker genes using paired-end short reads. *Bioinformatics*. 2022;38(15):3684–8. <https://doi.org/10.1093/bioinformatics/btac398> PMID: [35713513](https://pubmed.ncbi.nlm.nih.gov/35713513/)
93. Pericard P, Dufresne Y, Couderc L, Blanquart S, Touzet H. MATAM: reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes. *Bioinformatics*. 2018;34(4):585–91. <https://doi.org/10.1093/bioinformatics/btx644> PMID: [29040406](https://pubmed.ncbi.nlm.nih.gov/29040406/)