Targeted isolation and cultivation of uncultivated bacteria by reverse genomics

Karissa L. Cross1,2,7, James H. Campbell1,6,7, Manasi Balachandran1, Alisha G. Campbell1,6,3, Connor J. Cooper1,3, Ann Griffen4, Matthew Heaton5, Snehal Joshi1, Dawn Klingeman1, Eugene Leys4, Zamin Yang1, Jerry M. Parks1,3 and Mircea Podar6,1,2,3*

Most microorganisms from all taxonomic levels are uncultured. Single-cell genomes and metagenomes continue to increase the known diversity of Bacteria and Archaea; however, while ‘omics can be used to infer physiological or ecological roles for species in a community, most of these hypothetical roles remain unvalidated. Here, we report an approach to capture specific microorganisms from complex communities into pure cultures using genome-informed antibody engineering. We apply our reverse genomics approach to isolate and sequence single cells and to cultivate three different species-level lineages of human oral Saccharibacteria (TM7). Using our pure cultures, we show that all three Saccharibacteria species are epibions of diverse Actinobacteria. We also isolate and cultivate human oral SR1 bacteria, which are members of a lineage of previously uncultured bacteria. Reverse-genomics-enabled cultivation of microorganisms can be applied to any species from any environment and has the potential to unlock the isolation, cultivation and characterization of species from as-yet-uncultured branches of the microbial tree of life.

Our understanding of the microbial tree of life has undergone continuous reshaping and refinement since the advent of molecular, culture-independent methods to identify and classify microorganisms. For example, 16S ribosomal RNA gene sequencing resulted in the discovery of nearly two dozen candidate Bacteria and Archaea phyla1, and more recently single-cell amplified genomes (SAGs) and metagenome-assembled genomes (MAGs) have led to a deluge of proposed new lineages2–7. Notably, however, SAGs and MAGs of specific taxa are often retrieved by chance, depending on their abundance in the community. Furthermore, although microbial cultivation has been successful for a select few microbes (reviewed in ref. 8), these authors contributed equally: Karissa L. Cross, James H. Campbell. *e-mail: podar@ornl.gov

The causes of microbial ‘uncultivability’ are numerous9–11 and have been associated with a requirement for factors produced by other microbes12–13, strict interspecies interactions14–15, slow growth16, competition/inhibition and dormancy17. Various high-throughput approaches have been developed to alleviate cultivation recalcitrance. Encapsulation and incubation in microdroplets under media flow, for example, enable interspecies signaling and potential cross-feeding18–19, and integrating multiple media and cultivation conditions with rapid strain identification (culturomics)20–22 have both been successful in expanding the repertoire of cultured species and strains from open environments and human-associated microbiota. Restricting potential competition by stochastic dilutions of complex microbial samples to a few or single cells, isolated in micro-compartments, while exposing them to native environmental conditions through porous membranes (iChip and similar devices), has also resulted in pure cultures of microbes that were refractory to isolation using standard techniques23,24. One important limitation of these cultivation approaches is stochasticity, which means that success in isolating new organisms is driven partly by chance. Existing methods cannot target specific groups of as-yet-uncultured microbes, especially if those organisms are present at low abundance or have no readily selectable phenotypes. The availability of a plethora of sequence data from all the lineages of life presents an opportunity to design an approach to directly link a genotype to actual cells. A physical marker selected based on genomic data, and compatible with maintaining viability, would enable isolation of microbes, regardless of abundance, for potential cultivation or for selective genomic sequencing. Here, we introduce an approach named reverse genomics isolation that is directed at, and uses antibodies against, predicted cell surface proteins to target efforts to cultivate selected bacteria or archaea (Fig. 1a).

TM7 was one of the first candidate bacterial phyla proposed based on 16S rRNA gene sequences15,26 and has been found in many environments. After near-complete MAGs were obtained for several TM7 lineages from activated sludge, the group was renamed Saccharibacteria27. In humans, approximately a dozen distinct species-level Saccharibacteria are recognized28, mostly in the oral cavity, and increased abundance of Saccharibacteria is correlated with the development of periodontitis29 and inflammatory bowel disease30. Related lineages are also present in dogs, cats and dolphins31–33. The first genomic data for a human oral TM7 were generated by single-cell sequencing following micro-isolation from supragingival plaque34. Recently, the first representative (TM7x) was cultured as an obligate epibiont on a human oral Actinomyces odontolyticus, based on the serendipitous streptomycin resistance of both organisms35.

1Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA. 2Department of Microbiology, University of Tennessee, Knoxville, TN, USA. 3Genome Science and Technology Program, University of Tennessee, Knoxville, TN, USA. 4College of Dentistry, The Ohio State University, Columbus, OH, USA. 5Knoxville Periodontics, Knoxville, TN, USA. 6Present address: Department of Natural Sciences, Northwest Missouri State University, Maryville, MO, USA. 7These authors contributed equally: Karissa L. Cross, James H. Campbell. *e-mail: podar@ornl.gov
Considering the large taxonomic diversity and ubiquitous environmental presence of these organisms, generally at low abundance (<1%), we decided to develop and test a reverse genomics isolation method to isolate and cultivate Saccharibacteria.

Results
Selection of cell-surface antibody targets. We used the single-cell genomic data from a human oral saccharibacterium named TM7a (ref. 34) to identify genes coding for predicted membrane-associated proteins for which a general function could be inferred based on bacterial homologs. We avoided hypothetical proteins, because in the absence of gene expression information we considered them less likely to be present on the cell surface. However, if functional genomic data are available, any membrane protein would be a potential target. We screened for proteins with extracellular domains that could serve as antigens for antibody development, and that had homologs with solved three-dimensional (3D) structures, to facilitate searching for surface-exposed epitopes. One candidate protein was a predicted penicillin-binding protein (penicillin-binding protein 2 (PBP2)). In pathogenic bacteria, PBP2 can be an effective vaccine target, indicating that it is expressed and recognizable by an antibody36. Crystal structures for Staphylococcus aureus (Protein Data Bank (PDB) entry 3DWK) and Escherichia coli (3VMA) PBP2 homologs were used to model TM7a PBP2 and identify potentially exposed epitopes. PBP2 is anchored in the membrane by a single α-helix and consists of two extracellular domains, an N-terminal glycosyltransferase domain and a C-terminal transpeptidase domain connected by a β-rich linker region29 (Fig. 1b). A 19-amino acid sequence located in the glycosyltransferase domain was selected as an antigen based on surface exposure, predicted immunogenicity and solubility, and was used for raising polyclonal antibodies (see Methods).

Antibody-enabled sorting of oral Saccharibacteria (TM7). Human oral samples (saliva, subgingival fluid) from healthy donors or individuals with periodontitis were incubated with fluorescently labeled anti-PBP2 or anti-CspC IgGs. Stained samples were analyzed by flow cytometry, revealing a distinct population of fluorescent cells (0.1–5%, depending on the sample) (Fig. 2a,b). To determine whether Saccharibacteria were labeled by either IgG, we initially sorted pools of 10–100 fluorescent cells and used multiple displacement genomic amplification (MDA) followed by PCR with specific primers (see Methods) targeting the small subunit rRNA (SSU rRNA) gene. PCR products were analyzed by MiSeq amplicon sequencing of the V4 region of the SSU rRNA gene amplified with universal primers (Supplementary Table 1, SRA data). Between 25% and 100% of the cell pools were positive for TM7, depending on individual subjects/samples and gating parameters, with both anti-PBP2 and anti-CspC retrieving TM7 cells. To enhance the fluorescence signal and to cover as much as possible of the sequence variation space, all experiments directed at cultivation were subsequently performed by dual labeling using both anti-PBP2 and anti-CspC antibodies. When we analyzed such labeled cell pools for taxonomic representation using MiSeq amplicon sequencing, Saccharibacteria (TM7) comprised up to 95% of the sequences and predominantly represented groups 1, 3 and 6 phylotypes (Fig. 2c,d, Supplementary Fig. 2 and Supplementary Table 1, SRA data). Other bacteria were primarily members of Fusobacteria, Actinobacteria, Bacteroidetes, Haemophilus and Streptococcus. The identification of species other than the target TM7 may indicate cross-reactivity of the antibodies, free genomic DNA or interspecies cell–cell adhesion (specific or non-specific), the latter being extensive in the oral microbiota39,40.

Antibody-enabled cultivation of Saccharibacteria. Having demonstrated that anti-TM7 antibodies can label target cells in biological samples and enable sorting by flow cytometry, we next tested whether antibody-sorted cells were viable and could be

**Fig. 1 | Overview of targeted microbial isolation through reverse genomics.** a, Diagram showing the steps in reverse genomics-enabled microbiology: (1) Identification of membrane protein-encoding genes in SAGs and MAGs, (2) selection of predicted exposed epitopes, (3) antibody production, (4) purification and fluorescent labeling, (5) staining of target cells from microbiome samples and (6) isolation for (7) genomic sequencing or (8) cultivation. b, X-ray structure of penicillin-binding protein from E. coli, with the modeled TM7 glycosyltransferase domain colored blue and the selected epitope region highlighted in yellow–brown. c, Structural model of TM7a CpsC, with the peptide epitope region in purple.
also harbored different TM7 species (both types of organisms are contained multiple Cellulosimicrobium with based on 93% rRNA identity) were identified in the same colony co-isolates. Surprisingly, TM7 HOTs 346 and 348 (distinct species, Streptococcus colonies also contained species of Actinobacteria and most (80%) related to family/order taxonomic range): HOTs 346, 348, 352, 952/TM7x (Supplementary Datasets 1–3). Cultured cells were assayed using 16S rRNA amplicon sequencing (Methods). Following incubation in anaerobic or hypoxic conditions, T71 antibodies with anti-PBP2 and anti-CpsC antibodies and sorted individual fluorescent cells onto different liquid and solid culture media (see Methods). Following incubation in anaerobic or hypoxic conditions, cultured cells were assayed using 16S rRNA amplicon sequencing (Supplementary Datasets 1–3).

On solid media, we identified colonies containing five distinct TM7 phyotypes (human oral taxa (HOTs), representing species to family/order taxonomic range): HOTs 346, 348, 352, 952/TM7x (group 1) and HOT351 (group 3) (Figs. 3 and 4). All TM7-positive colonies also contained species of Actinobacteria and most (80%) also contained Streptococcus. We hypothesized that Actinobacteria were host organisms, as has been reported for TM7x (ref. 34), and that Streptococcus species, known to be highly adherent cells41, were co-isolates. Surprisingly, TM7 HOTs 346 and 348 (distinct species, based on 93% rRNA identity) were identified in the same colony with Cellulosimicrobium cellulans, an actinobacterium present in healthy saliva but occasionally associated with bacteremia and other diseases42. As it is unlikely that the fluorescent particle could have propagated in culture. Importantly, cultivation could enable identification of potential physiological interactions between various Saccharibacteria (TM7) lineages and other bacteria in the oral microbiota. Using freshly collected oral samples, we labeled bacteria with anti-PBP2 and anti-CpsC antibodies and sorted individual fluorescent cells onto different liquid and solid culture media (see Methods). Following incubation in anaerobic or hypoxic conditions, cultured cells were assayed using 16S rRNA amplicon sequencing (Supplementary Datasets 1–3).

The sort gate area indicates selection characteristics for cell sorting. Highly fluorescent particles (RFU > 10^4) are autofluorescent precipitates/minerals. The sort gate area indicates selection characteristics for cell sorting. Highly fluorescent particles (RFU > 10^4) are autofluorescent precipitates/minerals. Scale bar, substitutions per site. FSC, forward scatter, relative units; RFU, relative fluorescence units. Indentification of multiple oral TM7–host pairs provides additional evidence for interbacterial ecosymbiosis within this candidate phylum. Interestingly, based on four oral TM7–Actinobacteria association pairs, there is no apparent correlation in terms of phylogenetic distance between the TM7/Saccharibacteria species and their Actinobacteria hosts that would indicate cospeciation (Fig. 3a). This suggests that these interactions evolved separately.
perchament after oral colonization, and were not necessarily linked to speciation events. Host specificity appears somewhat limited for TM7x, as closely related Actinomyces cannot substitute the strain with which it was co-isolated 35. The findings from our study suggest that cells from distinct TM7 lineages may also simultaneously attach to the same host cell, and that the same TM7 species can use distinct hosts in vivo. This indicates some degree of flexibility in the establishment of the interspecies interaction. The molecular and cellular mechanisms by which different TM7 species interact with their hosts are unknown.

All of the TM7 species reported in this study are epibiont cocci of $<0.5\mu m$ in diameter present on the surface of host cells. We observed these species using both fluorescence in situ hybridization (FISH) and staining with the TM7 antibody (Fig. 4b). TM7 HOT351 cells are primarily associated with the pole of the rod-shaped Actinomyces host cells, while for TM7 HOTs 346, 352 and 952, there does not appear to be a localization preference, as previously noted 43. As with TM7x–Actinomyces sp. HOT897, we found that growth conditions affected the relative abundance of each TM7 species in coculture differently. Whereas TM7x abundance markedly increased in the presence of oxygen, growth of TM7 HOTs 351 and 346 was inhibited by micro-oxic conditions. The physiology of each TM7 lineage is linked to that of their hosts and is under current investigation.

**Reverse genomics for other uncultured microbial lineages.** We also applied our reverse genomics strategy to cultivate members of a microbial lineage that has no cultured representatives. We selected candidate bacterial phylum SR1/Abosconditabacteria. SR1 members are present at low abundance in a variety of environments including the human oral cavity. SR1 bacteria have highly reduced genomes, indicating likely dependence on interacting organism(s), and have recoded theopal stop codon (UGA) for glycine 44. Using single-cell genomic data of human oral SR1 HOT345, we selected a predicted surface-exposed transglycosylase, also annotated as a surface antigen involved in cell wall biogenesis (GenBank accession RAL55607). This protein is primarily extracellular, with one predicted transmembrane helix. We chose the entire exposed ~300-amino acid domain for antibody production (Methods). The antibody we generated was effective for labeling and flow sorting of SR1 cells from human oral samples (Fig. 5). We independently obtained two SR1 cultures. Sequencing of SSU rRNA amplicons using SR1-specific primers (see Methods and Supplementary Table 1) revealed that both contained SR1 HOT875, one of three known human oral phylotypes (Fig. 5). The cocultured bacteria, however, were different and were *Fusobacterium periodonticum* and *Parvimonas micra*, respectively. Species related to these bacteria have previously been shown to produce growth factors required by other fastidious bacteria 10,11. However, it is unknown whether oral SR1 bacteria are epibiotic on other bacteria, as we have not yet found conditions that could increase their abundance to a level that would enable detailed microscopic characterisation. Cultivation optimization conditions as well as characterisation of binding specificity of the SR1 antibody on bacteria in samples from various environments are under active study.

**Single-cell genomics enabled by reverse genomics.** Because not every organism may be readily cultivable, we also tested our targeted isolation approach to obtain single-cell genomes for various oral Saccharibacteria (TM7) lineages. Conceptually, targeted collection of cells representing specific species could provide enriched genome sequence information that might underpin physiological hypotheses or support population genomics studies. Flow cytometry enables analysis of more than 5,000 cells s$^{-1}$; therefore, taxa present at extremely low abundance ($<0.01\%$) could be isolated. Retrieving such organisms by standard, stochastic single-cell sorting and sequencing is impractical. Cell sorting combined with phylogenetic-specific labeling by FISH has occasionally been applied including for TM7 bacteria 11, but is inefficient and kills the cells. Our reverse genomics approach may be better suited for population...
Fig. 4 | Diversity of cultivated and uncultivated TM7 bacteria. a, Maximum-likelihood phylogenies of TM7 bacteria and of selected human oral Actinobacteria, based on SSU rRNA genes. TM7 bacteria include HOTs as well as related host-associated lineages from open environments, identified by sequence accession numbers and associated with six operational groupings (G1–6). HOTs detected using antibodies, isolated as single cells or identified in cultures are indicated by colored stars. Circles at nodes indicate bootstrap support (>80% filled, 50–80% open, only shown for major groupings). Lines connecting TM7 and Actinobacteria species indicate epibiont–host systems identified previously (black) and in this study (green). b, Epifluorescence microscopy images of TM7–host associations identified in this study using immunolabeling (TM7 HOT351–Actinomyces HOT897 and TM7 HOT346–C. cellulans) or FISH (TM7352–A. odontolyticus OR). Arrows point to green fluorescent TM7 cocc. Scale bar, 5 μm.
genome reduction has led to the loss of numerous functions, including one previously uncultured Actinomyces species. Antibody-enabled approaches may therefore be especially useful for microbiota whose members are refractory to traditional cultivation methods.

Although previous studies have attempted to bring TM7 into stable laboratory cultures, the only successful report achieved cultivation of TM7x based on serendipitous antibiotic resistance. Our method, dubbed ‘reverse genomics,’ exploits culture-independent sequence data for physical retrieval of specific microorganisms and does not depend on any physiological selection criteria. Therefore, it should be broadly applicable to any target organisms, as long as robust cultivation conditions can be identified for isolated cells.

We also applied our approach to isolate oral SR1 bacteria, members of a candidate phylum with no previously cultured representatives. Based on their highly reduced genomes, oral SR1 bacteria also likely depend on other microbes. Following antibody-targeted single-particle sorting, we detected growth of SR1 HOT875, one of three distinct human oral phylotypes. The genome sequences of many putative bacterial and archaeal lineages indicate partnerships such as exo- or endosymbiosis. Cultivation is essential for the characterization of such interspecies interactions. Physical retrieval of numerous live cells of even low-abundance target microbial taxa from complex communities enables a high-throughput search for optimal cultivation parameters and could be applicable to any organism for which a suitable surface antigen can be identified.
While for the Saccharibacteria targets we opted to select short peptides, entire protein domains could also be used as antigens for generating polyclonal antibodies, which would target multiple regions of the membrane protein. This may increase the chance of binding, especially if some residues are sterically hindered (for example, due to glycosylation). We used this approach for generating an effective antibody against SR1 bacteria. However, depending on the degree of sequence conservation, antibodies raised against large protein domains might be less specific.

Antibody-isolated cells could also be funneled into other platforms that enable high-throughput media selection and optimization protocols, for example, microdroplets or chips. Clearly, not every organism will be immediately culturable, but single-cell genomics data from targeted cells could be used to underpin population studies and make additional physiological inferences, and may ultimately result in more complete genome information and ensuing successful cultivation. Applying and further developing our strategy to other yet-uncultured taxa could begin to fill the many cultivation gaps in the tree of life.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41587-019-0260-6.

**References**

1. DeLong, E. F. & Pace, N. R. Environmental diversity of bacteria and archaea. *Syst. Biol.* 50, 470–478 (2001).
2. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* 1, 16048 (2016).
3. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358 (2017).
4. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2, 1533–1542 (2017).
5. Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437 (2013).
6. Guteleben, J. et al. The multi-omics promise in context: from sequence to microbial isolate. *Crit. Rev. Microbiol.* 44, 212–229 (2017).
7. Epstein, S. S. The phenomenon of microbial uncultivability. *Curr. Opin. Microbiol.* 16, 636–642 (2013).
8. Overmann, J., Abt, B. & Sikorski, J. Present and future of culturing bacteria. *Annu. Rev. Microbiol.* 71, 711–730 (2017).
9. D’Onofrio, A. et al. Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* 17, 254–264 (2010).
10. Strandwitz, P. et al. GABA-modulating bacteria of the human gut microbiota. *Nat. Microbiol.* 4, 396–403 (2018).
11. Vartoukian, S. R. et al. In vitro cultivation of ‘unculturable’ oral bacteria, facilitated by community culture and media supplementation with siderophores. *PLoS ONE* 11, e0146926 (2016).
12. Cross, K. L. et al. Insights into the evolution of host association through the isolation and characterization of a novel human periodontal pathobiont, *Desulfovibrio oralis*, *mBio* 9, e02061–17 (2018).
13. Huber, H. et al. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67 (2002).
14. St. John, E. et al. A new symbiotic nanoarchaeote (*Candidatus Nanocefa*) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst. Appl. Microbiol.* 42, 94–106 (2018).
15. Wurch, L. et al. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nat. Commun.* 7, 1–10 (2016).

16. Pulschen, A. A. et al. Isolation of uncultured bacteria from antarctica using long incubation periods and low nutritional media. *Front. Microbiol.* 8, 1346 (2017).

17. Oliver, J. D. Recent findings on the viable but nonculturable state in pathogenic bacteria. *FEBS Microbiol. Rev.* 34, 415–425 (2010).

18. Terekhov, S. S. et al. Microfluidic droplet platform for ultra-high-throughput single-cell screening of biodiversity. *Proc. Natl Acad. Sci. USA* 114, 2550–2555 (2017).

19. Zengler, K. et al. High-throughput cultivation of microorganisms using microcapsules. *Methods Enzymol.* 397, 124–130 (2005).

20. Browne, H. P. et al. Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature* 533, 543–546 (2016).

21. Lagier, J. C. et al. Culture of previously uncultured members of the human gut microbiota by culturoomics. *Nat. Microbiol.* 1, 16203 (2016).

22. Oberhardt, M. A. et al. Harnessing the landscape of microbial culture media to predict new organism-media pairings. *Nat. Commun.* 6, 8493 (2015).

23. Bentz, B., Spiering, A. I., Ling, L. L. & Epstein, S. S. In situ cultivation of previously uncultivable microorganisms using the ichip. *Nat. Protoc.* 12, 2232–2242 (2017).

24. Sizova, M. V. et al. New approaches for isolation of previously uncultivated oral bacteria. *Appl. Environ. Microbiol.* 78, 194–203 (2012).

25. Rheims, H., Rainey, F. A. & Stackebrandt, E. A molecular approach to search for diversity among bacteria in the environment. *J. Ind. Microbiol.* 17, 159–169 (1996).

26. Hugenholtz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180, 4765–4774 (1998).

27. Albertsen, M. et al. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538 (2013).

28. Dewhirst, F. E. et al. The human oral microbiome. *J. Bacteriol.* 192, 5002–5017 (2010).

29. Brinig, M. M., Lepp, P. W., Ouverney, C. C., Armitage, G. C. & Relman, D. A. Prevalence of bacteria of division TM7 in human subgingival plaque and their association with disease. *Appl. Environ. Microbiol.* 69, 1687–1694 (2003).

30. Kuehbacher, T. et al. Intestinal TM7 bacterial phylogenies in active inflammatory bowel disease. *J. Med. Microbiol.* 57, 1569–1576 (2008).

31. Dewhirst, F. E. et al. The feline oral microbiome: a provisional 16S rRNA gene based taxonomy with full-length reference sequences. *Vet. Microbiol.* 175, 294–303 (2015).

32. Dewhirst, F. E. et al. The canine oral microbiome. *PLoS ONE* 7, e36087 (2012).

33. Dudak, N. K. et al. Novel microbial diversity and functional potential in the marine mammal oral microbiome. *Curr. Biol.* 27, 3752–3762.e6 (2017).

34. Marcy, Y. et al. Dissecting biological ‘dark matter’ with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc. Natl Acad. Sci. USA* 104, 11889–11894 (2007).

35. He, X. et al. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc. Natl Acad. Sci. USA* 112, 244–249 (2015).

36. Haghhighat, S., Siadat, S. D., Sorkhahadi, S. M. R., Sepahi, A. A. & Mahdavi, M. A novel recombinant vaccine candidate comprising PBP2a and autolysin against Methicillin Resistant *Staphylococcus aureus* confers protection in the experimental mouse. *Mol. Immunol.* 91, 1–7 (2017).

37. Lovering, A. L., de Castro, L. H., Lim, D. & Strynadka, N. C. Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis. *Science* 315, 1402–1405 (2007).

38. Simons, K. T. et al. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins* 34, 82–95 (1999).

39. Palmer, R. J., Ir. et al. Interbacterial adhesion networks within early oral biofilms of single human hosts. *Appl. Environ. Microbiol.* 83 (2017).

40. Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E. & Borisy, G. G. Biogeography of a human oral microbiome at the micron scale. *Proc. Natl Acad. Sci. USA* 113, E791–E800 (2016).

41. Jakabolics, N. S., Vass, S. A. & Rickard, A. H. Community interactions of oral streptococci. *Adv. Appl. Microbiol.* 87, 43–110 (2014).

42. Heym, B. et al. Molecular detection of *Cellulosimicrobium cellulosum* as the etiological agent of a chronic tongue ulcer in a human immunodeficiency virus-positive patient. *J. Clin. Microbiol.* 43, 4269–4271 (2005).

43. Bor, B. et al. Phenotypic and physiological characterization of the epibiotic interaction between TM7x and its basibiont actinomyces. *Microb. Ecol.* 71, 243–255 (2016).

44. Campbell, J. H. et al. USA is an additional glycinol codon in uncultured SR1 bacteria from the human microbiota. *Proc. Natl Acad. Sci. USA* 110, 5540–5545 (2013).

45. Poulos, M. et al. Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl. Environ. Microbiol.* 73, 3205–3214 (2007).

46. Campbell, A. G. et al. Diversity and genomics insights into the uncultured *Chloroflexi* from the human microbiota. *Environ. Microbiol.* 16, 2635–2643 (2014).

47. Castelle, C. J. et al. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* 16, 629–645 (2018).

48. Ferrari, B. C., Binnerup, S. J. & Gillings, M. Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. *Appl. Environ. Microbiol.* 71, 8714–8720 (2005).

49. Soro, V. et al. Axenic culture of a candidate division TM7 bacterium from the human oral cavity and biofilm interactions with other oral bacteria. *Appl. Environ. Microbiol.* 80, 6480–6489 (2014).

50. Zengler, K. et al. Cultivating the uncultured. *Proc. Natl Acad. Sci. USA* 99, 15681–15686 (2002).

**Acknowledgements**

We thank S. Allman, S. Kauffman, S. Lebreux, L. Sukharnikov and M. Robeson for technical assistance. Support for this work was provided by grants (nos. R56DE021567 and R01DE024463) from the National Institute of Dental and Craniofacial Research of the US National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. J.M.P. was supported by the Laboratory Directed Research and Development program at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC for the US Department of Energy (contract no. DE-AC05-00OR22755). C.J.C. was supported by a National Science Foundation Graduate Research Fellowship (grant no. 2017219379). This work used resources of the Compute and Data Environment for Science at Oak Ridge National Laboratory.

**Author contributions**

M.P. conceived the study, designed and performed experiments and analyzed data. K.L.C. and J.H.C. designed and performed experiments and analyzed data. M.B., A.G.C., S.J., D.K. and Z.Y. designed and performed experiments. C.C. and J.P. performed protein sequence analyses and structure modeling. M.P., M.H., A.G. and E.L. recruited human subjects and collected oral samples. K.L.C. and M.P. wrote the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41587-019-0260-6.

**Correspondence and requests for materials** should be addressed to M.P.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Ethics statement. Human subject recruitment and sampling protocols were approved by the Ohio State University Institutional Review Board and by the Oak Ridge Site Wide Institutional Review Board. Written, informed consent was obtained from all participants. The use of commercial custom antibodies was approved by the Oak Ridge National Laboratory Animal Care and Use Committee. A list of participants and samples is provided in Supplementary Table 1.

Sample collection and processing. Saliva samples were self-collected by healthy donors by passive drooling, using a SalivaBio collection aid (Salivometrics), into sterile cryovials. Saliva samples of 1–2 ml were collected at least 2 h after eating or drinking, kept at room temperature and processed immediately or within 2 h after collection. Recurring samples were collected from the same donors as needed for individual experiments. Gingival crevicular fluid from patients with periodontitis was collected from affected teeth using sterile endodontic paper points that were placed after irrigation for 1 min or by scraping the gingival tissues with a 1 ml syringe and dental transport medium (Aneura Systems), transported to the laboratory and further processed within 24 h. Patients with periodontitis were sampled once.

Oral microbiota samples were diluted 1:5 in dental transport medium (per liter: Na thioglycollate 1.0 g, Na2HPO4 1.15 g, NaCl 3.0 g, KCl 0.2 g, KH2PO4 0.2 g, MgSO4·7H2O 0.1 g, L-cysteine HCl 0.05 g, pH 7.3), mixed by vortexing for 1 min then centrifuged at 1,000g for 2 min to remove large particles. The supernatant was passed through 10-μm CellTrics filters (Symex) and the cells were collected by centrifugation (12,000g for 15 min), followed by resuspension in 1 ml PBS. An aliquot was used for total bacterial DNA extraction using proteinase K digestion and phenol-chloroform extraction15. For fixation, an aliquot was fixed by addition of an equal volume of ice-cold 100% ethanol and stored at −20°C overnight. For live cell sorting, the processed microbiota samples were used immediately.

TM7a genome mining, epitope selection and antibody generation. For selection of candidate proteins to serve as immunogens, we used the single-cell genomics data generated by Marcy et al.11, the only oral TM7 genomes available at the time we initiated this work. We used the TM7a (6,806 protein genes) and TM7c (590 protein genes) datasets, based on the annotations available in IMG (https://img.jgi.doe.gov)54, with the expectation that some of the genes may represent organisms other than TM7 (ref. 1). We first selected genes encoding proteins associated with membrane processes, based on clusters of orthogonal groups (COG) categories (transport functions, cell wall/membrane/envelope biogenesis, extracellular structures), using the IMG assignments. The corresponding proteins were analyzed for the presence of transmembrane-anchoring helical domains using TMHMM v.2.0 (ref. 1). From the list of proteins with such domains, we eliminated the ones annotated as hypothetical as well as those shorter than 100 amino acids, as we aimed for a final selection of relatively large proteins with known functions. The proteins that passed these criteria (117 combined between TM7a and TM7c) were analyzed by BLASTP against the proteomes of all oral bacteria downloaded from the Human Oral Microbiome Database (http://www.homd.org/). Forty-eight of the proteins had high sequence similarity (>90%) to proteins from species of Leptotrichia (Fusobacteria), indicating that they are likely contaminants, as previously reported3. We then individually analyzed the remaining sequences for number of transmembrane domains, number and size of predicted extracellular domains, availability of 3D structures for homologs from other bacteria and experimental data related to antigenicity. The list of proteins with a predicted transmembrane domain and an extracellular region composed of LysM and CHAP domains was set to avoid sorting such particles. For genomic and community diversity sorting for cultivation experiments, after 30 min, the samples were analyzed by flow cytometry. Negative controls included unstained oral sample (to check for auto-fluorescence), and samples processed in parallel with the oral sample but without microbiota (to check for fluorescent antibody precipitates).

For flow cytometry and cell sorting, we used a Cytopeia/BD Inflx Model 2085 (Cytopeia, now BD) in a biological safety cabinet designed specifically for the Inflx cell sorter (Baker). Preparation of the flow cytometer for single-cell sorting for genomic amplification and sequencing was as we and others have previously described55,56. For sorting, we used a 70-μm nozzle and the 488-nm laser for forward–side scatter (FSC–SSC) analyses as well as fluorescence detection and data acquisition. For sorting trigger, parameters were chosen for an average of ≤10,000 SSQ events for each experiment and sample. A very low level of highly autofluorescent particles (<0.01%, relative fluorescence units >10) was observed in all samples in the absence of staining with the antibody, and gating parameters were set to avoid sorting such particles. For genomic and community diversity sorting, single-particle data were processed as described (ref. 56) with the Tri-colors (Tri-Hc10 (10 mM PH 8.0) and 1 mM EDTA (TE) in individual wells of 96-well plates. We observed sufficient fluorescence signal with primary labeled antibody preparations. However, if the signal was low, we used a secondary labeled antibody (for example, goat anti-rabbit IgG) would provide signal amplification.

Isolation and cultivation of oral TM7 bacteria. For cultivation, we used saliva samples collected recurrently from three healthy donors and processed them as detailed in the section “Sample collection and processing”. Antibody staining was performed as described under “Immunofluorescence labeling of oral microbiota samples and flow cytometry”. For the initial cultivation tests, in 200 μl liquid medium in 96-well plates, we cultured TM7 bacteria as basal media commercial Tryptic Soy Broth (TSB; Difco) supplemented with L-cysteine HCl 0.05 g, pH 7.3, as well as basal heart infusion (BHI; Difco), Oral Treponeema Enrichment Broth (OTEB; Aneura Systems), Membrane Tryptone Glucose Extract (MTGE) Broth (Aneura Systems) and Tryptic Soy Broth (TSB; Difco), supplemented with various additional factors (ATCC vitamins, trace minerals, clarified filtered salmon, pig gastric mucin, sugars, amino acids and nucleotides, N-acetylmuramic acid, N-acetylglucosamine, pyruvate). Following flow sorting, performed as described under “Immunofluorescence labeling of oral microbiota samples and flow cytometry,” the plates were incubated in an anaerobic chamber (COY) at 37°C, under 85% N2, 10% CO2, 5% H2, or in anaerobic jars to which we added air via a syringe port to reach 2% O2. To test for TM7 growth, after 48–72 h we retrieved 100 μl cultures using multichannel pipettes and vacuum filtering through 0.2 μm filter 96-well plates (Millipore). The cells were washed with 200 μl PBS. Then, 20 μl TE was added to the filter. The plates were placed on an orbital shaker at 100 rpm for 5 min at room temperature. Then, 10 μl of the cell suspension was recovered in 96-well PCR plates, to which we added 10 μl lysis solution (0.13 M

Immunofluorescence labeling of oral microbiota samples and flow cytometry. Aliquots of oral microbiota samples, processed as described above and suspended in PBS with 1% serum, were blocked for 30 min with 5% goat serum and a mixture of IgGs (1 μg/ml each rabbit anti-IgGc IgGδ, anti-Cladostrium IgGδ and human IgG, I506, Sigma) for blocking potential nonspecific antibody-binding sites on some oral bacteria. We also used 5% rabbit prebleed serum instead of the goat serum, with no observed differences. Following blocking, fluorescently labeled anti-PBP2, anti-CpsC or anti-SR1 IgG was added to samples at 0.1–1 μg/ml. It was important to centrifuge the antibody solution (15,000g for 5 min) and carefully recover supernatant before addition to the microbiota samples, to reduce the background of fluorescent particles. Antibody labeling was conducted for 1 h at room temperature, followed by either centrifugation (12,000g for 15 min) and resuspension in 2 ml PBS supplemented with 1% serum or directly diluting in 2 ml PBS with 1% serum (mainly for live cell sorting for cultivation experiments). After 30 min, the samples were analyzed by flow cytometry. Negative controls included unstained oral sample (to check for auto-fluorescence), and samples processed in parallel with the oral sample but without microbiota (to check for fluorescent antibody precipitates).

For flow cytometry and cell sorting, we used a Cytopeia/BD Inflx Model 2085 (Cytopeia, now BD) in a biological safety cabinet designed specifically for the Inflx cell sorter (Baker). Preparation of the flow cytometer for single-cell sorting for genomic amplification and sequencing was as we and others have previously described55,56. For sorting, we used a 70-μm nozzle and the 488-nm laser for forward–side scatter (FSC–SSC) analyses as well as fluorescence detection and data acquisition. For sorting trigger, parameters were chosen for an average of ≤10,000 SSQ events for each experiment and sample. A very low level of highly autofluorescent particles (<0.01%, relative fluorescence units >10) was observed in all samples in the absence of staining with the antibody, and gating parameters were set to avoid sorting such particles. For genomic and community diversity sorting, single-particle data were processed as described (ref. 56) with the Tri-colors (Tri-Hc10 (10 mM PH 8.0) and 1 mM EDTA (TE) in individual wells of 96-well plates. We observed sufficient fluorescence signal with primary labeled antibody preparations. However, if the signal was low, we used a secondary labeled antibody (for example, goat anti-rabbit IgG) would provide signal amplification.
We did not pursue characterization of mixed SAGs. Also, as Gram-positive bacteria (such as Actinobacteria) cannot be efficiently lysed by the alkaline treatment step of the MDA protocol, identifying the potential hosts of those TM7 cells would be unlikely.

To characterize the microbial diversity in cell populations sorted following antibody staining (10–100 cells), genomic amplification by MDA was conducted in the same way as for single cells. The amplification product was first purified using a Zymo genomic DNA cleanup kit (Zymo Research). To determine the microbial composition of the original oral microbiota samples we used genomic DNA purified from the saliva and subgingival samples. Microbial diversity was determined by amplification of the hypervariable V4 region of the SSU rRNA gene followed by MiSeq amplicon sequencing. We followed the protocol described by Lundberg et al., with a modification in the universal primers. Specifically, because the canonical 515F primer sequence (5′-GTCCACGCMGCCGCGGTAA-3′) does not recognize the TM7 rRNA, we replaced the PCR primer set with the 5′-515F-derivative that can bind TM7 rRNA (5′-GTCCACGCMGCCGCGGTTC-3′). The reverse primer, 806R (5′-GGACTACHVGGGTWTCTAAT-3′), can recognize the full diversity of oral bacterial SSU rRNA. The purified amplicons were sequenced on the MiSeq platform (Illumina) on v2 or v3 sequencing chips using 2×250-nucleotide reads. Read processing and microbial diversity analysis were performed using the software programs cutadapt16, usearch17 and qiime17, as well as the HOMD SSU rRNA taxonomy database, and following the commands and parameters listed in Supplementary Dataset 4. To determine whether the 254-nucleotide V4 amplicon provides sufficient resolution to enable distinction between TM7 and other microbial species (OTUs) and performed phylogenetic analysis including the TM7 OTUs identified based on sequencing the V4 amplicon (TM7HOTs_OTUs.fasta) into SequenceMining file and files deposited in the SRA archive linked to NCBI BioProject PRJNA472398). As shown in Supplementary Fig. 2 (sequences provided in the online Data S1 table containing only TM7 reads of the 254- and 322-nucleotide paired reads) at the Hudson Alpha Institute for Biotechnology (Huntsville, AL, USA) or on MiSeq (250-nucleotide paired reads) at Oak Ridge National Laboratory. Approximately 100 million reads were obtained for each SAG sequenced on HiSeq and 10 million for the MiSeq SAGs.

SAG assembly, annotation and comparative genomic analysis. For individual assembly, the reads for each SAG were trimmed using Trim Galore v0.4 (https://github.com/FelixKrueger/TrimGalore) and assembled using SPAdes (v3.9.0).18

Removal of non-TM7 contigs (human and other bacterial sequences) was performed by a combination of GC-content-based binning and Emergent Self-Organizing Maps (v3) based on k-mer frequency.19 Further contamination was removed using DNA and protein BLAST and single-copy gene analysis.20 Lastly, further assembly in Geneious (v8.9-10)21 consolidated some contigs and the final cleaned genomes were analyzed with Prokka (v1.11)22 for gene prediction and annotation. Completion and contamination estimates were calculated using KEGG MapViewer (v5). Analysis of genome workflow and a manually adapted gene workflow containing 42 single-copy genes conserved in bacteria of the candidate phyla radiation (Supplementary Table 1, Supplementary Note and Supplementary Figs. 3 and 4). Pfam (v31), eggNOG (v4.5.1) and Interproscan (v5.27-66.0) databases were used for protein assignment to families and orthologous groups. Anvio (v1.5.1) was used for core genome analysis and identification of conserved COGs following the pangenome pipeline with mapping of eggNOG classifications (Supplementary Table 1 and Supplementary Figs. 3 and 5). NCBI, IMG and Geneious were used for basic homology searches and for obtaining genomic data for further metabolic inference analyses (Supplementary Note). SSU rRNA and concatenated single-copy protein gene trees were generated in Geneious using RAxML and then 1–250 bootstraps (Supplementary Dataset 6). The TM7 SAG sequences and associated metadata were deposited in GenBank under BioProject PRJNA472398. SSU rRNA sequence alignments that include reference strains as well as detected phylotypes, SAGs and isolates obtained in this study are provided as Supplementary Datasets 1–3.

Protein structure modeling. The glycosyltransferase domain of PB2 from TM7 (residues 95–278) and TM7x (residues 145–334) was modeled using a combination of homology modeling and coevolution-based residue–residue contact prediction. We used the GREMLIN web server23 to identify structural templates, predict residue contacts based on coevolution analysis and generate contact restraints for MODELLER. We then used 1–250 bootstraps. Briefly, the server uses HHMINTO generate a multiple sequence alignment, which is then filtered to include only sequences that cover at least 75% of the query sequence, have fewer than 75% gaps and have a maximum mutual sequence identity of 90%. From this multiple sequence alignment, GREMLIN learns a global statistical model that accounts for both sequence conservation and coevolution, and predicts residue–residue contacts.
Articles used as templates for modeling PBP2: 3VMA (ref. 90) (which is missing part of initial threaded models generated by map_align were then used as input for align (https://github.com/gjoni/map_align) to align the contact maps predicted additional benefits compared with standard homology modeling. We used map_template structures are available, we included coevolution information to provide various oral sample types (saliva, subgingival crevicular fluid) to broaden the diversity of target TM7 bacteria that could be found across a small sampled population. Sorting and cultivation of TM7 bacteria. Fluorescence imaging of TM7 bacteria and their Actinobacteria hosts was performed in two separate experiments, with Fig. 4 showing a combination of panels from both experiments. The cell size, morphology and observable interactions between the TM7 bacteria and their host cells are similar to those reported for TM7x (ref. 17). Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article. Data availability Annotated TM7 SAGs are deposited in GenBank under the BioProject PRJNA472398. Raw and aligned SSU rRNA sequences are provided as Supplementary Datasets 1–3. MiSeq amplicon data are deposited in SRA linked to BioProject PRJNA472398.

References

51. Markowitz, V. M. et al. The integrated microbial genomes (IMG) system. Nucleic Acids Res. 34, D344–D348 (2006).

52. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580 (2001).

53. Sauvage, E., Kerrf, F., Terrak, M., Ayala, J. A. & Charlier, P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEBS Microbiol. Rev. 32, 234–258 (2008).

54. Haghighat, S., Siadat, S. D., Sorkhabadi, S. M., Sepahi, A. A. & Mahdavi, M. Cloning, expression and purification of penicillin binding Protein2a (PBPA2) from Methicillin resistant Staphylococcus aureus; a study on immunoreactivity in Balb/c mouse. Avicenna J. Med. Biotechnol. 5, 204–211 (2013).

55. Zarantonelli, M. L. et al. Immunogenicity of meningococcal PBP2 during natural infection and protective activity of anti-PBP2 antibodies against meningococcal bacteriaemia in mice. J. Antimicrob. Chemother. 57, 924–930 (2006).

56. Bocchini, J. P., Morona, J. K., Paton, J. C. & Morona, R. Identification of Streptococcus pneumoniae Cps2C residues that affect capsular polysaccharide polymerization, cell wall ligation, and Cps2D phosphorylation. J. Bacteriol. 193, 2341–2346 (2011).

57. Tonio, C. et al. Streptococcus agalactiae capsule polymer length and attachment is determined by the proteins CpsABC and J. Biol. Chem. 290, 9521–9532 (2015).

58. Morona, R., Purins, L., Tocilj, A., Matte, A. & Cyliger, M. Sequence-structure relationships in polysaccharide co-polymer (PCP) proteins. Trends Biochem. Sci. 34, 78–84 (2009).

59. Saha, S. & Raghava, G. P. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65, 40–48 (2006).

60. Larsen, J. E., Lund, O. & Nielsen, M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2, 2 (2006).

61. Heimel, T. et al. A complex endorembosome system in the archeaon Ignicoccus hospitalis tapped by Nanoarchaeum equitans. Front. Microbiol. 8, 1566 (2017).

62. Miller, L. D. et al. Establishment and metabolic analysis of a model microbial community for understanding trophic and electron accepting interactions of subsurface anaerobic environments. BMC Microbiol. 10, 149 (2010).

63. Campbell, A. G. et al. Multiple single-cell genomes provide insight into functions of uncultured Deltoproteobacteria in the human oral cavity. PLoS ONE 8, e59361 (2013).

64. Podar, M. et al. Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. Biol. Direct 8, 9 (2013).
65. Rinke, C. et al. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* **9**, 1038–1048 (2014).

66. Wong, L. & Sissons, C. A comparison of human dental plaque microcosm biofilms grown in an undefined medium and a chemically defined artificial saliva. *Arch. Oral Biol.* **46**, 477–486 (2001).

67. Pernthaler, A. & Pernthaler, J. Fluorescence in situ hybridization for the identification of environmental microbes. *Methods Mol. Biol.* **353**, 153–164 (2007).

68. Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D. & Dangl, J. L. Practical innovations for high-throughput amplicon sequencing. *Nat. Methods* **10**, 999–1002 (2013).

69. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

70. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

71. Caporaso, J., Kuczynski, J., Stombaugh, J. & Bittinger, K. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

72. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

73. Brown, C. T. et al. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).

74. Hug, L. A. et al. Critical biogeochemical functions in the subsurface are interconnected biogeochemical processes in an aquifer system. *Proc. Natl Acad. Sci. USA* **110**, 15674–15679 (2013).

75. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

76. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

77. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

78. Anantharaman, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).

79. Finn, R. D. et al. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).

80. Huerta-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).

81. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

82. Delmont, T. O. & Eren, A. M. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* **6**, e4320 (2018).

83. Eren, A. M. et al. Anvi’o: an advanced analysis and visualization platform for ‘omics data. *PeerJ* **3**, e1319 (2015).

84. Balakrishnan, S., Kameshree, H., Carbonell, J. G., Lee, S. I. & Langmead, C. J. Learning generative models for protein fold families. *Proteins* **79**, 1061–1078 (2011).

85. Kamisetty, H., Ovchinnikov, S. & Baker, D. Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. *Proc. Natl Acad. Sci. USA* **110**, 15674–15679 (2013).

86. Ovchinnikov, S. et al. Protein structure determination using metagenome sequence data. *Science* **355**, 294–298 (2017).

87. Remmert, M., Biegert, A., Hauser, A. & Soding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2011).

88. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).

89. Burkley, S. K. et al. Protein Data Bank (PDB): the single global macromolecular structure archive. *Methods Mol. Biol.* **1607**, 627–641 (2017).

90. Sung, M. T. et al. Crystal structure of the membrane-bound bifunctional transglycosylase PBP1b from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **106**, 8824–8829 (2009).

91. Han, S. et al. Distinctive attributes of beta-lactam target proteins in *Acinetobacter baumannii* relevant to development of new antibiotics. *J. Am. Chem. Soc.* **133**, 20536–20545 (2011).

92. Yuan, Y. et al. Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. *Proc. Natl Acad. Sci. USA* **104**, 5348–5353 (2007).

93. Huang, C. Y. et al. Crystal structure of *Staphylococcus aureus* transglycosylase in complex with a lipid II analog and elucidation of peptidoglycan synthesis mechanism. *Proc. Natl Acad. Sci. USA* **109**, 6496–6501 (2012).

94. Gront, D., Kulp, D. W., Vernon, R. M., Strauss, C. E. & Baker, D. Generalized fragment picking in Rosetta: design, protocols and applications. *PLoS ONE* **6**, e23294 (2011).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|-----|-----------|
| ☐   | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐   | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐   | The statistical test(s) used AND whether they are one- or two-sided |
|     | Only common tests should be described solely by name; describe more complex techniques in the Methods section. |
| ☐   | A description of all covariates tested |
| ☐   | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐   | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐   | For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted |
|     | Give P values as exact values whenever suitable. |
| ☐   | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐   | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐   | Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated |

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
Flow cytometry data and was collected using the BD FACS™ Software Sorter Software (v1) installed on the BD Influx flow cytometer system

Data analysis
We used the following publicly available software for data analysis: Trim Galore v0.4, SPADES v3.9.0, Geneious v8-v10, Prokka v1.11, CheckM v1.0.11, Anvi’o v5.1, HHsearch, usearch v8, qiime v1, PyMOL v2, TMHMM v.2.0, ABCpred v1, BepiPred v1, IEBD Analysis Resource v3.6.0, Bio-Synthesis peptide design online tool [www.biosyn.com], Adobe Photoshop vCS6. The use of every software is described in the Methods section. A file with commands used for MiSeq data processing is provided as Supplementary file. No custom algorithm or software was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Single cell genomic data that support the findings of this study have been deposited in GenBank under the Bioproject accession number PRJNA472398. 16S rRNA sequences [alignments and raw Sanger data chromatograms] are provided as Supplementary files. MiSeq raw amplicon sequence data was deposited in the GenBank SRA archive under Bioproject accession number PRJNA472398.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-list.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The number of human subjects used was based on biological sample requirements for individual experiments. No statistical comparisons were made between health-disease state or sample types (saliva-subgingival fluid). The biological samples were collected and used to develop the molecular methods, to isolate and culture bacteria, and did not serve to make health-disease correlations or other statistical inferences.

Data exclusions

No data was excluded.

Replication

All attempts at molecular and microbiological replication were successful within known ranges of biological variations across human oral microbiomes. Sequencing of recurring samples revealed reproducible presence of target organisms. Antibody staining and flow cytometry was reproducible across experiments, time series and samples. The target bacteria isolated using antibody-based sorting were obtained multiple times independently, as further detailed in the Methods and the Statistics and Reproducibility section.

Randomization

Human participants were grouped based on health or disease status. The study did not aim at comparing health and disease, those categories were only used to cover potential biological variability in the microbiome.

Blinding

Blinding was not relevant as the study did not compare different types of samples or human donors.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

| n/a | Involved in the study | n/a | Involved in the study |
|-----|----------------------|-----|----------------------|
| ☒ | Antibodies | ☒ | ChiP-seq |
| ☒ | Eukaryotic cell lines | ☒ | Flow cytometry |
| ☒ | Palaeontology | ☒ | MRI-based neuroimaging |
| ☒ | Animals and other organisms | ☒ | |
| ☒ | Human research participants | ☒ | |
| ☒ | Clinical data | ☒ | |

Antibodies

Custom made polyclonal antibodies against synthetic peptides (TM7 antibodies), protein (SR1 antibody) and whole cells (Staphylococcus, Clostridium antibodies) were generated commercially by Bio-Synthesis, Inc, General Bioscience Corporation and GenScript Corporation. Dilutions were determined empirically and are specified in the Methods section. Commercially available human IgG used for blocking was from Sigma, St. Louis MO, USA (I4506).

Validation

The antibodies were tested against the immunogenic peptides and proteins by ELISA by the manufacturer. The antibodies labeled target uncultured organisms and also stained some of the isolated bacteria. The antibodies may cross react with other species as well, but that cross-reactivity was not specifically characterized.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Oral samples were collected from healthy adults and from adults with periodontitis. Both males and females were sampled but gender, age, race or ethnic origin were not factors in the study. Other The results of the study are not correlated to specific participants or sample type.

Recruitment

The purpose of the clinical samples collection was to provide a fresh supply of saliva or subgingival fluid for the laboratory studies. Participants volunteered for the study following IRB approved advertisement and were sampled based on availability on
specific dates. Because the target bacteria are universally present across the human population regardless of age, gender, race, the selection of volunteers is unlikely to have influenced the results of this study.

**Ethics oversight**

Human subjects recruitment and sampling protocols were approved by the Ohio State University Institutional Review Board and by the Oak Ridge Site-Wide Institutional Review Board. Written informed consent was obtained from all participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

### Flow Cytometry

**Plots**

- Confirm that:
  - The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
  - The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
  - All plots are contour plots with outliers or pseudocolor plots.
  - A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

**Sample preparation**

From the Methods section: Aliquots of oral microbiota samples suspended in PBS were blocked for 30 minutes with 5% goat serum and a mixture of IgGs (for blocking potential non-specific antibody-binding sites on some oral bacteria. We also used 5% rabbit pre-bleed serum instead of the goat serum, with no observed differences. Following blocking, fluorescently labelled anti-PBP2 or anti-CpsC IgG were added to the samples at 0.1-1 µg/ml. Antibody labelling was conducted for 1 hour at room temperature, followed by either centrifugation (12,000 x g for 15 minutes) and resuspension in 2 ml PBS supplemented with 1% serum or directly diluting in 2 ml PBS with 1% serum (mainly for live cell sorting for cultivation experiments). After 30 minutes, the samples were analyzed by flow cytometry. Negative controls included unstained oral sample (to check for auto-fluorescence), and samples processed in parallel with the oral sample but without microbiota (to check for fluorescent antibody precipitates).

**Instrument**

Cytopeia/BD Influx Model 208S

**Software**

BD Influx software

**Cell population abundance**

The sorted population of bacteria represented 1-5 % of the total depending on sample type and experiment. The precise abundance was not a factor in results interpretation.

**Gating strategy**

Sorted fluorescent cell population was based on comparison with negative controls (no antibody staining).

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.