Improvement of Microbial Detection and Analysis Techniques in Complex Biological Environments

Thesis by

Michael K Porter

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2024

(Defended 23 May 2024)

Michael K Porter
ORCID: 0000-0002-0777-7563

ACKNOWLEDGEMENTS

There are many people I need to thank for my time here at Caltech. First, I would like to thank my adviser, Rustem Ismagilov, for all of the support and mentorship I received through my seven years in your lab. Thank you for taking me into your lab and always challenging me to be better, both as a scientist and a person. I have changed immensely during my time working with you, and I am extremely grateful for this.

I would like to thank my thesis committee, Zhen-Gang Wang, David Tirrell, and Jared Leadbetter, for their continued guidance and support.

I was very lucky to have a robust support group of friends and colleagues I met at Caltech. First, I want to thank Alex Viloria Winnett; not enough can be said about our friendship. Thank you for being a consistent support and being someone who always pushed me during grad school and life. I don't know if I could've finished this without you. Reid Akana, thank you for being hilarious yet kind. You are one of the smartest people I know; I could always expect to learn something new and crazy (even if not necessarily useful or productive) hanging out with you. I always enjoyed your colorful commentary (a good use of an English degree), and you were a great partner in crime. To Minkyo Lee, thank you for your chill presence you provided in lab and the late-night food and drink runs we would often go on. I will treasure all of the memories of us laughing and ranting about whatever craziness was happening in our life at the time. To Asher Preska Steinberg, thank you for taking me under your wing and being such an awesome mentor all these years. You really helped me grow as a person and scientist during my early graduate school years, and even now I appreciate your guidance and friendship. To Anna Romano, thank you for being such a calm and stable support presence in the lab. Your technical expertise is obviously unmatched, but I have learned so much more from you about how to be a better scientist. To Tasha Shelby, thank you for all your help writing and idea generating through the years. You always manage to be so upbeat and uplifting, even in the face of a crazy impending deadline. To Mi Kyung Kim, thank you for all your support. I will miss coming to your office to chat and eat candy. Finally, I would like to thank all my lab mates that I have worked with through the years for their support.

During my time at Caltech, I did an internship at Sangamo Therapeutics. This opportunity opened my world to the interesting science and problems being tackled in biopharma and ultimately pushed me towards a career as a scientist in the biotech industry. I am extremely grateful for my manager and mentor, Sandeep Yadav; thank you for taking me onto your team and helping me navigate my early career uncertainties. I would also like to thank the formulation development team, Momina Mahmood, Madhura Som, and Katie Esposto, for showing me the ropes and supporting me during my internship and beyond. The positive environment fostered by all of you played an immense impact in shaping my early career decisions.

I want to thank my tennis family, GFT Club, for taking me in and being a constant support system outside of Caltech. I will treasure the memories of our victories, and I'm sure the club

will have many more successes in the future. I want to thank everyone who I've had the chance to play and hit with over the years for contributing to the vibrant SoCal tennis community that I've been able to enjoy during my time here.

Lastly, I want to thank my family for their unwavering support through my education. I want to thank my amazing wife, Songnan Wang, for putting up with seven years of long distance and my incessant complaining through the ups and downs of grad school. For some reason you watched me do this and decided you wanted to pursue a PhD yourself; I look forward to returning the favor and supporting you as we grow together. To my mom and dad, who also listened to me constantly complain but still supported me anyway, thank you. And Thomas, you probably thought you could graduate before me, but you won't. Your thesis is probably going to be boring, too.

I apologize to anyone whom I may have left off, as this was written somewhat last minute.

ABSTRACT

Human bodies are home to a vast assortment of microbes, including bacteria, fungi, and viruses. These microbes live within their human hosts, interacting with each other and influencing states of health and disease. Despite their prevalence and importance, studying host-microbe interactions has been limited by the dearth of appropriate tools and approaches, and an underappreciation for the role of biophysics.

This thesis describes the development and application of novel tools and approaches for studying bacteria, fungi, and viruses to uncover their potential roles in human health and disease.

In my first project, we investigated bacterial aggregation, a phenomenon related to important host-microbe interactions such as biofilm formation and the clearance of pathogens from the gastrointestinal tract. We found that bacteria aggregate in the presence of polymers (such as dietary fiber) via a mechanism that is qualitatively consistent with depletion-type forces under gut-like conditions. Surprisingly, motile bacteria aggregate more than nonmotile bacteria in viscous, high-polymer concentrations due to the higher effective diffusivity and inter-bacterial collisions enabled by motility. These two results give insight on how the foods (such as fiber) that we consume can physically affect the structure of microbes and other matter in the gut.

In my next projects, we investigated viral-load kinetics to understand the best testing modality for early detection of SARS-CoV-2 via a large community-based household transmission study. By collecting longitudinal, paired saliva and nasal-swab specimens from SARS-CoV-2 patients starting from the incident of infection, we quantified the viral-load trajectories of COVID-19-positive participants in each specimen type over time. Our results revealed that viral loads increased quickly and reached a higher peak in nasal-swab specimens, whereas viral loads were detectable earlier but reached a lower maximum in saliva. Both specimen types exhibited a temporal trend whereby viral loads were higher in specimens collected in the morning compared with the evening. In samples where infectious viral titer was measured, we found that the ratio of *N* gene viral load and infectious viral titer

did not remain consistent throughout the course of infection. These three results help us understand the heterogeneity of SARS-CoV-2 disease progression in different individuals, and how the analytical sensitivity of a diagnostic, the specimen type, and time of sampling can be crucial in conducting community surveillance programs during a pandemic.

Finally, we extended and co-validated for fungi a novel sample-preparation method that enriches fungal cells in host-rich samples to enable the first demonstration of deep metagenomic sequencing of fungal communities directly from clinical samples (without a culture step). Our results show that this method depletes host DNA by over 1000-fold by mass, improving taxonomic classification and gene calling, as well as enabling *de novo* metagenome assembled genome (MAG) assembly in samples dominated by human biomass.

PUBLISHED CONTENT AND CONTRIBUTIONS

1. Porter, M.K., Steinberg, A.P. and Ismagilov, R.F., 2019. Interplay of motility and polymer-driven depletion forces in the initial stages of bacterial aggregation. *Soft Matter*, *15*(35), pp. 7071–7079.

   This article investigates the interactions between polymers and *E. coli*, specifically on the role of bacterial motility in polymer-driven depletion aggregation. My role was as follows (text reproduced from page 15 of this thesis):

   Designed experiments, interpreted results, and analyzed all data. Optimized aggregation measuring experiment for bacteria. Performed experiments to generate all date for figures 2.2, 2.3, 2.4, 2.S1, and 2.S2. Drew schematic in figure 2.1 and obtained images for figure 2.1. Performed numerical calculations for figure 2.3 using code adapted from APS Contributed to writing the abstract, introduction, results/discussion, and conclusion sections of the manuscript. Contributed to writing of supplemental information.

2. Savela, E.S.*, Viloria Winnett, A.*, Romano, A.E.*, Porter, M.K., Shelby, N., Akana, R., Ji, J., Cooper, M.M., Schlenker, N.W., Reyes, J.A. and Carter, A.M., 2022. Quantitative SARS-CoV-2 viral-load curves in paired saliva samples and nasal swabs inform appropriate respiratory sampling site and analytical test sensitivity required for earliest viral detection. *Journal of clinical microbiology*, *60*(2), pp. e01785–21.

   This article investigates the viral load trajectories of SARS-CoV-2 in paired saliva and nasal swab specimens. My role was as follows (text reproduced from page 46 of this thesis):

   Performed specimen logging and QC, RNA extractions, RT-qPCR, data processing. Performed data acquisition and analysis for and made Figure 3.S1 with AW. Prepared participant sample collection materials and helped with supplies acquisition. Assisted in literature analysis with ES, RA, AW.

3. Viloria Winnett, A.*, Porter, M.K.*, Romano, A.E., Savela, E.S., Akana, R., Shelby, N., Reyes, J.A., Schlenker, N.W., Cooper, M.M., Carter, A.M. and Ji, J., 2022. Morning SARS-CoV-2 testing yields better detection of infection due to higher viral loads in saliva and nasal swabs upon waking. *Microbiology Spectrum*, *10*(6), pp. e03873–22.

   This article investigates the difference in SARS-CoV-2 viral loads between specimens collected in the morning versus those collected in the evening. My role was as follows (text reproduced from page 111 of this thesis):

   Performed specimen logging and QC, RNA extractions, RT-qPCR, data processing. Performed data acquisition and analysis for and made Figure 4.S2 with AVW. Prepared

participant sample collection materials and helped with supplies acquisition. Assisted in literature analysis with ES, RA, AVW. Performed data analysis and prepared Figures 4.1B, 4.1D, with AVW. Performed data analysis and prepared Figures 4.2B, 4.2D with AVW. Verified the underlying data with AVW and NS. Outlined and wrote manuscript with SVW.

4. Porter, M.K., Viloria Winnett, A., Hao, L., Shelby, N., Reyes, J.A., Schlenker, N.W., Romano, A.E., Tognazzini, C., Feaster, M., Goh, Y.Y. and Gale, Jr, M., 2023. The ratio between SARS-CoV-2 RNA viral load and culturable viral titer differs depending on stage of infection. *Access Microbiology*, pp. 000732–v1.

   This article investigates the infectious viral titer of SARS-CoV-2 in longitudinally collected nasal swab specimens. My role was as follows (text reproduced from page 138 of this thesis):

   Conceptualization of nested study with AVW and RFI. Literature review with AVW. Co-wrote biosafety protocols for receipt and processing of live specimens with AVW. Reagent and supply acquisition. Determined and validated live specimen collection methods with AVW. Received, logged and aliquoted live specimens with AVW. Established academic collaboration with colleagues at University of Washington. Performed nucleic acid extractions and RNaseP, N, E, and RdRP gene RT-qPCR. Coordinated and prepared shipments of live specimens to collaborators. Interpreted paired viral load and culture data with AVW and LH. Outlined and co-wrote manuscript with AVW. Prepared Figures 5.1, 5.S1, 5.S2, and 5.S3. Validated underlying data with AVW. Managed author citations and references.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Number *Page*

Chapter 1

INTRODUCTION

Bacteria, fungi, and viruses colonize most of the human body, including the oral cavity (1), gastrointestinal (GI) tract (2), skin (3), vagina (4), and the respiratory system (5). Microbes are implicated in a wide range of health outcomes, from maintaining homeostasis to contributing to the pathogenesis of many diseases (6, 7). Yet, despite their ubiquity and importance to human health, the roles microbes play in human health, including biophysical and pathological, are still largely unknown (8). The motivation behind this thesis is to study host-microbe interactions and to develop novel methods and approaches to gain insights into microbes' contributions to human health and disease.

Bacterial Aggregation in the Gut

The intestines are home to more than a trillion commensal bacteria, living within the complex and diverse environment of the gastrointestinal (GI) mucosa (9). Many of the functions carried out by these bacteria are mediated by their aggregation behavior. One such mechanism is through quorum sensing, a regulatory pathway that is modulated by the local bacterial cell density (10). Quorum sensing has been shown to regulate virulence (10, 11), biofilm formation (12, 13), and metabolism (14, 15), depending on the biological context. Therefore, one step towards elucidating the roles of bacteria in gut pathogenicity and homeostasis includes understanding the biophysical mechanisms by which bacteria aggregate in the gut.

Polymer-driven Bacterial Aggregation.

Gut bacteria live among a wide variety of macromolecules and polymers, such as those from host cell secretions (e.g., mucus, immunoglobulins) and food (e.g., dietary fibers) (16). Host cell secretions in the GI tract primarily act to protect the host from invading pathogens (17).

For example, the mucus layer contains secreted mucin glycoproteins used to protect the host epithelium (18, 19). In addition to mucins, secretory Immunoglobulin A (sIgA) is another prominent intestinal secretion involved in host defense (20).

The GI tract also contains polymers in the form of dietary fibers, which are mostly from plant material. Dietary fibers cannot be digested by human enzymes, so they reach the large intestine relatively intact and are metabolized there by resident microbes. Dietary fibers have a wide range of physicochemical properties, including different sizes, molecular weights, degrees of solubility, and gel-forming capabilities (16).

Generally, polymers can aggregate bacteria via two mechanisms: chemically mediated or through physical forces. One example of chemically mediated aggregation is the binding of mucin glycoproteins to *Lactobacillus plantarum* in the intestines (21). Similarly, in the presence of pathogenic bacteria, sIgA has been shown to bind to the O-antigen component of lipopolysaccharides such as those found on *Salmonella typhimurium* (20).

Previous work from our lab demonstrated that particulate matter in the gut may aggregate through a physical mechanism consistent with depletion forces (16). Depletion forces arise when particles are suspended in polymer solutions, and polymers avoid the surface of the particles within length scales on the order of the polymer size to minimize entropic penalties. This creates a volume around the particles that is lower in polymer concentration compared to the bulk solution, called the depletion zone. Depletion forces arise when two particles approach each other and their depletion zones overlap (called the overlap region), creating an osmotic pressure difference between the overlap region and the bulk solution. This pressure differential pulls the particles together, causing aggregation (22–26).

Unlike chemically mediated forces, which rely on specific chemical epitopes to bind to bacteria, physical forces such as depletion are agnostic to surface chemistry. Instead, depletion forces are a function of the physicochemical properties of the polymer, such as its concentration, molecular weight, and size (22). In our lab's previous work, it was

demonstrated that particulate matter can aggregate in the presence of soluble, high molecular weight dietary fibers in a gut-like environment, through a mechanism qualitatively consistent with polymer-driven depletion aggregation (16).

Motility and Bacterial Aggregation

The gut is a dynamic environment. In addition to polymer-driven aggregation forces, bacteria in the gut are also subjected to forces generated from their own motility (27, 28). Bacteria swim using the rotation of flagella, a slender projection outside of the bacterium that rotates to generate sufficient torque and drag on a bacterium's body (27).

Flagella has been linked to host epithelial cell binding and pathogenicity. For example, enteropathogenic *E. coli* utilize a type III protein secretory complex during flagellar assembly which mobilizes virulence factors from within the bacterium across its membrane (29). As flagella are implicated in virulence, environmental pressures from the gut microbiota select for non-flagellated, nonmotile commensal bacteria (30). Flagellar motility can be used as a distinguishing characteristic between pathogenic and commensal bacteria in the gut; Therefore, it is of biological interest to distinguish between their behaviors in polymer-driven aggregation.

Very little is understood about how gut polymers affect motile bacteria and their early aggregation behavior at time scales relevant to the gut. In Chapter I of this thesis, I investigate the polymer-driven aggregation of bacteria through the depletion mechanism using a polymer physics framework. Furthermore, I investigate the interplay of these polymer-driven depletion forces with bacterial motility to understand how swimming forces and depletion forces affect bacterial aggregation.

SARS-CoV-2 Viral Load Kinetics to Educate COVID-19 Public Healthy Strategy.

In March 2020, the World Health Organization declared COVID-19 a global health emergency, marking the official start of a pandemic that had already infected millions

worldwide. By the end of the year, the virus had claimed over a million lives (31). Shortly thereafter, various states in the US began implementing stay-at-home orders to prevent the spread of SARS-CoV-2, shutting down schools and businesses that were not deemed essential (32).

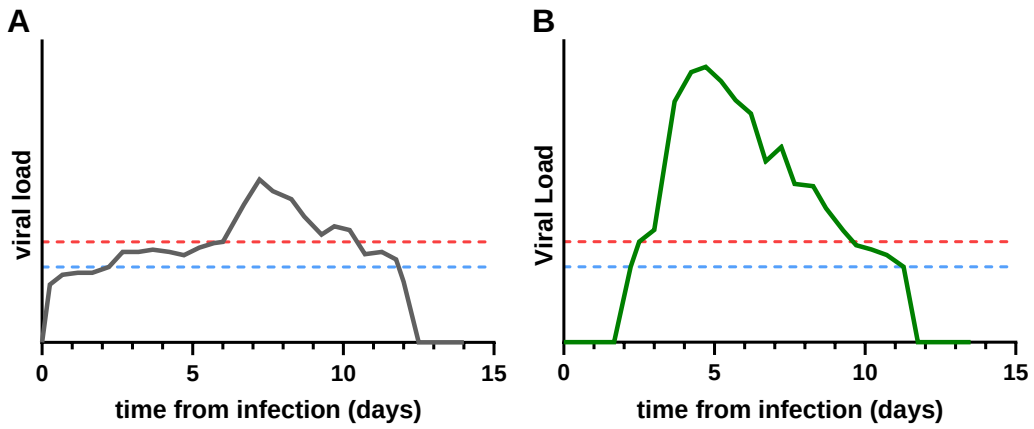


Figure 1.1: Two hypothetical scenarios resulting from different viral-load trajectories. Hypothetical viral-load trajectories for two different specimen types (A, gray and B, green) show the rise and fall of viral loads over the course of an infection. Horizontal lines indicate the hypothetical limits of detection (LODs) for diagnostic tests with high (blue) or low (red) analytical sensitivity.

Due to the significant strain the COVID-19 pandemic was placing on the healthcare system and the economy, community testing was initiated to help mitigate the spread of disease. This testing typically utilized nucleic acid amplification-based diagnostic platforms. Nucleic acid amplification tests (NAATs) use quantitative polymerase chain reaction (qPCR) to measure the amount of SARS-CoV-2 RNA (viral load) in a person's specimen sample (33). An effective surveillance testing program requires implementing a diagnostic platform that that has sufficient analytical sensitivity, quantified by the limit of detection (LOD; defined as the lowest amount of analyte that can be detected with statistical certainty), to detect disease early in infection and thus enable timely quarantine and treatment (34-36).

However, test sensitivity is not sufficient; to be effective, the test must be used on the specimen type (e.g., saliva or nasal swab) containing enough viral particles to be detected. Selection of the best specimen type requires knowledge of the viral load trajectories in different anatomical locations over time.

The importance of implementing a diagnostic with the appropriate LOD can be illustrated in two hypothetical viral load trajectories (Figure 1.1).

1. For example, if a patient's viral load trajectory for a given specimen type stays low upon incident of infection (Figure 1A), a test with poor analytical sensitivity (red dotted line) would not yield a positive result until day 6. In contrast, a test with good analytical sensitivity (blue dotted line) would yield a positive result on day 2.

2. However, if a patient's viral load trajectory increases quickly, (Figure 1B), then using a test with a good analytical sensitivity does not yield a much faster result compared to a test with poor analytical sensitivity (the infection would be detected just one-half day later using a test with poor sensitivity).

In the case of the viral load trajectory illustrated in Figure 1A, a more sensitive test (low LOD) would be needed to be able to detect SARS-CoV-2 RNA early in infection. If instead the SARS-CoV-2 viral load trajectory resembled the one illustrated in Figure 1B, where viral load rises rapidly, a test with low sensitivity (high LOD) would achieve similar results to a high sensitivity test. An important consideration is that different specimen types in the same individual may exhibit completely different viral-load trajectories.

In Chapters II and III, we quantify and characterize the *N* gene viral load kinetics of participants infected with SARS-CoV-2 starting from the incidence of infection. From these datasets, we infer sensitivities of hypothetical diagnostics at various LODs to underscore the importance of considering both viral load trajectories and specimen type when implementing a surveillance testing program.

Finally, although measuring viral RNA in a specimen can tell us if the patient has been recently infected by SARS-CoV-2, it does not indicate whether the patient is infectious at that point in time. Infectiousness is often determined by the presence of replicable virus, measured by viral titer. Unfortunately, measuring viral titer can take up to several days and requires specialized biosafety considerations, whereas an NAAT can yield results within a few hours (37). Establishing the correlative relationship between viral RNA load and infectious titer could help provide a guideline for determining the infectivity of a patient simply by quantifying viral load (without undergoing a lengthy viral titer measurement).
In Chapter IV, we measure the viral titer of two participants from one household to understand the relationship between viral RNA load and infectious viral titer.

Extending the Microbial Enrichment Method (MEM) for Metagenomic Characterization of Fungi in Clinical Specimens

The human mycobiome is an often overlooked but highly consequential component of human health (38, 39). Fungal pathogens and microbial dysbiosis have been identified ass the direct causative agent in conditions including aspergillosis lung infections (40), vaginal candidiasis (41), *Candida auris* infections (42), and invasive candidiasis in the gut (43), and highly implicated in conditions such as inflammation (44) and cancer (45).

Mycobiome studies often use sequencing methods, such amplicon or shotgun sequencing to detect and identify fungi (46–52). Although amplicon sequencing is highly sensitive, it can only target a single gene if the sequence is known *a priori*; it cannot reveal functional information regarding the genes present in the fungal genome (53). In contrast with amplicon sequencing, shotgun sequencing involves sequencing all DNA in a sample. This approach reveals a broader range of genetic information beyond just taxonomic classification, including genes of interest (54).

However, a significant challenge in using shotgun sequencing to studying microbial

communities in clinical samples is the over-abundance of host DNA. Because shotgun sequencing sequences all DNA fragments in a sample, a substantial portion of the sequencing reads are allocated to host DNA, leaving few, if any, for fungi. Host reads can comprise anywhere between 85-99.9% of the total DNA in host-rich clinical samples such as vaginal swabs (55), saliva (56), bronchoalveolar lavage (57), and colonic biopsy (54).

A previous study in our lab developed a novel microbial enrichment method (MEM) for improving the relative abundance of bacterial DNA in host-rich clinical samples, such as colonic biopsies (54). By reducing the total host DNA mass in a sample by over 1000x, MEM enabled deeper shotgun metagenomics and the construction of metagenome assembled genomes (MAGs) to for strain level identification of the bacteria that reside in the colon (54).

In Chapter VI of this thesis, we extend and validate MEM to show that it can be used to enrich the relative abundance of fungi in host-rich clinical samples, such as bronchoalveolar lavage and vaginal swabs. Using this method, we show the first fungal MAGs that were assembled directly from host-rich human samples.

References

1. Baker, J.L., Mark Welch, J.L., Kauffman, K.M., McLean, J.S. and He, X., 2024. The oral microbiome: diversity, biogeography and human health. *Nature Reviews Microbiology*, *22*(2), pp. 89–104.

2. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P. and Heath, A.C., 2012. Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), pp. 222–227.

3. Byrd, A.L., Belkaid, Y. and Segre, J.A., 2018. The human skin microbiome. *Nature Reviews Microbiology*, *16*(3), pp. 143–155.

4. Gao, X., Lee, V., Prom-Wormley, E., Girerd, P., Mo, H., Xie, B., Harwich, M., Neale, M., York, T., Hendricks, S. and Puma, F., 2011. The vaginal microbiome: disease, genetics and the environment. *Nature Precedings*, pp. 1–1.

5. Wypych, T.P., Wickramasinghe, L.C. and Marsland, B.J., 2019. The influence of the microbiome on respiratory health. *Nature immunology*, *20*(10), pp. 1279–1290.

6. Lee, S.M., Donaldson, G.P., Mikulski, Z., Boyajian, S., Ley, K. and Mazmanian, S.K., 2013. Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature*, *501*(7467), pp. 426–429.

7. Das, B. and Nair, G.B., 2019. Homeostasis and dysbiosis of the gut microbiome in health and disease. *Journal of biosciences*, *44*, pp. 1–8.

8. Qian, G. and Ho, J.W., 2020. Challenges and emerging systems biology approaches to discover how the human gut microbiome impact host physiology. *Biophysical Reviews*, *12*(4), pp. 851–863.

9. Round, J.L. and Mazmanian, S.K., 2009. The gut microbiota shapes intestinal immune responses during health and disease. *Nature reviews immunology*, *9*(5), pp. 313–323.

10. Antunes, L.C.M., Ferreira, R.B., Buckner, M.M. and Finlay, B.B., 2010. Quorum sensing in bacterial virulence. *Microbiology*, *156*(8), pp. 2271–2282.

11. Chadha, J., Harjai, K. and Chhibber, S., 2022. Revisiting the virulence hallmarks of Pseudomonas aeruginosa: a chronicle through the perspective of quorum sensing. *Environmental microbiology*, *24*(6), pp. 2630–2656.

12. Hammer, B.K. and Bassler, B.L., 2003. Quorum sensing controls biofilm formation in Vibrio cholerae. *Molecular microbiology*, *50*(1), pp. 101–104.

13. Pezzoni, M., Pizarro, R.A. and Costa, C.S., 2020. Role of quorum sensing in UVA-induced biofilm formation in Pseudomonas aeruginosa. *Microbiology*, *166*(8), pp. 735–750.

14. An, J.H., Goo, E., Kim, H., Seo, Y.S. and Hwang, I., 2014. Bacterial quorum sensing and metabolic slowing in a cooperative population. *Proceedings of the National Academy of Sciences*, *111*(41), pp. 14912–14917.

15. Goo, E., An, J.H., Kang, Y. and Hwang, I., 2015. Control of bacterial metabolism by quorum sensing. *Trends in microbiology*, *23*(9), pp. 567–576.

16. Preska Steinberg, A., Datta, S.S., Naragon, T., Rolando, J.C., Bogatyrev, S.R. and Ismagilov, R.F., 2019. High-molecular-weight polymers from dietary fiber drive aggregation of particulates in the murine small intestine. *Elife*, *8*, p.e40387.

17. Brandtzaeg, P., 2011. The gut as communicator between environment and host: immunological consequences. *European journal of pharmacology*, *668*, pp. S16–S32.

18. Johansson, M.E., Larsson, J.M.H. and Hansson, G.C., 2011. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host–microbial interactions. *Proceedings of the national academy of sciences*, *108*(supplement_1), pp. 4659–4665.

19. Johansson, M.E., Phillipson, M., Petersson, J., Velcich, A., Holm, L. and Hansson, G.C., 2008. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proceedings of the national academy of sciences*, *105*(39), pp. 15064–15069.

20. Moor, K., Diard, M., Sellin, M.E., Felmy, B., Wotzka, S.Y., Toska, A., Bakkeren, E., Arnoldini, M., Bansept, F., Co, A.D. and Völler, T., 2017. High-avidity IgA protects the intestine by enchaining growing bacteria. *Nature*, *544*(7651), pp. 498–502.

21. Sun, J., Le, G.W., Shi, Y.H. and Su, G.W., 2007. Factors involved in binding of Lactobacillus plantarum Lp6 to rat small intestinal mucus. *Letters in applied microbiology*, *44*(1), pp. 79–85.

22. Asakura, S. and Oosawa, F., 1954. On interaction between two bodies immersed in a solution of macromolecules. *The Journal of chemical physics*, *22*(7), pp. 1255–1256.

23. Vrij, A., 1976. Polymers at interfaces and the interactions in colloidal dispersions. *Pure and Applied Chemistry*, *48*(4), pp. 471–483.

24. Gast, A.P., Hall, C.K. and Russel, W.B., 1983. Polymer-induced phase separations in nonaqueous colloidal suspensions. *Journal of Colloid and Interface Science*, *96*(1), pp. 251–267.

25. Prasad, V., 2002. *Weakly interacting colloid polymer mixtures*. Harvard University.

26. Ilett, S.M., Orrock, A., Poon, W.C.K. and Pusey, P.N., 1995. Phase behavior of a model colloid-polymer mixture. *Physical Review E*, *51*(2), p. 1344.

27. Gabel, C.V. and Berg, H.C., 2003. The speed of the flagellar rotary motor of Escherichia coli varies linearly with protonmotive force. *Proceedings of the National Academy of Sciences*, *100*(15), pp. 8748–8751.

28. Schwarz-Linek, J., Winkler, A., Wilson, L.G., Pham, N.T., Schilling, T. and Poon, W.C., 2010. Polymer-induced phase separation in Escherichia coli suspensions. *Soft Matter*, *6*(18), pp. 4540–4549.

29. Yip, C.K., Finlay, B.B. and Strynadka, N.C., 2005. Structural characterization of a type III secretion system filament protein in complex with its chaperone. *Nature structural & molecular biology*, *12*(1), pp. 75–81.

30. Sevrin, G., Massier, S., Chassaing, B., Agus, A., Delmas, J., Denizot, J., Billard, E. and Barnich, N., 2020. Adaptation of adherent-invasive E. coli to gut environment: Impact on flagellum expression and bacterial colonization ability. *Gut Microbes*, *11*(3), pp. 364–380.

31. The true death toll of COVID-19 estimating global excess mortality (2021) World Health Organization. Available at: https://www.who.int/data/stories/the-true-death-toll-of-covid-19-estimating-global-excess-mortality (Accessed: 28 May 2024).

32. CDC Museum Covid-19 Timeline (2023) Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/museum/timeline/covid19.html (Accessed: 28 May 2024).

33. Lai, C.K. and Lam, W., 2021. Laboratory testing for the diagnosis of COVID-19. *Biochemical and biophysical research communications*, *538*, pp. 226–230.

34. Afzal, A., 2020. Molecular diagnostic technologies for COVID-19: Limitations and challenges. *Journal of advanced research*, *26*, pp. 149–159.

35. Viloria Winnett, A., Akana, R., Shelby, N., Davich, H., Caldera, S., Yamada, T., Reyna, J.R.B., Romano, A.E., Carter, A.M., Kim, M.K. and Thomson, M., 2023. Daily SARS-CoV-2 nasal antigen tests miss infected and presumably infectious people due to viral load differences among specimen types. *Microbiology Spectrum*, *11*(4), pp. e01295-23.

36. Viloria Winnett, A., Akana, R., Shelby, N., Davich, H., Caldera, S., Yamada, T., Reyna, J.R.B., Romano, A.E., Carter, A.M., Kim, M.K. and Thomson, M., 2023.

Extreme differences in SARS-CoV-2 viral loads among respiratory specimen types during presumed pre-infectious and infectious periods. *PNAS nexus*, *2*(3), p.pgad033.

37. Vandergaast, R., Carey, T., Reiter, S., Lathrum, C., Lech, P., Gnanadurai, C., Haselton, M., Buehler, J., Narjari, R., Schnebeck, L. and Roesler, A., 2021. IMMUNO-COV v2. 0: Development and validation of a high-throughput clinical assay for measuring SARS-CoV-2-neutralizing antibody titers. *MSphere*, *6*(3), pp. 10–1128.

38. Huseyin, C.E., O'Toole, P.W., Cotter, P.D. and Scanlan, P.D., 2017. Forgotten fungi—the gut mycobiome in human health and disease. *FEMS Microbiology Reviews*, *41*(4), pp. 479–511.

39. Seed, P.C., 2015. The human mycobiome. *Cold Spring Harbor perspectives in medicine*, *5*(5), p. a019810.

40. Palmieri, F., Koutsokera, A., Bernasconi, E., Junier, P., von Garnier, C. and Ubags, N., 2022. Recent Advances in Fungal Infections: From Lung Ecology to Therapeutic Strategies With a Focus on Aspergillus spp. *Frontiers in Medicine*, *9*, p.832510.

41. Sobel, J.D., 2016. Recurrent vulvovaginal candidiasis. *American journal of obstetrics and gynecology*, *214*(1), pp. 15–21.

42. Cortegiani, A., Misseri, G., Fasciana, T., Giammanco, A., Giarratano, A. and Chowdhary, A., 2018. Epidemiology, clinical characteristics, resistance, and treatment of infections by Candida auris. *Journal of intensive care*, *6*, pp. 1–13.

43. Kumamoto, C.A., Gresnigt, M.S. and Hube, B., 2020. The gut, the bad and the harmless: Candida albicans as a commensal and opportunistic pathogen in the intestine. *Current opinion in microbiology*, *56*, pp. 7–15.

44. Ott, S.J., Kühbacher, T., Musfeldt, M., Rosenstiel, P., Hellmig, S., Rehman, A., Drews, O., Weichert, W., Timmis, K.N. and Schreiber, S., 2008. Fungi and

inflammatory bowel diseases: alterations of composition and diversity. *Scandinavian journal of gastroenterology*, *43*(7), pp. 831–841.

45. Galloway-Peña, J., Iliev, I.D. and McAllister, F., 2024. Fungi in cancer. *Nature Reviews Cancer*, pp. 1–4.

46. Gomez, C.A., Budvytiene, I., Zemek, A.J. and Banaei, N., 2017. Performance of targeted fungal sequencing for culture-independent diagnosis of invasive fungal disease. *Clinical Infectious Diseases*, *65*(12), pp. 2035–2041.

47. Lau, A., Chen, S., Sorrell, T., Carter, D., Malik, R., Martin, P. and Halliday, C., 2007. Development and clinical application of a panfungal PCR assay to detect and identify fungal DNA in tissue specimens. *Journal of clinical microbiology*, *45*(2), pp. 380–385.

48. Larkin, P.M., Lawson, K.L., Contreras, D.A., Le, C.Q., Trejo, M., Realegeno, S., Hilt, E.E., Chandrasekaran, S., Garner, O.B., Fishbein, G.A. and Yang, S., 2020. Amplicon-based next-generation sequencing for detection of fungi in formalin-fixed, paraffin-embedded tissues: Correlation with histopathology and clinical applications. *The Journal of Molecular Diagnostics*, *22*(10), pp. 1287–1293.

49. Usyk, M., Peters, B.A., Karthikeyan, S., McDonald, D., Sollecito, C.C., Vazquez-Baeza, Y., Shaffer, J.P., Gellman, M.D., Talavera, G.A., Daviglus, M.L. and Thyagarajan, B., 2023. Comprehensive evaluation of shotgun metagenomics, amplicon sequencing, and harmonization of these platforms for epidemiological studies. *Cell reports methods*, *3*(1).

50. Ali, N.A.T.B.M., Mac Aogáin, M., Morales, R.F., Tiew, P.Y. and Chotirmall, S.H., 2019. Optimisation and benchmarking of targeted amplicon sequencing for mycobiome analysis of respiratory specimens. *International journal of molecular sciences*, *20*(20), p. 4991.

51. Qi, C., Hountras, P., Pickens, C.O., Walter, J.M., Kruser, J.M., Singer, B.D., Seed, P., Green, S.J. and Wunderink, R.G., 2019. Detection of respiratory pathogens in

clinical samples using metagenomic shotgun sequencing. *Journal of medical microbiology*, *68*(7), pp. 996–1002.

52. Wensel, C.R., Pluznick, J.L., Salzberg, S.L. and Sears, C.L., 2022. Next-generation sequencing: insights to advance clinical investigations of the microbiome. *The Journal of clinical investigation*, *132*(7).

53. Nilsson, R.H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P. and Tedersoo, L., 2019. Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews Microbiology*, *17*(2), pp. 95–109.

54. Wu-Woods, N.J., Barlow, J.T., Trigodet, F., Shaw, D.G., Romano, A.E., Jabri, B., Eren, A.M. and Ismagilov, R.F., 2023. Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples. *Nature Methods*, *20*(11), pp. 1672–1682.

55. Ruiz-Perez, D., Coudray, M.S., Colbert, B., Krupp, K., Kumari, H., Stebliankin, V., Mathee, K., Cook, R.L., Schwebke, J., Narasimhan, G. and Madhivanan, P., 2021. Effect of metronidazole on vaginal microbiota associated with asymptomatic bacterial vaginosis. *Access microbiology*, *3*(5), p.000226.

56. Marotz, C.A., Sanders, J.G., Zuniga, C., Zaramela, L.S., Knight, R. and Zengler, K., 2018. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome*, *6*, pp. 1–9.

57. Leo, S., Gaïa, N., Ruppé, E., Emonet, S., Girard, M., Lazarevic, V. and Schrenzel, J., 2017. Detection of bacterial pathogens from broncho-alveolar lavage by next-generation sequencing. *International journal of molecular sciences*, *18*(9), p.2011.

Chapter 2

INTERPLAY OF MOTILITY AND POLYMER-DRIVEN DEPLETION FORCES IN THE INITIAL STAGES OF BACTERIAL AGGREGATION

1. Porter, M.K., Steinberg, A.P. and Ismagilov, R.F., 2019. Interplay of motility and polymer-driven depletion forces in the initial stages of bacterial aggregation. *Soft Matter*, *15*(35), pp. 7071–7079.

Contributions of non-corresponding authors

M.P.

1. Designed experiments, interpreted results, and analyzed data
2. Optimized aggregation measuring experiment for bacteria
3. Performed experiments to generate all data for Figures 2, 3, 4, S1, and S2
4. Drew schematic in figure 1 and obtained images for Figure 1
5. Performed numerical calculations for Figure 3 using code adapted from A.P.S.
6. Contributed to writing abstract, introduction, results/discussion, and conclusion sections of the manuscript
7. Contributed to writing of supplemental information

A.P.S.

1. Contributed to experimental design, interpretation of results, and analysis of data.
2. Contributed to adaptation of image analysis pipeline for use in this system.
3. Co-developed numerical methods and contributed to adaptation of theory for use in Figure 3.
4. Contributed to writing abstract, introduction, results/discussion, and conclusion sections of the manuscript.
5. Contributed to writing of supplemental information.

Abstract

Motile bacteria are often found in complex, polymer-rich environments in which microbes can aggregate via polymer-induced depletion forces. Bacterial aggregation has many biological implications; it can promote biofilm formation, upregulate virulence factors, and lead to quorum sensing. The steady state aggregation behavior of motile bacteria in polymer solutions has been well studied and shows that stronger depletion forces are required to aggregate motile bacteria as compared with their nonmotile analogs. However, no one has studied whether these same trends hold at the initial stages of aggregation. We use experiments and numerical calculations to investigate the polymer-induced depletion aggregation of motile Escherichia coli in polyethylene glycol solutions on short experimental timescales (~10 min). Our work reveals that in the semi-dilute polymer concentration regime and at short timescales, in contrast to what is found at steady state, bacterial motility actually enhances aggregate formation by increasing the collision rate in viscous environments. These unexpected findings have implications for developing models of active matter, and for understanding bacterial aggregation in dynamic, biological environments, where the system may never reach steady state.

Introduction

Bacteria thrive in a wide range of biological and ecological contexts and play important roles in the human gut, soil, wastewater sludge, and other complex environments. In these environments, bacterial motility has implications for biofilm formation and virulence. For example, in the gut, *Salmonella typhimurium* uses its flagella to burrow through the intestinal mucus layer and penetrate the host epithelium, causing infection (2). Certain species of *Pseudomonas aeruginosa* require motility to form biofilms, such as those found on medical devices (3). Microbial motility is further influenced by polymers, which are abundant in many environments (4). For example, in the gut, polymers are secreted by the host (5–8) and dietary fibers are ingested regularly (1). In wastewater treatment plants, sludge used to collect microbes and particulate waste also contains polymers (9, 10).

Polymers are well known to aggregate bacteria, which is important because aggregation precedes biofilm formation (11), correlates with altered gene expression (3, 12), including antibiotic resistance genes (13), and induces phenotypic changes such as quorum sensing (14). Polymers can bind to bacteria via specific chemical interactions and cause aggregation through agglutination. These interactions are found in biological settings such as the gut, where mucins (5), immunoglobulins (6), and other host-secreted proteins (7) can aggregate bacteria via chemically mediated interactions. Polymers can also aggregate microbes via depletion forces. This mechanism of aggregation does not depend on microbes binding to specific chemical functional groups but is instead only a function of the physical parameters of the polymer (molecular weight (MW), hydrodynamic radius) and bacteria size (15–17). In the presence of non-adsorbing polymers such as polyethylene glycol (PEG), bacteria aggregate through depletion interactions, which occur when two bacteria approach each other at a close enough distance that the polymer is excluded from the space between the bacteria, a region called the depletion zone (13). The difference in polymer concentration between the depletion zone and the bulk solution results in an osmotic pressure difference that generates an attractive force between bacteria. Because depletion forces depend on the physical properties of polymers and bacteria, these forces can drive bacterial aggregation irrespective of bacterial surface chemistry. In the case of nonmotile bacteria (which are not auto-aggregating) in solutions of non-adsorbing polymers, the only driving forces for aggregation are polymer-induced depletion forces (17,18). However, for motile bacteria in solutions of non-adsorbing polymers, there are forces due to the motility of the bacteria and due to depletion forces; the competition between these forces determines the steady-state aggregation behavior (18). At sufficient polymer concentrations and long time scales, when the system reaches steady state, polymer-induced depletion attractions between bacteria can result in aggregation (i.e., phase separation) (18). The addition of motility has been found to require significantly stronger depletion attraction to achieve the aggregation as compared with nonmotile bacteria (18). Active matter is an area of intense research, and the field is currently working on a unified theoretical framework to understand these systems (19,20). In particular, the aggregation behavior of motile bacteria in polymeric solutions at long time

scales is widely studied (18,21–24); however, to our knowledge there are no published studies on aggregation at short time scales.

The initial stages of aggregate formation are of particular interest because microbial responses to aggregation (e.g., upregulation of quorum sensing and virulence pathways) occur on short time scales (tens of minutes) and would be influenced by the initial stages of aggregation. Furthermore, the initial stages of aggregation are particularly relevant for biological systems where the system may never reach steady state; for example, in the gastrointestinal (GI) tract where food and ingested material are constantly in transit (1, 25) In these environments, aggregate formation is constantly disrupted because of shear, peristaltic contractions, and other forces, making early aggregate formation relevant to these biological systems.

In this study, we investigate how motility influences the polymer-induced depletion aggregation of bacteria at short time scales ($t \sim 10$ min). We quantify this experimentally by using confocal fluorescence microscopy to measure the size distribution of bacterial aggregates in PEG solutions with molecular weights and concentrations relevant to the murine small intestine (1). Furthermore, we develop an understanding of which physical parameters influence the initial formation of these aggregates. We use a physical model for motile bacteria in PEG solutions that focuses on the balance of depletion and swim forces as well as the effective diffusivity of the bacteria.

Experimental

Bacteria cell culture

Overnight cultures of naturally motile *E. coli* K12 MG1655 (ATCC 47076) were prepared in liquid lysogeny broth (LB) culture incubated at 35 ˚C to mid-exponential phase. These cultures were combined to reach the desired cell concentration for the experiment (10^9 cells per mL). Cells were first centrifuged at 4.8 kG for 10 min and then resuspended in motility buffer (MB; 10 mM potassium phosphate buffer, 0.1 mM EDTA, pH 7.0) to stain with SYTO

9 (live, ex/em 480/500 nm). Following staining, cells were centrifuged again to wash out excess stain. To obtain nonmotile *E. coli*, cells were treated with 0.5% (75 mM) sodium azide in MB after washing.

Confocal microscope imaging and bacterial aggregation in polyethylene glycol (PEG).

All images and z-stacks were obtained using a Zeiss LSM 800 confocal fluorescence microscope (488 nm excitation, detection at 490–540 nm). Each stack was 200 200 45 mm in volume and contained about 135 slices. Nonmotile and motile *E. coli* K12 were prepared and stained using the method described above. PEG solutions were prepared by dissolving four times the overlap concentration into MB (for motile conditions) or by using MB with 0.5% sodium azide (for nonmotile condition). A range of 10 kDa, 100 kDa, and 1 MDa PEG solutions were achieved by serial dilution. A 5 mL aliquot of each respective PEG solution was combined with 0.5 mL *E. coli* for a final cell concentration of 10^9 cells per mL. We pipetted 2 mL of each combined suspension into an imaging chamber made from SecureSeal imaging spacer (Electron Microscopy Sciences; 0.12 mm depth and 9 mm diameter) and a glass slide, and the top of the chamber was immediately sealed with a #1.5 glass coverslip. A single z-stack of each PEG dilution sample was taken approximately 10 min after the imaging chamber was sealed. Each biological replicate was conducted with a new bacterial cell culture.

FIJI macro imaging and empirical bootstrapping

All imaging analysis was performed as previously described in ref. 1.

Measuring mean-squared displacement (MSD) of E. coli

E. coli K12 were cultured and prepared as described above. The final cell concentration was diluted to $5x10^8$ cells per mL when added to each PEG solution in MB. A Leica DMI6000 with a Visitech Infinity3 confocal microscope was used to obtain 20 s videos of the cells at about 16 frames per second. Videos were analyzed using an ImageJ plugin developed by the MOSAIC group for 2D/3D particle tracking using an algorithm developed in ref. 26. At least 1000 bacteria were analyzed per condition. Data output from the ImageJ plugin was further

analyzed and the MSD was calculated using MATLAB code used in ref. 27 in conjunction with an in-house script. (Script will be provided by request.)

Estimating overlap concentration for PEG

The polymer overlap concentration c_P^* was estimated using the following relation (26, 27):

$$c_P^* = \frac{MW}{\frac{4\pi}{3} N_{Avo} R_g(0)^3}$$

(Eq. 2.1)

where MW is the polymer molecular weight in kDa, N_{Avo} is Avogadro's number, and $R_g(0)$ is the radius of gyration given in m.

Estimating solution viscosity

We estimated the solution viscosity via a virial expansion (28):

$$\eta = \eta_s(1 + [\eta]c_P + k_H[\eta]^2 c_P^2 + \cdots)$$

(Eq. 2.2)

where η_s is the solvent viscosity in Pa s, c_P is the polymer mass concentration in kg m^{-3}, $[\eta]$ is the intrinsic viscosity measured to be 452.8 mL g^{-1} (using literature measurements for PEG 1 MDa),30 and k_H is the Huggins parameter for PEG 1 MDa, approximated to be 0.4. Using this equation, η at c_P = 6.5 mg mL^{-1} is ~7 mPa s compared with ~1 mPa s at c_P = 0.4 mg mL^{-1}. Literature measurements report similar values for the zero shear rate viscosity (η_0) of PEG 1 MDa at high concentrations, where η_0 = 10 mPa s at c_P = 5 mg mL^{-1} (our experiments were conducted our quiescent conditions, *i.e.*, no shear) (31).

Results and Discussion

Measuring E. coli aggregation at short time scales

To understand how motility affects the depletion-driven aggregation of bacteria at short time scales, we measured the initial formation of bacterial aggregates. As a model organism, we used *E. coli* K12 MG1655. This naturally motile strain of *E. coli* displays ‘‘run and tumble’’

dynamics, and can be rendered nonmotile by treating with 0.5 wt% sodium azide (32, 33). This method of rendering cells nonmotile has been used before in research focusing on the flagellar motility of *E. coli* in studies conducted by Christopher V. Gabel and Howard Berg, (32). Briefly, sodium azide acts as a respiratory poison that disrupts the cell's protonmotive force, which inhibits flagellar motility within 1-2 minutes (32). Plating showed that the azide treatment killed nearly all of the cells 1 h after the initial treatment (Figure 2.S1). To ensure we selected biologically relevant physical parameters for our experiments, we used data from our previous gel permeation chromatography experiments on luminal fluid from the murine small intestine to determine the range of the polymer molecular weights and concentrations (1). In these previous experiments, the polymers we found in the murine small intestine ranged in size from a few kDa to a few MDa (1). Therefore, we chose to tune the depletion potential in this work with polymers within this size range. As our test polymer, we chose to use PEG solutions in a motility buffer (MB, 0.1 mM EDTA, 10 mM potassium phosphate, pH 7.0). PEG is a linear, chemically inert polymer (33, 34) that is well characterized in inducing depletion forces in passive colloid solutions (35–42). We used a range of 1 MDa PEG concentrations (0.05–6.5 mg mL^{-1}) to adjust the depletion potential and rheology of the solution, which span both the dilute and semi-dilute polymer concentration regimes (the transition between these regimes is denoted by the overlap concentration, $c_P^* = 1.6$ mg mL^{-1} (see calculations in Experimental)). We also measured bacterial aggregation in 10 kDa and 100 kDa PEG near their respective overlap concentrations ($c_P^* = 8.5$ mg mL^{-1} and 85 mg mL^{-1} for 100 kDa and 10 kDa PEG, respectively). We previously detected polymers of a similar size and concentration in the murine small intestine (1). A motility buffer control was implemented to confirm that the cells were not auto-aggregating and that the aggregation measured in each sample containing PEG was the result of the PEG in the solution.

To quantify the initial aggregation of bacteria, we measured the volume-weighted average aggregate sizes (N) using fluorescence confocal microscopy (Figure 2.1a). After mixing the *E. coli* with the PEG, the bacterial suspension was placed into an imaging chamber, and sealed with a glass coverslip to eliminate drifting and evaporation effects. Z-stacks of cells in solutions of PEG at various concentrations were obtained after 10 min to focus on the

behavior on short timescales (Figure 2.1b and c). Separate experiments were performed for motile and nonmotile cells. Imaging at short timescales also reduces the effects of sedimentation from gravity (43). Imaging analysis was performed (using an ImageJ pipeline that we developed previously (1)) to count each object, measure the volumes of each aggregate, and normalize by the singlet volume to obtain the volume-weighted average aggregate size.

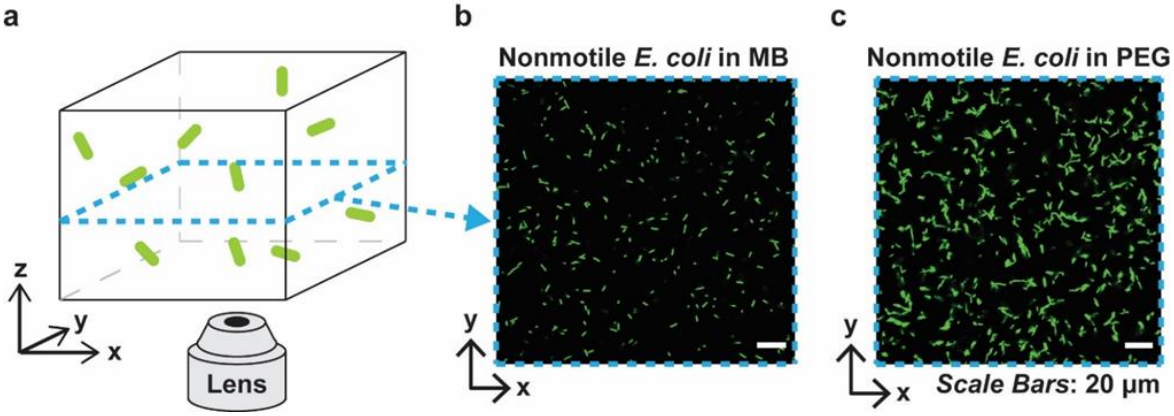


Figure 2.1: Measuring the volume-weighted average size of bacterial aggregates in PEG solution. (a) Cartoon depicting experimental setup of motile or nonmotile *E. coli* (green) in buffer or PEG. (b) Representative slices of nonmotile *E. coli* taken in motility buffer with no aggregation and in (c) 0.8 mg mL^{-1} PEG showing aggregation.

In these experiments, we found that nonmotile bacterial aggregation in the presence of 1 MDa PEG (Figure 2.2a) was qualitatively consistent with depletion-driven aggregation with similar trends observed at half (Figure 2.2b) and double (Figure 2.2c) the bacterial concentration. Additionally, we tested the effect of changing PEG MW and found trends that were qualitatively consistent with depletion-driven aggregation (Figure 2.S2); the extent of aggregation generally decreased with MW. Nonmotile *E. coli* aggregated less in the presence of PEG 100 kDa (Figure 2.S2b) as compared to in the presence of PEG 1 MDa (Figure 2.2a) and no aggregation was observed in the presence of PEG 10 kDa. Somewhat counterintuitively, as the 1 MDa PEG concentration increased, aggregate size increased up

to a limit and then started decreasing. We previously measured similar aggregation profiles with particles in PEG solutions (1); the shape of those profiles was at least in part attributed to the increase in solution viscosity, which hindered the Brownian motion of the particles, limiting the inter-particle collisions that initiate aggregation. We suspect that a similar mechanism may be at play for the nonmotile bacteria in this study. The depletion attractions increase as a function of polymer concentration due to the contribution of osmotic pressure, which is mirrored by the increasing aggregate size of the nonmotile bacteria in the dilute PEG concentration regime (Eq. (2.3) and (2.4)). We observed that as the PEG concentration increased and approached the semi-dilute concentration regime, the aggregate sizes became larger. In the semi-dilute regime, the depletion attractions continued to increase with the increase in PEG concentration, but the range of these attractions decreased with PEG concentration. At PEG concentrations far above the overlap concentration, we observed a decrease in aggregate sizes. The use of sodium azide to render E. coli nonmotile could alter cell surface properties in a manner that affects their depletion-driven aggregation. However, due to their lack of aggregation in the motility buffer control and the aggregation curve being qualitatively similar to bioinert particles in similar polymer solutions (1), we assume that this effect is minimal. Similar aggregation behavior was measured using a nonmotile mutant of E. coli (without the use of sodium azide), further suggesting that the aggregation measured in the nonmotile conditions is minimally affected by azide treatment (Figure 3.S3). We hypothesized that at higher PEG concentrations, the enhanced viscosity decreased inter-bacterial collisions, thus hindering aggregate formation.

Motile bacteria demonstrated different aggregation trends compared with their nonmotile analogs. In the PEG 1 MDa solutions, we observed no aggregation in the dilute PEG concentration regime, but at c_P = 0.8 mg mL^{-1}, motile *E. coli* abruptly began to aggregate, and continued to aggregate through the semidilute polymer concentration regime (Figure 2.2a). Similar trends were observed when we halved (Figure 2.2b) or doubled (Figure 2.2c) the bacterial concentrations. In the lower MW PEG solutions, we found minimal aggregation for the motile bacteria in PEG 100 kDa (Figure 2.S2b) and no aggregation for motile bacteria in PEG 10 kDa (Figure 2.2a). Because the main physical difference between the motile and

nonmotile bacteria is their motion, we hypothesized that the differences in depletion attractions required to aggregate motile versus nonmotile bacteria at short time scales are due to cell motility. Previously, researchers have modeled the steady-state depletion aggregation of motile bacteria by assuming that the swim force produced by bacterial motility directly counteracts the depletion force (18). We hypothesized that a similar framework could be applied here; the swim force counteracts the depletion force at low PEG concentrations, negating any attractive force to aggregate the cells. Therefore, a stronger depletion force, or higher PEG concentration, is necessary to aggregate the motile bacteria to the same extent as the nonmotile bacteria. Support for this hypothesis was experimentally demonstrated at c_P = 0.4 mg mL^{-1} whereby we observed the nonmotile bacteria aggregate but the motile do not.

Effective potentials describe E. coli aggregation in the dilute polymer concentration regime

To investigate the interplay of swim and depletion forces that give rise to the differences in the observed aggregation behavior between motile and nonmotile bacteria, we began by using effective potentials to describe aggregation in the dilute polymer concentration regime. We focused on explaining the discrepancy in aggregation at c_P= 0.4 mg mL^{-1}, where we observed substantial aggregation of nonmotile bacteria but no aggregation of motile bacteria (Figure 2.3a). In The dilute regime, nonmotile bacterial aggregation increased with the PEG polymer concentration from the osmotic pressure contribution in the depletion potential (Figure 2.3b). The effective potential between nonmotile bacteria can be described by the Asakura–Oosawa depletion potential (15, 16, 44):

$$U_{dep}(r) = \begin{cases} +\infty \; for \; r \leq 0 \\ -2\pi\Pi_P a\left(R_P - \frac{r}{2}\right)^2 \; for \; 0 < r < 2R_P \\ 0 \; for \; r > 2R_P \end{cases}$$

(Eq. 2.3)

where U_{dep} is the depletion potential (in J), Π_P is the osmotic pressure of the polymer solution (in Pa), a is the radius of the bacteria (approximated as a sphere, in m), R_P is the characteristic polymer length scale (in m), and r is the separation distance between two bacteria surfaces (in m). The contribution of polymer concentration to the depletion potential

is implicit in the osmotic pressure. PEG behaves as a polymer in good solvent in buffer (45), and we can use the following crossover equation for the osmotic pressure of a polymer solution which spans the dilute and semi-dilute polymer concentration regimes (46, 47):

$$\Pi_P = \frac{N_{Avo} kT}{MW} c_P \left(1 + \left(\frac{c_P}{c_P^*}\right)^{1.3}\right)$$

(Eq. 2.4)

where N_{Avo} is Avogadro's number, k is the Boltzmann constant, T is the temperature (in Kelvin), MW is the molecular weight of the polymer (in kDa), c_P is the polymer mass concentration (in kg m^{-3}), and c_P^* is the overlap mass concentration (in kg m^{-3}). We use the concentration-dependent radius of gyration to estimate the characteristic polymer length scale (48, 49):

$$R_P(c_P) = R_g(0) \left(\frac{MW}{N_{Avo} kT} \frac{d\Pi_P}{dc_P}\right)^{-1/2}$$

(Eq. 2.5)

where $R_P(c_P)$ is the characteristic polymer length scale (in m), and $R_g(0)$ is the radius of gyration at dilute concentrations (in m).

To estimate the depletion potential in polymer concentrations around the overlap concentration, equations 2.4 and 2.5 are used to obtain the correct scaling arguments with respect to polymer concentration (c_P). In the dilute polymer concentration regime ($c_P < c_P^*$), the polymer osmotic pressure (Π_P) scales linearly with polymer concentration (matching the van't Hoff Law) and the characteristic polymer length scale (R_P) is given by the radius of gyration at dilute concentrations ($R_g(0)$) and is independent of polymer concentration (c_P). In the semi-dilute polymer concentration regime ($c_P > c_P^*$), the polymer osmotic pressure (Π_P) scales as $\Pi_P \propto c_P^{1.3}$ as according to de Gennes scaling theory, and the characteristic polymer length scale (R_P) goes with the correlation length of the polymer solution (ξ), which scales as $\xi \propto c_P^{-0.77}$ for a good solvent (28). Thus, equations 2.4 and 2.5 reasonably captures the scaling for the polymer concentrations regimes around the overlap concentration, which is the polymer concentration regime studied in this chapter. Previous studies have derived

similar crossover equations to describe experimentally observed depletion driven aggregation in polymer concentrations that span the dilute and semi-dilute concentration regimes (53).

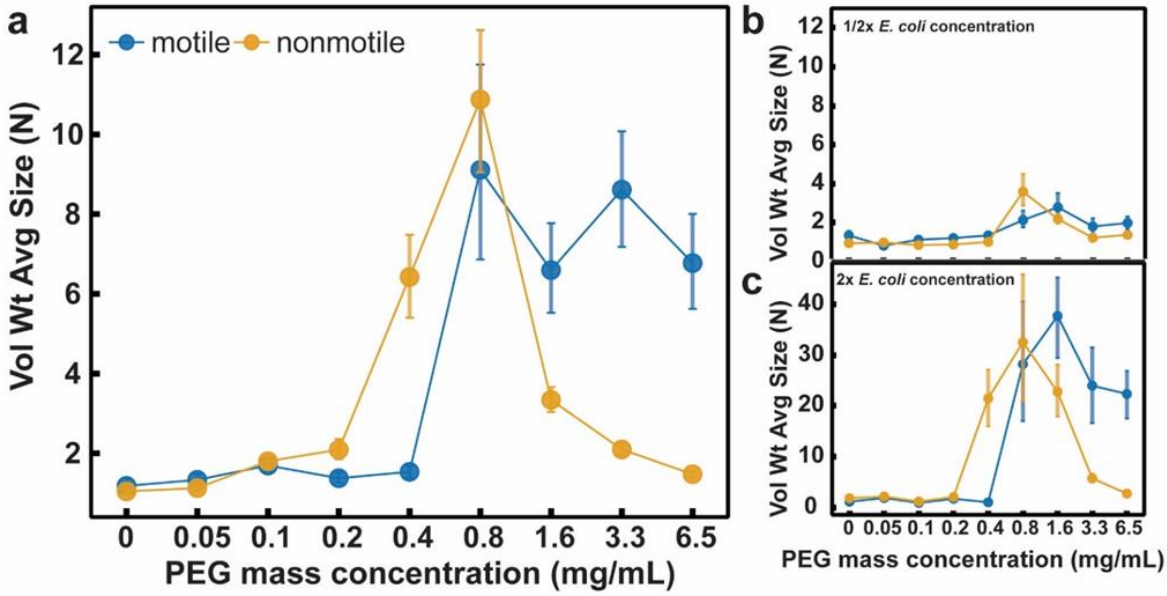


Figure 2.2: A comparison of the aggregation of motile and nonmotile *E. coli* K12 at a range of concentrations of 1 MDa PEG. Volume-weighted average aggregate sizes (Vol Wt Avg Size) of nonmotile and motile *E. coli* K12 for serial dilutions of 1 MDa PEG using a bacteria concentration of $1 \times 10^9\ CFU/mL$ (a) and for *E. coli* at half (b) and twice (c) this concentration. Aggregate sizes were measured 10 min after cells were mixed with PEG. Volume-weighted average sizes in terms of bacteria per aggregate (N) are plotted against polymer mass concentration (c_P) in mg/mL. Vertical error bars are 95% empirical bootstrap confidence intervals using the bootstrapping protocol described in [Imaging Analysis] in Methods of Ref.[1] Data for each PEG concentration were compiled from at least four (a), at least three (b), or at least two (c) biological replicates in these experiments (where a replicate is a new bacterial culture). For each concentration of PEG, each replicate was obtained from one z-stack that was comprised of about 135 slices.

$R_g(0)$ was estimated using literature values of the hydrodynamic radius of PEG (30) and the Kirkwood–Riseman relation (50, 51) We estimated $R_g(0)$ to be 62.6 nm for 1 MDa PEG,

and using this value and the molecular weight of the polymer, we estimated c_P^* to be 1.6 mg mL^{-1} (see Experimental for calculation). Combining Eq. (2.3)–(2.5) gives us an expression for the depletion potential that closely approximates the Asakura– Oosawa potential in the dilute polymer concentration regime and the potential derived by Joanny, Leibler, and De Gennes in the semi-dilute regime (52) Similar crossover equations for the depletion potential have been previously used to quantitatively describe experimentally observed depletion-driven colloid aggregation in polymer solutions that span the dilute and semi-dilute concentration regimes (1, 53) As the polymer concentration increases, the depletion potential also increases. We observed that the aggregate size of the nonmotile *E. coli* increases with PEG concentration in the dilute polymer concentration regime, which suggests that depletion forces drive aggregation.

We found that motile bacteria do not aggregate at low PEG concentrations until a certain PEG concentration threshold is reached (Figure 2.3c). To estimate the effective potential of motile bacteria and the effect of the swim force on the aggregation of motile bacteria, we used a previously established theoretical framework (18). We began by considering the forces that bacteria experience in solution. This model accounts for the swim force that arises from bacterial motility and the polymer-induced depletion force. The swim force can be described from the ellipsoid approximation to the Stokes–Einstein drag coefficient (17, 54):

$$F_{swim} = -\frac{4\pi\eta b}{\ln(2b/a) - 1/2}V$$

(Eq. 2.6)

where η is the solution viscosity in Pa s (see Experimental for details of estimate), a and b are the lengths of the semi-minor and semi-major axes for E. coli (in m), and V is the speed (in m s^{-1}). For E. coli, a and b are approximated to be about 0.5 mm and 2 mm (55) respectively, and V is assumed to be constant at about 10 mm s^{-1} (4). The effective force is then calculated using a force balance on the bacteria, assuming that the swim force directly counteracts the depletion force:

$$F_{eff} = F_{swim} - \frac{\partial U_{dep}(r)}{\partial r}$$

(Eq. 2.7)

where F_{eff} is the effective force (in Newtons) and the depletion force is given by the negative first derivative of $U_{dep}(r)$ with respect to r. To find the effective potential U_{eff}, we integrate Eq. (2.5) with respect to r:

$$U_{eff} = \begin{cases} \infty \ for\, r \leq 0 \\ -F_{swim} r + U_{dep}(r) + U_0 \ \ for\ 0 < r < r^* \\ 0\ for\ r^* \geq 0 \end{cases}$$

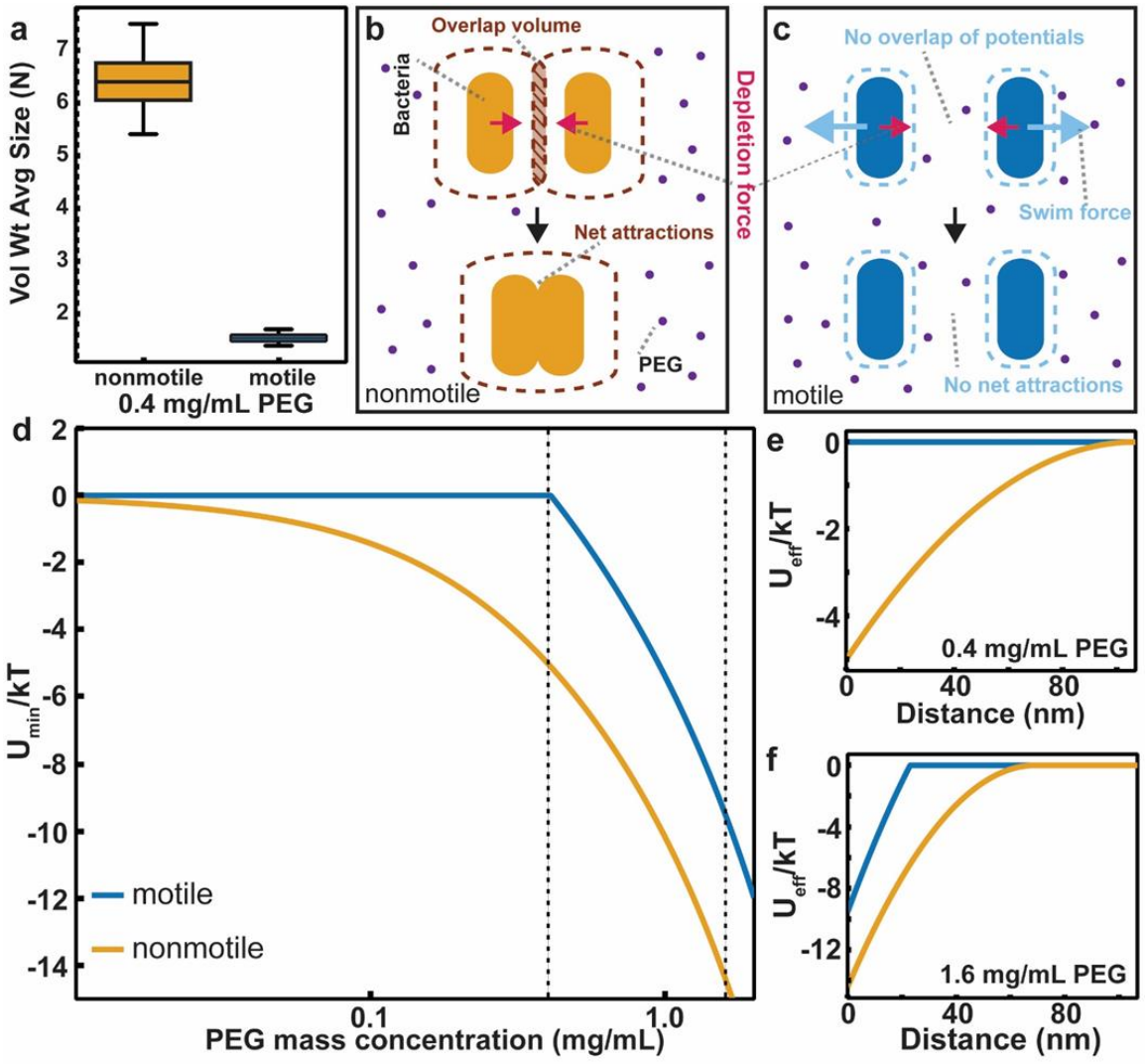
(Eq. 2.8)

The integration constant, $U_0(J)$, is defined as described previously (18). In a condition where two bacteria are swimming in the exact opposite direction, there is a separation distance r^* where the effective force acting on each bacterium is zero. Beyond this range, the swim force overwhelms the depletion potential, and the effective potential on the bacteria is zero. U_0 is defined such that U_{eff} is zero beyond r^*.

In the dilute polymer concentration regime, we observed that the aggregation trends for both motile and nonmotile bacteria were qualitatively consistent with expectations based on the changes of the minima of their respective effective potentials ($U_{eff}\ (r = 0)$) at each PEG concentration (Figure 2.3d). For nonmotile bacteria, the effective potential consists of only the depletion potential, which increases with polymer concentration. Our observations of the aggregation of nonmotile bacteria were qualitatively consistent with what is predicted from the depletion potential. In contrast, for motile bacteria, theoretical calculations suggest that swim force will exceed the depletion force at low polymer concentrations, resulting in no effective potential. Our experimental observations were consistent with these calculations; we saw no aggregation in motile bacteria at PEG 1 MDa concentrations less than 0.8 mg mL^{-1}. This effect is further illustrated by looking at the shape of the effective potentials for motile and nonmotile bacteria at c_P = 0.4 mg mL^{-1} (Figure 2.3e). For nonmotile bacteria, the total depletion range spans about 120 nm from the bacteria surface, and at inter-bacterial contact, the depletion well is at its minimum at approximately -5kT. In contrast, theory suggests that the swim force of motile bacteria will completely dominate over the depletion potential, resulting in no net attractive potential. The predicted lack of net attractions between motile

bacteria is supported experimentally as we did not observe aggregation of motile bacteria at $c_P \leq 0.4$ mg mL^{-1}; the depletion forces are simply not strong enough to compete with the swimming force of the motile bacteria to induce aggregation.

[Figure on next page] Figure 2.3: Effective potentials describe aggregation of motile and nonmotile *E. coli* in the dilute concentration regime of PEG 1 MDa. (a) The volume-weighted average aggregate size (Vol Wt Avg Size, N) are plotted for both motile and nonmotile *E. coli* at cP = 0.4 mg mL^{-1} PEG. The box plots depict the 95% empirical bootstrap confidence intervals of the Vol Wt Avg Size calculated using the method described in the ''Imaging analysis'' section of the Methods of Ref. 1. The line bisecting the box is the 50th percentile; the upper and lower edges of the box are the 25th and 75th percentile, respectively; and the whiskers are the 2.5th and 97.5th percentiles. Data were compiled from at least three biological replicates. (b) A schematic of nonmotile bacteria (orange) in PEG (purple) solution at 0 min and at (c) 10 min. The PEG is excluded from the inter-bacterial volume inducing an effective potential (brown dotted line) due to depletion (pink arrows). (d) A schematic of motile bacteria (blue) in PEG (purple) solution at 0 min and at (e) 10 min. Although the PEG induces the same depletion potential (pink arrows) at a given concentration, the swim force (white arrows) from the bacterial motility decreases both the well depth and the range, reducing their effective potential (blue dotted line) and preventing aggregation in the dilute PEG concentration regime. (f) The effective potential at contact (U_{min}/kT) is plotted for motile and nonmotile *E. coli* against PEG concentration (mg mL^{-1}). The vertical black dotted lines at c_P = 0.4 and c_P = 1.6 mg mL^{-1} denote the potential minima taken from the complete potentials plotted in (g) and (h), respectively. (g) The full effective potential (U_{eff}/kT) is plotted against distance from the bacterial surface for both motile and nonmotile *E. coli* at cP = 0.4 mg mL^{-1} PEG. (h) The full effective potential (U_{eff}/kT) is plotted against distance from the bacterial surface for both motile and nonmotile *E. coli* at c_P = 1.6 mg mL^{-1} PEG.


a
Vol Wt Avg Size (N)
nonmotile
motile
0.4 mg/mL PEG
b
Overlap volume
Bacteria
Depletion force
Net attractions
PEG
nonmotile
c
No overlap of potentials
Swim force
No net attractions
motile
d
U_{min}/kT
motile
nonmotile
PEG mass concentration (mg/mL)
e
U_{eff}/kT
0.4 mg/mL PEG
Distance (nm)
f
U_{eff}/kT
1.6 mg/mL PEG
Distance (nm)

At $c_P \geq$ 0.4 mg mL^{-1} and as the polymer concentration enters the semi-dilute regime, effective potentials are insufficient to explain the experimentally observed aggregation results. At c_P = 1.6 mg mL^{-1} (overlap concentration), where both motile and nonmotile *E. coli* aggregate, the depletion potential is strong enough to induce attractions between motile bacteria (Figure 2.3f). Beyond the overlap concentration, in which the solution enters the semi-dilute regime, the magnitude of the effective potential minima continues to increase (Figure 2.3d). Based on these predictions, we expect to see greater aggregation of nonmotile bacteria in the semi-dilute regime. However, in our experiments, we observed that aggregation actually began to decrease for the nonmotile bacteria at PEG concentrations of 40.8 mg mL^{-1} (Figure 2.2a). Because the theory did not match with our experimental observations at these short time scales, it suggests that effective potentials are insufficient to explain the observed aggregation behavior at PEG concentrations greater than 0.8 mg mL^{-1}. We therefore proposed an alternate explanation for bacterial aggregation in the semi-dilute polymer concentration regime that relies on effective diffusivity. We test this hypothesis next.

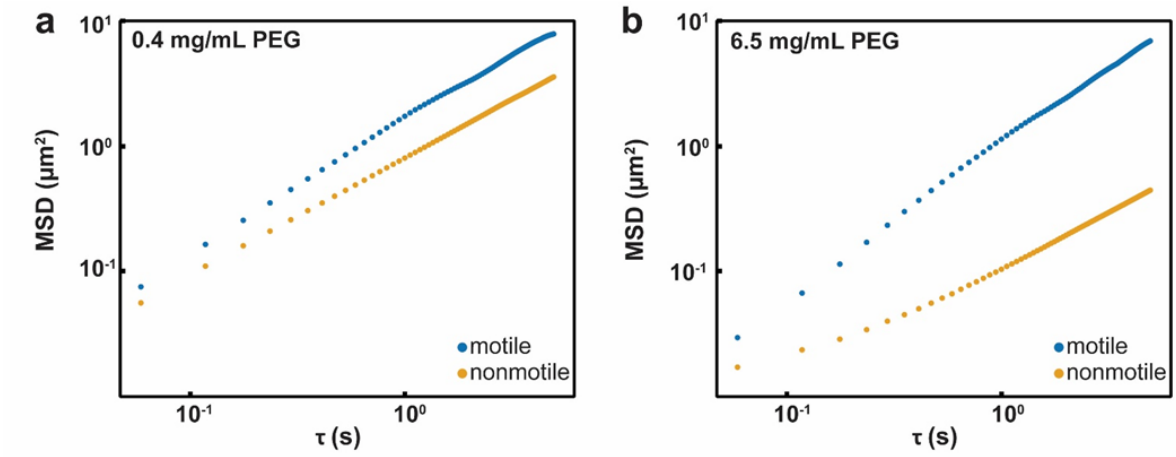


Figure 2.4: The mean-squared displacement (MSD) of *E. coli* in 1 MDa PEG solutions. (a) The MSD (μm^2) of both motile (blue) and nonmotile (orange) *E. coli* are plotted against lag time t (s) in c_P = 0.4 mg mL^{-1} and (b) c_P = 6.5 mg mL^{-1} PEG.

Reduction in bacterial diffusivity explains aggregation in the semi-dilute polymer concentration regime where effective potentials are insufficient

We hypothesized that in the semi-dilute concentration regime, the higher solution viscosity limits Brownian motion more appreciably than it limits bacterial motility. If this were the case, one would expect a reduced collision frequency among nonmotile bacteria leading to reduced aggregation at short time scales in viscous environments. In contrast, because motile bacteria are able to swim in viscous environments (4), their rate of collision would remain high and they would aggregate at short time scales even in highly viscous environments.

To test our hypothesis, we compared the displacement of both motile and nonmotile *E. coli* in the dilute concentration regime (c_P = 0.4 mg mL^{-1}) and in the semi-dilute concentration regime (6.5 mg mL^{-1}) by measuring their MSD (Figure 2.4a). The MSD (in μm^2) is defined as $MSD(\tau) = \langle|\boldsymbol{r}(t_0+\tau) - \boldsymbol{r}(t_0)|^2\rangle$, where r (in μm) is a vector describing the position of the bacteria at a given time scale t (in s). At cP = 0.4 mg mL^{-1}, where nonmotile bacteria aggregate but motile do not, displacement of both motile and nonmotile cells is high. At this PEG concentration, we had already established that the depletion force is strong enough to aggregate nonmotile bacteria but is not strong enough to overcome the swim force induced by the motile cells. Although motile cells are likely undergoing more collisions due to their motility, there are insufficient attractions to initiate aggregation. The displacement of the nonmotile bacteria at c_P = 6.5 mg mL^{-1} was substantially lower than their displacement at c_P = 0.4 mg mL^{-1}, supporting the assertion that their collision frequency, and therefore aggregation rate, was decreased. The displacement of motile bacteria at this PEG concentration was much larger than the nonmotile bacteria, confirming that the higher viscosity did not impede movement and therefore collisions (Figure 2.4b). The higher collision frequency of the motile bacteria, combined with the high depletion potential at the c_P = 6.5 mg mL^{-1} PEG concentration, resulted in aggregation of the motile bacteria (Figure 2.2). These results suggest that the at short time scales, aggregation becomes kinetically limited and is controlled by collision frequency of bacteria and their aggregates.

Conclusions

Bacterial aggregation often occurs in complex physiological environments such as the gastrointestinal tract or the lungs; in these settings it is important to understand the initial formation of aggregates before physiological factors such as peristalsis or swallowing play a role. Bacterial aggregation is thought to be the first step in colonization and biofilm formation; therefore, understanding the mechanism by which aggregates initially form could reveal important patterns in these processes. In this paper, we investigated how the interplay of swim forces that arise from motility and polymer-induced depletion forces influence bacterial aggregation. It has been reported previously that, at steady state, *E. coli* aggregate in the presence of non-adsorbing polymers, due to polymer-induced depletion forces (18). At steady state, an effective potential incorporating contributions from swim forces and depletion forces adequately describes the aggregation behavior of the system (18), wherein motility diminishes depletion-driven aggregation. Thus, polymers in greater quantities and/or sizes (i.e., a stronger depletion potential) are required for motile bacteria to aggregate compared with their nonmotile analogs. Here, we investigated the effect of bacterial motility on aggregation at short time scales (before the system reaches steady state).

We found that, in contrast to what has been demonstrated at steady state, at short time scales, motility can actually enhance aggregation, resulting in the formation of aggregates that are larger than those formed by nonmotile bacteria under the same conditions. For polymer-induced depletion forces to result in bacterial aggregation at short time scales, two conditions must be met: (i) sufficient inter-bacterial attractions, and (ii) sufficient inter-bacterial collisions. In the dilute polymer concentration regime, swimming competes with depletion to reduce the effective potential of a bacteria, resulting in less aggregation. Above the overlap concentration, motility allows the bacteria to overcome kinetic hindrances that may otherwise prevent aggregation at high viscosity environments, resulting in an increase in aggregation.

Future work should explore whether differences in motility impact aggregation of microbes and microbial physiology in more complex environments, e.g., in the gastrointestinal tract and the lungs. Beyond bacterial aggregation, this work may inform other studies of the behavior of active matter at early timescales.

Conflicts of Interest

The technology described in this publication is the subject of a patent application filed by Caltech.

Acknowledgements

This work was funded in part by the Army Research Office (ARO) Multidisciplinary University Research Initiative (MURI) contract #W911NF-17-1-0402, the Jacobs Institute for Molecular Engineering for Medicine, an NSF Graduate Research Fellowship DGE-144469 (to APS), and a Center for Environmental Microbial Interactions (CEMI) Caldwell Graduate Fellowship (to APS). We thank Andres Collazo and the Beckman Imaging Facility for help with imaging, and we thank Natasha Shelby for contributions to writing and editing this manuscript.

References

1. Preska Steinberg, A., Datta, S.S., Naragon, T., Rolando, J.C., Bogatyrev, S.R. and Ismagilov, R.F., 2019. High-molecular-weight polymers from dietary fiber drive aggregation of particulates in the murine small intestine. *Elife*, *8*, p.e40387.

2. Stecher, B., Hapfelmeier, S., Müller, C., Kremer, M., Stallmach, T. and Hardt, W.D., 2004. Flagella and chemotaxis are required for efficient induction of Salmonella enterica serovar Typhimurium colitis in streptomycin-pretreated mice. *Infection and immunity*, *72*(7), pp. 4138–4150.

3. O'Toole, G.A. and Kolter, R., 1998. Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Molecular microbiology*, *30*(2), pp. 295–304.

4. Patteson, A.E., Gopinath, A., Goulian, M. and Arratia, P.E., 2015. Running and tumbling with E. coli in polymeric solutions. *Scientific reports*, *5*(1), p. 15761.

5. Puri, S., Friedman, J., Saraswat, D., Kumar, R., Li, R., Ruszaj, D. and Edgerton, M., 2015. Candida albicans shed Msb2 and host mucins affect the candidacidal activity of salivary Hst 5. *Pathogens*, *4*(4), pp. 752–763.

6. Moor, K., Diard, M., Sellin, M.E., Felmy, B., Wotzka, S.Y., Toska, A., Bakkeren, E., Arnoldini, M., Bansept, F., Co, A.D. and Völler, T., 2017. High-avidity IgA protects the intestine by enchaining growing bacteria. *Nature*, *544*(7651), pp. 498–502.

7. Bergström, J.H., Birchenough, G.M., Katona, G., Schroeder, B.O., Schütte, A., Ermund, A., Johansson, M.E. and Hansson, G.C., 2016. Gram-positive bacteria are held at a distance in the colon mucus by the lectin-like protein ZG16. *Proceedings of the National Academy of Sciences*, *113*(48), pp. 13833–13838.

8. Caldara, M., Friedlander, R.S., Kavanaugh, N.L., Aizenberg, J., Foster, K.R. and Ribbeck, K., 2012. Mucin biopolymers prevent bacterial aggregation by retaining

cells in the free-swimming state. Current Biology, 22(24), pp. 2325–2330.

9. Ramesh, A., Lee, D.J. and Hong, S.G., 2006. Soluble microbial products (SMP) and soluble extracellular polymeric substances (EPS) from wastewater sludge. *Applied microbiology and biotechnology*, *73*, pp. 219–225.

10. Liu, H. and Fang, H.H., 2002. Extraction of extracellular polymeric substances (EPS) of sludges. *Journal of biotechnology*, *95*(3), pp. 249–256.

11. Hojo, K., Nagaoka, S., Ohshima, T. and Maeda, N., 2009. Bacterial interactions in dental biofilm development. *Journal of dental research*, *88*(11), pp. 982–990.

12. Laganenka, L., Colin, R. and Sourjik, V., 2016. Chemotaxis towards autoinducer 2 mediates autoaggregation in Escherichia coli. *Nature communications*, *7*(1), p. 12984.

13. Secor, P.R., Michaels, L.A., Ratjen, A., Jennings, L.K. and Singh, P.K., 2018. Entropically driven aggregation of bacteria by host polymers promotes antibiotic tolerance in Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, *115*(42), pp. 10780–10785.

14. Solano Goñi, C., Echeverz Sarasúa, M. and Lasa Uzcudun, Í., 2014. Biofilm dispersion and quorum sensing. *Current Opinion in Microbiology 2014, 18: 96–104*.

15. Asakura, S. and Oosawa, F., 1954. On interaction between two bodies immersed in a solution of macromolecules. *The Journal of chemical physics*, *22*(7), pp. 1255–1256.

16. Asakura, S. and Oosawa, F., 1958. Interaction between particles suspended in solutions of macromolecules. *Journal of polymer science*, *33*(126), pp. 183–192.

17. Dorken, G., Ferguson, G.P., French, C.E. and Poon, W.C., 2012. Aggregation by depletion attraction in cultures of bacteria producing exopolysaccharide. *Journal of The Royal Society Interface*, *9*(77), pp. 3490–3502.

18. Schwarz-Linek, J., Valeriani, C., Cacciuto, A., Cates, M.E., Marenduzzo, D., Morozov, A.N. and Poon, W.C.K., 2012. Phase separation and rotor self-assembly in active particle suspensions. *Proceedings of the National Academy of Sciences*, *109*(11), pp. 4052–4057.

19. Takatori, S.C. and Brady, J.F., 2016. Forces, stresses and the (thermo?) dynamics of active matter. *Current Opinion in Colloid & Interface Science*, *21*, pp. 24–33.

20. Solon, A.P., Fily, Y., Baskaran, A., Cates, M.E., Kafri, Y., Kardar, M. and Tailleur, J., 2015. Pressure is not a state function for generic active fluids. *Nature physics*, *11*(8), pp. 673–678.

21. Cates, M.E. and Tailleur, J., 2015. Motility-induced phase separation. *Annu. Rev. Condens. Matter Phys.*, *6*(1), pp. 219–244.

22. Cates, M.E. and Tailleur, J., 2013. When are active Brownian particles and run-and-tumble particles equivalent? Consequences for motility-induced phase separation. *Europhysics Letters*, *101*(2), p. 20010.

23. Stenhammar, J., Tiribocchi, A., Allen, R.J., Marenduzzo, D. and Cates, M.E., 2013. Continuum theory of phase separation kinetics for active Brownian particles. *Physical review letters*, *111*(14), p. 145702.

24. Marchetti, M.C., Joanny, J.F., Ramaswamy, S., Liverpool, T.B., Prost, J., Rao, M. and Simha, R.A., 2013. Hydrodynamics of soft active matter. *Reviews of modern physics*, *85*(3), p. 1143.

25. Huizinga, J.D. and Lammers, W.J., 2009. Gut peristalsis is governed by a multitude of cooperating mechanisms. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *296*(1), pp. G1–G8.

26. Sbalzarini, I.F. and Koumoutsakos, P., 2005. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of structural biology*, *151*(2), pp. 182–195.

27. Gao, Y. and Kilfoil, M.L., 2009. Accurate detection and complete tracking of large populations of features in three dimensions. *Optics express*, *17*(6), pp. 4685–4704.

28. Rubinstein, M. and Colby, R.H., 2003. *Polymer physics*. Oxford university press.

29. Dobrynin, A.V. and Rubinstein, M., 2005. Theory of polyelectrolytes in solutions and at surfaces. *Progress in polymer science*, *30*(11), pp. 1049–1118.

30. Armstrong, J.K., Wenby, R.B., Meiselman, H.J. and Fisher, T.C., 2004. The hydrodynamic radii of macromolecules and their effect on red blood cell aggregation. *Biophysical journal*, *87*(6), pp. 4259–4270.

31. Ebagninin, K.W., Benchabane, A. and Bekkour, K., 2009. Rheological characterization of poly (ethylene oxide) solutions of different molecular weights. *Journal of colloid and interface science*, *336*(1), pp. 360–367.

32. Gabel, C.V. and Berg, H.C., 2003. The speed of the flagellar rotary motor of Escherichia coli varies linearly with protonmotive force. *Proceedings of the National Academy of Sciences*, *100*(15), pp. 8748–8751.

33. Valentine, M.T., Perlman, Z.E., Gardel, M.L., Shin, J.H., Matsudaira, P., Mitchison, T.J. and Weitz, D.A., 2004. Colloid surface chemistry critically affects multiple particle tracking measurements of biomaterials. *Biophysical journal*, *86*(6), pp. 4004–4014.

34. Wang, Y.Y., Lai, S.K., Suk, J.S., Pace, A., Cone, R. and Hanes, J., 2008. Addressing the PEG mucoadhesivity paradox to engineer nanoparticles that “slip” through the human mucus barrier. *Angewandte Chemie (International ed. in English)*, *47*(50), p. 9726.

35. Vincent, B., Edwards, J., Emmett, S. and Jones, A., 1986. Depletion flocculation in dispersions of sterically-stabilised particles (“soft spheres”). *Colloids and Surfaces*, *18*(2-4), pp. 261–281.

36. Cowell, C., Li-In-On, R. and Vincent, B., 1978. Reversible flocculation of sterically-stabilised dispersions. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, *74*, pp. 337–347.

37. Vincent, B., Luckham, P. and Waite, F.A., 1980. The effect of free polymer on the stability of sterically stabilized dispersions. *Journal of Colloid and Interface Science*, *73*(2), pp. 508–521.

38. Clarke, J. and Vincent, B., 1981. Nonaqueous silica dispersions stabilized by terminally anchored polystyrene: The effect of added polymer. *Journal of Colloid and Interface Science*, *82*(1), pp. 208–216.

39. Feigin, R.I. and Napper, D.H., 1980. Depletion stabilization and depletion flocculation. *Journal of Colloid and Interface Science*, *75*(2), pp. 525–541.

40. Vincent, B., Clarke, J. and Barnett, K.G., 1986. The Flocculation fo non-aqueous, sterically-stabilised latex dispersions in the presence of free polymer. *Colloids and surfaces*, *17*(1), pp. 51–65.

41. Gast, A.P. and Leibler, L., 1986. Interactions of sterically stabilized particles suspended in a polymer solution. *Macromolecules*, *19*(3), pp. 686–691.

42. Jones, A. and Vincent, B., 1989. Depletion flocculation in dispersions of sterically-stabilised particles 2. Modifications to theory and further studies. *Colloids and Surfaces*, *42*(1), pp. 113–138.

43. Allain, C., Cloitre, M. and Wafra, M., 1995. Aggregation and sedimentation in colloidal suspensions. *Physical review letters*, *74*(8), p. 1478.

44. Vrij, A., 1976. Polymers at interfaces and the interactions in colloidal dispersions. *Pure and Applied Chemistry*, *48*(4), pp. 471–483.

45. Kawaguchi, S., Imai, G., Suzuki, J., Miyahara, A., Kitano, T. and Ito, K., 1997. Aqueous solution properties of oligo-and poly (ethylene oxide) by static light scattering and intrinsic viscosity. *Polymer*, *38*(12), pp. 2885–2891.

46. Rubinstein, M. and Colby, R.H., 2003. *Polymer physics*. Oxford university press.

47. Cai, L., 2012. *Structure and function of airway surface layer of the human lungs & mobility of probe particles in complex fluids* (Doctoral dissertation, The University of North Carolina at Chapel Hill).

48. Burchard, W., 2001. Structure formation by polysaccharides in concentrated solution. *Biomacromolecules*, *2*(2), pp. 342–353.

49. Prasad, V., 2002. *Weakly interacting colloid polymer mixtures*. Harvard University.

50. Tanford, C., 1966. *Physical chemistry of macromolecules*. John Wiley & Sons, Incorporated.

51. Lee, H., Venable, R.M., MacKerell, A.D. and Pastor, R.W., 2008. Molecular dynamics studies of polyethylene oxide and polyethylene glycol: hydrodynamic radius and shape anisotropy. *Biophysical journal*, *95*(4), pp. 1590–1599.

52. Joanny, J.F., Leibler, L. and De Gennes, P.G., 1979. Effects of polymer solutions on colloid stability. *Journal of Polymer Science: Polymer Physics Edition*, *17*(6), pp. 1073–1084.

53. Verma, R., Crocker, J.C., Lubensky, T.C. and Yodh, A.G., 1998. Entropic colloidal interactions in concentrated DNA solutions. *Physical review letters*, *81*(18), p. 4004.

54. Yan, W. and Brady, J.F., 2015. The swim force as a body force. *Soft Matter*, *11*(31), pp. 6235–6244.

55. Cates, M.E., 2012. Diffusive transport without detailed balance in motile bacteria: does microbiology need statistical physics?. *Reports on Progress in Physics*, *75*(4), p. 042601.

Supplemental Information

Supplemental experimental section. Viability of *E. coli* K12 after treatment with sodium azide was tested by plating cells onto lysogeny broth (LB) with 1.5% agar. Cells were prepared as described in the Experimental section, and were left to incubate at room temperature for 1 h before plating. Duplicate agar plates were incubated at 35 °C overnight and then colonies were counted. None of the cells in the azide-treated group survived, whereas the cells washed in motility buffer maintained viability.

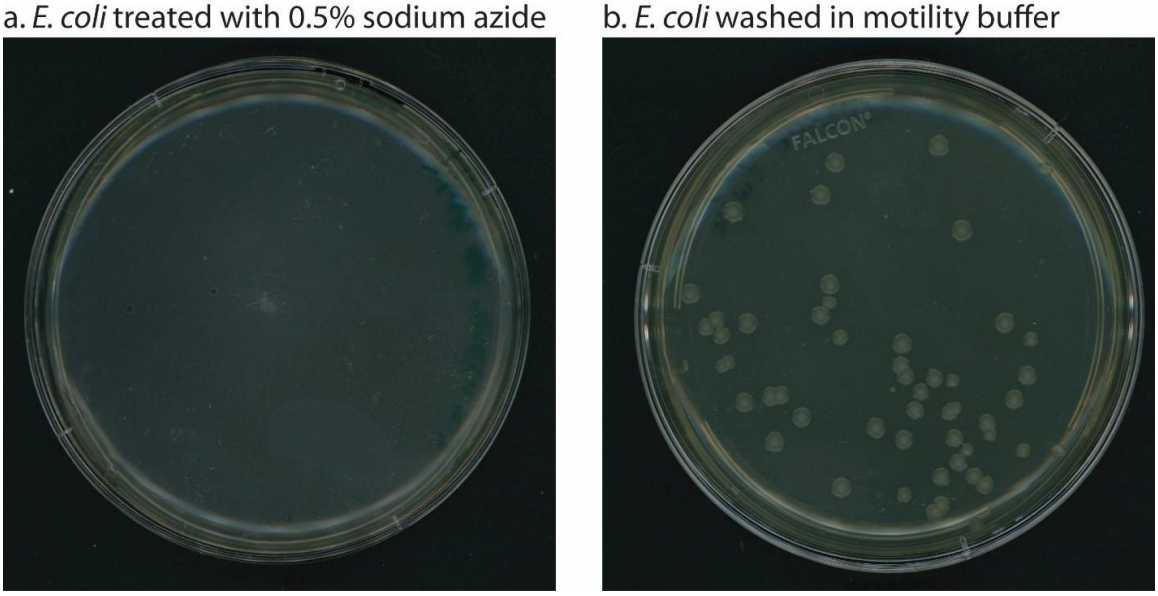


Figure 2.S1: Confirming the deactivation of *Escherichia coli* via plating. *Escherichia coli* washed (a) with sodium azide in motility buffer and (b) motility buffer (no sodium azide) which were then plated on LB and 1.5% agar.

Aggregation of motile and nonmotile *E. coli* was measured in a serial dilution of 10 kDa and 100 kDa PEG around their respective overlap concentrations. In 10 kDa PEG, neither motile nor nonmotile bacteria aggregated at the concentrations tested (c_P/c_P^*= 1/32 to 4). A similar observation was made using PEG-coated particles instead of bacteria in a similar MW PEG solution in our previous publication (Preska Steinberg *et al*. 2019).

In the 100 kDa PEG solutions, nonmotile *E. coli* aggregated in a manner qualitatively consistent with polymer-driven depletion aggregation, as reported with PEG-coated particles tested with the same MW PEG in the same concentration range (c_P/c_P^*= = 1/32 to 4). The nonmotile bacteria aggregation measured in the 100 kDa PEG is similar but smaller in magnitude compared to the aggregation measured in the 1 MDa PEG, indicative of the smaller depletion potential exerted from the smaller MW PEG. Minimal aggregation was measured for the motile *E. coli* in 100 kDa PEG, suggesting that at this MW and concentration range, the swim force is strong enough to overcome the depletion potential.

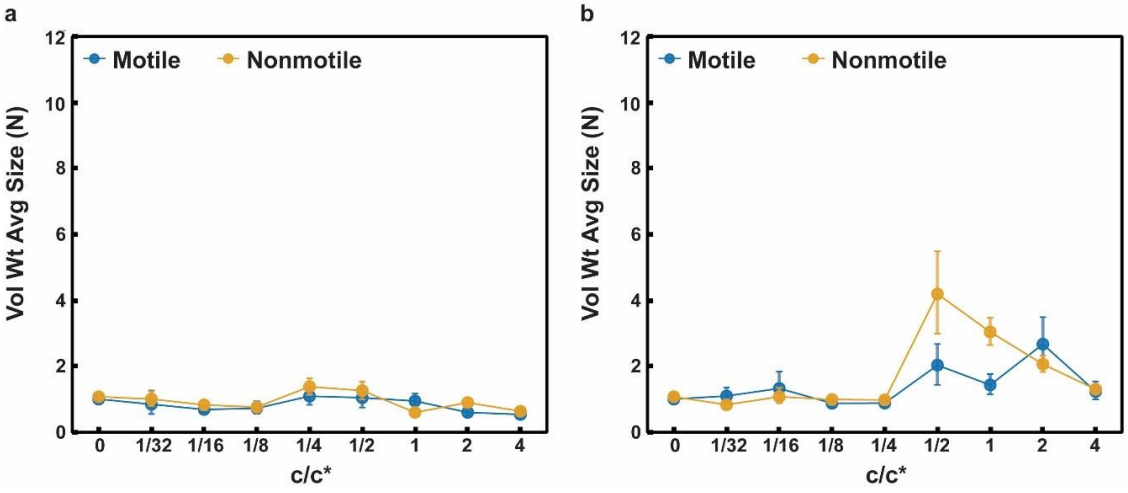


Figure 2.S2: A comparison of the aggregation of motile and nonmotile *E. coli* K12 in a range of concentrations of 10 kDa and 100 kDa PEG. Volume-weighted average aggregate sizes (Vol Wt Avg Size) of nonmotile and motile *E. coli* K12 for serial dilutions of (a) 10 kDa PEG and (b) 100 kDa PEG. Aggregate sizes were measured 10 min after cells were mixed with PEG using a cell concentration of 10^9 cells/mL. Volume-weighted average sizes in terms of bacteria per aggregate (N) are plotted against polymer mass concentration (c_P) normalized by overlap concentration (c_P^*). Overlap concentration was estimated to be 85 mg/mL for 10 kDa PEG and 8.5 mg/mL for 100 kDa PEG. Vertical error bars are 95% empirical bootstrap confidence intervals using the bootstrapping protocol described in the "Imaging Analysis" section of the Methods in Ref.1 Data for the 10 kDa PEG concentration were compiled from one biological replicate and data for the 100 kDa PEG concentration were compiled from two biological replicates (where each replicate is a separate bacterial culture). For each concentration of PEG, each replicate was obtained from one z-stack that was comprised of at least 120 slices.

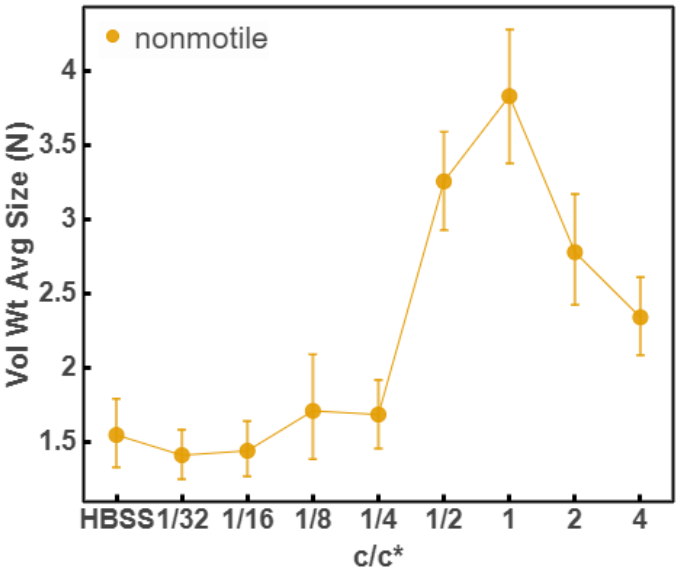


Figure 2.S3: Aggregation of nonmotile *E. coli* without azide treatment. A comparison of the aggregation of *E. coli* ORN225 (1) grown until stationary phase at a range of concentrations of 1 MDa PEG. Volume-weighted average aggregate sizes (Vol Wt Avg Size) of *E. coli* ORN225 for serial dilutions of 1 MDa PEG using a bacteria concentration of $1 \times 10^9\ CFU/mL$. Aggregate sizes were measured 10 min after cells were mixed with PEG. Volume-weighted average sizes in terms of bacteria per aggregate (N) are plotted against polymer concentration normalized by overlap concentration (c_P/c_P^*) in mg/mL. Vertical error bars are 95% empirical bootstrap confidence intervals using the bootstrapping protocol described in [Imaging Analysis] in Methods of Supplemental Ref (2). Data for each PEG concentration was compiled from one biological replicate in these experiments (where a replicate is a new bacterial culture). For each concentration of PEG, each replicate was obtained from one z-stack that was comprised of about 135 slices.

Supplemental References

1. Orndorff, P.E., Devapali, A., Palestrant, S., Wyse, A., Everett, M.L., Bollinger, R.R. and Parker, W., 2004. Immunoglobulin-mediated agglutination of and biofilm formation by Escherichia coli K-12 require the type 1 pilus fiber. *Infection and immunity*, *72*(4), pp. 1929–1938.

2. Preska Steinberg, A., Datta, S.S., Naragon, T., Rolando, J.C., Bogatyrev, S.R. and Ismagilov, R.F., 2019. High-molecular-weight polymers from dietary fiber drive aggregation of particulates in the murine small intestine. *Elife*, *8*, p.e40387.

Chapter 3

QUANTITATIVE SARS-COV-2 VIRAL-LOAD CURVES IN PAIRED SALIVA AND NASAL SWABS INFORM APPROPRIATE RESPIRATORY SAMPLING SITE AND ANALYTICAL TEST SENSITIVITY REQUIRED FOR EARLIEST VIRAL DETECTION

1. Savela, E.S.*, Viloria Winnett, A.*, Romano, A.E.*, Porter, M.K., Shelby, N., Akana, R., Ji, J., Cooper, M.M., Schlenker, N.W., Reyes, J.A. and Carter, A.M., 2022. Quantitative SARS-CoV-2 viral-load curves in paired saliva samples and nasal swabs inform appropriate respiratory sampling site and analytical test sensitivity required for earliest viral detection. *Journal of clinical microbiology*, *60*(2), pp. e01785–21.

Author Contributions

Listed alphabetically by name

Reid Akana (RA): Assisted in literature analysis with ES, MKP, AVW, MC; collaborated with AVW in creating digital participant symptom surveys; assisted with data quality control/curation with JJ, NWS, NS; collaborated with ES, JJ to write data analysis/visualization code; created current laboratory information management system (LIMS) for specimen logging and tracking. Creation of iOS application for sample logging/tracking. Configured a SQL database for data storage. Created an Apache server and websites to view study data. Configured FTPS server to catalog PCR data. Wrote a Python package to access study data. Worked with ES, AVW, AR to implement logic that prioritized specimen extraction order. Collaborated with ES, MKP, AVW, AR in analyzing RNA stability. Created supplementary Figures 3.3B and 3.3C.

Jacob T. Barlow (JTB): Created initial specimen tracking database to aid in specimen logging and tracking. Maintenance of database and implementation of corrections. Feedback on manuscript draft.

Alyssa M. Carter (AMC): Received and logged specimens, and performed sample QC. Prepared reagents for and assisted with RNA extractions. Performed RT-qPCR and analyzed RT-qPCR data for both time series and screening experiments. Performed some of the initial experiments that assessed RNA stability in nasal swab samples

Matthew M. Cooper (MMC): Collaborated with AVW, MF, NS, YG, RFI, on study design and recruitment strategies. Co-wrote initial IRB protocol and informed consent with AVW and NS; assisted in the writing of the enrollment questionnaire; developed laboratory sample processing workflow for saliva with AVW and AER; performed sample processing on subset of samples; funding acquisition; collaborated with AR to write data processing/visualization code for observing household transmission events for active study participants. Contributor to the design of the calibration curve for saliva LOD experiments.

Matthew Feaster (MF): Co-investigator; collaborated with AVW, MMC, NS, YG, RFI on study design and recruitment strategies; provided guidance and expertise on SARS-CoV-2 epidemiology and local trends.

Ying-Ying Goh (YYG): Co-investigator; collaborated with AVW, MMC, NS, MF, RFI on study design and recruitment strategies; provided guidance and expertise on SARS-CoV-2 epidemiology and local trends.

Rustem F. Ismagilov (RFI): Co-investigator; collaborated with AVW, MMC, NS, MF, YYG on study design and recruitment strategies; provided leadership, technical guidance, oversight, and was responsible for obtaining funding for the study.

Jenny Ji (JJ): Contributed to study design and study organization and implementation with NS and JAR; co-wrote enrollment questionnaire with NS and AVW. Major contributor to curation of participant symptom data. Provided quality control of participant data with RA, NS, NWS. Major contributor to the symptom data analysis and visualization shown in Figure 3.2.

Michael K. Porter (MKP): Performed specimen logging and QC, RNA extractions, RT-qPCR, data processing. Performed data acquisition and analysis for and made Figure 3.S1 with AVW. Prepared participant sample collection materials and helped with supplies acquisition. Assisted in literature analysis with ES, RA, AVW.

Jessica A. Reyes (JAR): Study coordinator; collaborated with NS, AVW, NWS, and RFI on recruitment strategies, translated study materials into Spanish, co-wrote informational sheets with AVW and NS; created instructional videos for participants; enrolled and maintained study participants with NS and NWS.

Anna E. Romano (AER): Developed laboratory swab sample processing workflow with ES. Optimized extraction and ddPCR protocols working with vendor scientists. Created budgets and managed, planned, and purchased reagents and supplies; developed and validated method for RT-qPCR and RT-digital droplet PCR analysis for saliva & swab samples with MMC, and AVW. Performed specimen logging and QC, RNA extractions, RT-qPCR and RT-digital droplet PCR; Design of saliva calibration curve experiment. Analyzed ddPCR data for participant and calibration curve data included in Figure 1. Interpretation of sequence data with AVW. Prepared Figure 3.1 and 3.SI Figure 3.S2 with ES and AVW; Collaborated with ES and RA to generate and curate data for RNA stability analysis. Managing logistics for the expansion of the BSL-2+ lab space with ESS. Edited manuscript.

Emily S. Savela (ESS): Coordinated the laboratory team and division of lab work, coordinated lab schedules to ensure completion of time-sensitive analyses of participant samples while complying with COVID-19 lab occupancy restrictions and biosafety requirements. Performed initial nasal-swab workflow validation experiments with AER. Major contributor to workflow validation, methods, biosafety SOPs, and sample storage. Developed a plan for, and executed, the long term sample storage for efficient, safe, storage. Performed specimen logging and QC, RNA extractions, RT-qPCR, data processing, and conducted biosafety training. Performed the data curation and data analysis for Figure 3.2.

Made Figure 3.2. Minor contributor to symptoms data analysis and visualization with JJ and RA for Figure 3.2. Experimental design and RNA extractions of the samples, to Figure 1A and minor contributor to Figure S2A with AER. Managing logistics for the expansion of the BSL-2+ lab space with AER and biohazardous waste pickups. Collaborated with ES, RA, and AVW to generate data for and curated data set to assess viral RNA stability (Figure 3.S3). Prepared Figure 3.S4. Co-wrote the manuscript. Verified the underlying data with AVW.

Noah W. Schlenker (NWS): Study coordinator; collaborated with NS, AVW, JAR, and RFI on recruitment strategies; enrolled and maintained study participants with NS and JAR; study-data quality control, curation and archiving with RA, JJ, and NS.

Natasha Shelby (NS): Study administrator; collaborated with AVW, MMC, RFI, YG, MF on initial study design and recruitment strategies; co-wrote IRB protocol and informed consent with AVW and MMC; co-wrote enrollment questionnaire with AVW and JJ; co-wrote participant informational sheets with AVW and JAR; enrolled and maintained study participants with JAR and NWS; study-data quality control, curation and archiving with RA, JJ, and NWS; reagents and supplies acquisition; assembled Table S1; managed citations and reference library; co-wrote and edited the manuscript.

Colten Tognazzini (CT): Coordinated the recruitment efforts at PPHD with case investigators and contact tracers; provided guidance and expertise on SARS-CoV-2 epidemiology and local trends.

Alexander Viloria Winnett (AVW): Collaborated with MMC, NS, RFI, YG, MF on initial study design and recruitment strategies; co-wrote IRB protocol and informed consent with MMC and NS; co-wrote enrollment questionnaire with NS and JJ; co-wrote participant informational sheets with NS and JAR and digital survey; developed and validated methods for saliva and nasal-swab sample collection; developed and validated methods for RT-qPCR and RT-digital droplet PCR analysis for saliva and swab samples with AER, ESS, MMC;

reagents and supplies acquisition; funding acquisition; developed laboratory sample processing workflow with AER, ESS, and MMC; performed specimen logging and QC, nucleic acid extraction, RT-qPCR, data processing – including experimental data generation for saliva calibration curve (Figure 3.1, Figure 3.S2) designed with MMC and AER, establishment of nasal swab limit of detection (Fig. S1), and viral load timeseries data (Figure 3.2) with ESS, AER, MKP, and AMC; interpreted sequencing data with AER; analyzed viral load timeseries data to visualize trends (Figure 3.3) with ESS; generated, analyzed and visualized data to assess degradation of viral RNA in saliva and nasal swab samples (Figure 3.S3) with RA, ESS, and AER; literature analysis with RA, ESS, and MKP; co-wrote sections of the manuscript outlined by ESS and RFI, edited the manuscript. Verified the underlying data with ESS.

Abstract

Early detection of SARS-CoV-2 infection is critical to reduce asymptomatic and presymptomatic transmission, curb the spread of variants, and maximize treatment efficacy. Low-analytical-sensitivity nasal-swab testing is commonly used for surveillance and symptomatic testing, but the ability of these tests to detect the earliest stages of infection has not been established. In this study, conducted between September 2020 and June 2021 in the greater Los Angeles County, California, area, initially SARS-CoV-2 negative household contacts of individuals diagnosed with COVID-19 prospectively self-collected paired anterior-nares nasal-swab and saliva samples twice daily for viral-load quantification by high-sensitivity reverse-transcription quantitative PCR (RT-qPCR) and digital-RT-PCR assays. We captured viral-load profiles from the incidence of infection for seven individuals and compared diagnostic sensitivities between respiratory sites. Among unvaccinated persons, testing saliva with a high-analytical-sensitivity assay detected infection up to 4.5 days before viral loads in nasal swabs reached concentrations detectable by low-analytical-sensitivity nasal-swab tests. For most participants, nasal swabs reached higher peak viral loads than saliva but were undetectable or at lower loads during the first few days of infection. High-analytical-sensitivity saliva testing was most reliable for earliest detection. Our study illustrates the value of acquiring early (within hours after a negative high-sensitivity test) viral-load profiles to guide the appropriate analytical sensitivity and respiratory site for detecting earliest infections. Such data are challenging to acquire but critical to designing optimal testing strategies with emerging variants in the current pandemic and to respond to future viral pandemics.

Introduction

Early detection of SARS-CoV-2 infection is needed to reduce asymptomatic and presymptomatic transmission, including the introduction and spread of new viral variants. More than half of transmission events occur from presymptomatic or asymptomatic persons (1). Early detection enables individuals to isolate sooner, reducing transmission within households and local communities and to vulnerable populations. Rapid antigen or molecular

tests performed on nasal swabs are common for both SARS-CoV-2 screening and symptomatic testing (2) but can have low analytical sensitivity compared with lab-based molecular tests. As new variants of concern emerge with increased transmissibility (3–5), high viral loads (4, 6), and breakthrough infections (7), these testing strategies (analytical sensitivity and sample type) need to be assessed and adjusted to ensure detection of early infection. It is still unclear which testing strategy can detect SARS-CoV-2 infection at the earliest stages. Does one need a high-sensitivity test, or would a test with low analytical sensitivity suffice? Which sample site contains detectable virus first?

Tests with high analytical sensitivity can detect low levels of molecular components of the virus (e.g., RNA or proteins) in a sample. Analytical sensitivity is described by the limit of detection (LOD) of a test (defined as the lowest concentration of the viral molecules that produces 95% or better probability of detection). LOD of SARS-CoV-2 diagnostic tests are described in various units; the most directly comparable among tests are units that report the number of viruses (viral particles) or viral RNA copies per milliliter of sample. Viral RNA copies/mL are roughly equivalent to genome copy equivalents/mL (GCE/mL) or nucleic-acid detectable units/mL (NDU/mL). These LOD values are tabulated by the U.S. Food and Drug Administration (FDA) and vary by ≥5 orders of magnitude between tests (8). Tests with high analytical sensitivity have LOD values equivalent to $\sim 10^2$ to 10^3 copies/mL of sample, whereas tests with low analytical sensitivity have LOD values equivalent to $\sim 10^5$ to 10^7 copies/mL (9–12). Importantly, test types (e.g., reverse transcription-quantitative PCR [RT-qPCR], antigen) are often incorrectly equated with a certain analytical sensitivity, despite an FDA analysis (8) demonstrating that the sensitivity of different RT-qPCR tests ranges from highly sensitive (e.g., LOD of 180 NDU/mL for PerkinElmer and 450 NDU/mL for Zymo Research) to substantially less sensitive (e.g., LOD of 180,000 NDU/mL for TaqPath COVID-19 combo kit and 540,000 NDU/mL for Lyra Direct SARS-CoV-2 assay). The low end of this range (corresponding to the higher LOD values) overlaps with the range of low-analytical-sensitivity rapid isothermal nucleic acid tests (e.g., LOD of 180,000 NDU/mL for Atila BioSystems and 300,000 NDU/mL for Abbott ID NOW tests) and approaches the analytical sensitivity range of antigen tests (9, 10). To choose the appropriate

test for reliable early detection, one needs to measure viral loads present in samples collected early in the course of infection (13) and then choose a test with an LOD below that viral load. Initial data by us (14) and others (15, 16) show that, at least in some humans, SARS-CoV-2 viral load can be low (in the range of 10^3 to 10^5 copies per mL of saliva sample) early in infection; therefore, only tests with high analytical sensitivity would reliably detect virus in saliva.

Sampling site or specimen type may also be critical to early detection. Other respiratory viruses have been shown to have detection rates that vary by sampling site (17), which have occasionally been linked to viral tropism. For example, the cellular receptor for entry of Middle East respiratory syndrome coronavirus (MERS-CoV) is expressed nearly exclusively in the lower respiratory tract, prompting recommendations for diagnostic testing of specific sample types (bronchoalveolar lavage fluid, sputum, and tracheal aspirates) (18). A previous study on SARS-CoV found high levels of viral RNA in saliva and throat-wash early in the infection course (before the development of lung lesions), suggesting saliva as a promising sample type for early detection (19). Although nasopharyngeal (NP) swab is often considered the gold standard for SARS-CoV-2 detection, it requires collection by a health care worker and is not well tolerated. Furthermore, the performance of NP swabs for early detection of current SARS-CoV-2 variants is unknown. Sample types such as anterior-nares or mid-turbinate nasal swabs (20–23) and saliva (24–27) are more practical, especially for repeated sampling in screening.

To understand the required test sensitivity and the optimal sample type for earliest SARS-CoV-2 detection, we designed a case-ascertained study of household transmission with high-frequency sampling of both saliva and anterior-nares nasal swabs. Building on our earlier work (14), we enrolled individuals ages 6 years and older who had recently tested positive (household index case) and their exposed household contacts at risk of infection. Negative samples preceding the first positive result are needed to confirm that a participant is within the first days of detectable SARS-CoV-2 RNA. All participants self-collected saliva and anterior-nares nasal swabs twice daily, in the morning upon waking and before bed.

Importantly, all samples were immediately placed in a guanidinium-based inactivating and RNA-stabilizing solution (see Materials and Methods) Samples were screened for SARS-CoV-2 *N1* and *N2* gene positivity using a high-sensitivity assay. When a transmission event was observed (a previously SARS-CoV-2-negative participant tested positive in at least one sample type), we quantified viral loads in all samples prospectively collected from that participant for at least 2 weeks from their first positive result. Quantification was performed via quantitative reverse-transcription PCR (RT-qPCR), with a subset of measurements validated by reverse-transcription droplet digital PCR (RT-ddPCR), capturing the early and full course of acute SARS-CoV-2 infection with high sensitivity.

Materials and Methods

Refer to the supplemental material for detailed methods.

Questionnaires and sample collection

Acquisition of participant data was performed as described previously (14). Symptoms (including those listed by the Centers for Disease Control and Prevention [CDC]) were reported by participants twice daily at the time of sample collection (28).

Participants self-collected nasal-swab and saliva samples in the Spectrum SDNA-1000 Saliva Collection Kit (Spectrum Solutions LLC, Draper, UT, USA), which contains 1.5 mL of liquid buffer, at home twice per day (after waking up and before going to bed), per the manufacturer's guidelines. A parent or legal guardian assisted all minors with collection and were instructed to wear a face covering during supervision.

Samples were stored at 4°C and equilibrated to room temperature before being processed with extraction protocols.

RNA extraction and nucleic acid quantification

Participant saliva and anterior-nares swab samples were extracted using the KingFisher Flex

96 instrument (Thermo Fisher Scientific) with the MagMax Viral Pathogen I Nucleic Acid Isolation Kit (catalog [cat.] no. A42352; Applied Biosystems, Waltham, MA, USA) guided by Thermo Fisher technical notes for SARS-CoV-2 modification and saliva.

RT-qPCR was performed as previously described (14) using the CDC 2019-novel coronavirus (2019-nCoV) real-time RT-PCR diagnostic panel (29), with duplicate reactions. See the supplemental material methods for establishing the extraction to RT-qPCR assay workflow LOD of 1,000 copies/mL (Figures 3.S1A and B in the supplemental material).

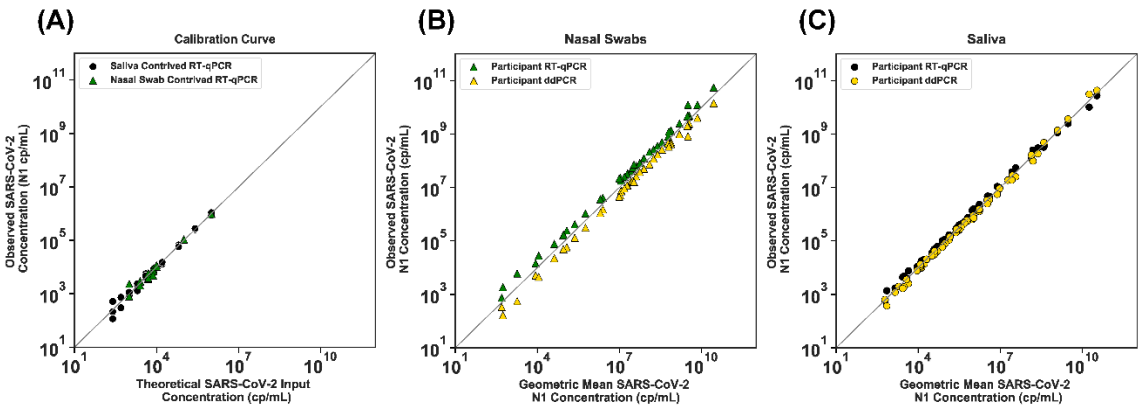


Figure 3.1: SARS-CoV-2 viral-load quantification measured with RT-ddPCR and RT-qPCR. (A) Calibration curves were prepared with contrived saliva and nasal-swab samples. The theoretical SARS-CoV-2 concentration was calculated from a dilution series of contrived samples that were prepared using commercial, inactivated SARS-CoV-2 and commercially available SARS-CoV-2-negative saliva (black circles) or nasal fluid (green triangles) and run with the CDC SARS-CoV-2 RT-qPCR assay. Detailed calibration curves are shown in Figure 3.S2. (B) Participant nasal-swab or (C) saliva samples positive for SARS-CoV-2 RNA at a range of viral loads were selected. SARS-CoV-2 *N1* concentrations (copies/mL) by detection method of RT-ddPCR (gold triangles in panel B, gold circles in panel C) and RT-qPCR (green triangles in panel B, black circles in panel C) are plotted against the geometric mean of RT-qPCR and RT-ddPCR viral-load concentrations (the square root of the product of the two viral-load measurements). A total of 42 nasal-swab and 63 saliva samples from study participants were quantified with both methods. The gray line represents $x = y$. See the supplemental material methods for details of contrived samples, calibration curves, and calculations.

For samples defined as positive by assay guidelines from the CDC (29), viral load was quantified by conversion of the mean quantification cycle (C_q) of duplicate RT-qPCRs using the equations obtained from calibration curves of contrived samples—healthy human saliva or nasal fluid spiked with heat-inactivated SARS-CoV-2 particles. See the supplemental material methods for additional details.

Quantification was also performed by reverse-transcription droplet digital PCR (RT-ddPCR) on both the calibration curve samples (Figures 3.1, Figures 3.S2) and participant samples (Figure 3.1) using the Bio-Rad SARS-CoV-2 droplet digital PCR kit (cat. no. 12013743; Bio-Rad). Droplets were created using the QX200 Droplet Generator (cat. no. 1864002; Bio-Rad); thermocycling was performed on a Bio-Rad C1000 and detected using the QX200 droplet digital PCR system (cat. no. 1864001; Bio-Rad). Samples were analyzed with QuantaSoft analysis Pro 1.0.595 software following Bio-Rad's research-use only (RUO) SARS-CoV-2 guidelines (30).

Viral sequencing

Saliva and nasal-swab samples with an N1 gene C_q of below 26 were sent to Chan Zuckerberg Biohub for SARS-CoV-2 viral genome sequencing, a modification of Deng et al. (31) as described in Gorzynski et al. (32). Sequences were assigned pangolin lineages described by Rambaut et al. (33) using Phylogenetic Assignment of Named Global outbreak LINeages software 2.3.2 (github.com/cov-lineages/pangolin). Chan Zuckerberg Biohub submitted the resulting genomes to GISAID.

Data availability

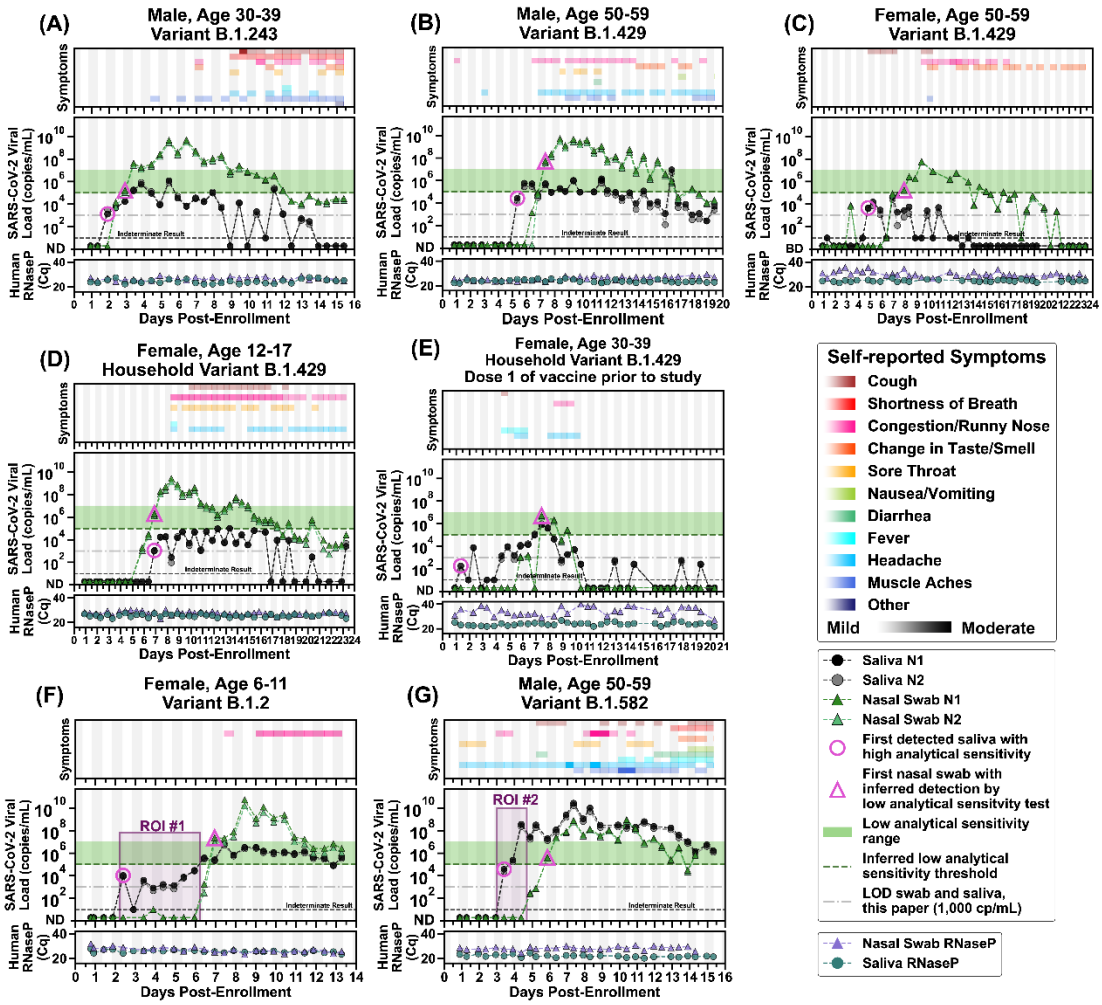
Data are available on CaltechDATA at https://data.caltech.edu/records/1942.

Results

We first established and validated two independent quantitative assays to measure SARS-CoV-2 viral load, an RT-qPCR based on the assay put forth by the U.S. Centers for Disease

Control and Prevention (CDC) (29) and an RT-ddPCR assay developed by Bio-Rad (30). Both of these assays received an emergency use authorization (EUA) for qualitative, but not quantitative, detection of SARS-CoV-2. We optimized the extraction and each quantitative assay protocol (see supplemental material methods) to obtain more reliable quantification of SARS-CoV-2 viral load. The LOD of the modified assay was determined to be 1,000 copies/mL or better by following FDA guidelines (see Materials and Methods; Figure 3.S1). Commercial, heat-inactivated SARS-CoV-2 virus was used to establish calibration curves to convert RT-qPCR quantification cycle values (C_q) to viral load (Figure 3.1A; full details in Figure 3.S2 and supplemental material methods). The linearity of these calibration curves was assessed with 43 participant nasal-swab (Figure 3.1B) and 63 participant saliva samples (Figure 3.1C) across a wide dynamic range of viral loads.

[Figure on next page] Figure 3.2: Symptoms and SARS-CoV-2 viral loads in paired saliva and nasal-swab samples of seven participants who became SARS-CoV-2 positive during study participation. (A to G) Self-reported twice-daily symptom data over the course of enrollment are shown as the top panel for each of the participants (see color-coded legend for symptom categories). Details of symptoms are included in the raw data files. Demographic data including any reported medical conditions are included in Table S1. Viral loads are reported for the *N1* and *N2* genes of SARS-CoV-2 for both saliva (black and gray circles) and nasal-swab samples (dark-green and light-green triangles); ND, not detected for C_qs of ≥40. Samples with an indeterminate result by the CDC RT-qPCR assay are shown along the horizontal black dashed line (see Materials and Methods for details). The limit of detection (LOD) of the assay used for high-analytical-sensitivity measurements is shown with a horizontal gray dashed line. The inferred low-analytical-sensitivity threshold (1.0×10^5 copies/mL) is indicated by the horizontal green dashed line; the low-analytical-sensitivity range (horizontal green bar) is shown. A diagnostic test does not provide reliable detection for samples with viral loads below its LOD. For each participant, the first detected saliva point is emphasized with a pink circle (high analytical sensitivity), and the first nasal-swab point with a viral-load concentration at or above 1.0×10^5 copies/mL (low-analytical-sensitivity threshold) is emphasized with a pink triangle. Vertical shading in gray indicates nighttime (8 pm to 8 am). Internal controls of *RNase P* gene C_qs from the CDC primer set are provided for each sample to compare self-sampling consistency and sample integrity (failed samples, where *RNase P* C_q is ≥40, are not plotted). Participant sex, age range, and SARS-CoV-2 variant are given in each panel's title. Two regions of interest (ROI) are indicated by purple-shaded rectangles and discussed in the main text.


(A)
Male, Age 30-39
Variant B.1.243
(B)
Male, Age 50-59
Variant B.1.429
(C)
Female, Age 50-59
Variant B.1.429
(D)
Female, Age 12-17
Household Variant B.1.429
(E)
Female, Age 30-39
Household Variant B.1.429
Dose 1 of vaccine prior to study
(F)
Female, Age 6-11
Variant B.1.2
(G)
Male, Age 50-59
Variant B.1.582
Symptoms
SARS-CoV-2 Viral Load (copies/mL)
Human RNaseP (Cq)
Indeterminate Result
ROI #1
ROI #2
Days Post-Enrollment
Self-reported Symptoms
Cough
Shortness of Breath
Congestion/Runny Nose
Change in Taste/Smell
Sore Throat
Nausea/Vomiting
Diarrhea
Fever
Headache
Muscle Aches
Other
Mild
Moderate
Saliva N1
Saliva N2
Nasal Swab N1
Nasal Swab N2
First detected saliva with high analytical sensitivity
First nasal swab with inferred detection by low analytical sensitivity test
Low analytical sensitivity range
Inferred low analytical sensitivity threshold
LOD swab and saliva, this paper (1,000 cp/mL)
Nasal Swab RNaseP
Saliva RNaseP

Next, to quantify the viral load at the earliest stage of infection, we analyzed the viral loads in the saliva and nasal swabs of participants who were negative in both sample types upon enrollment and became positive during their participation in the study (Figure 3.2). We extended each participant's enrollment in our study to acquire 14 days of paired saliva and nasal-swab samples starting from the first positive sample. The data in Figure 3.2 report the viral-load concentrations as measured on the day of extraction. All samples were stored at 4°C before extraction; time of storage varied between 0 and 27 days. The stability of SARS-CoV-2 RNA and impact on our conclusions is discussed in the supplemental material methods, Figure 3.S3, and Figure 3.S4.

Here, we report complete viral-load curves in saliva and anterior-nares nasal swabs from seven individuals (Figure 3.2). Each participant tested negative (ND, not detected; Figure 3.2) in both saliva and nasal swabs upon study enrollment, demonstrating that we captured the earliest days of infection. *RNase P* C_q values remained consistent throughout the collection period for saliva and swabs for most participants (Figure 3.2A, B, D, F, and G), indicating that observed changes in viral loads were likely not a sampling artifact but reflected the underlying biology of the infection. Because nasal swabs are commonly used with tests of low analytical sensitivity, and because such tests are proposed to be utilized for SARS-CoV-2 serial screening testing (34, 35), we wanted to compare whether low-analytical-sensitivity testing with nasal swabs could provide equivalent performance to high-analytical-sensitivity testing with saliva (26, 36, 37). We did not run any tests with low analytical sensitivity; our quantitative viral-load measurements were used to infer the performance of a test with an LOD representing low analytical sensitivity. When viral loads in nasal swabs crossed a threshold of 1.0×10^5 copies/mL, entering the low-analytical-sensitivity range, shown as the inferred low-analytical-sensitivity threshold (Figure 3.2), we marked the sample with a pink triangle.

In six out of seven participants, high-analytical-sensitivity saliva testing would have been superior for early detection of SARS-CoV-2 infection compared with the predicted performance of nasal-swab tests with low analytical sensitivity. This prediction was made by

evaluating when nasal-swab viral loads entered the LOD range of nasal-swab tests with low analytical sensitivity. In the seventh participant, the first positive high-analytical-sensitivity saliva test was detected at the same time point that the first nasal-swab test reached a viral load likely to be detected by a low-analytical-sensitivity nasal-swab test (Figure 3.2D). In the first participant (Figure 3.2A), detection occurred first in saliva at low viral load (1.3×10^3 copies/mL *N1* gene, pink circle), while the nasal swab remained negative, and days before the participant reported any symptoms. As measured, viral load in nasal-swab samples reached the level of LOD of low-analytical-sensitivity tests 1.0 days after the first saliva positive samples (pink triangle). This same pattern of earlier detection in high-sensitivity saliva was observed in five of the other six participants; high-sensitivity saliva was 2.5 days earlier (Figure 3.2B), 3.0 days earlier (Figure 3.2C), 6.0 days earlier (Figure 3.2E), 4.5 days earlier (Figure 3.2F), and 2.5 days earlier (Figure 3.2G). The maximum delay in detection between saliva and nasal swab in an unvaccinated person was observed in the youngest participant in our study (see region of interest [ROI] no. 1 of Figure 3.2F). This participant had detectable but low viral load (10^3 to 10^4) in saliva for 4 days, while nasal swabs remained negative by high-sensitivity measurements. Nasal viral loads spiked above 10^{10} copies/mL while the participant's only symptoms were mild congestion/runny nose.

Even with high-analytical-sensitivity nasal-swab testing, only one participant tested positive in nasal swab before saliva (Figure 3.2D). In this participant, SARS-CoV-2 RNA was detectable with a high-analytical-sensitivity nasal swab 1 day before it was detectable in a high-analytical-sensitivity saliva test. Nasal swabs reached the detection range of low-analytical-sensitivity tests (pink triangle) on the same day as the first saliva sample was detected by high-analytical sensitivity testing (pink circle). For all seven participants, high-analytical-sensitivity saliva testing would have detected SARS-CoV-2 RNA either the same day or up to 6 days before viral loads in nasal swab reached the detection limits of low-sensitivity nasal-swab tests.

Two participants (Figure 3.2C and E) had low viral load in both saliva and nasal swabs. Their viral-load measurements were near the LOD of our assay, and therefore, as expected, many

samples from these participants had indeterminate results. One participant (Figure 3.2E) had received one dose of the Pfizer-BioNTech COVID-19 vaccine (38) 13 days prior to her first sample, though observations here are not powered to make conclusions about viral load due to vaccination.

Remarkably (see ROI no. 2 in Figure 3.2G), in one participant, saliva viral load spiked to 3.7×10^8 viral copies/mL (*N1* gene target) while SARS-CoV-2 RNA remained undetectable in nasal swab, even by the high-analytical-sensitivity assay used here.

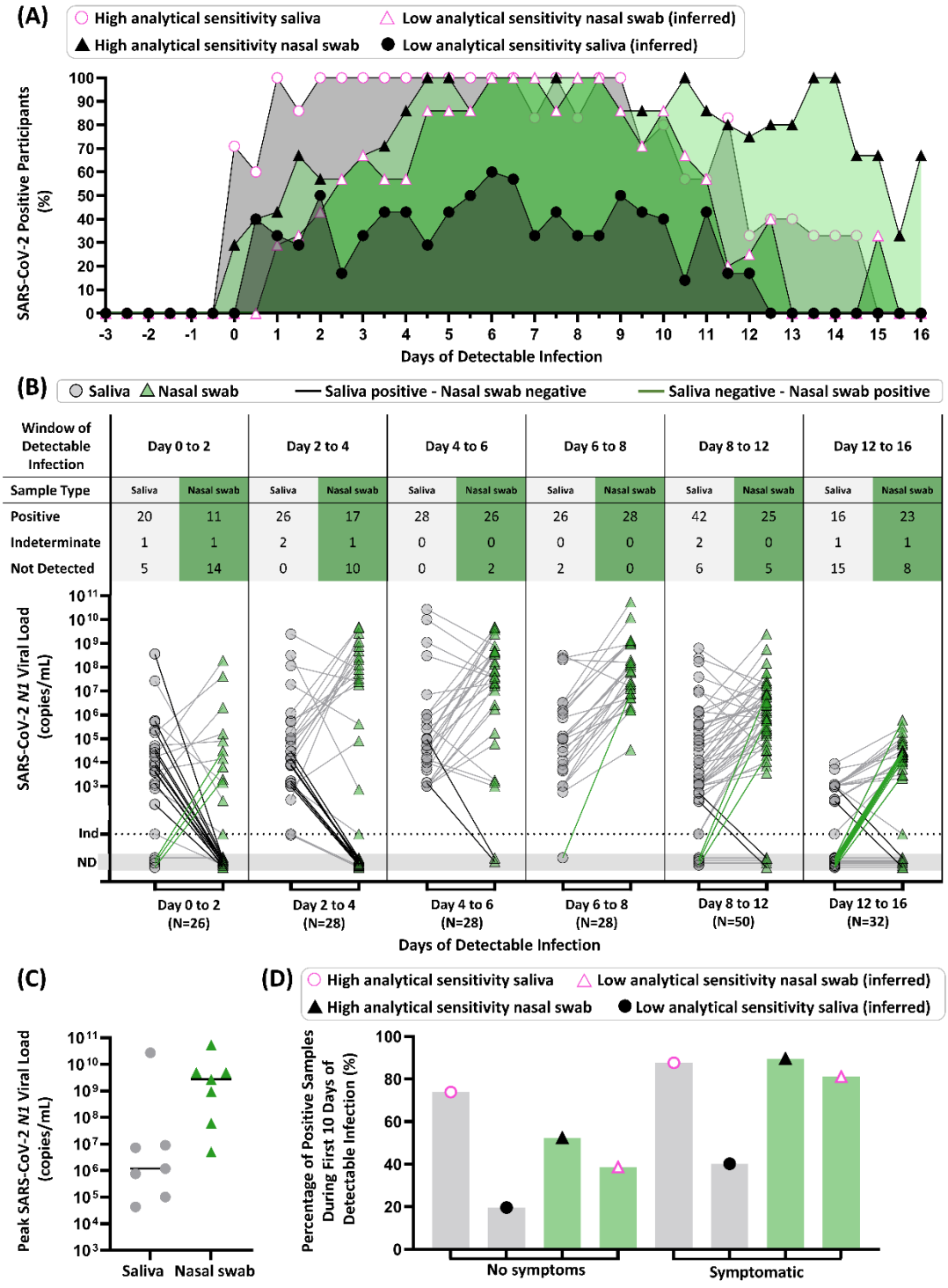
Compiled data from all seven participants highlight the importance of the interplay among anatomical sampling site, infection stage, and diagnostic test sensitivity (Figure 3.3). Participant results were aligned to the first positive result from either sample type (day 0). If a saliva or nasal-swab sample had a SARS-CoV-2 viral load above 1.0×10^5 copies/mL, entering the low-analytical-sensitivity range (39), we inferred that a low-analytical-sensitivity test would have correctly determined that sample to be positive. The percentage of participants with either observed or inferred positive results at each time point (0.5-day intervals) from the first positive sample revealed that high-analytical-sensitivity saliva testing outperformed low-analytical-sensitivity nasal-swab testing for the first 5.5 days of detectable infection (Figure 3.3A) and high-analytical-sensitivity nasal-swab testing during the first 4 days (Figure 3.3A). Analytical sensitivity affects the overall test performance in each sample type. Based on early viral loads in saliva, we inferred that low-sensitivity saliva testing was outperformed by high-sensitivity saliva and both high- and low-sensitivity nasal-swab testing (Figure 3.3A).

Next, we plotted paired viral loads in each respiratory site starting from the first positive test (Figure 3.3B). From day 0 to day 6, using high-sensitivity testing for both sample types, saliva was more frequently positive than nasal swabs (Figure 3.3B). Comparison of paired samples between day 6 and day 12 for both sample types showed highly concordant detection. In a later time interval, between days 12 and 16, nasal swabs were more frequently positive than saliva (Figure 3.3B). The median of peak viral loads was higher in nasal swabs

than saliva (Figure 3.3C), consistent with the literature (21, 23, 40).

Many testing strategies and decisions are based on the presence or absence of symptoms (2, 41). We considered the positivity rate of high- or low-analytical-sensitivity testing methods with each sample type during the first ten days of test-positive infection (to capture the presymptomatic and symptomatic phases of infection for this cohort, not the postsymptomatic phase), separating them into categories of no symptoms or symptomatic if the participant reported at least one COVID-19-like symptom (Figure 3.3D). For samples collected while participants were asymptomatic, high-sensitivity saliva testing was more effective (74% positivity) than high- (52%) or low-sensitivity (39%) nasal-swab testing and low-sensitivity saliva testing (20%). In contrast, during symptomatic phases, which are often concurrent with peak nasal viral loads (Figure 3.2), high-sensitivity saliva (88%) and high-sensitivity nasal-swab testing (89%) have similar positivity rates (Figure 3.3D). Additionally, based on our measured viral loads, low-sensitivity nasal-swab testing is predicted to perform better in symptomatic cohorts (81%) than in asymptomatic persons (39%), consistent with how these tests were originally authorized.

These data reveal a more nuanced view than “saliva is better than swab”. Using tests with high analytical sensitivity, SARS-CoV-2 RNA is more detectable in saliva than nasal swab during the early phase of the infection (Figure 3.3B). However, because viral loads in saliva generally remained lower than those in nasal swabs (Figure 3.3C), we infer that positivity by a low-analytical-sensitivity saliva test would be outperformed by both high- and low-analytical-sensitivity nasal-swab testing (Figure 3.3A), independent of symptom status (Figure 3.3D). It was the combination of test analytical sensitivity along with sample type that determined the overall test performance.


(A)
High analytical sensitivity saliva
Low analytical sensitivity nasal swab (inferred)
High analytical sensitivity nasal swab
Low analytical sensitivity saliva (inferred)
SARS-CoV-2 Positive Participants (%)
Days of Detectable Infection
(B)
Saliva
Nasal swab
Saliva positive - Nasal swab negative
Saliva negative - Nasal swab positive
Window of Detectable Infection
Day 0 to 2
Day 2 to 4
Day 4 to 6
Day 6 to 8
Day 8 to 12
Day 12 to 16
Sample Type
Saliva
Nasal swab
Positive 20 11 26 17 28 26 26 28 42 25 16 23
Indeterminate 1 1 2 1 0 0 0 0 2 0 1 1
Not Detected 5 14 0 10 0 2 2 0 6 5 15 8
SARS-CoV-2 N1 Viral Load (copies/mL)
Ind
ND
Day 0 to 2 (N=26)
Day 2 to 4 (N=28)
Day 4 to 6 (N=28)
Day 6 to 8 (N=28)
Day 8 to 12 (N=50)
Day 12 to 16 (N=32)
Days of Detectable Infection
(C)
Peak SARS-CoV-2 N1 Viral Load (copies/mL)
Saliva
Nasal swab
(D)
High analytical sensitivity saliva
Low analytical sensitivity nasal swab (inferred)
High analytical sensitivity nasal swab
Low analytical sensitivity saliva (inferred)
Percentage of Positive Samples During First 10 Days of Detectable Infection (%)
No symptoms
Symptomatic

[Figure on previous page] Figure 3.3: Summary of diagnostic insights from study participants who became infected with SARS-CoV-2 while enrolled in the study. (A) Participant infection time courses were aligned to the first high-sensitivity (LOD of $\leq 1 \times 10^3$ copies/mL) positive result from either saliva or nasal-swab sample type (day 0), and the percentage of positive tests was calculated for each time point (0.5-day intervals) from the first positive sample. The predicted performance of nasal swabs or saliva with low analytical sensitivity was determined using the SARS-CoV-2 *N1* viral-load values for each participant shown in Figure 3.2, above a viral-load threshold of 1.0×10^5 copies/mL, entering the low-analytical-sensitivity range. We show the percentage of participants who were detected by our high-analytical-sensitivity saliva test (pink circle), high-analytical-sensitivity nasal-swab test (black triangle), or that could be inferred to be detectable by a low-analytical-sensitivity test nasal-swab (pink triangle) or saliva test (black circle) at a given time point. (B) Quantitative SARS-CoV-2 *N1* viral loads of paired samples collected during time windows of the infection (aligned to first positive result by high-sensitivity testing of either sample type) are shown for saliva (gray circles) and nasal swabs (green triangles). Paired samples for a given time point are connected with gray lines, with emphasis on paired samples where only saliva (black connecting line) or nasal swab (green connecting line) was positive. ND, not detected; Ind, indeterminate result. (C) Peak SARS-CoV-2 *N1* viral loads measured in saliva (gray circles) and nasal swab (green triangles) for each of the seven participants are shown. Horizontal black line indicates the median. (D) Percentage of positive samples (out of all samples of that type and symptom status) are shown for the first ten days of detectable infection for each participant. Saliva (gray bars with circles) and nasal swab (green bars with triangles) are shown. Positivity by a high-analytical-sensitivity test was observed by our assay, while positivity by a low-analytical-sensitivity test was inferred if the sample had a viral load above 1.0×10^5 copies/mL. The symptom status was classified as symptomatic if the participant reported one or more COVID-19-like symptoms at the time of sample collection. Details of the data analysis are included in the supplemental material methods.

Discussion

Limitations

Our study needs to be interpreted in the context of its limitations. First, our results capture viral-load dynamics from a limited number of individuals from one region of one country with limited SARS-CoV-2 diversity. Follow-up studies with a larger sample size, including individuals of diverse ages, genetic backgrounds, medical conditions, COVID-19 severity, and SARS-CoV-2 lineages would be ideal to provide a more nuanced and representative understanding of viral dynamics in saliva and nasal-swab samples. Second, the commercial inactivating buffer used here (Spectrum SDNA-1000) is not authorized (at the time of this writing) for the sample collection of nasal swabs. The solution in the Spectrum SDNA-1000 kits is a guanidinium-based inactivating and stabilizing buffer that preserves viral RNA but

eliminated the opportunity to also perform viral culture. Third, we have paired data for saliva and anterior-nares nasal swabs but do not compare nasopharyngeal (NP) swabs, sputum, or other lower-respiratory specimens. We do not know whether other sampling sites, such as NP swabs or oropharyngeal swabs, would have provided earlier or later detection than saliva. Fourth, we are inferring the ability of tests with low analytical sensitivity to detect infections based on the quantified viral load in the participant samples and the LODs reported by the FDA for the diagnostic platforms. Fifth, some degradation may have occurred in some samples (see supplemental material for a complete analysis of RNA stability). Sixth, all samples were self-collected, which may result in lower-quality specimens.

Conclusions

By rapidly enrolling household members at high risk for contracting COVID-19 and having them self-sample twice daily in paired respiratory sites, we observed patterns in SARS-CoV-2 viral load in the earliest days of infection. All seven participants tested negative in saliva and nasal swabs upon enrollment, demonstrating that we captured the earliest detectable SARS-CoV-2 viral load (within 12 h) in both sample types. Our data set helps inform diagnostic testing strategies by showing differences in viral loads in paired nasal swabs and saliva samples at high temporal resolution during the early days and presymptomatic phases of infection.

We made five conclusions from our study.

First, choosing the correct respiratory sampling site is critical for earliest detection of SARS-CoV-2 infection. In our study, alignment of longitudinal data to the first day of positivity clearly shows the superiority of high-sensitivity saliva testing for detection in the first 5.5 days of infection (Figure 3.3A). Given our data, early infection viral-load dynamics in multiple sampling sites should be investigated and compared with saliva as new SARS-CoV-2 variants emerge.

Second, our data explain the conflicting results in the literature comparing test performance

in paired respiratory sites, with some studies showing that nasal swabs outperform saliva (21, 23, 40) and others showing that saliva (or oral fluid) has detection equivalent to or better than that of nasal swabs (16, 25, 42–50). Through longitudinal rather than cross-sectional sampling, we show that the relative viral loads in each respiratory site are a factor of infection stage (shown in time intervals in Figure 3.3B), and the kinetics of viral load may be quite distinct in each sample type for an individual (Figure 3.2). Most studies examining paired sample types enrolled participants after a positive test or symptom onset; as our data show, detectable viral loads precede symptoms, in most cases (5 of 7 participants) by several days (Figure 3.2).

Third, peak viral load measured in nasal swabs (Figure 3.3C) is not representative of detectable viral load in the earliest days of infection (Figure 3.3A) or during the presymptomatic phase (Figure 3.3D). Early in an infection, it is inappropriate to assume that a person is "not infectious" or "has low viral load" based on a measurement from a single sample type, such as a nasal swab, given that saliva is known to carry infectious virus (51). In our study, we observed a participant with very high ($>10^7$ to 10^8 copies/mL) viral load in saliva samples while the paired nasal swab was either negative (Figure 3.2G, ROI no. 2) or had low ($\sim 10^3$ copies/mL) viral load (Figure 3.2G, day after ROI no. 2). Quantitative SARS-CoV-2 culture from paired saliva and swab samples is still needed to understand infectiousness during the early stages of SARS-CoV-2 infection.

Fourth, using a diagnostic test with high analytical sensitivity (Figure 3.3D), rather than a test of a particular detection method (RT-qPCR, antigen, next-generation sequencing, etc.), is essential to early detection. With many strategies for asymptomatic screening/surveillance testing in use, it is critically important to consider whether the LODs of the tests would be able to detect early infection and to prompt actions that minimize transmission.

Fifth, our data show the utility of combining knowledge of the appropriate respiratory site and the appropriate test analytical sensitivity for achieving earliest detection. Among our unvaccinated participants, when a high-sensitivity test was combined with saliva as a sample

type, SARS-CoV-2 infection was detected up to 4.5 days before viral loads in nasal swabs reached the LODs of low-analytical-sensitivity tests (Figure 3.2F). Although high-sensitivity saliva testing was usually able to detect virus earlier than nasal swabs (Figure 3.3A), during the peak of the infection viral loads in nasal swabs were usually higher than in saliva (Figure 3.3C). Furthermore, SARS-CoV-2 was detected in saliva with high-sensitivity methods, and the viral loads were low (Figure 3.2, 3.3A and D); low-sensitivity saliva tests would likely not have been able to detect these infections early. These observations support the preferred use of nasal swabs in environments where only low-sensitivity testing is available, although the performance of such testing for early detection is poor (Figure 3.3D). These observations also show that the optimal respiratory sampling site is nuanced and depends on the phase of the infection being detected (early versus peak) and on the analytical sensitivity of the test being used with each sample type.

Our work suggests four steps to improve the effectiveness of diagnostic tests in early detection and preventing transmission of SARS-CoV-2 as new variants emerge and as infections spread to additional segments of the global population. (i) Additional longitudinal studies are needed that include high-frequency collection from multiple respiratory sites using quantitative assays with high analytical sensitivity. (ii) Policy makers need to use such quantitative data to revise and optimize screening testing guidelines to ensure early detection of SARS-CoV-2 infections and reduction of transmission. (iii) Innovation is needed to produce rapid point-of-care tests with high analytical sensitivity for a range of sample types (including saliva) at a price point to enable global distribution. (iv) Quantitative studies of the kinetics of early stage viral loads in each respiratory site (collected in parallel with viral culture data) must be updated with the emergence of new SARS-CoV-2 variants.

We hope our data, important work by others in this area (15, 16, 51, 52), and future quantitative studies of early viral-load kinetics will lead to improved testing strategies to combat the current COVID-19 pandemic. The methodology for performing such studies efficiently and quickly will likely be extendable to defining strategies for early detection of causative pathogens in subsequent pandemics.

Acknowledgements

We thank Lauriane Quenee, Junie Hildebrandt, Grace Fisher-Adams, RuthAnne Bevier, Chantal D'Apuzzo, Ralph Adolphs, Victor Rivera, Steve Chapman, Gary Waters, Leonard Edwards, Gaylene Ursua, Cynthia Ramos, and Shannon Yamashita for their assistance and advice on study implementation and/or administration. We thank Jessica Leong, Jessica Slagle, Mika Walton, Angel Navarro, Daniel Brenner, and Ojas Pradhan for volunteering their time to help with this study, Si Hyung Jin for helping with a literature review, and Mary Arrastia for providing biosafety support. We thank Maira Phelps, Lienna Chan, Lucy Li, Dan Lu, and Amy Kistler at the Chan Zuckerberg Biohub for performing SARS-CoV-2 sequencing. We thank Angie Cheng, Susan Magdaleno, Christian Kis, Monica Herrera, and Zaina Lemeir for technical discussions regarding saliva extraction and ddPCR detection. We thank Jennifer Fulcher, Debika Bhattacharya, and Matthew Bidwell Goetz for their ideas on potential study populations and early study design. We thank Omai Garner and David Beenhouwer for providing materials for initial nasal-swab validation. We thank Martin Hill, Alma Sanchez, Scott Kim, Debbie Noble, Nina Paddock, Whitney Harrison, Emily Holman, Isaac Turner, Vivek Desai, Luke Wade, Tom Mayell, Stu Miller, and Jennifer Howes for their support with recruitment. We thank Allison Rhines, Karen Heichman, and Dan Wattendorf for valuable discussions. Finally, we thank all the case investigators and contact tracers at the Pasadena Public Health Department and the City of Long Beach Department of Health & Human Services for their efforts in study recruitment and their work in the pandemic response.

R.F.I. is a cofounder, consultant, and a director and has stock ownership of Talis Biomedical Corp. In addition, R.F.I. is an inventor on a series of patents licensed by the University of Chicago to Bio-Rad Laboratories, Inc., in the context of ddPCR.

This study is based on research funded in part by the Bill & Melinda Gates Foundation (INV-023124). The findings and conclusions contained within are those of the authors and do not

necessarily reflect positions or policies of the Bill & Melinda Gates Foundation. This work was also funded by the Ronald and Maxine Linde Center for New Initiatives at the California Institute of Technology and the Jacobs Institute for Molecular Engineering for Medicine at the California Institute of Technology. A.V.W. is supported by a National Institutes of Health NIGMS predoctoral training grant (GM008042) and a UCLA DGSOM Geffen fellowship; M.M.C. is supported by a Caltech graduate student fellowship, and M.K.P. and J.T.B. are each partially supported by a National Institutes of Health Biotechnology Leadership Predoctoral Training Program (BLP) fellowship from Caltech's Donna and Benjamin M. Rosen Bioengineering Center (T32GM112592).

References

1. Johansson, M.A., Quandelacy, T.M., Kada, S., Prasad, P.V., Steele, M., Brooks, J.T., Slayton, R.B., Biggerstaff, M. and Butler, J.C., 2021. SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA network open*, *4*(1), pp. e2035057–e2035057.

2. Recommendations for national SARS-CoV-2 testing strategies and diagnostic capacities (2021) WHO. Available at: https://apps.who.int/iris/bitstream/handle/10665/342002/WHO-2019-nCoV-lab-testing-2021.1-eng.pdf (Accessed: 28 May 2024).

3. SPI-M-O: Consensus statement on covid-19, 30 June 2021 (2021) GOV.UK. Available at: https://www.gov.uk/government/publications/spi-m-o-consensus-statement-on-covid-19-30-june-2021 (Accessed: 28 May 2024).

4. Li, B., Deng, A., Li, K., Hu, Y., Li, Z., Shi, Y., Xiong, Q., Liu, Z., Guo, Q., Zou, L. and Zhang, H., 2022. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. *Nature communications*, *13*(1), p. 460.

5. Fisman, D.N. and Tuite, A.R., 2021. Progressive increase in virulence of novel SARS-CoV-2 variants in Ontario, Canada. *MedRxiv*, pp. 2021–07.

6. Yadav, P.D., Mohandas, S., Shete, A.M., Nyayanit, D.A., Gupta, N., Patil, D.Y., Sapkal, G.N., Potdar, V., Kadam, M., Kumar, A. and Kumar, S., 2021. SARS CoV-2 variant B. 1.617. 1 is highly pathogenic in hamsters than B. 1 variant. *BioRxiv*, pp. 2021–05.

7. Brown, C.M., 2021. Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings—Barnstable County, Massachusetts, July 2021. *MMWR. Morbidity and mortality weekly report*, *70*.

8. SARS-CoV-2 reference panel comparative data (2020) FDA. Available at: https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/sars-cov-2-reference-panel-comparative-data (Accessed: 28 May 2024).

9. Caruana, G., Croxatto, A., Kampouri, E., Kritikos, A., Opota, O., Foerster, M., Brouillet, R., Senn, L., Lienhard, R., Egli, A. and Pantaleo, G., 2021. ImplemeNting SARS-CoV-2 Rapid antigen testing in the Emergency wArd of a Swiss univErsity hospital: the INCREASE study. *Microorganisms*, *9*(4), p. 798.

10. Olearo, F., Nörz, D., Heinrich, F., Sutter, J.P., Roedl, K., Schultze, A., Zur Wiesch, J.S., Braun, P., Oestereich, L., Kreuels, B. and Wichmann, D., 2021. Handling and accuracy of four rapid antigen tests for the diagnosis of SARS-CoV-2 compared to RT-qPCR. *Journal of Clinical Virology*, *137*, p. 104782.

11. Corman, V.M., Haage, V.C., Bleicker, T., Schmidt, M.L., Mühlemann, B., Zuchowski, M., Jo, W.K., Tscheak, P., Möncke-Buchner, E., Müller, M.A. and Krumbholz, A., 2021. Comparison of seven commercial SARS-CoV-2 rapid point-of-care antigen tests: a single-centre laboratory evaluation study. *The Lancet Microbe*, *2*(7), pp. e311–e319.

12. Cubas-Atienzar, A.I., Kontogianni, K., Edwards, T., Wooding, D., Buist, K., Thompson, C.R., Williams, C.T., Patterson, E.I., Hughes, G.L., Baldwin, L. and Escadafal, C., 2021. Limit of detection in different matrices of 19 commercially available rapid antigen tests for the detection of SARS-CoV-2. *Scientific reports*, *11*(1), p. 18313.

13. Suess, T., Remschmidt, C., Schink, S.B., Schweiger, B., Heider, A., Milde, J., Nitsche, A., Schroeder, K., Doellinger, J., Braun, C. and Haas, W., 2012. Comparison of shedding characteristics of seasonal influenza virus (sub) types and influenza A (H1N1) pdm09; Germany, 2007–2011. *PloS one*, *7*(12), p. e51653.

14. Winnett, A., Cooper, M.M., Shelby, N., Romano, A.E., Reyes, J.A., Ji, J., Porter, M.K., Savela, E.S., Barlow, J.T., Akana, R. and Tognazzini, C., 2020. SARS-CoV-2 viral load in saliva rises gradually and to moderate levels in some

humans. *medRxiv*.

15. Kissler, S., Fauver, J.R., Mack, C., Tai, C.G., Breban, M.I., Watkins, A.E., Samant, R.M., Anderson, D.J., Ho, D.D., Grubaugh, N.D. and Grad, Y., 2021. Densely sampled viral trajectories suggest longer duration of acute infection with B. 1.1. 7 variant relative to non-B. 1.1. 7 SARS-CoV-2.

16. Kissler, S.M., Fauver, J.R., Mack, C., Olesen, S.W., Tai, C., Shiue, K.Y., Kalinich, C., Jednak, S., Ott, I., Vogels, C. and Wohlgemuth, J., 2020. Viral dynamics of SARS-CoV-2 infection and the predictive value of repeat testing.

17. Kim, Y.G., Yun, S.G., Kim, M.Y., Park, K., Cho, C.H., Yoon, S.Y., Nam, M.H., Lee, C.K., Cho, Y.J. and Lim, C.S., 2017. Comparison between saliva and nasopharyngeal swab specimens for detection of respiratory viruses by multiplex reverse transcription-PCR. *Journal of clinical microbiology*, *55*(1), pp. 226–233.

18. Laboratory Testing for Middle East Respiratory Syndrome Coronavirus (2018) WHO. Available at: https://apps.who.int/iris/bitstream/handle/10665/259952/WHO-MERS-LAB-15.1-Rev1-2018-eng.pdf?sequence=1 (Accessed: 28 May 2024).

19. Wang, W.K., Chen, S.Y., Liu, I.J., Chen, Y.C., Chen, H.L., Yang, C.F., Chen, P.J., Yeh, S.H., Kao, C.L., Huang, L.M. and Hsueh, P.R., 2004. Detection of SARS-associated coronavirus in throat wash and saliva in early diagnosis. *Emerging infectious diseases*, *10*(7), p. 1213.

20. Ku, C.W., Shivani, D., Kwan, J.Q., Loy, S.L., Erwin, C., Ko, K.K., Ng, X.W., Oon, L., Thoon, K.C., Kalimuddin, S. and Chan, J.K., 2021. Validation of self-collected buccal swab and saliva as a diagnostic tool for COVID-19. *International Journal of Infectious Diseases*, *104*, pp. 255–261.

21. Barat, B., Das, S., De Giorgi, V., Henderson, D.K., Kopka, S., Lau, A.F., Miller, T., Moriarty, T., Palmore, T.N., Sawney, S. and Spalding, C., 2021. Pooled saliva specimens for SARS-CoV-2 testing. *Journal of clinical microbiology*, *59*(3), pp.

10–1128.

22. Manabe, Y.C., Reuland, C., Yu, T., Azamfirei, R., Hardick, J.P., Church, T., Brown, D.M., Sewell, T.T., Antar, A., Blair, P.W. and Heaney, C.D., 2021, February. Self-collected oral fluid saliva is insensitive compared with nasal-oropharyngeal swabs in the detection of severe acute respiratory syndrome coronavirus 2 in outpatients. In *Open Forum Infectious Diseases* (Vol. 8, No. 2, p. ofaa648). US: Oxford University Press.

23. Procop, G.W., Shrestha, N.K., Vogel, S., Van Sickle, K., Harrington, S., Rhoads, D.D., Rubin, B.P. and Terpeluk, P., 2020. A direct comparison of enhanced saliva to nasopharyngeal swab for the detection of SARS-CoV-2 in symptomatic patients. *Journal of clinical microbiology*, *58*(11), pp. 10–1128.

24. Al Suwaidi, H., Senok, A., Varghese, R., Deesi, Z., Khansaheb, H., Pokasirakath, S., Chacko, B., Abufara, I., Loney, T. and Alsheikh-Ali, A., 2021. Saliva for molecular detection of SARS-CoV-2 in school-age children. *Clinical Microbiology and Infection*, *27*(9), pp. 1330–1335.

25. Wyllie, A.L., Fournier, J., Casanovas-Massana, A., Campbell, M., Tokuyama, M., Vijayakumar, P., Warren, J.L., Geng, B., Muenker, M.C., Moore, A.J. and Vogels, C.B., 2020. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. *New England Journal of Medicine*, *383*(13), pp. 1283–1286.

26. Allicock, O.M., Petrone, M.E., Yolda-Carr, D., Breban, M., Walsh, H., Watkins, A.E., Rothman, J.E., Farhadian, S.F., Grubaugh, N.D. and Wyllie, A.L., 2022. Evaluation of saliva self-collection devices for SARS-CoV-2 diagnostics. *BMC Infectious Diseases*, *22*(1), p. 284.

27. Tan, S.H., Allicock, O., Armstrong-Hough, M. and Wyllie, A.L., 2021. Saliva as a gold-standard sample for SARS-CoV-2 detection. *The Lancet Respiratory Medicine*, *9*(6), pp. 562–564.

28. Symptoms of COVID-19 (2020) Centers for Disease Control and Prevention.

Available at: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html (Accessed: 28 May 2024).

29. CDC 2019 nCoV real-time RT-PCR diagnostic panel (2020) FDA. Available at: https://www.fda.gov/media/134922/download (Accessed: 28 May 2024).

30. Bio-Rad SARS-COV-2 ddpcr kit instructions for use (2020) Bio-Rad. Available at: https://www.bio-rad.com/webroot/web/pdf/lsr/literature/10000130776.pdf (Accessed: 28 May 2024).

31. Deng, X., Achari, A., Federman, S., Yu, G., Somasekar, S., Bártolo, I., Yagi, S., Mbala-Kingebeni, P., Kapetshi, J., Ahuka-Mundeke, S. and Muyembe-Tamfum, J.J., 2020. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nature microbiology*, *5*(3), pp. 443–454.

32. Gorzynski, J.E., De Jong, H.N., Amar, D., Hughes, C., Ioannidis, A., Bierman, R., Liu, D., Tanigawa, Y., Kistler, A.L., Kamm, J. and Kim, J., 2020. High-throughput SARS-CoV-2 and host genome sequencing from single nasopharyngeal swabs. *medRxiv*, pp. 2020–07.

33. Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L. and Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature microbiology*, *5*(11), pp. 1403–1407.

34. Larremore, D.B., Wilder, B., Lester, E., Shehata, S., Burke, J.M., Hay, J.A., Tambe, M., Mina, M.J. and Parker, R., 2021. Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Science advances*, *7*(1), p.eabd5393.

35. Mina, M.J., Parker, R. and Larremore, D.B., 2020. Rethinking Covid-19 test sensitivity—a strategy for containment. *New England Journal of Medicine*, *383*(22), p. e120.

36. Bloom, J.S., Sathe, L., Munugala, C., Jones, E.M., Gasperini, M., Lubock, N.B., Yarza, F., Thompson, E.M., Kovary, K.M., Park, J. and Marquette, D., 2020. Swab-

Seq: A high-throughput platform for massively scaled up SARS-CoV-2 testing. *MedRxiv*.

37. Smith, R.L., Gibson, L.L., Martinez, P.P., Ke, R., Mirza, A., Conte, M., Gallagher, N., Conte, A., Wang, L., Fredrickson, R. and Edmonson, D.C., 2021. Longitudinal assessment of diagnostic test performance over the course of acute SARS-CoV-2 infection. *The Journal of infectious diseases*, *224*(6), pp. 976–982.

38. Comirnaty and Pfizer-BioNTech COVID-19 Vaccine (2021) FDA. Available at: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine (Accessed: 28 May 2024).

39. ID now COVID-19 - instructions for use (2021) FDA. Available at: https://www.fda.gov/media/136525/download (Accessed: 28 May 2024).

40. da Costa Fernandes, P.A., da Conceicao Ferreira, F.A., Morais, O.M., Ramos, C.M.T., Fernandes, É.M.R., da Rocha, S.A.A., Rocha, R.J.A., Monteiro, V.J.P., Vilar, P.S.G., Romao, A.M. and Alves, M.R.A., 2021. Performance of saliva as a specimen to detect SARS-CoV-2. *Journal of Clinical Virology*, *142*, p. 104913.

41. Overview of Testing for SARS-CoV-2 (COVID-19) (2021) Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing-overview.html (Accessed: 28 May 2024).

42. Hanson, K.E., Barker, A.P., Hillyard, D.R., Gilmore, N., Barrett, J.W., Orlandi, R.R. and Shakir, S.M., 2020. Self-collected anterior nasal and saliva specimens versus health care worker-collected nasopharyngeal swabs for the molecular detection of SARS-CoV-2. *Journal of clinical microbiology*, *58*(11), pp. 10–1128.

43. De Santi, C., Jacob, B., Kroich, P., Doyle, S., Ward, R., Li, B., Donnelly, O., Dykes, A., Neelakant, T., Neary, D. and McGuinness, R., 2021. Concordance between PCR-based extraction-free saliva and nasopharyngeal swabs for SARS-CoV-2 testing. *HRB Open Research*, *4*.

44. Sasaki, T., Inoue, O., Ogihara, S., Kubokawa, K., Oishi, S., Shirai, T., Iwabuchi, K.

and Suzuki-Inoue, K., 2022. Detection of SARS-CoV-2 RNA using RT-qPCR in saliva samples and nasopharyngeal, lingual, and buccal mucosal swabs. *Japanese Journal of Infectious Diseases*, *75*(1), pp. 102–104.

45. Kojima, N., Turner, F., Slepnev, V., Bacelar, A., Deming, L., Kodeboyina, S. and Klausner, J.D., 2021. Self-collected oral fluid and nasal swabs demonstrate comparable sensitivity to clinician collected nasopharyngeal swabs for coronavirus disease 2019 detection. *Clinical Infectious Diseases*, *73*(9), pp. e3106–e3109.

46. Tu, Y.P., Jennings, R., Hart, B., Cangelosi, G.A., Wood, R.C., Wehber, K., Verma, P., Vojta, D. and Berke, E.M., 2020. Patient-collected tongue, nasal, and mid-turbinate swabs for SARS-CoV-2 yield equivalent sensitivity to health care worker collected nasopharyngeal swabs. *MedRxiv*, pp. 2020–04.

47. Iwasaki, S., Fujisawa, S., Nakakubo, S., Kamada, K., Yamashita, Y., Fukumoto, T., Sato, K., Oguri, S., Taki, K., Senjo, H. and Sugita, J., 2020. Comparison of SARS-CoV-2 detection in nasopharyngeal swab and saliva. *Journal of Infection*, *81*(2), pp. e145–e147.

48. To, K.K.W., Tsang, O.T.Y., Yip, C.C.Y., Chan, K.H., Wu, T.C., Chan, J.M.C., Leung, W.S., Chik, T.S.H., Choi, C.Y.C., Kandamby, D.H. and Lung, D.C., 2020. Consistent detection of 2019 novel coronavirus in saliva. *Clinical Infectious Diseases*, *71*(15), pp. 841–843.

49. Kernéis, S., Elie, C., Fourgeaud, J., Choupeaux, L., Delarue, S.M., Alby, M.L., Quentin, P., Pavie, J., Brazille, P., Néré, M.L. and Minier, M., 2021. Accuracy of antigen and nucleic acid amplification testing on saliva and naopharyngeal samples for detection of SARS-CoV-2 in ambulatory care. *MedRxiv*, pp. 2021–04.

50. Callahan, C., Ditelberg, S., Dutta, S., Littlehale, N., Cheng, A., Kupczewski, K., McVay, D., Riedel, S., Kirby, J.E. and Arnaout, R., 2021. Saliva is comparable to nasopharyngeal swabs for molecular detection of SARS-CoV-2. *Microbiology spectrum*, *9*(1), pp. 10–1128.

51. Huang, N., Pérez, P., Kato, T., Mikami, Y., Okuda, K., Gilmore, R.C., Conde, C.D., Gasmi, B., Stein, S., Beach, M. and Pelayo, E., 2021. SARS-CoV-2 infection of the oral cavity and saliva. *Nature medicine*, *27*(5), pp. 892–903.

52. Ke, R., Martinez, P.P., Smith, R.L., Gibson, L.L., Mirza, A., Conte, M., Gallagher, N., Luo, C.H., Jarrett, J., Conte, A. and Liu, T., 2021. Daily sampling of early SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness. *medRxiv*, pp. 2021–07.

Supplementary Information

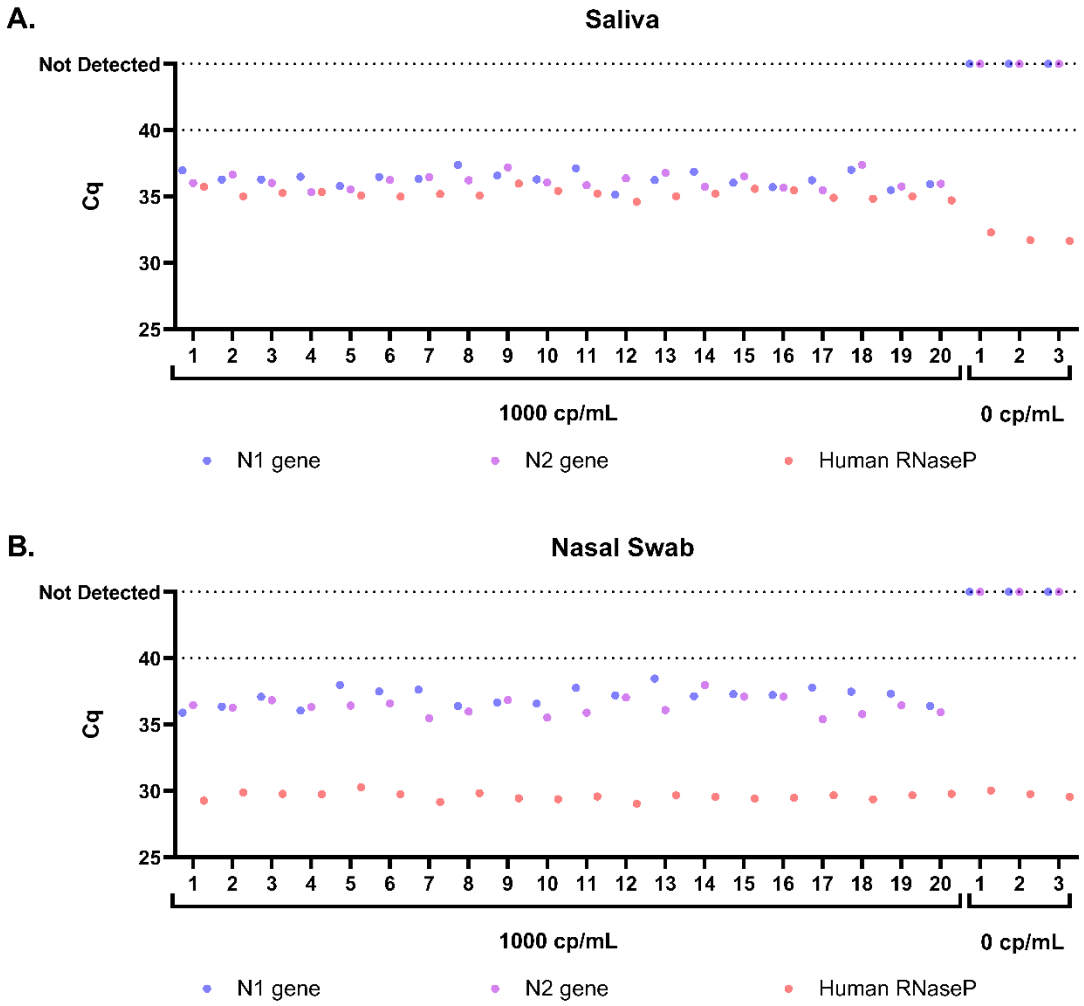
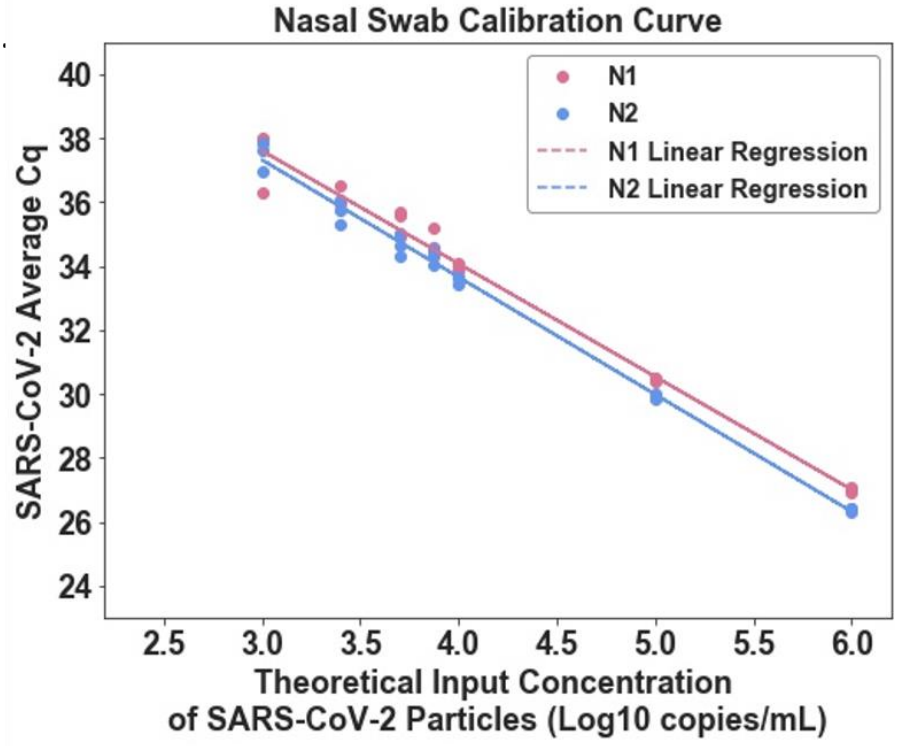
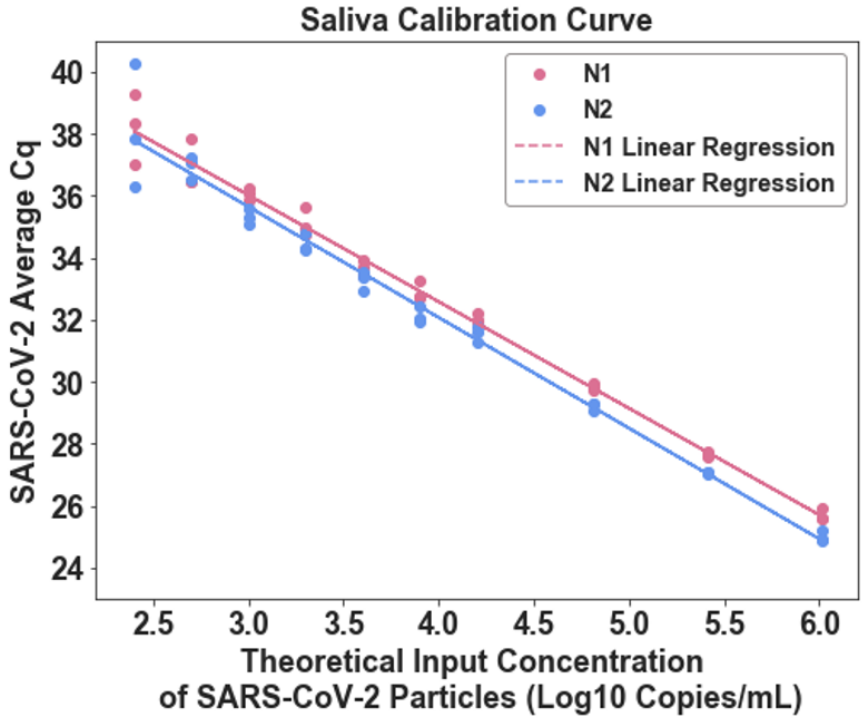


Figure 3.S1: Limit of detection of saliva and nasal-swab RT-qPCR assays used in this study. RT-qPCR quantification cycle (Cq) for SARS-CoV-2 N1 gene (blue circle), N2 gene (purple circle), and human RNase P gene (orange circle) in 20 replicates of pooled matrix spiked with 1000 copies/mL (cp/mL) heat-inactivated SARS-CoV-2 RNA and three replicates of pooled matrix spiked with a buffer blank for saliva (A) and nasal-swab (B) samples. Duplicate RT-qPCR reactions were performed for each extraction replicate and the averages are shown, with the following three exceptions: replicate 9 (saliva), in which the N1 gene only amplified in one of the duplicate runs (N2 in this run was positive, so per CDC EUA guidelines1 this run was interpreted as inconclusive), replicate 10 (nasal swab) in which the N2 gene only amplified in one of the duplicate runs (N1 in this run was positive, so this run was interpreted as inconclusive), and replicate 18 (nasal swab) in which the N1 gene only amplified in one of the duplicate runs (N2 in this run was positive, so this run was also interpreted as inconclusive). None of the samples spiked at 1000 copies/mL gave a negative detection result.

A.


Nasal Swab Calibration Curve
N1
N2
N1 Linear Regression
N2 Linear Regression
SARS-CoV-2 Average Cq
40
38
36
34
32
30
28
26
24
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
Theoretical Input Concentration
of SARS-CoV-2 Particles (Log10 copies/mL)

B.


Saliva Calibration Curve
N1
N2
N1 Linear Regression
N2 Linear Regression
SARS-CoV-2 Average Cq
40
38
36
34
32
30
28
26
24
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
Theoretical Input Concentration
of SARS-CoV-2 Particles (Log10 Copies/mL)

[Figure on previous page] Figure 3.S2: Calibration curve of SARS-CoV-2 inactivated particles to establish viral-load conversion equations. Linear regression of RT-qPCR quantification cycle (Cq) for N1 (red circle) and N2 (blue circles) genes at known concentrations of inactivated SARS-CoV-2 particles for saliva (A) or nasal swab (B) using this study's collection and laboratory workflows. Triplicate replicates per concentration were performed. Linear regression for N1 represented by red line and N2 represented by blue line. Linear regression R2 values are 0.986 for N1 in nasal swabs, 0.994 for N2 in nasal swabs, 0.989 for N1 in saliva, and 0.979 for N2 in saliva.

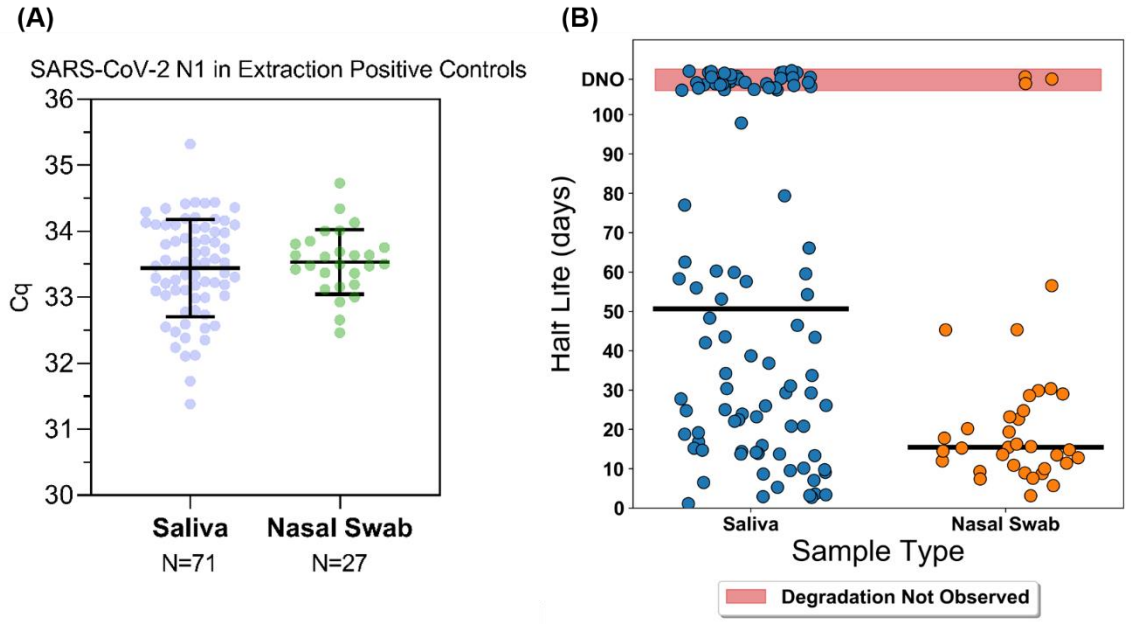
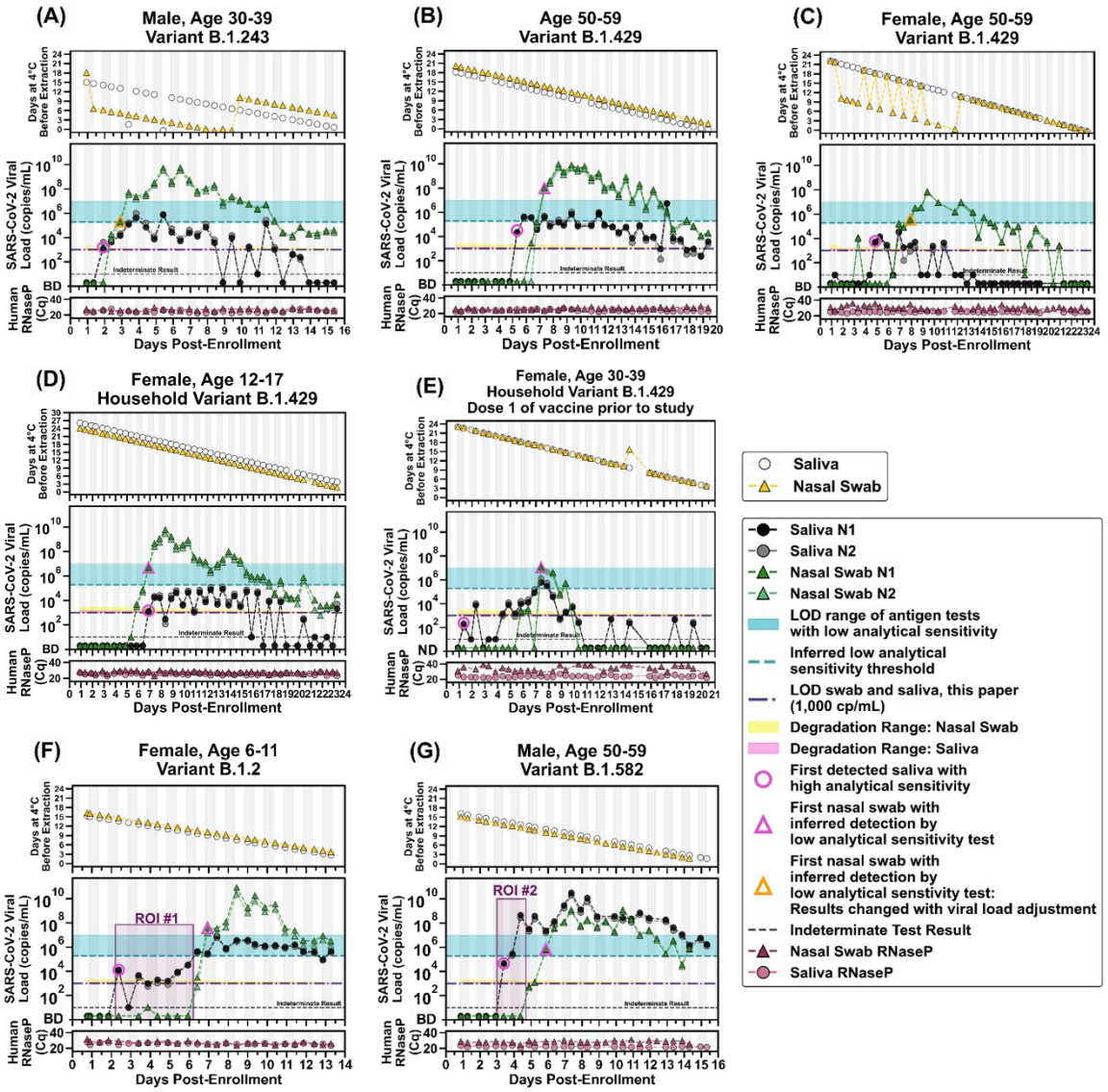


Figure 3.S3: SARS-CoV-2 RNA stability over time in Spectrum SDNA-1000 buffer at 4 °C. (A) Positive extraction control samples from 71 saliva extraction runs and 27 nasal-swab extraction runs are included to show the measurement noise in the quantification workflow. The standard deviation for the positive control measurements was 0.74 Cq for saliva and 0.49 for nasal swab. (B) The observed half-life (days) of participant saliva (blue circles) and nasal-swab (orange circles) samples in Spectrum SDNA-1000 buffer stored at 4 °C. Individual samples were extracted at multiple time points. Half-life in this context refers to the time required to observe a 1 Cq increase (representing a 2-fold decrease) in RNA detected by RT-qPCR. The median point is identified for each sample type (black bars), at 15.0 days for nasal swabs (red circle) and 51.0 days for saliva (green circle). Of the 110 total participant saliva samples plotted in panel B, 36 samples had no evidence of degradation (DNO) under the time frame measured. Only 3 of the 36 total participant nasal-swab samples plotted in panel B had no evidence of degradation under the timeframe measured. DNO = degradation not observed, meaning that the difference in extraction Cq values of the same sample at multiple time points was within 1 standard deviation observed in replicate extraction positive controls for the respective sample type, as shown in panel A.


(A)
Male, Age 30-39
Variant B.1.243
(B)
Age 50-59
Variant B.1.429
(C)
Female, Age 50-59
Variant B.1.429
(D)
Female, Age 12-17
Household Variant B.1.429
(E)
Female, Age 30-39
Household Variant B.1.429
Dose 1 of vaccine prior to study
(F)
Female, Age 6-11
Variant B.1.2
(G)
Male, Age 50-59
Variant B.1.582
Days at 4°C Before Extraction
SARS-CoV-2 Viral Load (copies/mL)
Indeterminate Result
Human RNaseP (Cq)
BD
ND
Days Post-Enrollment
ROI #1
ROI #2
Saliva
Nasal Swab
Saliva N1
Saliva N2
Nasal Swab N1
Nasal Swab N2
LOD range of antigen tests with low analytical sensitivity
Inferred low analytical sensitivity threshold
LOD swab and saliva, this paper (1,000 cp/mL)
Degradation Range: Nasal Swab
Degradation Range: Saliva
First detected saliva with high analytical sensitivity
First nasal swab with inferred detection by low analytical sensitivity test
First nasal swab with inferred detection by low analytical senstivity test: Results changed with viral load adjustment
Indeterminate Test Result
Nasal Swab RNaseP
Saliva RNaseP

[Figure on previous page] Figure 3.S4. Predicted impact of SARS-CoV-2 RNA stability on quantitative viral loads shown in Figure 3.2. (A-G) The time [days] of sample storage at 4 °C between sample collection and RNA extraction is shown in the topmost panels. Open circles represent saliva samples and yellow triangles represent the nasal swabs. Viral-load calculations are corrected for the median half-life (1 Cq decrease in RNA detected by RT-qPCR) of each sample type and the duration of storage at 4 °C before quantification (15 days for 2-fold decrease in detected RNA in nasal swabs and 51 days for 2-fold decreased in detected RNA in saliva). The degradation ranges, represented by a shaded yellow (nasal swab) or pink (saliva) region to represent how a measured value of 1,000 copies/mL may have degraded from concentrations in this range. As in Figure 3.2, ND = not detected for Cqs ≥40 (see Methods for details). The LOD of the saliva and nasal-swab assays used here (1,000 cp/mL) is indicated with the purple dashed line; low-analytical-sensitivity threshold Supplemental Information page 6 is indicated by the horizontal green dashed line; the low-analytical-sensitivity range (horizontal blue bar) is shown for reference. A diagnostic test does not provide reliable detection for samples with viral loads below its LOD. For each participant, the first detected saliva point is emphasized with a pink circle and their first nasal-swab point above the LOD of the ID NOW is emphasized with a pink triangle. Vertical shading in gray indicates nighttime (8pm – 8am). Internal control of RNase P gene Cqs from the CDC primer set are provided for each sample to compare self-sampling consistency and sample integrity (failed samples, where RNase P Cq ≥40, are not plotted). Samples with an indeterminate result by the CDC RT-qPCR assay are shown along the horizontal black dashed line. Participant gender, age range, and SARS-CoV-2 variant is given in each panel's title. Two regions of interest (ROI) are indicated by purple-shaded rectangles and discussed in the main text. For the two points that change interpretation with the viral-load adjustment, orange triangles show which new data points become the first nasal-swab point in range of low-analytical-sensitivity tests.

Table 3.S1. Study participant demographic data. Figure 2 shows viral loads and symptoms data for the seven participants for whom we observed transmission during their enrollment in the study.

	Age Range (Years)	**Sex**	**Race; Ethnicity**	**Reported Medical Conditions Associated with Increased Risk of Severe COVID-19**[88]
Fig. 2A, Fig. S4A	30-39	Male	Other; Mexican/Mexican-American/Chicano (Salvadoran)	Diabetes
Fig. 2B, Fig. S4B	50-59	Male	Do not wish to respond; Mexican/Mexican-American/Chicano	None
Fig. 2C, Fig. S4C	50-59	Female	White; Mexican/Mexican-American/Chicano (Spanish-American from Spain)	None
Fig. 2D, Fig. S4D	12-17	Female	White; Mexican/Mexican-American/Chicano	None
Fig. 2E, Fig. S4E	30-39	Female	White; Mexican/Mexican-American/Chicano	None
Fig. 2F, Fig. S4F	6-11	Female	White; Non-Hispanic	None
Fig. 2G, Fig. S4G	50-59	Male	American Indian or Alaskan Native, White; Other Hispanic, Latinx or Spanish origin	Obesity

Supplementary Methods

Participant population

This study is an extension of our previous study examining viral load in saliva (3). Both studies were reviewed and approved by the Institutional Review Board of the California Institute of Technology, protocol #20-1026. All participants provided either written informed consent (or for minors ages 6–17, assent accompanied by parental permission), prior to enrollment. Household index cases were eligible for participation if they had recently (within 7 days) been diagnosed with COVID-19 by a CLIA laboratory test. Individuals were ineligible if they were hospitalized or if they were not fluent in either Spanish or English. All participant data were collected and managed using REDCap (Research Electronic Data Capture) on a server hosted at the California Institute of Technology. Demographic and health information for the seven participants can be found in Table 3.S1.

Questionnaires and symptom monitoring

Acquisition of participant data was performed as described previously (3). Briefly, upon enrollment each participant completed an online questionnaire regarding demographics, health factors, prior COVID-19 tests, COVID-19- like symptoms since February 2020, household infection-control practices, and perceptions of COVID-19 risk. Participants also filled out a post-study questionnaire in which they documented medications taken and their interactions with each household member during their enrollment.

Information on symptoms was collected twice daily in parallel with sample collection. Participants recorded any COVID-19-like symptoms (as defined by the CDC4) they were experiencing at the time of sample donation on a symptom-tracking card or on a custom app run on REDCap (4). Whenever possible, participants indicated the self-reported severity of each symptom. Participants were also given the opportunity to write-in additional symptoms or symptom details not otherwise listed.

Collection of respiratory specimens

Participants self-collected both their nasal-swab and saliva samples using the Spectrum SDNA-1000 Saliva Collection Kit (Spectrum Solutions LLC, Draper, UT, USA), which contains 1.5 mL of liquid buffer, at home twice per day (after waking up and before going to bed), per manufacturer guidelines. Of note, at the time of this writing, Spectrum devices are not approved for the collection of nasal-swab samples. Participants self-collected nasal-swab (1 swab) and saliva (~1.5 mL) samples in the Spectrum SDNA-1000 Saliva Collection Kit (Spectrum Solutions LLC, Draper, UT, USA), which contains 1.5 mL of liquid buffer, at home twice per day (after waking up and before going to bed), per manufacturer's guidelines. Of note, at the time of this writing, Spectrum devices are not approved for the collection of nasal-swab samples.

Participants were instructed not to eat, drink, smoke, brush their teeth, use mouthwash, or chew gum for at least 30 min prior to donating. Prior to nasal-swab donation, participants were asked to gently blow their noses to remove debris. Participants were provided with one of the following types of sterile flocked swabs: Nest Oropharyngeal Specimen Collection Swabs (Cat. NST-202003, Stellar Scientific, Baltimore, MD, USA) Puritan HydraFlock Swab (Cat. 25-3000-H E30, Puritan, Guilford, ME, USA) or Copan USA FLOQSwab (Cat. 520CS01, VWR International, Radnor, PA, USA). Participants were instructed to swab each nostril for four complete rotations using the same swab while applying gentle pressure, then to break the tip of the swab into the Spectrum tube and securely screw on the cap. A parent or legal guardian assisted all minors with swab collection and they were instructed to wear a face covering during supervision. Tubes were labeled and packaged by the participants and transported at room temperature by a touch-free medical courier to the California Institute of Technology daily for analysis.

Upon receipt of the samples in the California Institute of Technology laboratory, each sample was inspected for quality. A sample failed quality control if the preservation buffer was not released from the Spectrum SDNA1000 cap, or if sample tubes were leaking or otherwise unsafe to handle. Samples that failed quality control were not processed. Inactivated samples

were stored at 4 °C and were equilibrated to room temperature before being processed with extraction protocols.

RNA extraction protocols

In initial testing, when combined with standard KingFisher MagMax sample-preparation protocols, these assays performed well to quantify heat-inactivated SARS-CoV-2 viral particles spiked into commercially available SARS-CoV-2 negative saliva and nasal fluid from pooled donors. However, the assay did not provide reliable quantification from freshly collected individual saliva samples with varying viscosity from positive participants in this study. Carryover of materials from some of the mucus-rich samples was inhibitory, as determined by RTddPCR analysis of dilutions of eluted RNA (data not shown). Following recommendations from ThermoFisher, the protocol was adjusted and described below. Briefly, we added a centrifugation step after proteinase k treatment to pellet the mucus-rich cell debris. We also include a third wash to improve RNA quality for viral genome sequencing. These steps reduced bead carryover into the eluate, as well as ddPCR inhibition.

Participant saliva and anterior-nares swab samples were extracted using the KingFisher Flex 96 instrument (ThermoFisher Scientific) with the MagMax Viral Pathogen I Nucleic Acid Isolation kit (Cat. A42352, Applied Biosystems, Waltham, MA, USA) guided by ThermoFisher technical notes for SARS-CoV-2 modification and saliva. Each extraction batch, depending on the sample type being extracted, contained a contrived SARS-CoV-2 negative control sample containing either 225 µL of Spectrum buffer mixed with 225 µL of commercial pooled human saliva (Lee Bio 991-05-P-PreC) or 240 µL of Spectrum buffer with 10 µL of pooled commercial nasal fluid (Lee Bio 991-13-P-PreC); a contrived SARS-CoV-2 positive control sample was also included in each extraction batch, with the formulations above, but with the Spectrum buffer spiked with 7,500 genomic copy equivalents/mL of heat-inactivated SARS-CoV-2 particles (BEI NR-52286).

Saliva and anterior-nares swab samples were prepared for purification by transferring 550 µl (for saliva) or 250 µl (for nasal swab) of each sample from its corresponding Spectrum buffer

tube into a 1.5 mL lo-bind Eppendorf tube containing 10 µl (for saliva) or 5 µl (for nasal swab) of proteinase K. To maximize recovery of RNA off swabs, prior to transfer, pipet mixing was performed 5–7 times near the swab in the Spectrum tube before aliquoting into an Eppendorf tube. Saliva samples were vortexed for 30 sec in the Eppendorf tube. Samples were incubated at 65 ℃ for 10 min, then centrifuged at 13,000 x g for 1 min. Aliquots of 400 µl (for saliva) or 200 µl (for nasal swab) were transferred into a KingFisher 96 deep well plate (Cat. 95040450, ThermoFisher Scientific) and processed following KingFisher protocols MVP_400ul_3washes.bdz (for saliva) or MVP_200ul_2washes.bdz (for nasal swab). Ethanol washes were performed with 80% ethanol. Both sample types were eluted into 100 µl of MagMax viral pathogen elution buffer.

RT-qPCR

Quantification of SARS-CoV-2 was performed as previously described.3 Briefly, the CDC5 SARS-CoV-2 N1 and N2 gene primers and probes with an internal control targeting RNase P gene primer and probe were run in a multiplex RT-qPCR reaction using TaqPath 1-Step Rt-qPCR Mastermix (Cat. A15299, ThermoFisher Scientific). Reactions were run in duplicate on a CFX96 Real-Time Instrument (Bio-Rad Laboratories, Hercules, CA, USA).

RT-ddPCR

Reverse-transcription droplet digital PCR (RT-ddPCR) was performed using the Bio-Rad SARS-CoV-2 Droplet Digital PCR kit (Cat. 12013743, Bio-Rad). Swab samples were processed following the manufacturer's RUO protocol with 5.5 µl template per 22 µl reaction. A total of 42 participant nasal-swab samples were characterized by RT-ddPCR. Modifications were made for saliva samples by reducing the template addition to 2.75 µl per 22 µl reaction. A total of 63 participant saliva samples were characterized by RT-ddPCR. Prior to adding template, samples were diluted into digital range using nuclease-free water. Droplets were created using the QX200 Droplet Generator (Cat #1864002, Bio-Rad), thermocycling performed on Bio-Rad C1000 and detected using the QX200 Droplet Digital PCR system (Cat. 1864001, Bio-Rad). Samples were analyzed with QuantaSoft analysis Pro 1.0.595 software following Bio-Rad's RUO SARS-CoV-2 guidelines (6).

Viral-load calibration curves

A calibration curve was prepared for both the saliva and nasal-swab protocols. Contrived samples were prepared with known concentrations (based on the certificate of analysis, COA) of heat-inactivated SARS-CoV-2 virus ($3.75x10^8$ GE/mL, Batch 70034991, Cat. NR-52286, BEI Resources, Manassas, VA, USA) in the inactivating buffer from the Spectrum SDNA-1000 Saliva Collection Kit (Spectrum Solutions LLC, Draper, UT, USA) and commercial, healthy human fluids were used as healthy human samples. Commercial pooled human saliva collected prior to November 2019 (Cat, 991-05-P, Lee Biosolutions, Maryland Heights, MO, USA) for the contrived saliva samples or commercial human nasal fluid collected prior to November 2019 (Cat No 991-13-PPreC, Lee Biosolutions) for the contrived nasal-swab samples. Details of reagent volumes are described in the following two paragraphs for how the samples were prepared for both nasal swab and saliva calibration curves.

To establish the nasal-swab calibration curve (Figure 3.S2A), contrived samples were prepared by creating a dilution series of commercial heat-inactivated SARS-CoV-2 virus from BEI ($3.75x10^8$ GE/mL) in a 10-fold dilution series from $1x10^6$ to $1x10^4$ copies/mL with finer resolution down to our LOD at $1x10^3$ copies/mL. Dilutions were prepared in Spectrum device inactivation buffer, to a volume of 768 µL, at concentrations of 0 copies/mL, 1,000 copies/mL, 2,500 copies/mL, 5,000 copies/mL, 7,500 copies/mL, 10,000 copies/mL, 100,000 copies/mL, and 1,000,000 copies/mL. To bring the volume to 800 µL total, 32 µL of healthy human nasal fluid collected prior to November 2019 (Cat No 991-13-P-PreC, Lee Biosolutions) was added. Triplicate extractions, 250 µL each, were performed according to the nasal-swab RNA extraction protocol (described above). Each extraction was run in triplicate RT-qPCR reactions.

To establish the saliva calibration curve (Figure 3.S2B), contrived samples were prepared by creating a dilution series of commercial BEI heat-inactivated SARS-CoV-2 virus in Spectrum device inactivation buffer at concentrations of 0 copies/mL, 1,000 copies/mL,

2,000 copies/mL, 4,000 copies/mL, 8,000 copies/mL, 16,000 copies/mL, 64,000 copies/mL, 256,000 copies/mL, and 1,020,000 copies/mL. Contrived samples were made by mixing 620 µL of each concentration of the dilution series with 620 µL of healthy pooled human saliva (Cat, 991-05-P, Lee Biosolutions). Triplicate extractions, 550 µL each, were performed according to the saliva RNA extraction protocol. Each extraction was run in triplicate RT-qPCR reactions.

Equations, calculated from the linear regression of the calibration curves, are shown below as Equations 3.1–3.4. These calibration curves are used to convert the Cq values obtained by RT-qPCR to viral load in each participant sample. For saliva, viral load is a calculation of viral copies/mL in the saliva corrected for dilution with the Spectrum buffer. We assumed that participants donate saliva to the fill line, matching the 1:1 dilution in Spectrum buffer recreated when preparing contrived samples for the saliva calibration curve. For nasal swabs, viral load is a calculation of the concentration of viral copies/mL released from the swab into the 1.5 mL of inactivating buffer (which is a similar volume as the 1–3 mL of viral transport media typically used for sample collection). Concentrations higher than 1,000,000 copies/mL could not be characterized due to a limitation of the available stock concentration of commercial inactivated SARS-CoV-2. To validate linear conversion was acceptable at concentrations higher than 1,000,000 copies/mL, we compared RT-ddPCR and RT-qPCR quantification on some participant samples (Figure 3.1) as described in the next section "Viral-load Quantification between qPCR and ddPCR assays."

$$Saliva\ N1\ gene\ viral\ load\ \left[\frac{copies}{mL}\right] = 2^{((Cq-46.349)/-1.0357)}$$

(Eq. 3.1)

$$Saliva\ N2\ gene\ viral\ load\ \left[\frac{copies}{mL}\right] = 2^{((Cq-46.374)/-1.0759)}$$

(Eq. 3.2)

$$Nasal\ Swab\ N1\ gene\ viral\ load\ \left[\frac{copies}{mL}\right] = 2^{((Cq-48.221)/-1.0643)}$$

(Eq. 3.3)

$$Nasal\ Swan\ N2\ gene\ viral\ load\ \left[\frac{copies}{mL}\right] = 2^{((Cq-48.330)/-1.1044)}$$

(Eq. 3.4)

Viral-load quantification between qPCR and ddPCR assays

Contrived saliva and nasal-swab calibration curve RT-qPCR data was converted into viral load (N1 copies/mL) using Eq. 3.1 and 3.3 listed in the above section. Calculated viral load was plotted against the theoretical input of heat-inactivated SARS-CoV-2.

Extending quantification capabilities above $1x10^6$ N1 copies/mL was achieved using SARS-CoV-2-positive participant samples. Due to the limitation of the commercial SARS-CoV-2 standard concentration, we were not able to prepare contrived samples with SARS-CoV-2 input concentrations greater than $1x10^6$ copies/mL. To capture a range of participant samples over 7 orders of magnitude ($1x10^3$ to $1x10^{10}$ copies/mL SARS-CoV-2 *N1* gene), 63 saliva and 42 nasal-swab samples from SARS-CoV-2-positive participants were selected based on RTqPCR data to quantify using RT-ddPCR. Using the geometric mean of the viral load computed from RT-qPCR and the calibration curves and the concentration measured by RT-ddPCR, we were able to evaluate the linearity of the calibration curve across the seven orders of magnitude viral load seen in the participant samples (Figure 3.1B–C). Samples were selected to capture the range of viral concentrations within our calibration curve and to the highest viral loads recorded for each sample type (nasal and saliva). The geometric means of RT-qPCR and RT-qPCR viral concentrations were calculated by taking the square root of RT-qPCR *N1* concentrations × RT-ddPCR *N1* concentration.

We observed excellent concordance between the calibration curve (Figure 3.1A, complete data in Figure 3.S2), RT-qPCR and RT-ddPCR assays over the entire dynamic range of input

concentrations (Figure 3.1B-C), even though RT-qPCR eluents were run as-is and RT-ddPCR eluents from high-concentration samples were significantly diluted. For nasal-swab samples, RT-ddPCR values were slightly below the RT-qPCR values, however this difference was consistent across the entire dynamic range, indicating no concentration-dependent biases like enzymatic inhibition. We chose not to adjust the calibration curve to fit RT-ddPCR values; we reported the concentrations based on the calibration curves derived from the certificate of analysis from the BEI Resources reference material. For saliva samples, all points tightly clustered around the x=y line (Figure 3.1 A-C).

Establishment of Limit of Detection

Results of the calibration curve (Figure 3.S2 A, B) demonstrated 3 of 3 replicates detected at 1,000 copies/mL saliva (for saliva) and 1,000 copies/mL buffer (for nasal swabs). For each sample type (saliva, nasal swab), 20 contrived samples with the equivalent of 1,000 copies/mL were prepared as described above, individually extracted as described above, and subjected to RT-qPCR as described above. The LOD for each sample type through the workflow was considered established if a positive result for detection (as defined in the EUA for the CDC RT-qPCR assay) was obtained for ≥ 19 of 20 (≥95% as required by FDA EUA guidelines for determining LOD) of replicates at the input concentration (Figure 3.S1 A, B).

Three of three replicate sample extractions included in the calibration curves for both contrived nasal-swab samples and contrived saliva samples spiked with heat-inactivated SARS-CoV-2 particles at a concentration of 1,000 copies/mL were detected by RT-qPCR, prompting testing of additional 20 replicates of each sample type spiked at that concentration, individually extracted, and tested by RT-qPCR to establishment of the LOD for our RT-qPCR assay. For both sample types (saliva and nasal swabs), 20 of 20 replicates were positive for SARSCoV-2 (Figure 3.S1A, B), establishing 1,000 copies/mL of saliva and 1,000 copies/mL of swab buffer as the high-sensitivity LOD for our RT-qPCR assays.

Threshold to infer Performance of tests with low analytical sensitivity

The threshold of $1.0x10^5$ copies/mL is applied generally to both saliva and nasal swabs viral

loads (copies/mL) to infer detection by a test with low analytical sensitivity. The rationale to use this threshold is to demonstrate a best-case scenario performance of tests with low-analytical-sensitivity (from the low-analytical-sensitivity range $1.0x10^5 – 1.0x10^7$ copies/mL used in this paper). The comparisons in the paper would be more dramatic if a low analytical-sensitivity threshold greater than this number was selected.

Data analysis

Before we converted Cq values to viral load, we used Cq cutoffs based on the CDC guidelines5 to define samples as positive, negative, indeterminate, or invalid (fail), and then excluded from the viral-load plots any points that failed, and any samples whose RNase P Cq values ≥40. Because we ran duplicate RT-qPCR reactions, the mean Cq of positive reactions was used for conversion to viral load.

Figure 3A percentages are calculated by Equation 3.5, where the percentage positive by a test of a given analytical sensitivity (high-analytical-sensitivity results are all measured values, by our internal test with an LOD ≤ 1000 cp/mL; low-analytical-sensitivity results are measured values at or above a threshold of $1.0x10^5$ cp/mL):

$$Percent\ Positive_{as} = \frac{n}{N} \times 100$$

(Eq. 3.5)

Where “as” refers to the analytical sensitivity. In Eq. 3.5, “N” is defined as the total number of participants with saliva samples passing quality-control evaluations (see Methods) for safety and human *RNaseP* gene Cq threshold at the corresponding aligned time point in column “Days from First Positive Results in Either Sample Type.” Maximum denominator of number of 7, corresponding to the number of participants in the study and each participant has a maximum of one sample per time point. Numbers may vary before day 0 as each participant had a variable number of negative test results before first detected SARS-CoV-2 RNA. In Eq. 3.5, “n” represents the number of participants, at a given time point, with detectable SARS-CoV-2 RNA (see RT-qPCR methods) in the sample type (saliva or nasal

swab) using a high-analytical-sensitivity assay. For predicting performance of each sample type (saliva or nasal swab) with a test of low analytical sensitivity, "n" is defined as the number of participants with a SARS-CoV-2 *N1* gene viral load of SARS-CoV-2 greater than $1.0x10^5$ copies/mL (cp/mL) in samples, which would indicate an inferred positive result using a low-sensitivity assay with an LOD of SARSCoV-2 *N1* gene viral load of $1.0x10^5$ copies/mL. Details of the calculations are included in the Data_Annotation file on CaltechDATA.

Figure 3.3D considers only samples collected within ten days after the assigned first positive result were analyzed to consider symptoms relevant to an early infection. The first date of positive result observed using our high analytical-sensitivity assay (either sample type) was assigned for each participant shown in the panels of Figure 3.2 and days 0–10 were analyzed for panel D.

Samples were designated as being collected while symptomatic if the participant reported experiencing one or more COVID-19-like symptoms at the time of sample collection; if no COVID-19 like symptoms were reported, the sample was designated as "No Symptoms Reported". Samples were defined as either positive, negative, or indeterminate by our high-analytical-sensitivity assay, based on the criteria from the manufacturer of the RT-qPCR assay, detailed above. Samples were inferred as either positive or negative by a low-analytical-sensitivity assay if the viral load measured in the sample was greater than our inferred low-analytical-sensitivity threshold, $1.0x10^5$ copies/mL.

Figure 3.3D utilizes Equation 5, where "N" is defined as the number of participant samples positive for SARSCoV-2 RNA within the symptomatic categories defined in the first 10 days of detectable infection (criteria above). There were 97 saliva and 95 nasal-swab samples collected while symptomatic, and 46 saliva and 44 nasal-swab samples collected with the participant reporting no symptoms. The value of "n" corresponds to the percent positive by either observed positivity by our high-analytical-sensitivity assay or inferred positive by a

low analytical-sensitivity assay as the numerator over the denominator corresponding to that sample type and symptom status, multiplied by 100%.

RNAseq

Saliva and nasal-swab samples below N1 Cq of 26 were sent to Chan Zuckerberg Biohub for SARS-CoV-2 viral genome sequencing, a modification of Deng et al. (2020)7 as described in Gorzynski et al. (2020).8 Sequences were assigned pangolin lineages described by Rambaut et al. (2020)9 using Phylogenetic Assignment of Named Global outbreak LINeages software v2.3.2 (github.com/cov-lineages/pangolin). Consequences viral genomes were submitted to GISAID by Chan Zuckerberg Biohub, see data availability section for accession id details.

SARS-CoV-2 RNA stability at 4 ˚C

As described above, each extraction batch included a contrived sample spiked with SARS-CoV-2 heat-inactivated particles. For all available saliva or nasal-swab extraction batches, the Cq value of the SARS-CoV-2 *N1* gene in the contrived SARS-CoV-2 positive extraction control was collected. The standard deviation of these measurements was calculated and used to establish a threshold for expected noise between repeat extractions of the same sample. To assess samples for evidence of SARS-CoV-2 RNA degradation, any participant sample that had more than one extraction replicate performed were analyzed. Samples where the difference in Cq values between the extractions was less than the threshold of expected noise between replicate extractions were defined as degradation not observed, (DNO). For samples where the difference was above this threshold, the time for 1 Cq increase (2-fold decrease) in RNA detected by RT-qPCR is described by the term half-life, which was calculated according to Eq. 3.6, below:

$$t_{1/2} = \frac{-\ln 2}{k}$$

(Eq. 3.6)

Where “k” is defined as the slope of the linear regression of the natural logarithm of the viral

load vs. extraction date (relative to sample collection date). The median over the entire dataset (saliva or swab) was used as a point estimate of RNA half-life. The median point was determined to be 15.0 days for nasal swabs and 51.0 days for saliva.

Calculations that predict the impact of storage time at 4 °C and RNA stability on viral load are calculated according to Eq. 3.7, below.

$$y_{adj} = y_{deg} 2^{\frac{\Delta t}{t_{1/2}}} \quad (3.7)$$

Where y_{adj} is defined as the adjusted viral load, y_{deg} is defined as the viral load before adjustment for degradation (as calculated by Eq. 3.1–3.4), and $t_{1/2}$ is defined as the RNA half-life, shown in Eq. 3.5.

All samples were stored at 4 °C before extraction; time of storage varied between 0–27 days. The stability of SARS-CoV-2 RNA in nasal-swab samples was slightly lower (1 Cq loss of RNA detected after a median of 15 days) than the stability of SARS-CoV-2 RNA in saliva samples (1 Cq loss of RNA detected after a median of 51 days) (Figure 3.S3). An assessment of how viral-load measurements in Figure 3.2 may have been affected by time between sample collection and quantification is included in Figure 3.S4. Given the large dynamic range of the viral loads in these samples (~24 Cq or about 10,000,000 fold), we considered stability corresponding to a 1 Cq (2 fold) loss to be adequate.

The predicted impact of RNA degradation on the comparisons of high-analytical-sensitivity saliva to inferred low-analytical-sensitivity nasal testing is shown in Figure 3.S4. Accounting for potential decreases of viral RNA in the nasal swab resulting from delays between sample collection and quantification only impact the interpretation of two points, conservatively decreasing the delay from 2.0 to 1.5 day for the first participant (Figure 3.2B and Figure 3.S4B) and from 3.0 to 2.0 days for the third participant (Figure 3.2C and Figure 3.S4C).

Supplementary Discussion

Three participants (Figure 3.1C–E) were infected with the same variant, B.1.429 (CAL20), classified as a variant-of concern at the time of this study. The SARS-CoV-2 variant for the participants in Figure 3.1D and Figure 3.1E were inferred from the sequenced sample of the household's presumed index case. Saliva viral loads for each of these participants (Figure 3.2C-E) were low. Of note, the participants in Figure 3.2C and 3.2E showed high *RNase* P Cq values (indicating low concentration of the human control target); and variability of *RNase* P Cq values across the nasal swab samples suggests that inconsistent swab-sampling quality could have impacted these participants' viral-load data and should be taken into account when interpreting those data.

Beyond outbreak prevention and control, early detection of COVID-19 may also be useful for individual patient care, as high-risk patients who are identified early can be monitored and have treatment initiated swiftly if it becomes appropriate. Several treatments show exclusive or increased efficacy only when given early in the infection. The advantage of earlier treatment initiation is likely due to reduction of viral replication either directly or by promotion of an early effective immune response, which prevents a later exaggerated inflammatory response (10). Results of the ACTT-1 trial demonstrated a survival benefit in patients for whom Remdesivir was initiated in the early stages of treatment (supplemental oxygen only), but that benefit was lost once disease progressed, and advanced respiratory support was needed (10–12). More recently, the MOVe-OUT clinical trial demonstrated the efficacy of molnupiravir when (per trial inclusion criteria) initiated among outpatients within the first five days from symptom onset, whereas the inpatient study (MOVe-IN) did not proceed to Phase 3, as clinical benefit was not observed for hospitalized patients with a longer duration of symptoms prior to initiation of the treatment (10). Other therapies, such as plasma and monoclonal antibody therapies (bamlanivimab or casirivimab plus imdevimab) show similar clinical benefits in early initiation of treatment (11–16).

Although national vaccination efforts are reducing severe COVID-19 outcomes in the U.S.,

a sizable portion of the world's population is likely to remain unvaccinated due to limited vaccine availability, medical ineligibility (in the U.S., children under 5 years of age are not yet eligible), or personal preference. Thus, testing remains an important tool for preventing outbreaks among children in schools and daycare facilities (where children under age 2 generally do not wear masks), which may spread to the community and increase rates of infection among high-risk and unvaccinated individuals. Tests that detect early infections are also important to prevent viral transmission in congregate settings with high-risk or unvaccinated populations, such as hospitals, college dormitories, homeless shelters, correctional facilities, summer camps for children, elementary schools, and long-term care facilities.

As new SARS-CoV-2 variants emerge, quantitative studies of the kinetics of early stage viral loads must be continually updated in follow-up studies. Importantly, such studies should be undertaken in people of a wide range of ethnicities, races, health conditions, vaccination status, and ages. Breakthrough cases are often asymptomatic (17) and recent evidence suggests that vaccinated individuals may transmit infections from the new variants, including Delta (18). Another reason for continued monitoring of early viral kinetics is that viral evolution, including of host tropism, can markedly diminish the effectiveness of a diagnostic strategy. In one study, decreased clinical sensitivity of NP swabs was observed in SARS-CoV-2 variant B.1.616 (19) which may indicate a tropism shift of the virus into lower-respiratory compartments. Finally, quantitative data must be acquired in parallel with viral-culture data to understand the viral loads and phases of infection that are most relevant to transmission.

Early detection of infection clearly reduces community transmission, however, for most of the COVID-19 pandemic, policy makers have had to develop testing strategies in the absence of quantitative data on viral kinetics from the earliest stage of infection. Lacking such data-based guidance, diagnostic tests have been used incorrectly (with false-negative results due to using tests with insufficient sensitivity) in several scenarios, resulting in outbreaks that could have been prevented with an appropriate testing strategy (20–26).

One barrier to implementing such more advanced testing strategies is availability of appropriate tests. Because the optimal sample type for early detection might be different for different populations, or might change as new variants emerge, tests with robust high analytical sensitivity across multiple sample types are needed. Developing such tests is challenging because it requires incorporating robust sample-preparation technology to purify and concentrate pathogen nucleic acids from diverse human matrices, from upper respiratory (e.g., fluids from the nasal, nasopharyngeal, oral and oropharyngeal compartments, captured in swabs or saliva) to lower respiratory samples (e.g., sputum, tracheal aspirate, bronchoalveolar lavage). It is even more challenging to incorporate such sample-preparation technology into tests that can be broadly deployed—at very low cost—at the point of care in limited-resource settings (such as schools, homes, and businesses, and especially in the developing countries). Development of such highly sensitive, rapid, and inexpensive tests with broad sample-type compatibility is urgently needed.

Supplemental references

1. Centers for Disease Control and Prevention, 2020. Real-time RT-PCR panel for detection 2019-Novel Coronavirus. 2020.

2. Centers for Disease Control and Prevention, 2021. COVID-19: People with certain medical conditions. *Centers for Disease Control and Prevention, Atlanta, GA. https://www. cdc. gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html*.

3. Winnett, A., Cooper, M.M., Shelby, N., Romano, A.E., Reyes, J.A., Ji, J., Porter, M.K., Savela, E.S., Barlow, J.T., Akana, R. and Tognazzini, C., 2020. SARS-CoV-2 viral load in saliva rises gradually and to moderate levels in some humans. *medRxiv*.

4. Symptoms of COVID-19 (2020) Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html (Accessed: 28 May 2024).

5. CDC 2019 nCoV real-time RT-PCR diagnostic panel (2020) FDA. Available at: https://www.fda.gov/media/134922/download (Accessed: 28 May 2024).

6. Bio-Rad SARS-COV-2 ddpcr kit instructions for use (2020) Bio-Rad. Available at: https://www.bio-rad.com/webroot/web/pdf/lsr/literature/10000130776.pdf (Accessed: 28 May 2024).

7. Deng, X., Achari, A., Federman, S., Yu, G., Somasekar, S., Bártolo, I., Yagi, S., Mbala-Kingebeni, P., Kapetshi, J., Ahuka-Mundeke, S. and Muyembe-Tamfum, J.J., 2020. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nature microbiology*, *5*(3), pp. 443–454.

8. Gorzynski, J.E., De Jong, H.N., Amar, D., Hughes, C., Ioannidis, A., Bierman, R., Liu, D., Tanigawa, Y., Kistler, A.L., Kamm, J. and Kim, J., 2020. High-throughput

SARS-CoV-2 and host genome sequencing from single nasopharyngeal swabs. *medRxiv*, pp. 2020–07.

9. Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L. and Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature microbiology*, *5*(11), pp. 1403–1407.

10. Merck, M., 2021. Merck and Ridgeback Biotherapeutics Provide Update on Progress of Clinical Development Program for Molnupiravir, an Investigational Oral Therapeutic for the Treatment of Mild-to-Moderate COVID-19.

11. Simonovich, V.A., Burgos Pratx, L.D., Scibona, P., Beruto, M.V., Vallone, M.G., Vázquez, C., Savoy, N., Giunta, D.H., Pérez, L.G., Sánchez, M.D.L. and Gamarnik, A.V., 2021. A randomized trial of convalescent plasma in Covid-19 severe pneumonia. *New England Journal of Medicine*, *384*(7), pp. 619–629.

12. Libster, R., Marc, G.P., Wappner, D., Coviello, S., Bianchi, A., Braem, V., Esteban, I., Caballero, M.T., Wood, C., Berrueta, M. and Rondan, A., 2020. Prevention of severe COVID-19 in the elderly by early high-titer plasma. *MedRxiv*, pp. 2020–11.

13. Activ-3/Tico Ly-CoV555 Study Group, 2021. A neutralizing monoclonal antibody for hospitalized patients with Covid-19. *New England Journal of Medicine*, *384*(10), pp. 905–914.

14. Chen, P., Nirula, A., Heller, B., Gottlieb, R.L., Boscia, J., Morris, J., Huhn, G., Cardona, J., Mocherla, B., Stosor, V. and Shawa, I., 2021. SARS-CoV-2 neutralizing antibody LY-CoV555 in outpatients with Covid-19. *New England Journal of Medicine*, *384*(3), pp. 229–237.

15. Weinreich, D.M., Sivapalasingam, S., Norton, T., Ali, S., Gao, H., Bhore, R., Musser, B.J., Soo, Y., Rofail, D., Im, J. and Perry, C., 2021. REGN-COV2, a neutralizing antibody cocktail, in outpatients with Covid-19. *New England Journal of Medicine*, *384*(3), pp. 238–251.

16. Food and Drug Administration, 2020. Fact sheet for health care providers; Emergency Use Authorization (EUA) of casirivimab and imdevimab. *Available at: Fact Sheet For Health Care Providers Emergency Use Authorization (Eua) Of Bamlanivimab And Etesevimab (fda.gov)*.

17. Covid, C.D.C., Team, V.B.C.I., Birhane, M., Bressler, S., Chang, G., Clark, T., Dorough, L., Fischer, M., Watkins, L.F., Goldstein, J.M. and Kugeler, K., 2021. COVID-19 vaccine breakthrough infections reported to CDC—United States, January 1–April 30, 2021. *Morbidity and Mortality Weekly Report*, *70*(21), p. 792.

18. Brown, C.M., 2021. Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings—Barnstable County, Massachusetts, July 2021. *MMWR. Morbidity and mortality weekly report*, *70*.

19. Fillâtre, P., Dufour, M.J., Behillil, S., Vatan, R., Reusse, P., Gabellec, A., Velmans, N., Montagne, C., Du Coudret, S.G., Droumaguet, E. and Merour, V., 2022. A new SARS-CoV-2 variant with high lethality poorly detected by RT-PCR on nasopharyngeal samples: an observational study. *Clinical Microbiology and Infection*, *28*(2), pp. 298–e9.

20. Pray, I.W., 2020. COVID-19 outbreak at an overnight summer school retreat—Wisconsin, July–August 2020. *MMWR. Morbidity and mortality weekly report*, *69*.

21. Moreno, G.K., Braun, K.M., Pray, I.W., Segaloff, H.E., Lim, A., Poulsen, K., Meiman, J., Borcher, J., Westergaard, R.P., Moll, M.K. and Friedrich, T.C., 2021. Severe acute respiratory syndrome coronavirus 2 transmission in intercollegiate athletics not fully mitigated with daily antigen testing. *Clinical Infectious Diseases*, *73*(Supplement_1), pp. S45–S53.

22. Winichakoon, P., Chaiwarith, R., Liwsrisakun, C., Salee, P., Goonna, A., Limsukon, A. and Kaewpoowat, Q., 2020. Negative nasopharyngeal and oropharyngeal swabs do not rule out COVID-19. *Journal of clinical microbiology*, *58*(5), pp. 10–1128.

23. Schwartz, N.G., 2020. Adolescent with COVID-19 as the source of an outbreak at a 3-week family gathering—Four states, June–July 2020. *MMWR. Morbidity and mortality weekly report*, *69*.

24. Cao, G., Tang, S., Yang, D., Shi, W., Wang, X., Wang, H., Li, C., Wei, J. and Ma, L., 2020. The potential transmission of SARS-CoV-2 from patients with negative RT-PCR swab tests to others: two related clusters of COVID-19 outbreak. *Japanese Journal of Infectious Diseases*, pp. JJID–2020.

25. Hagan, L.M., 2020. Mass testing for SARS-CoV-2 in 16 prisons and jails—six jurisdictions, United States, April–May 2020. *MMWR. Morbidity and mortality weekly report*, *69*.

26. Szablewski, C.M., 2020. SARS-CoV-2 transmission and infection among attendees of an overnight camp—Georgia, June 2020. *MMWR. Morbidity and mortality weekly report*, *69*.

Chapter 4

MORNING SARS-COV-2 TESTING YIELDS BETTER DETECTION OF INFECTION DUE TO HIGHER VIRAL LOADS IN SALIVA AND NASAL SWABS UPON WAKING

1. Viloria Winnett, A.*, Porter, M.K.*, Romano, A.E., Savela, E.S., Akana, R., Shelby, N., Reyes, J.A., Schlenker, N.W., Cooper, M.M., Carter, A.M. and Ji, J., 2022. Morning SARS-CoV-2 testing yields better detection of infection due to higher viral loads in saliva and nasal swabs upon waking. *Microbiology Spectrum*, *10*(6), pp. e03873–22.

Author Contributions

Listed alphabetically by last name

Reid Akana (RA): collaborated with AVW in creating digital participant symptom surveys; assisted with data quality control/curation with JJ, NWS, NS; created current laboratory information management system (LIMS) for specimen logging and tracking. Creation of iOS application for specimen logging/tracking. Configured an SQL database for data storage. Created an Apache server and websites to view study data. Configured FTPS server to catalog PCR data. Wrote a Python package to access study data. Worked with ESS, AVW, AER to implement logic that prioritized specimen extraction order.

Jacob T. Barlow (JTB): Created initial specimen tracking database to aid in specimen logging and tracking. Maintenance of database and implementation of corrections.

Alyssa M. Carter (AMC): Received, performed QC on, and logged specimens. Performed nucleic acid extractions and RT-qPCR. Aliquoting and preparing study reagents as needed. Performed preliminary experiments to assess RNA stability in our saliva and nasal swab specimens. Summarizing daily RT-qPCR data of participant time courses to inform participant keep/drop decisions. Provided feedback on early figure drafts.

Matthew M. Cooper (MMC): Collaborated with AVW, MF, NS, YG, RFI, on study design and recruitment strategies. Co-wrote initial IRB protocol and informed consent with AVW and NS; assisted in the writing of the enrollment questionnaire; developed laboratory specimen processing workflow for saliva with AVW and AER; performed specimen processing on subset of specimens; funding acquisition; collaborated with AER to write data processing/visualization code for observing household transmission events for active study participants. Contributor to the design of the calibration curve for saliva LOD experiments. Performed specimen logging and QC.

Matthew Feaster (MF): Co-investigator; collaborated with AVW, MMC, NS, YG, RFI on study design and recruitment strategies; provided guidance and expertise on SARS-CoV-2 epidemiology and local trends.

Ying-Ying Goh (YG): Co-investigator; collaborated with AVW, MMC, NS, MF, RFI on study design and recruitment strategies; provided guidance and expertise on SARS-CoV-2 epidemiology and local trends.

Rustem F. Ismagilov (RFI): Co-investigator; collaborated with AVW, MMC, NS, MF, YYG on study design and recruitment strategies; provided leadership, technical guidance, oversight, and was responsible for obtaining funding for the study.

Jenny Ji (JJ): Researched epidemiological survey structures, performed epidemiological literature review with MMC and AVW, and co-wrote enrollment questionnaire with NS and AVW. Major contributor to curation of participant symptom data. Provided quality control of participant data with RA, NS, NWS.

Michael K. Porter (MKP): Performed specimen logging and QC, RNA extractions, RT-qPCR, data processing. Performed data acquisition and analysis for and made Figure S2 with AVW. Prepared participant specimen collection materials and helped with supplies acquisition. Assisted in literature analysis with ESS, RA, AVW. Performed data analysis and

prepared Figure 4.2 with AVW. Assisted in preparation of Figures 4.1, 4.3, 4.S2, 4.S3 with AVW. Verified the underlying data with AVW and NS. Outlined and wrote manuscript with AVW.

Jessica A. Reyes (JAR): Lead study coordinator; collaborated with NS, AVW, NWS, and RFI on recruitment strategies, translated study materials into Spanish, co-wrote informational sheets with AVW and NS; created instructional videos for participants; enrolled and maintained study participants with NS and NWS.

Anna E. Romano (AER): Developed laboratory swab specimen processing workflow with ESS. Optimized extraction protocols working with vendor scientists. Created budgets and managed, planned, and purchased reagents and supplies; developed and validated method for RT-qPCR analysis for saliva and nasal-swab specimens with MMC, and AVW. Performed specimen logging and QC, RNA extractions, RT-qPCR; Design of saliva calibration curve experiment. Managing logistics for the expansion of the BSL-2+ lab space with ESS. Provided feedback on earlier manuscript revision and provided a few references.

Emily S. Savela (ESS): Coordinated laboratory team schedules and division of lab work. Performed initial nasal swab workflow validation experiments with AER. Major contributor to workflow validation, methods, biosafety SOPs. Developed/implemented system for specimen archiving. Performed specimen logging and QC, RNA extractions, RT-qPCR, and data processing. Managing logistics for the expansion of the BSL-2+ lab space with AER and biohazardous waste pickups.

Natasha Shelby (NS): Study administrator; collaborated with AVW, MMC, RFI, YG, MF on initial study design and recruitment strategies; co-wrote IRB protocol and informed consent with AVW and MMC; co-wrote enrollment questionnaire with AVW and JJ; co-wrote participant informational sheets with AVW and JAR; enrolled and maintained study participants with JAR and NWS; study-data quality control, curation and archiving with RA, JJ, and NWS; reagents and supplies acquisition; assembled Table 1 with AVW; managed

citations and reference library; verified the underlying data with MKP and AVW; edited the manuscript.

Noah W. Schlenker (NWS): Study coordinator; collaborated with NS, AW, JAR, and RFI on recruitment strategies; enrolled and maintained study participants with NS and JAR; study-data quality control, curation and archiving with RA, JJ, and NS. Colten Tognazzini (CT): Coordinated the recruitment efforts at PPHD with case investigators and contact tracers; provided guidance and expertise on SARS-CoV-2 epidemiology and local trends.

Alexander Viloria Winnett (AVW): Collaborated with MMC, NS, RFI, YG, MF on initial study design and recruitment strategies; co-wrote IRB protocol and informed consent with MMC and NS; co-wrote enrollment questionnaire with NS and JJ; co-wrote participant informational sheets with NS and JAR and digital survey; developed and validated methods for saliva and nasal-swab specimen collection; developed and validated methods for RT-qPCR analysis for saliva and swab specimens with AER, ESS, MMC; reagents and supplies acquisition; funding acquisition; developed laboratory specimen processing workflow with AER, ESS, and MMC; performed specimen logging and QC, nucleic acid extraction, RT-qPCR; assembled Table 1 with NS; analyzed viral load time series data to visualize trends (Figure 4.1) with MKP; assisted MKP in the preparation of Figure 4.2; analyzed viral load data to generate Figure 4.3 with MKP; analyzed specimen collection data for Figure 4.S1; generated longitudinal RT-qPCR plot array for Figure 4.S2, Figure 4.S3 with MKP; analyzed data for Figure 4.S4. Verified the underlying data with MKP and NS. Outlined and wrote manuscript with MKP.

Abstract

Optimizing specimen collection methods to achieve the most reliable SARSCoV-2 detection for a given diagnostic sensitivity would improve testing and minimize COVID-19 outbreaks. From September 2020 to April 2021, we performed a household-transmission study in which participants self-collected specimens every morning and evening throughout acute SARS-CoV-2 infection. Seventy mildly symptomatic participants collected saliva, and of those, 29 also collected nasal swab specimens. Viral load was quantified in 1,194 saliva and 661 nasal swab specimens using a high analytical-sensitivity reverse transcription-quantitative PCR (RT-qPCR) assay. Viral loads in both saliva and nasal swab specimens were significantly higher in morning-collected specimens than in evening-collected specimens after symptom onset. This aspect of the biology of SARS-CoV-2 infection has implications for diagnostic testing. We infer that morning collection would have resulted in significantly improved detection and that this advantage would be most pronounced for tests with low to moderate analytical sensitivity. Collecting specimens for COVID-19 testing in the morning offers a simple and low-cost improvement to clinical diagnostic sensitivity of low- to moderate-analytical-sensitivity tests.

Importance

Our findings suggest that collecting saliva and nasal swab specimens in the morning immediately after waking yields higher SARS-CoV-2 viral loads than collection later in the day. The higher viral loads from morning specimen collection are predicted to significantly improve detection of SARS-CoV-2 in symptomatic individuals, particularly when using moderate- to low-analytical-sensitivity COVID-19 diagnostic tests, such as rapid antigen tests.

Introduction

Although vaccination has substantially reduced hospitalizations and death from COVID-19, limited vaccine uptake and availability and the potential for breakthrough infections

(particularly with novel viral variants) support the continued necessity for diagnostic testing and subsequent isolation of infected individuals (1, 2). Optimizing how diagnostics are used can enhance our ability to combat the COVID-19 pandemic.

Nasopharyngeal swab, anterior nares swab, mid-turbinate swab, oropharyngeal swab, buccal swab, gingival crevicular fluid, sputum, tracheal aspirate, and saliva have all been utilized and compared as diagnostic specimens for the detection of SARS-CoV-2 infection. Work done by many groups (3–5), including ours (6), has suggested that SARS-CoV-2 is detectable, albeit at low viral loads, in saliva before anterior nares nasal swab specimens. However, conflicting results have been reported in head-to-head comparisons of saliva to other specimen types in cross-sectional studies.

Lack of clarity on which specimen type is most reliable for SARS-CoV-2 detection is likely due to the dynamic nature of viral loads in different specimen types through the course of an infection (3, 6–11) and the differences in analytical sensitivity of diagnostic assays used in the comparisons. Currently available SARS-CoV-2 diagnostics span a wide (6 orders of magnitude) range of analytical sensitivities, from the reverse transcription-PCR (RT-PCR) PerkinElmer new coronavirus nucleic acid detection kit (LOD of 180 nucleic acid amplification test detectable units [NDU]/mL) (12) to the Coris BioConcept rapid antigen lateral flow assay COVID-19 Ag Respi-Strip (LOD of ~4×10^7 copies/mL) (13). Tests with relatively moderate analytical sensitivity (LOD of 10^4 to 10^5 copies/mL of specimen) or low analytical sensitivity (LOD of 10^5 to 10^8 copies/mL of specimen) are being increasingly used, particularly for at-home and rapid screening testing and in areas of the world with limited laboratory capacity (14–16).

How specimens are collected can also affect the detectability of SARS-CoV-2 in a specimen. Because SARS-CoV-2, like other pathogens, may exhibit circadian rhythms to replication kinetics (17, 18), we hypothesized that collection time may impact SARS-CoV-2 viral load in respiratory specimens and therefore detectability of infection. Simple, low-cost changes to specimen collection protocols that significantly improve the

clinical sensitivity of COVID-19 diagnostics offer an immediately actionable opportunity to improve existing diagnostics, which would be particularly valuable in settings that rely on tests with low analytical sensitivity.

We conducted a COVID-19 household transmission study (9, 19) where participants prospectively self-collected saliva and nasal swab specimens twice per day (in the morning and in the evening). From mildly symptomatic participants, we compared SARS-CoV-2 viral loads in morning- and evening-collected specimens to determine if the time of day affected viral load, and if this could be leveraged to improve detection of SARS-CoV-2 infection.

Results

Timing of morning and evening specimen collection

Viral load was quantified in 1,194 saliva specimens from 70 individuals and 661 nasal swab specimens from 29 individuals (Figure 4.1). The distribution of collection times was roughly bimodal. Although each participant's specimen collection time varied slightly throughout enrollment, nearly all (92%) participants had an average morning specimen collection time between 7 a.m. and 10 a.m. Evening collection time was more variable, but most participants (74%) had an average specimen collection time between 8 p.m. and 11 p.m. These patterns were used to delineate the morning and evening periods in the study: we defined sampling upon waking (4 a.m. to 12 p.m.) as morning and sampling before bed (3 p.m. to 3 a.m.) as evening (see Figure 4.S1 in the supplemental material).

Saliva and nasal swab specimens exhibit higher viral loads in morning than evening collection across the course of acute, symptomatic illness

Saliva and nasal swab viral load profiles from most individuals (Figure 4.S2 and 4.S3) revealed a pattern of higher viral loads in specimens collected in the morning than in those collected in the evening. In specimens from some individuals (e.g., Figure 4.S2A

and 4.S3E), fluctuations in both SARS-CoV-2 and human *RNase P* markers were observed, whereas in others *RNase P* remained stable and SARS-CoV-2 viral load appeared to be independent of the host marker (e.g., Figures 4.S2AH and 4.S3N).

Although direct comparison between all positive morning or evening specimens demonstrates greater target abundance for both SARS-CoV-2 *N1* (Figures 4.S4A and C) and human *RNase P* (Figures 4.S4B and D), this comparison would be skewed by participants who contributed more specimens and biased by sampling at different stages of the infection. To minimize these potential biases, the time of each specimen collection was aligned relative to the date of symptom onset for that participant before plotting both individual viral load datapoints (Figure 4.S2 and 4.S3) and the average of log-transformed viral load values (Figure 4.1A and B) for all saliva and nasal swab specimens in twelve-hour time bins.

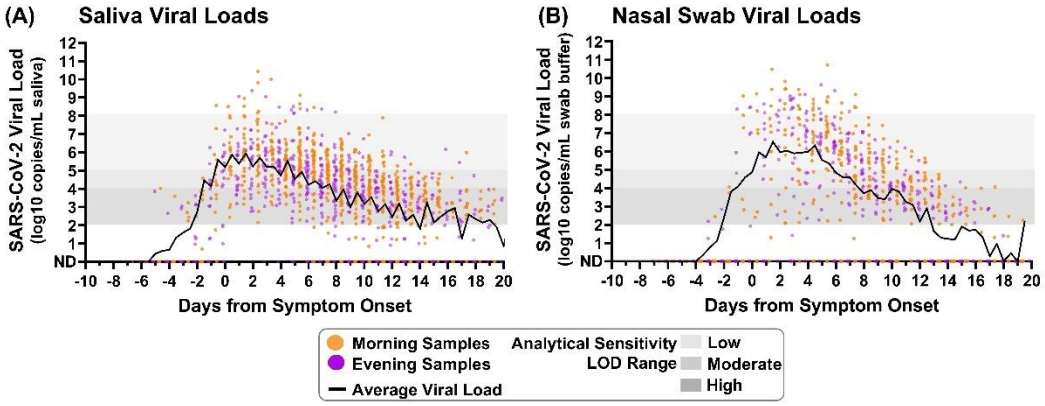


Figure 4.1: Saliva and nasal swab specimens collected in the morning and evening through the course of infection demonstrate differences in SARS-CoV-2 viral load. Black lines on each plot indicate the average viral load for each daily morning or evening specimen collection window. (A) Saliva specimen viral load (SARS-CoV-2 *N1* copies/milliliter of saliva) as measured by RT-qPCR is plotted relative to symptom onset for 1,194 specimens. (B) Nasal swab specimen viral load (*N1* copies/milliliter of swab buffer) as measured by RT-qPCR is plotted relative to symptom onset for 661 specimens. Specimens were designated morning (orange) if collected between 4 a.m. and 12 p.m. or evening (purple) if collected between 3 p.m. and 3 a.m. ND, not detected. Additional specimen details are provided in the supplemental material.

The averaged salivary viral load during each collection time point visually suggests higher viral loads in specimens collected in the morning than in the evening during both the presymptomatic and symptomatic phases of infection. This pattern was less apparent in the averaged nasal swab viral loads but can be seen when comparing the *N1* threshold cycle (C_T) values between successive time points by calculating differences in C_T (Figure 4.2A and B). Only reverse transcription-quantitative PCR (RT-qPCR) C_T values for pairs of successively collected morning-to-evening or evening-to-morning specimen were used to calculate the C_T difference; negative or indeterminate specimens were included only if directly followed by a positive specimen collected in the presymptomatic phase of infection. A negative difference in C_T values indicates that viral load was increasing relative to the previous measurement, whereas a positive difference indicates that viral load was decreasing relative to the previous measurement. Starting from symptom onset (day 0), saliva specimens collected in the morning typically exhibited a negative difference in C_T values relative to their preceding evening specimens, whereas evening specimens consistently had a positive difference in C_T values relative to their preceding morning specimens. This indicates that throughout the course of symptomatic infection, morning specimens typically result in relatively lower C_T values (higher viral loads) than evening specimens.

To further illustrate the pattern observed in viral loads and changes in C_T values, specimens were binned by infection stage: prior to symptom onset and in 4-day intervals relative to symptom onset. The 4-day interval was selected to capture reasonable resolution for infection stage while also providing sufficient measurements to observe potential differences. Significantly higher morning viral loads were not observed prior to symptom onset in either specimen type in the limited number of specimens collected during this period. However, significantly higher viral loads ($P < 0.05$, Wilcoxon matched-pair signed-rank test) were observed in saliva specimens collected in the morning for the first 16 days of symptomatic infection (Figure 4.2C). Differences in C_T values were also significantly lower ($P < 0.05$, Wilcoxon matched-pair signed-rank

test) in morning nasal swab specimens from day 4 to day 16 of symptomatic infection (Figure 4.2D). Of note, nasal swab viral load appears to increase more quickly to peak than does salivary viral load (Figures 4.1A and B), and nasal swabs also achieve higher peak viral loads (Figure 4.S4C) than does saliva (Figure 4.S4A); the high rate of increase in viral load in nasal swabs likely obscures subtle daily fluctuations that are more apparent in saliva, where viral load rises more gradually (19). Nasal swabs appear to also be subject to more sampling variability (Figures 4.S3 and 4.S4D) than saliva (Figures 4.S2 and 4.S4B), evidenced by *RNase P* control marker C_T values.

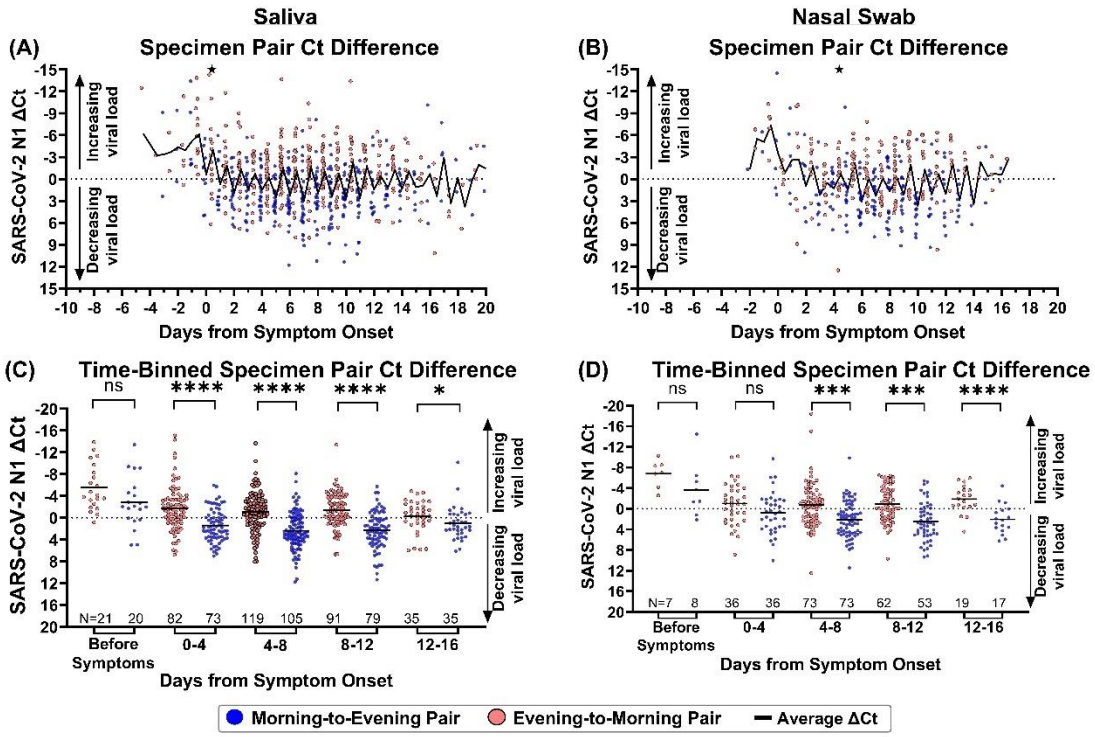


Figure 4.2: Morning viral loads are significantly higher than evening viral loads during most of SARS-CoV-2 infection. (A and B) The difference in *N1* C_T values (ΔC_T) in 703 morning-to-evening and evening-to-morning successive saliva specimen pairs (A) and 365 morning-to-evening and evening-to-morning successive nasal swab specimen pairs (B), plotted relative to symptom onset. One point in panel A and one point in panel B had ΔC_T values outside the *y* axis of the plot; these are represented as black stars at −15. (C and D) The difference in *N1* C_T values in 703 morning-to-evening and evening-to-morning sequential saliva (C) and nasal swab (D) specimen pairs relative to symptom onset. Morning-to-evening or evening-to-morning ΔC_T values were then binned into presymptomatic or 4-day infection stages. The distributions of morning-to-evening and evening-to-morning ΔC_T values for each infection stage bin were then statistically compared using the Wilcoxon matched-pair signed-rank test; ns, nonsignificant or insufficient data points to perform analysis; *, $P < 0.05$; **, $P < 0.01$; ****, $P < 0.001$. Black lines indicate average viral load. ND, not detected.

Saliva and nasal swab viral loads in the range of moderate- and low-sensitivity tests underscore utility of morning sampling

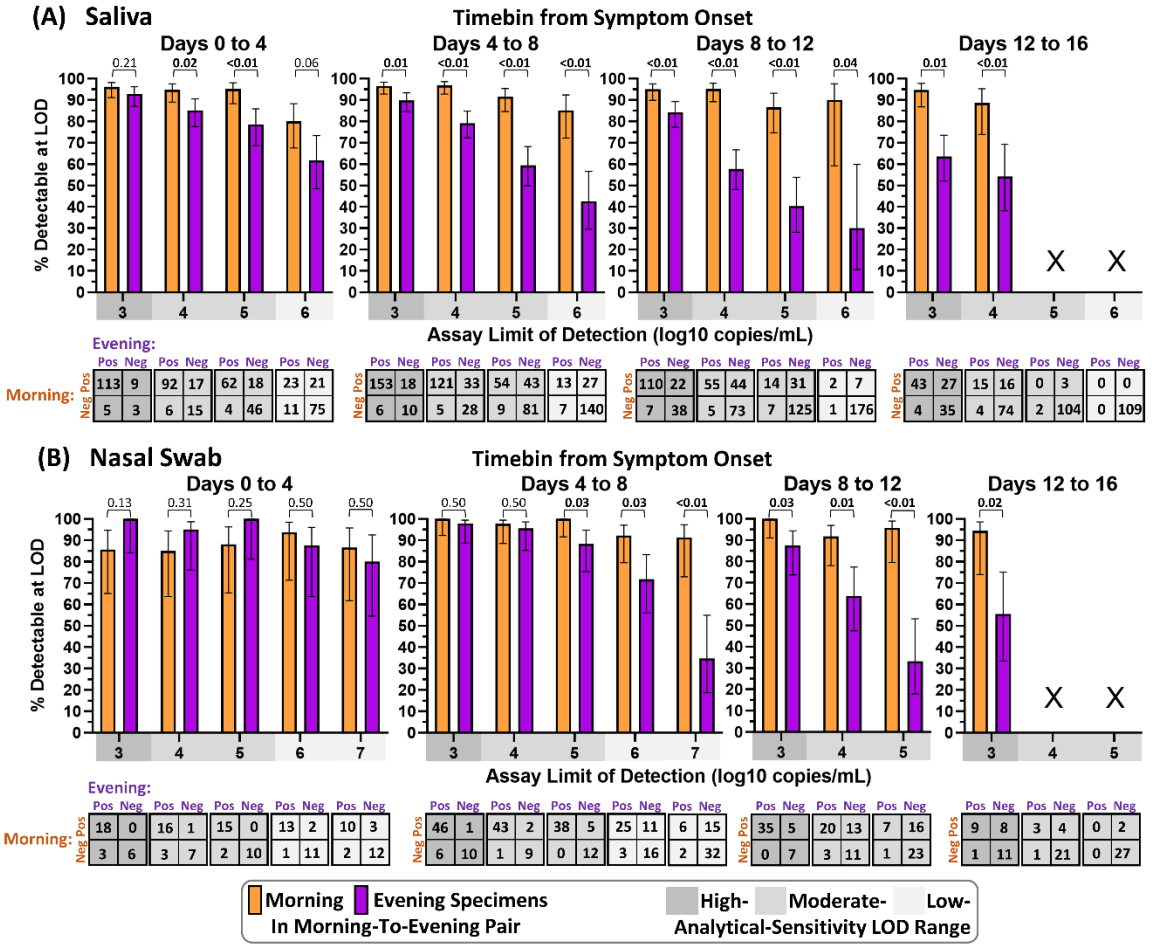
The observed higher viral loads in specimens collected in the morning upon waking than in those collected later in the day led us to hypothesize that sampling in the morning could detect significantly more infected individuals than sampling in the evening. Because viral loads rise and decline throughout the course of the infection (Figure 4.1), we assessed this hypothesis during discrete 4-day time bins following symptom onset. The presymptomatic period was not assessed, as few specimens from this period were available for analysis. Additionally, because COVID-19 diagnostics have analytical sensitivities that span several orders of magnitude, we tested this hypothesis for assays with LODs of 10^3, 10^4, 10^5, 10^6, 10^7 copies/mL; quantitative viral loads measured in each specimen were used to predict whether each specimen would reliably yield a positive result when tested by an assay of each LOD. For each time bin and each LOD, we generated two-by-two matrices to assess the detectability of morning or evening sampling within pairs of sequentially collected morning-to-evening (Figure 4.3) specimens. Each time bin and LOD that did not contain at least ten positive samples from saliva or nasal swab were excluded from this analysis.

For saliva specimens, the advantage of morning sampling was statistically significant in all but two comparisons (Figure 4.3A); the two comparisons for which a nonsignificant advantage was observed occurred in the first 4 days of infection, at the LODs of the lowest- and highest-analytical-sensitivity assays (LODs of 10^6 and 10^3 copies/mL, respectively). As LOD increases, fewer pairs are predicted to have detectable virus in either the morning or evening specimen; for this reason, confidence intervals widen as the LOD increases, which results in decreased power to detect significant differences in detection by assays with higher LODs. Additionally, assays with lower LODs are able to reliably detect lower viral concentrations, decreasing the impact of fluctuations in viral load from morning to evening sampling on detection.

Morning sampling with nasal swab specimens also exhibited an advantage over evening sampling after 4 days from symptom onset, for all LODs (Figure 4.3B). In the first 4 days of infection, a nonsignificant advantage of evening over morning sampling was observed; in this phase of the infection, viral loads in nasal swab specimens typically rise rapidly from undetectable to high (Figure 4.1). Therefore, during this rapid rise, the specimen collected later within a pair of successively collected specimens would improve detection; indeed, when morning-to-evening pairs were assessed (Figure 4.3), the later (evening) time point had improved detection but when evening-to-morning pairs were assessed (Figure 4.S5), the later (morning) time point resulted in improved detection.

Similarly, when viral loads are declining, one may expect the earlier time point within a pair of successively collected specimens to exhibit improved detection. We assessed whether this effect was responsible for the improved performance of morning sampling over evening sampling when pairs of successively collected morning-to-evening specimens were compared by performing an equivalent analysis of pairs of successively collected evening-to-morning specimens (Figure 4.S5). Even with evening-to-morning pairing, morning sampling exhibited an advantage over evening sampling for all comparisons with saliva and nearly all comparisons with nasal swabs. In the three of 12 comparisons where morning sampling with nasal swabs did not exhibit an advantage, two comparisons had equivalent detection by morning or evening sampling, and in the third comparison evening sampling exhibited only a nonsignificant advantage of less than 2% over morning sampling.

This supports that the advantage of morning sampling over evening sampling for both saliva and nasal swabs was robust to whether the morning specimen is collected prior to or following the evening specimen. These results suggest that collecting saliva or nasal swab specimens for SARS-CoV-2 testing in the morning, immediately after waking, can significantly improve detection of symptomatic, infected individuals.



[Figure on previous page] Figure 4.3: Morning saliva or nasal swab specimen collection yields improved detection across infection stages and assay analytical sensitivities. For each 4-day time bin relative to symptom onset, pairs of sequentially collected morning-to-evening specimens were assessed. In each pair, the viral load in each specimen was used to predict a positive or negative result if tested by an assay with a given limit of detection (LOD) below or above the viral load, respectively. Bar plots show the fraction of pairs with a positive result in either the morning or evening specimen that would be detectable if the morning specimen (orange) or evening specimen (purple) were tested at a given LOD. Error bars indicate the 95% confidence interval. Bars are not shown (X) when fewer than 10 pairs had positive results at the given LOD during the infection time bin. Among LODs and infection time bins with more than 10 positive pairs, the percents detectable for morning versus evening specimens were compared by an upper-tailed McNemar exact test, applied to the 2×2 table shown below each comparison. Resulting *P* values are shown above each comparison. Boldfaced values indicate significantly higher detection with morning sampling than with evening sampling. Analysis was performed on saliva specimens (A) and nasal swab specimens (B). Equivalent analysis for evening-to-morning pairs is shown in Figure 4.S5 in the supplemental material. Pos, positive; Neg, negative.

Discussion

In this study, we quantitatively measured SARS-CoV-2 viral load with high frequency (twice per day) longitudinally through the course of mild COVID-19 infection in saliva for 70 individuals and in nasal swabs for 29 individuals. From these measurements, we identified a pattern of higher viral loads in saliva and nasal swab specimens collected in the morning after waking than in those collected in the evening. Although similar observations have been reported for nasopharyngeal swabs (20, 21), early morning versus spot oropharyngeal specimens (22), and early morning saliva versus nasopharyngeal swabs (23) and in wastewater surveillance (24), our study is unique and clinically relevant for three reasons: (i) we measured viral load in specimen types relevant to at-home testing using a high-analytical-sensitivity RT-qPCR assay, which enabled us to infer the performance of diagnostic tests of different analytical sensitivities at each stage of infection; (ii) we collected specimens at high temporal resolution (morning and evening) longitudinally for 2 weeks, starting from early in the course of the infection via prospective sampling of high-risk populations; and (iii) our study provides the largest data set to date that investigates daily patterns in SARS-CoV-2 viral loads, with 1,194 saliva and 661 nasal swab specimens collected longitudinally. From these data, we find compelling evidence that collecting samples for COVID-19 testing in the morning upon waking can significantly improve detection of infected individuals.

The biological and physiological reasons for higher SARS-CoV-2 viral loads in the morning remain unknown but may be due to accumulation of viral material overnight or related to viral replication and immune function. Similar to the improved performance of at-home pregnancy tests with morning urine due to accumulation of human chorionic gonadotropin (25), improved detection of SARS-CoV-2 may be the result of physical accumulation of material (e.g., cells, virions, and nucleic acids) in the upper respiratory tract due to supine positioning (aiding mucociliary clearance) and/or the decreased rate of swallowing at night (26). Higher morning viral loads being due to physical accumulation of nucleic acids is supported by an increased abundance of the constitutive

human *RNase P* target in saliva and nasal swab specimens collected in the morning (see Figures 4.S2A and 4.S3A in the supplemental material). Human salivary production decreases overnight (27), suggesting that higher morning viral loads could be due to a concentration of virus when saliva volume is lower. Given that some individuals exhibit this phenomenon independently of human *RNase P* target abundance (Figures 4.S2B and 4.S3B), a circadian rhythm in viral replication may also contribute. Regulation and responsiveness of the immune system have been linked to circadian rhythms (28, 29), shown to affect SARS-CoV-2 infection of monocytes in cell culture (30) and proposed as a modulating factor for COVID-19 severity and management (31). Others have proposed cellular interactions between viral proteins and circadian rhythm-dependent host signals (32) and demonstrated circadian rhythm-dependent entry and proliferation of SARS-CoV-2 in lung epithelial cell types in culture (33). Regardless of mechanism, because higher viral loads are associated with replication-competent culturable virus (34, 35), these findings may also suggest a higher risk of transmission in the morning.

As many individuals remain unvaccinated and new variants emerge, it remains critical to identify infections, promptly isolate infected persons, trace and quarantine contacts, and initiate early treatment to improve efficacy. Much of the world lacks access to tests with high analytical sensitivity (36–38). Our findings suggest that strategically collecting specimens in the morning immediately after waking up may improve the performance of available low- to moderate-analytical-sensitivity tests. Morning sampling will not raise the performance of tests with low analytical sensitivity to the levels of those with higher analytical sensitivity; however, even marginal improvements in detection have been shown to reduce deaths from COVID-19 (39).

This study is subject to five main limitations. First, we had a limited number of specimens collected prior to the onset of symptoms, limiting our ability to discern a difference in detectability with morning or evening specimens during the presymptomatic phase of infection. Second, this study was performed prior to the dominance of the Delta and Omicron variants of SARS-CoV-2, which may exhibit different viral load kinetics. Host

factors, including vaccination status, may also influence viral load kinetics; nearly all individuals in this cohort were unvaccinated. Third, specimens were self-collected without supervision and thus may have had a different quality from those collected by a health care professional. However, many COVID-19 diagnostics in use utilize self-collected specimens, and measurements of the human *RNase P* gene suggest consistent sampling without failure to collect sufficient material. Fourth, we quantified viral load using RT-qPCR with SARS-CoV-2 *N* gene target. Many COVID-19 diagnostics utilize *N* gene targets, and *N* gene viral loads have been shown to track with other gene targets, suggesting that *N* gene quantification to viral load conversion would be representative to demonstrate a general phenomenon relevant for diagnostics detecting other viral targets. Fifth, this analysis involves inferring positivity by assays with various analytical sensitivities (LODs), based on the quantitatively measured viral loads. A direct comparison with a specific test is needed to test real-world efficacy.

Materials and Methods

Study design

Participants were recruited for participation in a COVID-19 household transmission study as previously described (9, 19). Briefly, if at least one member of a household with two or more persons had a positive COVID-19 test result within 7 days or was suspected to be positive, all household members aged 6 years and older were eligible to participate. Participants began collecting saliva or saliva and nasal swab specimens on the evening of enrollment and each subsequent morning and evening (as described below). COVID-19-like symptoms were reported via questionnaire with each specimen collection time point.

For participants who were SARS-CoV-2 positive when initially enrolled in the study, symptom onset was defined as the date of first symptoms reported in an enrollment questionnaire. For participants who entered the study SARS-CoV-2 negative but had unrelated symptoms, symptom onset was the first instance of a new COVID-19-like

symptom or an increase in symptom severity following their first SARS-CoV-2-positive specimen.

Specimen collection

Participants self-collected anterior nares nasal swab and saliva specimens in the Spectrum SDNA-1000 Saliva Collection Kit (Spectrum Solutions LLC, Draper, UT), at home twice per day (after waking up and before going to bed), per manufacturer's guidelines (although Spectrum devices are not currently authorized for the collection of nasal swab specimens). One participant self-collected both anterior nares nasal swab and saliva specimen in Nest viral transport medium (VTM) (catalog no. NST-NST-202117; Stellar Scientific, Baltimore, MD), and three individuals collected their nasal swab specimens in VTM and their saliva specimens in the Spectrum SDNA-1000 Saliva Collection Kit. Participants were instructed not to ingest anything, smoke, or brush their teeth for at least 30 min prior to collection. For nasal swab collection, participants were asked to gently blow their noses before swabbing (four complete rotations with gentle pressure in each nostril) with sterile flocked swabs. A parent/guardian assisted minors with collection. At collection, participants recorded the date and time and any symptoms experienced in the previous 12 h. Specimens collected between 4 a.m. and 12 p.m. were defined as morning; specimens collected between 3 p.m. and 3 a.m. were defined as evening (see Figure 4.S1 in the supplemental material).

Cohort of individuals with SARS-CoV-2 infection

Between September 2020 and April 2021, 72 participants from 39 households in southern California had acute SARS-CoV-2 infection. Of these, two never reported experiencing symptoms and were not included in subsequent analyses where viral loads are aligned with date of symptom onset. Of the 70 symptomatic individuals from 37 households included in the analyses (Table 4.1), all 70 collected saliva specimens while a subset of 29 individuals collected both saliva and nasal swab specimens every morning and every evening while enrolled, from which we quantified viral loads.

Individuals were enrolled at various stages of infection. Of the 70 infected, symptomatic individuals, 58 were positive for SARS-CoV-2 in the first saliva or saliva and nasal swab specimen collected upon enrollment while twelve were initially negative but became positive while enrolled in the study; of these twelve individuals, seven were collecting both saliva and nasal swabs, and the viral loads and symptoms of these individuals have been previously reported (6). Of the 58 cases positive on enrollment, 50 (86.2%) were already experiencing mild COVID-19-like symptoms and 8 (13.8%) were presymptomatic. Of the 20 individuals who were either presymptomatic (8) or negative for SARS-CoV-2 (12) on enrollment, COVID-19 symptom onset occurred an average of 1.2 days after the first SARS-CoV-2-positive saliva specimen.

The mean age of the saliva cohort was 32.8 years (standard deviation [SD], ±16.0 years), and the mean age was 33.9 years (SD, ±15.2 years) among those collecting both saliva and nasal swabs. Health conditions and medications that may have impacted viral load kinetics are provided for individual participants in the supplemental material. No participants required hospitalization. At the time of these participants' enrollment in the study (September 2020 to April 2021), vaccines were either unavailable or limited to priority groups. Only one individual (Figures 4.S2H and 4.S3H) reported receiving a COVID-19 vaccine (first dose of Pfizer-BioNTech COVID-19, ~3 weeks before enrollment).

Table 4.1: Demographic and medical information was collected via online questionnaire upon study enrollment. All participants (No.=70) collected saliva; of these 70, 29 additionally collected nasal swabs

	Participants Contributing Sample Type			
	Saliva		**Saliva and Nasal Swabs**	
	70		29	
Sex*				
Male	25	35.7%	9	31.0%
Female	45	64.3%	20	69.0%
Age				
6–11	6	8.6%	1	3.4%
12–17	9	12.9%	4	13.8%
18–24	9	12.9%	3	10.3%
25–35	17	24.3%	10	34.5%
36–45	12	17.1%	3	10.3%
46–55	11	15.7%	6	20.7%
56–65	5	7.1%	2	6.9%
65+	1	1.4%	0	0.0%
Race				
Asian / Pacific Islander	6	8.6%	2	6.9%
Black / African American	2	2.9%	2	6.9%
Native American	0	0.0%	0	0.0%
White	33	47.1%	15	51.7%
Multiple Races	4	5.7%	3	10.3%
Other/Unknown†	25	35.7%	7	24.1%
Ethnicity				
Hispanic	52	74.3%	21	72.4%
Non-Hispanic	17	24.3%	8	27.6%
Unknown	1	1.4%	0	0.0%

Tobacco Smoker or Vape User History				
Current	5	7.1%	3	10.3%
Former	15	21.4%	9	31.0%
Never	43	61.4%	16	55.2%
Unknown	7	10.0%	1	3.4%
Active Medications and Supplements				
Vitamins/Supplements	47	67.1%	21	72.4%
Acetaminophen/NSAIDs‡	33	47.1%	13	44.8%
Allergy medications/Antihistamines	11	15.7%	3	10.3%
Antibiotics/Antivirals	3	4.3%	0	0.0%
Steroid drug	3	4.3%	1	3.4%
Medical Comorbidities				
Asthma	6	8.6%	1	3.4%
Anxiety or Depression	4	5.7%	2	6.9%
Diabetes	4	5.7%	3	10.3%
Obesity	4	5.7%	2	6.9%
Hypertension	3	4.3%	1	3.4%
Immunocompromise	0	0.0%	0	0.0%
SARS-CoV-2 Vaccination Status§				
Partially vaccinated	1	1.4%	1	3.4%
Completed vaccination	0	0.0%	0	0.0%
No vaccines reported	69	98.6%	28	96.6%

Extraction and quantification of viral load by RT-qPCR

Specimen processing was performed as previously described (9). Briefly, 400 or 200 µL of fluid from each saliva or nasal swab specimen, respectively, was extracted using the MagMAX Viral/Pathogen nucleic acid isolation kit (ThermoFisher Scientific; catalog no. A42352), followed by the CDC 2019-novel coronavirus (2019-nCoV) real-time RT-PCR diagnostic panel, which targets the SARS-CoV-2 *N1* and *N2* genes, as well as a human *RNase P* control. *N1* gene C_T values were converted to viral load using an equation derived from a standard curve of heat-inactivated SARS-CoV-2 particles spiked into human specimen matrix validated previously by independent RT-double differential PCR (ddPCR) measurement (6).

Statistical analyses

Initial processing was performed in Python v3.8.2, with calculation of log-transformed averages (Figure 4.1). Data were exported, and differences in C_T from sequential specimens were calculated in Microsoft Excel (Figure 4.2A to D). Plots were prepared in GraphPad Prism 9.2.0, including calculation of medians (Figure 4.2). For comparison of the differences between morning and evening viral loads and differences in C_T values, the Wilcoxon matched-pair signed-rank test was performed using GraphPad (Figure 4.2). An upper-tailed McNemar test to compare inferred percentages of infections detectable by assays with various LODs for specimens collected in the morning or evening (Figure 4.3 and Figure 4.S5) was performed in Python v3.8.2 using the scipy.stats package (40).

Data availability

The data underlying the results presented in the study are available at CaltechDATA at https://data.caltech.edu/records/20049.

Acknowledgements

We sincerely thank the study participants for making this work possible. We thank Lauriane Quenee, Grace Fisher-Adams, Junie Hildebrandt, Megan Hyashi, RuthAnne Bevier, Chantal D'Apuzzo, Ralph Adolphs, Victor Rivera, Steve Chapman, Gary Waters,

Leonard Edwards, Gaylene Ursua, Cynthia Ramos, and Shannon Yamashita for their assistance and advice on study implementation and/or administration. We thank Jessica Leong, Jessica Slagle, Mika Walton, Angel Navarro, Daniel Brenner, Ojas Pradhan, Si Hyung Jin, and Mary Arrastia for volunteering their time to help with this study. We thank Angie Cheng, Susan Magdaleno, Christian Kis, Monica Herrera, and Zaina Lemeir for technical discussions regarding saliva extraction and detection. We thank Jennifer Fulcher, Debika Bhattacharya, and Matthew Bidwell Goetz for their ideas on potential study populations and early study design. We thank Omai Garner and David Beenhouwer for providing materials for initial nasal swab validation. We thank Martin Hill, Alma Sanchez, Scott Kim, Debbie Noble, Nina Paddock, Whitney Harrison, Emily Holman, Isaac Turner, Vivek Desai, Luke Wade, Tom Mayell, Stu Miller, Jennifer Howes, and Nari Shin for their support with recruitment. We thank Allison Rhines, Karen Heichman, and Dan Wattendorf for valuable discussions and guidance. We thank David Prober for discussions of circadian rhythms and feedback on the manuscript. Finally, we thank all the case investigators and contact tracers at the Pasadena Public Health Department, the City of Long Beach Department of Health & Human Services, and Caltech Student Wellness Services for their efforts in study recruitment and their work in the pandemic response.

R.F.I. is a cofounder, consultant, and a director and has stock ownership of Talis Biomedical Corp.

This study is based on research funded in part by the Bill & Melinda Gates Foundation (INV-023124). The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation. This work was also funded by the Ronald and Maxine Linde Center for New Initiatives at the California Institute of Technology and the Jacobs Institute for Molecular Engineering for Medicine at the California Institute of Technology. A.V.W. is supported by a UCLA DGSOM Geffen Fellowship, and M.M.C. is supported by a Caltech Graduate Student Fellowship.

References

1. Wadhwa, A., Fisher, K.A., Silver, R., Koh, M., Arons, M.M., Miller, D.A., McIntyre, A.F., Vuong, J.T., Kim, K., Takamiya, M. and Binder, A.M., 2021. Identification of presymptomatic and asymptomatic cases using cohort-based testing approaches at a large correctional facility—Chicago, Illinois, USA, May 2020. *Clinical Infectious Diseases*, *72*(5), pp. e128–e135.

2. Dora, A.V., 2020. Universal and serial laboratory testing for SARS-CoV-2 at a long-term care skilled nursing facility for veterans—Los Angeles, California, 2020. *MMWR. Morbidity and mortality weekly report*, *69*.

3. Jianyu Lai, B., German, J.R., Hong, F.H., Tai, S.H.S., McPhaul, K.M., Milton, D.K. and University of Maryland StopCOVID Research Group, 2021. Comparison of Saliva and Mid-Turbinate Swabs for Detection of COVID-19 (preprint).

4. Wyllie, A.L., Fournier, J., Casanovas-Massana, A., Campbell, M., Tokuyama, M., Vijayakumar, P., Warren, J.L., Geng, B., Muenker, M.C., Moore, A.J. and Vogels, C.B., 2020. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. *New England Journal of Medicine*, *383*(13), pp. 1283–1286.

5. Ke, R., Martinez, P.P., Smith, R.L., Gibson, L.L., Mirza, A., Conte, M., Gallagher, N., Luo, C.H., Jarrett, J., Conte, A. and Liu, T., 2021. Daily sampling of early SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness. *medRxiv*, pp. 2021–07.

6. Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C. and Hoelscher, M., 2020. Virological assessment of hospitalized patients with COVID-2019. *Nature*, *581*(7809), pp. 465–469.

7. Smith, R.L., Gibson, L.L., Martinez, P.P., Ke, R., Mirza, A., Conte, M., Gallagher, N., Conte, A., Wang, L., Fredrickson, R. and Edmonson, D.C., 2021. Longitudinal

assessment of diagnostic test performance over the course of acute SARS-CoV-2 infection. *The Journal of infectious diseases*, *224*(6), pp. 976–982.

8. Singanayagam, A., Hakki, S., Dunning, J., Madon, K.J., Crone, M.A., Koycheva, A., Derqui-Fernandez, N., Barnett, J.L., Whitfield, M.G., Varro, R. and Charlett, A., 2022. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B. 1.617. 2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. *The lancet infectious diseases*, *22*(2), pp. 183–195.

9. Savela, Emily S., Alexander Viloria Winnett, Anna E. Romano, Michael K. Porter, Natasha Shelby, Reid Akana, Jenny Ji et al. "Quantitative SARS-CoV-2 viral-load curves in paired saliva samples and nasal swabs inform appropriate respiratory sampling site and analytical test sensitivity required for earliest viral detection." *Journal of clinical microbiology* 60, no. 2 (2022): e01785–21.

10. Kissler, S.M., Fauver, J.R., Mack, C., Tai, C.G., Breban, M.I., Watkins, A.E., Samant, R.M., Anderson, D.J., Metti, J., Khullar, G. and Baits, R., 2021. Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated persons. *New England Journal of Medicine*, *385*(26), pp. 2489–2491.

11. Kissler, S.M., Fauver, J.R., Mack, C., Olesen, S.W., Tai, C., Shiue, K.Y., Kalinich, C.C., Jednak, S., Ott, I.M., Vogels, C.B. and Wohlgemuth, J., 2021. Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. *PLoS biology*, *19*(7), p.e3001333.

12. SARS-CoV-2 reference panel comparative data (2020) FDA. Available at: https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/sars-cov-2-reference-panel-comparative-data (Accessed: 28 May 2024).

13. Brümmer, L.E., Katzenschlager, S., Gaeddert, M., Erdmann, C., Schmitz, S., Bota, M., Grilli, M., Larmann, J., Weigand, M.A., Pollock, N.R. and Macé, A., 2021. Accuracy of novel antigen rapid diagnostics for SARS-CoV-2: A living systematic review and meta-analysis. *PLoS medicine*, *18*(8), p.e1003735.

14. Global Partnership to make available 120 million affordable, quality covid-19 rapid tests for low- and middle-income countries (2022) WHO. Available at: https://www.who.int/news/item/28-09-2020-global-partnership-to-make-available-120-million-affordable-quality-covid-19-rapid-tests-for-low--and-middle-income-countries (Accessed: 28 May 2024).

15. Connor, A., Hariharan, N., Carson, S., Sanders, K.C., Vosburg, K.B. and Sabot, O., 2021. Access to COVID-19 testing in low-and middle-income countries is still critical to achieving health equity. *Health Affairs Forefront*.

16. Global availability of COVID-19 diagnostic tests (2022) PATH. Available at: https://www.path.org/programs/diagnostics/covid-dashboard-global-availability-covid-19-diagnostic-tests/ (Accessed: 28 May 2024).

17. Borrmann, H., McKeating, J.A. and Zhuang, X., 2021. The circadian clock and viral infections. *Journal of biological rhythms*, *36*(1), pp. 9–22.

18. Diallo, A.B., Coiffard, B., Leone, M., Mezouar, S. and Mege, J.L., 2020. For whom the clock ticks: clinical chronobiology for infectious diseases. *Frontiers in immunology*, *11*, p. 533252.

19. Winnett, A., Cooper, M.M., Shelby, N., Romano, A.E., Reyes, J.A., Ji, J., Porter, M.K., Savela, E.S., Barlow, J.T., Akana, R. and Tognazzini, C., 2020. SARS-CoV-2 viral load in saliva rises gradually and to moderate levels in some humans. *medRxiv*.

20. McNaughton, C.D., Adams, N.M., Hirschie Johnson, C., Ward, M.J., Schmitz, J.E. and Lasko, T.A., 2021. Diurnal variation in SARS-CoV-2 PCR test results: test accuracy may vary by time of day. *Journal of biological rhythms*, *36*(6), pp. 595–601.

21. Liu, M., Li, Q., Zhou, J., Ai, W., Zheng, X., Zeng, J., Liu, Y., Xiang, X., Guo, R., Li, X. and Wu, X., 2020. Value of swab types and collection time on SARS-COV-2 detection using RT-PCR assay. *Journal of virological methods*, *286*, p. 113974.

22. Hung, D.L.L., Li, X., Chiu, K.H.Y., Yip, C.C.Y., To, K.K.W., Chan, J.F.W., Sridhar, S., Chung, T.W.H., Lung, K.C., Liu, R.W.T. and Kwan, G.S.W., 2020, June. Early-morning vs spot posterior oropharyngeal saliva for diagnosis of SARS-CoV-2 infection: implication of timing of specimen collection for community-wide screening. In *Open Forum Infectious Diseases* (Vol. 7, No. 6, p. ofaa210). US: Oxford University Press.

23. Rao, M., Rashid, F.A., Sabri, F.S., Jamil, N.N., Zain, R., Hashim, R., Amran, F., Kok, H.T., Samad, M.A.A. and Ahmad, N., 2021. Comparing nasopharyngeal swab and early morning saliva for the identification of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Clinical Infectious Diseases*, *72*(9), pp. e352–e356.

24. Bivins, A., North, D., Wu, Z., Shaffer, M., Ahmed, W. and Bibby, K., 2021. Within- and between-day variability of SARS-CoV-2 RNA in municipal wastewater during periods of varying COVID-19 prevalence and positivity. *Acs Es&T Water*, *1*(9), pp. 2097–2108.

25. Center for Devices and Radiological Health (2019) Pregnancy, U.S. Food and Drug Administration. Available at: https://www.fda.gov/medical-devices/home-use-tests/pregnancy (Accessed: 28 May 2024).

26. Uludag, I.F., Tiftikcioglu, B.I. and Ertekin, C., 2016. Spontaneous swallowing during all-night sleep in patients with Parkinson disease in comparison with healthy control subjects. *Sleep*, *39*(4), pp. 847–854.

27. Dawes, C., 1972. Circadian rhythms in human salivary flow rate and composition. *The Journal of physiology*, *220*(3), pp. 529–545.

28. Mazzoccoli, G., Vinciguerra, M., Carbone, A. and Relógio, A., 2020. The circadian clock, the immune system, and viral infections: the intricate relationship between biological time and host-virus interaction. *Pathogens*, *9*(2), p. 83.

29. Niemeyer, J.E., 2017. Viruses and circadian rhythms. *Lab Animal*, *46*(1), pp. 7–7.

30. Diallo, A.B., Gay, L., Coiffard, B., Leone, M., Mezouar, S. and Mege, J.L., 2021. Daytime variation in SARS-CoV-2 infection and cytokine production. *Microbial pathogenesis*, *158*, p. 105067.

31. Ray, S. and Reddy, A.B., 2020. COVID-19 management in light of the circadian clock. *Nature Reviews Molecular Cell Biology*, *21*(9), pp. 494–495.

32. e Cruz, M.M., Miyazawa, M. and Gozal, D., 2020. Putative contributions of circadian clock and sleep in the context of SARS-CoV-2 infection. *European Respiratory Journal*, *55*(6).

33. Zhuang, X., Tsukuda, S., Wrensch, F., Wing, P.A., Schilling, M., Harris, J.M., Borrmann, H., Morgan, S.B., Cane, J.L., Mailly, L. and Thakur, N., 2021. The circadian clock component BMAL1 regulates SARS-CoV-2 entry and replication in lung epithelial cells. *IScience*, *24*(10).

34. Gallichotte, E.N., Quicke, K.M., Sexton, N.R., Fitzmeyer, E., Young, M.C., Janich, A.J., Dobos, K., Pabilonia, K.L., Gahm, G., Carlton, E.J. and Ebel, G.D., 2020. Longitudinal surveillance for SARS-CoV-2 among staff in six Colorado long-term care facilities: epidemiologic, virologic and sequence analysis. *medRxiv*.

35. Perera, R.A., Tso, E., Tsang, O.T., Tsang, D.N., Fung, K., Leung, Y.W., Chin, A.W., Chu, D.K., Cheng, S.M., Poon, L.L. and Chuang, V.W., 2020. SARS-CoV-2 virus culture and subgenomic RNA for respiratory specimens from patients with mild coronavirus disease. *Emerging infectious diseases*, *26*(11), p.2701.

36. Okeke, I.N. and Ihekweazu, C., 2021. The importance of molecular diagnostics for infectious diseases in low-resource settings. *Nature Reviews Microbiology*, *19*(9), pp.547-548.

37. Sander, A.L., Yadouleton, A., Moreira-Soto, A., Tchibozo, C., Hounkanrin, G., Badou, Y., Fischer, C., Krause, N., Akogbeto, P., F. de Oliveira Filho, E. and Dossou, A., 2021. An observational laboratory-based assessment of SARS-CoV-2 molecular diagnostics in Benin, Western Africa. *MSphere*, *6*(1), pp. 10–1128.

38. Health, T.L.G., 2021. Essential diagnostics: mind the gap. *The Lancet. Global Health*, *9*(11), p. e1474.

39. Terriau, A., Albertini, J., Montassier, E., Poirier, A. and Le Bastard, Q., 2021. Estimating the impact of virus testing strategies on the COVID-19 case fatality rate using fixed-effects models. *Scientific Reports*, *11*(1), p. 21650.

40. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J. and Van Der Walt, S.J., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), pp. 261–272.

Supplemental Information

Additional participant and specimen details from figure 4.1

Viral load was quantified from an average of 32 saliva specimens (SD ±6 specimens) each from the twelve participants in the negative-on-enrollment cohort, while on average thirteen saliva specimens (SD ±10 specimens) each were processed from 58 participants positive-on-enrollment (Figure 4.S2). For nasal swabs, an average of 35 specimens (SD ±7 specimens) were quantified from seven participants in the negative-on-enrollment cohort, while viral load was quantified in an average of seventeen nasal-swab specimens (SD ±9 specimens) from 22 participants who were positive on-enrollment (Figure 4.S3).

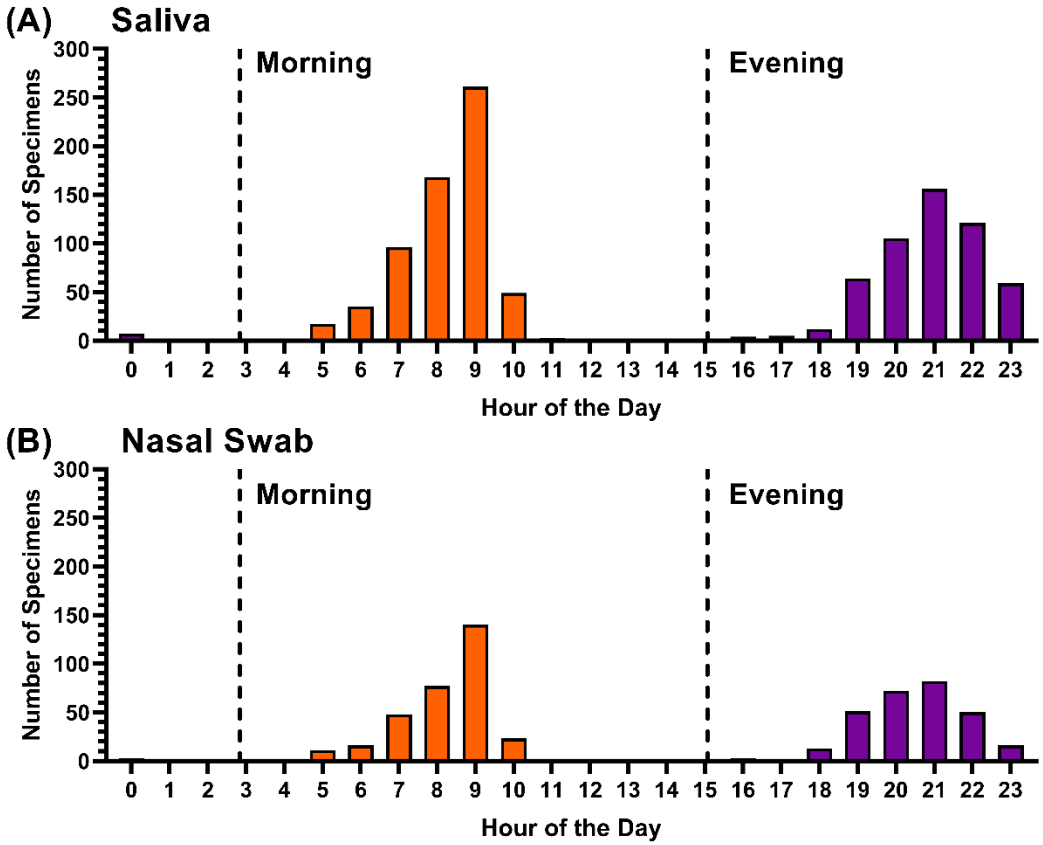
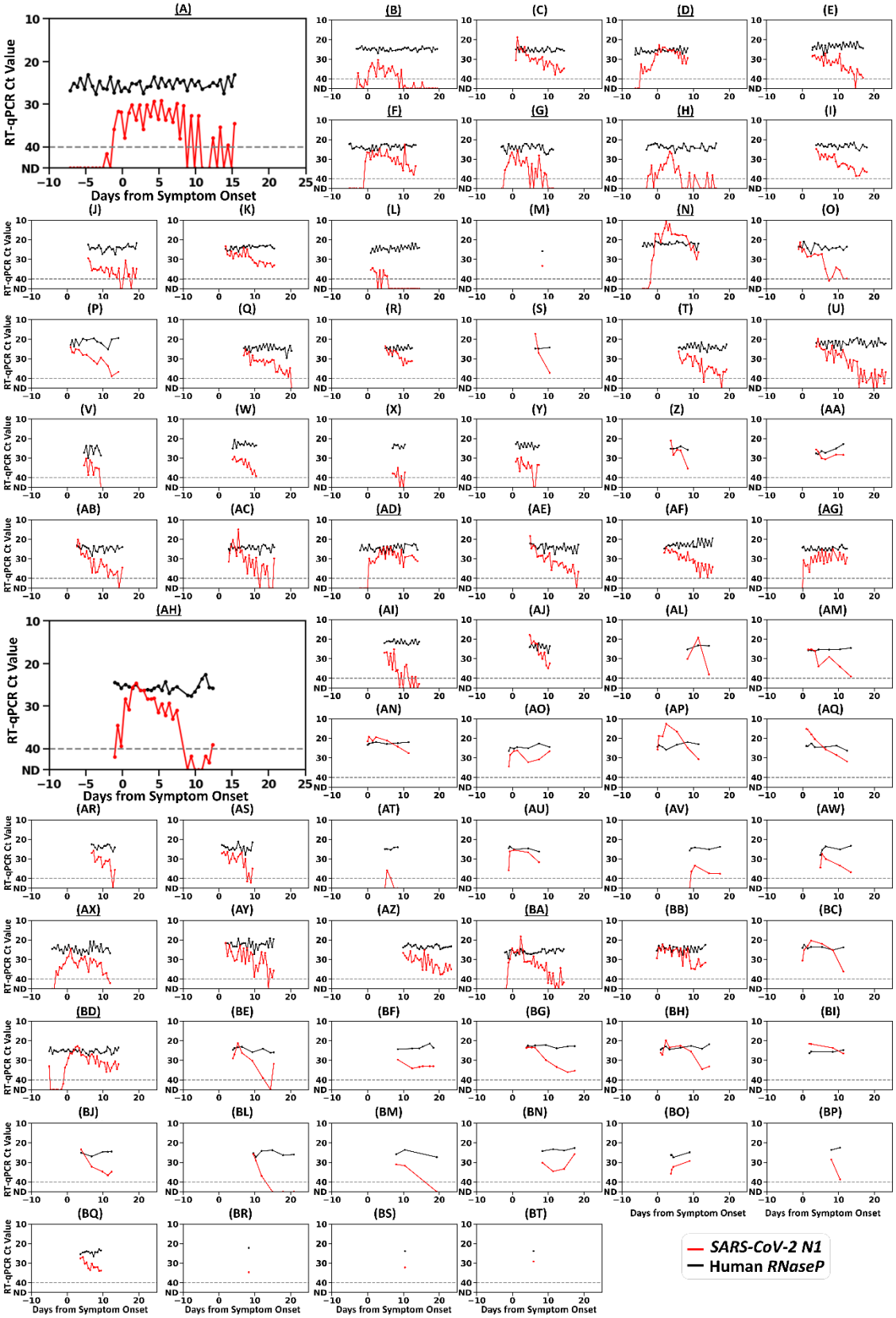


Figure 4.S1: Frequency of Saliva and Nasal-Swab Specimen Collection Times. Study participants either collected saliva only, or saliva then anterior nasal swab at the same time point, and were instructed to collect specimens immediately after waking up and immediately before bed (see Methods for detailed instructions). The frequency of specimens collected by each hour of the day is plotted for 1194 saliva specimens (A) and 661 nasal-swab specimens (B). Dashed vertical line indicates cutoff for morning (3 AM to 12 PM) and evening (3 PM to 3 AM) collected specimens used in this study.

[figure on next page] Figure 4.S2: Individual salivary RT-PCR Ct measurements, for SARS-CoV-2 N1 gene target (red) and human RNase P control gene target (black), relative to symptom onset. Matching panel labels correspond to the same participant shown in Figure S3. Underlined panel labels indicate that the participant converted from SARS-CoV-2- negative to -positive while enrolled in the study. Gray dashed line indicates Ct threshold for positivity. ND indicates not detected.


(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
(K)
(L)
(M)
(N)
(O)
(P)
(Q)
(R)
(S)
(T)
(U)
(V)
(W)
(X)
(Y)
(Z)
(AA)
(AB)
(AC)
(AD)
(AE)
(AF)
(AG)
(AH)
(AI)
(AJ)
(AL)
(AM)
(AN)
(AO)
(AP)
(AQ)
(AR)
(AS)
(AT)
(AU)
(AV)
(AW)
(AX)
(AY)
(AZ)
(BA)
(BB)
(BC)
(BD)
(BE)
(BF)
(BG)
(BH)
(BI)
(BJ)
(BL)
(BM)
(BN)
(BO)
(BP)
(BQ)
(BR)
(BS)
(BT)
RT-qPCR Ct Value
Days from Symptom Onset
ND
SARS-CoV-2 N1
Human RNaseP

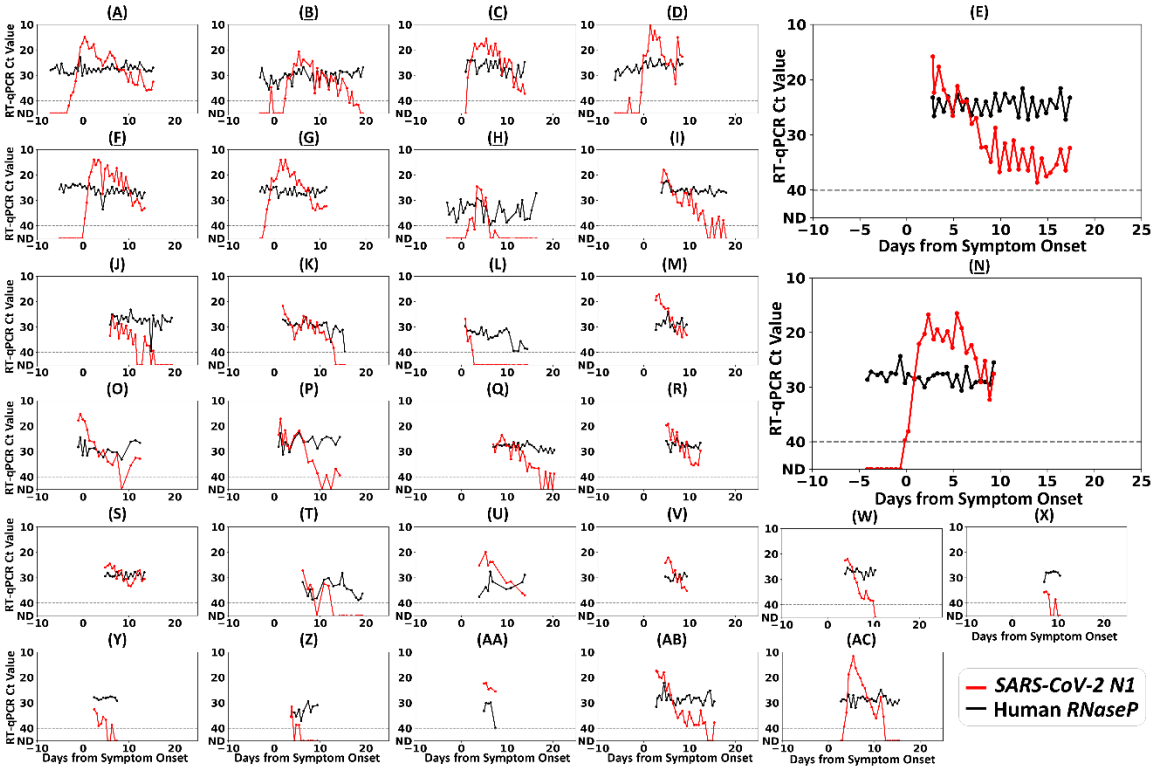


Figure 4.S3: Individual nasal-swab RT-PCR Ct measurements, for SARS-CoV-2 N1 gene target and human RNase P control gene target. Each panel shows the measured SARS-CoV-2 N1 Ct values (red), and human RNase P Ct values (black) for an individual participant, relative to symptom onset. Matching panel labels correspond to the same participant shown in Figure S2. Underlined panel labels indicate that the participant converted from SARS-CoV-2-negative to -positive while enrolled in the study. Gray dashed line indicates Ct threshold for positivity. ND indicates not detected.

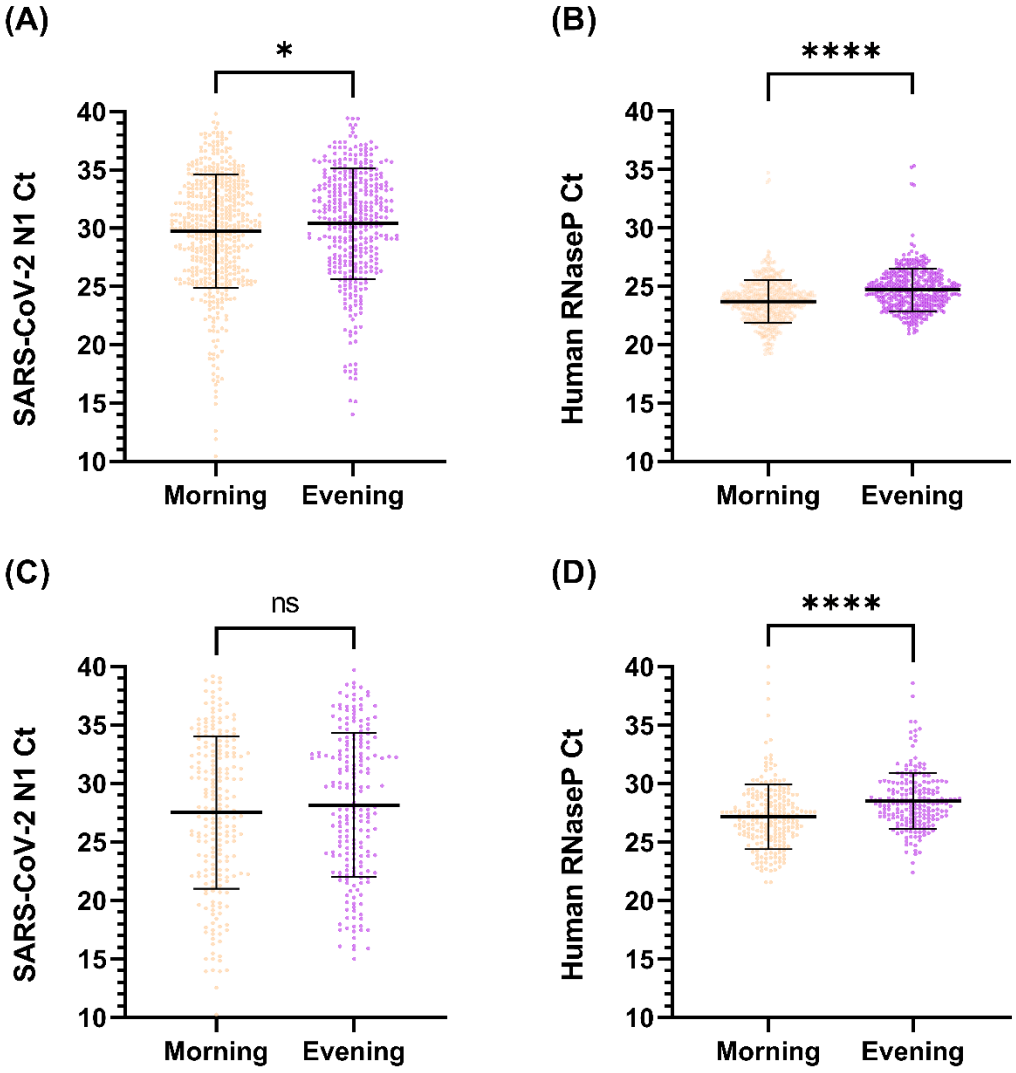


Figure 4.S4: Aggregated SARS-CoV-2 N1 and human RNase P Ct values grouped by specimens collected in the morning and evening. A) Direct comparison of aggregated Ct values for SARS-CoV-2 N1 gene target, measured from all SARS-CoV-2 positive saliva specimens from all participants, by either morning or evening collection time B) Direct comparison of aggregated Ct values for human RNase P target from all SARS-CoV-2 positive saliva specimens from all participants, by either morning or evening collection time C) Direct comparison of aggregated Ct values for SARS-CoV-2 N1 gene, measured from all SARS-CoV-2 positive nasal-swab specimens from all participants, by either morning or evening collection time D) Direct comparison of aggregated Ct values for human RNase P target from all SARS-CoV-2 positive nasal-swab specimens from all participants, by either morning or evening collection time. Specimens with morning collection times are shown as orange points, while evening are shown as purple points. Black lines indicate mean Ct value, with error bars representing standard deviation. Statistical comparison of Ct values for groups performed by unpaired t-test without correction: ns indicates nonsignificant difference, * indicates P <0.001.

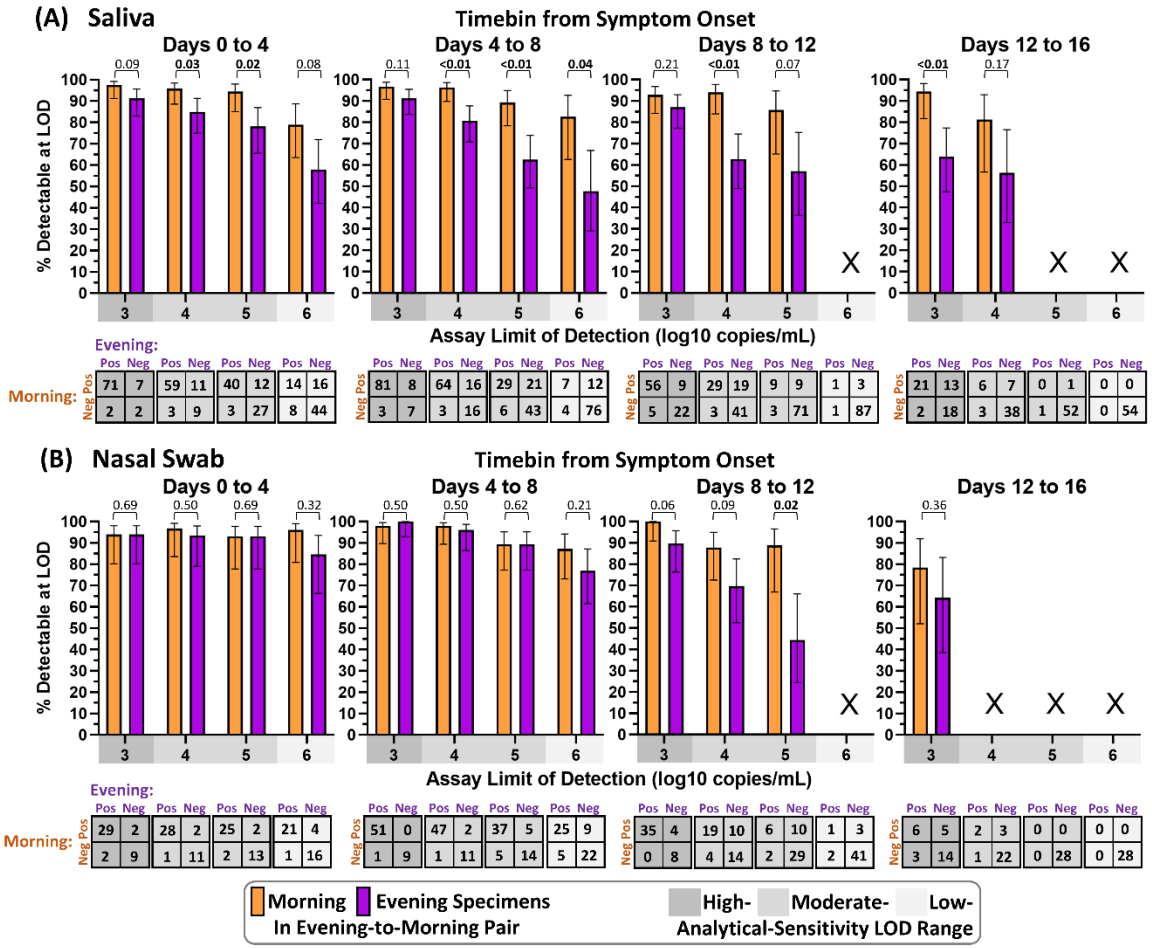


Figure 4.S5: Morning saliva or nasal-swab specimen collection yields improved detection across infection stages and assay analytical sensitivities. For each four-day time bin relative to symptom onset, pairs of sequentially collected evening-to-morning specimens were assessed. In each pair, the viral load in each specimen was used to predict positivity if tested by an assay with a given limit of detection (LOD). Bar plots show the fraction of pairs with a positive result in either the morning or evening specimen that would be detectable if the morning specimen (orange) or evening specimen (purple) were tested at a given LOD. Error bars indicate the 95% confidence interval. Bars are not shown (X) when fewer than ten pairs had positive results at the given LOD during the infection time bin. Among LODs and infection time bins with more than ten positive pairs, the percent detectable for morning versus evening specimens were compared by upper-tailed McNemar Exact Test, applied to the 2x2 table shown below each comparison. Resulting P-values are shown above each comparison. Bolded values indicate significantly higher detection with morning sampling over evening sampling. Analysis was performed on (A) saliva specimens and (B) nasal swab specimens. Equivalent analysis for morning-to-evening pairs is shown in Figure 3.

Chapter 5

THE RATIO BETWEEN SARS-COV-2 RNA VIRAL LOAD AND CULTURABLE VIRAL TITER DIFFERS DEPENDING ON STAGE OF INFECTION

1. Porter, M.K., Viloria Winnett, A., Hao, L., Shelby, N., Reyes, J.A., Schlenker, N.W., Romano, A.E., Tognazzini, C., Feaster, M., Goh, Y.Y. and Gale, Jr, M., 2023. The ratio between SARS-CoV-2 RNA viral load and culturable viral titer differs depending on stage of infection. *Access Microbiology*, pp. 000732–v1.

Author Contribution Statements

Michael K. Porter (MKP) – Conceptualization of nested study with AVW and RFI. Literature review with AVW. Co-wrote biosafety protocols for receipt and processing of live specimens with AVW. Reagent and supply acquisition. Determined and validated live specimen collection methods with AVW. Received, logged and aliquoted live specimens with AVW. Established academic collaboration with colleagues at University of Washington. Performed nucleic acid extractions and *RNaseP, N, E, and RdRP* gene RT-qPCR. Coordinated and prepared shipments of live specimens to collaborators. Interpreted paired viral load and culture data with AVW and LH. Outlined and co-wrote manuscript with AVW. Prepared Figures 1, S1, S2, and S3. Validated underlying data with AVW. Managed author citations and references.

Alexander Viloria Winnett (AVW) – Conceptualization of nested study with MKP and RFI. Literature review with MKP. Co-wrote biosafety protocols for receipt and processing of live specimens with MKP. Reagent and supply acquisition. Determined and validated live specimen collection methods with MKP. Received, logged and aliquoted live specimens with MKP. Performed nucleic acid extractions and *RNaseP* and *N* gene RT-qPCR. Coordinated and prepared shipments of nucleic acids for viral sequencing. Coordinated and prepared shipments of live specimens to collaborators for viral culture. Interpreted paired viral load and culture data with MKP and LH. Outlined and co-wrote manuscript with MKP. Provided feedback on the design of Figures 1, S1, and S2. Validated underlying data with MKP.

Linhui Hao (LH) – Received live specimens for viral culture. Performed viral culture at BSL3 and analyzed viral culture data. Interpreted paired viral load and culture data with AVW and MKP. Wrote viral culture methods section. Edited the manuscript.

Natasha Shelby (NS) – Study Administrator; collaborated on study design and recruitment strategies; created instructions for live viral-load sampling with JAR; enrolled and maintained study participants with JAR and NS. Edited the manuscript.

Jessica A. Reyes (JAR) – Lead Study Coordinator; created instructions for live viral-load sampling with NS; enrolled and maintained study participants with NS and NWS.

Noah W. Schlenker (NWS) – Study Coordinator; enrolled and maintained study participants with NS and JAR.

Anna E. Romano (AER) – Analyzed and summarized viral sequencing data.

Colton Tognazzini (CT) – Coordinated the recruitment efforts at PPHD with case investigators and contact tracers and provided guidance and expertise on SARS-CoV-2 epidemiology and local trends.

Matthew Feaster (MF) – Co-investigator; collaborated on study design and recruitment strategies; provided guidance and expertise on SARS-CoV-2 epidemiology and local trends.

Ying-Ying Goh (YYG) – Co-investigator; collaborated on study design and recruitment strategies; provided guidance and expertise on SARS-CoV-2 epidemiology and local trends.

Michael Gale (MG) – Co-investigator; coordinated and oversaw viral culture method development and execution. Edited the manuscript.

Rustem F. Ismagilov (RFI) – Co-investigator; collaborated on study design and recruitment strategies; provided leadership, technical guidance, oversight, and was responsible for obtaining funding for the study.

Abstract

Analysis of incident, longitudinal RNA viral loads in saliva and nasal swabs and culturable viral titers in nasal swabs collected twice-daily by a tricenarian male infected with SARS-CoV-2 revealed the ratio between viral load and viral titer can be five orders of magnitude higher during early infection than late infection.

Introduction

Throughout the COVID-19 pandemic, the relationship between the detection of viral RNA and replication-competent virus has been used as guiding evidence for infection-control strategies. For example, studies suggesting that low viral load specimens are unlikely to have observable replication-competent virus (1) were used to argue that low-analytical-sensitivity antigen tests (which only detect high viral loads (2)) would more specifically identify infectious individuals (3, 4). Additionally, the lack of replication-competent virus in specimens collected more than a week after symptom onset (5–10) was used as evidence to release individuals from isolation despite persistently detectable viral RNA (11).

Assessment of replication-competent virus in clinical specimens is technically challenging (12) and therefore not routinely performed to determine whether an individual is infectious. Rather, the studies which have generated viral-culture data are often applied broadly to guide infection-control strategies (13). However, the design of such studies influences the data, conclusions, and resulting policies.

Many studies that assess presence of replication-competent virus in specimens from individuals with SARS-CoV-2 infection are primarily cross-sectional, include data from only one specimen type, and are biased toward specimens collected late in the course of infection (e.g., after symptom onset) (4, 14–18). However, during the earliest phase of infection, detection of infected individuals can help reduce subsequent transmission (19, 20) and improve clinical outcomes (21). Few studies report viral loads starting from the incidence of acute SARS-CoV-2 infection (13, 22–29), and of these, few report both viral-load and viral-

culture data (25, 27). If studies of replication-competent virus during SARS-CoV-2 infection are insufficiently representative of early infection, resulting infection control policies may not be optimally effective.

As part of the Caltech COVID-19 Study (23, 24, 30), we attempted to fill this gap by capturing both viral load and viral titer measurements longitudinally from the incidence of acute SARS-CoV-2 infection in a subset of participants at risk of becoming infected. Within this subset, one individual was found to have incident infection with the B.1.243 lineage of SARS-CoV-2 while enrolled and collecting twice-daily specimens, from which we measured both anterior-nares (nasal) swab viral load and viral titer. This participant also collected saliva specimens for viral-load measurements. SARS-CoV-2 *N* gene viral loads and human *RNaseP* marker Cq values in saliva and nasal swab specimens from this individual (Participant AC) have previously been reported (30). Here, we provide additional quantifications of SARS-CoV-2 *E* and *RdRp* gene viral loads and viral-titer measurements from this participant's nasal-swab specimens to investigate the relationship of RNA viral load and infectious virus longitudinally from the incidence of naturally acquired infection.

Results

We report the case (Figure 5.1A) of a 30–39-year-old male (Participant AC), who does not smoke/vape and is otherwise healthy (no chronic medical conditions and self-reported health as "very good"). The participant did not report evidence of prior SARS-CoV-2 infection nor receipt of any SARS-CoV-2 vaccine doses. The participant reported taking Vitamin C and fish oil supplements, and no other medications. In late-January 2021, six days prior to enrollment in this study, the participant reported exposure to SARS-CoV-2. Three days prior to enrollment, the participant began experiencing a sore throat, but two days prior to enrollment tested negative on an outpatient, non-rapid nasopharyngeal test. At this time, a household contact of Participant AC (Participant AB, Figure 5.S1) tested positive, prompting eligibility of both Participant AC for enrollment in this study.

Upon enrollment, Participant AC had detectable and rising salivary viral loads, but was negative in anterior-nares nasal-swab specimens collected over the next day. During this time, the participant remained symptomatic with only a sore throat. In the subsequent day the participant developed shortness of breath and low ($<10^5$copies/mL) nasal viral loads without replication-competent virus detected by culture. After this point, the participant's nasal swab specimens achieved high ($>10^7$ copies/mL) viral loads and high ($>10^6$ TCID50/mL) viral titers for approximately 3 days before gradually declining. Throughout this time, headaches, cough, congestion, change in taste/smell, muscle aches, and one event of severe nausea were reported, all of which resolved before completion of enrollment.

Cross-sectional SARS-CoV-2 viral loads from different gene targets in nasal swab specimens correlated closely with each other (Figure 5.1A, Figure 5.S2A) and the relationship between viral loads from different gene targets remained proportional throughout the course of infection (Figure 5.S2B). Cross-sectional analysis of viral load and viral titer revealed that only high viral load nasal swab specimens ($>10^8$ *N* cp/mL) would contain replication competent virus (Figure 5.1B). Additionally, saliva viral load is less distinguishable between samples with and without replication competent virus in nasal swab specimens (Figure 5.1B). However, longitudinal analysis revealed that the ratio of nasal swab viral load and viral titer changed by over five orders of magnitude throughout the course of acute infection (Figure 5.1C). This relationship indicates that RNA viral load alone, without considering infection stage, may not represent whether a specimen or a person is likely to be infectious or not.

Discussion

High-frequency nasal swab and saliva sampling from the incidence of infection, and paired measurements of viral load and viral titer in nasal swab specimens revealed four key findings uniquely enabled by this study design.

First, saliva exhibited higher *N* gene viral loads than in nasal swabs for approximately the first two days of incident infection, after which nasal swab viral loads rose and remained

subsequently higher than saliva viral loads. This supports previous observations that SARS-CoV-2 often presents first in oral specimen types before anterior nares swabs (23, 24), and that testing a single specimen type (e.g., nasal swabs) may yield false negative results during early infection.

Second, replication-competent virus was observed in nasal swabs at many timepoints when saliva viral loads were low. This suggests that the low viral load of one specimen type is not necessarily indicative of the absence of replication competent virus in another specimen type.

Third, nasal-swab viral-load measurements from different gene targets (*N*, *E*, and *RdRP* genes) correlated strongly with each other longitudinally, such that measurement of any one viral RNA target was indicative of other viral RNA targets (31).

Fourth, we note that the ratio between RNA viral load and culturable viral titer in nasal swabs decreased substantially (greater than five orders of magnitude) through the first week of infection. Cross-sectional analyses of data from Participant AC and in other studies (4, 15, 18, 25, 32) have suggested a correlation between viral load and the presence of infectious virus. However, these cross-sectional analyses overlook that the relationship between viral load and infectious virus is dynamic, and that early viral loads are more indicative of viral titer than viral loads later in the infection. Therefore, earlier in the infection, individuals with lower viral loads could actually be more infectious than expected based on cross-sectional data.

Data from a SARS-CoV-2 human challenge study (25) supported these conclusions (Figure 5.S3). In that study, 36 human participants were inoculated intranasally with 10 $TCID_{50}$ virus, and 18 participants had subsequent sustained detectable infection. We reanalyzed longitudinal nasal swab viral load and viral culture data graciously provided by the study authors to compare to what was observed in Participant AC's naturally acquired infection. Indeed, among specimens with replication competent virus, the average ratio between viral

titer and viral load at each timepoint after inoculation decreased by nearly four orders of magnitude in the five days following inoculation.

Taken together, these results caution against conclusions about infectiousness that assume a constant ratio of RNA viral load and culturable viral titer, commonly inferred based on cross-sectional data or from single specimen types (4, 33–35). Assuming a constant ratio of RNA viral load and culturable viral titer may not reflect early infection or all anatomical sites from which transmissible virus can be shed, and therefore may be suboptimal evidence for public health policies that seek to reduce transmission.

We acknowledge three main limitations. First, data are from a single unvaccinated person with acute SARS-CoV-2 B.1.243 infection, prior to the availability of COVID-19 vaccines and the emergence of currently circulating variants. Infection characteristics may exhibit substantial person-to-person variation, and vaccination status and/or viral variant may affect the relationship between viral load and viral titer (36). Second, Participant AC collected saliva specimens in a preservation buffer that precluded the ability to perform viral culture, thereby prohibiting inferences on the relationship between saliva viral load and viral titer, or saliva viral titer and nasal viral titer. Third, the lack of detection of replication-competent virus by viral culture may not reflect a true absence of replication-competent virus in the specimen or shedding of infectious virus by the individual as specimen collection, handling, and storage affect virion viability (37, 38). Moreover, both the methods of attempted viral culture and viral characteristics can affect the analytical sensitivity to detect replication-competent virus (39). Therefore, it is possible that replication-competent virus was present in the first two nasal-swab specimens with detectable viral RNA collected by this participant, but at a concentration below the LOD by viral culture.

The data presented here is rare and challenging to obtain. We hope that similar datasets of viral load and viral titer in paired specimen types collected longitudinally starting from early infection can be made accessible for meta-analysis and guide optimized public health strategies that reduce the burden of SARS-CoV-2 or other pathogens.

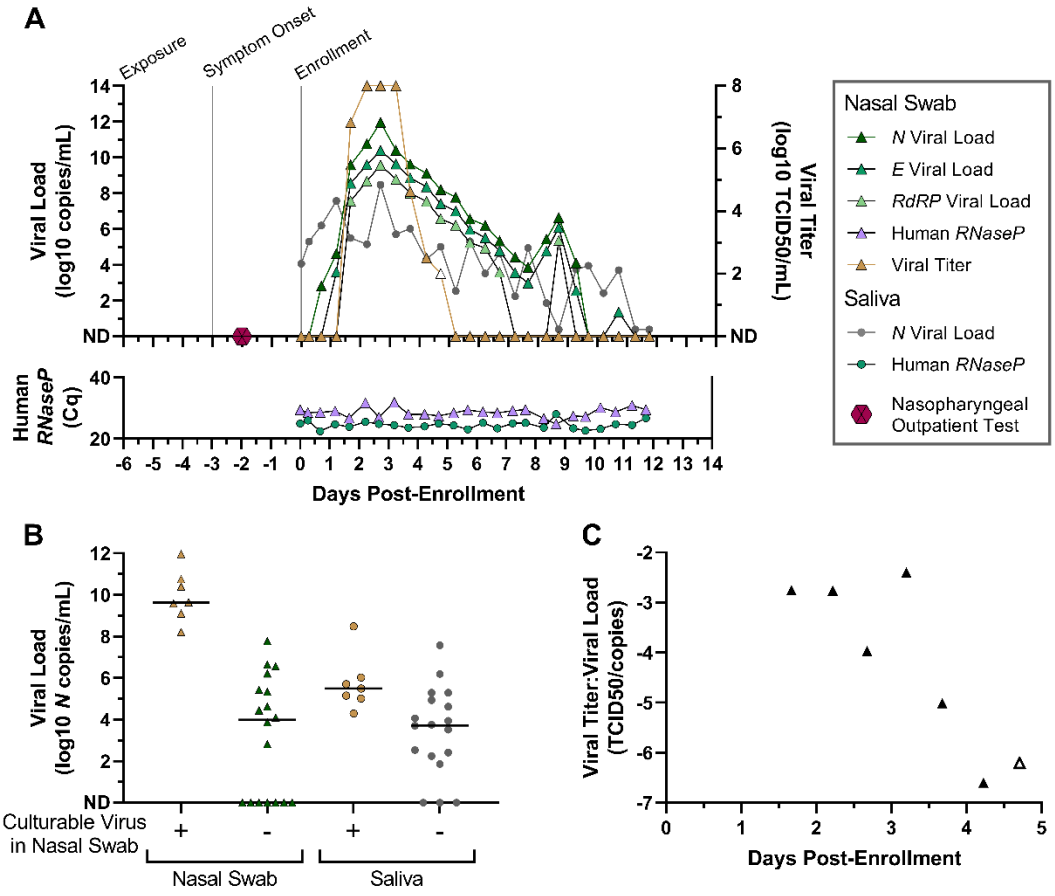


Figure 5.1: The viral load and viral titer trajectories from a single study participant from the incidence of infection. (A) A timeline of Participant AC's infection is shown with notable case events (exposure, symptom onset, study enrollment), as well as SARS-CoV-2 viral loads in saliva (circles) and anterior-nares nasal swabs (triangles) on the left y-axis, and SARS-CoV-2 viral titer (log_{10} $TCID_{50}$/mL) on the right y-axis. Human *RNaseP* Cq values are shown as a measure of sampling consistency and specimen RNA integrity. (B) Cross-sectional relationship of SARS-CoV-2 viral load (log_{10} *N* copies/mL, y axis) in nasal swab specimens (triangles) or saliva specimens (circles) based on whether viral culture positivity (yellow) of the nasal swab from the same timepoint. Black horizontal bars indicate median viral load. **(C)** For specimens with detectable viral titer and viral load, the ratio of viral titer ($TCID_{50}$/mL) over *N* gene viral load (copies/mL) in nasal swab specimens collected by the participant is plotted through days of enrollment. The open symbol indicates a specimen with detectable but not quantifiable viral titer, for which 100 $TCID_{50}$/mL was imputed. ND, not detected.

Methods

Participant consent statement

This COVID-19 household transmission study was approved under California Institute of Technology Institutional Review Board under protocol #20-1026, as previously described (23, 30).

Study design and specimen collection

Enrolled participants began self-collecting saliva and nasal swab specimens immediately upon receipt of specimen collection materials at enrollment, and then each subsequent morning (immediately after waking), and evening (prior to bed). Participants self-collected anterior-nares nasal swabs in Nest VTM (catalog no. NST-NST-202117; Stellar Scientific, Baltimore, MD) and saliva specimens in the Spectrum SDNA-1000 Saliva Collection Kit (Spectrum Solutions LLC, Draper, UT). Study participants were instructed not to eat, smoke, chew gum, or brush their teeth for at least 30 min prior to collection and asked to gently blow their noses before nasal swabbing (four complete rotations with gentle pressure in each nostril) with sterile flocked swabs. Specimens were transported daily by medical courier to the Caltech laboratory for analysis. Additional reagent information is tabulated in Table S1.

Nucleic acid extraction, quantification of viral load by RT-qPCR, and viral variant determination

Nucleic-acid extraction was performed as previously described (23). Conversion from RT-qPCR Cq to viral load (in copies/mL) was determined via calibration curves, reported for *N* gene previously (23), and built for *E* and *RdRP* gene using standard positive controls (IDT 10006896, IDT 10006897):

$$E\left[\frac{cp}{uL}\right] = 2^{\frac{38.241-Cq}{0.9841}}$$

$$RdRP\left[\frac{cp}{uL}\right] = 2^{\frac{39.085-Cq}{0.8981}}$$

Nucleic acids extracted from the 7th saliva and nasal swab specimens collected by the

participant underwent viral sequencing and variant determination as previously described (23).

Measurement of viral titer

Tissue culture infection dose to infect 50% of test cultures ($TCID_{50}$) assay was performed to measure the viral titer in VTM samples. Briefly, 500 µl VTM sample was filter-cleaned with a spin column (CLS-8160, Corning). VeroE6 cells ectopically expressing human ACE2 and TMPRSS2 (VeroE6-AT cells; a gift from Dr. Barney Graham, National Institutes of Health, Bethesda MD) were seeded confluent in a 96-well plate, after replacing the seeding medium with 90 µL of assay medium (Dulbecco's Modified Eagle Media (DMEM) + 2% heat inactivated Fetal Bovine Serum (FBS) + 10 mM HEPES + 1% Penicillin/Streptomycin), 10 µL of filtered VTM sample was added to the first row of the plate as the starting inoculation. Then, 10-fold serial dilutions were performed in the 2nd through 7th rows, leaving the 8th row as the negative control. Each sample was tested with five replicates. Cells were fixed with 10% formaldehyde and stained with 1% crystal violet three days post infection. Digital photographs were taken, and cell death indicated by clear areas in a well, were scored to calculate $TCID_{50}$.

Acknowledgements

We wish to thank the participant who contributed these specimens for analysis, the contact tracers at the Pasadena Public Health Department, as well as Dr. Matthew Bidwell Goetz and Dr. David Beehouwer for their thoughts on this data. We thank Maira Phelps, Lienna Chan, Lucy Li, Dan Lu, and Amy Kistler at the Chan Zuckerberg Biohub for performing SARS-CoV-2 sequencing.

Funding Information

This study is based on research funded in part by the Bill & Melinda Gates Foundation (INV-023124). The findings and conclusions contained within are those of the authors and do not

necessarily reflect positions or policies of the Bill & Melinda Gates Foundation. This work was also funded by the Ronald and Maxine Linde Center for New Initiatives at the California Institute of Technology and the Jacobs Institute for Molecular Engineering for Medicine at the California Institute of Technology. A.V.W. is supported by a UCLA DGSOM Geffen Fellowship. MG and LH were supported by National Institutes of Health grant AI151698 for the United World Antiviral Research Network (UWARN) component of the Centers for Research in Emerging Infectious Disease (CREID).

Data Availability

The data underlying the results presented in the study are available at CaltechDATA at https://data.caltech.edu/records/cgf4q-byr92.

Statements and Declarations

R.F.I. is a cofounder, consultant, and a director and has stock ownership of Talis Biomedical Corp. All other co-authors report no competing interests.

Consent to publish

The two adult participants in this report provided written informed consent.

References

1. Walsh, K.A., Jordan, K., Clyne, B., Rohde, D., Drummond, L., Byrne, P., Ahern, S., Carty, P.G., O'Brien, K.K., O'Murchu, E. and O'Neill, M., 2020. SARS-CoV-2 detection, viral load and infectivity over the course of an infection. *Journal of Infection*, *81*(3), pp. 357–371.

2. Drain, P.K., 2022. Rapid diagnostic testing for SARS-CoV-2. *New England journal of medicine*, *386*(3), pp. 264–272.

3. Mina, M.J., Parker, R. and Larremore, D.B., 2020. Rethinking Covid-19 test sensitivity—a strategy for containment. *New England Journal of Medicine*, *383*(22), p.e120.

4. Pickering, S., Batra, R., Merrick, B., Snell, L.B., Nebbia, G., Douthwaite, S., Reid, F., Patel, A., Ik, M.T.K., Patel, B. and Charalampous, T., 2021. Comparative performance of SARS-CoV-2 lateral flow antigen tests and association with detection of infectious virus in clinical specimens: a single-centre laboratory evaluation study. *The Lancet Microbe*, *2*(9), pp. e461–e471.

5. Perera, R.A., Tso, E., Tsang, O.T., Tsang, D.N., Fung, K., Leung, Y.W., Chin, A.W., Chu, D.K., Cheng, S.M., Poon, L.L. and Chuang, V.W., 2020. SARS-CoV-2 virus culture and subgenomic RNA for respiratory specimens from patients with mild coronavirus disease. *Emerging infectious diseases*, *26*(11), p. 2701.

6. Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C. and Hoelscher, M., 2020. Virological assessment of hospitalized patients with COVID-2019. *Nature*, *581*(7809), pp. 465–469.

7. Siedner, M.J., Boucau, J., Gilbert, R.F., Uddin, R., Luu, J., Haneuse, S., Vyas, T., Reynolds, Z., Iyer, S., Chamberlin, G.C. and Goldstein, R.H., 2022. Duration of viral shedding and culture positivity with postvaccination SARS-CoV-2 delta variant infections. *JCI insight*, *7*(2).

8. Murata, T., Sakurai, A., Suzuki, M., Komoto, S., Ide, T., Ishihara, T. and Doi, Y., 2021. Shedding of viable virus in asymptomatic SARS-CoV-2 carriers. *MSphere*, *6*(3), pp. 10–1128.

9. Jefferson, T., Spencer, E.A., Brassey, J. and Heneghan, C., 2021. Viral cultures for coronavirus disease 2019 infectivity assessment: a systematic review. *Clinical Infectious Diseases*, *73*(11), pp. e3884–e3899.

10. van Kampen, J.J., van de Vijver, D.A., Fraaij, P.L., Haagmans, B.L., Lamers, M.M., Okba, N., van den Akker, J.P., Endeman, H., Gommers, D.A., Cornelissen, J.J. and Hoek, R.A., 2021. Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). *Nature communications*, *12*(1), p. 267.

11. *Centers for Disease Control and Prevention. "Ending isolation and precautions for people with COVID-19: interim guidance." URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html [accessed 2022-05-01]© Nicole Zahradka, Juliana Pugmire, Jessie Lever Taylor, Adam Wolfberg, Matt Wilkes. Originally published in JMIR Formative Research (https://formative.jmir.org) 29 (2022).*

12. Santos Bravo, M., Berengua, C., Marín, P., Esteban, M., Rodriguez, C., Del Cuerpo, M., Miró, E., Cuesta, G., Mosquera, M., Sánchez-Palomino, S. and Vila, J., 2022. Viral culture confirmed SARS-CoV-2 subgenomic RNA value as a good surrogate marker of infectivity. *Journal of clinical microbiology*, *60*(1), pp. e01609–21.

13. Ke, R., Martinez, P.P., Smith, R.L., Gibson, L.L., Achenbach, C.J., McFall, S., Qi, C., Jacob, J., Dembele, E., Bundy, C. and Simons, L.M., 2022, July. Longitudinal analysis of SARS-CoV-2 vaccine breakthrough infections reveals limited infectious virus shedding and restricted tissue distribution. In *Open Forum Infectious Diseases* (Vol. 9, No. 7, p. ofac192). Oxford University Press.

14. Pekosz, A., Parvu, V., Li, M., Andrews, J.C., Manabe, Y.C., Kodsi, S., Gary, D.S., Roger-Dalbert, C., Leitch, J. and Cooper, C.K., 2021. Antigen-based testing but not

real-time polymerase chain reaction correlates with severe acute respiratory syndrome coronavirus 2 viral culture. *Clinical Infectious Diseases*, *73*(9), pp. e2861–e2866.

15. Kirby, J.E., Riedel, S., Dutta, S., Arnaout, R., Cheng, A., Ditelberg, S., Hamel, D.J., Chang, C.A. and Kanki, P.J., 2023. SARS-CoV-2 antigen tests predict infectivity based on viral culture: comparison of antigen, PCR viral load, and viral culture testing on a large sample cohort. *Clinical Microbiology and Infection*, *29*(1), pp. 94–100.

16. Korenkov, M., Poopalasingam, N., Madler, M., Vanshylla, K., Eggeling, R., Wirtz, M., Fish, I., Dewald, F., Gieselmann, L., Lehmann, C. and Fätkenheuer, G., 2021. Evaluation of a rapid antigen test to detect SARS-CoV-2 infection and identify potentially infectious individuals. *Journal of clinical microbiology*, *59*(9), pp. 10–1128.

17. Kohmer, N., Toptan, T., Pallas, C., Karaca, O., Pfeiffer, A., Westhaus, S., Widera, M., Berger, A., Hoehl, S., Kammel, M. and Ciesek, S., 2021. The comparative clinical performance of four SARS-CoV-2 rapid antigen tests and their correlation to infectivity in vitro. *Journal of clinical medicine*, *10*(2), p. 328.

18. Bullard, J., Dust, K., Funk, D., Strong, J.E., Alexander, D., Garnett, L., Boodman, C., Bello, A., Hedley, A., Schiffman, Z. and Doan, K., 2020. Predicting infectious severe acute respiratory syndrome coronavirus 2 from diagnostic samples. *Clinical infectious diseases*, *71*(10), pp. 2663–2666.

19. Ji, J., Viloria Winnett, A., Shelby, N., Reyes, J.A., Schlenker, N.W., Davich, H., Caldera, S., Tognazzini, C., Goh, Y.Y., Feaster, M. and Ismagilov, R.F., 2023. Index cases first identified by nasal-swab rapid COVID-19 tests had more transmission to household contacts than cases identified by other test types. *Plos one*, *18*(10), p. e0292389.

20. McGarry, B.E., Gandhi, A.D. and Barnett, M.L., 2023. Covid-19 surveillance testing and resident outcomes in nursing homes. *New England Journal of Medicine*, *388*(12), pp. 1101–1110.

21. COVID-19 Treatments and Medications. (2023) CDC.

22. Kissler, S.M., Fauver, J.R., Mack, C., Olesen, S.W., Tai, C., Shiue, K.Y., Kalinich, C.C., Jednak, S., Ott, I.M., Vogels, C.B. and Wohlgemuth, J., 2021. Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. *PLoS biology*, *19*(7), p. e3001333.

23. Savela, E.S., Viloria Winnett, A., Romano, A.E., Porter, M.K., Shelby, N., Akana, R., Ji, J., Cooper, M.M., Schlenker, N.W., Reyes, J.A. and Carter, A.M., 2022. Quantitative SARS-CoV-2 viral-load curves in paired saliva samples and nasal swabs inform appropriate respiratory sampling site and analytical test sensitivity required for earliest viral detection. *Journal of clinical microbiology*, *60*(2), pp. e01785–21.

24. Viloria Winnett, A., Akana, R., Shelby, N., Davich, H., Caldera, S., Yamada, T., Reyna, J.R.B., Romano, A.E., Carter, A.M., Kim, M.K. and Thomson, M., 2023. Extreme differences in SARS-CoV-2 viral loads among respiratory specimen types during presumed pre-infectious and infectious periods. *PNAS nexus*, *2*(3), p.pgad033.

25. Killingley, B., Mann, A.J., Kalinova, M., Boyers, A., Goonawardane, N., Zhou, J., Lindsell, K., Hare, S.S., Brown, J., Frise, R. and Smith, E., 2022. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nature Medicine*, *28*(5), pp. 1031–1041.

26. Kissler, S.M., Fauver, J.R., Mack, C., Tai, C.G., Breban, M.I., Watkins, A.E., Samant, R.M., Anderson, D.J., Metti, J., Khullar, G. and Baits, R., 2021. Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated persons. *New England Journal of Medicine*, *385*(26), pp. 2489–2491.

27. Ke, R., Martinez, P.P., Smith, R.L., Gibson, L.L., Mirza, A., Conte, M., Gallagher, N., Luo, C.H., Jarrett, J., Zhou, R. and Conte, A., 2022. Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness. *Nature microbiology*, *7*(5), pp. 640–652.

28. Karita, H.C.S., Dong, T.Q., Johnston, C., Neuzil, K.M., Paasche-Orlow, M.K.,

Kissinger, P.J., Bershteyn, A., Thorpe, L.E., Deming, M., Kottkamp, A. and Laufer, M., 2022. Trajectory of viral RNA load among persons with incident SARS-CoV-2 G614 infection (Wuhan strain) in association with COVID-19 symptom onset and severity. *JAMA Network Open*, *5*(1), pp. e2142796–e2142796.

29. Hakki, S., Zhou, J., Jonnerby, J., Singanayagam, A., Barnett, J.L., Madon, K.J., Koycheva, A., Kelly, C., Houston, H., Nevin, S. and Fenn, J., 2022. Onset and window of SARS-CoV-2 infectiousness and temporal correlation with symptom onset: a prospective, longitudinal, community cohort study. *The Lancet Respiratory Medicine*, *10*(11), pp. 1061–1073.

30. Viloria Winnett, A., Porter, M.K., Romano, A.E., Savela, E.S., Akana, R., Shelby, N., Reyes, J.A., Schlenker, N.W., Cooper, M.M., Carter, A.M. and Ji, J., 2022. Morning SARS-CoV-2 testing yields better detection of infection due to higher viral loads in saliva and nasal swabs upon waking. *Microbiology Spectrum*, *10*(6), pp. e03873–22.

31. Dimcheff, D.E., Valesano, A.L., Rumfelt, K.E., Fitzsimmons, W.J., Blair, C., Mirabelli, C., Petrie, J.G., Martin, E.T., Bhambhani, C., Tewari, M. and Lauring, A.S., 2021. Severe acute respiratory syndrome coronavirus 2 total and subgenomic RNA viral load in hospitalized patients. *The Journal of infectious diseases*, *224*(8), pp. 1287–1293.

32. Stanley, S., Hamel, D.J., Wolf, I.D., Riedel, S., Dutta, S., Contreras, E., Callahan, C.J., Cheng, A., Arnaout, R., Kirby, J.E. and Kanki, P.J., 2022. Limit of detection for rapid antigen testing of the SARS-CoV-2 Omicron and Delta variants of concern using live-virus culture. *Journal of Clinical Microbiology*, *60*(5), pp. e00140–22.

33. Babady, N.E., Cohen, B., McClure, T., Chow, K., Caldararo, M., Jani, K., McMillen, T., Taur, Y., Shah, M., Robilotti, E. and Aslam, A., 2022. Variable duration of viral shedding in cancer patients with coronavirus disease 2019 (COVID-19). *Infection Control & Hospital Epidemiology*, *43*(10), pp. 1413–1415.

34. Puhach, O., Adea, K., Hulo, N., Sattonnet, P., Genecand, C., Iten, A., Jacquérioz, F., Kaiser, L., Vetter, P., Eckerle, I. and Meyer, B., 2022. Infectious viral load in unvaccinated and vaccinated individuals infected with ancestral, Delta or Omicron

SARS-CoV-2. *Nature medicine*, *28*(7), pp. 1491–1500.

35. Smith, R.L., Gibson, L.L., Martinez, P.P., Ke, R., Mirza, A., Conte, M., Gallagher, N., Conte, A., Wang, L., Fredrickson, R. and Edmonson, D.C., 2021. Longitudinal assessment of diagnostic test performance over the course of acute SARS-CoV-2 infection. *The Journal of infectious diseases*, *224*(6), pp. 976–982.

36. Tassetto, M., 2022. Detection of higher cycle threshold values in culturable SARS-CoV-2 Omicron BA. 1 sublineage compared with pre-omicron variant specimens—San Francisco Bay Area, California, July 2021—March 2022. *MMWR. Morbidity and Mortality Weekly Report*, *71*.

37. Chin, A.W., Chu, J.T., Perera, M.R., Hui, K.P., Yen, H.L., Chan, M.C., Peiris, M. and Poon, L.L., 2020. Stability of SARS-CoV-2 in different environmental conditions. *The Lancet Microbe*, *1*(1), p. e10.

38. McAuley, J., Fraser, C., Paraskeva, E., Trajcevska, E., Sait, M., Wang, N., Bert, E., Purcell, D. and Strugnell, R., 2021. Optimal preparation of SARS-CoV-2 viral transport medium for culture. *Virology Journal*, *18*, pp. 1–6.

39. Mautner, L., Hoyos, M., Dangel, A., Berger, C., Ehrhardt, A. and Baiker, A., 2022. Replication kinetics and infectivity of SARS-CoV-2 variants of concern in common cell culture models. *Virology journal*, *19*(1), p. 76.

Supplementary Information

Table 5.S1. Reagent list. Table includes all reagents utilized in this study.

Step	ReagentName	Description	Manufacturer	Catalogue Number
Specimen Collection	Spectrum SDNA1000 Saliva Collection Device	For at-home collection of spit saliva into a guandinium-thiocyanate based preservation buffer	Spectrum Solutions LLC	SDNA1000
Specimen Collection	NEST Scientific 10mL Sterile Screw-Cap Transport Tube with 3mL VTM	For at-home collection of nasal swab specimens into media that maintains live virions	Stellar Scientific	NST-NST-202117
Nucleic Acid Extraction	MagMAX™ Viral/Pathogen Nucleic Acid Isolation Kit	For extraction of nucleic acids from clinical upper respiratory specimens	ThermoFisher Scientitif	A42352
Viral Load Quantification	TaqPath™ COVID-19 Combo Kit	For RT-qPCR measurement of human RNaseP and SARS-CoV-2 N gene	ThermoFisher Scientific	A47814
Viral Load Quantification	Heat-inactivated SARS-CoV-2 2019-nCoV/USA-WA1/2020	Extraction control and standard for RT-qPCR quantification	BEI	NR-52286
Viral Load Quantification	2019-nCoV_E Positive Control	Standard for RT-qPCR quantification	IDT	10006896
Viral Load Quantification	2019-nCoV_RdRp (ORF1ab) Positive Control	Standard for RT-qPCR quantification	IDT	10006897
Viral Load Quantification	E_Sarbeco_F1 Forward Primer, 50 nmol	Forward primer for RT-qPCR measurement of SARS-CoV-2 E gene	IDT	10006888
Viral Load Quantification	E_Sarbeco_R2 Reverse Primer, 50 nmol	Reverse primer for RT-qPCR measurement of SARS-CoV-2 E gene	IDT	10006890
Viral Load Quantification	E_Sarbeco_P1 (FAM) Probe, 25 nmol	Probe for RT-qPCR measurement of SARS-CoV-2 E gene	IDT	10006892
Viral Load Quantification	RdRP_SARSr_F2 Forward Primer, 50 nmol	Forward primer for RT-qPCR measurement of SARS-CoV-2 RdRp gene	IDT	10006860
Viral Load Quantification	RdRP_SARSr_R1 Reverse Primer, 50 nmol	Reverse primer for RT-qPCR measurement of SARS-CoV-2 RdRp gene	IDT	10006881

Viral Load Quantification	RdRP_SARSr_P2 (SUN) Probe, 25 nmol	Probe for RT-qPCR measurement of SARS-CoV-2 RdRp gene	IDT	10007063
Viral Culture	Culture Media	DMEM 2%FBS (heat inactivated) 1% Penicillin-Streptomycin 1% HEPES (1M)	Fisher Scientific Fisher Scientific	MT10013CV SH30071.03 MT30002CI MT25060CI
Viral Culture	Cell Line	VeroE6-AT	A gift from Dr. Barney Graham (NIH)	
Viral Culture	Stain for Readout	1% crystal violet 20% Ethanol	Sigma_Aldrich Fisher	C-6158 4355222

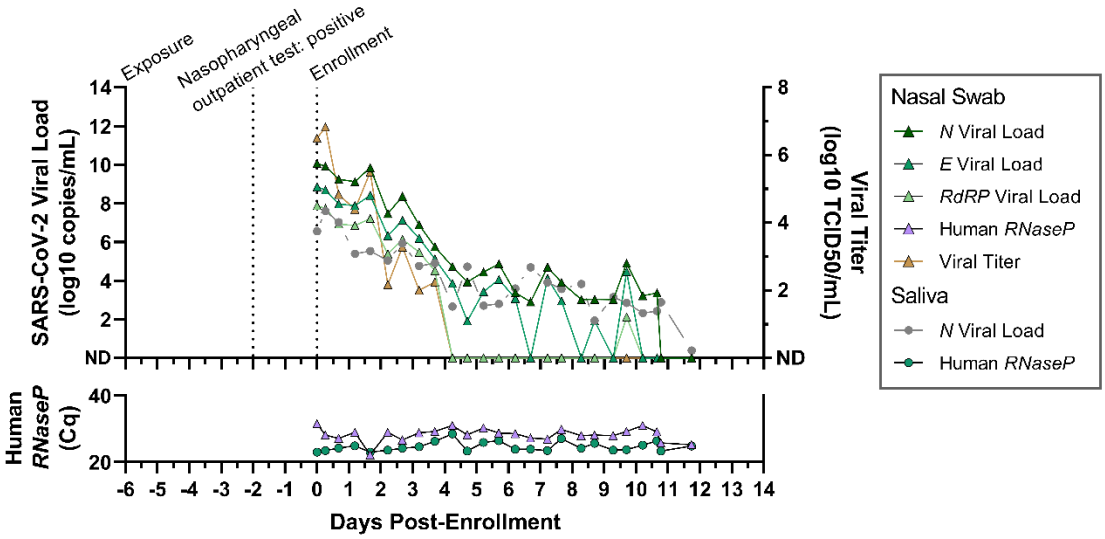


Figure 5.S1: The viral load and viral titer trajectories from a single study participant from the incidence of infection. (A) A timeline of this participant's infection is shown with notable case events (e.g. exposure, positive nasopharyngeal outpatient test, study enrollment), as well as SARS-CoV-2 viral loads (log10 copies/mL) in saliva (circles) and anterior nares nasal swab (triangles) on the left y-axis, and SARS-CoV-2 viral titer (log10 TCID50/mL) on the right y-axis. Human *RNaseP* Cq values are shown as a measure of sampling consistency and specimen RNA integrity. ND, not detected.

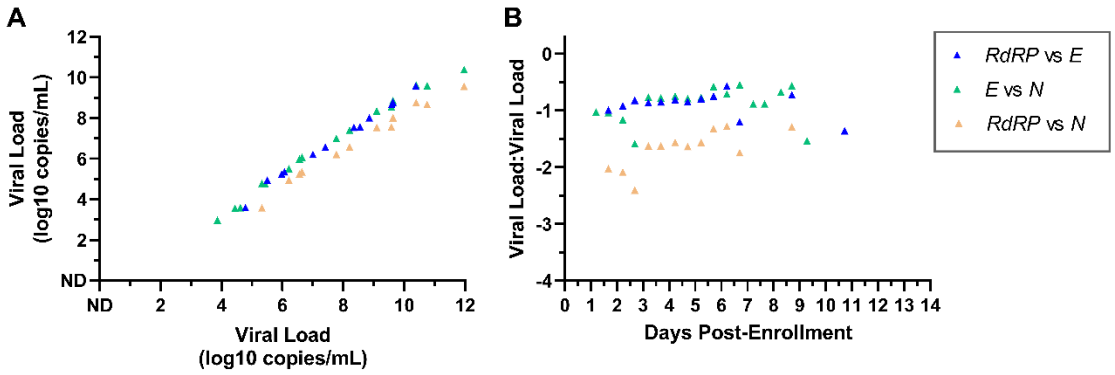


Figure 5.S2: Swab viral loads measured from *N*, *E*, and *RdRP* genes remain constant with respect to each other through the course of infection. (A) The viral load from one gene is plotted on the y axis with respect to another gene comparing *RdRP* and *E* genes (blue triangle), *E* and *N* genes (green triangles), and *RdRP* and *N* genes (tan triangles). (B) The ratios of viral loads are plotted over days post-enrollment for RdRP and *E* genes (blue triangle), *E* and *N* genes (green triangles), and *RdRP* and *N* genes (tan triangles). Viral loads that were not detected were omitted from analysis. ND, not detected.

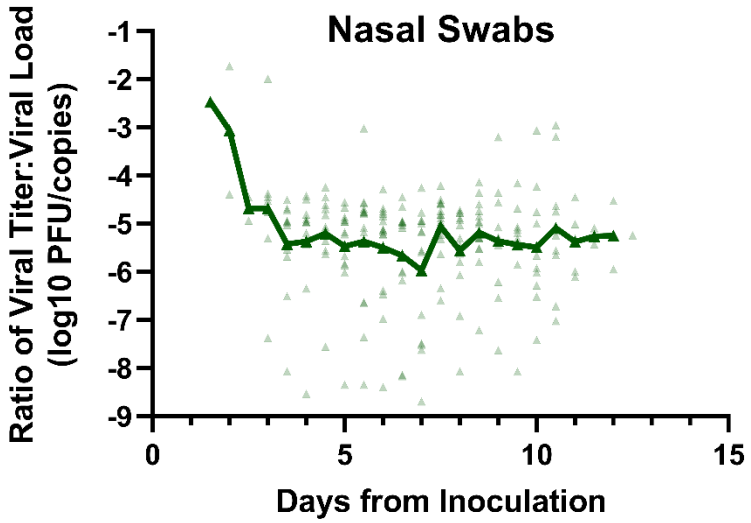


Figure 5.S3: Longitudinal ratio of viral titer to viral load from participants in SARS-CoV-2 human challenge study. As part of a SARS-CoV-2 human challenge study performed in (1), participants were inoculated intranasally with 10 $TCID_{50}$ virus. Eighteen participants had subsequent sustained detectable infection in nasal swab and throat swab specimens collected daily after inoculation. Viral load and viral culture data from these specimens was graciously provided by the authors of this study. We plotted the log10 transformed ratio of viral titer to viral load in nasal swabs, for all specimens with replication competent virus, by the time from inoculation (green triangles). Green line represents the average log10 transformed ratio of viral titer to viral load among culture-positive nasal swab specimen, for each day following inoculation.

Supplemental References

1. Killingley, B., Mann, A.J., Kalinova, M., Boyers, A., Goonawardane, N., Zhou, J., Lindsell, K., Hare, S.S., Brown, J., Frise, R. and Smith, E., 2022. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nature Medicine*, *28*(5), pp.1031-1041.

Chapter 6

GENOME ANALYSIS OF FUNGI DIRECTLY FROM CLINICAL SAMPLES[1]

List of contributors:
Michael K Porter, Reid Akana, Bishoy Kamel, Anna Romano, Igor Grigoriev, Kurt Labutti, Natalie Wu-Woods, Xinyue (Penny) Pei, Omai Garner, David Underhill, Rustem Ismagilov
Acknowledgements: SPEC

1. Sections of this chapter are adapted form a manuscript in preparation.

Authorship Contributions

MKP: Designed overall study. Executed experiments and analyzed all the data for wet lab experiments. Made Figures 6.1, 6.2, 6.3, 6.4, 6.5, 6.S1, 6.S2, 6.S3, 6.S4, 6.S5, 6.S6. Generated data for the following claims:

- Performing MEM on BAL samples significantly reduces host (human) DNA mass by over 1000-fold and preserves fungal DNA within 10x in BAL (Figure S2)
- MEM preserves fungal DNA and community composition in vitro according to qPCR and ITS/18S amplicon sequencing. Preservation of 18S rRNA demonstrates whole, intact fungal cells are preserved through MEM (Figure 2, Figure S4).
- Fungal MAGs were assembled *de novo* with high completion in BAL samples with fungal loads as low as 10 pg per 200 μL of BAL (Figure S1).

RA: Assisted with overall study design. Helped with BAL nucleic acid extractions. Performed bioinformatics to generate data for Figures 6.1c–e., Figure 6.S3, Figure 6.2B, Table 6.S1, Figure 6.S5, Figure 6.S6, Figure 6.S7, Figures 6.3–6.44. Generated data for the following claims:

- Generated sequencing data for Table S1
- Reads mapping to the host were reduced between 10–60% and fungal reads were enriched by about 0.2–50% (Figure 6.1c, d).
- Fungal contigs from MEM processed samples generated contigs that were greater than 40 times longer across all four BAL samples and generated greater than 75-fold

more for larger contigs (>10^3.5bp) than those not processed by MEM (Figure 6.1e, Figures 6.S3a, b, c, d, e, f, g).

- Fungal reads from BAL specimens (MEM processed and without) mapped back to fungal isolate genomes, suggesting that enrichment does not create fungal reads from contamination artifacts, etc.
    - The reads obtained from the +MEM BAL were generally recruited back to the isolate MAG with high depth of coverage, of 16.9x, 162.3x, 97.6x, and 5.3x depth for BAL_011, BAL_148, BAL_160, and BAL_170 (C. glabrata), respectively.
    - In contrast, reads obtained from the -MEM BAL recruited to the isolate MAG with minimal coverage: 0.03x, 3.95x, 0.31x, and 0.06x for BAL_011, BAL_148, BAL_160, and BAL_170 (C. glabrata), respectively.
    - For BAL_170 (C. tropicalis), depth of coverage was low for both MEM processed BAL and control. However, MEM treatment still showed improvement of depth of coverage (0.06x vs. 0.00x for with and without MEM processing, respectively).
- Summary statistics of the MAGs
    - ANI, completion, contamination for each of the isolate genomes and reference-assembled MAGs from BAL (Figure 6.S5, Figure 6.4, Figure 6.S6, Figure 6.S7).
        - De novo MAGs were assembled from BAL_148 and BAL_160 with high completion (97.4% and 93.5%, respectively) and low contamination (0.1% and 1.5%, respectively), as determined by BUSCO (Figure 6.4 a, b)
        - High quality reference-assembled MAGs were assembled from the MEM processed BAL samples with high completion (97.4% and 93.5% for BAL_148 and BAL_160, respectively) and minimal contamination (0.1% and 1.5% for BAL_148 and BAL_160, respectively) (Figure 6.4c, d)

- The reference assembled MAGs have high ANI (100.0% and 100.0% for BAL_148 and BAL_160, respectively) to the isolate genome. Fungal MAGs assembled directly from the BAL samples were taxonomically identical to the fungal isolate genomes and have high concordance (Figure 6.4c, d)
- Comparatively, the reference assembled MAGs have lesser ANI (96.3% and 99.3% for BAL_148 and BAL_160, respectively) to the reference genomes compared to the isolate genomes, suggesting that they are genetically more similar to their cultured isolate than the NCBI reference genome (Figure 6.4c, d)
- With referenced-based assembly, the MAG for *C. lusitaniae* from BAL_011 could be completed with 95.4% and 0.0% contamination, and the MAG for *C. glabrata* from BAL_170 could be completed with 48.8% with 0.5% completion (Figure 6.S6)
- For both sets of MAGs, the ANI between the reference-assembled MAG and the isolate genome was higher than the ANI between the reference-assembled MAG and the NCBI genome (99.9% vs. 98.2% for BAL_011, and 98.3% vs. 97.6%). (Figure 6.S6).

AR: Contributor to the implementation of fungal amplicon sequencing workflow and performed amplicon sequencing. Provided technical guidance for experimental design and data interpretation. Assisted with background literature research. Helped with amplicon and shotgun library preparation. Assisted with overall study design. Provided edits to manuscript and figures.

BK: Provided guidance on data analysis and interpretation. Analyzed sequencing data and assembled genomes and MAGs assembled in Figures 6.4, 6.S5, and 6.S6.

IG: Provided guidance on data analysis and interpretation.

KL: Provided guidance on data interpretation.

NW: Helped with amplicon and shotgun library preparation. Assisted with overall study design.

XPP: Helped with amplicon and shotgun library preparation.

OG: Provided guidance on study design and provided BAL samples and cultured isolates for analysis

DU: Provided guidance on experimental design and provided cell cultures for in vitro validation (Figure 2).

RFI: Investigator, provided feedback on study design, provided leadership, technical guidance, oversight, and was responsible for obtaining funding for the study.

Abstract

Fungi have been heavily implicated in human health and disease, but their exact roles, whether pathogenic or commensal, are not fully understood. One reason for this dearth in understanding is due to the lack of tools the microbiome field has for investigating the interactions between fungi and the host. Current methods for studying fungi (culture, qPCR, and amplicon sequencing) provide mostly taxonomic information; they do not yield the functional information needed to discern mechanistic insights between fungi and host. Shotgun metagenomic sequencing provides both taxonomic and functional information but is challenging to perform directly from clinical specimens (without a culture step) because clinically relevant fungi are often present at low relative abundance (<0.05%) compared with host cells. In this study, we expand a recently developed microbial enrichment method (MEM) to enrich for low-abundance fungi to enable metagenomic sequencing directly from bronchoalveolar lavage (BAL) specimens. We show that MEM enables high-quality characterization of fungal metagenomes in host-rich BAL samples by reducing host DNA mass by more than 1000-fold while preserving fungal load within 10x of the original DNA mass. Shotgun sequencing of MEM-processed clinical specimens enabled characterization of low-abundance fungal taxa (as low as 10 ng fungal DNA per 200 µL BAL). We report the assembly of fungal contigs and metagenome-assembled genomes (MAGs) *de novo* from MEM-processed BAL specimens, with high completion (at least 93.4% as determined by BUSCO) and low contamination (at most 1.5% as determined by BUSCO). MEM is a tool that can enable metagenomic sequencing and deep characterization of fungal communities directly from samples rich in host DNA. We anticipate that MEM processing will have utility in numerous applications relevant to public health (such as tracking new fungal pathogens), clinical treatment (such as identifying antifungal resistance genes), and research into complex diseases with suspected fungal roles (such as inflammatory bowel disease and cancer).

Introduction

Fungi play significant roles in both human disease and maintenance of health (1–5). Fungal pathogens cause some of humanity's most prevalent diseases (such as candidiasis) and deadly infections (such as aspergillosis); however, most human-associated fungi are benign or commensal members of the human microbiome (4, 6–10). Adaptable and opportunistic, fungi can colonize or infect diverse human anatomical locations, including mucosal surfaces (respiratory tract (5), vagina (11), gastrointestinal tract (12–14), skin and nails (15), urine (16), and blood (fungemia) (17, 18). Most of the 1000 to 1 billion fungal spores introduced to the human body daily (19) are rapidly cleared by immune cells; however, some individuals, such as those who are immune compromised, are at high risk for deadly fungal infections (20).

Despite their prevalence and impacts on human health, fungal infections are commonly under-diagnosed (21, 22) and/or incorrectly treated. The gold standard for fungal diagnostics is culture, which is unacceptably slow (up to 4 weeks) (23) and has poor sensitivity (50%) (24, 25). Faster and more sensitive diagnostics for identification of both known and emerging fungal pathogens are needed, particularly for vulnerable populations. Additionally, because a rising number of fungal pathogens are resistant to antifungals—including more than 90% of *Candida auris* isolates (22, 26)—it is also critical to be able to quickly identify any resistance genes to select the most appropriate antifungal treatment. Genome level characterization is necessary to identify where the mutations that confer these antidrug resistance and virulence properties are taking place on the genome and trace these mutations phylogenetically to see how they may have been acquired to help understand their evolution (27).

Similarly, there are many knowledge gaps in our understanding of both the extent and the exact roles of fungi in human microbiomes. Some fungi seem to contribute positively—maintaining gut homeostasis (10), interacting with the host immune response (4, 28, 29), or blocking pathogens from invading (10, 30). Other fungi in the human microbiome seem to

drive disease under certain conditions (31), as mycobiome dysbiosis has been implicated in various diseases (32), such as IBD (31, 33), cancer (34–36), and celiac (37, 38). In fact, the mycobiome literature is often contradictory; a particular fungal species appears protective in some individuals or under certain conditions, but harmful in a different context. For example, *Candida albicans* seems to contribute directly to the pathogenesis in two of the most common vaginal diseases, BV and VVC, yet in some individuals this same species seems to be a benign vaginal commensal, even when present at high loads (39). Such contradictory findings highlight the need for genome-level understanding and functional characterization of fungi, which requires analyses performed directly on clinical samples (not via a culture step).

Both of these gaps in the mycobiome field (the inability to rapidly diagnose and correctly treat fungal pathogens, and our poor understanding of the functional role of fungi in human microbiomes) are due to technological limitations (40, 41). Fungal diagnostic methods are lacking (40,41), especially for rare, emerging, and diverging fungal pathogens. As mentioned above, the gold standard for fungal diagnostics is culture, which can take days to weeks (23), and is limited to the fungi that can grow outside of the human body. Culture alone also cannot inform on morphologically indistinguishable strains or subpopulations. Additionally, genotypic subpopulations, such as those that are more virulent or that contain resistance genes, may compete poorly in culture (42).

PCR is a faster modality and has much higher sensitivity compared with culture-based methods. Primers have been developed for many clinically relevant fungi (43, 44) and for some well-characterized virulence genes (45, 46). However, as a gene-targeted method, PCR still requires *a priori* knowledge of the sample to select appropriate primers, and PCR cannot provide functional information.

Amplicon sequencing, which targets highly conserved genetic segments (such as ITS and the 18S rRNA gene) yields more comprehensive fungal taxonomic information and can sometimes resolve down to species-level (47). However, as another gene-targeted method,

amplicon sequencing cannot provide strain-level taxonomy nor can it inform on functional differences, such as the presence or absence of any resistance, virulence, or metabolism genes without *a priori* knowledge. Moreover, unlike for bacteria, for which databases have been constructed and curated for many genes of interest, fungal genomic data are relatively scarce. Thus, it is difficult to obtain *a priori* information of fungal genes, strains, and SNPs.

Metagenomic shotgun sequencing overcomes all these limitations by analyzing the genomic content of the entire microbial community in a sample. Shotgun sequencing yields millions of DNA sequences (reads) that can be assembled into longer sequences or onto reference genomes, providing strain-level information and the ability to characterize functional capabilities (48). However, performing metagenomic sequencing of fungi directly from host-rich clinical samples has been challenging to impossible due to the high amount of human host DNA relative to fungal DNA (48). For example, the relative abundances of human DNA can be as high as 95% in saliva (49), 99% in BAL (50), 99.99% in mucosal biopsies (48), and 85% in vaginal swabs (51). In this study, the fungal DNA in these clinically relevant samples was low as ~10 pg in 200 µL in a BAL sample, which is 100x lower than the typical input for library preparation for Illumina sequencing. Sequencing more deeply is expensive and still does not overcome the limitations posed by high host content because the overwhelming majority of the reads will still belong to the host (48). This technical limitation needs to be overcome because both pathogenic and mutualistic fungi are often found at low abundance in host-rich mucosal tissues, such as the oral cavity (49), GI tract (12–14) and lungs (50).

Genomic characterization of fungi *directly* from host-rich clinical samples (without a culture step) is critically needed to enable: (1) complete characterization of fungi in a clinical samples, even species that are unculturable; (2) faster and higher resolution taxonomic identification; and (3) the identification of single-nucleotide variants (SNVs) in key functional genes, such as those that may confer virulence, antifungal resistance, biofilm formation, or the ability to colonize. These three capabilities would both guide appropriate

treatment of infections and greatly improve our understanding of the mechanistic role fungi play in human health (52, 52).

Although a few previous studies directly shotgun-sequenced host-rich clinical samples to study fungi, they lacked a pre-sequencing fungal-enrichment step, so the resulting sequencing assemblies yielded only taxonomic information (54-57). A wide variety of host-removal (aka host-depletion) methods have been developed (58-63), but none have been sufficiently effective to enable fungal MAG construction from relevant sample types high in host material (such as vaginal mucosal swabs or respiratory samples).The only studies to date that have assembled high-quality fungal genomes directly from clinical samples (without culture) used samples with low host loads, such as stool, where only ~10% of sequencing reads are attributed to host (64).

To take full advantage of the capabilities of metagenomic sequencing to study fungi, the field critically needs a method to simultaneously deplete the human component and enrich the fungal component in challenging human-rich (and mucus-rich) clinical samples. Our lab recently developed a MEM that was effective in reducing host DNA by over 1000x in tissue biopsies while preserving bacterial DNA and community composition (48). This method enabled the first construction of MAGs of bacteria and archaea directly from host-rich human intestinal biopsies. In this paper, we extend and validate MEM to fungi to perform the first fungal metagenomic analysis and MAG assemblies directly from host-rich clinical sample types. We selected for our demonstrations two clinically relevant and mucus-rich sample types (BAL, vaginal swabs) that contained low (down to 10 pg of fungal DNA per 200 μL BAL) fungal biomass.

Results

Application of MEM to study fungi

We first wished to test if the MEM previously validated on bacteria (48) can also be used to enrich fungi in clinical samples (Figure 6.1a). Briefly, MEM utilizes selective mechanical

lysis and nuclease and proteinase treatment to enrich for microbial DNA. Mechanical lysis with bead beating exploits the differences in mechanical properties and size of host and microbial cells. We predicted that fungal cell walls, which are reinforced by chitin, glucans, and glycoproteins, will be sufficiently more physically robust against mechanical forces compared to mammalian cells (Young's modulus of fungal yeasts are 5 - 1e6 times greater than that of human cells) (65–67).

The MEM protocol uses gentle bead beating using larger spherical beads (1.4 mm) to preserve harder and smaller particulate matter (such as fungal cells). In contrast, standard nucleic acid extraction techniques for fungal cells typically use smaller glass or ceramic beads (0.1–0.5 mm) (68, 69) or garnet beads (70). Next, proteinase K is added to further lysing the mechanically disrupted host cells, and Benzonase is added to degrade the nucleic acids made accessible from the mechanical lysis and proteinase K treatment.

As a clinical demonstration, we selected bronchoalveolar lavage (BAL) samples that were culture-positive for fungal pathogens. BAL is an appropriate sample type for testing the effectiveness of MEM because they are rich in host material (10:1 ratio to fungal cells; up to 1000:5 ratio to fungal genome size; Figure 1b) and known to contain fungal pathogens (50). BAL samples can additionally be rich in mucus, requiring additional pre-processing steps including a longer incubation step with 1% saponin to release nucleic acids (71).

We processed 22 BAL samples with and without the MEM protocol followed by quantification of fungal DNA using real-time PCR (qPCR) targeting the 18S rRNA gene. Four of the 22 samples had at least 0.001 ng fungal DNA / 200 uL BAL with MEM (Figure 6.S1). In these four samples, MEM reduced host DNA by over 1000x according to qPCR (Figure 6.S2), while preserving fungal DNA within 10x of the original fungal DNA (Figure 6.S2a, b).

Next, these four samples then underwent Illumina shotgun sequencing to characterize the fungal metagenome in BAL. First, we wanted to test whether host reads are depleted and

fungal reads are sufficiently enriched from using MEM. Both control and MEM-treated samples were sequenced to quantify fungal enrichment by MEM (Table 6.S1). After processing, host reads were filtered using a complexity-masked GRCh38 human genome (BBMask) and BBDuk (72). In MEM processed samples, reads mapping to the host were reduced between 10–60% and fungal reads were enriched by about 0.2–50% (Figure 6.1c, d).

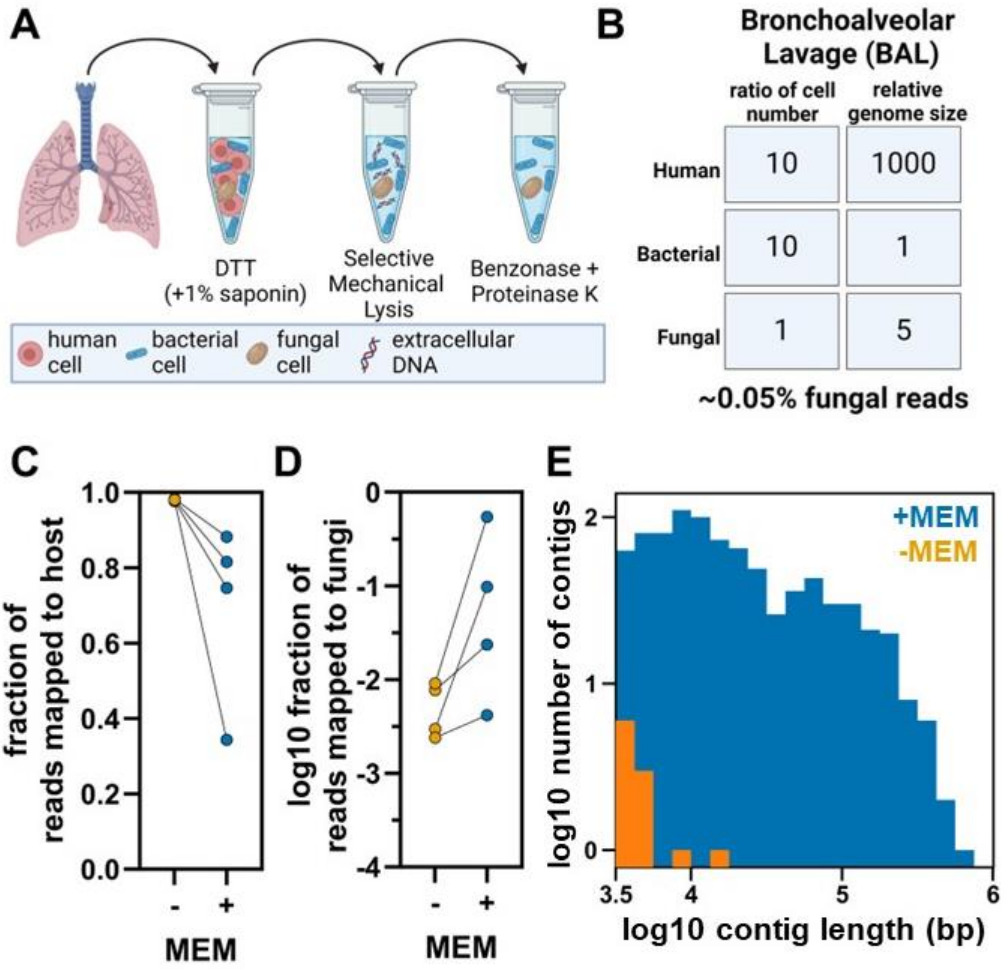


Figure 6.1. Microbial enrichment method (MEM) performed on bronchoalveolar lavage (BAL) fluid improves fungal reads from shotgun sequencing. (A) A schematic of MEM on a BAL clinical sample with fungus. (B) Tabulated values for the ratio of cell number and relative genome size of the three main components of a BAL sample (human, bacteria, and fungi), giving an estimated percentage of 0.05% fungal reads obtained when BAL fluid is sequenced without MEM (50). The fraction of reads mapping to host (C) and fungi (D) in BAL fluid with (blue) and without (orange) MEM. (E) The log-log plot of the number of fungal contigs across four BAL samples as a factor of the length of each fungal contig (bp) for BAL fluid with (blue) and without (orange) MEM treatment.

The number and size of fungal contigs generated from MEM and non-MEM processed samples were then compared. Fungal contigs from MEM processed samples generated contigs that were at least 40 times longer across all four BAL samples and generated at least 75-fold more for larger contigs (>10^3.5bp) than those not processed by MEM (Figure 6.1e, Figure 6.S3a, b, c, d). Fungal contigs from MEM processed samples still had improved number and length when reads were rarefied to 45M (in samples for which contigs still remained) (Figure 6.S3 e, f, g).

MEM preserves fungal DNA load and community composition

To test whether MEM preserves total fungal cells and community composition, we used a 10-taxa community standard purchased from ATCC as a model community (Figure 6.2a, b and Figure 6.S4 a, b). ATCC MSA2010 was selected based on its inclusion of medically relevant fungi and fungi representing different physical properties.

First, the fungal community standard was spun down and resuspended in saline prior to MEM to remove extracellular fungal DNA that may have been present due to lysis prior to sample processing (e.g., from freeze-thaw). Next, we used qPCR (targeting the 18S rRNA gene and ITS) to confirm that total fungal DNA was preserved in the community mix following MEM (Figure 6.2a, Figure 6.S4a). Then, we used 18S rRNA gene (Figure 6.2b) and ITS amplicon sequencing (Figure 6.S4b) to confirm the preservation of community composition in the mix following MEM.

Next, to test whether MEM preserves viable fungal cells, we used five cultured ATCC strains (see Methods). We measured differences with and without MEM on both the amount of the 18S rRNA gene and the amount of 18S rRNA RNA transcript on five different pure cultures of fungi using reverse-transcriptase real-time PCR (RT-qPCR) (Figures 6.2 c–g). The 18S rRNA RNA transcript was used as a proxy to assess the intact-ness of five fungal taxa after MEM processing. RNA degrades more rapidly than DNA, so the presence of rRNA RNA

suggests that fungal cells were intact and rRNA was inaccessible prior to nucleic-acid extraction.

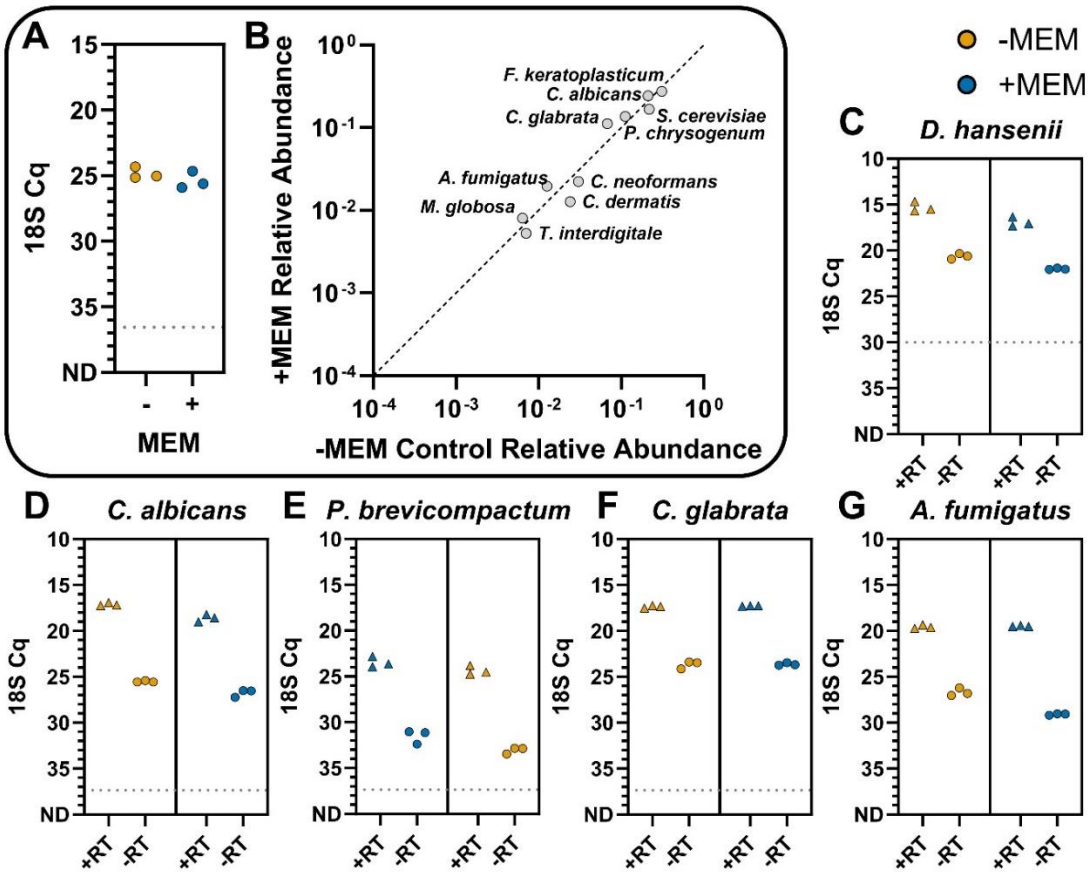


FIGURE 6.2: Microbial enrichment method preserves fungal cells and community composition in contrived samples. (A) The fungal 18S rRNA gene cycles of quantification (Cq) of the 10-taxa community mix with (blue) and without (orange) MEM treatment. (B) The relative abundance of each fungal taxon in the 10-taxa community mix after MEM treatment as a factor of its relative abundance before MEM treatment, as measured by 18S rRNA gene sequencing. Cq as measured by reverse-transcriptase real-time PCR (RT-qPCR) for the fungal 18S rRNA RNA transcript (+RT, triangles) and fungal 18S rRNA gene (-RT, circles) for MEM treated (+MEM, blue) and no-MEM controls (-MEM, orange) for (C) *Debaryomyces hansenii*, (D) *Candida albicans*, (E) *Penicillium brevicompactum*, (F) *Candida glabrata*, and (G) *Aspergillus fumigatus*.

Enriched fungal reads recruit back to isolate genome with high depth of coverage

We next wanted to establish that the fungal reads enriched from the BAL samples matched the genome from the paired cultured isolate and were not artifacts (such as from lab or regent contamination). We first cultured the clinical isolates from the four BAL samples used in Figure 6.1. Three samples had one dominant fungus each (BAL_011 "*Candida lusitaniae*", BAL_148 "*Cryptococcus gattii*", and BAL_160 "*Candida glabrata*"), and one BAL sample had two fungal species identified (BAL_170, "*Candida tropicalis*" and "*Candida glabrata*"). Quotations around the fungal species name indicates the name that the clinic reported in their ID.

Next, we sequenced the genomes from each of the five cultures. Five genomes were assembled with high completion (94.8–97.5%) and minimal contamination (0.0–1.5%) (Figure 6.S5). We recruited the raw fungal reads from the respective BAL sample (with and without MEM processing) back to the isolate genome to confirm that they matched and to compare average depth of coverage (Figure 6.3). Most reads obtained from the +MEM BAL were generally recruited back to the isolate MAG with moderate to high depth of coverage, of 16.91x, 162.28x, 98.15x, and 5.26x depth for BAL_011, BAL_148, BAL_160, and BAL_170 (*C. glabrata*), respectively. In contrast, reads obtained from the -MEM BAL recruited to the isolate MAG had low coverage: 0.03x, 3.95x, 0.31x, and 0.01x for BAL_011, BAL_148, BAL_160, and BAL_170 (*C. glabrata*), respectively. For BAL_170 (*C. tropicalis*), processing without MEM resulted in an average depth of coverage of 0.00, whereas +MEM processing improved of depth of coverage to 0.06x.

MEM enables de novo *assembly of fungal MAGs directly from clinical samples*

We next wished to test whether MEM can enable the *de novo* assembly of high-quality fungal MAGs directly from a BAL sample. First, using the MEM-processed samples, we assembled fungal reads into MAGs of the dominant fungal species directly from two BAL samples *de novo* (from BAL_148 and BAL_160) with high completion (97.4% and 93.5%, respectively) and low contamination (0.1% and 1.5%, respectively), as

determined by BUSCO (Figure 6.4a, b). These MAGs had high synteny to the closest-related NCBI reference genome.

Next, we took assemblies from the two *de novo* MAGs and improved them using reference-based scaffolding from reference genomes on NCBI, resulting in high-quality reference-assembled MAGs with high completion (97.4% and 93.4% for BAL_148 and BAL_160, respectively) and minimal contamination (0.1% and 1.5% for BAL_148 and BAL_160, respectively (Figures 6.4c, d).

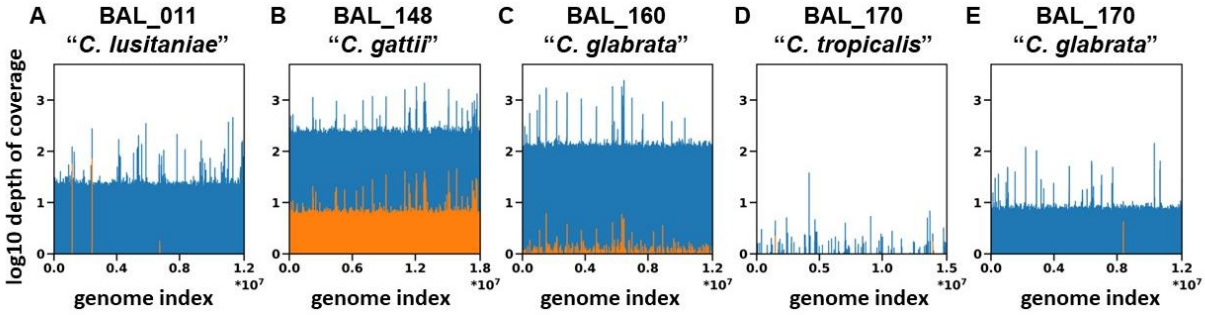


Figure 6.3: Fungal reads from BAL samples with and without MEM processing mapped back to the isolate genome. The depth of coverage from BAL samples with (blue) and without (orange) MEM processing recruited onto a fungal genome assembled from the isolate for (A) BAL_011 "*Candida lusitaniae"*, (B) BAL_148 "*Cryptococcus gattii"*, (C) BAL_160 "*Candida glabrata"*, (D) BAL_170 "*Candida tropicalis",* and (E) BAL_170 "*C. glabrata"*. Quotations are used around the species name to signify the classification of the isolate given by the clinic. Sample BAL_170 contained two fungal isolates. The closest taxonomic classification by NCBI is given in figure S7.

To assess the quality of the reference-assembled fungal MAGs, we compared them to the isolate genome (Figure 6.4c, d). The reference-assembled MAGs had high average nucleotide identity (ANI) (99.98% and 99.97% for BAL_148 and BAL_160, respectively) to the isolate genome. Comparatively, the reference-assembled MAGs had lesser ANI (96.29% and 99.24% for BAL_148 and BAL_160, respectively) to the NCBI reference genomes compared to the isolate genomes, suggesting that they are genetically more similar to the cultured isolate from the same sample than they are to the NCBI reference genome.

For the two BAL samples where *de novo* MAG assembly was not possible (BAL_011 and BAL_170), reference-based assembly using NCBI reference genomes was used to construct the fungal MAGs (Figure 6.S6). With reference-based assembly, the MAG for *C. lusitaniae* from BAL_011 was 95.4% complete and had 0.0% contamination, and the MAG for *C. glabrata* from BAL_170 was 48.8% complete with 0.5% contamination. For both samples (BAL_011 and BAL_170), the reference-assembled MAGs were more similar to the genomes of their paired isolate than the NCBI genome (ANI values of 99.9% vs. 98.2% for BAL_011, and 98.3% vs. 97.6%) (Figure 6.S6).

MAG construction from BAL samples enables genome analysis of infecting fungi

Following assembly, MAGs were annotated to analyze the fungal genes present in the BAL fungi.

Phylogenetic analysis was performed on each of the fungal MAGs that were assembled to high completion (>90%) to identify the ancestral relationship of the fungi we identified in BAL from fungal genomes in the NCBI database (Figure S7). "*Candida lusitaniae*" from BAL_011 is most closely related to *Clavispora lusitaniae*, "*Cryptococcus gattii*" from BAL_148 is most closely related to *Cryptococcus decagattii*, and "*Candida glabrata*" from BAL_160 is most closely related to *Nakaseomyces glabratus* from NCBI (Figure S7).

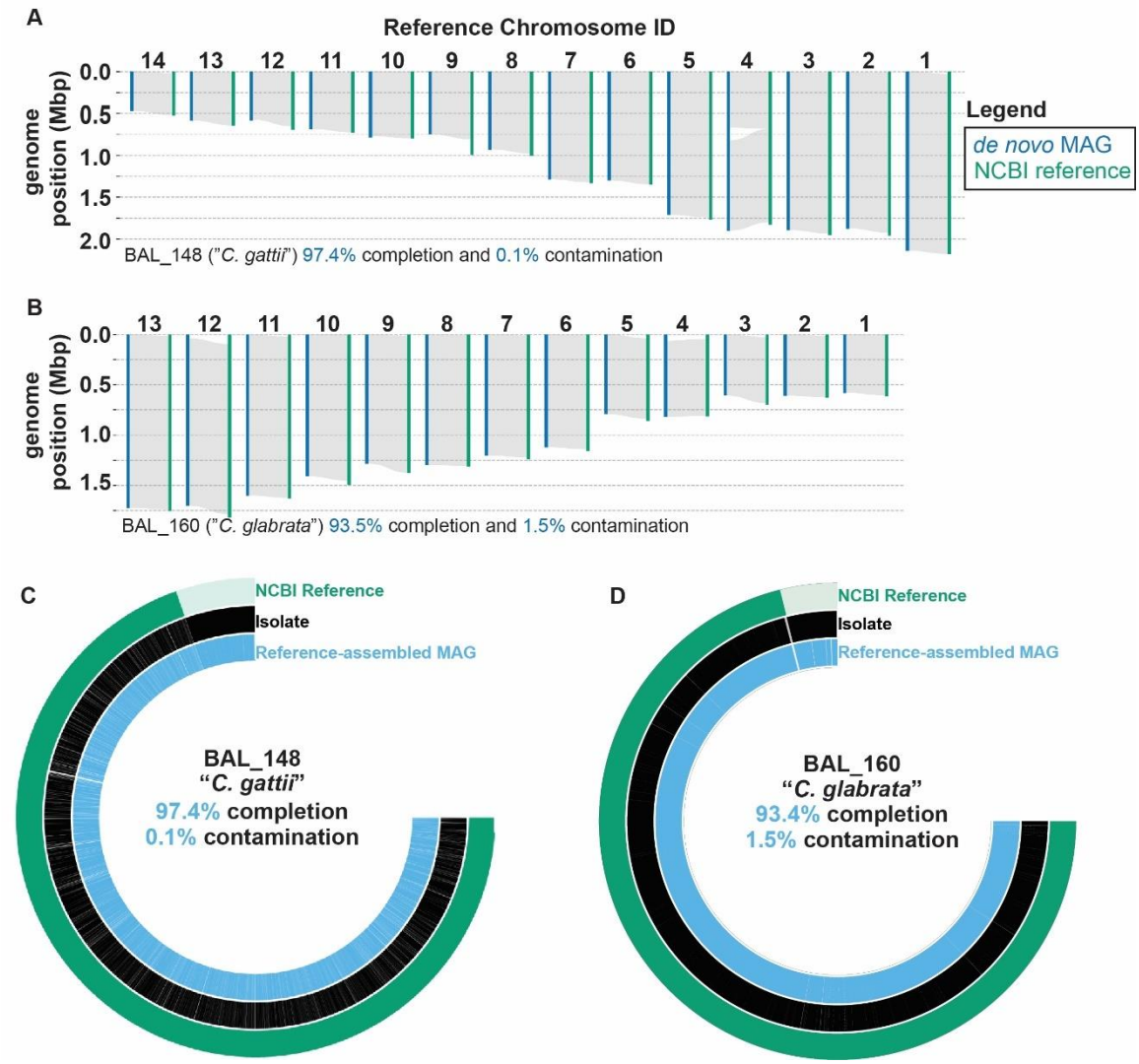


Figure 6.4: MEM enables high quality *de novo* fungal MAG assembly directly from BAL samples. Synteny plots comparing the *de novo*-assembled MAG (blue) and the NCBI reference MAG (green) for (A) BAL_148 "*Cryptococcus gattii*" and (B) "*Candida glabrata.*" The reference-based assembled MAGs (light blue) depicted alongside the assembled genome derived from the isolates (black), and closely related NCBI reference genome (green) for (C) BAL_148 *Cryptococcus gattii* and (D) BAL_160 *Candida glabrata*. Percent completion and contamination as determined by BUSCO are stated for each of the *de novo*-assembled (A, B) and referenced-assembled (C, D) MAGs. Quotations are used around the species name to signify the classification of the isolate given in the clinic. The closest taxonomic classification by NCBI is given in Figure S7.

Discussion and Conclusions

MEM can be used on host-rich samples to enable metagenomic shotgun sequencing of fungi directly from clinical samples. We have demonstrated that MEM depletes host DNA mass by over 1000-fold while preserving fungal DNA within 10x in BAL samples. MEM is compatible with downstream applications of RT-qPCR, amplicon sequencing, and Illumina shotgun sequencing, even in clinical specimens with very low fungal loads (as low as 10 pg per 200 µL of BAL). Analysis of BAL samples that underwent MEM using Illumina shotgun sequencing generated contigs that were at least 40 times longer across all four BAL samples and generated at least 75-fold more larger contigs compared to those not processed by MEM. Additionally, we showed that fungal reads from MEM-processed BAL samples were recruited back to isolate genomes with high average depth of coverage (up to 162.28x). Fungal MAGs were assembled *de novo* with high completion (up to 97.4%) in BAL samples with fungal loads as low as 10 pg per 200 µL of BAL. In cases where it was not possible to assemble MAGs *de novo* (due to low fungal reads and high bacterial load), processing with MEM still enabled the assembly of fungal MAGs using the NCBI reference genome.

Our study has three main limitations. First, this method does not deplete bacterial DNA, which can comprise up to 99% of the relative abundance in certain clinical samples, such as stool. Second, our method has only been validated on yeasts and not filamentous fungi, which may be less robust to the mechanical lysis steps used in MEM. Although most clinically relevant fungi are found as yeasts (73), some applications would benefit from being able to capture other fungal phenotypes (74). MEM would need to be optimized to preserve hyphal fungi during host depletion. Third, not all clinical samples analyzed after MEM processing contained sufficient fungal reads to be assembled into a fungal MAG *de novo*. Among the BAL samples used in this study, only those that contained fungal species with at least 40M reads could be used to assemble de novo fungal MAGs (Table 6.S1).

In this study, fungal MEM was validated on BAL samples, which are an important sample type for diagnosing and studying infections such as aspergillosis (75). Future work in this

area is needed to be able to improve the diagnosis and study of the role of both pathogenic and commensal fungi in relevant clinical contexts. Importantly, as new fungi emerge and problematic pathogens develop antifungal resistance, it is important to develop tools for acquiring mechanistic insights and functional and genomic information. Here we demonstrated fungal MEM with a challenging sample type relevant to fungal lung infections (BAL); MEM should also be demonstrated with additional clinically relevant sample types, such as vaginal swabs. Fungal MEM enables MAG assembly directly from samples without a culture step; future work should demonstrate its utility for acquiring clinically relevant functional and genomic information.

Once fully validated, MEM processing will enable numerous capabilities relevant to public health, clinical treatment, and disease research. For example, with the deep metagenomic profiling of fungal pathogens enabled by MEM, strain/clade typing can be more easily performed to track outbreaks of fungal infections (76). In the clinic, MEM could be used to quickly identify mutations in key virulence or antifungal resistance genes, expediting the course of treatment (77). Additionally, fungal genomes assembled from samples that were MEM processed can be used to capture new fungal genomes and expand reference databases. Finally, when MEM is used to process complex clinical specimens, it may help untangle previously intractable host-microbe interactions, such as in IBD/IBS (78, 79), celiac disease (37), and cancer (35), to improve the field's understanding of fungal evolution (80, 81), and to greatly improve the diagnosis and treatment of emerging and diverging fungal pathogens.

Methods

Human bronchoalveolar lavage (BAL) samples

Human BAL samples were acquired from immunocompromised adults at the University of California, Los Angeles, and sent to Caltech to be analyzed under California Institute of Technology Institutional Review Board (IRB) protocols #19-000308 and #19-0909. Before undergoing MEM, BAL samples underwent a DTT pretreatment (10 mM DTT in autoclaved 0.9% NaCl), vortexed, and incubated for 30 min at room temperature before undergoing

MEM. BAL samples that were visibly mucous and opaque underwent DTT (10 mM DTT in autoclaved 0.9% NaCl) and 1% saponin treatment at 37C for 30 min.

For cultured isolates, 10 mL of BAL samples are vortexed and centrifuged at 1500 xg for 10 minutes. The supernatant is removed and vortexed for 15–20 seconds before 30–60 uL is added to inhibitory mold agar and SabHI agar. The plates are streaked for isolation and incubated at 30C.

Cultured single species fungal samples

Single species fungal cultures were generously provided by Dr. David Underhill from Cedars Sinai. Culturing conditions for each strain were as follows: *Aspergillus fumigatus* isolate AF293 and *Penicillium brevicompactum* ATCC 9056 were grown on Potato Dextrose Agar (ATCC Medium 337) plates at room temperature. *Candida albicans* ATCC 90028, *Candida glabrata* ATCC 2001, and *Debaryomyces hansenii* ATCC 36239 were grown on yeast extract peptone dextrose (YPD) plates (ATCC media 1245) at 30C.

Microbial Enrichment Method (MEM) depletion protocol

Samples that underwent MEM treatment were processed as previous described (48) with minor modifications. Briefly, BAL and VS liquid samples were added into 2-ml 1.3-mm ceramic bead beating tubes (Lysing Matrix D from MP Biomedical, catalog no. 116913050-CF) and supplemented with saline (0.9% NaCl, autoclaved) so that the final volume in the bead beating tube was 400 uL. Samples were homogenized using FastPrep-24 (MP Biomedical catalog no. 116004500) for 30 s at 4.5 m/s. A total of 183 uL of homogenized sample was removed and placed into a clean microcentrifuge tube containing 10 uL of buffer (100 mM Tris + 40 mM MgCl2, pH 8.0 and 0.22 um sterile filtered), 5 uL of Proteinas K (NEB catalog no. P8107S), and 2 uL of Benzonase Nuclease HC (EMD Millipore catalog no. 71205). Sample tubes were placed on a dry block incubator for 15 min at 37C while shaking at 600 rpm. Following incubation, samples were centrifuged at 10,000 g for 1 min and the supernatant was discarded. BAL samples that were visibly mucous and opaque underwent an additional enzyme wash. Pellets were resuspended in either 750 µL Zymo

DNA/RNA Shield (catalog no. R1100-50) for contrived samples or BAL samples and stored on ice until nucleic acid extraction. VS samples were resuspended in 600 µL PM1 solution (Qiagen) and stored on ice until nucleic acid extraction.

Nucleic acid extraction of BAL, cultured clinical isolates, and contrived samples

Nucleic acids from BAL specimens, contrived fungal communities (ATCC MSA2010), and cultured strains were isolated following Zymo's ZymoBIOMICS MagBead DNA/RNA kit (catalog no. R2135) using the KingFisher Flex 96 instrument (Thermo Fisher Scientific). Samples were first homogenized in a 2 mL tube containing 1.4 mm ceramic beads, 0.1 mm silica spheres, and one 4 mm glass bead (Lysing Matrix E from MP Biomedical, catalog no. 116914050-CF) for 1 min at 6 m/s for three cycles using FastPrep-24 (MP Biomedical catalog no. 116004500), incubating the samples on ice for five minutes between each cycle.

Quantification of host DNA

Human host DNA in extracted samples was characterized by quantitative PCR (qPCR) using the Zymo Femto Human DNA Quantification Kit (catalog no. E2005) according to the manufacturer's protocol.

Quantification of fungal rRNA and DNA

Fungal rRNA and DNA in extracted samples were characterized by qPCR. The 18S DNA and rRNA was first reverse transcribed using a High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher catalog no. 4368814). Following reverse transcription, 18S cDNA was amplified based on the following primers developed previously (82). Thermocycling was performed on a CFX96 Real-time PCR system (Bio-Rad Laboratories) using the following protocol: 94˚C for 5 min, 40 cycles of 94˚C for 30 s, 58.3˚C for 50 s, and 72˚C for 1 min. The concentrations of the components in the qPCR mix used in this study are as follows: 1x AccuStart II PCR SuperMix (Quantabio catalog no. 95137-04K), 1x EvaGreen Dye (biotium catalog no. #31000), 500 nM forward primer, and 500 µM reverse primer for a total reaction volume of 10 µL.

A calibration curve relating fungal 18S rDNA Cq to ng of fungal DNA was constructed using a dilution series of ATCC MSA1010 10-taxa fungal genomic DNA standard:

$$18S\,\frac{ng}{rxn} = 2^{-0.8748Cq+12.175}$$

The ITS DNA was amplified based on the following primers from literature (83): (Forward: 5'- TCCGTAGGTGAACCTGCGG-3' and Reverse: 5'-GCTGCGTTCTTCATCGATGC-3', Integrated DNA Technologies). Thermocycling was performed on a CFX96 Real-time PCR system (Bio-Rad Laboratories) using the following protocol: 94C for 5 min, 40 cycles of 94C for 30 s, 61C for 50s, and 72C for 1min. The concentrations of the components in the qPCR mix used in this study are as follows: 1x AccuStart II PCR SuperMix (Quantabio catalog no. 95137-04K), 1x EvaGreen Dye (biotium catalog no. #31000), 500nM forward primer, 500 uM reverse primer, and 1 uL of template for a total reaction volume of 10uL.

Fungal 18S and ITS gene amplicon sequencing

Extracted DNA was amplified and sequenced using 18S and ITS primers described above with Illumina barcode adapter regions (Integrated DNA Technologies). The concentrations of the components in the qPCR mix used for both primers are as follows: 1x KAPA HiFi HotStart ReadyMix (Roche catalog no. 09420398001), 1x EvaGreen Dye (biotium catalog no. #31000), 500 nM forward primer, 500 μM reverse primer, and 2.5 μL of template for a total reaction volume of 25 μL. Amplification was monitored on a CFX96 Real-time PCR system (Bio-Rad Laboratories) and samples were removed once fluorescence measurements reached late exponential phase. The following amplification protocol was used for ITS: 94˚C for 3 min, up to 40 cycles of 95˚C for 30 s, 61.4˚C for 30 s, and 72˚C for 2 min. For 18S: 94˚C for 5 min, up to 40 cycles of 94˚C for 30 s, 62.5˚C for 50 s, and 72˚C for 1 min. Illumina barcode ligation was performed after initial amplification using the following protocol: 95˚C for 3 min, 8 cycles of 95˚C for 3 min, 55˚C for 30 s, 72˚C for 30 s, followed by 72˚C for 5 min. The concentrations of the components in the reaction mix used are as follows: 1x KAPA

HiFi HotStart ReadyMix (Roche catalog no. 09420398001), 200 nM of IDT for Illumina DNA/RNA UD Indexes (Illumina, catalog no. 20026121, 20026930, and 20043019), and 5 μL of template, for a final reaction volume of 50 μL. Duplicate reactions that amplified were pooled together and quantified with Kapa library quantification kit (Kapa Biosystems catalog no. KK4824) before equimolar sample mixing. Libraries were concentrated and cleaned using AMPureXP beads (Beckman Coulter) following each amplification step. The final library was quantified using a High Sensitivity D1000 Tapestation Chip (Agilent catalog nos. 5076–5585, 5067–5584) and Qubit dsDNA Quantification Assay Kit (Thermo Fisher catalog no. Q32851). Sequencing was performed using the Illumina MiSeq platform using V2 chemistry kit (Illumina, PN:MS-103-1003) and 2x300 bp paired-end sequencing. 15% PhiX v3 was added to each sequencing run.

Fungal 18S and ITS gene amplicon data processing

Amplicon sequencing data was primarily analyzed using qiime2 (84). Primer sequences and low-quality regions were removed using Cutadapt (85) using linked primers to eliminate read-through of short amplicons. Resulting reads were denoised using DADA2 (86) and taxonomically classified using a Naive Bayes classifier (sklearn; (87) trained on a database of amplicon sequences obtained by performing in silico PCR (ecoPCR; (88)) on NCBI's RefSeq and GenBank fungal databases (89, 90).

Shotgun sequencing

Extracted DNA was prepared for sequencing as previously described (Natalie) using Illumina DNA prep kit (catalog no. 20018704). Estimations of DNA input were made using fungal 18S primers, the Zymo Femto Human DNA Quantification Kit (catalog no. E2005), and Qubit dsDNA Quantification Assay Kit (Thermo Fisher catalog no. Q32851). For samples with DNA concentrations below Illumina's recommended input, additional PCR cycles were added to the amplification step based on DNA input.

Finished libraries were quantified using Qubit dsDNA Quantification Assay Kit (Thermo Fisher catalog no. Q32851) and a High Sensitivity D1000 Tapestation Chip (Agilent catalog nos. 5076–5585, 5067–5584).

MAG assembly

First, adapter sequences and low quality/homopolymer sequences were trimmed from reads using fastp (0.23.4) (90). Host reads were then filtered using a masked version of the GRCh38 (91) human genome with BBMask (v39.06) (92) and BBDuk. Reads were assembled using Spades (v3.15.5) (93). Fungal contigs were extracted and used to filter reads. Contigs were then reassembled with filtered reads using Spades. Completeness and contamination of assemblies were assessed using BUSCO (v5.7.1) (94). Depth of coverage was determined for each base in an assembly by recruiting reads back onto assemblies using bwa (v0.7.17-r1188) (95) and depth was calculated with samtools depth (v1.17) (96). Average depth of coverage was calculated by averaging over the depth of coverage of each base in a genome (97) Synteny was calculated across closely related reference genomes obtained from NCBIs RefSeq (89) database and visualized using anvio (v8 “Marie”) (98) and matplotlib (v3.7.2) (99) and syri (v1.6.3) (100) and plotsr (v1.1.0) (101). Phylogenomic trees were generated using OrthoFinder (v2.5.5) (102) and iqtree (v2.3.3) (103).

MAG annotation

The resulting MAG genomes assemblies were annotated using the standard JGI annotation pipeline (104, 105). RepeatMasker (106) was used in combination with the RepBase library (107) and repeats larger than 150 bp generated by RepeatScout (108) to mask the assemblies. Masked assemblies were staged into the gene prediction pipeline. *Ab initio* gene prediction was used using Fgenesh, Fgenesh+ (109) and GeneWise (110) which use seeds from protein to genome alignments. Best models at each locus were selected using the strength of homology support. Functional annotation of the proteins included, using eggNOG-Mapper (111) for EC number assignment, InterproScan (112), for protein domains, TCDB (113) for transporter prediction, TMHMM (114) for transmembrane domains and SignalP (115, 116) for signal peptide prediction, in addition comparisons to NCBI NR, swissprot, KOG (117),

KEGG (118) were incorporated. The information from InterPro and Swissprot were used to confidently assign gene ontology terms.

References

1. Kong, H.H. and Segre, J.A., 2020. Cultivating fungal research. *Science*, *368*(6489), pp. 365–366.

2. Caetano, C.F., Gaspar, C., Martinez-de-Oliveira, J., Palmeira-de-Oliveira, A. and Rolo, J., 2023. The role of yeasts in human health: A review. *Life*, *13*(4), p. 924.

3. Brown, G.D., Denning, D.W., Gow, N.A., Levitz, S.M., Netea, M.G. and White, T.C., 2012. Hidden killers: human fungal infections. *Science translational medicine*, *4*(165), pp. 165rv13–165rv13.

4. Underhill, D.M. and Iliev, I.D., 2014. The mycobiota: interactions between commensal fungi and the host immune system. *Nature Reviews Immunology*, *14*(6), pp. 405–416.

5. Cui, L., Morris, A. and Ghedin, E., 2013. The human mycobiome in health and disease. *Genome medicine*, *5*, pp. 1–12.

6. Wang, L., Zhang, K., Zeng, Y., Luo, Y., Peng, J., Zhang, J., Kuang, T. and Fan, G., 2023. Gut mycobiome and metabolic diseases: the known, the unknown, and the future. *Pharmacological Research*, p. 106807.

7. Zhao, X., Hu, X., Han, J., Yin, R., Zhang, S. and Liu, H., 2023. Gut mycobiome: A "black box" of gut microbiome–host interactions. *WIREs Mechanisms of Disease*, *15*(5), p. e1611.

8. Zhou, X., Zhang, X. and Yu, J., 2023. Gut mycobiome in metabolic diseases: mechanisms and clinical implication. *Biomedical Journal*, p.100625.

9. Zhang, F., Aschenbrenner, D., Yoo, J.Y. and Zuo, T., 2022. The gut mycobiome in health, disease, and clinical applications in association with the gut bacterial microbiome assembly. *The Lancet Microbe*, *3*(12), pp. e969–e983.

10. Iliev, I.D. and Leonardi, I., 2017. Fungal dysbiosis: immunity and interactions at mucosal barriers. *Nature Reviews Immunology*, *17*(10), pp. 635–646.

11. Bradford, L.L. and Ravel, J., 2017. The vaginal mycobiome: A contemporary perspective on fungi in women's health and diseases. *Virulence*, *8*(3), pp. 342–351.

12. Mukherjee, P.K., Sendid, B., Hoarau, G., Colombel, J.F., Poulain, D. and Ghannoum, M.A., 2015. Mycobiota in gastrointestinal diseases. *Nature reviews Gastroenterology & hepatology*, *12*(2), pp. 77–87.

13. Lombardi, A. and Ouanounou, A., 2020. Fungal infections in dentistry: Clinical presentations, diagnosis, and treatment alternatives. *Oral surgery, oral medicine, oral pathology and oral radiology*, *130*(5), pp. 533–546.

14. Naglik, J.R., Tang, S.X. and Moyes, D.L., 2013. Oral colonization of fungi. *Current Fungal Infection Reports*, *7*, pp. 152–159.

15. Gupta, A.K., MacLeod, M.A., Foley, K.A., Gupta, G. and Friedlander, S.F., 2017. Fungal skin infections. *Pediatrics in Review*, *38*(1), pp. 8–22.

16. Guze, L.B. and Haley, L.D., 1958. Fungus infections of the urinary tract. *The Yale Journal of Biology and Medicine*, *30*(4), p. 292.

17. Nawrot, U., Kowalska-Krochmal, B., Sulik-Tyszka, B., Kozak, M., Świętek, K., Pajączkowska, M., Piątkowska, E., Rosiak, D. and Swoboda-Kopeć, E., 2015. Evaluation of blood culture media for the detection of fungi. *European Journal of Clinical Microbiology & Infectious Diseases*, *34*, pp. 161–167.

18. Arendrup, M.C., Sulim, S., Holm, A., Nielsen, L., Nielsen, S.D., Knudsen, J.D., Drenck, N.E., Christensen, J.J. and Johansen, H.K., 2011. Diagnostic issues, clinical characteristics, and outcomes for patients with fungemia. *Journal of clinical microbiology*, *49*(9), pp. 3300–3308.

19. [figure, humans inhale between 1,000 and 10 billion spores on a daily basis] - one health: Fungal pathogens of humans, animals, and plants - NCBI bookshelf (2017) One Health: Fungal Pathogens of Humans, Animals, and Plants: Report on an American Academy of Microbiology Colloquium held in Washington, DC, on October 18, 2017. Available at:

https://www.ncbi.nlm.nih.gov/books/NBK549988/figure/F6/ (Accessed: 28 May 2024).

20. Low, C.Y. and Rotstein, C., 2011. Emerging fungal infections in immunocompromised patients. *F1000 medicine reports*, *3*.

21. Lass-Flörl, C., 2017. Current challenges in the diagnosis of fungal infections. *Human fungal pathogen identification: methods and protocols*, pp. 3–15.

22. Coordination, G., Alastruey-Izquierdo, A., World Health Organization and World Health Organization, 2022. *WHO fungal priority pathogens list to guide research, development and public health action*. Organización Mundial de la Salud (OMS).

23. Bosshard, P.P., 2011. Incubation of fungal cultures: how long is long enough?. *Mycoses*, *54*(5), pp. e539–e545.

24. Velasquez-Agudelo, V. and Cardona-Arias, J.A., 2017. Meta-analysis of the utility of culture, biopsy, and direct KOH examination for the diagnosis of onychomycosis. *BMC infectious diseases*, *17*, pp. 1–11.

25. Chamilos, G., Luna, M., Lewis, R.E., Bodey, G.P., Chemaly, R., Tarrand, J.J., Safdar, A. and Kontoyiannis, D.P., 2006. Invasive fungal infections in patients with hematologic malignancies in a tertiary care cancer center: an autopsy study over a 15-year period (1989-2003). *haematologica*, *91*(7), pp. 986–989.

26. Fungal diseases (2024) Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/fungal/index.html (Accessed: 28 May 2024).

27. Huang, X., Welsh, R.M., Deming, C., Proctor, D.M., Thomas, P.J., NISC Comparative Sequencing Program, Gussin, G.M., Huang, S.S., Kong, H.H., Bentz, M.L. and Vallabhaneni, S., 2021. Skin metagenomic sequence analysis of early Candida auris outbreaks in US nursing homes. *Msphere*, *6*(4), pp. 10–1128.

28. Doron, I., Leonardi, I., Li, X.V., Fiers, W.D., Semon, A., Bialt-DeCelie, M.,

Migaud, M., Gao, I.H., Lin, W.Y., Kusakabe, T. and Puel, A., 2021. Human gut mycobiota tune immunity via CARD9-dependent induction of anti-fungal IgG antibodies. Cell, 184(4), pp. 1017–1031.

29. Rozaliyani, A., Antariksa, B., Nurwidya, F., Zaini, J., Setianingrum, F., Hasan, F., Nugrahapraja, H., Yusva, H., Wibowo, H., Bowolaksono, A. and Kosmidis, C., 2023. The Fungal and Bacterial Interface in the Respiratory Mycobiome with a Focus on Aspergillus spp. *Life*, *13*(4), p. 1017.

30. Peleg, A.Y., Hogan, D.A. and Mylonakis, E., 2010. Medically important bacterial–fungal interactions. *Nature reviews microbiology*, *8*(5), pp. 340–349.

31. Jain, U., Ver Heul, A.M., Xiong, S., Gregory, M.H., Demers, E.G., Kern, J.T., Lai, C.W., Muegge, B.D., Barisas, D.A., Leal-Ekman, J.S. and Deepak, P., 2021. Debaryomyces is enriched in Crohn's disease intestinal tissue and impairs healing in mice. *Science*, *371*(6534), pp. 1154–1159.

32. Chen, Y., Chen, Z., Guo, R., Chen, N., Lu, H., Huang, S., Wang, J. and Li, L., 2011. Correlation between gastrointestinal fungi and varying degrees of chronic hepatitis B virus infection. *Diagnostic microbiology and infectious disease*, *70*(4), pp. 492–498.

33. Ha, C.W., Martin, A., Sepich-Poore, G.D., Shi, B., Wang, Y., Gouin, K., Humphrey, G., Sanders, K., Ratnayake, Y., Chan, K.S. and Hendrick, G., 2020. Translocation of viable gut microbiota to mesenteric adipose drives formation of creeping fat in humans. *Cell*, *183*(3), pp. 666–683.

34. Hosseini, K., Ahangari, H., Chapeland-Leclerc, F., Ruprich-Robert, G., Tarhriz, V. and Dilmaghani, A., 2022. Role of fungal infections in carcinogenesis and cancer development: a literature review. *Advanced Pharmaceutical Bulletin*, *12*(4), p. 747.

35. Gamal, A., Elshaer, M., Alabdely, M., Kadry, A., McCormick, T.S. and Ghannoum, M., 2022. The mycobiome: cancer pathogenesis, diagnosis, and therapy. *Cancers*, *14*(12), p. 2875.

36. Narunsky-Haziza, L., Sepich-Poore, G.D., Livyatan, I., Asraf, O., Martino, C., Nejman, D., Gavert, N., Stajich, J.E., Amit, G., González, A. and Wandro, S., 2022. Pan-cancer analyses reveal cancer-type-specific fungal ecologies and bacteriome interactions. *Cell*, *185*(20), pp. 3789–3806.

37. El Mouzan, M., Al-Hussaini, A., Fanelli, B., Assiri, A., AlSaleem, B., Al Mofarreh, M., Al Sarkhy, A. and Alasmi, M., 2022. Fungal dysbiosis in children with celiac disease. *Digestive Diseases and Sciences*, *67*, pp. 216–223.

38. Ott, S.J., Kühbacher, T., Musfeldt, M., Rosenstiel, P., Hellmig, S., Rehman, A., Drews, O., Weichert, W., Timmis, K.N. and Schreiber, S., 2008. Fungi and inflammatory bowel diseases: alterations of composition and diversity. *Scandinavian journal of gastroenterology*, *43*(7), pp. 831–841.

39. Pramanick, R., Mayadeo, N., Warke, H., Begum, S., Aich, P. and Aranha, C., 2019. Vaginal microbiota of asymptomatic bacterial vaginosis and vulvovaginal candidiasis: Are they different from normal microbiota?. *Microbial pathogenesis*, *134*, p. 103599.

40. Fang, W., Wu, J., Cheng, M., Zhu, X., Du, M., Chen, C., Liao, W., Zhi, K. and Pan, W., 2023. Diagnosis of invasive fungal infections: challenges and recent developments. *Journal of biomedical science*, *30*(1), p. 42.

41. Zhang, S.X., Babady, N.E., Hanson, K.E., Harrington, A.T., Larkin, P.M., Leal Jr, S.M., Luethy, P.M., Martin, I.W., Pancholi, P., Procop, G.W. and Riedel, S., 2021. Recognition of diagnostic gaps for laboratory diagnosis of fungal diseases: expert opinion from the Fungal Diagnostics Laboratories Consortium (FDLC). *Journal of clinical microbiology*, *59*(7), pp. 10–1128.

42. Desnos-Ollivier, M., Patel, S., Spaulding, A.R., Charlier, C., Garcia-Hermoso, D., Nielsen, K. and Dromer, F., 2010. Mixed infections and in vivo evolution in the human fungal pathogen Cryptococcus neoformans. *MBio*, *1*(1), pp. 10–1128.

43. White, P.L., Alanio, A., Cruciani, M., Gorton, R., Millon, L., Rickerts, V., Barnes,

R.A., Donnelly, J.P. and Loeffler, J., 2020. Nucleic acid tools for invasive fungal disease diagnosis. *Current Fungal Infection Reports*, *14*, pp. 76–88.

44. Zhang, J., Hung, G.C., Nagamine, K., Li, B., Tsai, S. and Lo, S.C., 2016. Development of Candida-specific real-time PCR assays for the detection and identification of eight medically important Candida species. *Microbiology insights*, *9*, pp. MBI–S38517.

45. Monroy-Pérez, E., Rodríguez-Bedolla, R.M., Garzón, J., Vaca-Paniagua, F., Jiménez, E.A.R. and Paniagua-Contreras, G.L., 2020. Marked virulence and azole resistance in Candida albicans isolated from patients with periodontal disease. *Microbial Pathogenesis*, *148*, p. 104436.

46. Stielow, J.B., Levesque, C.A., Seifert, K.A., Meyer, W., Irinyi, L., Smits, D., Renfurm, R.G.J.M.V., Verkley, G.J.M., Groenewald, M., Chaduli, D. and Lomascolo, A., 2015. One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. *Persoonia-Molecular Phylogeny and Evolution of Fungi*, *35*(1), pp. 242–263.

47. De Filippis, F., Laiola, M., Blaiotta, G. and Ercolini, D., 2017. Different amplicon targets for sequencing-based studies of fungal diversity. *Applied and Environmental Microbiology*, *83*(17), pp. e00905–17.

48. Wu-Woods, N.J., Barlow, J.T., Trigodet, F., Shaw, D.G., Romano, A.E., Jabri, B., Eren, A.M. and Ismagilov, R.F., 2023. Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples. *Nature Methods*, *20*(11), pp. 1672–1682.

49. Marotz, C.A., Sanders, J.G., Zuniga, C., Zaramela, L.S., Knight, R. and Zengler, K., 2018. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome*, *6*, pp. 1–9.

50. Leo, S., Gaïa, N., Ruppé, E., Emonet, S., Girard, M., Lazarevic, V. and Schrenzel, J., 2017. Detection of bacterial pathogens from broncho-alveolar lavage by next-

generation sequencing. *International journal of molecular sciences*, *18*(9), p. 2011.

51. Elnaggar, J.H., Lammons, J.W., Taylor, C.M., Toh, E., Ardizzone, C.M., Dong, A., Aaron, K.J., Luo, M., Tamhane, A., Lefkowitz, E.J. and Quayle, A.J., 2023. Characterization of vaginal microbial community dynamics in the pathogenesis of incident bacterial vaginosis, a pilot study. *Sexually Transmitted Diseases*, *50*(8), pp. 523–530.

52. Tiew, P.Y., Mac Aogain, M., Ali, N.A.T.B.M., Thng, K.X., Goh, K., Lau, K.J. and Chotirmall, S.H., 2020. The mycobiome in health and disease: emerging concepts, methodologies and challenges. *Mycopathologia*, *185*, pp. 207–231.

53. Fisher, M.C., Alastruey-Izquierdo, A., Berman, J., Bicanic, T., Bignell, E.M., Bowyer, P., Bromley, M., Brüggemann, R., Garber, G., Cornely, O.A. and Gurr, S.J., 2022. Tackling the emerging threat of antifungal resistance to human health. *Nature reviews microbiology*, *20*(9), pp. 557–571.

54. Qi, C., Hountras, P., Pickens, C.O., Walter, J.M., Kruser, J.M., Singer, B.D., Seed, P., Green, S.J. and Wunderink, R.G., 2019. Detection of respiratory pathogens in clinical samples using metagenomic shotgun sequencing. *Journal of medical microbiology*, *68*(7), pp. 996–1002.

55. Irinyi, L., Hu, Y., Hoang, M.T.V., Pasic, L., Halliday, C., Jayawardena, M., Basu, I., McKinney, W., Morris, A.J., Rathjen, J. and Stone, E., 2020. Long-read sequencing based clinical metagenomics for the detection and confirmation of Pneumocystis jirovecii directly from clinical specimens: a paradigm shift in mycological diagnostics. *Medical mycology*, *58*(5), pp. 650–660.

56. Zhang, J., Gao, L., Zhu, C., Jin, J., Song, C., Dong, H., Li, Z., Wang, Z., Chen, Y., Yang, Z. and Tan, Y., 2022. Clinical value of metagenomic next-generation sequencing by Illumina and Nanopore for the detection of pathogens in bronchoalveolar lavage fluid in suspected community-acquired pneumonia patients. *Frontiers in Cellular and Infection Microbiology*, *12*, p. 1021320.

57. Huang, J., Jiang, E., Yang, D., Wei, J., Zhao, M., Feng, J. and Cao, J., 2020. Metagenomic next-generation sequencing versus traditional pathogen detection in the diagnosis of peripheral pulmonary infectious lesions. *Infection and drug resistance*, pp. 567–576.

58. Marotz, C.A., Sanders, J.G., Zuniga, C., Zaramela, L.S., Knight, R. and Zengler, K., 2018. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome*, *6*, pp. 1–9.

59. Bruggeling, C.E., Garza, D.R., Achouiti, S., Mes, W., Dutilh, B.E. and Boleij, A., 2021. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. *Microbiologyopen*, *10*(3), p. e1191.

60. Ganda, E., Beck, K.L., Haiminen, N., Silverman, J.D., Kawas, B., Cronk, B.D., Anderson, R.R., Goodman, L.B. and Wiedmann, M., 2021. DNA extraction and host depletion methods significantly impact and potentially bias bacterial detection in a biological fluid. *Msystems*, *6*(3), pp. 10–1128.

61. Avanzi, C., Del-Pozo, J., Benjak, A., Stevenson, K., Simpson, V.R., Busso, P., McLuckie, J., Loiseau, C., Lawton, C., Schoening, J. and Shaw, D.J., 2016. Red squirrels in the British Isles are infected with leprosy bacilli. *Science*, *354*(6313), pp. 744–747.

62. Charalampous, T., Kay, G.L., Richardson, H., Aydin, A., Baldan, R., Jeanes, C., Rae, D., Grundy, S., Turner, D.J., Wain, J. and Leggett, R.M., 2019. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nature biotechnology*, *37*(7), pp. 783–792.

63. Cheng, W.Y., Liu, W.X., Ding, Y., Wang, G., Shi, Y., Chu, E.S., Wong, S., Sung, J.J. and Yu, J., 2022. High sensitivity of shotgun metagenomic sequencing in colon tissue biopsy by host DNA depletion. *Genomics, Proteomics & Bioinformatics*.

64. Sharon, I., Morowitz, M.J., Thomas, B.C., Costello, E.K., Relman, D.A. and Banfield, J.F., 2013. Time series community genomics analysis reveals rapid shifts

in bacterial species, strains, and phage during infant gut colonization. *Genome research*, *23*(1), pp. 111–120.

65. Smith, A.E., Zhang, Z., Thomas, C.R., Moxham, K.E. and Middelberg, A.P., 2000. The mechanical properties of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, *97*(18), pp. 9871–9874.

66. Wang, J., Liu, M., Shen, Y., Sun, J., Shao, Z. and Czajkowsky, D.M., 2018. Compressive force spectroscopy: from living cells to single proteins. *International Journal of Molecular Sciences*, *19*(4), p. 960.

67. Gibbs, E., Hsu, J., Barth, K. and Goss, J.W., 2021. Characterization of the nanomechanical properties of the fission yeast (Schizosaccharomyces pombe) cell surface by atomic force microscopy. *Yeast*, *38*(8), pp. 480–492.

68. Haugland, R.A., Brinkman, N. and Vesper, S.J., 2002. Evaluation of rapid DNA extraction methods for the quantitative detection of fungi using real-time PCR analysis. *Journal of microbiological methods*, *50*(3), pp. 319–323.

69. Scharf, S., Bartels, A., Kondakci, M., Pfeffer, K., Henrich, B. and Haas, R., 2020. Introduction of a bead beating step improves fungal DNA extraction from selected patient specimens. *International Journal of Medical Microbiology*, *310*(6), p. 151443.

70. Koide, R.T., 2005. Nucleic acid isolation from ecological samples—fungal associations, mycorrhizae. In *Methods in enzymology* (Vol. 395, pp. 58–72). Academic Press.

71. Lazarovitz, R., Romano, AE., Wu-Woods, N., Pradhan, O., Naydenkov, P., Pei, XP., Martin, A., Devkota, S., Ismagilov, R., 2024 Improved Microbial Enrichment Method Enables High Resolution Metagenomic Analysis of Low Microbial Load Tissue Samples. *In preparation.*

72. Duncan, A., Barry, K., Daum, C., Eloe-Fadrosh, E., Roux, S., Schmidt, K., Tringe, S.G., Valentin, K.U., Varghese, N., Salamov, A. and Grigoriev, I.V., 2022.

Metagenome-assembled genomes of phytoplankton microbiomes from the Arctic and Atlantic Oceans. *Microbiome*, *10*(1), p. 67.

73. Boyce, K.J. and Andrianopoulos, A., 2015. Fungal dimorphism: the switch from hyphae to yeast is a specialized morphogenetic adaptation allowing colonization of a host. *FEMS microbiology reviews*, *39*(6), pp. 797–811.

74. Zhao, G. and Rusche, L.N., 2021. Genetic analysis of sirtuin deacetylases in hyphal growth of Candida albicans. *Msphere*, *6*(3), pp. 10–1128.

75. Imbert, S., Portejoie, L., Pfister, E., Tauzin, B., Revers, M., Uthurriague, J., Hernandez-Grande, M., Lafon, M.E., Jubert, C., Issa, N. and Dumas, P.Y., 2023. A multiplex PCR and DNA-sequencing workflow on serum for the diagnosis and species identification for invasive aspergillosis and mucormycosis. *Journal of Clinical Microbiology*, *61*(1), pp. e01409–22.

76. Etienne, K.A., Roe, C.C., Smith, R.M., Vallabhaneni, S., Duarte, C., Escandón, P., Castañeda, E., Gómez, B.L., De Bedout, C., López, L.F. and Salas, V., 2016. Whole-genome sequencing to determine origin of multinational outbreak of Sarocladium kiliense bloodstream infections. *Emerging Infectious Diseases*, *22*(3), p. 476.

77. Schikora-Tamarit, M.À. and Gabaldón, T., 2022. Using genomics to understand the mechanisms of virulence and drug resistance in fungal pathogens. *Biochemical Society Transactions*, *50*(3), pp. 1259–1268.

78. Beheshti-Maal, A., Shahrokh, S., Ansari, S., Mirsamadi, E.S., Yadegar, A., Mirjalali, H. and Zali, M.R., 2021. Gut mycobiome: The probable determinative role of fungi in IBD patients. *Mycoses*, *64*(5), pp. 468–476.

79. Hong, G., Li, Y., Yang, M., Li, G., Qian, W., Xiong, H., Bai, T., Song, J., Zhang, L. and Hou, X., 2020. Gut fungal dysbiosis and altered bacterial-fungal interaction in patients with diarrhea-predominant irritable bowel syndrome: An explorative study. *Neurogastroenterology & Motility*, *32*(11), p. e13891.

80. Rokas, A., 2022. Evolution of the human pathogenic lifestyle in fungi. *Nature*

Microbiology, *7*(5), pp. 607–619.

81. Taylor, J.W., 2015. Evolutionary perspectives on human fungal pathogens. *Cold Spring Harbor perspectives in medicine*, *5*(9), p. a019588.

82. White, T.J., Bruns, T., Lee, S.J.W.T. and Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications*, *18*(1), pp. 315–322.

83. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F. and Bai, Y., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature biotechnology*, *37*(8), pp. 852–857.

84. Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, *17*(1), pp. 10–12.

85. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods*, *13*(7), pp. 581–583.

86. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, pp. 2825–2830.

87. Ficetola, G.F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P. and Pompanon, F., 2010. An in silico approach for the evaluation of DNA barcodes. *BMC genomics*, *11*, pp. 1–10.

88. Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D. and Karsch-Mizrachi, I., 2020. GenBank. *Nucleic acids research*, *48*(D1), p. D84.

89. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D. and Astashyn, A., 2016.

Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*, *44*(D1), pp. D733–D745.

90. Chen, S., 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta*, *2*(2), p. e107.

91. Cole, C.G., McCann, O.T., Collins, J.E., Oliver, K., Willey, D., Gribble, S.M., Yang, F., McLaren, K., Rogers, J., Ning, Z. and Beare, D.M., 2008. Finishing the finished human chromosome 22 sequence. *Genome Biology*, *9*, pp. 1–11.

92. Bushnell, B., Rood, J. and Singer, E., 2017. BBMerge–accurate paired shotgun read merging via overlap. *PloS one*, *12*(10), p. e0185056.

93. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. and Korobeynikov, A., 2020. Using SPAdes de novo assembler. *Current protocols in bioinformatics*, *70*(1), p. e102.

94. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A. and Zdobnov, E.M., 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular biology and evolution*, *38*(10), pp. 4647–4654.

95. Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, *25*(14), pp. 1754–1760.

96. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. and Li, H., 2021. Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2), p. giab008.

97. McKinney, W., van der Walt, S. and Millman, J., 2010. Proceedings of the 9th Python in Science Conference.

98. Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N., Yousef, M., Fogarty, E.C. and Trigodet, F., 2021. Community-led,

integrated, reproducible multi-omics with anvi'o. *Nature microbiology*, *6*(1), pp. 3–6.

99. Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Computing in science & engineering*, *9*(03), pp. 90–95.

100. Goel, M., Sun, H., Jiao, W.B. and Schneeberger, K., 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome biology*, *20*, pp. 1–13.

101. Goel, M. and Schneeberger, K., 2022. Plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics*, *38*(10), pp. 2922–2926.

102. Emms, D.M. and Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, *20*, pp. 1–14.

103. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A. and Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, *37*(5), pp. 1530–1534.

104. Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F. and Smirnova, T., 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic acids research*, *42*(D1), pp. D699–D704.

105. Grigoriev, I.V., Hayes, R.D., Calhoun, S., Kamel, B., Wang, A., Ahrendt, S., Dusheyko, S., Nikitin, R., Mondo, S.J., Salamov, A. and Shabalov, I., 2021. PhycoCosm, a comparative algal genomics resource. *Nucleic Acids Research*, *49*(D1), pp. D1004–D1011.

106. Smit, A.F.A., Hubley, R. and Green, P., 2015. RepeatMasker Open-4.0. 2013–2015.

107. Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O. and Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research*, *110*(1-4), pp. 462–467.

108. Price, A.L., Jones, N.C. and Pevzner, P.A., 2005. De novo identification of repeat families in large genomes. *Bioinformatics*, *21*(suppl_1), pp. i351–i358.

109. Salamov, A.A. and Solovyev, V.V., 2000. Ab initio gene finding in Drosophila genomic DNA. *Genome research*, *10*(4), pp. 516–522.

110. Birney, E. and Durbin, R., 2000. Using GeneWise in the Drosophila annotation experiment. *Genome Research*, *10*(4), pp. 547–548.

111. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., Von Mering, C. and Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular biology and evolution*, *34*(8), pp. 2115–2122.

112. Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R., 2005. InterProScan: protein domains identifier. *Nucleic acids research*, *33*(suppl_2), pp. W116–W120.

113. Saier Jr, M.H., Reddy, V.S., Tamang, D.G. and Västermark, Å., 2014. The transporter classification database. *Nucleic acids research*, *42*(D1), pp. D251–D258.

114. Melen, K., Krogh, A. and von Heijne, G., 2003. Reliability measures for membrane protein topology prediction algorithms. *Journal of molecular biology*, *327*(3), pp. 735–744.

115. Nielsen, H., Engelbrecht, J., Brunak, S. and Von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein engineering*, *10*(1), pp. 1–6.

116. Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I.,

Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G. and Nielsen, H., 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature biotechnology*, *40*(7), pp. 1023–1025.

117. Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S. and Rogozin, I.B., 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome biology*, *5*, pp. 1–28.

118. Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), pp. 27–30.

Supplemental material

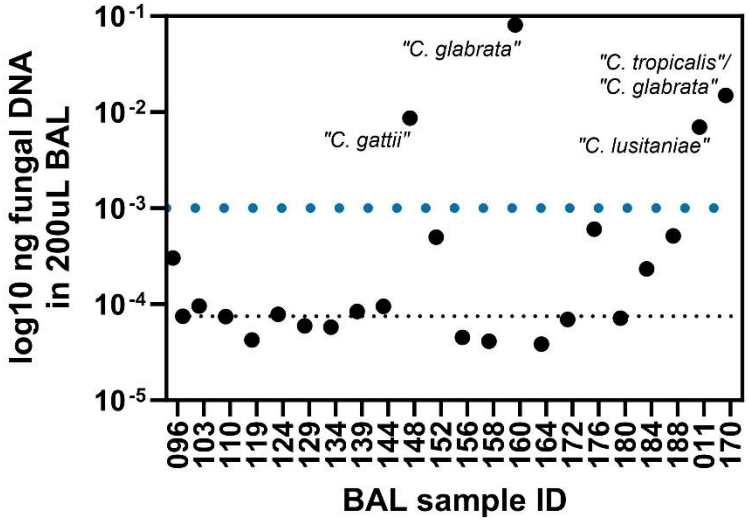


Figure 6.S1: Summary of BAL samples collected in this study. The log10 ng of fungal DNA extracted from 200 µL of BAL after MEM is plotted on the y axis for 22 samples collected in this study. The black horizontal dotted line indicates the extraction blank. The blue horizontal dotted line indicates the threshold drawn to proceed with sequencing (0.001 ng fungal DNA in 100 µL elution from 200 µL BAL specimen). The four BAL samples used in this study (011, 148, 160, and 170) are labeled with the dominant fungal species name(s) as determined by culture. Quotations are used around the species name to signify the classification of the isolate given by the clinic. The closest taxonomic classification by NCBI is given in Figure 6.S7.

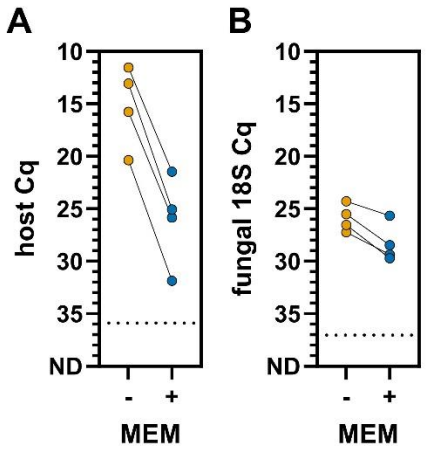


Figure S2: MEM reduces host DNA by 1000x while preserving fungal DNA within 10x. (A) The host DNA load in BAL fluid (measured with Zymo Human Femto qPCR, see methods) for seven samples with (blue) and without (orange) MEM treatment. (B) The fungal DNA load in BAL fluid (18S Cq) for five fungal culture-positive samples with (blue) and without (orange) MEM treatment. The horizontal dotted black lines indicate the Cq of the extraction blank negative control.

Table 6.S1: Distribution of Illumina sequencing reads from BAL samples before and after MEM.

sample (isolate)	MEM	total reads	unclassified reads	Unclassified reads mapping to fungal genome (%)	# unclassified reads mapping to fungal genome	# classified fungal reads	# host reads	# bacterial reads	% total fungal reads	% host reads	% bacterial reads
BAL_011 (C. lusitaniae)	-	40573734	748939	12.10%	90622	7350	39789705	27740	0.241%	98.068%	0.068%
BAL_011 (C. lusitaniae)	+	36303991	926586	11.96%	110820	40075	29599962	5737368	0.416%	81.534%	15.804%
BAL_148 (C. gattii)	-	57260744	1058802	35.88%	379898	142911	56039859	19172	0.913%	97.868%	0.033%
BAL_148 (C. gattii)	+	8999231	367140	83.00%	304726	4602943	3088702	940446	54.534%	34.322%	10.450%
BAL_160 (C. glabrata)	-	43077723	745298	13.96%	104044	23909	42301858	6658	0.297%	98.199%	0.015%
BAL_160 (C. glabrata)	+	50122420	1022970	19.52%	199684	4689936	44220293	189221	9.755%	88.225%	0.378%
BAL_170 (C. tropicalis/C. glabrata)	-	29619625	643434	34.27%	220505	7234	28940949	28008	0.769%	97.709%	0.095%
BAL_170 (C. tropicalis/C. glabrata)	+	102298040	3182796	65.87%	2096508	309935	76334472	22470837	2.352%	74.620%	21.966%

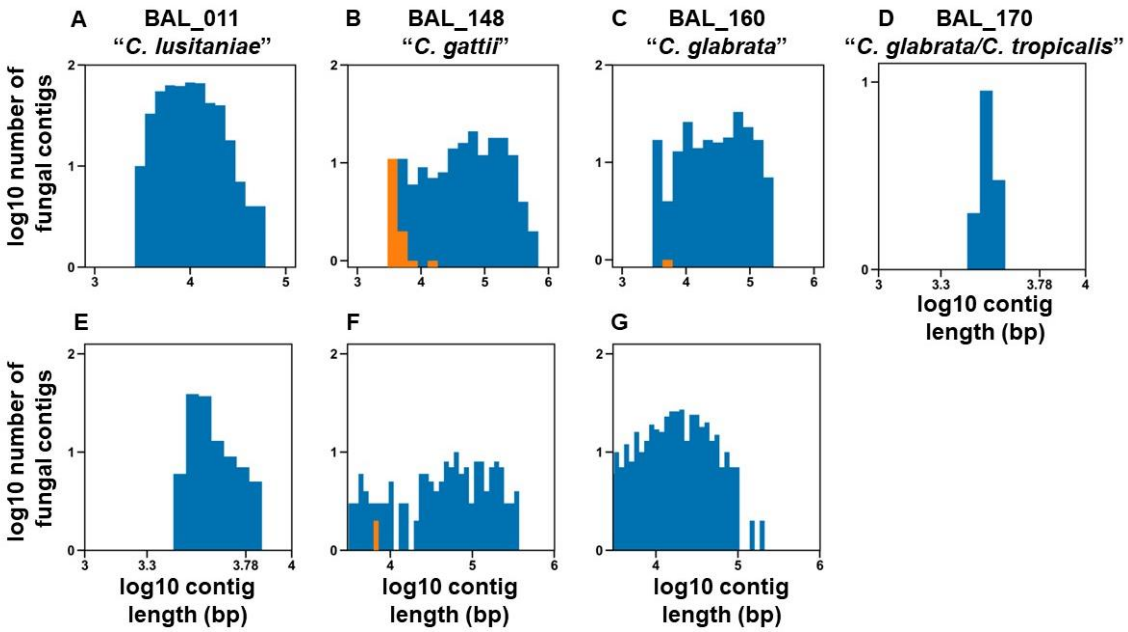


Figure 6.S3: Fungal contigs assembled from individual BAL samples. The log-log plot of the number of fungal contigs across four BAL samples as a factor of the length of each fungal contig (bp) for BAL fluid with (blue) and without (orange) MEM for (A) BAL_011 *Candida lusitaniae*, (B) BAL_148 *Cryptococcus gattii*, (C) BAL_160 *Candida glabrata*, and (D) BAL_170 *Candida glabrata*/*Candida tropicalis*. The log-log plot of fungal contigs across four BAL samples as a factor of the length of each fungal contig (bp) for BAL fluid with (blue) and without (orange) MEM for (E) BAL_011 *Candida lusitaniae*, (F) BAL_148 *Cryptococcus gattii*, and (G) BAL_160 *Candida glabrata* rarefied to 45M reads. The data for BAL_170 for rarefication is not shown as there were no contigs left. Quotations are used around the species name to signify the classification of the isolate given by the clinic. Sample BAL_170 contained two fungal isolates. The closest taxonomic classification by NCBI is given in Figure S7.

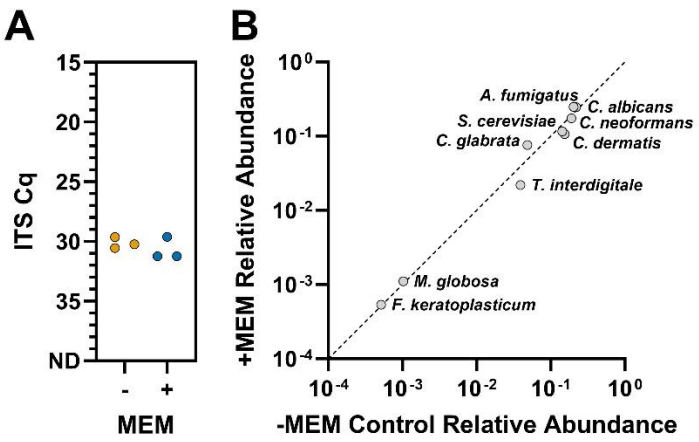


Figure 6.S4: MEM preserves fungal community composition as shown by ITS qPCR and amplicon sequencing. (A) The ITS Cq of the 10-taxa community mix with (blue) and without (orange) MEM processing. (B) The relative abundance of each fungal taxon in the 10-taxa community mix after MEM treatment as a factor of its relative abundance before MEM treatment, as measured by ITS amplicon sequencing. The relative abundances here reported for ITS are not the same as what is reported for 18S in Figure 6.2B due to mismatches in fungal primers against specific taxa.

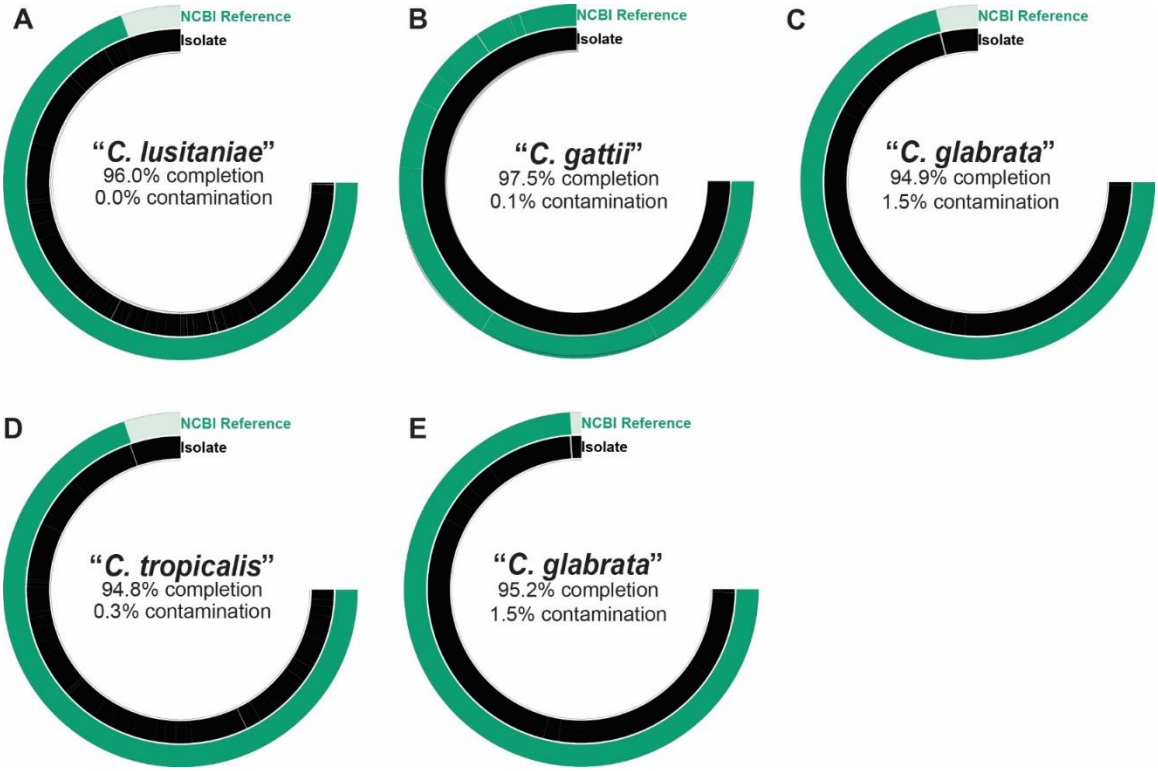


Figure 6.S5: Fungal genomes assembled from isolates cultured from BAL samples compared to NCBI reference genomes. The fungal genomes assembled from the clinical isolates (black) with the closely related NCBI reference genome (green) for each BAL sample. Corresponding genetic loci are darkly colored whole non-corresponding regions are lightly colored. (A) "*Candida lusitaniae"* from BAL_011 is being compared to *Clavispora lusitaniae* (GCF_000003835.1), (B) "*Cryptococcus gattii"* from BAL_148 is being compared to *Cryptococcus decagattii* (GCF_036417295.1), (C) "*Candida glabrata"* from BAL_160 is being compared to *Nakaseomyces glabratus* (GCF_000002545.3), (D) "*Candida tropicalis"* from BAL_170 is being compared to *Candida tropicalis* (GCF_000006335.3), and (E) "*C. glabrata"* BAL_170 is being compared to *Nakaseomyces glabratus* (GCF_000002545.3). Percent completion and contamination as determined by BUSCO are stated for each of the isolate genomes. Quotations are used around the species name to signify the classification of the isolate given by the clinic. Sample BAL_170 contained two fungal isolates. The closest taxonomic classification by NCBI is given in Figure 6.S7.

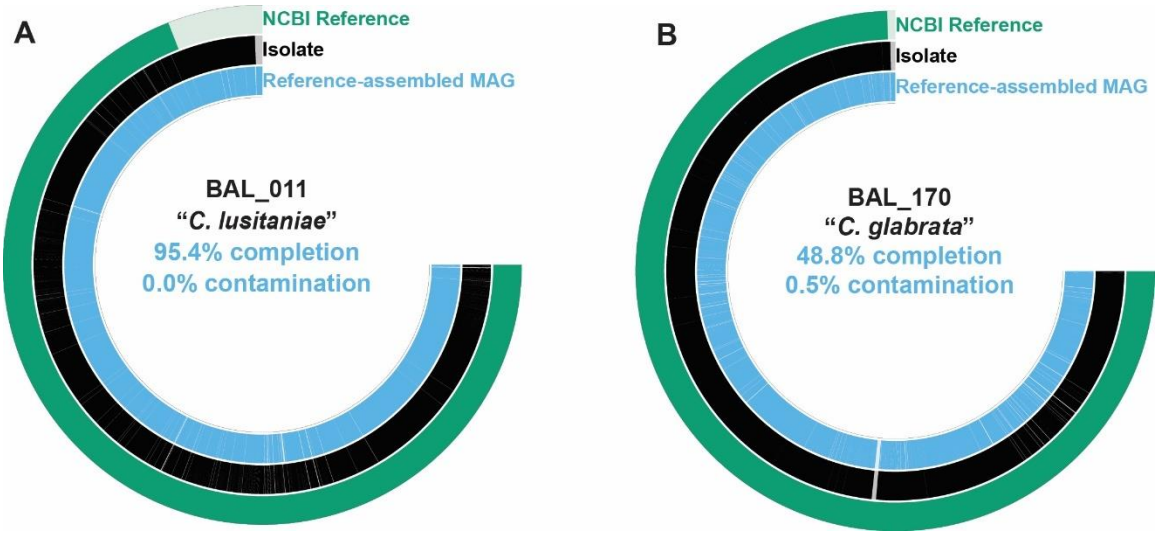


Figure 6.S6: MEM processing enables reference-based MAG assembly directly from BAL samples when *de novo* assembly is not possible. The reference-based assembled MAGs (light blue) are depicted alongside the assembled genomes derived from the isolates (black) and the most closely related NCBI reference genome (green) for (A) *Candida lusitaniae* from BAL_011 and (B) *C. glabrata* from BAL_170. Percentage completion and percentage contamination, as determined by BUSCO, are stated for each of the reference-assembled MAGs. Quotations are used around the species names to signify the classification of the isolate given in the clinic. The closest taxonomic classification by NCBI is given in Figure 6.S7.

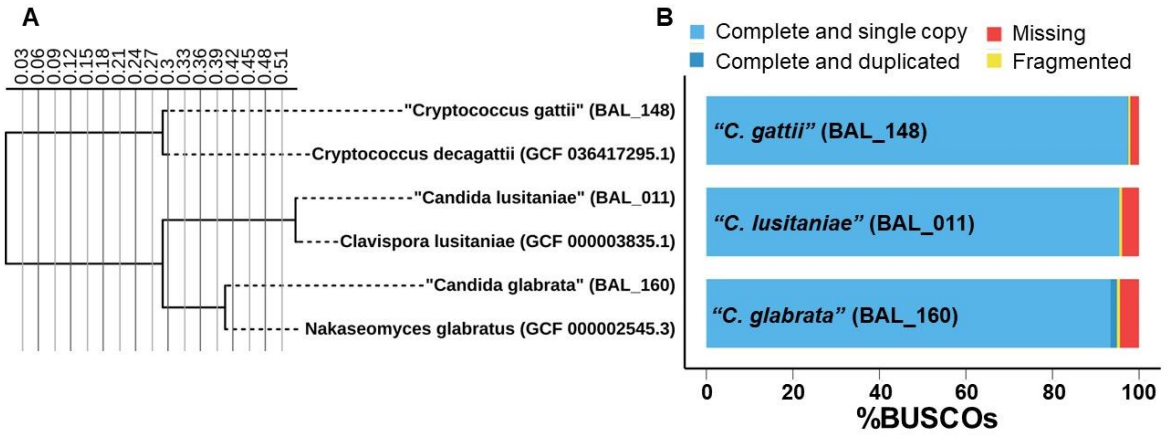


Figure 6.S7: Assessing taxonomic identities and the quality of the three fungal MAGs that were assembled to high (>90%) completion directly from BAL samples (from Figure 6.4 and Figure 6.S6) using reference genomes. (A) A phylogenetic tree indicating the relationship of the three fungal species to their most closely related NCBI reference genome (BAL sample ID and NCBI accession IDs indicated in parentheses). The scale bar indicates the number of nucleotide substitutions per site. (B) The %BUSCOs (Benchmarking Universal Single-Copy Orthologs) for the three fungal MAGs. Quotations around the species names signify the classification of the isolate given in the clinic.