Profiling the Microbial Communities of Brooks Cave in Pennington County, South Dakota

Riley S. Drake, Alex K. Shalek Massachusetts Institute of Technology

December 1, 2020

Introduction

Studies of cave bacteria have proven to be an incredibly promising research avenue for geology and biology alike – bacteria found in caves have shaped our understanding of everything from evolved antibiotic resistance (Pawlowski et al., 2016) to the vital role that certain microbes play in cave formation (Parker et al., 2017). This type of research may also prove useful for cave conservation: time-series monitoring of microbial communities within caves has been proposed as a sensitive method for identifying the effects of human travel through cave environments (Lü et al., 2018).

Motivated by the discovery of diverse, metabolically-active, microbial-only communities in the aquifer lakes of Wind Cave (Hershey et al., 2018), we began an investigation of the microbial communities in nearby Brooks Cave, which is also contained in the Madison limestone of the Black Hills (Long et al., 2012). We chose Brooks Cave for its lake; microgravity measurement studies have suggested that the lake in Brooks is a a perched lake (Roth and Long, 2012)—a body of water that is separate from nearby aquifers except at particularly high water levels. Recent studies have demonstrated that profiling the diversity and stability of microbial communities in underground water sources can inform researchers about the water's previous depth (Bougon et al., 2012), nutrient availability (Anderson et al., 2006), and position along hydrologic flow paths (Maamar et al., 2015). With continued time-series monitoring of the microbial communities in Brooks Lake, it will be possible to determine whether an intermittent connection exists between the lake and the aquifer.

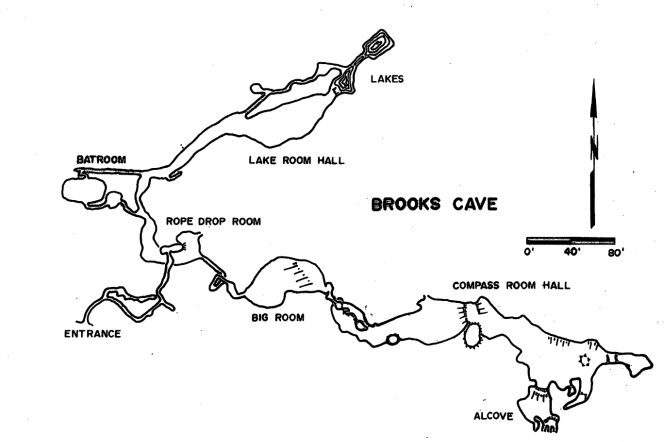

Fig. 1a. A historic map of Brooks Cave, most recently extended by Bruce Zerr who added 500 ft to the survey in 1971. Brooks Cave was discovered by Jim Brooks and John Dyer in 1954 (Black Hills Cave and Nature Conservancy, 2020).

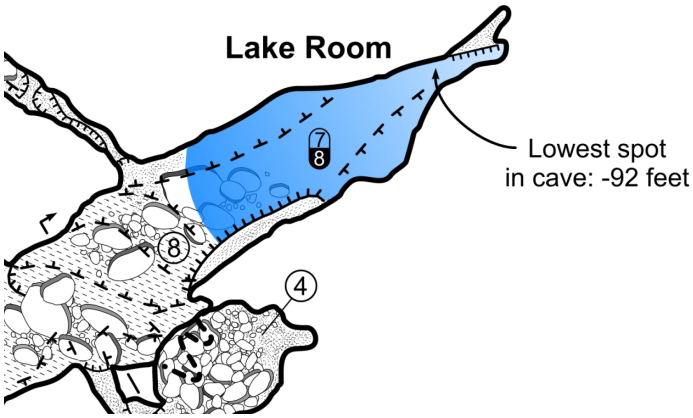

Fig. 1b. A rendering of the lake in Brooks Cave, cartography by Dan Austin. Compass and Tape Survey conducted by members of the Paha Sapa Grotto 2004-2007 (Austin et al., 2007).

Materials and Methods

We sought to characterize the microbial communities in Brooks Cave through unbiased genomic profiling. During two sampling campaigns in June 2019 and October 2020, 10 mL samples were collected outside of the path-of-human-travel using aseptic technique and immediately preserved in vivoPHIX (RNA Assist Ltd, RD-VIVO-20). To reduce the likelihood of confounding contamination, preparation controls were created as previously described (Hershey

Fig. 2a. Brooks Cave Lake, as it appeared during the first sampling campaign in June 2019.

Fig. 2b. Riley Drake (Shalek Lab, MIT) and Olivia Hershey (Barton Lab, University of Akron) taking water samples from Brooks Cave Lake in June 2019. Photo by Dave Springhetti.

et al., 2018) by subjecting all the sampling equipment to experimental processing in the absence of sample, including assembly both in the lab and at the field-site. Preparation controls were collected at every sampling site and processed in parallel with the samples, and remained site-separated through amplicon sequencing.

Samples were kept on ice for transportation and stored at -80°C before processing to prevent degradation. Cells were isolated via centrifugation at 1,000G for 10 minutes at 4°C. At an RNA clean workstation maintained using RNAzap (Invitrogen AM9780), cell pellets were washed 2x in 80% ethanol using sterile, low-retention filtered pipette tips (Rainin 30389166, 30389171, 30389173). Cell pellets were subsequently lysed with 10M Guanidinium Thiocyanate (Sigma Aldrich G9277) supplemented with 2% β-mercaptoethanol (Sigma Aldrich 63689) and agitated via bead-beating. Pulldown of genetic material was performed using RNAClean XP beads (Beckman Coulter A63987) which were subsequently washed 3x with 80% ethanol on a tube magnet (Eppendorf 12321D). After drying, the genetic material was eluted off of the AMPure beads using 15µL ultrapure water (Gibco 15230001). The DNA/RNA eluent was then PCR amplified using the universal primers 515F and 860R in three separate reactions (Caporaso et al., 2011)

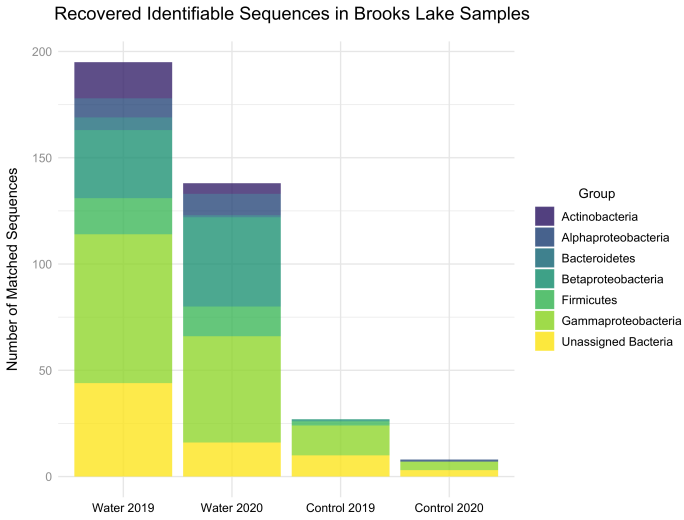

Fig. 3a. Number of sequences recovered from Brooks Lake water samples and corresponding controls, identified via 16S rRNA. Phyla representing less than 2.1% of total recovered sequences are omitted for clarity.

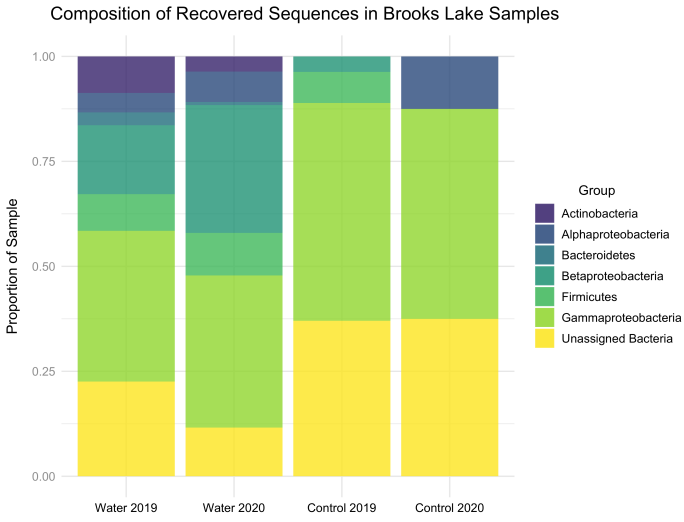

Fig. 3b. Relative composition of sequences recovered from Brooks Lake during the two sampling campaigns, identified via 16S rRNA. As above, phyla representing less than 2.1% of total recovered sequences are omitted for clarity.

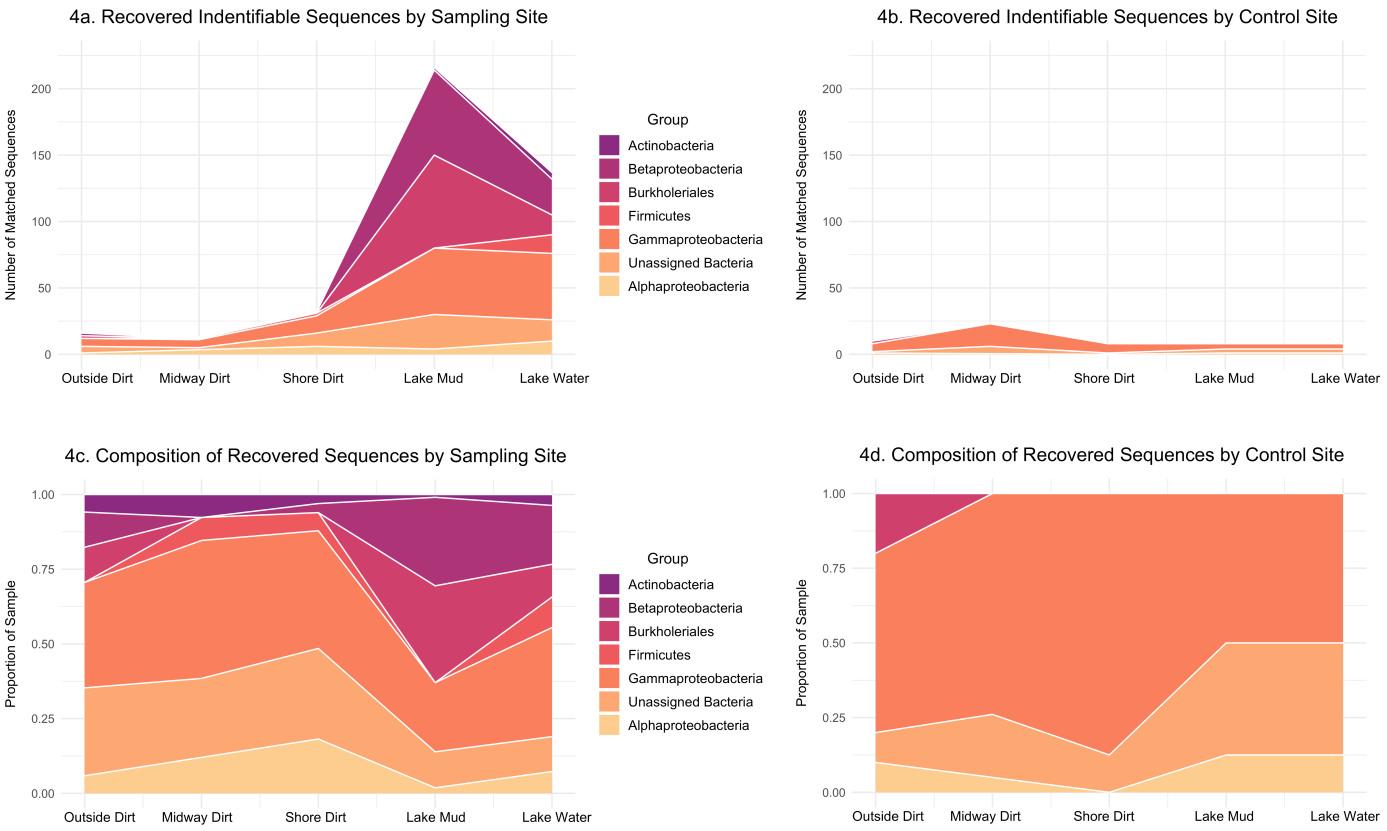

Fig. 4. (a) shows the number of sequences identifiable via 16S rRNA recovered from different sampling sites during the second sampling campaign in October 2020. (b) shows the number of identifiable sequences recovered from the preparation controls exposed at each site. Panels (c,d) normalize the data of (a,b) to show the relative composition of the microbial communities recovered from each sampling site. In all panels, phyla representing less than 1% of the sample population were omitted for clarity.

following the 16s Illumina Amplicon protocol developed by the Earth Microbiome Project (Earth Microbiome Project, 2016). Each sample was dual-indexed with Illumina N500 and N700 indices (Ilumina 20027213) using the Nextera XT DNA Library Preparation Kit (Illumina FC-131-1096) and the amplified DNA library was sequenced on an Illumina MiSeq using a MiSeq 600 Cycle Reagent Kit v3 (Illumina MS-102-3003) at Koch Institute BioMicro Center.

Analysis

The resulting read data were demultiplexed and sequencing artifacts were removed using the Galaxy Pipeline (Blankenberg et al., 2010). Quality filtering and trimming was performed using the FASTX toolkit (Gordon, 2010). Chimeric sequences were identified and removed using UCHIME (Schloss et al., 2009), and sampling-site specific assembly was performed using SPAdes (Bankevich et al., 2012). Subsequent assemblies were annotated using Kraken V2 (Wood and Salzberg, 2014; Lu and Salzberg, 2020), and figures were generated with custom R code (Version 3.6.3) using the Matplotlib (Hunter, 2007) and ggplot2 packages (Wickham, 2016).

Results

We identified considerable microbial diversity in Brooks Cave Lake; our initial sampling (June 2019) revealed 37 different orders of bacteria and archaea over 228 identifiable sequences as compared to the preparation control for that sample, which contained 13 orders of bacteria over 37 identifiable sequences. Similar diversity was observed in the resampling trip (October 2020), where we recovered 35 orders of bacteria over 152 identifiable sequences, with a preparation control that contained 6 orders of bacteria over 11 identifiable sequences. Between the two lake samples,

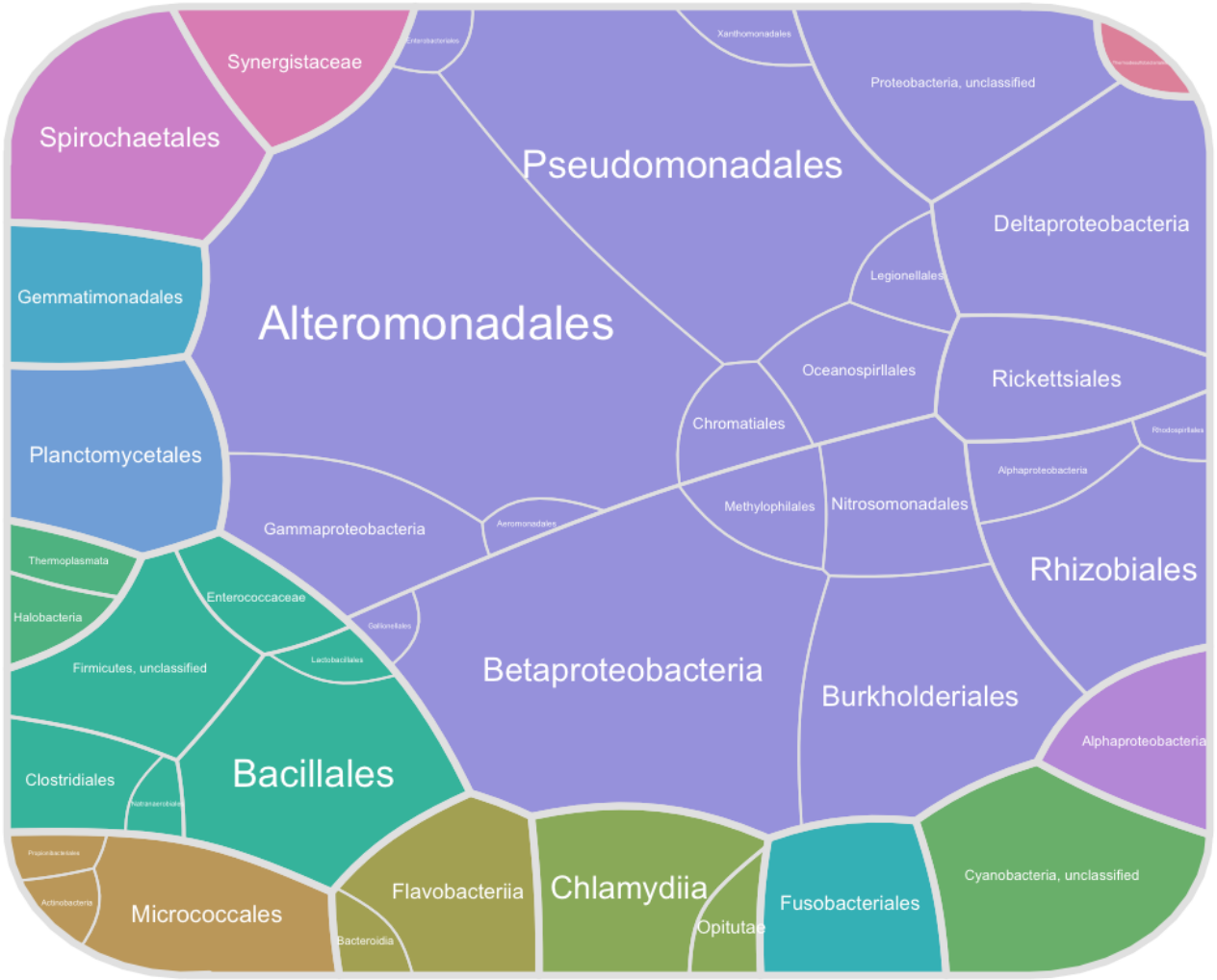

Fig. 5. Voronoi treemap diagram showing the relative abundance (indicated by the size of cells) and the class (indicated by cell color) of bacteria and archaea found in Brooks Lake during two sampling campaigns in June 2019 and October 2020. The image was generated by David Yang and Riley Drake using the SysbioTreemaps R package.

the relative composition of the microbial communities appeared relatively stable despite large observed fluctuations in the water level between sampling trips (**Fig. 3b**). If the water in Brooks Cave Lake were primarily surface water, we would have expected the composition of the microbial community to be unpredictable and subject to seasonal fluctuations (Griebler and Lueders, 2009). Brooks Cave is relatively short and shallow: the total surveyed length is 3,454.6' and the lake room, the lowest point in the cave, is just 92' below the datum (Austin et al., 2007). We wondered whether the microbial communities in such a short cave could be significantly different from those near the entrance. By progressively sampling several sites from the cave entrance to the lake, we found that the relative composition of the bacterial communities varied throughout the length of the cave (**Fig. 4c,d**), with a far higher absolute number of microbes further from the entrance (**Fig. 4a,b**).

The data we collected clearly demonstrates the existence of microbial communities which are compositionally distinct from the surface communities. Since the samples, taken over a year apart, had a relatively consistent composition, we suggest that the source of the water in Brooks Cave Lake is unlikely to be primarily surface water and is

more likely to be ground-water or aquifer derived. Because our data spans just two sampling trips, we do not yet believe it is sufficient to make broad conclusions about the stability of the microbial communities within Brooks Cave Lake. Future study of proposed perched ponds continues to be an exciting research avenue, especially if microbial monitoring proves useful in making predictions about local surface-groundwater mixing phenomena.

Acknowledgements

We are very grateful to Nick Anderson and Dave Springetti for their diligent completion of the second sampling campaign; Adam Weaver for coordinating access to the cave; Marilyn Borgeson for allowing us to access the cave; Rene Ohms and the Black Hills Cave and Nature Conservancy for facilitating and permitting the continuation of this research; Lee-Gray Boze for helping guide our geological inquiries and sending books; the Koch Institute BioMicro Core, especially Stuart Levine and Noelani Kamelamela, for their unrelenting support; Daniel Kramnik for lots of help with layout; Britt Goods, Maya Kaul and Sarah K. Nyquist for their substantial comments and edits.

This work was supported in part by the National Institute of Environmental Health Sciences of the NIH under award P30-ES002109.

References

Anderson, C. R., James, R. E., Fru, E. C., Kennedy, C. B., and Pedersen, K. (2006). In situ ecological development of a bacteriogenic iron oxide-producing microbial community from a subsurface granitic rock environment. *Geobiology*, 4(1):29–42.

Austin, D., Armstrong, B., Austin, S., Busch, M., Boze, L.-G., Foust, D., Geu, K., Ohms, M., Ohms, R., Resch, G., and et al. (2007).

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., and et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19(5):455–477.

Black Hills Cave and Nature Conservancy (2020). Brooks Cave.

Blankenberg, D., Gordon, A., Kuster, G. V., Coraor, N., Taylor, J., and Nekrutenko, A. (2010). Manipulation of FASTQ data with Galaxy. *Bioinformatics*, 26(14):1783–1785.

Bougon, N., Aquilina, L., Molénat, J., Marie, D., Delettre, Y., Chancerel, E., and Vandenkoornhuyse, P. (2012). Influence of depth and time on diversity of free-living microbial community in the variably saturated zone of a granitic aquifer. *FEMS Microbiology Ecology*, 80(1):98–113.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, 108(Supplement 1):4516–4522.

Earth Microbiome Project (2016). 16S Illumina Amplicon Protocol : Earth Microbiome Project. https://press.igsb.anl.gov/earthmicrobiome/protocols-and-standards/16s/.

Gordon, A. (2010). Fastx-toolkit: Fastq/a short-reads pre-processing tools.

Griebler, C. and Lueders, T. (2009). Microbial biodiversity in groundwater ecosystems. *Freshwater Biology*, 54(4):649–677.

Hershey, O. S., Kallmeyer, J., Wallace, A., Barton, M. D., and Barton, H. A. (2018). High Microbial Diversity Despite Extremely Low Biomass in a Deep Karst Aquifer.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Long, A. J., Ohms, M. J., and McKasey, J. D. (2012). Groundwater Flow, Quality (2007-10), and Mixing in the Wind Cave National Park Area, South Dakota. *U.S. Geological Survey*.

Lu, J. and Salzberg, S. L. (2020). Ultrafast and accurate 16S microbial community analysis using Kraken 2.

Lü, X.-F., He, Q.-F., Wang, F.-K., Zhao, R.-Y., and Zhang, H. (2018). Impact of Tourism on Bacterial Communities of Karst Underground River: A Case Study from Two Caves in Fengdu, Chongqing. *Environmental Science [Chinese]*, page 2389–2399.

Maamar, S. B., Aquilina, L., Quaiser, A., Pauwels, H., Michon-Coudouel, S., Vergnaud-Ayraud, V., Labasque, T., Roques, C., Abbott, B. W., Dufresne, A., and et al. (2015). Groundwater Isolation Governs Chemistry and Microbial Community Structure along Hydrologic Flowpaths. *Frontiers in Microbiology*, 6.

Parker, C. W., Auler, A. S., Barton, M. D., Sasowsky, I. D., Senko, J. M., and Barton, H. A. (2017). Fe(III) Reducing Microorganisms from Iron Ore Caves Demonstrate Fermentative Fe(III) Reduction and Promote Cave Formation. *Geomicrobiology Journal*, 35(4):311–322.

Pawlowski, A. C., Wang, W., Koteva, K., Barton, H. A., Mcarthur, A. G., and Wright, G. D. (2016). A diverse intrinsic antibiotic resistome from a cave bacterium. *Nature Communications*, 7(1).

Roth, K. R. and Long, A. J. (2012). Microgravity Methods for Characterization of Groundwater-Storage Changes and Aquifer Properties in the Karstic Madison Aquifer in the Black Hills of South Dakota, 2009–12. *U.S. Geological Survey*.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., and et al. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, 75(23):7537–7541.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wood, D. E. and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15(3).