Bacterial diversity in the waterholes of the Kruger National Park: an eDNA metabarcoding approach

Maxwell J. Farrell[1*], D. Govender[2,3], M. Hajibabaei[4], M. van der Bank[5], T. J. Davies[5,6]

[1]Department of Biology, McGill University
[2]Scientific Services, Kruger National Park, SANParks
[3]Department of Paraclinical Sciences, Faculty of Veterinary Science, University of Pretoria
[4]Integrative Biology & Biodiversity Institute of Ontario, University of Guelph
[5]African Centre for DNA Barcoding, University of Johannesburg
[6]Botany, Forest & Conservation Sciences, University of British Columbia
[*]To whom correspondence should be addressed; e-mail: maxwell.farrell@mail.mcgill.ca

September 12, 2018

Abstract

Bacteria are essential components of natural environments. They contribute to ecosystem functioning through roles as mutualists and pathogens for larger species, and as key components of food webs and nutrient cycles. Bacterial communities respond to environmental disturbances, and the tracking of these communities across space and time may serve as indicators of ecosystem health in areas of conservation concern. Recent advances in DNA sequencing of environmental samples allow for rapid and culture-free characterization of bacterial communities. Here we conduct the first metabarcoding survey of bacterial diversity in the waterholes of the Kruger National Park, South Africa. We show that eDNA can be amplified from waterholes and find strongly structured microbial communities, likely reflecting local abiotic conditions, animal ecology, and anthropogenic disturbance. Over timescales from days to weeks we find increased turnover in community composition, indicating bacteria may represent host-associated taxa of large vertebrates visiting the waterholes. Through taxonomic annotation we also identify pathogenic taxa, demonstrating the utility of eDNA metabarcoding for surveillance of infectious diseases. These samples serve as a baseline survey of bacterial diversity in the Kruger, and in the future, spatially distinct microbial communities may be used as markers of ecosystem disturbance, or biotic homogenization across the park.

Introduction

Traditional programs that monitor for early signs of ecosystem degradation require baseline data on the distributions and ecology of species in an ecosystem. DNA barcoding uses differences in conserved regions of genomes to classify sequences as belonging to particular taxonomic units, regardless of whether or not they have been described formally by taxonomists (Hebert et al., 2003; Blaxter et al., 2005; Ratnasingham and Hebert, 2013). DNA barcoding is thus a particularly powerful tool for exploring microbial diversity, where there are many undescribed taxa that cannot be cultured using traditional methods (Rappé and Giovannoni, 2003). Molecular barcoding coupled with recent advances in genetic sequencing have allowed for unprecedented exploration of microbial communities and the ability to characterize organisms of interest from environmental samples with great sensitivity (Shokralla et al., 2012). In particular, sequencing of cellular and extracellular DNA that can be extracted from environmental samples, collectively known as environmental DNA (eDNA) (Taberlet et al., 2012), is an emerging approach for exploring diversity in aquatic ecosystems (Rees et al., 2014; Lodge et al., 2012).

Microbial diversity in freshwater systems responds to environmental conditions (Lozupone and Knight, 2007), and perturbations (Zeglin, 2015) including multiple anthropogenic impacts such as urbanization (Fisher et al., 2015) and pollution (Bouskill et al., 2010). In addition to acting as indicators of ecosystem health, changes in microbial diversity may be important in themselves. Bacteria are essential components of ecosystems and play important roles in food webs, nutrient recycling, disease, and as important mutualists for larger multicellular species. Bacteria are thus integral to maintaining the natural balance of ecosystems and shifts in taxonomic composition due to environmental change may severely impact connectivity, functioning (Laforest-Lapointe et al., 2017; Delgado-Baquerizo et al., 2016) and increase exposure to pathogens (Cabral, 2010). However, in most ecosystems of conservation priority, microbial diversity is poorly described.

Surface waters are a vital resource for savannah ecosystems (Redfern et al., 2005; Owen-Smith, 1996), but frequent use by a large variety of species means they can also be a source of cross-species infection and spread of harmful pathogens (Bengis and Erasmus, 1988). These ecosystems provide an ideal context for refining eDNA metabarcoding approaches as they act as sources and sinks of

microbial species for larger animals, however baseline information about microbial diversity in these systems is lacking. Here we conduct a survey of bacterial diversity among watering holes of Kruger National Park, South Africa (KNP) through spatio-temporal sampling and sequencing of the V3-V4 region of 16S genes present in water. Water can be scarce in the KNP throughout the dry season and periods of drought (Redfern et al., 2005), and the park has a long history of water provisioning that included the construction of a series of more than 300 artificial waterholes beginning in the 1930s (Smit et al., 2007). These waterholes were intended to increase game numbers by stabilizing water availability year-round and are frequently visited by a diversity of birds and mammals (Smit and Grant, 2009). However, they have proven to alter the distributions of wildlife, which in turn have negative impacts on vegetation dynamics and the park-wide ecosystem (Smit et al., 2007; Smit and Grant, 2009). As a result, a number of artificial waterpoints have been closed since 1994 as the park began reverting to a more natural cycle of water availability (Smit et al., 2007; Van Wyk, 2011). A subset of the waterholes still open are small concrete troughs which are well mixed, largely mud and silt-free, and experience limited inflow from nearby surface waters. This means that eDNA samples will largely represent microbes in sourcewater and those dispersed by air and animals, allowing us to capture snapshots of local bacterial diversity across the park.

This study provides the first survey of bacterial diversity in the waterholes of the KNP, and is among the first studies using next-generation sequencing to describe aquatic microbial diversity in Africa (see Jordaan and Bezuidenhout (2016, 2013); Tekere, M., Prinsloo, A., Olivier, J., Jonker, N., Venter, S. (2012); Mwirichia, R., Cousin, S., Muigai, A. W., Boga, H. I., Stackebrandt, E. (2011); Tekere, M., Lötter, A., Olivier, J., Jonker, N., Venter, S. (2011)). Here we explore bacterial diversity across the southern half of the park and describe variation across, space, time, sample volume, and abiotic influences.

Methods

Study Site

Waterholes were sampled in June and July of 2015 in the Kruger National Park, South Africa (KNP), a large protected savannah ecosystem and a global diversity hotspot (Lahaye et al., 2008). Sampling was conducted during the dry season when natural sources of surface water are largely dry and watering hole visitation rates by medium and large vertebrates are highest. The park is divided into twenty-two ranger sections, which range in size from roughly 520 to 1,170 square kilometers. Across the southern half of the reserve below the Olifants river, ten concrete bottom artificial waterholes were selected from five of these sections (Table 1, Fig. 1). The waterholes varied in shape with some mimicking the contours of natural pans, making volume estimations difficult. However, the generic design included longer and shorter axes, with comparable dimensions across waterholes. Each waterhole is equipped with a ball-valve, which regulates water levels and re-fills the trough from nearby reservoirs when water levels drop. Water is sourced predominantly from groundwater via boreholes, but three sites use pipeline troughs filled with diverted river water.

Water Sampling and Processing

At each site, samples were taken once per week for three weeks. For one site (NWA), samples were also taken every day for five consecutive days. Sampling consisted of two 1L water samples collected in autoclaved, UV sterilized glass jars from opposite ends of the waterhole, approximately one foot from the nearest edge. These two within-waterhole samples (A/B) were taken along the waterhole's longest axis that maximized the distance and upwind-downwind gradient between them, if a strong wind was present. Water samples were placed on icepacks in a cooler and kept between 4-8°C until returning to the laboratory, where they were placed in the fridge.

Water quality parameters were taken during each sampling period using a YSI 650QS multi-parameter sonde. Temperature (°C), conductivity (mS/cm), dissolved oxygen (in mg/L and % saturation), and pH were recorded. Three measurements were taken along the same axis that the A/B water samples were drawn, and then averaged to measure quality per sample-time.

In the lab, the outside of water sample collection bottles were washed with ELIMINase (Decon Labs) and rinsed with deionized (DI) water to limit contamination. For each A/B sample, 150 mL of water was sub-sampled and filtered through gamma-irradiated 0.2 μm Supor hydrophilic polyethersulfone membranes (Pall no. 66234). The filtration apparatus consisted of three 300 mL Advantec polysulfone 47mm filter funnels fitted to a Pall vacuum manifold with vacuum pressure maintained by a Pall filtration vacuum/pressure pump (model no. 13158). After filtration, filters were stored in sterile 15 mL Falcon tubes and placed in a freezer at -60°C. On one sampling date for six sites, additional volumes of 50 ml and 15 ml were filtered from each 1L sample to asses the impact of sample volume filtered. Twice throughout sampling, BLANK samples were generated by filtering 1L of deionized water used in the laboratory.

Prior to and between filtrations, all funnel components and tweezers used to manipulate the filters were sterilized by soaking with 10% bleach for 10 minutes, rinsing with DI water, washing with ELIMINase, rinsing with DI water, and subsequent exposure to UV radiation for a minimum of 30 minutes. Gloves were worn at all times and changed between samples to minimize cross-sample contamination. To avoid sample freezing and bacterial growth in collection jars, all samples were processed within 12 hours of collection. Frozen 50 mL unfiltered voucher samples were kept and placed at -80°C at the University of Johannesburg's African Centre for DNA Barcoding for long term storage.

DNA Extraction, Amplification, and Sequencing

DNA was isolated from filter papers using MO BIO PowerWater DNA Isolation Kits. Universal bacterial primer sets designed by Sundquist et al. (2007) (V3-F: 5'ACTCCTACGGGAGGCAGCAG 3'; V4-R: 5'GGACTACARGGTATCTAAT 3') tagged with an Illumina adapter sequence were used to amplify the V3-V4 hypervariable region of the 16S ribosomal RNA gene through polymerase chain reaction (PCR). The PCR used a standard mix of 17.8μL molecular grade water, 2.5μL 10 reaction buffer (200mM Tris HCl, 500mM KCl, pH 8.4), 1μL $MgCl_2$ (50mM), 0.5μL dNTP (10mM), 0.5μL forward primer (10mM), 0.5μL reverse primer (10mM), 0.2μL Platinum Taq DNA polymerase (Invitrogen), and 2μL DNA as template for a total volume of 25μL. PCRs underwent the following cycler

conditions: initial 94°C for 5 minutes, then 30 cycles of 94°C for 40 seconds, 46°C for 1 minute, 72°C for 30 seconds, and a final temperature of 72°C for 2 minutes. Amplification success was confirmed through gel electrophoresis, using a 1.5% agarose gel. PCR products were purified using MinElute PCR purification kit (Qiagen), and quantified through flurometry using a Quant-iT PicoGreen dsDNA assay kit (Invitrogen). Samples were normalized, then multiplexed with the Nextera XT Index kit (96 indexes) (Illumina) and sequenced on an Illumina MiSeq flowcell using a V2 sequencing chemistry kit (2 x 250) making up approximately 1/8th of the run.

Sequence Processing, Taxonomy Assignment, and Phylogeny Construction

Across all samples, we generated a total of 2,164,262 Illumina reads. Primer sequences were removed using the trim.seqs function in mothur (Schloss et al., 2009). Reads were then processed in R (version 3.4.3) (R Development Core Team, 2008) using the package *dada2* version 1.6.0 (Callahan et al., 2016) following a modified version of the DADA2 Bioconductor workflow (Callahan et al., 2017) and online tutorials v1.6 and workflow for big data v1.4 (benjjneb.github.io/dada2/tutorial.html). Reads were filtered by quality, removing sequences with maximum expected error (maxEE) greater than 6 for both forward and reverse reads, and reads with any base pair having Q of 6 or lower. Reads were truncated to a length of 230bp and 220bp for forward and reverse reads respectively, consistent with dropoffs in quality profiles, and reads shorter than this were removed. Since the samples were sequenced across four different runs, subsequent steps of learning error rates, dereplication, denoising and Amplicon Sequence Variant (ASV) calling (Callahan et al., 2017) using pooled samples, and merging of paired reads were performed separately for each run. Tables of ASV sequences per sample within each run were then combined and chimera detection using all pooled samples was performed (see SM Table 2 for the number of reads retained across each step). In total 1,184,831 reads were retained, representing 3533 ASVs.

Taxonomy assignment from Kingdom to Genus was performed using the RDP classifier and SILVA nr v128 reference database (Quast et al., 2013) formatted for DADA2 (available at benjjneb.github.io/dada2/training.html), using the assignTaxonomy function (Fig. 5). ASVs assigned as Archaea, Eukarya, Chloroplast, or Mitochondria were removed. Species level assignments were

added by exact sequence matching using the addSpecies function. ASV sequences were aligned with the pynast algorithm via align_seqs.py in QIIME (Caporaso et al., 2010) and sequences with poor alignment automatically removed. A phylogenetic tree was constructed using the GTRCAT model in FastTree version 2.1.3 (Price et al., 2010) after filtering nucleotides with greater than 90% gap fraction and removing the 5% highest entropy positions with filter_alignment.py in QIIME (Caporaso et al., 2010). This resulted in a phylogenetic tree of 3393 ASVs which were used in subsequent community analyses.

Community Analyses

The ASV sequence table was merged with the phylogeny and sample metadata using the R pacakge *phyloseq* version 1.22.3 (McMurdie and Holmes, 2013). Negative controls (BLANK samples) used to investigate contamination during sample filtration contained 43 ASVs collectively (with 9 ASVs found in both samples). The sequence reads in each filtration blank were both dominated by the same ASV (53% and 86% respectively), however none of the 43 ASVs identified in the blanks were identified in any of the other samples. These control samples were removed prior to community analyses.

A subset of core samples was created by removing the first four daily NWA samples and samples of differential volume (S & XS samples), resulting in 54 samples of 150 mL each (Table 2). An ASV accumulation curve for core samples was generated using the specaccum function in the R package *vegan* version 2.4.6 (Oksanen et al., 2018) using the "exact" method, and extrapolated to total ASV richness using the Chao and Bootstrap methods in *vegan*'s specpool function. Alpha diversity was calculated for the core samples as observed ASV richness and Shannon diversity using the *phyloseq* package. Additive partitioning of Shannon diversity across core samples was investigated using the adipart function in *vegan* (Table 3).

Taxonomic composition was assessed by merging core samples at each site, and plotting relative abundances of reads for the most common taxa at levels of phylum, class, and order (Fig. 6). Temporal variation in taxonomic composition across core samples was assessed by merging A and B samples and plotting relative abundances of reads for sites with two or more weekly samples, for the levels of phylum (Fig. 7), and class (Fig. S6). To further investigate fine-scale temporal variation in

taxonomic composition (phylum, class, and order), relative abundance of reads were plotted across the daily samples at site NWA (Fig. 8). We also explored temporal turnover among samples with Sorensen's dissimilarity calculated using the beta.pair function from the *betapart* package (Baselga et al., 2018) (Fig. S9) and significant differences among daily and weekly samples was assessed using permutationsl multivariate analysis of variance (adonis in *vegan*) with 999 permutations each.

Community composition across sites in the core samples was described with nonmetric multidimentional scaling (NMDS) ordinations on relative ASV abundances per sample using the Bray-Curtis dissimilarity, and the abundance weighted Unifrac dissimilarity (Figs. 9, S10 & S11). Statistically, associations between dissimilarities and both water quality properties and common taxonomic groups were assessed using the envfit function in *vegan* for bacterial classes (Fig. 10) and orders (Fig. S12).

Phylogenetic community structure across core samples was calculated using standardized effect sizes of mean pairwise phylogenetic distances (MPD) and mean nearest taxon distances (MNTD) in the R package *picante* version 1.6.2 (Kembel et al., 2010) using the abundance weighted "richness" null model and 999 randomizations in the ses.mpd and ses.mntd functions (Fig. S13). For a given sample, MPD calculates the mean phylogenetic distance among each pair of taxa present, while MNTD calculates the mean phylogenetic distance from each taxa to its closest relative. These raw metrics give an estimate of how closely related community members are to each other, and are then compared to randomized communities to determine whether the observed metrics are different than what would expected if communities were assembled at random from taxa pooled across all samples.

To assess the effect of differential sample volumes, S (50 mL) and XS (15 mL) samples were subset along with their corresponding full volume samples (150 mL). Alpha diversity, calculated as observed ASV richness and Shannon diversity were calculated as described above (Fig. S14). Variation in taxonomic composition was investigated by comparing relative read abundances of bacterial phyla in A/B samples across sites and differential volumes (Fig. S15).

Raw reads with primers removed are available via the NCBI Sequence Read Archive BioProject PRJNA490450 (accession numbers SRR7822814 to SRR7822901). The ASV table, taxonomic assignments, phylogenetic tree, sample metadata, and scripts necessary to reproduce the results are included in the supplemental materials.

Results

Our sampling design aimed to sequence a core set of samples with all ten sites being sampled once per week for three weeks. Due to logistic constraints of sample storage and extremely low water levels from drawdown by animals, we were only able to process 27 of the planned 30 weekly samples (Table 2). In addition to this core sampling, we sequenced differential volumes for six samples, and an additional four daily samples from Nwaswitshaka (site NWA). A/B samples were taken at each site-time, resulting in a total of 88 sequenced samples, including the two filtration blanks. Across all 88 samples, we identified a total of 3393 ASVs. Roughly 15% of ASVs (n=524) were represented by a single read, together comprising fewer than 0.05% of all reads. The DADA2 approach infers the biological sequences in the sample prior to the introduction of amplification and sequencing errors, and can distinguish sequence variants differing by as little as one nucleotide. As such, we included all ASVs, including those represented by single reads, in subsequent analyses of biodiversity.

The ASV accumulation curve generated for the core sample set (2603 ASVs) does not appear to saturate (Fig. 2). Estimates of total richness using the Chao estimator predicts 6164 ASVs (+/- 262 SE) among the core samples, indicating we may be capturing less than half of the total bacterial diversity present among our sites. However, estimates of total diversity using the Bootstrap method were more conservative, with 3260 (+/- 146 SE) estimated ASVs. ASV diversity varied across sites (Figs. 3, Fig. 4), but the largest turnover (β diversity) was observed among park sections (Table 3). Variation among A/B samples contributed very small amounts to β diversity, indicating that at a particular time, microbial diversity within each waterhole was fairly well mixed.

In terms of taxonomic composition, 99.2% of ASVs were assigned to a known phylum, with the proportion of assignments decreasing at lower taxonomic levels (Fig. 5). The majority of bacterial ASVs were classified as Proteobacteria ($\sim$ 59%), followed by Bacteroidetes ($\sim$ 14%), Firmicutes ($\sim$ 9%), Actinobacteria ($\sim$ 6%), and Verrumicrobia ($\sim$ 2%). For bacterial classes, ASVs were largely classified as Betaproteobacteria ($\sim$ 34%), Alphaproteobacteria ($\sim$ 11%), Gammaproteobacteria ($\sim$ 10%), and Sphingobacteriia ($\sim$ 6%). Among core samples, relative abundances of phyla, classes, and orders varied across sites (Figs. 6, S4, S5). Across weeks, relative abundances of phyla varied within each site (Fig. 7), with some sites displaying more stability (IMB & HOY) compared to others

(NYA & NGO). Patterns among bacterial classes (Fig. S6) largely reflected variation among phyla, though one site (HOY) displayed much more variation in relative abundances among classes, reflecting substantial turnover within Proteobacteria. Comparing weekly turnover with the five daily samples taken at Nwaswitshaka (NWA) (Fig. 8), taxonomic composition appeared more stable across days than weeks. Using hierarchical clustering of Sorensen's dissimilarity, we find that samples taken within a single week cluster together (Fig. S9). Permutational multivariate ANOVAs on these distances revealed a significant difference in beta diversity among weekly samples (NWA 2,7,8; $p = 0.02$), with 49% of the variance explained by sample date, but no significant difference among the additional daily samples (NWA 3,4,5,6; $p = 0.54$), with 14% of the variance explained by sample date.

Community composition visualized through NMDS ordinations reflected results from the additive partitioning of diversity, with core samples clustering by site (Fig. 9) and section (Fig. S10) for both Bray-Curtis and abundance weighted UniFrac dissimilarities. Interestingly, waterholes filled by water from pipeline troughs (NGO, NYA, WIT) grouped together (Fig. S10), although these three sites are situated on a different geological type than sites fed by boreholes, making us unable to differentiate the effects of each factor (Fig. 11). Bacterial community composition was significantly structured by conductivity and pH for both Bray-Curtis and UniFrac dissimilarities, and dissolved oxygen also had an influence on UniFrac dissimilarity (Figs. 10 & S12).The dissimilarity of high conductivity sites (particularly HOY & IMB in the Kingfisherspruit section) was associated with high abundances of Clostridia, Gammaproteobacteria, and Bacteroidia, while sites with high pH and dissolved oxygen were positively related to the abundances of Actinobacteria and Alphaproteobacteria (Fig. 10).

Reflecting the NMDS structure of the abundance weighted UniFrac dissimilarities, MNTD, which is most sensitive to phylogenetic structure towards the tips of the tree (Mazel et al., 2015), indicated strong phylogenetic clustering within the majority of samples (Fig. S13). The strength of clustering was weaker for MPD, which is more sensitive to phylogenetic structure deeper in the tree (Mazel et al., 2015).

We did not find any clear decrease in alpha diversity with smaller sample volumes (Fig. S14), and one of the 15 mL samples returned the largest richness of ASVs, though the median value for 15 mL samples was lower and had a larger interquartile range than the 50 mL and 150 mL samples.

The major phyla detected within samples was also relatively consistent, with most groups represented across different sample volumes, though not always in the same proportions (Fig. S15).

Discussion

Biological monitoring is an essential aspect of conservation for tracking contemporary changes in ecosystems as well as providing a historical baseline for making management decisions. The Kruger National Park, established in 1898, has a long history of management practices revolving around maintenance of large mammals (Venter et al., 2008). While bacterial diversity has been explored for important infectious agents in the system (Michel et al., 2007; Bengis and Erasmus, 1988; Smith et al., 2000), recent advances in next generation sequencing methods now allow for the rapid and culture-free description of bacterial diversity throughout the park.

Here we present the first description of bacterial diversity in the waterholes of the Kruger National Park. In total we identified over 3000 unique taxa (referred to as amplicon sequence variants, or ASVs), only about half of which could be assigned to a previously described genus. The relative dominance of bacterial phyla was consistent with bacterial surveys of the Vaal River in central South Africa (Jordaan and Bezuidenhout, 2016). However, bacterial diversity was strongly structured across space, with the largest turnover in diversity occurring among park sections. This is not surprising considering the distances from site to site range from 3km to 115km and represent a gradient in large animal density, rainfall, vegetation, and major subsurface geology (Chirima et al., 2012; Van Wilgen et al., 2000; Smit and Grant, 2009; Smit et al., 2013). Samples also clustered by site, displaying substantial variation in taxonomic composition across sites. This variation was associated with physico-chemical properties of the water, with conductivity and pH being important explanatory variables. In addition to water quality, variability in taxonomic composition is likely influenced by the origin of the water used to fill each waterhole, differences in the surrounding soil and vegetation types, and the particular species and populations of animals using the waterholes.

We assessed daily turnover in composition at Nwaswitshaka, which appeared to be more stable over this shorter timescale when compared to turnover across weeks. However, Nwaswitshaka was less variable across weeks than other sites, indicating that daily variation in bacterial communities

could be greater in other locations. Important water quality variables (conductivity and pH) were largely consistent across weeks (Table S1), suggesting that the observed temporal heterogeneity may be driven by differences in external factors influencing bacterial input and removal from the system, such as variation in animal visitation throughout the sampling period. Between sampling events, water levels would sometimes drop substantially, indicating major drawdown by animals and likely removing bacteria deposited by animals visiting earlier in the week. Large mammal communities vary across the sampled regions of the park (Chirima et al., 2012), which may contribute to observed spatial variation in bacterial communities. However, different species also differ in their dependence on water, which is reflected in their rates of visitation to water points (Redfern et al., 2005). Variation in samples taken across subsequent weeks may therefore reflect different components of local animal communities, each with their unique host-associated bacterial taxa (Ley et al., 2008). By pairing bacterial composition with animal visitation prior to sampling (either through direct observation, or presence of genetic material), it may be feasible to build statistical probabilities of associations using co-occurrences of microbial and animal signatures.

Across samples, patterns of phylogenetic clustering were consistent with observed taxonomic variation. Multiple phyla were present in all samples, consistent with an even representation of deep bacterial lineages. However, turnover at lower taxonomic levels shown by significantly low mean nearest taxon distances indicate that there are distinct subsets of closely related taxa present at each site. This structuring may reflect filtering of bacterial communities by local environmental conditions, or the deposition of microbes by particular animal populations or individuals. Many vertebrate species have expansive home ranges, but during the dry season drought-intolerant animals will restrict their movement so as to stay close to permanent water bodies (Redfern et al., 2005). Thus the maintenance of major bacterial taxa may reflect both free-living environmental bacteria, and the core microbiome of water-dependent species. By taking repeat temporal samples, it may be possible to build association networks between bacterial taxa and host species, or even their local populations, solely from environmental DNA.

Through examination of taxonomic assignments we identified taxa belonging to genera that include important pathogens (*Arcobacter, Bacillus, Burkholderia, Coxiella, Legionella, Neisseria, Pas-*

teurella, Rickettsia, and *Yersinia*). While many of these genera include both pathogenic as well as benign species found in environmental samples, some of these genera are comprised solely of pathogenic species. For example, the genus *Coxiella* is represented by one species, *Coxiella burnetii*, the causative agent of Q fever, which has previously been documented as causing disease in the park (Van Heerden et al., 1995). Additionally, taxa in the order *Chlamydiales* are all obligate intracellular pathogens of eukaryotes (Ball et al., 2015), and taxa in the genus *Neisseria* colonize mucosal surfaces of animals, some of which are pathogenic in humans (Liu et al., 2015). Interestingly, we also identified sequences classified as *Streptococcus urinalis*, a recently described species linked to urinary tract infections in humans (Peltroche-Llacsahuanga et al., 2012). While hypervariable regions of the 16S gene may not be the optimal genomic regions for detecting the presence of particular pathogenic species or strains, our findings indicate that broad scale surveys of microbial diversity may be useful in determining the presence of potential pathogens across vastly divergent groups of bacteria. This can in turn guide more targeted sampling of both pathogenic and commensal bacteria across the park.

In addition to the presence of pathogens, surveys of bacterial diversity may be used to detect anthropogenic influences in the park. For example, we found that bacterial diversity was quite different for the two sites sampled in the Kingfisherspruit section (Hoyo Hoyo and Imbali). These sites were dominated by an ASV assigned to the genus *Arcobacter*, but which did not exactly match any sequence in the SILVA reference database. Three of the five described members of this genus are known to be pathogenic (Fera et al., 2004) and include *A. butzleri*, which can cause severe diarrhea (Lerner et al., 1994) and was detected in two of the weekly samples at Hoyo Hoyo. The two waterholes in Kingfisherspruit are fed in part with greywater that is passed through reed beds. Greywater is untreated household wastewater that typically has not been contaminated by toilet waste and is often used as year-round sources of water, especially in water scarce areas (Nganga et al., 2012). Compared to source water, kitchen and laundry greywater can have elevated conductivity (Nganga et al., 2012), which may explain the high conductivity of water at these sites, and present a strong selective environment driving their unique bacterial communities.

We did not have sufficient sampling of sites to explore all possible drivers of differences in bacterial communities. Nonetheless, some features differed obviously among sites. For example, Witpens

is a heavily vulture-dominated site. Bathing by vultures likely results in large influxes of nutrients such as blood, and vulture feces has been found to alter soil bacterial communities through elevated nitrogen and decreased pH (Ganz et al., 2012). Visually, water from Witpens was bright green, indicating high abundance of photosynthetic species and consistent with the large variations observed in dissolved oxygen (Table S1). However, we found no evidence of elevated abundances of *Microcystis* or other Cyanobacteria, though samples from Witpens strongly clustered together and had relatively high abundances of Rhizobiales and Rhodocylales, both of which include species known to fix nitrogen (Carvalho et al., 2010; Loy et al., 2005).

Witpens, along with Nyamarhi and Ngotso North are filled by pipeline troughs that divert river water to waterholes many kilometers away. While pipeline troughs are likely to reflect a subsample of the diversity found in river water, the acts of pipeline transport and storage themselves may have strong filtering effects on bacterial communities. Our results indicate that waterholes represent locally unique bacterial communities, thus the practice of diverting river water to waterholes kilometers away may homogenize microbial diversity across the landscape, ultimately disrupting local communities. The consequences of such shifts in community structure are difficult to assess without comparing pipeline troughs with their source waters, but the diversion of river water may have unintended impacts on microbial diversity. For example, genes conferring antimicrobial resistance have been shown to spread from river water to impala in the Kruger National Park (Mariano et al., 2009).

Here we show that eDNA can be amplified from waterholes in the Kruger National Park, and find strongly structured microbial communities, likely reflecting local abiotic conditions, animal ecology, and anthropogenic disturbance. We suggest that disruption of spatially distinct microbial communities may be used as a marker of ecosystem disturbance, or biotic homogenization across the park. We find that for artificial waterholes, bacterial diversity is surprisingly insensitive to sample volume, with even small volumes useful for capturing major components of bacterial communities, though larger volumes are necessary to detect rare taxa. Replicating this study across different seasons, and expanding sampling to include natural waterpoints will provide improved understanding of the roles micro-organisms play in ecosystem stability and resilience, and offer an effective method for monitoring of shifting species interactions in the face of environmental change. Just as studies of the

microbiome have revolutionized our understanding of human health, metagenomic analysis of environmental DNA have the potential to revolutionize our understanding of ecosystem health. Tracking of bacterial communities can provide a template for monitoring ecosystem disturbance through their response to biological contaminants, documenting the spread of invasive species or infectious organisms, and better understanding the impacts ecological disturbances have on the composition of native communities.

Acknowledgements

This work would not have been possible without the logistical support of the African Centre for DNA Barcoding and Purvance Shikwambana, molecular work by Shannon Eagle, encouragement and support from Lea Berrang-Ford, and the incredible knowledge and protection provided by Kruger game guards Thomas, Velly, and Martin. We also thank Steven Kembel, his lab, Alexis Carteron, and Benjamin Callahan for constructive feedback on bioinformatic analyses. Special thanks are due to B.S. Bezeng for volunteering his friendship, dedication, and long days in the field. This work was performed with the support and permission of SANParks (project DAVIJ1256). MJF was supported by a Vanier NSERC CGS and the CIHR Systems Biology Training Program, with project funding supported by the Quebec Centre for Biodiversity Science, the McGill Biology Department, and an NSERC Discovery Grant awarded to TJD.

References

Ball, S. G., C. Colleoni, D. Kadouche, M. Ducatez, M. C. Arias, and C. Tirtiaux (2015). Toward an understanding of the function of Chlamydiales in plastid endosymbiosis. *Biochimica et Biophysica Acta - Bioenergetics 1847*(6-7), 495–504.

Baselga, A., D. Orme, S. Villeger, J. De Bortoli, and F. Leprieur (2018). *betapart: Partitioning Beta Diversity into Turnover and Nestedness Components*. R package version 1.5.0.

Bengis, R. G. and J. M. Erasmus (1988). Wildlife diseases in South Africa: a review. *Revue Scientifique et Technique, (OIE) 7*(4), 807–821.

Blaxter, M., J. Mann, T. Chapman, F. Thomas, C. Whitton, R. Floyd, and E. Abebe (2005). Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences 360*(1462), 1935–1943.

Bouskill, N. J., J. Barker-Finkel, T. S. Galloway, R. D. Handy, and T. E. Ford (2010). Temporal bacterial diversity associated with metal-contaminated river sediments. *Ecotoxicology 19*(2), 317–328.

Cabral, J. P. (2010). Water microbiology. Bacterial pathogens and water. *International Journal of Environmental Research and Public Health 7*(10), 3657–3703.

Callahan, B. J., P. J. McMurdie, and S. P. Holmes (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal 11*(12), 2639–2643.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods 13*(7), 581–583.

Callahan, B. J., K. Sankaran, J. A. Fukuyama, P. J. McMurdie, and S. P. Holmes (2017). Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research 5*(3), 1492.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. a. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. a. Lozupone, D. Mcdonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. a. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight (2010). QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nature Publishing Group 7*(5), 335–336.

Carvalho, F. M., R. C. Souza, F. G. Barcellos, M. Hungria, and A. T. R. Vasconcelos (2010). Genomic and evolutionary comparisons of diazotrophic and pathogenic bacteria of the order Rhizobiales. *BMC Microbiology 10*.

Chirima, G. J., N. Owen-Smith, and B. F. Erasmus (2012). Changing distributions of larger ungulates in the Kruger National Park from ecological aerial survey data. *Koedoe 54*(1), 1–11.

Delgado-Baquerizo, M., F. T. Maestre, P. B. Reich, T. C. Jeffries, J. J. Gaitan, D. Encinar, M. Berdugo, C. D. Campbell, and B. K. Singh (2016). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nature Communications 7*, 1–8.

Fera, M. T., T. L. Maugeri, C. Gugliandolo, C. Beninati, E. L. Camera, M. Carbone, and M. Giannone (2004). Detection of Arcobacter spp . in the Coastal Environment of the Mediterranean Sea Detection of Arcobacter spp . in the Coastal Environment of the Mediterranean Sea. *Applied and Environmental Microbiology 70*(3), 1271–1276.

Fisher, J. C., R. J. Newton, D. K. Dila, and S. L. McLellan (2015). Urban microbial ecology of a freshwater estuary of Lake Michigan. *Elementa: Science of the Anthropocene 3*, 000064.

Ganz, H. H., U. Karaoz, W. M. Getz, W. Versfeld, and E. L. Brodie (2012). Diversity and structure of soil bacterial communities associated with vultures in an African savanna. *Ecosphere 3*(6), art47.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard (2003). Biological identifications through DNA barcodes. *Proceedings. Biological sciences / The Royal Society 270*(1512), 313–21.

Jordaan, K. and C. C. Bezuidenhout (2013). The impact of physico-chemical water quality parameters on bacterial diversity in the Vaal River, South Africa. *Water SA 39*(3), 385–396.

Jordaan, K. and C. C. Bezuidenhout (2016). Bacterial community composition of an urban river in the North West Province, South Africa, in relation to physico-chemical water quality. *Environmental Science and Pollution Research 23*(6), 5868–5880.

Kembel, S., P. Cowan, M. Helmus, W. Cornwell, H. Morlon, D. Ackerly, S. Blomberg, and C. Webb (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics 26*, 1463–1464.

Laforest-Lapointe, I., A. Paquette, C. Messier, and S. W. Kembel (2017). Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature 546*(7656), 145–147.

Lahaye, R., M. van der Bank, D. Bogarin, J. Warner, F. Pupulin, G. Gigot, O. Maurin, S. Duthoit, T. G. Barraclough, and V. Savolainen (2008). DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences 105*(8), 2923–2928.

Lerner, J., V. Brumberger, and V. Preac-Mursic (1994). Severe diarrhea associated with Arcobacter butzleri. *European Journal of Clinical Microbiology & Infectious Diseases 13*(8), 660–662.

Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight, and J. I. Gordon (2008). Evolution of mammals and their gut microbes. *Science 320*(5883), 1647–1651.

Liu, G., C. M. Tang, and R. M. Exley (2015). Non-pathogenic Neisseria: members of an abundant, multi-habitat, diverse genus. *Microbiology (Reading, England) 161*(7), 1297–1312.

Lodge, D., C. Turner, and C. Jerde (2012). Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology 21*(11), 2555–8.

Loy, A., C. Schulz, S. Lücker, A. Schöpfer-Wendels, K. Stoecker, C. Baranyi, A. Lehner, and M. Wagner (2005). 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". *Applied and Environmental Microbiology 71*(3), 1373–1386.

Lozupone, C. A. and R. Knight (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America 104*(27), 11436–11440.

Mariano, V., C. M. E. McCrindle, B. Cenci-Goga, and J. A. Picard (2009). Case-control study to determine whether river water can spread tetracycline resistance to unexposed impala (aepyceros melampus) in Kruger National Park (South Africa). *Applied and Environmental Microbiology 75*(1), 113–118.

Mazel, F., T. J. Davies, L. Gallien, J. Renaud, M. Groussin, T. Munkemuller, and W. Thuiller (2015). Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. *Ecography*.

McMurdie, P. J. and S. Holmes (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE 8*(4).

Michel, A. L., L.-M. de Klerk, N. van Pittius, R. M. Warren, and P. D. van Helden (2007). Bovine tuberculosis in African buffaloes: observations regarding Mycobacterium bovis shedding into water and exposure to environmental mycobacteria. *BMC Veterinary Research 3*(1), 23.

Mwirichia, R., Cousin, S., Muigai, A. W., Boga, H. I., Stackebrandt, E. (2011). Bacterial diversity in the haloalkaline lake elmenteita, Kenya. *Current Microbiology 62*(1), 209–221.

Nganga, V. G., F. W. Kariuki, and K. Kotut (2012). A comparison of the physico-chemical and bacteriological quality of greywater from water deficient households in Homabay town and Githurai estates in Kenya. *Open Environmental Engineering Journal 5*, 110–118.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner (2018). *vegan: Community Ecology Package*. R package version 2.4-6.

Owen-Smith, N. (1996). Ecological guidelines for waterpoints. *South African Journal of Wildlife Research 26*(4), 107–112.

Peltroche-Llacsahuanga, H., B. Frye, and G. Haase (2012). Isolation of Streptococcus urinalis from a human blood culture. *Journal of Medical Microbiology 61*(5), 740–742.

Price, M. N., P. S. Dehal, and A. P. Arkin (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE 5*(3).

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research 41*(D1), 590–596.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rappé, M. S. and S. J. Giovannoni (2003). The Uncultured Microbial Majority. *Annual Review of Microbiology 57*(1), 369–394.

Ratnasingham, S. and P. D. N. Hebert (2013). A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE 8*(7).

Redfern, J., C. Grant, a. Gaylard, and W. Getz (2005). Surface water availability and the management of herbivore distributions in an African savanna ecosystem. *Journal of Arid Environments 63*(2), 406–424.

Rees, H. C., B. C. Maddison, D. J. Middleditch, J. R. M. Patmore, and K. C. Gough (2014). The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*.

Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology 75*(23), 7537–7541.

Shokralla, S., J. L. Spall, J. F. Gibson, and M. Hajibabaei (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular ecology 21*(8), 1794–805.

Smit, I. and C. Grant (2009). Managing surface-water in a large semi-arid savanna park: Effects on grazer distribution patterns. *Journal for Nature Conservation 17*(2), 61–71.

Smit, I. P., C. C. Grant, and B. J. Devereux (2007). Do artificial waterholes influence the way herbivores use the landscape? Herbivore distribution patterns around rivers and artificial surface water sources in a large African savanna park. *Biological Conservation 136*(1), 85–99.

Smit, I. P., E. S. Riddell, C. Cullum, and R. Petersen (2013). Kruger National Park research supersites: Establishing long-term research sites for cross-disciplinary, multiscaled learning. *Koedoe 55*(1), 1–7.

Smith, K. L., V. DeVos, H. Bryden, L. B. Price, M. E. Hugh-Jones, and P. Keim (2000). Bacillus anthracis diversity in Kruger National Park. *Journal of Clinical Microbiology 38*(10), 3780–3784.

Sundquist, A., S. Bigdeli, R. Jalili, M. L. Druzin, S. Waller, K. M. Pullen, Y. Y. El-Sayed, M. M. Taslimi, S. Batzoglou, M. Ronaghi, R. Amann, W. Ludwig, K. Schleifer, M. Rappé, S. Giovannoni, S. Tringe, E. Rubin, B. Anderson, J. Dawson, D. Jones, K. Wilson, R. Verhelst, H. Verstraelen, G. Claeys, G. Verschraegen, J. Delanghe, L. V. Simaey, C. D. Ganck, M. Temmerman, M. Vaneechoutte, R. Goldenberg, J. Hauth, W. Andrews, M. Gravett, M. Novy, R. Rosenfeld, A. Reddy, T. Jacob, M. Turner, A. Sbarra, R. Selvaraj, C. Cetrulo, M. Feingold, E. Newton, G. Thomas, J. McGregor, D. Lawellin, A. Franco-Buff, J. Todd, E. Makowski, W. Andrews, J. Hauth, R. Goldenberg, R. Gomez, R. Romero, G. Cassell, S. Fortunato, R. Menon, K. Swan, R. Menon, S. Fortunato, R. Menon, S. Lombarda, M. Margulies, M. Egholm, W. Altman, S. Attiya, J. Bader, L. Bemben, J. Berka, M. Braverman, Y. Chen, Z. Chen, S. Dewell, L. Du, J. Fierro, X. Gomes, B. Godwin, W. He, S. Helgesen, C. Ho, G. Irzyk, S. Jando, M. Alenquer, T. Jarvie, K. Jirage, J.-B. Kim, J. Knight, J. Lanza, J. Leamon, S. Lefkowitz, M. Lei, J. Li, K. Lohman, H. Lu, V. Makhijani, K. McDade, M. McKenna, E. Myers, E. Nickerson, J. Nobile, R. Plant, B. Puc, M. Ronan, G. Roth, G. Sarkis, J. Simons, J. Simpson, M. Srinivasan, K. Tartaro, A. Tomasz, K. Vogt, G. Volkmer, S. Wang, Y. Wang, M. Weiner, P. Yu, R. Begley, J. Rothberg, Y. Rogers, J. Venter, M. Ronaghi, S. Karamohamed, B. Petterson, M. Uhlen, P. Nyren, R. Hyman, M. Fukushima, L. Diamona, J. Kumm, L. Giudice, R. Davis, M. Sogin, H. Morrison, J. Huber, D. Welch, S. Huse, P. Neal, J. Arrieta, G. Herndl, W. Kent, J. Cole, B. Chai, R. Farris, Q. Wang, S. Kulam, D. McGarrell, G. Garrity, J. Tiedje, T. DeSantis, I. Dubosarskiy, S. Murray, G. Andersen, F. V. Wintzingerode, U. Gobel, E. Stackebrandt, H. Monstein, S. Nikpour-Badi, J. Jonasson, M. Gray, D. Sankoff, R. Cedergren, J. Carey, M. Klebanoff, J. Hauth, S. Hillier, E. Thom, J. Ernest, R. Heine, R. Nugent, M. Fischer, K. Leveno, R. Wapner, M. Varner, W. Trout, A. Moawad, B. Sibai, M. Miodovnik, M. Dombrowski, M. O'Sullivan, J. V. Dorsten, O. Langer, J. Roberts, M. Klebanoff, J. Carey, J. Hauth, S. Hillier, R. Nugent, E. Thom, J. Ernest, R. Heine, R. Wapner, W. Trout, A. Moawad, M. Miodovnik, B. Sibai, J. V. Dorsten, M. Dombrowski, M. O'Sullivan, M. Varner, O. Langer, D. McNellis, J. Roberts, K. Leveno, J. Neefs, Y. V. de Peer, P. D. Rijk, A. Goris, and R. D. Wachter (2007). Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiology 7*(1), 108.

Taberlet, P., E. Coissac, M. Hajibabaei, and L. H. Rieseberg (2012). Environmental DNA. *Molecular*

ecology 21(8), 1789–93.

Tekere, M., Lötter, A., Olivier, J., Jonker, N., Venter, S. (2011). Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa. *African Journal of Biotechnology 10*(78), 18005–18012.

Tekere, M., Prinsloo, A., Olivier, J., Jonker, N., Venter, S. (2012). An evaluation of the bacterial diversity at Tshipise, Mphephu and Sagole hot water springs, Limpopo Province, South Africa. *African Journal of Microbiology Research 6*(23), 4993–5004.

Van Heerden, J., M. G. Mills, M. J. Van Vuuren, P. J. Kelly, and M. J. Dreyer (1995). An investigation into the health status and diseases of wild dogs (Lycaon pictus) in the Kruger National Park. *Journal of the South African Veterinary Association 66*(1), 18–27.

Van Wilgen, B. W., H. C. Biggs, S. P. O'Regan, and N. Mare (2000). A fire history of the savanna ecosystems in the Kruger National Park, South Africa, between 1941 and 1996.

Van Wyk, L. (2011). Review of the ecological implications of artificial waterhole closures in the Kruger National Park and the effect thereof on tourism. pp. 1–80.

Venter, F. J., R. J. Naiman, H. C. Biggs, and D. J. Pienaar (2008). The evolution of conservation management philosophy: Science, environmental change and social adjustments in Kruger National Park. *Ecosystems 11*(2), 173–192.

Zeglin, L. H. (2015). Stream microbial diversity in response to environmental changes: Review and synthesis of existing research. *Frontiers in Microbiology 6*(MAY), 1–15.

Figures & Tables

Section	Site	Type	Geology
Tshokwane (TSH)	Nhlanguleni (NHL)	Borehole	Granite
Skukuza (SKZ)	Nwaswitshaka (NWA)	Borehole	Granite
Skukuza (SKZ)	De LaPorte (DLP)	Borehole	Granite
Skukuza (SKZ)	Kwaggas Pan (KWA)	Borehole	Granite
Satara (SAT)	Girivana (GIR)	Borehole	Granite
SaTara (SAT)	Witpens (WIT)	Pipeline trough	Basalt
Kingfisherpruit (KFI)	Imbali (IMB)	Borehole	Granite
Kingfisherpruit (KFI)	Hoyo Hoyo (HOY)	Borehole	Granite
Houtboschrand (HOU)	Nyamarhi (NYA)	Pipeline trough	Basalt
Houtboschrand (HOU)	Ngosto North (NGO)	Pipeline trough	Basalt

Table 1: Sample locations

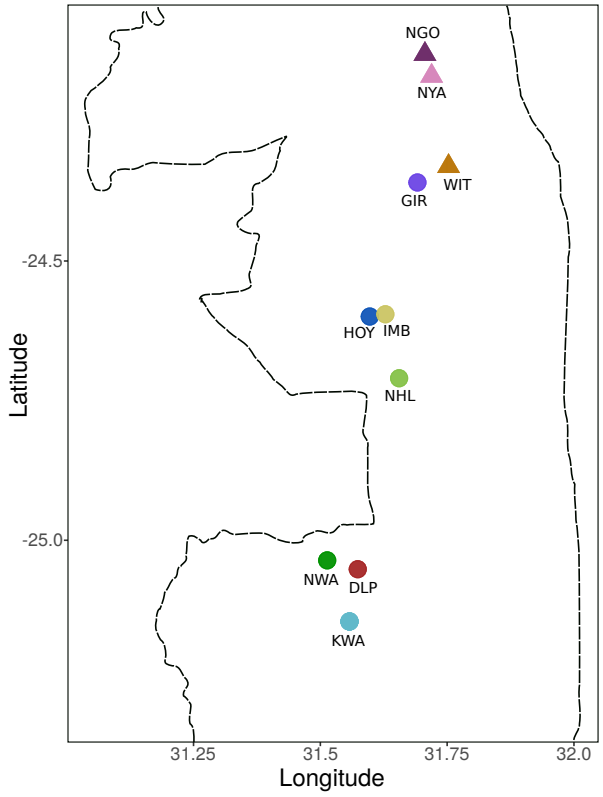


Figure 1: Map of site locations with park boundary indicated by dashed line. Circles represent sites filled by boreholes while triangles represent sites filled by river water via pipeline troughs.

Site	Weeks	S	XS	Daily	A/B	Total
Nhlanguleni (NHL)	3	0	0	0	Yes	6
Nwaswitshaka (NWA)	3	1	1	4	Yes	18
De LaPorte (DLP)	1	1	1	0	Yes	6
Kwaggas Pan (KWA)	2	1	1	0	Yes	8
Girivana (GIR)	3	0	0	0	Yes	6
Witpens (WIT)	3	0	0	0	Yes	6
Imbali (IMB)	3	0	0	0	Yes	6
Hoyo Hoyo (HOY)	3	1	1	0	Yes	10
Nyamarhi (NYA)	3	1	1	0	Yes	10
Ngosto North (NGO)	3	1	1	0	Yes	10
BLANK	2	0	0	0	No	2
	29	6	6	4		88

Table 2: Samples sequences, broken down by number of weekly samples, number of site-times for which S (50 mL) and XS (15 mL) samples were filtered, additional daily samples taken, whether A/B samples were taken, and the resulting total number of samples sequenced per site.

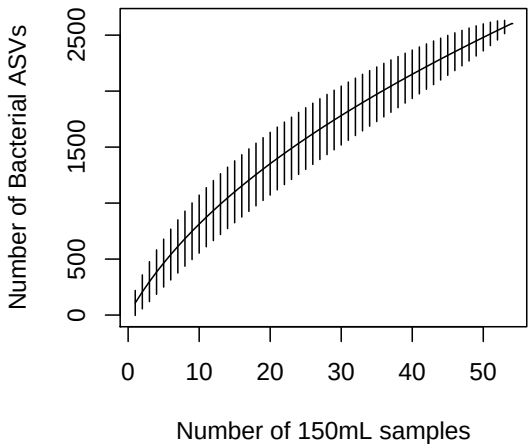


Figure 2: ASV accumulation curve of bacterial ASV richness using the “exact” method. Bars represent two standard deviations around mean estimates.

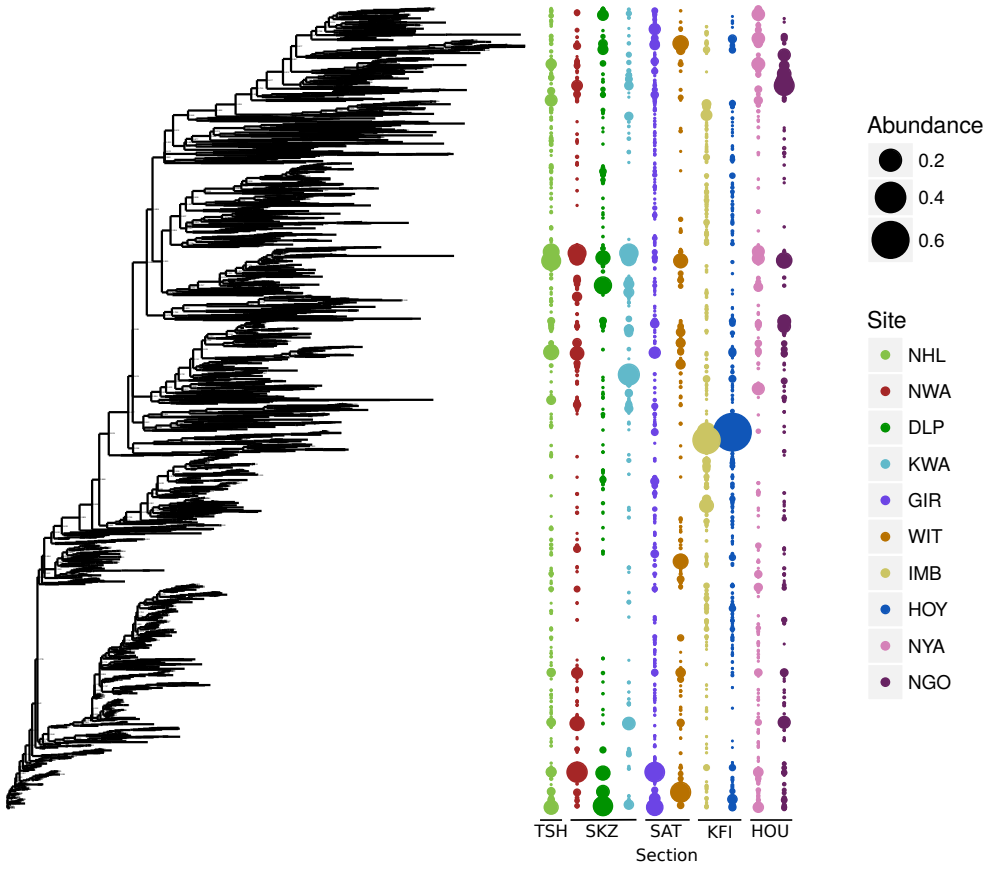


Figure 3: Phylogenetic tree of 16S ASV sequences in the core samples, paired with their relative abundances at each site. Sites are ordered by section.

Diversity	Level	Shannon	SES	2.5%	97.5%	Pr(sim.)
α	A/B samples	2.93	-9335.9	5.10	5.10	0.01
α	Temporal samples	2.99	-12535.9	5.13	5.13	0.01
α	Sites	3.39	-16873.6	5.15	5.15	0.01
α	Sections	3.94	-14919.9	5.16	5.16	0.01
γ	(Total)	5.17	0.0	5.17	5.17	1.00
β	A/B samples	0.05	153.3	0.027	0.027	0.01
β	Temporal samples	0.40	2688.2	0.022	0.022	0.01
β	Sites	0.55	6528.7	0.008	0.008	0.01
β	Sections	1.23	14919.9	0.008	0.008	0.01

Table 3: Additive partitioning of Shannon diversity into α, β, and γ diversities across sections, sites, temporal samples, and within site-time samples (A/B) as components of the total diversity observed across all core samples. Observed diversity is compared to 99 simulations and the standardized effect size (SES) using "r2dtable" null model with the *adipart* function in the R package *vegan*.

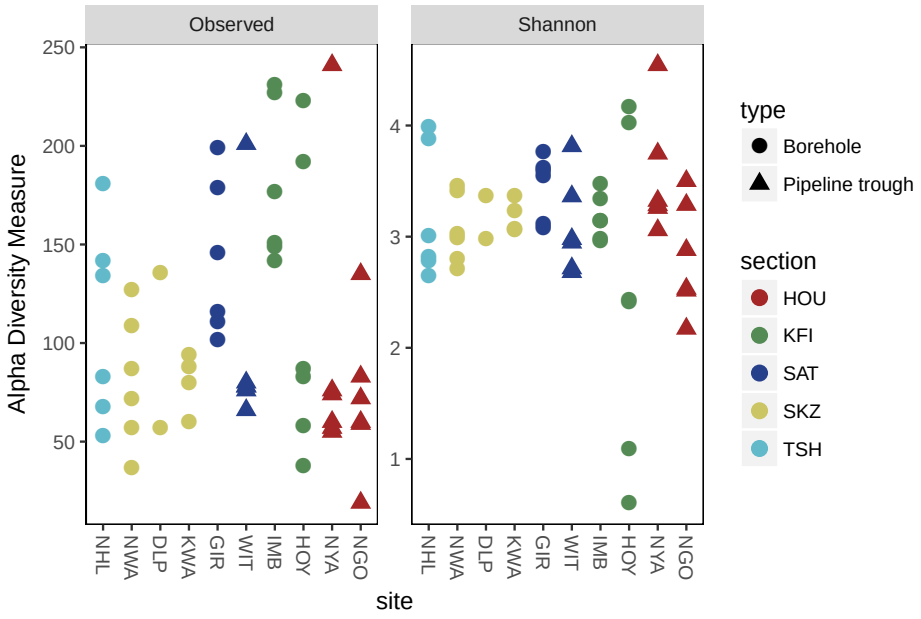


Figure 4: Plots of observed ASV richness and Shannon diversity across samples. Samples are grouped and coloured by park section and with shape indicating waterhole type.

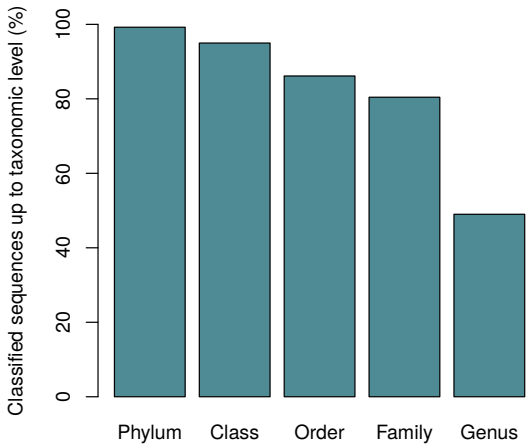


Figure 5: The proportion of ASVs assigned a given taxonomic level using the SILVA database v128 and a pre-trained RDP classifier with minimum 50% bootstrap support.

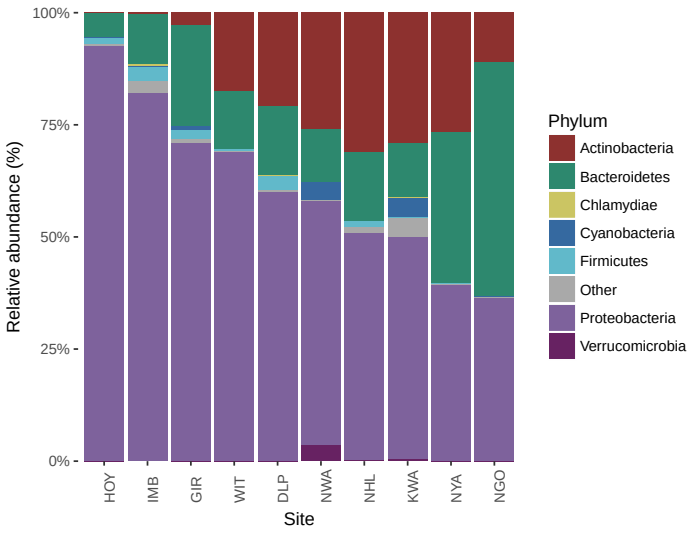


Figure 6: Relative abundances of bacterial phyla across sites. Sites are ordered by relative abundance of phylum Proteobacteria.

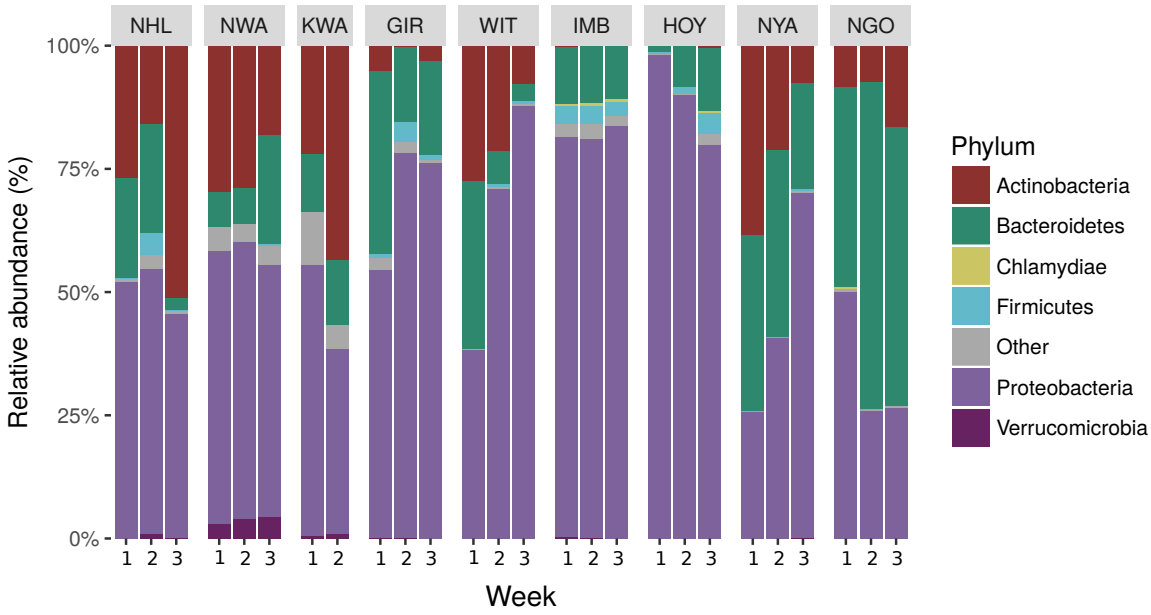


Figure 7: Relative abundances of bacterial phyla across weekly samples.

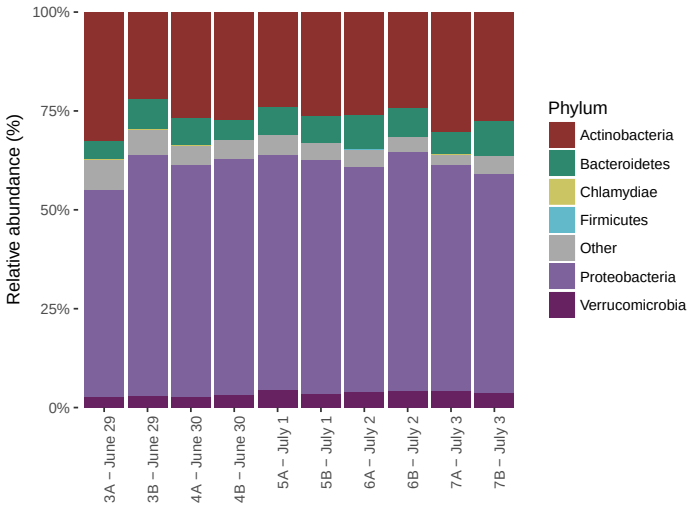


Figure 8: Relative abundances of bacterial phyla across five days at a single site (NWA).

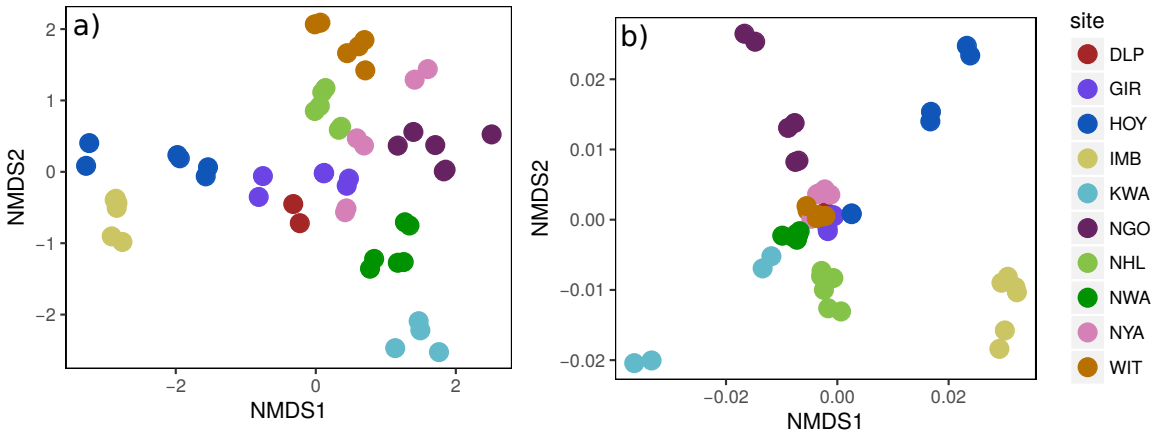
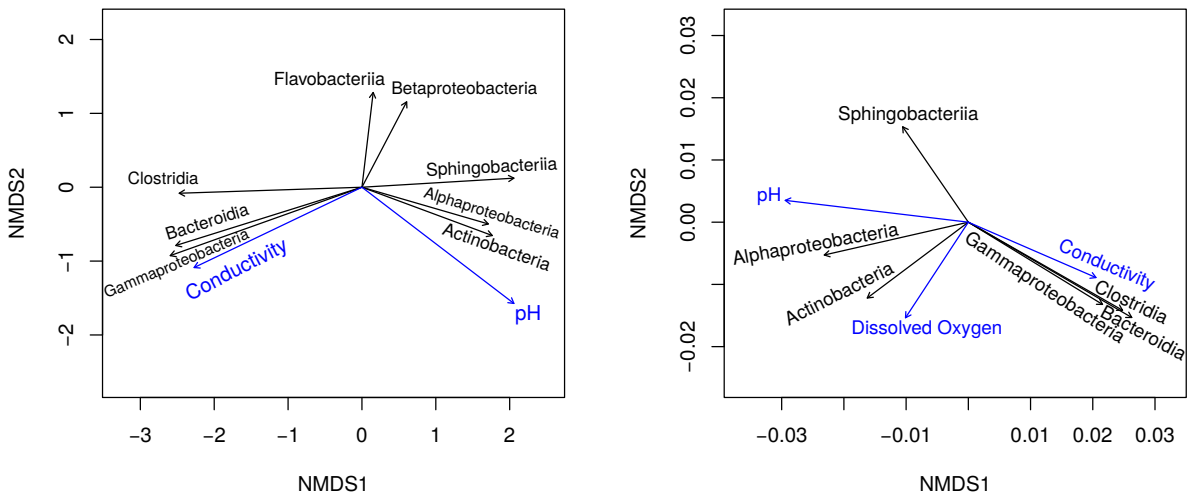


Figure 9: NMDS plots of a) Bray-Curtis and b) abundance-weighted UniFrac distances. Colours represent site.



(a) Bray-Curtis NMDS

(b) Weighted UniFrac NMDS

Figure 10: Nonmetric multidimensional scaling (NMDS) ordination of variation in bacterial community structure across 54 samples based on a) Bray-Curtis and b) abundance-weighted UniFrac distances. Arrows indicate the direction of significant ($p < 0.05$) correlations among variables and the NMDS axes, with arrow length indicating the strength of the correlation. Blue arrows indicate environmental variables, while black arrows indicate relative abundances of sequences from different microbial classes. The ordination axes explain 96.8% (a) and 98.1% (b) of the variance in the dissimilarities (Fig. S17).