Shallow shotgun sequencing of the microbiome recapitulates 16S amplicon results and provides functional insights

Mason Stothart[1], Philip McLoughlin[2], and Jocelyn Poissant[1]

[1]University of Calgary
[2]University of Saskatchewan

April 8, 2022

Abstract

Prevailing 16S rRNA gene-amplicon methods for characterizing the bacterial microbiome are economical, but result in coarse taxonomic classifications, are subject to primer and 16S copy number biases, and do not allow for direct estimation of microbiome functional potential. While deep shotgun metagenomic sequencing can overcome many of these limitations, it is prohibitively expensive for large sample sets. We evaluated the ability of shallow shotgun metagenomic sequencing to characterize taxonomic and functional patterns in the fecal microbiome of a model population of feral horses (Sable Island, Canada). Since 2007, this unmanaged population has been the subject of an individual-based, long-term ecological study. Using deep shotgun metagenomic sequencing, we determined the sequencing depth required to accurately characterize the horse microbiome. In comparing conventional versus high-throughput shotgun metagenomic library preparation techniques, we validate the use of more cost-effective lab methods. Finally, we characterize similarities between 16S amplicon and shallow shotgun characterization of the microbiome, and demonstrate that the latter recapitulates biological patterns first described in a published amplicon dataset. Unlike amplicon data, we demonstrate how shallow shotgun metagenomic data also provided useful insights about microbiome functional potential which support previously hypothesized diet effects in this study system.

Shallow shotgun sequencing of the microbiome recapitulates 16S amplicon results and provides functional insights

Mason R. Stothart[1], Philip D. McLoughlin[2], Jocelyn Poissant[1]

[1]Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada

[2]Department of Biology, University of Saskatchewan, Saskatoon, SK, Canada

ORCID

Mason R. Stothart https://orcid.org/0000-0002-2863-908X

Jocelyn Poissant https://orcid.org/0000-0003-1546-3609

Running title: Evaluating shallow shotgun metagenomics

Corresponding Author:

Mason R. Stothart

Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada

Email: mason.stothart1@ucalgary.ca

Abstract

Prevailing 16S rRNA gene-amplicon methods for characterizing the bacterial microbiome of wildlife are economical, but result in coarse taxonomic classifications, are subject to primer and 16S copy number biases, and do not allow for direct estimation of microbiome functional potential. While deep shotgun metagenomic sequencing can overcome many of these limitations, it is prohibitively expensive for large sample sets. We evaluated the ability of shallow shotgun metagenomic sequencing to characterize taxonomic and functional patterns in the fecal microbiome of a model population of feral horses (Sable Island, Canada). Since 2007, this unmanaged population has been the subject of an individual-based, long-term ecological study. Using deep shotgun metagenomic sequencing, we determined the sequencing depth required to accurately characterize the horse microbiome. In comparing conventional versus high-throughput shotgun metagenomic library preparation techniques, we validate the use of more cost-effective lab methods. Finally, we characterize similarities between 16S amplicon and shallow shotgun characterization of the microbiome, and demonstrate that the latter recapitulates biological patterns first described in a published amplicon dataset. Unlike amplicon data, we further demonstrate how shallow shotgun metagenomic data provide useful insights about microbiome functional potential which support previously hypothesized diet effects in this study system.

Keywords: diet effects, metagenomic sequencing, horse, wildlife

Introduction

Our understanding of host-associated microbiomes has relied heavily upon 16S rRNA gene amplicon sequencing of bacterial communities . Amplicon-based profiling of the microbiome is affordable and supported by a suite of accessible bioinformatic and analytical tools. But despite their popularity, amplicon sequencing data: 1) are biased by 16S rRNA gene copy number variation, 2) often only resolve to the level of genus and 3) do not provide direct estimates of microbiome functional potential . Hence, our ability to interpret the causes and consequences of microbiome variation using 16S data is too often relegated to the realm of speculation based on coarse taxonomic profiles.

In response to the limitations inherent in 16S approaches, there is a growing interest in shotgun metagenomic sequencing. By randomly sequencing the entire microbial genomic content of a sample, shotgun sequencing can reconstruct microbial communities at finer resolutions than amplicon sequencing, and provide direct estimates of microbiome functional potential from microbial gene contents . However, the library preparations and deep sequencing (10-100 million of read-pairs per sample) required to generate shotgun metagenomic data can be an order of magnitude more expensive than amplicon sequencing on a per sample basis . These expenses make shotgun metagenomic sequencing infeasible for large sample-sets .

Recent reductions in the per-base-cost of DNA sequencing and the development of in-house and commercially available high-throughput library preparation techniques have made shotgun sequencing more affordable. Cost reductions notwithstanding, deep sequencing can still be prohibitively expensive for large sample-sets. However, depending on the research questions of interest, the characterization of major patterns in microbiota communities and functional profiles may not require deep sequencing.

Shotgun sequencing at lower depths than is conventional (shallow shotgun sequencing) has been proposed as a cost-effective method for characterizing microbial communities . Although not a synonymous substitution for deep shotgun sequencing, shallow shotgun sequencing can outcompete amplicon-based community characterization at comparable costs, and with the additional benefit of capturing major variation in microbial gene content. In an early test of shallow shotgun sequencing capabilities in humans (retroarticular creases, stool, sub/supragingival plaque, and tongue dorsum microbiomes), accurate species-level differential abundances were observed among taxa which occurred within samples at percent abundances as low as 0.05% of reads, in datasets rarefied to 0.5 million reads/sample. Furthermore, at depths of only 1000 reads/sample, coarse biological patterns in alpha and beta diversity were still evident

Despite its promise, shallow shotgun sequencing has limitations. For instance, a reliance on read-based profilers means that the same low sequencing depth which makes shallow shotgun sequencing economical,

renders *de novo* assembly approaches ineffective. Therefore, shallow shotgun sequencing data cannot be used for novel gene discovery, identification of rare taxa, or to create metagenome assembled genomes (MAGs) using single sample assemblies. An inability to use *de novo* assembly approaches means that shallow shotgun reads can only be classified if they match references within genome databases. Publicly available microbial genomes are heavily biased towards isolates or MAGs from humans, and lab or production animals Therefore, the utility of shallow shotgun sequencing needs to be assessed for host-associated microbial communities which are likely to be underrepresented within genome databases.

Here, we evaluate the ability of shallow shotgun sequencing to characterize the taxonomic composition and functional potential of the fecal microbiome of a free-ranging horse (*Equus ferus caballus*) population living on Sable Island, Nova Scotia, Canada. Although horses have been the subject of many 16S rRNA gene amplicon studies —including Sable Island horses —they have not benefitted from deep shotgun metagenomic studies. As a prevalent domesticated mammal, the major bacterial clades observed in horses might be similar to those observed in other domesticated species, which also originate from a human agricultural environment, but which have been the subject of deeply sequenced metagenomic studies, for example: cows , pigs , sheep , and chickens . Therefore, bacterial species unique to Sable Island horses are likely to have close relatives in available microbial genome reference databases, and so may be suitable for shallow shotgun sequencing.

First, to determine the depth at which shallow shotgun sequencing remains viable, we analyzed a successively rarefied deeply sequenced dataset of 16 fecal microbiome samples. Second, to validate the efficacy of more affordable library preparation methods, we compared sequencing results generated using prevailing library preparation methods (Illumina Nextera XT), to those created using a new high-throughput technique (iGenomx Riptide, now Twist Biosciences Riptide). Third, we compared shallow shotgun sequencing data to 16S rRNA gene amplicon sequencing of the same DNA extracts, to quantify the concordance between amplicon and shallow shotgun metagenomic based estimates of microbiota community structure. Fourth, using an expanded 83-sample dataset, we also tested whether biological patterns in the microbiome—which were first observed in a 16S rRNA gene amplicon dataset (e.g., diet effects and spatial structuring—could be replicated via shallow shotgun sequencing of the same samples. Fifth, we re-analyze this 83-sample dataset using profiles of microbiome functional potential derived from shallow shotgun sequencing, to evaluate the purported advantage of a shallow shotgun sequencing approach.

Methods

Study System

Sable Island is an oceanic sandbar, located 175 km off the eastern coast of Nova Scotia, Canada (Figure 1). Since the 18$^{\text{th}}$century, the island has been home to a population of free-living horses, which have remained largely unmanaged since their introduction. Importantly, no individual of the population has ever been administered antibiotics . Using 16S rRNA gene amplicon sequencing, we previously demonstrated evidence for dispersal limitation in horse-associated gut microbiota across the island's narrow (49 km) length We also characterized spatially structured variation in the microbiome, which we hypothesized is caused by dietary variation.

Marram grass (*Ammophila breviligulata*) comprises the most abundant component of the Sable Island horse diet across the island. However, the semi-succulent plant sea sandwort (*Honckenya peploides*), only found at the eastern and western extremes of Sable Island is preferred by the horses when present in their home-range. To generate coarse predictions of diet, we previously estimated the relative coverage of major Sable Island plant community types within 150-metre radius buffers centred on the point of fecal sample collection (Stothart et al., 2021). The radius of this spatial buffer is proportional to the median daily movement observed in the horses. Therefore, plant community composition within these spatial buffers are predicted to be coarsely indicative of the forage to which horses had access, during the 24 hours preceding defecation. Using these estimates, we observed greater diversity of amplicon sequence variants (ASVs) and greater phylogenetic clustering in the microbiome of horses without sea sandwort within their spatial buffer . Conversely, lower bacterial diversity and greater evidence for stochasticity was observed in the microbiome

of horses with access to sea sandwort. We hypothesized that sandwort did not support fibrolytic niches in the gut microbiome to the same extent as marram grass. However, we were unable to test this hypothesis, as 16S amplicon data do not allow for direct estimates of microbiome functional potential.

Sample Collection

Individual-linked fecal samples were collected in 2014 from adult female Sable Island horses, during an annual survey of the population. All samples were collected within 5 min of defecation and kept on ice in the field (maximum 6 hrs), before being homogenized, subset into 2.0-ml screw cap cryotubes, and frozen at -20 ^{o}C in the field for up to two months. Samples were stored at -80 oC for long-term storage, after transport to the mainland. Alongside the collection of fecal samples, data on horse location, social membership, and parental status were recorded. The study system and sampling protocol are described in greater detail elsewhere .

Sequencing

Deep Shotgun Metagenomic Dataset

We weighed 0.2-g portions of 16 freshly thawed fecal samples into 2-ml bead beating tubes containing Qiagen 0.7-mm garnet beads, added 750 µl of PowerBead Solution and 60 µl of C1 buffer from Qiagen's QIAamp PowerFecal DNA Kits to each tube, vortexed each tube for 2 seconds, and incubated tubes for 10 min at 60^{o}C. After incubation, we bead beat samples for 10 min on a Vortex-Genie 2 fitted with Qiagen's Vortex Adapter (cat. no 13000-V1-24). All other extraction steps were completed as per Qiagen kit recommendations (single-tube extraction protocol). Final DNA extracts were eluted in 100 µl of molecular grade water.

DNA extracts were quantified using Qubit dsDNA Broad Range Assays. Extracts were sent to the University of Calgary Centre for Health Genomics and Informatics, where shotgun metagenomic libraries were created using v1 Illumina Nextera XT Index Kit preparations, as per manufacturer specifications. Normalized libraries were sequenced on an Illumina NextSeq 500 High Output Flow Cell to a target depth of 30 million 2 x 150 bp reads per sample.

Shallow Shotgun Metagenomic Dataset

We weighed 0.16-g portions of freshly thawed fecal samples into 2-ml bead beating tubes containing Qiagen 0.7-mm dry garnet beads. Next, we added 680 µl of lysis buffer (pre-warmed to 60 ^{o}C) from Qiagen's QIAamp 96 PowerFecal QIAcube HT Kit to each tube, and bead beat samples for 10 min on a Vortex-Genie 2 fitted with Qiagen's Vortex Adapter (cat. no 13000-V1-24). All other extraction steps were completed as per Qiagen kit QIAcube HT recommendations (using a single-tube start protocol). Final DNA extracts were eluted in 100 µl of molecular grade water. A total of 2820 samples were extracted using this protocol, as part of a larger study. For this validation, we focus only an 83-sample subset of this larger sample-set. This 83-sample subset represents DNA extracts from technical replicates of fecal samples, which correspond to the 83 fecal samples included in an amplicon dataset (described below).

DNA extracts were quantified using Qubit dsDNA Broad Range Assays. Extracts were sent to the University of Calgary Centre for Health Genomics and Informatics, where shotgun metagenomic libraries were created using iGenomx Riptide preparations (1:1 mixture of low GC and High GC primers), as per manufacturer specifications. Unlike the Nextera XT protocol, the iGenomx Riptide protocol allows for multiplexing of up to 960 samples on a single Illumina sequencing lane. For this pilot study, we sequenced a total of 188 samples using an Illumina NovaSeq (300 cycle SP sequencing kit v1.5 2 lanes) to a target depth of 4.1 million read pairs. Two positive controls (ZymoBIOMICS Microbial Community Standard II, Log Distribution) and two negative controls (one pair per plate of extracted DNA) were sequenced alongside the biological samples, and provide no evidence of kit contamination (see supporting information). In addition to DNA from 83 samples extracted using the QIAamp 96 PowerFecal protocol described, we also used iGenomx Riptide protocols to prepare and sequence 13 of the same DNA extracts used to create the Nextera XT deeply sequenced dataset (see above). This allowed us to more directly compare the results generated using different library preparation techniques, by eliminating methodological and technical variation related to DNA extraction.

Three samples from the Nextera XT prepared dataset were excluded from the iGenomx Riptide preparation and sequencing, because the original DNA extract was depleted.

16S rRNA Gene Amplicon Dataset

This dataset is comprised of 83 samples, which have previously been described, analyzed, and archived on the NCBI SRA . The DNA extraction protocol for these samples is identical to the one used to create the Nextera XT prepared deep shotgun metagenomic dataset (described above). As above, we quantified DNA using Qubit dsDNA Broad Range Assays. Next, we standardized DNA concentration to 20 ng/µl and PCR amplified the v3–v4 region of the 16S rRNA gene using the 341f forward and 805r reverse universal primers. PCR products were sequenced on an Illumina MiSeq platform (v3 chemistry: 2 × 300 base-pair reads) by the University of Calgary Centre for Health Genomics and Informatics, to a target depth of >120,000 read pairs per sample.

Bioinformatics

All shotgun metagenomic reads underwent adapter trimming, quality control, and were filtered against the EquCab3 domestic horse reference genome , using default parameters in kneaddata , a wrapper function for trimmomatic and bowtie2 . Read pairs which passed quality filtering were used to estimate microbial taxon relative abundances, via default implementation of Kaiju . Unlike nucleotide-based profilers, Kaiju assigns taxon identities to reads by first translating sequences into all six reading frames, and mapping the resultant amino acid sequences to a protein reference database. Protein coding regions are more strongly conserved than non-coding regions, and amino acid sequences will not be affected by synonymous substitutions of nucleotide bases. This can allow for greater rates of classification for shotgun metagenomic reads which come from microbiota which are underrepresented in reference databases. Using the microbial subset of the NCBI BLAST non-redundant protein database, we first classified reads to 'species', a delineation which in actuality encompasses species, strains and co-abundant gene groups. Reads which could not be assigned to species were assigned to progressively coarser taxonomic levels (genus, family, order, class, phylum, kingdom). For example, reads which could be classified to family, but not genus or species would form a '*familyX unclassified* ' bin in our analyses.

To estimate microbial gene contents and metabolic potential, we used HUMAnN3 . HUMAnN3 maps quality-controlled and filtered shotgun metagenomic reads to UniProt Reference Clusters (UniRef; . In running HUMAnN3, we concatenated forward and reverse read files, by-passed the taxonomic classifier option, reduced subject sequence coverage thresholds to 0, classified reads to UniRef50 gene families, and mapped gene families to MetaCyc reactions and metabolic pathways . The methods used for quality-control filtering and taxonomic classification of 16S rRNA amplicon sequence data using DADA2 are described elsewhere .

Statistical Analysis

Comparisons of deep metagenomic, shallow metagenomic, and 16S amplicon sequencing

Unless specified, only reads identified as bacterial were retained for taxonomic analyses, and all reports of realized sampling depths hereafter refer to the fraction of reads classified as bacteria. All datasets were normalized prior to analysis via rarefaction, except for datasets used in ANCOM-BC differential abundance testing. A schematic of our study design can be found in the supporting information (Figure S1).

To determine the requisite shotgun metagenomic sequencing depth required to accurately characterize the Sable Island horse microbiome, we first analyzed a deeply sequenced 16 sample Nextera XT prepared library. Successively rarefied versions of this dataset were benchmarked against a minimally rarefied dataset (9.3 million read-pairs). Similarly, to determine the affect of sequencing depth on functional profile reconstruction, we compared successively rarefied datasets of MetaCyc reaction and pathway abundance tables.

Next, to identify discrepancies between prevailing and newer high-throughput library preparation methods, we compared the results of shotgun metagenomics libraries prepared from the same DNA extracts, using Nextera XT and iGenomx Riptide methods. For this comparison, datasets were rarefied to the lowest sequencing

depth observed amongst these 13 paired sample datasets, 1.6 million read pairs.

Finally, we identified differences between metagenomic and amplicon-based characterization of the bacterial microbiome using the same DNA extracts from 13 samples. As above, these datasets were rarefied to the same minimal sequencing depth, in this case 35,000 read pairs. Comparisons of taxon relative abundances between shotgun metagenomic and amplicon datasets occurred primarily at the level of family, since this was the finest resolution to which most 16S rRNA amplicon reads could be classified. Analyses across these methodological comparisons were comprised of general linear models to quantify correlations in alpha diversity (Shannon diversity indices) and taxon relative abundances between dataset types. We also tested for correlations between beta-diversity estimates using mantel tests of Bray-Curtis dissimilarity matrices.

Re-examination of reported biological patterns through a shallow shotgun metagenomic lens

In the second half of our analyses, we sought to determine whether shallow shotgun sequencing data (rarefied to 0.38 million bacterial read pairs) could replicate the biological patterns reported in a recently published 16S amplicon dataset We used mantel tests to test for correlations between spatial and Bray-Curtis dissimilarity matrices, PERMANOVAs to test for the effect of environmental variables on microbiome beta-diversity, and ANCOM-BC tests to identify taxa which differ in abundance between horses with access to sandwort versus horses without access to sandwort (with false discovery rate adjustments and a conservative variance estimate; To allow for comparisons between 16S amplicon and shotgun metagenomic results, ANCOM-BC tests for taxon differential abundance were performed on datasets binned to bacterial family. However, to prevent large imbalances in false discovery rate penalties between datasets, we further filtered the shallow shotgun dataset to only families which were present at a percent abundance of 0.1% in at least 1 sample. This resulted in the retention of 144 families, which was comparable to the 147 families observed in the 16S amplicon dataset. Notably, although Kaiju accurately estimates abundance weighted patterns in the microbiome , it overestimates richness and can result in a large number of spurious, low abundance taxa. Abundance filtering of count tables is therefore prudent. Similar analyses as described above were used to analyze differential abundances in MetaCyc reactions and metabolic pathways. For all functional analyses, reads which could not be classified or those which could not be integrated into pathways (in the case of pathway abundance estimates) were removed. Datasets were normalized via rarefaction for all analyses other than ANCOM-BC tests, which possess a built-in normalization process . Unless otherwise stated, all analyses were completed in R (v. 4.1.2) using the package phyloseq v. 1.38.0 and vegan v. 2.5–7

Results

On average, 87.2% ± 3.9% *SD* of shotgun metagenomic read pairs passed quality control. Of these, an average of 0.16% ± 0.8% *SD*of reads were removed after being filtered against the EquCab3 reference genome. While some samples had large amounts of host contamination (max = 27%), 95.5% of samples were found to contain < 2% horse DNA. However, investigators should be cognisant that other hosts have been shown to contain higher proportions of host DNA . Of the reads that passed quality control and filtering, 54.3% ± 4.1% *SD* were identified as bacterial, 0.5% ± 0.2% *SD* as archaeal, and 0.6% ± 0.1% *SD* as microbial eukaryotic. The unclassified fraction of reads might derive from: (1) microbiota not present in the reference database, (2) non-coding regions of microbial genomes, (3) parasitic nematodes in the intestinal tract, (4) other contaminant DNA, or (5) DNA derived from dietary sources.

Prior to the taxonomic analyses reported below, we removed reads which were not classified as bacteria. Therefore, when seeking to determine requisite sequencing depths, we caution investigators to account for sequence losses due to low-quality reads, or those derived from host, dietary, or microbial non-coding regions (if using a translated search method). Investigators should also be prepared for uneven sequencing depths when multiplexing large numbers of samples. Our sequencing of a 188-sample pilot dataset on an Illumina NovaSeq resulted in a median depth of 4,105,972 read pairs, which is remarkably close to our target depth of 4.1 million read pairs. However, observed sequencing depths were also highly variable (*SD* ± 1.9 million read pairs; Figure S2). Therefore, like , we would advise researchers to sequence more deeply than their identified minimal sequencing depth, to minimize sample drop-out.

Comparison of sequencing depth

To identify the sequencing depth required to accurately characterize fecal microbiomes in Sable Island horses, we successively rarefied a 16-sample deeply sequenced shotgun metagenomic dataset. Patterns in Shannon diversity remained indistinguishable from the minimally rarefied dataset down to depths of 0.4 million read pairs (Figure 2A), and a statistically significant decrease in Shannon diversity was only observed at depths below 40,000 read pairs per sample (t = -1.959, p = 0.05). However, despite a decrease in the accuracy of Shannon diversity estimates, Shannon diversity remained highly correlated between the full dataset and data rarefied to only 2,000 read pairs per sample (R^2 = 0.87, t = 10.24,$p < 0.01$). Similar patterns were observed with respect to microbiome beta-diversity. Bray-Curtis distances in the minimally rarefied dataset were highly correlated with those in more shallowly rarefied datasets (Figure 1B). For example, Pearson correlation coefficients remained above $r = 0.99$ (mantel tests, $p< 0.01$), at depths as low as 0.2 million read pairs per sample (Figure 2B). Even at 2,000 read pairs per sample, a strong correlation was observed with the full sequence dataset (mantel test: r = 0.81,$p < 0.01$). Similar results were observed with respect to MetaCyc reaction and pathway profiles (Figure S3).

Comparison of library preparation methods

Next, using a 13-sample dataset, we compared sequencing results generated using a prevailing library preparation technique (Illumina Nextera XT) with those created using a new commercially available method which allows for high-throughput preparation at a lower per-sample-price (iGenomx Riptide). Estimates of Shannon diversity between the Nextera XT and iGenomx Riptide generated datasets (both rarefied to 1.6 million bacterial read pairs per sample) were highly correlated (R^2 = 0.98, t = 22.988, $p< 0.01$). Shannon diversity estimates tended to be lower among the sequenced iGenomx Riptide libraries, but this effect was non-significant (t = -1.953 p = 0.08). Bray-Curtis estimates of beta-diversity were also correlated between Nextera XT and iGenomx Riptide created datasets (mantel: r = 0.86, $p< 0.01$; Figure 3A), as were log-transformed species-level estimates of taxa relative abundances (R^2 = 0.97, t = 1121.42, $p < 0.01$; Figure 3B).

Comparisons of 16S amplicon versus shallow shotgun metagenomic sequencing

We next sought to determine whether 16S amplicon and shallow shotgun (iGenomx Riptide prepared) sequencing approaches would yield similar descriptions of 13 fecal microbiome samples of Sable Island horses. Shannon index estimates of ASV (amplicon) and species (shotgun metagenomic) diversity derived from these sequencing methods were correlated ($R^2 = 0.46$, $t = 3.375$,$p < 0.01$), but shotgun metagenomic estimates of Shannon diversity (6.78 ± 0.34 SE) were consistently higher than amplicon-based estimates (5.78 ± 0.19 SE), even after rarefying datasets to the same sequencing depth (35,000 bacterial read-pairs). Despite correlations in alpha diversity, ASV-level Bray-Curtis distances calculated from amplicon data were not correlated with species-level Bray-Curtis distances in the shotgun metagenomic dataset (mantel test: r = 0.06, p = 0.38). However, the correlation between beta-diversity matrices increased as reads were grouped into successively coarser taxonomic bins: genus (mantel test: r = 0.46, p < 0.01), family (mantel test: r = 0.60, $p < 0.01$), order (mantel test: r = 0.81, $p < 0.01$), class (mantel test: r = 0.80, $p < 0.01$), and phylum (mantel test: r = 0.80, $p < 0.01$).

To characterize taxa which may be overrepresented or underrepresented in amplicon versus shotgun metagenomic sequence datasets, we compared bacterial family average relative abundance estimates. A total of 73% of reads could be assigned to one of 58 families in the amplicon dataset, while 70% of reads could be assigned to one of 277 families in the shotgun metagenomic dataset. Of these, 53 were observed in both datasets. Amplicon and shotgun metagenomic-based estimates of average (log-transformed) relative abundance were positively correlated (R^2 = 0.33, t = 5.174, $p< 0.01$; Figure 4A). Highly abundant families tended to be overrepresented within the amplicon dataset compared to the shotgun dataset, with the exception of *Fibrobacteraceae* . Conversely, many families which were present at moderate relative abundances in the shotgun metagenomic dataset, were either absent or present at lower relative abundances in the amplicon dataset. However, despite the departure of Families from a 1:1 relationship, amplicon and shotgun based estimate of

relative abundance were strongly correlated among families previously identified as dominant in the Sable Island horse microbiome: *Ruminococcaceae* ($R^2 = 0.49$,$t = 3.535$, $p < 0.01$), *Lachnospiraceae*($R^2 = 0.55$, $t = 3.954$, $p< 0.01$), *Prevotellaceae* ($R^2 = 0.73$, $t = 5.753$, $p < 0.01$),*Fibrobacteraceae* ($R^2 = 0.91$, $t = 10.806$, $p < 0.01$), *Spirochaetaceae*($R^2 = 0.84$, $t = 7.863$, $p< 0.01$), and *Rikenellaceae* (R^2= 0.76, $t = 6.314$, $p < 0.01$; Figure 4B).

Re-examination of a published sample-set using shallow shotgun metagenomics

After establishing that 16S amplicon and shallow shotgun metagenomic sequencing yielded similar estimates of taxonomic patterns, we sought to determine whether a shallow shotgun sequencing approach could recapitulate the findings reported in a recently published 16S rRNA amplicon sequence dataset .

A primary finding of this paper was an apparent relationship between microbiome ASV-level beta-diversity and the spatial separation of sample collection locations. Spatial patterns in the microbiome were still present in an 83-sample subset of the previously published amplicon dataset (mantel test: $r = 0.51$, $p < 0.01$; Figure 5), even after controlling for *sea sandwort availability* (partial mantel test: $r = 0.25$, $p < 0.01$), a heterogeneously distributed dietary component shown to correlate with microbiome variation. A much weaker spatial effect was observed in the shallow shotgun metagenomic dataset (mantel test: $r = 0.19$,$p < 0.01$), and this effect was only marginally significant after controlling for *sea sandwort availability*(partial mantel test: $r = 0.12$, $p = 0.04$). Despite an absence of correlation in ASV/species level beta-diversity in a 13-sample set of re-sequenced DNA extracts (see above), we did observe a modest correlation in amplicon and shotgun metagenomic estimated Bray-Curtis dissimilarity in this expanded 83-sample dataset (mantel test: $r = 0.44$, $p < 0.01$). However, dissimilarity estimates in the amplicon dataset were consistently higher than those in the shallow shotgun dataset. This could be caused by a zero-inflated ASV count-table in the amplicon dataset. This pattern is common in 16S rRNA gene amplicon studies , and was evidenced by an apparent bi-modal distribution in ASV prevalence (Figure S4).

In addition to spatial effects, we previously demonstrated that horse access to sandwort appeared to affect the Sable Island horse bacterial microbiome . This pattern was still evident in a reduced 83-sample subset of these data ($\text{PERMANOVA}_{\text{Bray-Curtis}}$: $F = 5.300$, $R^2 = 0.06$, $p < 0.01$), and we observed an effect of comparable magnitude within the shallow shotgun metagenomic dataset ($\text{PERMANOVA}_{\text{Bray-Curtis}}$:$F = 6.146$, $R^2 = 0.07$, $p< 0.01$).

Analysis of 16S rRNA amplicon sequence data using ANCOM-BC tests indicated that *Fibrobacteraceae* , *Spirochaetaceae* , and*Pasteurellaceae* were enriched in horses without access to sandwort, while horses with access to sandwort tended to possess greater abundances of *Christensenellaceae* , *Anaeroplasmataceae* ,*Mycoplasmataceae* , *Peptococcaceae* , *Burkholderiaceae* , and *Puniceicoccaceae* (Table S1A). Within the shallow shotgun dataset, *Fibrobacteraceae* and *Nanosyncoccacae* (candidate family) were significantly enriched among horses without access to sandwort, while *Oxalobacteraceae* and *Muribaculaceae* were of greater abundance in horses with access to sandwort (Table S1B). Despite non-identical patterns in the Families identified between datasets, there was a positive correlation in the test statistics (log-linear difference divided by the standard error) among families present in both datasets ($R^2 = 0.26$, $t = 3.717$, $p < 0.01$; Figure 6).

Functional analysis using shallow shotgun metagenomics

An advantage of shallow shotgun sequencing over amplicon sequencing is that it can allow for the direct estimation of microbiome functional potential. Mirroring patterns in taxonomic profiles, Bray-Curtis dissimilarities of estimated metabolic pathways significantly differed between horses with access to sandwort versus those without access to sandwort (PERMANOVA: $F = 1.987$, $R^2 = 0.02$, $p < 0.01$). ANCOM-BC analysis (Table S1C) indicated that horses without access to sandwort had greater relative abundances of pathways related to pyruvate fermentation to acetate and lactate (PWY-7357), thiamine phosphate formation (PWY-7357), thiamine diphosphate salvage (PWY-6897), and L-arginine biosynthesis (ARGSYNBSUB-PWY). Conversely, pathways for sucrose degradation (PWY-7345), alongside a suite of other metabolic pathways were enriched when sandwort was present (Figure 7). Interestingly, four of the metabolic pathways significantly associated with horse access to sandwort were Archaea specific. *Post hoc* re-analysis of the shal-

low shotgun dataset in which Archaea were retained alongside bacteria, indicated that the ratio of Archaea to Bacteria was nearly doubled when sandwort was present ($1.3\% \pm 0.3\%$ *CI*), compared to when sandwort was absent ($0.7\% \pm 0.04\%$ *CI* ; $t = 7.239$,$p < 0.01$).

In addition to metabolic pathways, reads which mapped to genes linked to 56 MetaCyc 'reactions' were more relatively abundant among horses with access to sandwort, while 40 were enriched among horses without access to sandwort (Table S1D). Most notably, reads which mapped to cellulase gene regions were more abundant among horses without access to sandwort (Figure 8A), while reads which mapped to genes for fructan beta-fructosidases (Figure 8B) and alpha-amylases (Figure 8C) synthesis were more abundant when sandwort was present.

Discussion

Comparisons of sequencing depth

We sought to test the ability of shallow shotgun sequencing to accurately characterize gut microbiome variation in the Sable Island feral horse population and whether it could recapitulate biological patterns observed in a twinned 16S rRNA gene amplicon dataset. We observed that sequencing depths at the scale of hundreds of thousands of read pairs yielded near identical patterns in microbiome alpha and beta-diversity as more deeply sequenced datasets. Similar results were observed with respect to HUMANn3 based estimates of MetaCyc pathway and reaction relative abundances. While not a synonymous substitution for deep shotgun sequencing, like previous studies , we conclude that shallow shotgun metagenomic sequencing can accurately estimate major biological patterns in the microbiome.

Comparisons of library preparation methods

The feasibility of shallow shotgun sequencing depends on cost-effective library preparation methods. We observed minor differences between datasets created using Nextera XT and iGenomx Riptide preparations. These differences may derive from random variation introduced during laboratory processing, or might be artefacts of the deeper initial sequencing depth of the Nextera XT versus iGenomx Riptide prepared samples (30 versus 4.1 million read-pairs). Additionally, the 16-sample Nextera XT prepared library was sequenced alone on an Illumina Nextseq 500 platform, while the 188-sample iGenomx Riptide library was sequenced on an Illumina NovaSeq6000. More extensive multiplexing during sequencing of the iGenomx Riptide library may have resulted in greater index hopping-based cross-contamination between samples. Nonetheless, iGenomx Riptide library preparations do not appear to quantitatively bias shotgun metagenomic results. We conclude that iGenomx Riptide preparations enables accurate cost-efficient library preparation for high-throughput shallow shotgun sequencing. However, we note recent benchmarking suggests it might be unsuitable for degraded DNA samples

Comparisons of 16S amplicon versus shallow shotgun metagenomic sequencing

Despite the methodological biases inherent in amplicon versus shotgun metagenomic sequencing approaches, these methods yielded similar biological patterns. In a 13-sample set of DNA extracts sequenced using both amplicon and shallow shotgun metagenomic methods (both rarefied to a depth of 35K bacterial read pairs) we observed a positive correlation in Shannon diversity. Bray-Curtis dissimilarities in these sequencing datasets were highly correlated, but declined at finer taxonomic resolutions, and were uncorrelated at the ASV/species level. This is unsurprising, since ASVs are binned based on sequence similarities, while classifications of shotgun metagenomics reads are constrained to the taxonomic demarcations present in reference databases. The modest correlation in genus-level Bray-Curtis dissimilarities could arise from an inability to classify amplicon reads to this taxonomic level. Points of discrepancy in Bray-Curtis dissimilarities may also derive from differences between taxonomies within reference databases, rather than biases in sequencing methods. For example, an abundant genus in the horse microbiome, *Oscillibacter* , is classified within the family *Ruminococcaceae* in the 16S Silva database , but within the family *Oscillospiraceae* in the NCBI non-redundant database.

Bacterial family average relative abundance estimates were positively correlated between sequencing me-

thods. But, abundant families in the Sable Island horse microbiome (*Lachnospiraceae* ,*Ruminococcaceae* , *Prevotellaceae* , *Spirochaetaceae* ,*Rikenellaceae*) tended to be over-represented in the amplicon dataset compared to the shotgun dataset. Amplicon sequencing results are biased by 16S rRNA gene copy number and primer design, while shotgun metagenomic estimates are biased by genome size, or in the context of Kaiju, the size of gene-coding regions. For example, the weakest correlations in relative abundance estimates were observed in*Ruminococcaceae* and *Lachnospiraceae* ; families known to possess large variation in 16S rRNA copy number . Conversely, the strongest correlation was observed between estimates of*Fibrobacteraceae* relative abundance, a narrow clade with low 16S rRNA copy number variation .

Many taxa of moderate relative abundance in the shotgun metagenomic dataset were either absent, or present at lower-than-expected values in the amplicon dataset. Additionally, we observed a clear bi-modal distribution of prevalence in the amplicon dataset, wherein families were either present in nearly all samples, or very few samples (Figure S4). This could suggest that amplicon-based sequencing under-represents some abundant bacterial clades, perhaps due to primer biases or 16S copy number variation . The discrepancies we observed between shallow shotgun metagenomic and 16S amplicon sequencing data are qualitatively similar to previous evaluations of shallow shotgun sequencing , but determining whether shotgun metagenomic or 16S amplicon data more accurately estimates microbiome features requires communities of known composition.

Kaiju-based classification of shotgun metagenomic reads is reported to more accurately estimate taxon relative abundances than 16S amplicon sequencing . Similarly, found that shallow shotgun sequencing more accurately recapitulated communities of a known composition than 16S amplicon data. These previous benchmarking studies, and similarities in the biological patterns described by our shotgun metagenomic and amplicon datasets lead us to conclude that shallow shotgun sequencing provides a suitable, if not superior, substitute for 16S rRNA gene amplicon-based characterization of the microbiome.

Re-examination of a published sample-set using shallow shotgun metagenomics

Amplicon and shotgun metagenomic sequencing painted similar portraits of the Sable Island horse microbiome, but we sought to determine whether shallow shotgun sequencing would recapitulate results first described in an amplicon dataset . A primary finding in this amplicon dataset—a clear beta-diversity by spatial distance relationship—was present (but weaker) in the shallow shotgun dataset. Although partial-length 16S amplicon reads cannot reliably be assigned to species, and frequently not to genus, grouping reads to ASVs can still provide a high-resolution picture of the microbiome. In fact, this approach can sometimes be too sensitive, separating 16S rRNA gene reads derived from the same bacterial genome, into separate ASV bins By contrast, the read-based classification of shotgun reads is constrained by taxonomic demarcations recorded in reference databases. Therefore, amplicon sequencing—or deep shotgun sequencing which allows for *de novo* MAG assembly and strain-resolved tracking—might be better suited than shallow shotgun sequencing for characterizing relationships in the microbiome which are mechanistically driven by microbiota dispersal.

The assembly of Sable Island horse specific MAGs could help increase the resolution of shallow shotgun read classification, however, increased taxonomic resolution is not always useful, since it can obscure biological patterns or cause spurious correlations if explanatory variables are spatially structured; like sea sandwort on Sable Island .

We observed that horse access to sea sandwort was correlated with microbiome beta-diversity differences in both the amplicon and shallow shotgun metagenomic datasets. Although the taxa identified by differential abundance were not identical, both broadly support the same biological conclusion. Namely, that marram grass better supports fibrolytic niches in the microbiome than sea sandwort. In the 16S amplicon dataset, plant fibre specialist bacteria were more abundant when sandwort was absent (*Fibrobacteraceae* , ;*Spirochaetaceae* ,), while simple sugar fermenters were more abundant when horses had access to sea sandwort (*Christensenellaceae* ,). Similarly, in the shallow shotgun metagenomic dataset, plant fibre specialists were more abundant when sandwort was absent (*Fibrobacteraceae* and a marginally non-significant trend in *Spirochaetaceae*) while a functionally diverse clade of plant, host, and alpha glycan foragers,*Muribaculaceae* , were more abundant when sea sandwort was present.

Interestingly, sea sandwort was associated with bacterial families which contain members capable of degrading the plant toxin oxalate, in both the 16S amplicon (*Burkholderiaceae* , Hervé, Junier, Bindschedler, Verrecchia, & Junier, 1982), and shallow shotgun datasets (*Muribaculaceae* , ; *Oxalobacteraceae* ,). Despite discrepancies in the taxa implicated, data from both sequencing methods suggest sea sandwort might be a source of oxalate in the Sable Island horse diet. Conventional analysis of 16S rRNA gene amplicon data would stop at this point; using taxon differential abundance results to make ecological inferences.

Shallow shotgun metagenomic functional analysis

In contrast to 16S amplicon data, shallow shotgun metagenomic data also provide direct estimates of microbiome functional potential. Horses without access to sandwort had greater relative abundances of reads which mapped to genes related to cellulases, enzymes which break down cellulose (the primary structural component of plant cell walls) into monosaccharides. Horses without access to sandwort were also enriched in metabolic pathways for pyruvate fermentation to acetate (and S-lactate). In horses, acetate is the primary fatty acid produced at the end of a metabolic chain which begins with the lysis of plant fibres . Conversely, metabolic pathways for sucrose and host glycan (N-Acetylneuraminic acid) degradation were enriched among horses with access to sandwort, alongside genes for amylase and fructosidase synthesis. Compared to taxonomic patterns, these functional results more directly support our hypothesis that sea sandwort consumption reduces fibrolytic niche space in the microbiome, and shifts communities towards the metabolism of simple sugars and host glycans.

Contrary to taxonomic indicators, we did not observe an association between oxalate decarboxylase genes and horse access to sandwort. Although reads which mapped to oxalate decarboxylase genes were observed, they were present at low abundances (< 10 reads) and not observed in all samples. This highlights a limitation of shallow shotgun sequencing. While shallow shotgun sequencing may be sufficient to detect patterns in genes which are abundant and widely distributed in the microbiome, genes related to specialized functions which are only found in a rare subset of taxa may require deep shotgun metagenomic sequencing or targeted sequencing.

Secondary uses of shallow shotgun metagenomic data

Our analysis of metabolic pathways revealed patterns initially overlooked by our taxonomic emphasis on bacteria; Archaea-associated pathways were more abundant among horses with access to sandwort. Re-analysis of shotgun metagenomic community profiles in which Archaea were retained, confirmed that Archaea were at greater average relative abundance when sandwort was present. Interestingly, *Christensenellaceae* is known to exist in tight co-occurrence patterns with Archaea in the mammalian gut . An association between Archaea and sandwort in the shallow shotgun dataset may parallel the association we observed between *Christensenellaceae* and sea sandwort in the 16S dataset. These results further highlight an ancillary advantage of shallow shotgun metagenomic sequencing; its ability to characterize abundant archaeal and eukaryotic constituents of the microbiome, alongside bacterial communities. Notably, the proportion of shallow shotgun sequencing reads which were classified as Bacteria (97%), Archaea (2.3%), and Fungus (0.3%) are in-line with expectations set by qPCR-based quantitation of these groups in horse feces . In contrast, the characterization of Archaea and Fungus communities using an amplicon sequencing approach would require separate primers, and additional sequencing.

In addition to more comprehensive profiling of the microbiome, shallow shotgun sequencing data might provide dietary insights. For example, we observed reads which mapped to genes related to the metabolic pathways for alginate degradation. In nature, alginate occurs primarily in brown seaweeds , which Sable Island horses are known to consume when it washes ashore. The presence of alginate degradation pathways could indicate a metabolic niche for the metabolism of brown seaweed biomass in the Sable Island horse microbiome. Alternatively, reads mapping to alginate degradation related genes could derive from undigested seaweed in the horse feces. This conclusion is partly supported by our observation of genes related to other metabolic pathways known to occur in brown seaweed, including the mannitol cycle .

The presence of dietary derived sequences in shallow shotgun metagenomic dataset provides both problem,

and opportunity. Dietary confounds of functional analyses might be circumvented by restricting functional profiling to shotgun reads previously identified as microbial, or by filtering reads to remove known dietary items, if reference genomes for dietary items are available. Although diet-derived sequences in shallow shotgun metagenomics datasets represent a potential confound, they might also provide researchers the opportunity to reconstruct host diet, alongside the microbiome. When benchmarked against dietary metabarcoding, dietary reconstruction using shotgun sequences performs well, but requires deep sequencing . However, deep sequencing has been required only because previous studies relied on small marker region reference databases, rather than mapping reads to the genomes of dietary items, which are frequently unavailable. As publicly available genomic data increase so too might the viability of shallow shotgun-based characterization of host diet from fecal samples.

Beyond the scope of any one study, we argue that shallow shotgun metagenomic sequences have greater long-term value than 16S rRNA gene amplicon datasets. Although the sequencing depth inherent of this approach is too low for *de novo* assembly on a single sample basis, shallow shotgun sequencing can still generate a tremendous wealth of metagenomic sequence data in aggregate. Co-assembly methods could be used to create system-specific microbial reference genomes, using the same sequencing data created to profile the microbiome —but at a cost comparable to amplicon sequencing. Study-system specific MAGs would allow for more precise classifications, and permit the use of functional profilers which infer total genomic contents . Incidentally, sequencing a breadth of samples at shallow depths, might allow for more efficient MAG assembly than deeply sequencing a handful of samples . Even if co-assembled shallow shotgun reads are insufficient for MAG recovery, they are useful for identifying samples to target with strategic supplemental deep short read, or long read sequencing.

Beyond the microbiome, if sequences of dietary origin are abundant in shotgun metagenomic datasets, co-assemblies might recover genomic contigs from the host diet. These contigs could be used for diet reconstruction (see previous paragraph) by relaxing our reliance on short marker regions for dietary reconstruction. However, this requires that the genomes of species which are closely related to dietary items are available, so that contigs can be classified. Alternately, in systems where an appreciable fraction of shotgun sequence reads are derived from the host, shallow shotgun sequencing may be capable of characterizing the microbiome, while also providing low-pass genotyping of the host .

Conclusion

We demonstrated that shallow shotgun sequencing yields characterization of the Sable Island horse microbiome which is comparable to both deeper shotgun sequencing (when using a read-based profiler) and 16S rRNA gene amplicon datasets. Shallow shotgun sequencing also replicated biological patterns reported in a previously published amplicon dataset, while also providing estimates of microbiome functional potential. Amplicon data may be better suited for some research questions; for example, those pertaining to biological patterns which are hypothesized to be the result of microbiota dispersal. However, the assembly of system-specific MAGs may help to narrow this gap. Researchers should still seek to validate shallow shotgun sequencing for their own study systems, especially if microbiota in their systems are likely to be underrepresented in genome databases. But, based on data from Sable Island feral horse fecal samples, we conclude that shallow shotgun sequencing can provide greater biological insight than amplicon data, at a comparable cost, and with greater long-term value.

Acknowledgements

We thank the numerous students, research assistants and volunteers who have contributed to sample collection and processing. We also thank D. Colville for permitting the use of the spatially explicit habitat data used in our analyses, and R. Greuel for using these data to create the map featured in the manuscript, and the University of Calgary Centre for Health Genomics and Informatics for assistance with library preparation and sequencing. In-kind and logistical support was provided by the Department of Fisheries and Oceans Canada (DFO), Canada Coast Guard, the Bedford Institute of Oceanography (DFO Science), Environment Canada, Parks Canada Agency, Maritime Air Charters Limited (Sable Aviation), and Sable Island Station

(Meteorological Service of Canada). Funding was provided by the Margaret Gunn Endowment for Animal Research, the L. David Dubé and Heather Ryan Veterinary Health and Research Fund, the University of Calgary, the Morris Animal Foundation (D20EQ-051), the Natural Sciences and Engineering Research Council of Canada (Discovery Grants numbers 2019-04388 to JP, 2016-06459 to PDM), and the Canada Foundation for Innovation (Leaders Opportunity Grant number 25046). MRS was supported by NSERC Vanier, Alberta Innovates, and Killam Pre-Doctoral scholarships. Sample collection and laboratory analyses were performed under Parks Canada Agency Research and Collections Permit SINP-2013-2014, University of Saskatchewan Animal Care Protocol 20090032, and University of Calgary Animal Care Protocol AC18-0078.

References

Author ContributionsJ.P. and P.D.M. secured research funding. J.P. and P.D.M. led sample collection and laboratory analysis. J.P. and M.R.S. designed the study. M.R.S. completed bioinformatic processing, analyses, and wrote the manuscript. All authors contributed to the writing of the final manuscript, and approved of its submission.

Data Availability Statement

Raw 16S amplicon sequence reads are deposited in the SRA (BioProject: PRJNA674675).

The remaining shotgun metagenomic sequence reads will also be archived on the NCBI SRA upon acceptance of the manuscript.

Figures

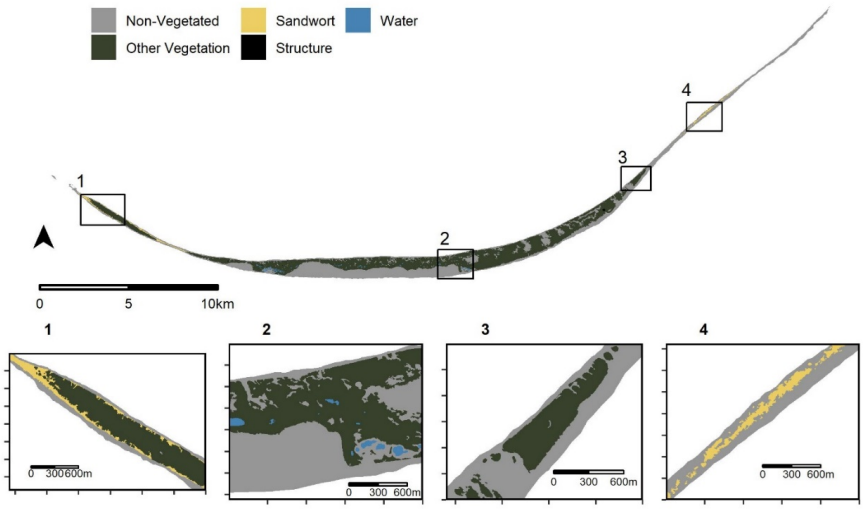

Figure 1. A map of Sable Island National Park Reserve, Nova Scotia (Canada). Habitat classes based on dominant plant assemblages were delineated through a combination of Light Detection And Ranging (LiDAR) surveys, high-resolution aerial photography, and ground truthing. Insets 1, 2, and 3 demonstrate habitat class heterogeneity across the island's length.

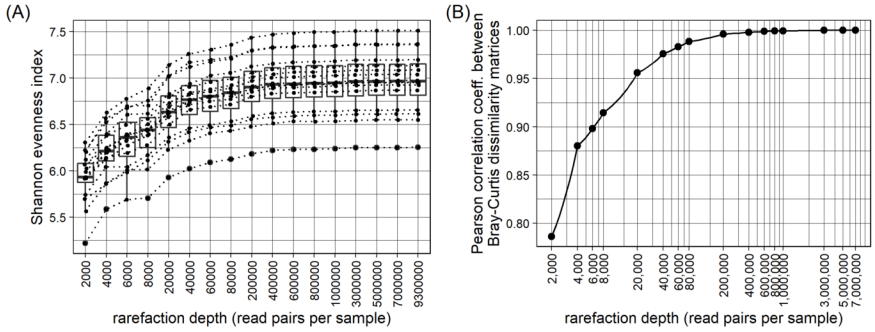

Figure 2. (A) Shannon index estimates of species diversity across a successively rarefied shotgun metagenomic sequencing dataset, in which reads could be classified as bacterial derived, and (B) Pearson correlation coefficients from Mantel tests of species Bray-Curtis dissimilarity matrices created from a minimally rarefied (9.3 million bacterial read-pairs) versus successively rarefied Nextera XT prepared shotgun metagenomic sequence datasets from Sable Island horse fecal samples.

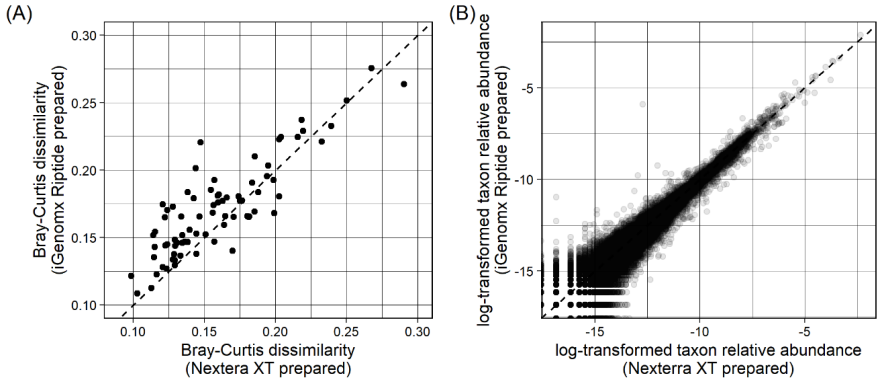

Figure 3. (A) Bray-Curtis dissimilarities of species count tables, and (B) species average relative abundance estimates (log-transformed) from Nextera XT versus iGenomx Riptide prepared shotgun metagenomic sequencing datasets of Sable Island horse fecal samples.

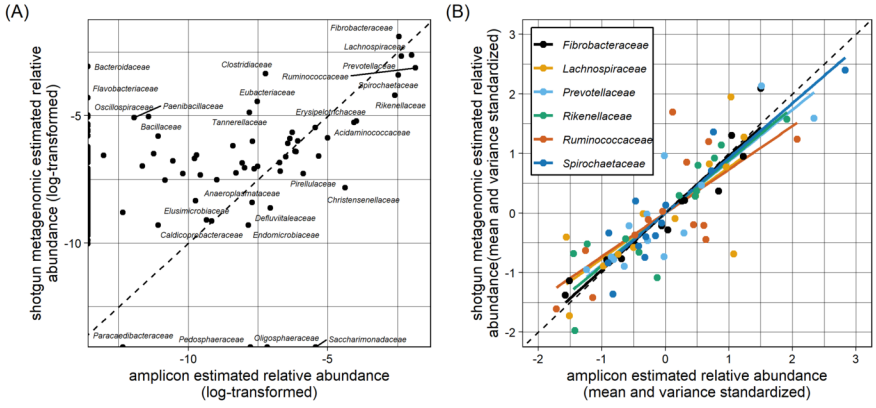

Figure 4: (A) amplicon versus shallow shotgun metagenomic estimates of bacterial family average relative abundance (log-transformed), and (B) amplicon versus shallow shotgun metagenomic relative abundance

estimates (mean and variance scaled) of the most 6 most abundant bacterial families, *Fibrobacteraceae, Lachnospiraceae* ,*Prevotellaceae* , *Rikenellaceae* , *Ruminococcaceae* ,*Spirochaetaceae* . Dashed diagonals are the 1:1 lines.

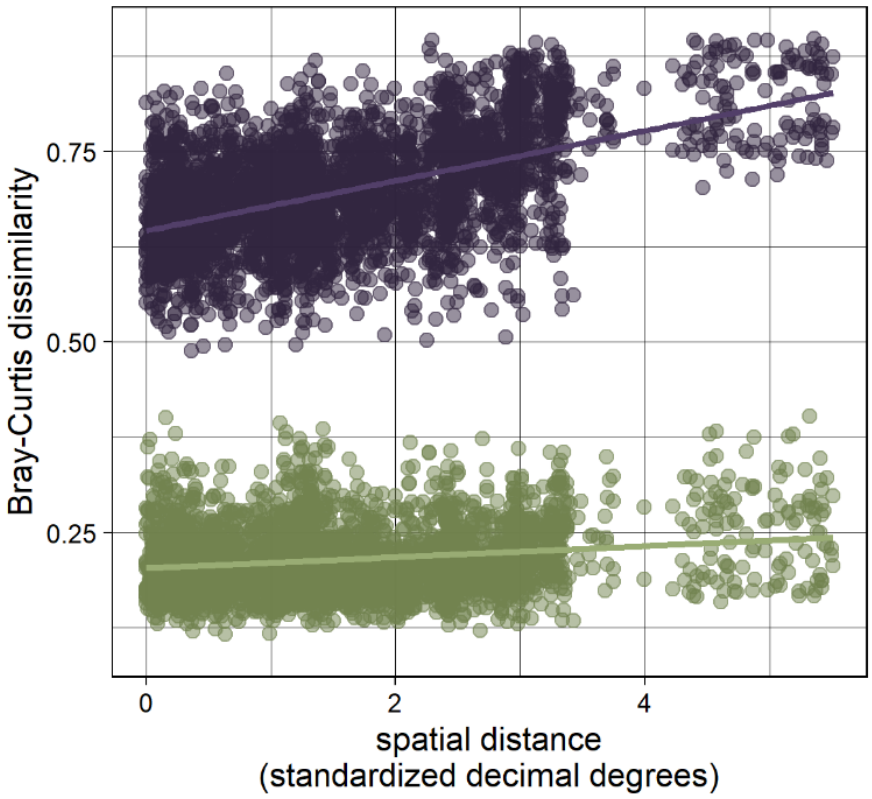

Figure 5: Physical distance separating locations of fecal sample collection versus Bray-Curtis dissimilarity of microbiome sample beta-diversity. Points coloured based on whether they derive from amplicon sequence variant (ASV) estimates from amplicon sequencing data (*), or species-level taxon assignments from shotgun metagenomic sequencing data (*).

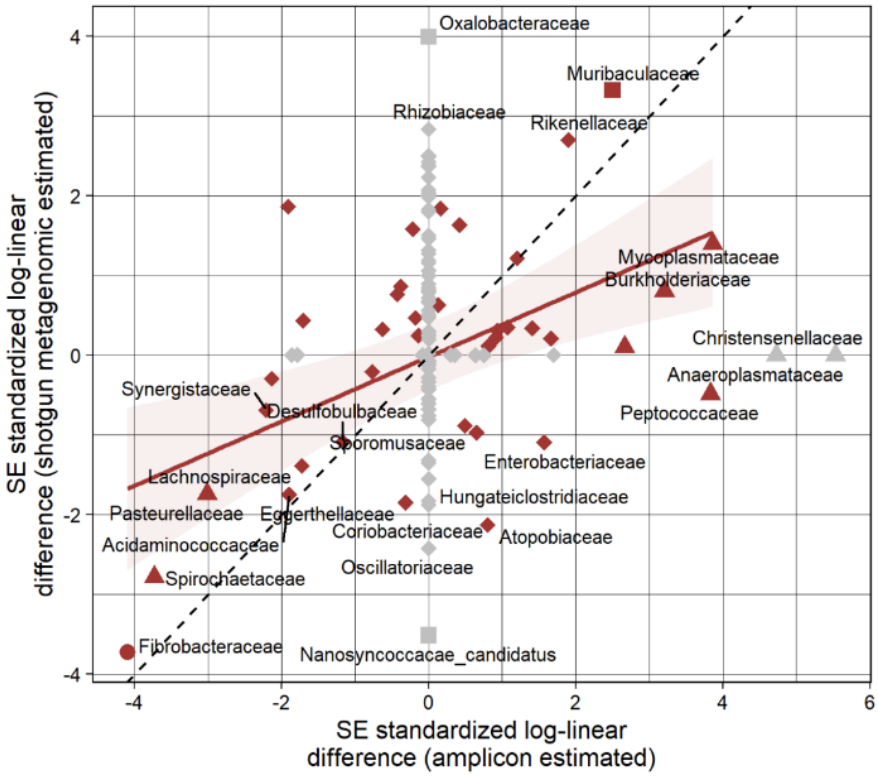

Figure 6. ANCOM-BC based estimates of bacterial family log-linear difference (SE standardized) when sandwort is present versus absent within a 150-metre radius spatial buffer, based on amplicon versus shotgun metagenomic sequencing data. Points coloured by whether family was observed using both (*) or only one (*) sequencing method. Shapes denote whether family significantly differed with sandwort access in the amplicon $(^{)}, shotgun metagenomic(), or both sequencing datasets(*). Dashed diagonal is the 1 : 1 line. Solid line is the best - fit line with 95\% confidence interval shading.$

Figure 7. ANCOM-BC based estimates of metabolic pathway log-linear difference (points) and standard error (bars) in response to sandwort availability. Points coloured according to whether pathways are less (*) or more (*) abundant when sandwort is present.

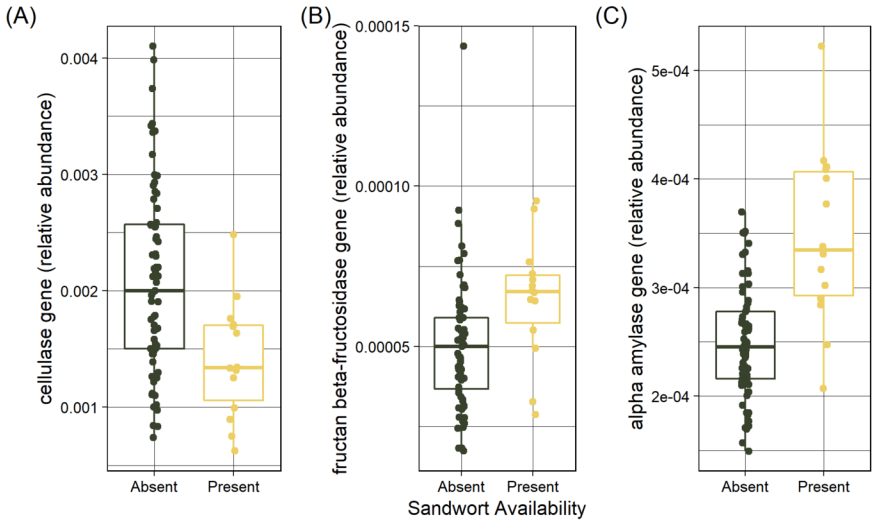

Figure 8. The relative abundance of reads mapped to A) cellulase, B) fructan beta-fructosidase, and C) alpha-amylase gene coding regions in DNA extracted from Sable Island horse feces, separated by whether sandwort was absent or present within a 150-metre radius spatial buffer.