



## Article

# Microbial Diversity Profiling of Gut Microbiota of *Macropus giganteus* Using Three Hypervariable Regions of the Bacterial 16S rRNA

Christian O’Dea <sup>1</sup>, Roger Huerlimann <sup>2</sup>, Nicole Masters <sup>1</sup>, Anna Kuballa <sup>1</sup>, Cameron Veal <sup>3</sup>, Paul Fisher <sup>3</sup>, Helen Stratton <sup>4</sup> and Mohammad Katouli <sup>1,\*</sup>

- <sup>1</sup> Genecology Research Centre, School of Health and Sport Sciences, University of the Sunshine Coast, Maroochydore, QLD 4558, Australia; christiancodea@gmail.com (C.O.); NMasters@USC.EDU.AU (N.M.); AKuballa@USC.EDU.AU (A.K.)
- <sup>2</sup> Marine Climate Change Unit, Okinawa Institute of Science and Technology (OIST), 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan; roger.huerlimann@jcu.edu.au
- <sup>3</sup> Seqwater, 117 Brisbane Street, Ipswich, QLD 4305, Australia; Cameron.Veal@seqwater.com.au (C.V.); Paul.Fisher@seqwater.com.au (P.F.)
- <sup>4</sup> School of Environment and Science, Griffith University, Nathan, QLD 4111, Australia; h.stratton@griffith.edu.au
- \* Correspondence: mkatouli@usc.edu.au; Tel.: +61-7-54302845



**Citation:** O’Dea, C.; Huerlimann, R.; Masters, N.; Kuballa, A.; Veal, C.; Fisher, P.; Stratton, H.; Katouli, M. Microbial Diversity Profiling of Gut Microbiota of *Macropus giganteus* Using Three Hypervariable Regions of the Bacterial 16S rRNA. *Microorganisms* **2021**, *9*, 1721. <https://doi.org/10.3390/microorganisms9081721>

Academic Editor: Flavia Indrio

Received: 8 July 2021

Accepted: 3 August 2021

Published: 12 August 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Animal faecal contamination of surface waters poses a human health risk, as they may contain pathogenic bacteria or viruses. Of the numerous animal species residing along surface waterways in Australia, macropod species are a top contributor to wild animals’ faecal pollution load. We characterised the gut microbiota of 30 native Australian Eastern Grey Kangaroos from six geographical regions (five kangaroos from each region) within South East Queensland in order to establish their bacterial diversity and identify potential novel species-specific bacteria for the rapid detection of faecal contamination of surface waters by these animals. Using three hypervariable regions (HVRs) of the 16S rRNA gene (i.e., V1–V3, V3–V4, and V5–V6), for their effectiveness in delineating the gut microbial diversity, faecal samples from each region were pooled and microbial genomic DNA was extracted, sequenced, and analysed. Results indicated that V1–V3 yielded a higher taxa richness due to its larger target region (~480 bp); however, higher levels of unassigned taxa were observed using the V1–V3 region. In contrast, the V3–V4 HVR (~569 bp) attained a higher likelihood of a taxonomic hit identity to the bacterial species level, with a 5-fold decrease in unassigned taxa. There were distinct dissimilarities in beta diversity between the regions, with the V1–V3 region displaying the highest number of unique taxa ( $n = 42$ ), followed by V3–V4 ( $n = 11$ ) and V5–V6 ( $n = 8$ ). Variations in the gut microbial diversity profiles of kangaroos from different regions were also observed, which indicates that environmental factors may impact the microbial development and, thus, the composition of the gut microbiome of these animals.

**Keywords:** gut microbiota; diversity profiling; macropods; 16S rRNA

## 1. Introduction

Faecal contamination of environmental waterways from warm-blooded animals is one of the leading causes of surface water degradation and risk to human health through a proliferation of zoonotic pathogens [1]. Understanding these animals’ possible faecal contribution is vital for water quality and catchments management from a Water Industry perspective. Wild, agricultural, and domestic animals may contaminate waterways through direct defecation into surface waters. Through transport in surface runoff during rainfall events [2,3], wild animals, who commensally exist alongside agricultural animals in open farmland acreages, contribute to the microbial load of adjacent surface waters [4]. Kangaroos (*Macropus* spp.) are considered to be the most prevalent Australian wild animal,

with their population estimated to comprise between 25–50 million individuals across Australia [4]. Kangaroos commonly occur in regions that are inhabited and used by live-stock and humans, and which include surface waters that are used as a drinking water source. As a ruminant animal, the kangaroo has been identified as hosting several human pathogens within their gut microbiota, with the potential to cause zoonotic harm to humans in rare cases [5,6].

Due to the diversity of gut microbiota and complex bacterial signatures between animals and humans, microbial source tracking (MST) methods have been developed to identify bacterial taxa that may be novel in order to distinguish animals that contribute to freshwater faecal pollution [7]. As animal faecal contamination is less distinguishable as a point source, identifying the faecal pollution's source once it enters the waterway is difficult [8]. Recent advances in MST methods, such as high-throughput sequencing (HTS) technologies, have allowed us to discover novel microorganisms associated with the host species by characterising the gut microbial populations of humans and animals [8,9]. This technology has dramatically improved the taxonomic classification of prokaryotes, mainly through the use of the ~1500 bp, 16S rRNA gene of the bacterial ribosome. While utilising the full-length 16S rRNA gene to characterise bacterial communities is considered to be the gold standard for the full taxonomic resolution of any sample, using one of the nine hypervariable regions (HVR) of the 16S rRNA gene allows for more cost-effective analyses [10–12]. However, limitations surround this HTS method, as there is no single region of the 16S gene that has been identified to date that can accurately represent its entire length or mimic its resolving power [10,13]. This has led to further research into the taxonomic classification method utilising other validation methods [14–17].

Before HTS, most researchers used Sanger sequencing to characterise bacterial communities, a process which often required time-consuming methods, such as cloning or bacterial isolation, before sequencing [18]. However, Sanger sequencing also allowed for “Meta” analyses of a sample, allowing for longer reads which were often greater than 500 base pairs. [19]. The increased speed, reduced cost, simpler protocols, and comparable read lengths of HTS have allowed this technology to supersede Sanger sequencing [20]. Regardless of the sequencing method, the results are represented in operational taxonomic units (OTUs), which usually identify an organism to the genus or species level [21,22].

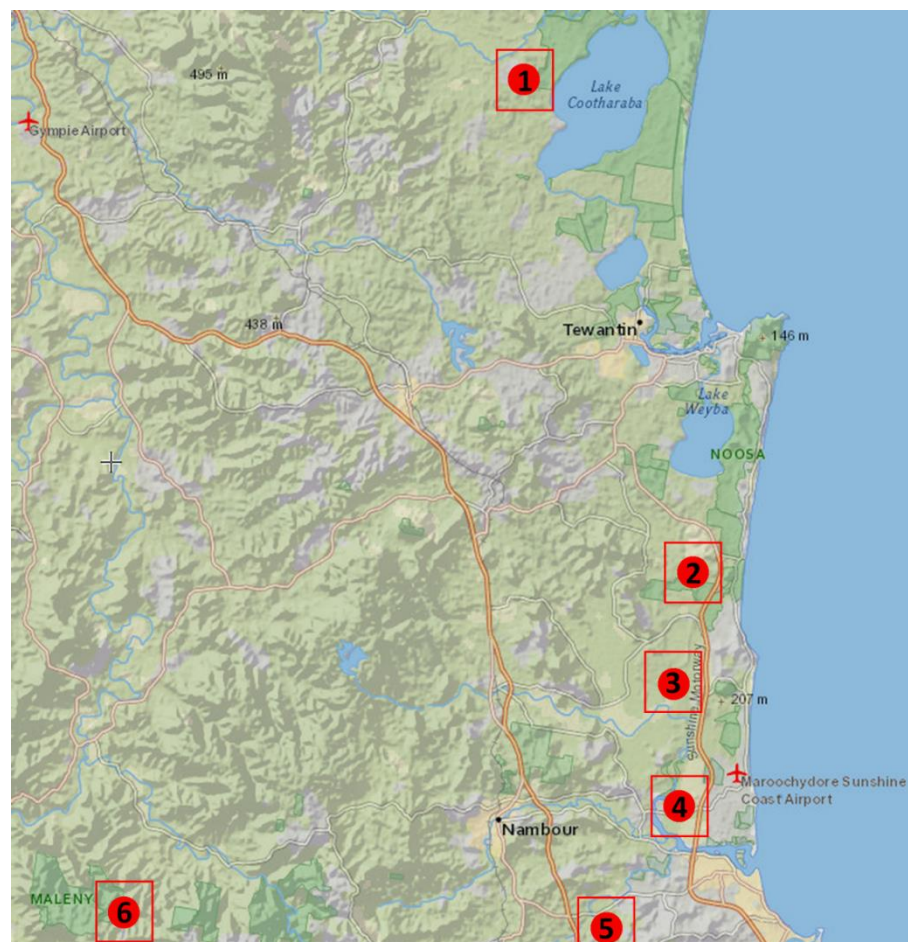
The nine HVRs of the 16S rRNA gene are within conserved regions, facilitating primer design and PCR amplification of the target sequences [23]. Previous studies have utilised only one or two of these regions due to shorter read lengths becoming more manageable than the total length ~1500 bp gene [24]. Conversely, a larger amplicon size will reduce misinterpretation of taxa [14,17,25–28]. These variables can make the selection of the appropriate HVRs troublesome to navigate. Additional considerations include accuracy in the sequencing platform analysis, filtering of data, and denoising the raw data using bioinformatics pipeline software, affecting the overall bacterial community and diversity outcome [29]. Although Australian macropods, including the Eastern Grey Kangaroo, are common to the South East Queensland (SEQ) region of Australia, there is very little information about the bacterial diversity of these animals' gut microbiota [30]. We postulated that selecting an appropriate HVR of the 16S rRNA gene for diversity profiling of gut microbiota would provide the discriminatory power to identify not only the same resolution as the full 16S gene, but also the unique bacterial signatures that can be used to identify the macropods' source of faecal pollution in waterways.

This study was undertaken in order to evaluate the potential of three popular HVRs of the 16S rRNA gene (V1–V3, V3–V4 and V5–V6) for detecting the phylogenetic diversity, species richness, and beta diversity of gut microbiota of Eastern Grey Kangaroos from six regions within SEQ, Australia. Furthermore, the study's purpose was to isolate and identify unique molecular genetic targets in the gut bacterial microbiome of the Eastern Kangaroo to be ultimately used in tracking their movement and faecal contamination in recreational water sources.

## 2. Materials and Methods

### 2.1. Collection of Faecal Samples

Thirty individual faecal samples were collected from known Eastern Grey Kangaroo mob sites across SEQ, Australia (Figure 1). All sites were home to discrete mobs, with no recently known mixing of populations. Sites included Elanda Point (site 1), Sippy Downs (site 2), Yaroomba (site 3), Maleny National Park (site 4), Cootharaba (site 5), and Twin Waters (site 6) (Figure 1). Where possible, fresh faecal samples were collected either after observing an individual animal defecating or from grounds where the animals were previously observed. Faecal samples were collected using sterile, wide-mouth sample containers and transported to the laboratory on ice for subsequent gDNA extraction within six hours of sampling. The kangaroos from which the faecal samples were collected were not observed to be of ill health at the time of sampling. Property owners where the kangaroos reside were also questioned regarding their observations of the mob's general health. Faecal scats (>10 g) were taken from the innermost section of the faecal samples; this internal portion of the scat was used in order to minimise external contaminants.



**Figure 1.** Map of Eastern Grey Kangaroo mob sites where faecal samples were collected in the southeast of Queensland, Australia. Sites in order: Cootharaba, Elanda Point, Yaroomba, Twin Waters, University of the Sunshine Coast, Maleny National Park.

### 2.2. DNA Extraction of Faecal Samples

In accordance with the manufacturer's protocol, DNA extraction of individual faecal samples was carried out using the Qiagen QIAamp Power Faecal DNA kit (Qiagen, Valencia, CA, USA). The elution of DNA was modified using 150  $\mu$ L of C6 buffer instead of the 100  $\mu$ L that is described in the protocol. The yield and purity of the genomic DNA were confirmed using NanoDrop Microvolume Quantitation (Thermo Scientific, Waltham,

MA, USA) and gel electrophoresis. Extracted DNA samples were stored at  $-20\text{ }^{\circ}\text{C}$  until further analysis could be performed. Extracted DNA from five individual faecal samples was pooled post-extraction according to the region where the sample was initially collected. An equal concentration ( $20\text{ }\mu\text{g}/\mu\text{L}$ ) of each faecal sample was used to create the final pooled sample to ensure equal representation and minimise bias. The pooled gDNA samples were submitted to the University of Minnesota Genomics Centre (UMGC), USA, for 16S rRNA amplicon sequencing of the V5–V6 region, while regions V1–V3 and V3–V4 were sequenced at the Australian Genome Research facility (AGRF), Melbourne, Australia.

### 2.3. 16S rRNA Amplification and Miseq Sequencing

Amplicon sequencing of the HVRs was sequenced utilising the Illumina MiSeq platform (San Diego, CA, USA) with two  $\times$  300 base pairs paired-end chemistry. Sequencing and library preparation was undertaken with the Australian Genome Research Facility, Brisbane, Australia and was performed in real-time using the MiSeq Control Software (MCS) v2.6.2.1 and Real-Time Analysis (RTA) v1.18.54, MiSeq instrument computer. Sequencing primers and reagents are described elsewhere [31] (Table 1). Resultant reads were subjected to quality control assessment ‘FASTQC’ (V3) and trimmed using the ‘Trimomatic’ adaptor trimming pre-processing tool (V0.39) [32]. The source libraries, sequencing primers, reagents, equipment, and pipelines were conducted as mentioned previously [33,34]. Primer sets used to amplify the HVRs of the 16S rRNA gene are shown in Table 2. KAPA HiFidelity Hot Start Polymerase was used for PCR amplification that utilised the following cycling conditions:  $95\text{ }^{\circ}\text{C}$  for 5 min, followed by 25 cycles at  $98\text{ }^{\circ}\text{C}$  for 20 s,  $55\text{ }^{\circ}\text{C}$  for 15 s, and  $72\text{ }^{\circ}\text{C}$  for 1 min. Illumina adapter and barcode sequences were added with an additional ten cycles of PCR using the dual index method [35]. Amplicons were gel purified, pooled, and paired-end sequenced at a read length of 300 nt on the Illumina MiSeq platform (Illumina, Inc., San Diego, CA, USA). Region V5–V6 Amplicon sequencing was performed by the University of Minnesota Genomics Centre (<http://genomics.umn.edu/> accessed 15 June 2018) using standard workflows and the Q.C. program as described above [7].

**Table 1.** Primers sequences used in this study for amplification of three hypervariable regions.

Region Name	Primer Name	Primer Sequence	Length	Reference
V1–V3	27F 519R	AGAGTTTGATCMTGGCTCAG GWATTACCGCGGCKGCTG	480 bp	[36]
V3–V4	341F 806R	CCTAYGGGRBGCASCAG GGACTACNNGGTATCTAAT	569 bp	[37]
V5–V6	784F 1064R	RGGATTAGATACCC CGACRRCCATGCANCACCT	280 bp	[7,38]

**Table 2.** Taxonomic composition of the most abundant bacterial phyla when tested for three HVRs (V1–V3, V3–V4, and V5–V6) of the 16S rRNA genes.

Taxa (Phylum)	V1–V3 Mean	Taxa (Phylum)	V3–V4 Mean	Taxa (Phylum)	V5–V6 Mean
Proteobacteria	49.65%	Firmicutes	47.39%	Firmicutes	45.14%
Firmicutes	28.07%	Proteobacteria	29.25%	Proteobacteria	19.55%
Bacteroidetes	12.02%	Bacteroidetes	14.50%	Bacteroidetes	18.18%

### 2.4. Bioinformatic Analysis of Data

The processing of raw sequence 16S rRNA data used Quantitative Insights into Microbial Ecology (QIIME2), an open source, a metagenomic bioinformatic pipeline analysis tool [39]. This bioinformatic methodology was also mentioned elsewhere [40]. Samples were indexed into categories that aligned with a metadata file. Samples were first sub-

jected to joining forward and reverse reads using artefact ‘Paired-End Sequences with Quality’ [39]. The joined reads were denoised and filtered of reads compromising the final output. The ‘DADA2 denoise-paired’ artefact trimmed forward and reverse primer ends of the amplicons and truncated the length of the region from the sequence reads [39]. Utilising a denoising algorithm also removed error in PCR recombination (chimeras), and a sequencing identity algorithm that removed reads with less than a 97% similarity cut-off. The resultant output QZA files underwent a ‘Feature table Merge’ of the denoised amplicon data to be analysed downstream. Sequence alignment, mask/filter alignment, and production of phylogenetic tree files were created. Biases were removed by ‘alpha rarefaction’ in order to accommodate and filter only quality reads at their maximum read length depths. Finally, the taxonomic assignment was undertaken with an open reference database, ‘SILVA-132-99 classifier’, for its regularly updated, quality-controlled resource data [41]. Phylogenetic diversity profiling was carried out in order to determine relative abundance -of bacterial genera between populations in different samples.

Exported taxonomy files were converted to the BIOM format for subsequent output analysis into the open-source metagenomic data facilitation software CALYPSO [42]. Data Filtering removed samples with short sequence read counts (<1000 sequence reads) and OTUs with low abundance (<1%). Furthermore, only the top 3000 taxa were included and filtered by mean. Data were normalised by dividing the number of sequence reads by the total number of reads within each sample by total sum normalisation (TSS).

Host-specific microbial richness was estimated using Alpha diversity matrices Chao1, ACE, Shannon’s diversity, and Simpson’s diversity indices. Distance relationships between microbial community groups were performed using Bray-Curtis [43]. Distance matrices by beta diversity measurements were visualised by Principle Coordinate Analysis (PCoA) using CALYPSO [42]. Unique taxa and incidence mapping comparing each of the three regions tested were calculated by group analysis at the genus level in order to identify unique and shared taxa between HRVs [42,44].

### 2.5. Statistical Analyses

Statistical analysis was performed using Graph Pad Prism software version 9.01 for Windows (GraphPad Software, San Diego, CA, USA). Differences between the bacterial richness obtained from the three HVRs were determined using the Kruskal-Wallis analysis of variance (ANOVA) and the values were considered statistically significant where  $p < 0.05$ .

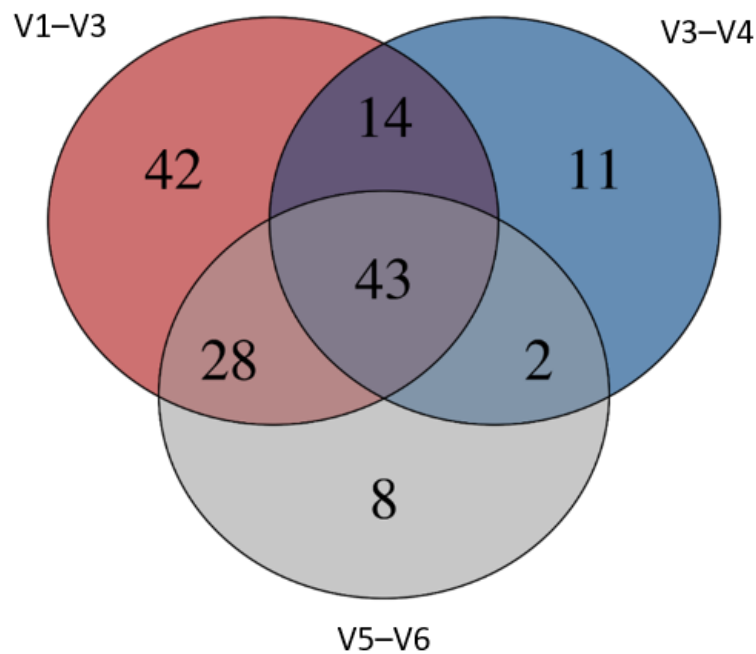
## 3. Results

### 3.1. Sequence Data Summary and Depth

Illumina sequencing of each of the three HVRs resulted in: 1,962,938 demultiplexed sequence reads for the V1–V3 region; 997,779 sequence reads for region V3–V4; and 271,184 reads for region V5–V6. Post rarefaction filtering to 10,000 reads per sample was employed before alpha diversity metrics were analysed. Chao1 and ACE bacterial richness indices showed a significantly higher bacterial richness in region V1–V3 out of all three regions ( $p < 0.05$ ) (Figure 2). Shannon’s and Simpson’s diversity indices displayed a significantly broader diversity within the V1–V3 region (Figure 2).

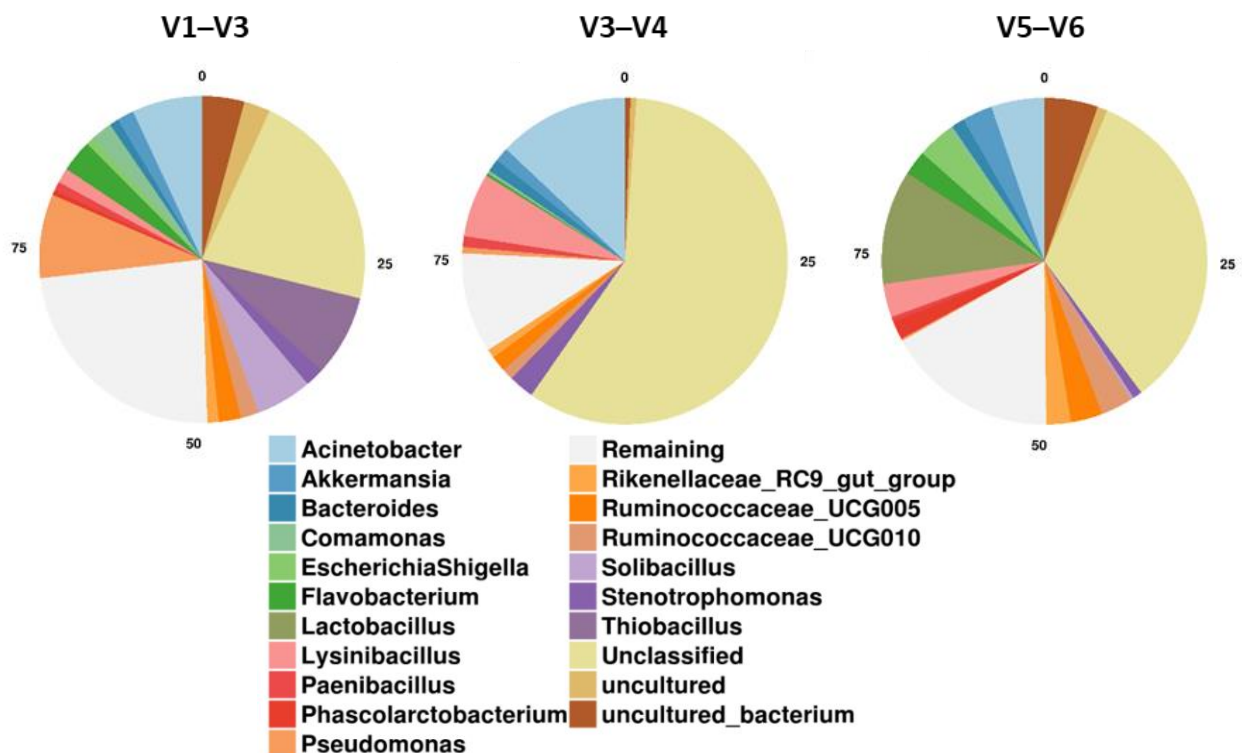
### 3.2. Taxonomic Composition by Hypervariable Region

The most abundant OTUs within the three HVRs of the 16S rRNA gene were identified as belonging to phyla Proteobacteria, Firmicutes and Bacteroidetes, which comprised up to 91% of the total bacterial coverage (Table 2). Proteobacteria was the predominant phyla in HVR V1–V3 (49.7%) and Firmicutes was predominant in HVR V3–V4 (47.4%) and V5–V6 (45.1%). Bacteroides, a known contributor to the upper portion of all mammalian’s gastrointestinal tracts [45], was also present as the third most abundant bacterial phyla within all three HVRs tested (Table 2).



**Figure 2.** Group visualisation of taxa abundances across three chosen hypervariable regions of the 16S rRNA gene. The top 300 most abundant genera were utilised for visualisation of shared and exclusive genera.

The highest proportion of unclassified bacterial OTUs was observed in HVR V3–V4 (58.5%). In comparison, HVRS V1–V3 and V5–V6 displayed much lower unclassified OTUs, 21.7% and 33.2%, respectively (Figure 3).



**Figure 3.** Microbial community composition at the genus level, within the three hypervariable regions of the 16S rRNA gene.

The top 40 genera underwent incidence analysis to determine if all three regions were able to detect any of the given genera. Of the top 40 genera, 95% were sensitive to being detected by two out of the three HVRs (Figure 4).

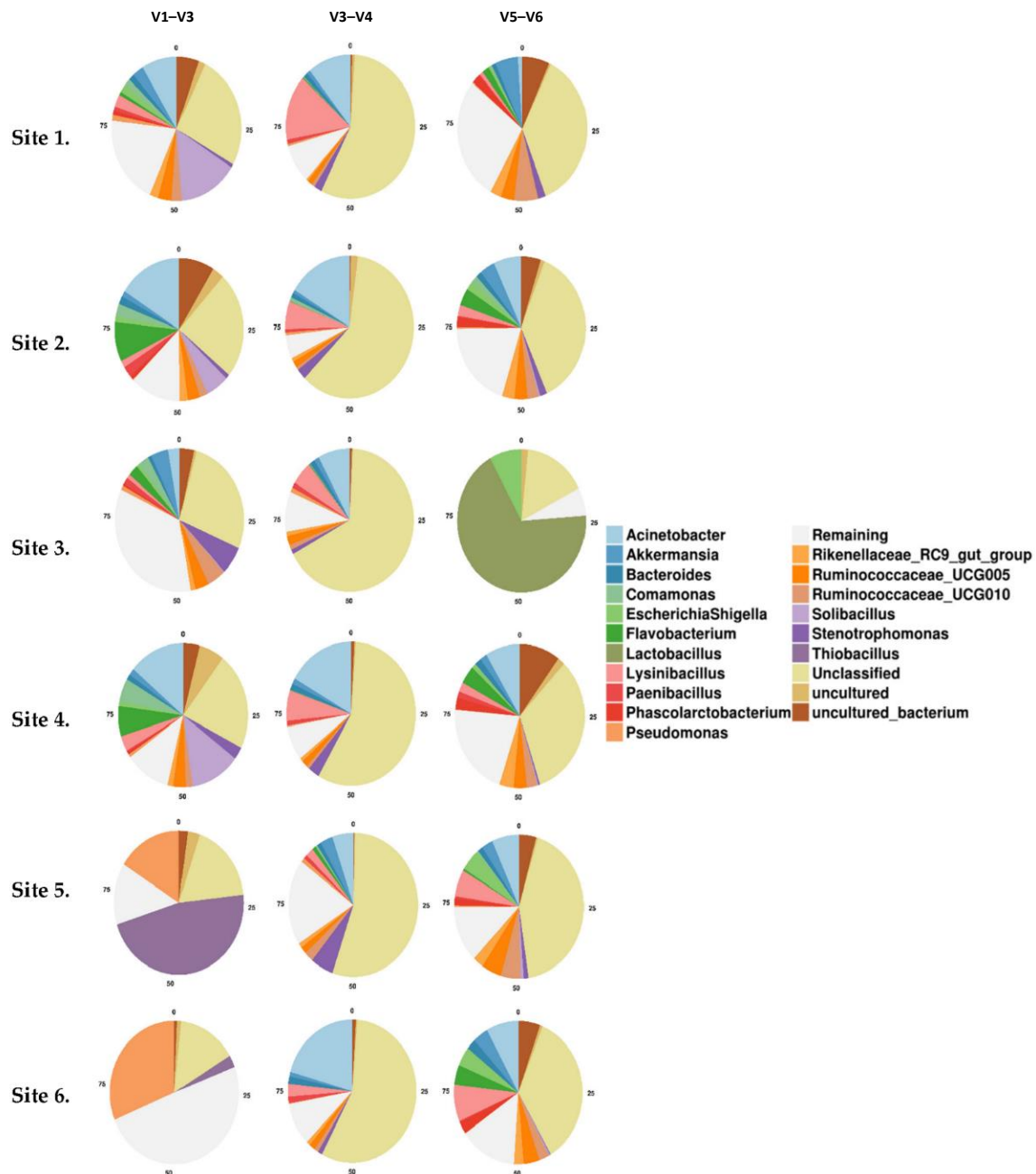
Genus	Region of 16S rRNA Gene		
	V1–V3	V3–V4	V5–V6
<i>Tyzzarella_4</i>	Red	White	Red
<i>Streptomyces</i>	Red	Red	White
<i>Sphingomonas</i>	Red	White	Red
<i>Solibacillus</i>	Red	White	White
<i>Ruminococcus_2</i>	Red	Red	White
<i>Ruminococcus_1</i>	Red	Red	White
<i>Ruminococcaceae_UCG002</i>	Red	White	Red
<i>Ruminococcaceae_NK4A214_group</i>	Red	Red	White
<i>Ruminiclostridium_9</i>	Red	White	Red
<i>Roseomonas</i>	Red	Red	White
<i>Pygmaibacter</i>	Red	Red	White
<i>Providencia</i>	Red	White	Red
<i>Prevotellaceae_UCG001</i>	Red	Red	Red
<i>Phascolarctobacterium</i>	Red	White	Red
<i>Parasutterella</i>	Red	White	Red
<i>Parabacteroides</i>	Red	White	Red
<i>p1088a5_gut_group</i>	Red	White	Red
<i>Olsenella</i>	Red	Red	White
<i>Mogibacterium</i>	Red	White	Red
<i>Methanocorypusculum</i>	White	Red	White
<i>Mailhella</i>	Red	White	Red
<i>Lactococcus</i>	Red	White	Red
<i>Lactobacillus</i>	White	Red	Red
<i>Kurthia</i>	Red	White	Red
<i>Fusobacterium</i>	Red	Red	White
<i>EscherichiaShigella</i>	Red	White	Red
<i>Enterobacter</i>	Red	White	Red
<i>dgA11_gut_group</i>	Red	Red	White
<i>Delftia</i>	Red	White	Red
<i>Defluviitaleaceae_UCG011</i>	Red	White	Red
<i>Cronobacter</i>	Red	White	Red
<i>Cohnella</i>	Red	Red	White
<i>Cerasicoccus</i>	Red	White	Red
<i>Cellulosilyticum</i>	Red	Red	White
<i>Cellulomonas</i>	Red	White	Red
<i>Candidatus_Soleaferrea</i>	Red	White	Red
<i>Bacteroidales_bacterium_55_9</i>	Red	White	Red
<i>Arthrobacter</i>	Red	White	Red
<i>Alloprevotella</i>	Red	Red	White
<i>Achromobacter</i>	Red	Red	White

**Figure 4.** Incidence map displaying bacteria found at the genus level in each of the three hypervariable regions of the 16S rRNA gene (V1–V3, V3–V4 and V5–V6). Presence (red), and absence (white).

### 3.3. Taxonomic Composition by Geospatial Region

The taxonomic composition of the faecal microbiota of Kangaroos from the six regions was compared using the results obtained from each of the three HVRs.

Bacterial OTU abundance was found to vary according to geographical location. Firmicutes were the predominant phyla within sites 5 (Cootharaba), 1 (Elanda Point), 6 (Twin Waters), and 2 (Yaroomba) for all three HVRs. However, V1–V3 analysis identified Proteobacteria as the most abundant at sites 4 (Maleny National Park) and 2 (Sippy Downs) (89.6% and 85.6%, respectively) (Figure 5).



**Figure 5.** Quantitative visualisation of bacterial phyla taken from kangaroo faecal material at six sample sites within South East Queensland. Samples were tested against each of the three hypervariable regions of the 16S rRNA gene, V1–V3, V3–V4, and V5–V6. Sites are displayed in order: 1—Cootharaba, 2—Elanda Point, 3—Yaroomba, 4—Twin Waters, 5—the University of the Sunshine Coast, 6—Maleny National Park.

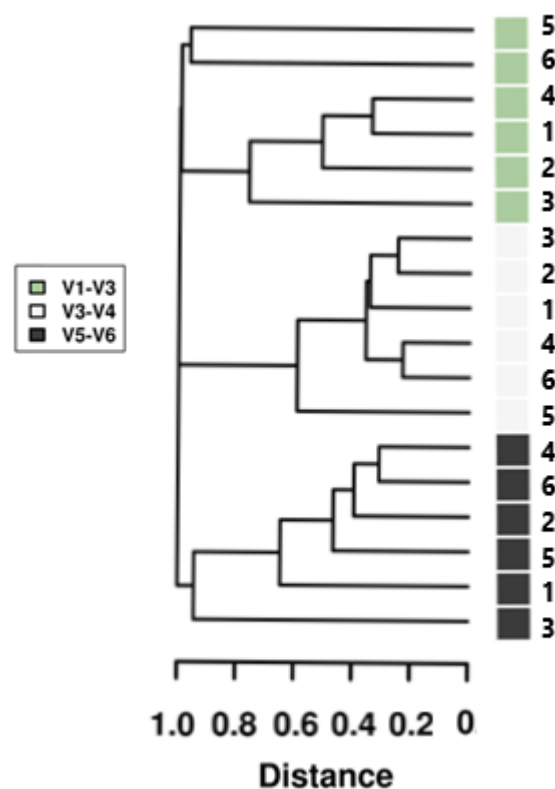


### 3.4. Group-Specific Bacterial Communities

Unique taxon differences between HVRs used in this study were conducted using the permutational multivariate analysis of variance (Figure 2). The V1–V3 HRV of the 16S gene showed the highest level of unique taxa identified among the three regions ( $n = 42$ ). The least unique genera were identified in HVR V5–V6, while all three HVRs shared a total of 43 genera (Figure 2).

### 3.5. Beta Diversity

Beta diversity profiling identified dissimilarities among bacterial communities within all three HVRs tested and across geographical sampling sites for the Eastern Grey Kangaroo populations. PCoA analysis revealed the V3–V4 region as the least differentiating diversity clustering of all the HVRs tested. Results indicate the most significant geographical diversity was observed between the ‘University of the Sunshine Coast’ (USC) & ‘Maleny National Park’ (MNP) sampling sites, when analysing HVR V1–V3. Similarity coefficients between the composition of the six pooled gut microbiota of EGK using three HVRs showed a higher similarity among the pooled gut microbiota of Kangaroos within each HVR analysis than between HVRs (Figure 6). Sample sites were also analysed for similarity; however, no significant similarity was observed between the six sample sites (Figure 6).



**Figure 6.** Hierarchical dendrogram displaying Bray-Curtis distances from samples of Eastern Grey Kangaroo faecal gDNA. Samples were analysed against the 16S rRNA bacterial gene regions V1–V3, V3–V4 and V5–V6. Sites are displayed in order: 1—Cootheraba, 2—Elanda Point, 3—Yaroomba, 4—Twin Waters, 5—the University of the Sunshine Coast, 6—Maleny National Park.

For the first time, this research presents insights into the effectiveness of the 16S rRNA gene sequence analysis to reconstruct phylogenies using three of the nine favoured, highly conserved regions of the 16S rRNA gene from macropod species *Macropus giganteus*; The Eastern Grey Kangaroo. The results of this experiment suggest that the bacterial phylogenies obtained from the testing of kangaroo gut-faecal material vary depending on which HRV of the 16S rRNA is analysed. Currently, no single HVR (neither in the

literature nor from our analysis) can be defined as a standard. Prior research of this high-throughput sequencing technology using various HRVs described certain regions as adept at distinguishing bacterial OTUs over others [46]. Our research aligns with other analyses that suggest that short read HRV's are not as effective in achieving the full taxonomic resolution of the full length 16S gene (~1500 bp) [13]. Analysis of gut microbiota and the diversity of bacteria in faeces may be affected by, but not limited to, several factors, which include (a) gDNA extraction method applied to faecal samples, (b) raw data processing, and (c) sequencing depth and microbiota improvement post-low-frequency sequence removal. Previous studies show that the efficacy of bacterial phylogenetic derivation may be compromised if the DNA extracted from animal faecal samples contains ethylenediamine-tetra acetic acid (EDTA) [46]. The use of EDTA has also increased the abundance of phylogenies such as Proteobacteria and decreased the abundance of Firmicutes [46]. Results from our study indicate that Proteobacteria are the predominant bacterial phyla when analysed using HVR V1–V3, while V3–V4 and V5–V6 predominantly identified Firmicutes. This phylogenetic shift in the most abundant bacterial communities is also documented in earlier research [47].

Illumina sequencing of the three HVRs in our study resulted in region V1–V3 displaying a 1.96-fold increase in the sequence read count over region V3–V4 and a seven-fold increase over region V5–V6. We also found that removing low frequency reads in-silico aids in a higher sensitivity of bacterial diversity. This removal of singletons decreases overall error rates, as also described elsewhere [29,48,49]. A loss of ~90% of the original reads post denoising protocol (DADA2) was observed for HVR V1–V3. This degree of denoising is a common trait for HVR V1–V3 when undertaking this method, as the original reads are removed from the analysis due to a failure to merge the paired reads. We used the forward reads only, as a higher yield in reading outputs can be attained in analysing our data. However, since single read runs are faster and cheaper, future sequencing projects using this primer set could use single-end sequencing.

It is acknowledged that the full length of the 16S rRNA gene is more appropriate in achieving a more definitive resolution of the taxa within a sample [50]. Identification of errors in amplicon sequencing are attended to by utilising quality filtering bioinformatic pipeline scripts [39,51,52]. There exists multiple denoising (error correcting) bioinformatic pipelines for resolving amplicon sequencing variants. By using the bioinformatic post denoising protocol (DADA2), the identification of PCR chimeras and correcting of sequence amplicon errors assists in decreasing the resultant unclassified taxa of up to a 97% similarity cut-off. This aids in a more discriminatory OTU identification [13,53,54].

The predominant bacterial phylogeny within the V1–V3 region was Proteobacteria. The other two HVRs tested were comprised of communities of Firmicutes as the dominant phyla. Research published previously indicates a high presence of Proteobacteria, evaluating the HVR's ability to quickly identify the bacterial sequences during high-throughput sequencing in faecal microbiome samples [47,55]. There exists no gold standard for utilising the correct HVR for the target bacterial community of interest. The sequencing of multiple regions exists in modern HTS research for a more comprehensive analysis of the sample of interest [50].

The literature suggests that the HVR preference in modern HTS depends on the microbiome source being tested. Analysis of alpha and beta diversity in our study showed that the V1–V3 region of the 16S rRNA gene held the most significant species richness when tested for alpha diversity. Conversely, beta diversity results displayed dissimilarity in the region holding the highest bacterial species between all three HVRs. A greater level of species richness can translate to a community composition that is not fundamentally similar.

Pitfalls relating to HTS technology pertain to the reproducibility of results using this technology. Sample processing and the extraction stage are related to the specificity of the bacteria present post metagenomic analysis. PCR based confirmation steps would hamper the effectiveness and affordability of attempts to confirm HTS analysis accuracy [56].

Measures to prevent this could include multiplex PCR which would enable the observable phyla within a DNA sample to help confirm HTS analysis [57].

This study hopes to show which HRV can be used as a comparative analysis in substitution of the full length ~1500 bp 16S rRNA gene. Through previous HRV comparison studies, it is evident that the conditions of a sample and the region of the 16S rRNA greatly dictate the microbial diversity of a sample. The literature is largely comprised of research into region comparisons across samples in isolation. It is noticed that through comparison of taxonomic percentages, not all HRVs display a consistent detection of genera across the 3 regions in this study, even if an equimolar mix of samples were used on the same HRV primer set [15,58]. Additionally, associated environmental factors such as diet, health and age can impact an animals bacterial microbiome [59].

It is not currently feasible to creating a database of suitable HRVs to use on a specific sample, in this case the gut microbiome of an animal species. Variability of any sample used against any HRV has the potential to differ in sequence variation, no matter how uniform the samples may be.

#### 4. Conclusions

In conclusion we revealed that the gut microbiota of kangaroos examined in this study was diverse. However, we could not identify unique taxa as potential kangaroo specific faecal markers that could be used as a tool for identifying faecal contamination of surface waters by macropods. Our results did suggest that, depending on which HVR of the 16S rRNA is used, significant differences in bacterial phylogeny will be obtained, making it difficult to suggest which bacteria taxa could be nominated for further investigation in order to pinpoint a candidate for MST purposes. Secondary confirmation methods such as quantitative PCR and alternative bioinformatic pipeline methods to discover variances in data could provide a better picture of the bacterial markers specific to these animals. Nonetheless, we found a combined use of three variable regions of 16S rRNA could substitute for the use of the entire ~1500 bp gene (the full-length gene). Although this was a more costly approach to bacterial phylogenetic testing, it is more suitable for a better understanding of the gut microbial diversity and the unique taxa of macropods. Future studies are needed to investigate direct comparison of the three HVRs used with samples analysed along the full length 16s gene.

**Author Contributions:** C.O.: Conceptualization, field research, data acquisition and validation, statistical analysis, bioinformatics, manuscript editing; R.H.: Bioinformatic data education, conceptualisation, data validation, manuscript writing, supervision; N.M.: Conceptualization, laboratory development, supervision; A.K.: Laboratory development, Manuscript editing; C.V.: Project funding; P.F.: Project funding, project administration, Manuscript editing; H.S.: Project conceptualization, resources, manuscript editing, project administration; M.K.: Funding acquisition, project administration, conceptualization, manuscript editing, resources. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded, in part, by the University of the Sunshine Coast scholarship and by Seqwater [top-up scholarship 4530e17] for the study design of local library sampling locations and target animals within the study.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the University of the Sunshine Coast (USC) Animal Ethics Committee (AN/A/17/117), 2017.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data are contained within the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Gyawali, P.; Hamilton, K.; Joshi, S.; Aster, D.; Ahmed, W. Identification of reliable marker genes for the detection of canine fecal contamination in sub-tropical Australia. *Sci. Total. Environ.* **2020**, *718*, 137246. [[CrossRef](#)]

2. Kirs, M.; Harwood, V.; Fidler, A.; Gillespie, P.; Fyfe, W.; Blackwood, A.; Cornelisen, C. Source tracking faecal contamination in an urbanised and a rural waterway in the Nelson-Tasman region, New Zealand. *N. Z. J. Mar. Freshw. Res.* **2011**, *45*, 43–58. [[CrossRef](#)]
3. Johnson, L.K.; Brown, M.B.; Carruthers, E.A.; Ferguson, J.A.; Dombek, P.E.; Sadowsky, M.J. Sample Size, Library Composition, and Genotypic Diversity among Natural Populations of *Escherichia coli* from Different Animals Influence Accuracy of Determining Sources of Fecal Pollution. *Appl. Environ. Microbiol.* **2004**, *70*, 4478–4485. [[CrossRef](#)] [[PubMed](#)]
4. Kangaroo and Wallaby Statistics Archive. Available online: <https://www.environment.gov.au/biodiversity/wildlife-trade/natives/wild-harvest/kangaroo-wallaby-statistics/kangaroo-population> (accessed on 29 December 2020).
5. Power, M.L.; Sangster, N.C.; Slade, M.B.; Veal, D.A. Patterns of *Cryptosporidium* Oocyst Shedding by Eastern Grey Kangaroos Inhabiting an Australian Watershed. *Appl. Environ. Microbiol.* **2005**, *71*, 6159–6164. [[CrossRef](#)]
6. Carossino, M.; Bauer, R.; Mitchell, M.A.; Cummings, C.O.; Stöhr, A.C.; Wakamatsu, N.; Harper, K.; Langohr, I.M.; Schultz, K.; Mitchell, M.S.; et al. Pathologic and immunohistochemical findings in an outbreak of systemic toxoplasmosis in a mob of red kangaroos. *J. Vet. Diagn. Investig.* **2021**, *33*, 554–565. [[CrossRef](#)]
7. O’Dea, C.; Zhang, Q.; Staley, C.; Masters, N.; Kuballa, A.; Fisher, P.; Veal, C.; Stratton, H.; Sadowsky, M.J.; Ahmed, W.; et al. Compositional and temporal stability of fecal taxon libraries for use with SourceTracker in sub-tropical catchments. *Water Res.* **2019**, *165*, 114967. [[CrossRef](#)]
8. Unno, T.; Staley, C.; Brown, C.M.; Han, D.; Sadowsky, M.J.; Hur, H.-G. Fecal pollution: New trends and challenges in microbial source tracking using next-generation sequencing. *Environ. Microbiol.* **2018**, *20*, 3132–3140. [[CrossRef](#)]
9. Ahmed, W.; Goonetilleke, A.; Powell, D.; Chauhan, K.; Gardner, T. Comparison of molecular markers to detect fresh sewage in environmental waters. *Water Res.* **2009**, *43*, 4908–4917. [[CrossRef](#)] [[PubMed](#)]
10. Dias, V.H.G.; Gomes, P.D.S.F.C.; Azevedo-Martins, A.C.; Cabral, B.C.A.; Woerner, A.E.; Budowle, B.; Moura-Neto, R.S.; Silva, R. Evaluation of 16S rRNA Hypervariable Regions for Bioweapon Species Detection by Massively Parallel Sequencing. *Int. J. Microbiol.* **2020**, *2020*, 1–11. [[CrossRef](#)]
11. Greay, T.L.; Gofton, A.W.; Paparini, A.; Ryan, U.M.; Oskam, C.L.; Irwin, P.J. Recent insights into the tick microbiome gained through next-generation sequencing. *Parasites Vectors* **2018**, *11*, 1–14. [[CrossRef](#)] [[PubMed](#)]
12. Bukin, Y.S.; Galachyants, Y.P.; Morozov, I.V.; Bukin, S.V.; Zakharenko, A.S.; Zemskaya, T. The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci. Data* **2019**, *6*, 190007. [[CrossRef](#)]
13. Johnson, J.S.; Spakowicz, D.J.; Hong, B.-Y.; Petersen, L.M.; Demkowicz, P.; Chen, L.; Leopold, S.R.; Hanson, B.M.; Agresta, H.O.; Gerstein, M.; et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **2019**, *10*, 5029. [[CrossRef](#)] [[PubMed](#)]
14. Yang, B.; Wang, Y.; Qian, P.-Y. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinform.* **2016**, *17*, 1–8. [[CrossRef](#)] [[PubMed](#)]
15. Claesson, M.J.; O’Sullivan, O.; Wang, Q.; Nikkilä, J.; Marchesi, J.R.; Smidt, H.; De Vos, W.M.; Paul Ross, R.; O’Toole, P.W. Comparative Analysis of Pyrosequencing and a Phylogenetic Microarray for Exploring Microbial Community Structures in the Human Distal Intestine. *PLoS ONE* **2009**, *4*, e6669. [[CrossRef](#)] [[PubMed](#)]
16. Myer, P.R.; Kim, M.; Freetly, H.C.; Smith, T.P. Evaluation of 16S rRNA amplicon sequencing using two next-generation sequencing technologies for phylogenetic analysis of the rumen bacterial community in steers. *J. Microbiol. Methods* **2016**, *127*, 132–140. [[CrossRef](#)] [[PubMed](#)]
17. Chakravorty, S.; Helb, D.; Burday, M.; Connell, N.; Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* **2007**, *69*, 330–339. [[CrossRef](#)]
18. Williams, E.L.; Bagg, E.A.L.; Mueller, M.; Vandrovcova, J.; Aitman, T.J.; Rumsby, G. Performance evaluation of Sanger sequencing for the diagnosis of primary hyperoxaluria and comparison with targeted next generation sequencing. *Mol. Genet. Genom. Med.* **2014**, *3*, 69–78. [[CrossRef](#)]
19. Vincent, A.T.; Derome, N.; Boyle, B.; Culley, A.I.; Charette, S.J. Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *J. Microbiol. Methods* **2017**, *138*, 60–71. [[CrossRef](#)]
20. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinf.* **2015**, *13*, 278–289. [[CrossRef](#)]
21. Dickson, R.P.; Erb-Downward, J.R.; Martinez, F.J.; Huffnagle, G.B. The microbiome and the respiratory Tract. *Annu. Rev. Physiol.* **2016**, *78*, 481–504. [[CrossRef](#)]
22. Ames, N.J.; Ranucci, A.; Moriyama, B.; Wallen, G.R. The Human Microbiome and Understanding the 16S rRNA Gene in Translational Nursing Science. *Nurs. Res.* **2017**, *66*, 184–197. [[CrossRef](#)] [[PubMed](#)]
23. Baker, G.C.; Smith, J.J.; Cowan, D.A. Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods* **2003**, *55*, 541–555. [[CrossRef](#)]
24. Birtel, J.; Walser, J.-C.; Pichon, S.; Bürgmann, H.; Matthews, B. Estimating Bacterial Diversity for Ecological Studies: Methods, Metrics, and Assumptions. *PLoS ONE* **2015**, *10*, e0125356. [[CrossRef](#)] [[PubMed](#)]
25. Zrimec, J.; Kopinč, R.; Rijavec, T.; Zrimec, T.; Lapanje, A. Band smearing of PCR amplified bacterial 16S rRNA genes: Dependence on initial PCR target diversity. *J. Microbiol. Methods* **2013**, *95*, 186–194. [[CrossRef](#)]
26. Ahmed, W.; Staley, C.; Sadowsky, M.; Gyawali, P.; Sidhu, J.P.S.; Palmer, A.; Beale, D.J.; Toze, S. Toolbox Approaches Using Molecular Markers and 16S rRNA Gene Amplicon Data Sets for Identification of Fecal Pollution in Surface Water. *Appl. Environ. Microbiol.* **2015**, *81*, 7067–7077. [[CrossRef](#)]

27. Fuks, G.; Elgart, M.; Amir, A.; Zeisel, A.; Turnbaugh, P.J.; Soen, Y.; Shental, N. Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. *Microbiome* **2018**, *6*, 1–13. [CrossRef]
28. Guo, W.; Li, Y.; Wang, L.; Wang, J.; Xu, Q.; Yan, T.; Xue, B. Evaluation of composition and individual variability of rumen microbiota in yaks by 16S rRNA high-throughput sequencing technology. *Anaerobe* **2015**, *34*, 74–79. [CrossRef] [PubMed]
29. Allen, H.K.; Bayles, D.O.; Looft, T.; Trachsel, J.; Bass, B.E.; Alt, D.P.; Bearson, S.M.D.; Nicholson, T.; Casey, T.A. Pipeline for amplifying and analyzing amplicons of the V1–V3 region of the 16S rRNA gene. *BMC Res. Notes* **2016**, *9*, 1–6. [CrossRef]
30. Gulino, L.-M.; Ouwerkerk, D.; Kang, A.Y.H.; Maguire, A.J.; Kienzle, M.; Klieve, A.V. Shedding Light on the Microbial Community of the Macropod Foregut Using 454-Amplicon Pyrosequencing. *PLoS ONE* **2013**, *8*, e61463. [CrossRef]
31. AGRF. Diversity Profiling Sample Submission Guidelines. 2016, Volume 1. Available online: <https://www.agrf.org.au/resources/next-gen-resources/divpro-faqs> (accessed on 15 June 2018).
32. Leggett, R.M.; Ramirez-Gonzalez, R.H.; Clavijo, B.J.; Ewaite, D.; Davey, R.P. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front. Genet.* **2013**, *4*, 288. [CrossRef]
33. Staley, C.; Kaiser, T.; Lobos, A.; Ahmed, W.; Harwood, V.J.; Brown, C.M.; Sadowsky, M.J. Application of SourceTracker for Accurate Identification of Fecal Pollution in Recreational Freshwater: A Double-Blinded Study. *Environ. Sci. Technol.* **2018**, *52*, 4207–4217. [CrossRef] [PubMed]
34. Ahmed, W.; Staley, C.; Hamilton, K.; Beale, D.; Sadowsky, M.; Toze, S.; Haas, C. Amplicon-based taxonomic characterization of bacteria in urban and peri-urban roof-harvested rainwater stored in tanks. *Sci. Total. Environ.* **2017**, *576*, 326–334. [CrossRef] [PubMed]
35. Brown, C.M.; Staley, C.; Wang, P.; Dalzell, B.; Chun, C.L.; Sadowsky, M.J. A High-Throughput DNA-Sequencing Approach for Determining Sources of Fecal Bacteria in a Lake Superior Estuary. *Environ. Sci. Technol.* **2017**, *51*, 8263–8271. [CrossRef]
36. Kim, M.; Morrison, M.; Yu, Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J. Microbiol. Methods* **2011**, *84*, 81–87. [CrossRef]
37. Klindworth, A.; Pruesse, E.; Schweer, T.; Peplies, J.; Quast, C.; Horn, M.; Glöckner, F.O. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **2012**, *41*, e1. [CrossRef] [PubMed]
38. Ahmed, W.; O’Dea, C.; Masters, N.; Kuballa, A.; Marinoni, O.; Katouli, M. Marker genes of fecal indicator bacteria and potential pathogens in animal feces in subtropical catchments. *Sci. Total. Environ.* **2019**, *656*, 1427–1435. [CrossRef]
39. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.; Costello, E.K.; Fierer, N.; Peña, A.G.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335–336. [CrossRef]
40. Veca, R.; O’Dea, C.; Burke, J.; Hatje, E.; Kuballa, A.; Katouli, M. A Comparative Study of the Adherent-Invasive *Escherichia coli* Population and Gut Microbiota of Healthy Vegans versus Omnivores. *Microorganisms* **2020**, *8*, 1165. [CrossRef] [PubMed]
41. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [CrossRef]
42. Zakrzewski, M.; Proietti, C.; Ellis, J.; Hasan, S.; Brion, M.-J.; Berger, B.; Krause, L. Calypso: A user-friendly web-server for mining and visualizing microbiome–environment interactions. *Bioinformatics* **2016**, *33*, 782–783. [CrossRef]
43. Bray, J.R.; Curtis, J.T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **1957**, *27*, 325–349. [CrossRef]
44. Ahasan, S.; Waltzek, T.; Huerlimann, R.; Ariel, E. Comparative analysis of gut bacterial communities of green turtles (*Chelonia mydas*) pre-hospitalization and post-rehabilitation by high-throughput sequencing of bacterial 16S rRNA gene. *Microbiol. Res.* **2018**, *207*, 91–99. [CrossRef] [PubMed]
45. Ahasan, S.; Waltzek, T.; Huerlimann, R.; Ariel, E. Fecal bacterial communities of wild-captured and stranded green turtles (*Chelonia mydas*) on the Great Barrier Reef. *FEMS Microbiol. Ecol.* **2017**, *93*, 93. [CrossRef]
46. Chen, Z.; Hui, P.C.; Hui, M.; Yeoh, Y.K.; Wong, P.Y.; Chan, M.C.W.; Wong, M.C.S.; Ng, S.C.; Chan, F.K.L.; Chan, P.K.S. Impact of Preservation Method and 16S rRNA Hypervariable Region on Gut Microbiota Profiling. *mSystems* **2019**, *4*, e00271-18. [CrossRef] [PubMed]
47. Cardoso, D.C.; Sandionigi, A.; Cretoiu, M.S.; Casiraghi, M.; Stal, L.; Bolhuis, H. Comparison of the active and resident community of a coastal microbial mat. *Sci. Rep.* **2017**, *7*, 1–10. [CrossRef]
48. Auer, L.; Mariadassou, M.; O’Donohue, M.; Klopp, C.; Hernandez-Raquet, G. Analysis of large 16S rRNA Illumina data sets: Impact of singleton read filtering on microbial community description. *Mol. Ecol. Resour.* **2017**, *17*, e122–e132. [CrossRef]
49. Chen, S.; Huang, T.; Zhou, Y.; Han, Y.; Xu, M.; Gu, J. AfterQC: Automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinform.* **2017**, *18*, 91–100. [CrossRef]
50. Barb, J.J.; Oler, A.J.; Kim, H.-S.; Chalmers, N.; Wallen, G.R.; Cashion, A.; Munson, P.J.; Ames, N.J. Development of an Analysis Pipeline Characterizing Multiple Hypervariable Regions of 16S rRNA Using Mock Samples. *PLoS ONE* **2016**, *11*, e0148047. [CrossRef] [PubMed]
51. Price, J.; Ledford, S.H.; Ryan, M.O.; Toran, L.; Sales, C.M. Wastewater treatment plant effluent introduces recoverable shifts in microbial community composition in receiving streams. *Sci. Total. Environ.* **2018**, *613–614*, 1104–1116. [CrossRef]
52. Lott, M.J.; Hose, G.; Power, M. Parasitic nematode communities of the red kangaroo, *Macropus rufus*: Richness and structuring in captive systems. *Parasitol. Res.* **2015**, *114*, 2925–2932. [CrossRef]

53. Callahan, B.J.; Mcmurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)] [[PubMed](#)]
54. Ellis, R.J.; Bruce, K.D.; Jenkins, C.; Stothard, R.; Ajarova, L.; Mugisha, L.; Viney, M.E. Comparison of the Distal Gut Microbiota from People and Animals in Africa. *PLoS ONE* **2013**, *8*, e54783. [[CrossRef](#)] [[PubMed](#)]
55. Moon, C.D.; Young, W.; MacLean, P.H.; Cookson, A.L.; Bermingham, E.N. Metagenomic insights into the roles of Proteobacteria in the gastrointestinal microbiomes of healthy dogs and cats. *MicrobiologyOpen* **2018**, *7*, e00677. [[CrossRef](#)]
56. Brooks, J.P.; Edwards, D.J.; Harwich, M.D., Jr.; Rivera, M.C.; Fettweis, J.M.; Serrano, M.G.; Reris, R.A.; Sheth, N.U.; Huang, B.; Girerd, P.; et al. The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.* **2015**, *15*, 1–14. [[CrossRef](#)]
57. Sinha, R.; Abnet, C.C.; White, O.; Knight, R.; Huttenhower, C. The microbiome quality control project: Baseline study design and future directions. *Genome Biol.* **2015**, *16*, 1–6. [[CrossRef](#)] [[PubMed](#)]
58. Vishnivetskaya, T.A.; Layton, A.C.; Lau, M.C.Y.; Chauhan, A.; Cheng, K.R.; Meyers, A.J.; Murphy, J.R.; Rogers, A.W.; Saaranya, G.S.; Williams, D.E.; et al. Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. *FEMS Microbiol. Ecol.* **2014**, *87*, 217–230. [[CrossRef](#)] [[PubMed](#)]
59. Vahjen, W.; Pieper, R.; Zentek, J. Bar-Coded Pyrosequencing of 16S rRNA Gene Amplicons Reveals Changes in Ileal Porcine Bacterial Communities Due to High Dietary Zinc Intake. *Appl. Environ. Microbiol.* **2010**, *76*, 6689–6691. [[CrossRef](#)]