Diversity and community succession of the essential soil microbial phylum Gemmatimonadetes in the context of a simulated urine patch

Erin Knox

A report submitted in partial fulfilment of the Degree of Bachelor of Science with Honours

Department of Microbiology and Immunology

University of Otago

Dunedin, New Zealand

October 2021

Acknowledgements

Firstly, thank you to my supervisor Dr. Sergio Morales for your much needed support and guidance throughout the year, the home baking, and introducing me to the world of bioinformatics. I hope I haven't caused you too many grey hairs. Thank you to the whole of Morales lab for your guidance throughout the year. In particular thanks to Syaliny for looking after me this year and guiding me every step of the way, and to Scott for showing me the ropes of bioinformatics and coding and answering all of my questions.

Thanks to David Rex for allowing me to use his samples and Syaliny Ganasmurthy for letting me use a subset of the metagenomic bins generated for her study.

Thanks to Mum, Dad, Reuben and Bruce for all of your support and encouragement and for putting up with my chats about coding. Thank you to my friends and flatmates for your encouragement and care during a rough year.

Lastly thank you to the Department of Microbiology and Immunology and the University of Otago for having me. I have learnt so much this year and it was a great experience.

Abstract

The increasing demand on agricultural production due to the growing world population, has resulted in large amounts of the potent greenhouse gas nitrous oxide (N_2O) being released into our atmosphere. In New Zealand, the majority of our greenhouse gas emissions come from agriculture and nitrogen loading through the urine of grazing ruminants results in the release of greenhouse gasses into the atmosphere. The addition of urea to the soil in what is known as a urine patch, greatly disrupts the soil microbial community leading to community succession and a decrease in alpha diversity.

Gemmatimonadetes is a bacterial phylum that is highly abundant in soil microbial communities, however, due to culturing limitations very little is known about the functional potential and true diversity of the phylum. We analysed metagenome assembled genomes from the phylum Gemmatimonadetes that were sequenced and assembled from soil samples treated with urea +/- antifungal and antibacterial inhibitors. We looked at the phylogenetic novelty and relationship of the metagenomes using Average Nucleotide Identity and Maximum likelihood phylogenetic trees from which we identified seven novel Gemmatimonadetes species. The functional potential of these species was then analysed using enrichM and showed a lot of diversity in functional potential between clusters. The species referred to as Cluster 1 was found to be of particular interest. This species was streamlined, had the highest pathway completeness for antimicrobial resistance modules, and was positively correlated with the addition of urea.

Community succession occurred within the Gemmatimonadetes community in the samples treated with urea. The negative control which did not have urea added kept a steady population. Therefore, we concluded that community succession occurred in the soil Gemmatimonadetes community within the context of a simulated urine patch.

Table of Contents

List of Figures

List of Tables

1. Introduction

Gemmatimonadetes is a bacterial phylum consisting of heterotrophic, gram-negative, rod-shaped bacteria. There are currently two classes within the Gemmatimonadetes phylum, Gemmatimonadetes and Longimicrobia (Pascual et al., 2016). Gemmatimonadetes are able to grow under aerobic and anaerobic conditions (Takaichi et al., 2010; Zhang et al., 2003) and are found in a large range of environments including soils, wastewater, permafrost, streams and oceans (Zeng et al., 2016). Gemmatimonadetes is one of the top nine most abundant soil bacteria, making up around 2% of total soil bacteria (Zeng et al., 2021). As a result of this it is expected that the phylum plays an important role in the function of the microbial community. Despite this notable abundance, not much is known about the role of these bacteria in the environment. There are very few cultured isolates of Gemmatimonadetes and these culturing limitations have resulted in the Gemmatimonadetes phylum being understudied (Pascual et al., 2016; Zeng et al., 2016; Zhang et al., 2003). The rise of metagenomic analysis has allowed for further study of the phylum, however the functions of different species of Gemmatimonadetes remain largely unknown.

1.1 Gemmatimonadetes functions and previously identified traits

The Gemmatimonadetes phylum has been found to have a wide range of functions present in different species within the phylum. These include phototrophy and methylotrophy in specific strains (Bay et al., 2021; Butterfield et al., 2016; Zeng et al., 2021). Gemmatimonadetes have also been found to have a preference for arid soils, suggesting that the bacteria may have adaptations to survive in environments with low moisture (M et al., 2011; Zeng et al., 2021). The difference in the functional potential displayed by different species of Gemmatimonadetes suggests that the phylum is very diverse. This could mean that the phylum may contain

specialists and generalists that perform different functions in the environment and are present in the soil at different times in order to utilise different niches. Gemmatimonadetes have been shown to be sensitive to carbon levels in the environment and have also been found to carry genes involved in carbon cycling (Baker et al., 2015) which may suggest that some Gemmatimonadetes species are involved in carbon fixation (Baker et al., 2015; Butterfield et al., 2016).

1.2 Carbon and nitrogen cycling

Microbial communities play an important role in the nitrogen and carbon cycles and therefore have a notable influence on the production and consumption of greenhouse gasses (Cavicchioli et al., 2019; Kolton et al., 2019). With the ever-increasing global population, agricultural demands are also growing meaning more livestock and increased fertiliser use resulting in large nitrogen loads being added to soils which has considerable consequences for greenhouse gas production (Tilman et al., 2011). This is of particular importance in New Zealand where agriculture makes nearly half of the country's total greenhouse gas emissions, with dairy farming in particular contributing almost three quarters of total agricultural emissions in New Zealand with the next largest source being nitrous oxide from nitrogen added to soils as fertiliser (*Agriculture Emissions and Climate Change | Ministry for the Environment*, n.d.). Intensive dairy farming is of particular concern as cows urinate on the paddocks causing nitrogen to leech into the soil creating a concentrated area of nitrogen known as a urine patch. The nitrogen leaching from the urine of grazing ruminants has been determined to be the largest cause of nitrogen losses in the environment, far greater than the nitrogen leaching caused by the addition of fertilisers to agricultural soils (Cichota et al., 2018; Selbie et al., 2015). Nitrous oxide (N_2O) is long lasting in the atmosphere and is the third most noxious greenhouse gas (Core Writing Team et al., 2014). Nitrogen loss due to urine patches results in increased

production of N_2O, as well as the indirect greenhouse gas ammonia, resulting in rising levels of these gasses in the atmosphere which has a marked impact on climate change (Hutchings et al., 2007; Selbie et al., 2015; Zaman et al., 2009).

1.3 Microbial composition and greenhouse gases

Due to the important role microbes play in the cycling of chemical compounds, including the production and consumption of greenhouse gasses, the composition of the microbial community plays a large role in what different compounds are converted to (Cavicchioli et al., 2019; Kolton et al., 2019). This means that there is a possibility that, with increased understanding, microbial communities could be manipulated to contain a greater proportion of microbial species that transform nitrogen into innocuous forms, such as dinitrogen gas, to reduce harmful emissions (Cavicchioli et al., 2019; Jansson & Hofmockel, 2019; Kuypers et al., 2018).

1.4 Metagenome assembly

Metagenome assembly is an important technique as it allows for the study of microorganisms that are not easily cultured and therefore have been overlooked in terms of their metabolic potential and roles in the community structure. Metagenome assembled genomes (MAGs) are draft genomes that are assembled from environmental samples. DNA from the environmental sample undergoes whole sequence shotgun sequencing. The short reads from the sequencing are then assembled into longer contigs which are sorted into bins, or groups, so that each bin contains contigs from the same organism. The bins undergo quality control to examine completeness and contamination, with bins of high enough quality considered as draft genomes or MAGs (Alneberg et al., 2018; Wilkins et al., 2019). The bins represent a single organism

and can be taxonomically classified and used to study the microbial community in the environment of interest.

1.5 ANI and maximum likelihood trees

Average nucleotide identity (ANI) is a method of nucleotide level, pairwise comparison of genomes. ANI match percentage describes the similarity of the genomes and can be used to cluster organisms of the same species together. An ANI score of >95% indicates same species, between 80-94% indicates closely related organisms, likely same family but not same species (Jain et al., 2018; Yoon et al., 2017). ANI has been shown to be a reliable method for calculating identity between a pair of genomes (Jain et al., 2018). It is also a high throughput method but remains relatively fast to run and can be easily visualised in a heatmap. The only con of ANI is that it does not calculate identities that are lower than 75% similarity(Jain et al., 2018b). Phylogenetic maximum likelihood (ML) trees are another high-resolution method of analysing the phylogenetic relationships between different organisms. Click or tap here to enter text.Maximum likelihood trees are made by recreating the phylogenetic tree may times to determine which phylogenetic tree as the greatest likelihood (Stamatakis, 2014). This information can then be visualised using a phylogenetic tree viewer, such as ITOL (Letunic & Bork, 2021). The nodes of the tree can show clustering of species with the distance between species equating to how similar or dissimilar the genomes are.

1.6 Functional potential analysis

The functional potential of different organisms can be analysed by looking at the completeness of metabolic pathways found in the genome and which particular genes are present. To do this a functional potential analysis tool, enrichM (Boyd, Joel A. ; Woodcroft, Ben J.; Tyson, 2019), is used to compare the metagenome sequences against gene and metabolic pathway databases.

We compared the sequences to the Kegg database (Kanehisa et al., 2016). Protein coding genes are then identified, classified and annotated. These annotations and the number of hits per gene for each genome or the average completeness of a pathway can then be used to perform statistical analysis of the functional potential of that organism (Song et al., 2021). By doing this we can compare different clusters and determine, if there is community succession occurring, whether the clusters of Gemmatimonadetes have different roles or whether they are different species that perform the same functions but differ in genome composition (redundant species).

1.7 Community succession and time series

Time series analysis allows for community succession to be studied. With changing environmental conditions, the microbial population structure changes (Coenen et al., 2020; Faust et al., 2015). By looking at the abundance of species under particular conditions over time we can gather a greater idea of what particular species' role in the community may be. We used time series analysis to determine whether or not Gemmatimonadetes community succession was occurring following the addition of urea to the soil samples. Large changes to environmental conditions trigger community successions as the microbes that are better adapted for the changed conditions will outcompete the other organisms, leading to a shift in the relative abundance of species in the community (Ganasamurthy et al., 2021).

1.8 Origin of samples and past studies

The samples for this project came from sandy loam soil samples that were collected for a previous study (Rex et al., 2018) from a representative area of pasture on the Lincoln University Dairy Farm, New Zealand. The rationale and methods of the origin study is explained in greater detail in the methods section Click or tap here to enter text.The study by Rex et al. used five different treatment types on the soil. An antibacterial treatment, antifungal treatment,

antibacterial + antifungal treatment, and a positive (+urea) and negative (ionised water). These treatments are explained in further detail in the methods section of this paper. Urea was then added to all of the samples except for the negative controls to replicate the effect of a urine patch. The soil was then destructively sampled at different time points (Day 9, Day 15, Day 21, Day 27, Day 33, Day 51). DNA was then extracted from the sampled and were then sequenced using 2x150 Illumina Novaseq. Community succession was seen within the prokaryotic community with the replacement of stable native species and a decrease in alpha diversity following nitrogen deposition (Ganasamurthy et al., 2021; Rex et al., 2018; Samad et al., 2017). Many different soil bacteria, archaea, and fungi were identified through metagenomic assembly and were binned. The bins from the Gemmatimonadetes phylum were chosen as the focus of our project. The Gemmatimonadetes bins had a high abundance of medium-quality draft genomes and phylum Gemmatimonadetes is known to be an abundant soil microbe although it has been identified in a wide range of environments. Despite this, much of the phylum remains unclassified and little is known about the true range of functions and diversity that can be found within the phylum. Gemmatimonadetes has been found to respond positively to the addition of urea (Ganasamurthy et al., 2021; Samad et al., 2017) therefore we hypothesised that we would identify community succession occurring over time within the Gemmatimonadetes community. This is important as there is little in the existing literature exploring Gemmatimonadetes role in nitrogen cycling in soil communities. Additionally, some species of Gemmatimonadetes have been found to contribute to nitrogen cycling, particularly through the *nosZ* gene which is found in nitrous oxide reducers. Taking into account all of these factors, we decided that the Gemmatimonadetes phylum would be a good candidate for deeper metagenome analysis under the context of a simulated urine patch.

1.9 Aims and hypotheses

The goal of this project was to characterize the poorly understood yet abundant phylum, Gemmatimonadetes. We analysed Gemmatimonadetes MAGs in order to better understand their role in the soil microbial community, particularly under the conditions of a simulated urine patch. We hypothesized that the changing environmental conditions due to the addition of urea would impact on the Gemmatimonadetes community structure and that some Gemmatimonadetes species may have adaptations that provide an advantage during periods of high nitrogen availability. Due to the diversity seen in the few studies focusing on members of the Gemmatimonadetes phylum, we also hypothesized that there would be significant differences in the functional potential of the Gemmatimonadetes species found.

2. Methods

2.1 Sample Collection and experimental design

Sample collection and experimental design was as per (Rex et al., 2018) . Briefly, soil samples were taken from a pasture on the Lincoln University dairy farm, New Zealand (43°38′33.73“S, 172°27′40.38″E) (Rex et al., 2018). The pasture had not been grazed by livestock or had fertilizer added to the soil for 12 months prior to sample collection to prevent nitrogen loading in the samples which could otherwise confound the results. The pasture contained perennial rye grass (*Lolium perenne L.*) and white clover (*Trifolium repens L.*) (Rex et al., 2018). The top 10cm of soil (Soil type: sandy loam, Soil classification: Typic Immature Pallic soil; soil pH: 5.9) (Rex, 2018)was sampled and then passed through a 4mm sieve to remove any debris such as rocks and plant matter. 50g of dry sieved soil was placed into 250ml jars, water was then added to 50% water holding capacity (Rex et al., 2015) and maintained throughout the experiment. Originally, 200 jars were prepared consisting of 5 treatments, spread across 8 destructive sampling times (Day 3,9,15,21,27,33,42 and 51) and replicated 5 times (Rex et al., 2018). Of these 5 replicates, 4 replicates were used for gas sampling and inorganic N measurements (5 treatments x 4 reps x 7 time points = 140 jars) and 1 replicate was used for molecular analysis [DNA] (5 treatments x 1 rep x 7 time points = 35 jars). Day 3 was excluded from molecular and gas analysis. The treatments were set up as follows: 1) Negative control (deionized water), 2) Positive control (+urea), 3) Antibacterial treatment (urea + streptomycin + penicillin), 4) Antifungal treatment (urea + cycloheximide) and 5) a combined Antibacterial + Antifungal treatment (Ganasamurthy et al., 2021; Rex et al., 2018). The urea treated samples were 15N enriched urea (50% atm) and was applied as a solution at 2141 kg urea ha−1 dry soil. Inhibitors were applied at 5 mg g^{-1}, 8 mg g^{-1}, and 13 mg g^{-1} for Antibacterial, Antifungal, and Antibacterial + Antifungal treatments respectively. Dry powdered forms were used for the application of streptomycin and cycloheximide whereas penicillin was dissolved in deionized

water and applied as a solution. The inhibitor concentration and efficacy were determined in a pilot study (Rex et al., 2018). All soils had the urea solution applied at Day 0 except for the negative control. All samples were then placed in an incubator with lids open in dark conditions at a temperature of 23% and 55% relative humidity for the duration of the experiment.

2.2 Destructive Sampling and Nucleic acid extraction

Destructive sampling and nucleic acid extraction were performed as per Rex et al. (2018) and Ganasamurthy et al. (2021), for each respective time point (Day 9, 15, 21, 27, 33, 42 and 51). 48hrs prior to destructive sampling, inhibitors were added to the respective samples, mixed using a spatula for 90 s (including control samples) and placed back in the incubator (Rex et al., 2015). After 48 hrs the soils were subsampled. At every destructive sampling event, DNA was extracted in triplicates from each treatment and analysed separately. 0.25g of soil from each sample was extracted using the Dneasy Powersoil Extraction Kit (Qiagen) following manufactures guidelines with some modifications (Ganasamurthy et al., 2021). A Genogrinder (1600 MiniG SPEX SamplePrep) set at 1500 strokes/min was used for two rounds of bead beating with a 1 min pause in between each round. The DNA was then eluted to 100µl and a Nanodrop Spectrophotometer, ND-1000 (Thermo Scientific) was used to determine DNA quantity and purity. DNA was stored at -80 °C until further analysis.

2.3 Short-read Illumina sequencing

Samples for shotgun sequencing (Illumina Novaseq) were selected from time points of interest based on the findings of amplicon data using the same soils (Ganasamurthy et al., 2021). The amplicon data was processed using QIIME (version 1.9.1) with parameters set to a minimum read length of 75 bp and sequences that were 0.75 of the total read length, a Phred quality score of 3, with no ambiguous bases allowed in a sequence (Bokulich et al., 2013). Based on the

amplicon data, Day 9 was identified as the most sensitive time point. Therefore, 3 individual replicates from 5 treatments were selected from Day 9 for shotgun sequencing (n=15) while the remaining samples were pooled by treatment and time point (n=35). A total of 50 samples were sent for short-read 2x150 bp Illumina sequencing to generate metagenomes to analyse the metabolic potential of organisms within the simulated urine patch.

2.4 Metagenome assembly and binning

The Illumina sequencing reads underwent quality control prior to assembly. All quality control steps were processed using paired-reads. First the overlapping reads were grouped into clumps using the Clumpify tool from BBTools (Bushnell, 2021), this was done to condense repetition in the sequence files to make future processing more efficient. Adapter trimming was done using the tool BBDuk from BBtools (Bushnell, 2021) (parameters: ktrim=r, k=23, mink=11, hdist=1, tpe, tbo) to remove artifacts left by the Illumina sequencing. BBDuk is also used for targeted removal of PhiX gene which is used by Illumina as a control library. This was done using a reference sequence database for PhiX (parameters: k=31, hdist=1). Contaminant removal to remove any artifacts of the human genome was done using the HG19 human genome as a reference database. We used the BBMap tool from BBTools (Bushnell, 2021) (parameters: minid=0.95, maxindel=3, bwr=0.16, bw=12, quickmatch, fast, minhits=2, qtrim=rl, trimq=10, untrim). Clean sequencing reads were the output from this step. The forward and reverse reads (not the clean reads) are then merged by overlap and non-overlap by kmer. BBMerge from BBtools (Bushnell et al., 2017) was used for merging (parameters: t=30, prealloc, rem, k=5, extend2=50, ecct) and this produced three output files: (1) merged reads, (2) unmerged forward reads, (3) unmerged reverse reads. These are the files that were used for error correction and assembly.

Error correction was then done using the merged reads and corresponding unmerged forward and reverse reads. The tool that was used was BayesHammer from the metaSPAdes pipeline (Nurk et al., 2017) (parameters: -k 21, 33, 55, 77 -phred-offset 33). Error corrected versions of all three input files were produced, which were then used as the input files for the assembly. The metaSPAdes (Nurk et al., 2017) tool was used for assembly (parameters=: -k 21, 33, 55, 77 –phred-offset 33). An assembled scaffold file was output from the assembly process. In preparation for binning, alignment was carried out using the assembled scaffolds and unmerged clean reads. First, BWA (H. Li, 2013) was used to create an index for the alignment. Next, the assembled scaffolds were aligned to the cleaned reads using BWA (H. Li, 2013). The output SAM file was then converted to a BAM file and sorted using samtools (Danecek et al., 2021). Lastly, a depth file is generated from the sorted BAM using MetaBAT (Kang et al., 2019).

The contigs then underwent binning to sort contigs that are highly likely from the same organism into groups (bins). The scaffold and depth file were used for binning and the tool MetaBAT2 (Kang et al., 2019) was used (parameters= --unbinned, -t 8). The output of this step was bins and a file name unbinned which contained un-grouped contigs. Assessment was then done using multiple processes from the tool CheckM (Parks et al., 2015) to assess the quality of our bins. To determine quality information including percent completion, percent contamination, and strain heterogeneity; the Lineage_wf process was used. The Coverge -> Profile module was used to estimate how populous each bin was in the community and Ssu_finder was used to identify and locate 16S genes. Lastly, taxonomy was assigned to each bin using GTDB-Tk (Chaumeil et al., 2020). This tool was used as it has a wide database of taxonomy and uses ANI to determine the closest match.

2.5 Statistical analysis

Average Nucleotide Identity (ANI) was calculated using the FastANI package (Jain et al., 2018) with the query list to reference list option. Phylogenetic trees were created using the PhyloPhlAn integrated pipeline (Asnicar et al., 2020) and refined and bootstrapped using the maximum likelihood tree tool RaXmL-NG (Stamatakis, 2014). Visualization of phylogenetic trees was done using the Interactive Tree of Life (ITOL) online tool (Letunic & Bork, 2021). Genes and pathways were classified using enrichM (Boyd, Joel A. ; Woodcroft, Ben J.; Tyson, 2019) and non-specific ORFs were called using OrfM (Woodcroft et al., 2016). DIAMOND (Buchfink et al., 2021) was used to align enrichM classified ORFs to the clean sequencing reads.

All statistical analyses were performed using RStudio version 1.4.1103 (RStudio Team, 2021) and visualized using the ggplot2 package (Wickham et al., 2016) unless otherwise stated. Mean and standard error for pathway completeness analysis were calculated using the ddply function of the plyr tool (Wickham, 2011). Significantly changing pathways were identified using the edgeR package (McCarthy et al., 2012; Robinson et al., 2010). P values ($p < 0.05$) were required for significance and the results were corrected for false discovery rates ($FDR < 0.05$). Significantly changing pathways required a two-fold increase or decrease in mean pathway completeness between clusters. The bash and R scripts are available on GITHUB (https://github.com/erin-knox/Honours_Thesis_2021).

3. Results

3.1 Description of Gemmatimonadetes dataset

General information on the quality and characteristics of all 96 Gemmatimonadetes bins in the dataset was used to create an informative graph to aid the decision of whether to refine the dataset going forward or not (Figure 1). The average completeness for the bins was 83.63% and ranged from 50-98.78%. The average contamination of the bins was 3.82 and ranged between 0-9.56. In order to maintain a good quality of bins for analysis whilst keeping a good number of bins, we decided to trim the Gemmatimonadetes dataset to only include bins with completeness >75% and contamination below 6.0 leaving us with medium-quality draft MAGs (Bowers et al., 2017) . This left 63 bins in the new trimmed dataset for analysis. From now on when Gemmatimonadetes bins are referred to in this paper, this is referring to the trimmed dataset of 63 bins. The bins range in size from 1.54 Mbp to 6.9 Mbp, and the average bin size is 3.57 Mbp. Bin size tends to correlate with the completeness of the bins as would be expected. Contamination is not biased to completeness, meaning the contamination of the bins is random.

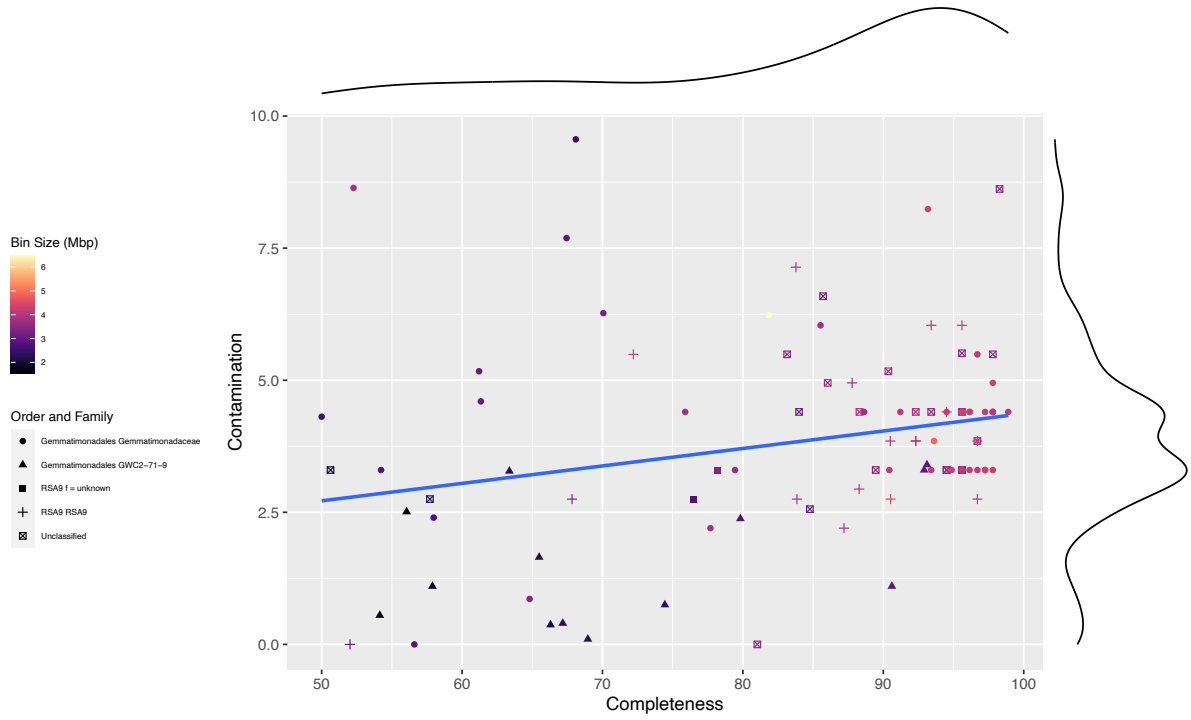

Figure 1: Description of all 96 Gemmatimonadetes bins from the dataset. The density plot at the top of the graph shows the spread of percentage completeness of the bins. The density curve on the right of the graph indicates the spread of the percentage of contamination of the bins.

3.2 Seven novel Gemmatimonadetes clusters, representing species, were identified from our bins.

Genome sequences of the Gemmatimonadetes bins were compared using average nucleotide identity (ANI). ANI uses a pairwise comparison of all the genomes and assigns an ANI match percentage to the pairs based on how similar they are. An ANI match of greater than 95% is generally accepted to indicate that the two genomes are from the same species (Jain et al., 2018b). We identified seven distinct clusters representing seven different species (Figure 2). Between clusters, there is a low identity percentage, below 85% ANI match. From this we hypothesised that each cluster represented a different species from the Gemmatimonadetes phylum. The clusters all vary in size from 23 bins in the biggest cluster (Cluster 1) to only one bin in the smallest cluster (Cluster 7). This potentially indicates that the clusters may have different life strategies e.g. Generalists vs specialists, and therefore the Gemmatimonadetes

may have undergone community succession. The grey squares represent results that were below the 75% match threshold of the fastANI tool, meaning that these results were not reported.

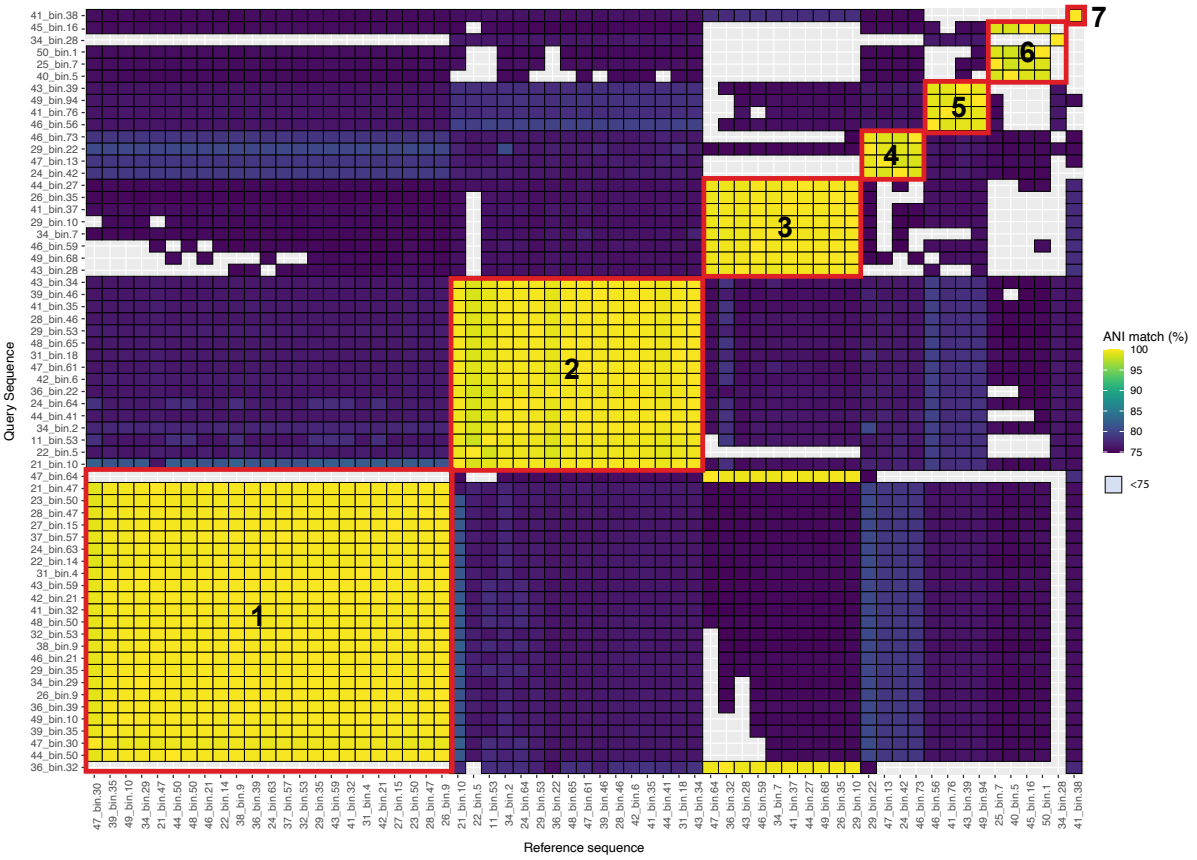

Figure 2: Average Nucleotide Identity (ANI) heatmap. Each sequence underwent a pairwise comparison with all the sequences to determine the ANI match percentage for the pair of bins. This percentage represents the percentage identity of the bins.

Reference Gemmatimonadetes sequences were obtained from JGI (Nordberg, H et al., 2014). All reference sequences used were full genomes sequenced from cultured isolates. The phylogenetic analysis was repeated with the reference sequences to analyse the novelty of our bins within the phylum.

The heatmap containing the reference genomes in addition to the bins shows the same general clustering of bins with the basis of the seven clusters remaining, although the smaller clusters

are less distinct (Figure 1Figure 3). The reference sequences were not grouped into any of the previously identified clusters except for the Gemm-3 bacterium (*Gemmatimonadetes-3* T17) which appeared to possibly be a part of Cluster 6. This indicated a wide range of diversity

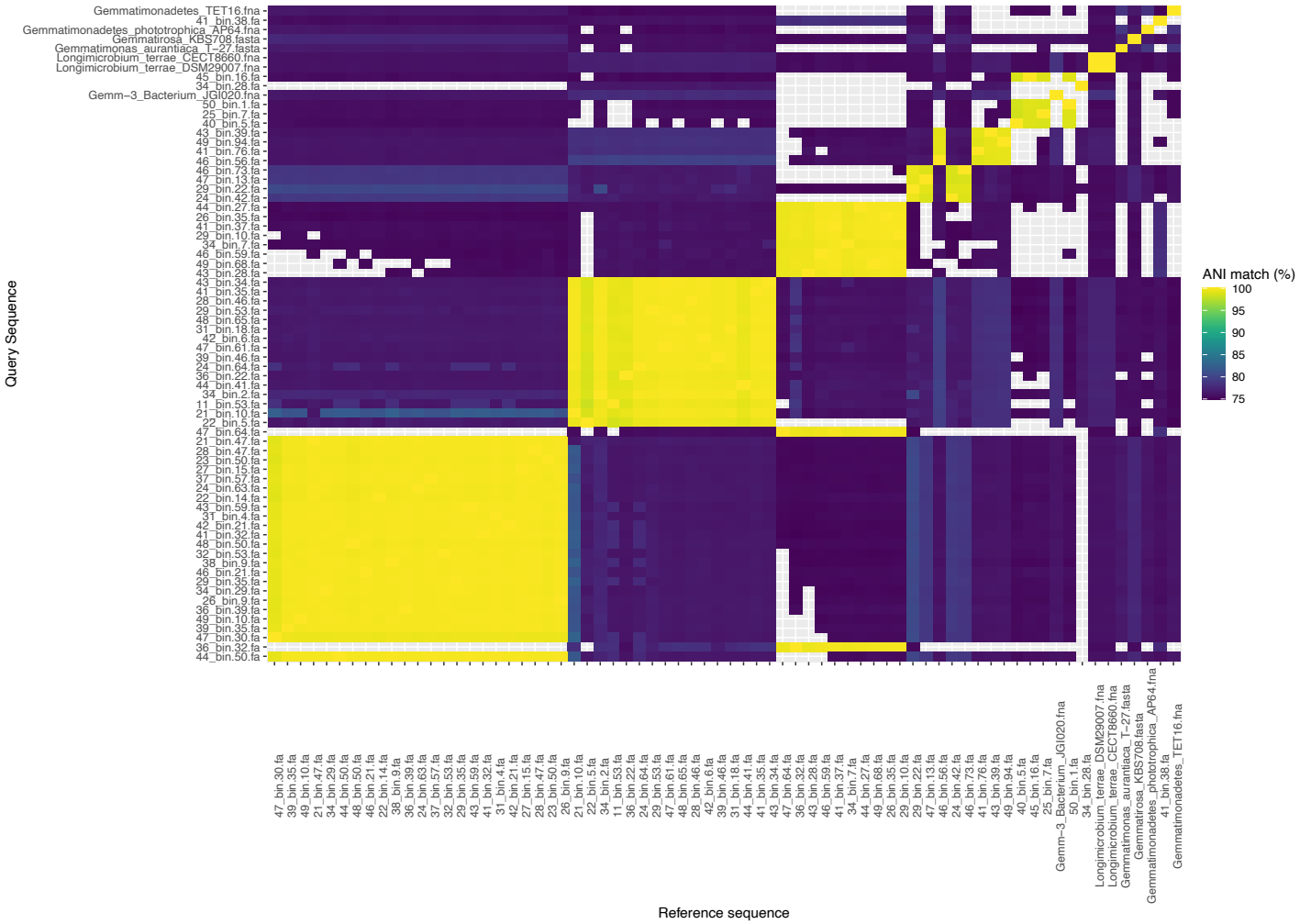

Figure 3: Average Nucleotide Identity (ANI) heatmap with reference sequences from cultured isolates in addition to the Gemmatimonadetes bin sequences. Each sequence underwent a pairwise comparison with all the sequences to determine the ANI match percentage for the pair of bins. This percentage represents the percentage identity of the bins.

within the phylum and showcased the novelty of our bins. This also confirmed that the species seen in our clusters have not been cultured.

A maximum likelihood phylogenetic tree was created to confirm the ANI results and further investigate the phylogenetic relationship between the Gemmatimonadetes bins (Figure 4). The

phylogenetic tree showed distinct division of the genomes into clades that matched the seven clusters seen in the ANI heatmap (Figure 2). The bootstrap value at the node where each clade branches off is 100 for all seven groups, suggesting that the phylogenetic arrangement is strongly supported. The smallest distance between separate clusters is 0.407 between Cluster 4 and Cluster 1 meaning there is notable genetic distance between all of the clusters. Within the clades containing bins from Cluster 1 and Cluster 2 respectively, the bootstrap values are much lower with some values below 50. This indicates that the phylogenetic arrangement within these clades is not strongly supported therefore, the bins inside the clades are likely highly genetically similar.

The Phylogenetic tree was made again to include the reference genomes (Figure 5). The same seven distinct clades are seen in the tree, even with the addition of the reference sequences. The reference sequences clustered into two groups separate from the clusters. The bootstrap values for the different nodes representing the clusters as well as for the reference groups were 100 for each of the nodes except for the node branching off to Cluster 1 which had a value of 59. The results seen in the ANI heatmap and the phylogenetic tree together indicate that the Gemmatimonadetes clusters are phylogenetically distinct from the cultured Gemmatimonadetes reference sequences.

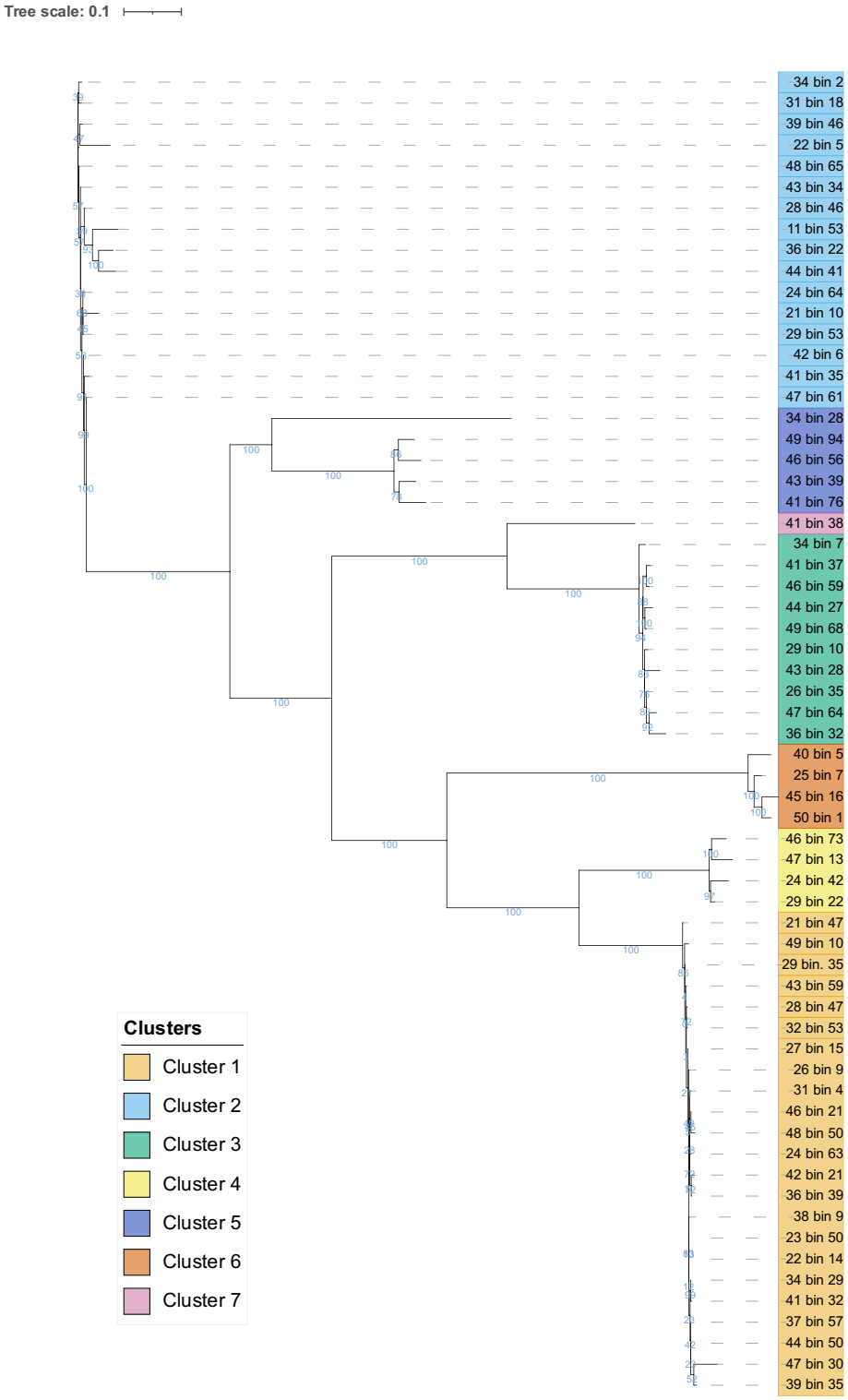

Figure 4: Maximum Likelihood Phylogenetic Tree of Gemmatimonadetes bins. The phylogenetic tree was created using the PhyloPhlAn tool and was then refined and bootstrapped using RaxML. Clades are coloured to indicate the cluster that it corroborates with on the ANI heatmap. Bootstrap values are displayed with blue text and were based on 200 replicates.

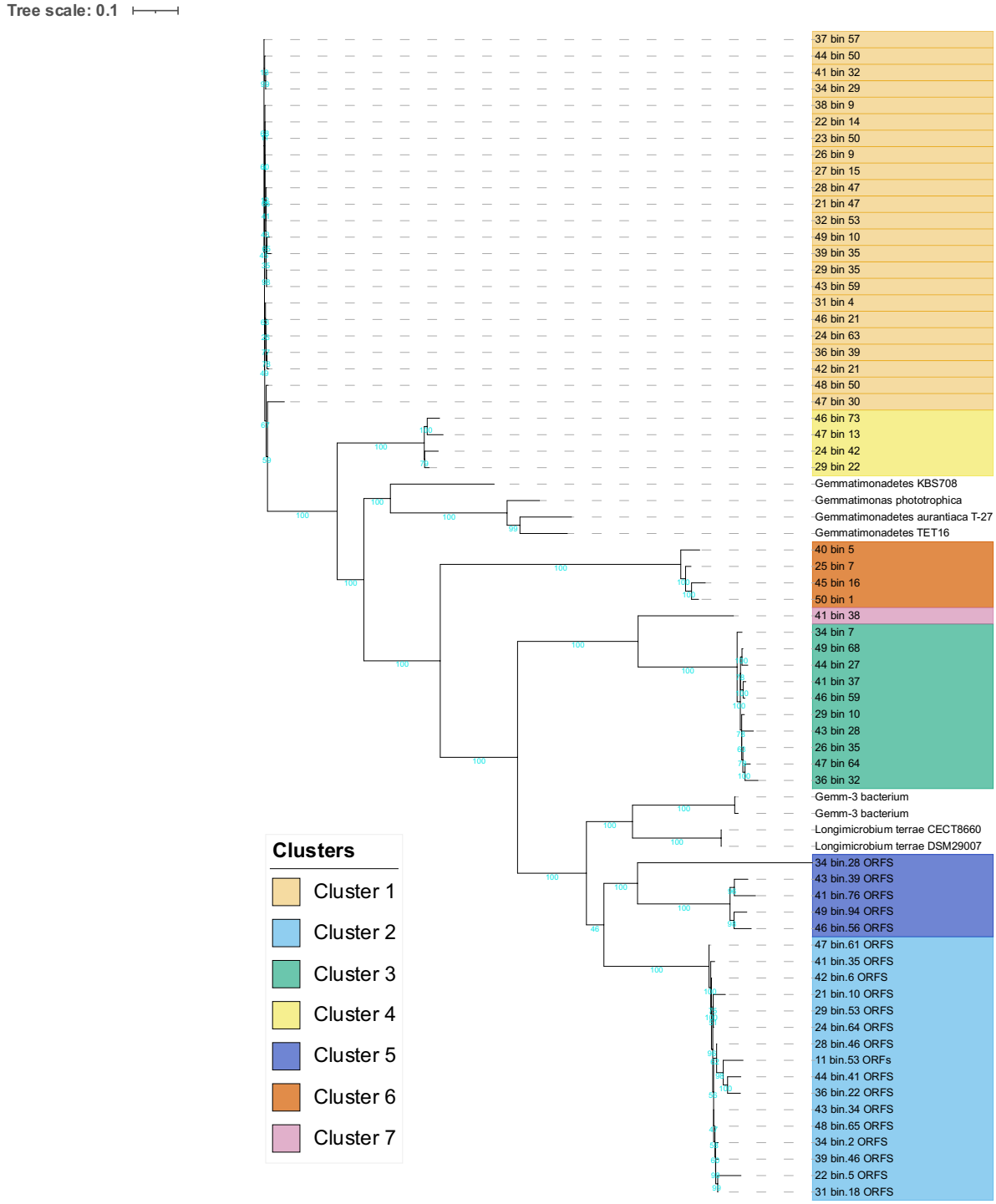

Figure 5: Maximum Likelihood Phylogenetic tree containing the Gemmatimonadetes bins and reference genomes from cultured isolates. The phylogenetic tree was created using the PhyloPhlAn tool and was then refined and bootstrapped using RaxML. Clades are coloured to indicate the cluster that it corroborates with on the ANI heatmap.

3.3 The functional potential of the Gemmatimonadetes clusters showcases the diversity within the phylum

The functional potential of the Gemmatimonadetes clusters was analysed by looking at expected vs found genes and analysing the completeness of the functional pathways found in the clusters, divided into overarching categories from the Kegg Orthology (Kanehisa et al., 2016).

We analysed the presence of different genes involved in key metabolic pathways in order to determine if there was any large difference in abundance between clusters (Table 1). Most notably, Cluster 1 and Cluster 2 had the *nosZ* gene present in all bins in the cluster whereas this gene was absent from all other clusters except for Cluster 5 which has the gene present in 60% of its bins. This is particularly interesting considering the urine patch context of our study as *nosZ* is involved in nitrous oxide reduction.

Table 1: Mean abundance of genes involved in key metabolic processes per cluster. The total number (n) of bins in each cluster used to calculate the mean were: 1 = 23, 2 = 16, 3=10, 4=4, 5=5, 6=4, 7=1.

Category	Gene	1	2	3	4	5	6	7
Sulfur	**fcc**	0	0	0	0	0	0	0
	soxb	0	0	0	0	0	0	0
	sqr	0	0	0	0	0.2	0	0
Carbon	**aclb**	0	0	0	0	0	0	0
	mcr	0	0	0	0	0	0	0
	rbcl	0	0	0	0	0	0	0
	sdha	1	0.875	0.9	1	1	0.5	1
Gas	**all_hyd**	1	1	1	0	0.8	0	1
	coxl	0	0	0	0	0	0	0
	pmoa	0	0	0	0	0	0	0
New	**cyc2**	0	0	0	0.25	0	0	0
	rdha	0	0	0	0	0	0	0
	norb	0.97	1	0	1	0	0	0
Nitrogen	**amoa**	0	0	0	0	0	0	0
	napa	0	1	0	0	0	0	0
	narg	0	0	0	0	0	0	0
	nifh	0	0	0.6	0	0	0	0
	nirk	1	0	0	1	0.6	0	1
	nirs	0	0	0	0	0	0	0
	nosz	1	1	0	0	0.6	0	0
	nrfa	0	0	1	0	0	0	1
	nxra	0	0	0	0	0	0	0
ATPase	**phoa**	0	0	0	0	0	0	0
	phod	1	0	0	1	0.6	0	0
	phox	0	0	0	0	0	0	0
Phosphonate	**pepm**	0	0	0	0	0	0	0
	phna	0	0	0	0	0	0	0
	phni	0	0	0	0	0	0	0
	phnj	0	0	0	0	0	0	0
	phnw	0.83	1	1	0	1	1	1
	phnx	0	0	0	0	0	0	0
	phnz	1	0	0	0	0	0	0
	pph	0	0	0	0	0	0	0

The mean pathway completeness for each cluster was calculated for different categories of pathways taken from the module categories from Kegg Orthology (Kanehisa et al., 2016). In general, there is a downwards trend in completeness with cluster one having the highest average

completeness and cluster seven having the lowest (Figure 6). Typically, one cluster will have

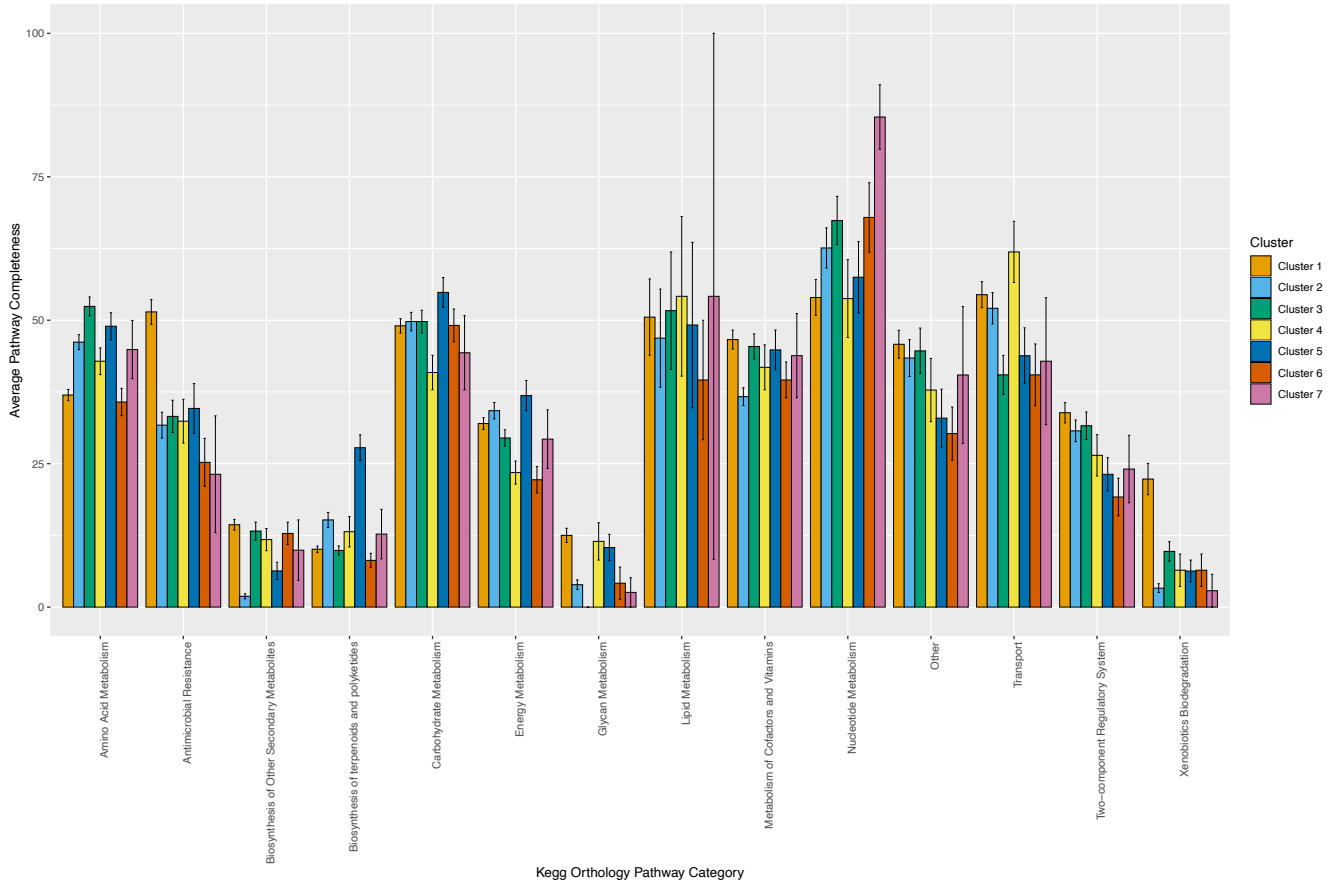

Figure 6: Mean pathway completeness of each cluster for different Kegg Orthology pathway categories. Standard error bars are supplied for each mean. Cluster 1: n=23, Cluster 2: n=16, Cluster 3: n=10, Cluster 4: n=4, Cluster 5: n=5, Cluster 6: n=4, Cluster 7: n=1.

a significantly higher average completeness for a particular category of pathways. This may indicate that the cluster is specialised for pathways in that category. There are three particularly notable categories that have one cluster with significantly higher average completeness, antimicrobial resistance, biosynthesis of terpenoids and polyketides, and nucleotide metabolism. Cluster 1 has a mean average completeness of 51.4% with a standard error of 2.15 for the antimicrobial resistance pathways. This is 1.5-fold higher than the next highest average of 34.6% standard error 4.3, for Cluster five. All other clusters range between 33.2-23.1%. For the biosynthesis of terpenoids and polyketides category Cluster five has a mean pathway

completeness of 27.8% with a standard error of 2.2. This is 1.8-fold higher than the next highest average of 15.2% with a standard error of 1.3 for Cluster two. Cluster seven has a notably higher average completeness for the nucleotide metabolism category. The mean completeness is 85.4% with a standard error of 5.6. This 1.25-fold higher than the next highest mean of 67.9% standard error 6.04 for Cluster six.

3.4 Genome streamlining occurred in certain Gemmatimonadetes species when compared with the average Gemmatimonadetes genome in our study

In order to determine whether or not genome streamlining may be occurring in the different species, we compared the number of ORFs that were called by a low confidence method vs a high confidence method of ORF calling. Low confidence refers to methods of ORF calling that are not specific, instead counting any possible ORF. This results in an inflated number of ORFs being found. In comparison, a high confidence method refers to a method of calling ORFs that are specific and have been called by referring to genetic databases. We used OrfM (Woodcroft et al., 2016) as a low confidence method and enrichM (Boyd, Joel A.; Woodcroft, Ben J.; Tyson, 2019) as a high confidence method. We looked at the ratio of ORFs called by OrfM to ORFs called by enrichM (Figure 7). The line on the graph represents the average ORF/enrichM ratio of all the Gemmatimonadetes bins. Clusters that fall above the average line have genomes containing more unclassified genetic material than the average for our clusters, indicating a bloated genome. In comparison, clusters that fall below the average line contain less unclassified ORFs to the number of classified ORFs, indicating a streamlined genome as it contains less unnecessary genetic material on average. The clusters tend to form distinct groups on the graph (Figure 7). Cluster 1 sits tightly just below the average ratio line, meaning bins in the cluster have more enrichM identified genes than average for the same number of ORFs

whilst remaining in the middle of the genome size range with an average size of . These results indicate that the genomes in Cluster 1 may have undergone some streamlining. Cluster 5 falls on the same area of the graph as Cluster 1 with the exception of one outlier bin that sits high above the average line, suggesting that this outlier bin may have a bloated genome. Cluster 2 sits close to the average ratio with some bins being above the line and others below. This is similarly seen with Cluster 4 and Cluster 6, although Cluster 6 contains notably smaller bins. Cluster 3 mostly sits above the average line with over 1000 more ORFs than the average for the same enrichM value and is made up of bins at around 4 Mbp in size. Cluster 7 is high in enrichM found genes as well as ORFs. The bin is large and together this suggests that the bin may be bloated.

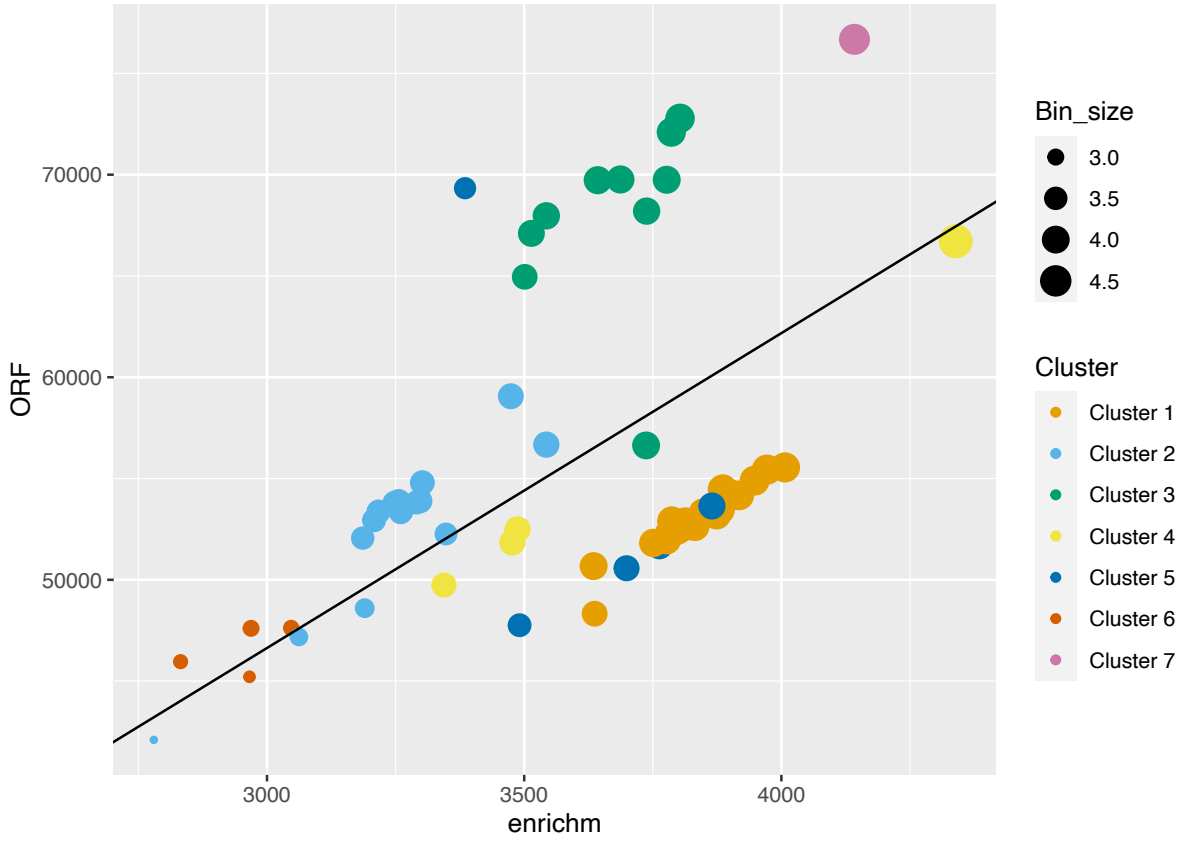

Figure 7: The average ratio of ORFs to enrichM derived genes indicates whether genome streamlining may have occurred. Each dot represents a separate bin and these are coloured by the cluster they belong to, as found in results section 3.2. The approximate bin size is shown by the size of the dot.

3.5 Community succession occurred in the Gemmatimonadetes community over time following the addition of urea

Bin specific ORFs were used to perform a DIAMOND blast against the clean sequencing reads from all samples. This enabled us to gauge the abundance of each of our Gemmatimonadetes species in the community over time without needing complete genome sequences. Community succession occurred to varying extents over time for all of the different treatments except the negative control (Figure 8). The Gemmatimonadetes community in the negative control samples remained consistent over time. Data for Day 27 of the antibacterial + antifungal treatment is missing as the original sample data is not available. The antibacterial (AB), antibacterial + antifungal (AB+AF) and antifungal (AF) treatments all showed similar community succession over time. Cluster 1 increased in relative abundance in each of these treatments following Day 9 whereas, Cluster 6 decreased in relative abundance over time after Day 9. Cluster 2 increased in relative abundance from Day 9 to Day 21in the AB and AB+AF treated samples. For the AF treatment, Cluster 2 increased in relative abundance on Day 15 and then decreased in relative abundance on Day 21. The relative abundance for Cluster 2 increased again on Day 42. The positive control (PC) treated samples showed clear community succession. Cluster 1 increased in relative abundance over time whilst Cluster 2 decreased. In all treatments except for the negative control, Cluster 6 appeared in highest relative abundance on Day 9 and then decreased in subsequent days. In the negative control, Cluster 6 remained as the dominant cluster over time with little change. Community succession occurred only in the urea treated samples.

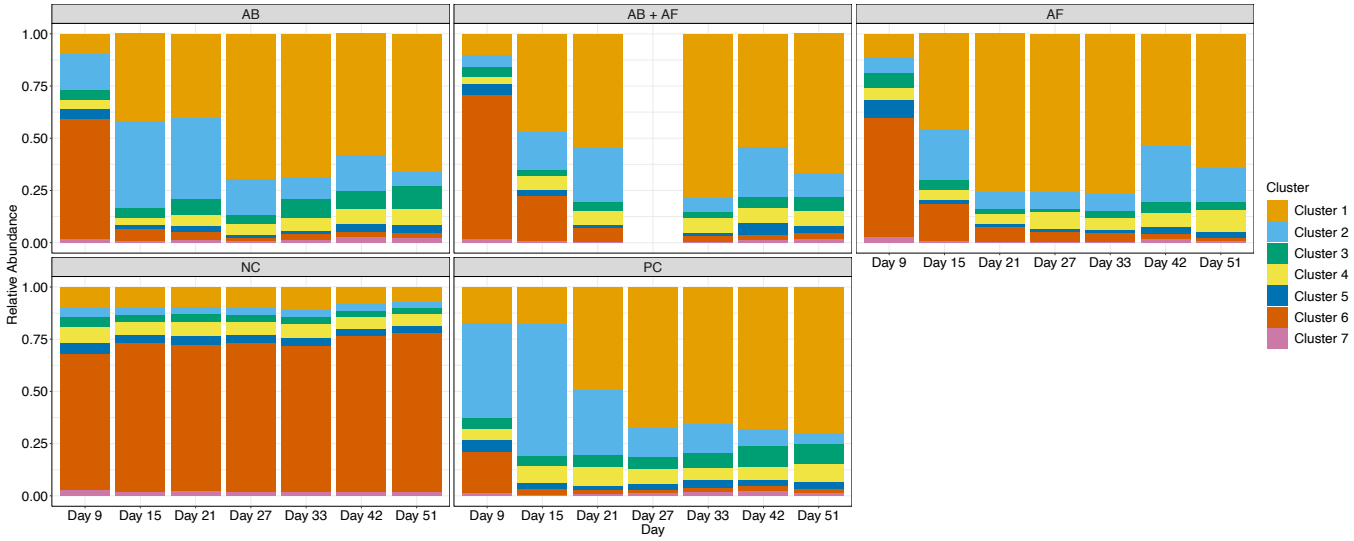

Figure 8: Relative abundance of Gemmatimonadetes clusters over time, facetted by treatment. Relative abundance is calculated from the total number of ORF hits from each cluster

Absolute abundance of the bin specific ORF matches was also calculated and used to create a line graph grouped by cluster (Figure 9). Log_2 of the abundances was used to better show community succession without the scale being skewed by the largest abundances (Figure 9B). In all of the urea treated samples (AB, AB+AF, AF) Cluster 1 increased in absolute abundance over time. Cluster 2 also increased in abundance over time, however the increase was less linear and on a smaller scale than Cluster 1. Cluster 6 remained at a relatively steady abundance for all of the treatments except for the AB treatment in which the abundance decreased after Day 9 until Day 27 where the abundance slowly started to recover. By Day 51 it had returned to the Day 9 abundance level. The negative control maintained a steady abundance for all clusters with Cluster 6 being present at the highest abundance. Additionally, the overall abundance of all clusters is much lower than for the urea treated samples (Figure 9A).

Figure 9: Absolute abundance of Gemmatimonadetes cluster-specific ORFs over time. Absolute abundance on a linear scale shows the different scale of abundances between clusters (A). Absolute abundance plotted with a Log_2 scale allows for changes in less abundant clusters to be visualised (B).

4. Discussion

The main goal of this thesis was to classify the understudied phylum Gemmatimonadetes under the conditions of a simulated urine patch. We hypothesized that the changing environmental conditions due to the addition of urea would impact on the Gemmatimonadetes community structure. Due to the diversity seen in the few studies focusing on members of the Gemmatimonadetes phylum, we also hypothesized that there would be significant differences in the functional potential of the Gemmatimonadetes species found.

4.1 Seven novel Gemmatimonadetes species were present in the metagenome bins

Seven clear, distinct clusters can be seen in both the ANI heatmap (Figure 2) and the maximum likelihood phylogenetic tree (Figure 3). The phylogenetic distance between the clusters is large however, still indicates that the clusters are related. The phylogenetic trees (Figure 3 and 5), showed high bootstrap values indicating that the tree is strongly supported. This allowed us to be confident that there is distinct genomic separation seen between the clusters. When the bins were compared to the reference sequences from clustered isolates, we found that the reference sequences did not match, or cluster with, any of our bins in the phylogenetic tree (Figure 5). We hypothesized that these results together indicated that the seven clusters we identified represented seven novel, uncultured species of phylum Gemmatimonadetes. In extension to this idea, the lack of matches to any of the current cultured Gemmatimonadetes isolates indicates that a large proportion of the Gemmatimonadetes phylum remains uncultured and unclassified, highlighting the need for further metagenomic studies to better understand the scope of the phylum. The under-studied nature of the phylum despite its observed widespread diversity is often an important point mentioned in previous literature focusing on Gemmatimonadetes (DeBruyn et al., 2011; Zeng et al., 2016, 2021). The finding that our bins

do not match any of the cultured isolates was therefore expected. The phylogenetic distance between our clusters and the novel nature indicated that our clusters likely differed in functional potential to other Gemmatimonadetes species. This led us to explore the difference in functional potential between our clusters and other Gemmatimonadetes through functional potential analysis.

4.2 Gemmatimonadetes is a diverse phylum of organisms that largely remain unclassified

Though present in many environments and particularly abundant in soils, phylum Gemmatimonadetes is greatly under classified. We identified diversity and notable differences in functional potential between Gemmatimonadetes species in the pasture soil environment (Figure 6) (Table 1). The Gemmatimonadetes species were found to differ in mean pathway completeness for all categories to varying extent, this showcases the diversity within members of the phylum that survive in the same community. Recent studies on Gemmatimonadetes agree with our findings of functional diversity and can help give us an idea of the scope of diversity within the phylum. To gives some examples on the range of functions; genes for Ni,Fe-hydrogenases (Baker et al., 2015), sulfate reduction (Baker et al., 2015), methanotrophy (Butterfield et al., 2016; M. Li et al., 2016), and photosynthesis (Zeng et al., 2016, 2021) have all been found in different Gemmatimonadetes species from different environments. The diversity of Gemmatimonadetes and the wide range of environments and conditions that it can be found in supports the idea that the Gemmatimonadetes phylum shows cosmopolitan success (DeBruyn et al., 2011). In addition to the diversity of functional potential observed, we also identified a range of different genome sizes with varying proportions of unnecessary genetic material. This led us to believe that genome streamlining may have occurred in some of our Gemmatimonadetes species.

4.3 Genome streamlining and ecological strategies

When comparing the seven Gemmatimonadetes species, some showed greater streamlining than average whilst others appeared to have more bloated genomes. One particular cluster of interest was Cluster 1 which was streamlined in comparison to the other clusters (Figure 7). Whilst the genome was streamlined, Cluster 1 had the highest mean pathway completeness for antimicrobial resistance pathways, suggesting that these pathways have provided the species with a selective advantage. In addition to this, Cluster 1 was also the most abundant Gemmatimonadetes cluster following the addition of urea to the different treated soils (Figure 8). We hypothesized from this that Cluster 1 may be a specialist organism that has adaptations that enable it to outcompete other microbes during periods of high nitrogen concentration. In comparison, Cluster 3 has a bloated genome in comparison to the average cluster and is also present in all treatments in a small abundance. From this we hypothesized that Cluster 3 may be a generalist, baseline Gemmatimonadetes species found in agricultural soil communities. Genome size has been found to be positively correlated with ubiquity of the microbe (Cobo-Simón & Tamames, 2017). Generalist species are often slow growing and outcompete other organisms in environments with limited nutrient availability (Monard et al., 2016; Sriswasdi et al., 2017) whereas specialist species tend to be fast growing and thrive in environments with large nutrient availability (Mariadassou et al., 2015; Monard et al., 2016). Bloated genomes would occur when there is no selective pressure to lose redundant genomic content. In contrast, streamlined genomes would occur when there is a selective advantage to losing unnecessary genetic content for example for lower energy requirements and faster replication.

On a whole, Gemmatimonadetes phylum appears to have maintained a strong generalist population (DeBruyn et al., 2011) that has allowed for the widespread dispersion of the phylum leading to the evolution of specialists in a wide range of environments (Sriswasdi et al., 2017). Cluster 1 contains nosZ unlike many of the other clusters. This gene is involved in nitrogen cycling and may contribute to the survival advantage that we can see following the addition of urea. In comparison, the baseline species cluster 6 has far less genes in general in it's genome. It may have a small genome which could give it an advantage under normal soil conditions as it requires less energy and is able to get nutrients etc. from other organisms in the soil community.

4.4 Succession and differentiation of niches and roles.

The diversity of the Gemmatimonadetes phylum can be showcased by the clear community succession in the soil samples that had urea added in comparison to the negative control with no urea added to the soil (Figure 8). The urea caused a large change in the environment which led to changes in the abundance of Gemmatimonadetes species in the samples over time. It has been found that Gemmatimonadetes positively responds to urea at the phylum level although the response differed among lower taxonomic levels (Samad et al., 2017). This corroborates with our findings as we observed that one species in particular, Cluster 1, positively reacted to the addition of urea whereas other species saw little to no change (Figure 7). The positive reaction of Cluster 1 to the addition to urea suggests that it may play a large role in nitrogen cycling during periods of high nitrogen loading as ammonia oxidizing bacteria have been found to have higher numbers than ammonia oxidizing archaea in time of high N loading (Samad et al., 2017).

Cluster 1 had the *nosZ* gene, which is found in nitrous oxide (N_2O) reducing microorganisms, in 100% of the bins making up the cluster (Table 1). This was similarly seen in Cluster 2 which

also responded positively to the addition of urea, albeit to a lesser degree. The other clusters had much lower presence of the gene or did not have it at all. We hypothesize that the presence of the *nosZ* gene in Cluster 1, alongside the presence of *nirK* (a nitrite reductase), provide Cluster 1 with a genetic advantage during periods of high nitrogen loading. Gemmatimonadetes are known to have one of the most abundant sources of *nosZ* genes in soil microbes. This is of particular importance as N_2O reducing microbes create the only environmentally relevant biological sink of N_2O in the environment (Park et al., 2021). N_2O is a potent greenhouse gas that contributes to ozone depletion therefore, the balance between microbial sources and sinks in the soil environment is important for determining total emissions of this greenhouse gas (Zaman et al., 2009). *Gemmatimonadetes aurantiaca* Strain T-27 was found to contain *nosZ* and was able to carry out N_2O reduction in the transient presence of oxygen (Park et al., 2021). Gemmatimonadetes originating from estuary sediment bacteria were also found to have *nosZ* in their genome (Baker et al., 2015).

4.5 Limitations of the study

As with any study there are limitations to the scope of our research. This analysis was carried out using samples from previous studies, therefore, although it provides a good overview of Gemmatimonadetes community succession within the wider soil microbial community under the conditions of a simulated urine patch, the study was not tailored specifically for Gemmatimonadetes therefore there is a limit to the depth of analysis we can carry out on the bins. Additionally, a very particular type of soil was studied and our findings may or may not be applicable to other soil types in different geographical locations although this would be an interesting further direction for this work. We were also limited as we only had medium-quality draft genomes to work with and were only able to analyze functional potential rather than carrying out a meta transcriptome analysis.

4.6 Future directions

Further studies focusing on the Gemmatimonadetes community in different environments and under different conditions would be helpful for increasing our understanding of the true diversity of the phylum and its roles and importance in different environments. It would also e interesting to carry out a meta-transcriptomic analysis of a Gemmatimonadetes community to determine whether the genes identified are utilized or not. Using what we know about the functional potential of particular Gemmatimonadetes species from metagenome study, we could also work towards culturing more isolates in an attempt to increase the number of complete Gemmatimonadetes reference sequences. It would be interesting to use the clusters identified in this study to compare to other Gemmatimonadetes found through metagenome studies to determine the scope of our identified species. It would also be interesting to further explore the genomes and functions of the species, particularly those represented by Cluster 1 and Cluster 6 as these species showed the greatest change in abundance following the addition of urea.

4.7 Conclusions

We identified seven novel, uncultured Gemmatimonadetes species. These species showcased a wide variety of functional potential which mirrors the diversity that has been seen in the Gemmatimonadetes phylum in previous studies. The addition of urea with and without inhibitors lead to community succession within the Gemmatimonadetes community. By analyzing the abundance of Gemmatimonadetes species over time we were identify a baseline Gemmatimonadetes soil species (Cluster 6) as well as a potential specialist that has adaptations to allow for greater efficiency in a nitrogen rich environment such as a urine patch (Cluster 1).

Agriculture emissions and climate change | Ministry for the Environment. (n.d.). Retrieved August 18, 2021, from https://environment.govt.nz/guides/agriculture-emissions-climate-change/

Alneberg, J., Karlsson, C. M. G., Divne, A.-M., Bergin, C., Homa, F., Lindh, M. v., Hugerth, L. W., Ettema, T. J. G., Bertilsson, S., Andersson, A. F., & Pinhassi, J. (2018). Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. *Microbiome 2018 6:1*, *6*(1), 1–14. https://doi.org/10.1186/S40168-018-0550-0

Asnicar, F., Thomas, A. M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., Sanders, J. G., Zolfo, M., Kopylova, E., Pasolli, E., Knight, R., Mirarab, S., Huttenhower, C., & Segata, N. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nature Communications 2020 11:1*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-16366-7

Baker, B. J., Lazar, C. S., Teske, A. P., & Dick, G. J. (2015). Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome*, *3*(1), 14. https://doi.org/10.1186/s40168-015-0077-6

Bay, S. K., Dong, X., Bradley, J. A., Leung, P. M., Grinter, R., Jirapanjawat, T., Arndt, S. K., Cook, P. L. M., LaRowe, D. E., Nauer, P. A., Chiri, E., & Greening, C. (2021). Trace gas oxidizers are widespread and active members of soil microbial communities. *Nature Microbiology*, *6*(2), 246–256. https://doi.org/10.1038/s41564-020-00811-w

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloe-Fadrosh, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., … Woyke, T. (2017). Minimum information about a single amplified genome

(MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology 2017 35:8*, *35*(8), 725–731. https://doi.org/10.1038/nbt.3893

Boyd, Joel A. ; Woodcroft, Ben J.; Tyson, G. W. (2019). *Comparative genomics using EnrichM. 2019. In preparation.* https://doi.org/In preparation

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods 2021 18:4*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Bushnell, B. (2021). *BBMap*. https://sourceforge.net/projects/bbmap/

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via overlap. *PLOS ONE*, *12*(10), e0185056. https://doi.org/10.1371/JOURNAL.PONE.0185056

Butterfield, C. N., Li, Z., Andeer, P. F., Spaulding, S., Thomas, B. C., Singh, A., Hettich, R. L., Suttle, K. B., Probst, A. J., Tringe, S. G., Northen, T., Pan, C., & Banfield, J. F. (2016). Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. *PeerJ*, *4*, e2687–e2687. https://doi.org/10.7717/peerj.2687

Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., Behrenfeld, M. J., Boetius, A., Boyd, P. W., Classen, A. T., Crowther, T. W., Danovaro, R., Foreman, C. M., Huisman, J., Hutchins, D. A., Jansson, J. K., Karl, D. M., Koskella, B., Mark Welch, D. B., … Webster, N. S. (2019). Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology 2019 17:9*, *17*(9), 569–586. https://doi.org/10.1038/s41579-019-0222-5

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925–1927. https://doi.org/10.1093/BIOINFORMATICS/BTZ848

Cichota, R., Vogeler, I., Snow, V., Shepherd, M., McAuliffe, R., & Welten, B. (2018). Lateral spread affects nitrogen leaching from urine patches. *Science of The Total Environment*, *635*, 1392–1404. https://doi.org/10.1016/J.SCITOTENV.2018.04.005

Cobo-Simón, M., & Tamames, J. (2017). Relating genomic characteristics to environmental preferences and ubiquity in different microbial taxa. *BMC Genomics*, *18*(1). https://doi.org/10.1186/S12864-017-3888-Y

Coenen, A. R., Hu, S. K., Luo, E., Muratore, D., & Weitz, J. S. (2020). A Primer for Microbiome Time-Series Analysis. *Frontiers in Genetics*, *0*, 310. https://doi.org/10.3389/FGENE.2020.00310

Core Writing Team, Pachauri, R. K., & Meyer, L. A. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Danecek, P., Bonfield, J., Liddle, J., Marshall, J., Ohan, V., Pollard, M., Whitwham, A., Keane, T., McCarthy, S., Davies, R., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/GIGASCIENCE/GIAB008

DeBruyn, J. M., Nixon, L. T., Fawaz, M. N., Johnson, A. M., & Radosevich, M. (2011). Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. *Applied and Environmental Microbiology*, *77*(17), 6295–6300. https://doi.org/10.1128/AEM.05005-11

Faust, K., Lahti, L., Gonze, D., de Vos, W. M., & Raes, J. (2015). Metagenomics meets time series analysis: unraveling microbial community dynamics. *Current Opinion in Microbiology*, *25*, 56–66. https://doi.org/10.1016/J.MIB.2015.04.004

Ganasamurthy, S., Rex, D., Samad, M. S., Richards, K. G., Lanigan, G. J., Grelet, G. A., Clough, T. J., & Morales, S. E. (2021). Competition and community succession link N

transformation and greenhouse gas emissions in urine patches. *Science of The Total Environment*, *779*, 146318. https://doi.org/10.1016/J.SCITOTENV.2021.146318

Glaeser, S. P., & Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Systematic and Applied Microbiology*, *38*(4), 237–245. https://doi.org/10.1016/J.SYAPM.2015.03.007

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I., & Dubchak, I. (2014). The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Research*, *42*(Database issue). https://doi.org/10.1093/NAR/GKT1069

Hutchings, N. J., Olesen, J. E., Petersen, B. M., & Berntsen, J. (2007). Modelling spatial heterogeneity in grazed grassland and its effects on nitrogen cycling and greenhouse gas emissions. *Agriculture, Ecosystems & Environment*, *121*(1–2), 153–163. https://doi.org/10.1016/J.AGEE.2006.12.009

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018a). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications 2018 9:1*, *9*(1), 1–8. https://doi.org/10.1038/s41467-018-07641-9

Jansson, J. K., & Hofmockel, K. S. (2019). Soil microbiomes and climate change. *Nature Reviews Microbiology 2019 18:1*, *18*(1), 35–46. https://doi.org/10.1038/s41579-019-0265-7

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. https://doi.org/10.1093/NAR/GKV1070

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, *7*(7). https://doi.org/10.7717/PEERJ.7359

Kolton, M., Marks, A., Wilson, R. M., Chanton, J. P., & Kostka, J. E. (2019). Impact of Warming on Greenhouse Gas Production and Microbial Diversity in Anoxic Peat From a Sphagnum-Dominated Bog (Grand Rapids, Minnesota, United States). *Frontiers in Microbiology*, *0*(APR), 870. https://doi.org/10.3389/FMICB.2019.00870

Kuypers, M. M. M., Marchant, H. K., & Kartal, B. (2018). The microbial nitrogen-cycling network. *Nature Reviews Microbiology 2018 16:5*, *16*(5), 263–276. https://doi.org/10.1038/nrmicro.2018.9

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/NAR/GKAB301

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. https://arxiv.org/abs/1303.3997v2

Li, M., Jain, S., & Dick, G. J. (2016). Genomic and Transcriptomic Resolution of Organic Matter Utilization Among Deep-Sea Bacteria in Guaymas Basin Hydrothermal Plumes. *Frontiers in Microbiology*, *7*, 1125. https://www.frontiersin.org/article/10.3389/fmicb.2016.01125

Mariadassou, M., Pichon, S., & Ebert, D. (2015). Microbial ecosystems are dominated by specialist taxa. *Ecology Letters*, *18*(9), 974–982. https://doi.org/10.1111/ELE.12478

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288–4297. https://doi.org/10.1093/NAR/GKS042

Monard, C., Gantner, S., Bertilsson, S., Hallin, S., & Stenlid, J. (2016). Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. *Scientific Reports 2016 6:1*, *6*(1), 1–10. https://doi.org/10.1038/srep37719

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, *27*(5), 824. https://doi.org/10.1101/GR.213959.116

Park, D., Kim, H., Yoon, S., & Kivisaar, M. (2021). Nitrous Oxide Reduction by an Obligate Aerobic Bacterium, Gemmatimonas aurantiaca Strain T-27. *Applied and Environmental Microbiology*, *83*(12), e00502-17. https://doi.org/10.1128/AEM.00502-17

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043. https://doi.org/10.1101/GR.186072.114

Pascual, J., García-López, M., Bills, G. F., & Genilloud, O. (2016). Longimicrobium terrae gen. nov., sp. nov., an oligotrophic bacterium of the under-represented phylum Gemmatimonadetes isolated through a system of miniaturized diffusion chambers. *International Journal of Systematic and Evolutionary Microbiology*, *66*(5), 1976–1985. https://doi.org/10.1099/ijsem.0.000974

Rex, D. (2018). *Codenitrification Under Ruminant Urine Patch Conditions: Microbial Contributions, Substrates, and Nitrogen Flux Kinetics*.

Rex, D., Clough, T. J., Richards, K. G., de Klein, C., Morales, S. E., Samad, M. S., Grant, J., & Lanigan, G. J. (2017). Fungal and bacterial contributions to codenitrification emissions of N2O and N2 following urea deposition to soil. *Nutrient Cycling in Agroecosystems 2017 110:1*, *110*(1), 135–149. https://doi.org/10.1007/S10705-017-9901-7

Rex, D., Schimmelpfennig, S., Jansen-Willems, A., Moser, G., Kammann, C., & Müller, C. (2015). Microbial community shifts 2.6 years after top dressing of Miscanthus biochar, hydrochar and feedstock on a temperate grassland site. *Plant and Soil 2015 397:1*, *397*(1), 261–271. https://doi.org/10.1007/S11104-015-2618-Y

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/BIOINFORMATICS/BTP616

RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. RStudio, PBC.

Samad, M. S., Johns, C., Richards, K. G., Lanigan, G. J., Klein, C. A. M. de, Clough, T. J., & Morales, S. E. (2017). Response to nitrogen addition reveals metabolic and ecological strategies of soil bacteria. *Molecular Ecology*, *26*(20), 5500–5514. https://doi.org/10.1111/MEC.14275

Selbie, D. R., Buckthought, L. E., & Shepherd, M. A. (2015). The Challenge of the Urine Patch for Managing Nitrogen in Grazed Pasture Systems. *Advances in Agronomy*, *129*, 229–292. https://doi.org/10.1016/BS.AGRON.2014.09.004

Song, B., Li, Z., Li, S., Zhang, Z., Fu, Q., Wang, S., Li, L., & Qi, S. (2021). Functional metagenomic and enrichment metatranscriptomic analysis of marine microbial activities within a marine oil spill area. *Environmental Pollution*, *274*, 116555. https://doi.org/10.1016/J.ENVPOL.2021.116555

Sriswasdi, S., Yang, C., & Iwasaki, W. (2017). Generalist species drive microbial dispersion and evolution. *Nature Communications 2017 8:1*, *8*(1), 1–8. https://doi.org/10.1038/s41467-017-01265-1

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/BIOINFORMATICS/BTU033

Takaichi, S., Maoka, T., Takasaki, K., & Hanada, S. (2010). Carotenoids of Gemmatimonas aurantiaca (Gemmatimonadetes): identification of a novel carotenoid, deoxyoscillol 2-rhamnoside, and proposed biosynthetic pathway of oscillol 2,2′-dirhamnoside. *Microbiology*, *156*(3), 757–763. https://doi.org/10.1099/mic.0.034249-0

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, *108*(50), 20260–20264. https://doi.org/10.1073/PNAS.1116437108

Vásquez-Ponce, F., Higuera-Llantén, S., Pavlov, M. S., Marshall, S. H., & Olivares-Pacheco, J. (2018). Phylogenetic MLSA and phenotypic analysis identification of three probable novel Pseudomonas species isolated on King George Island, South Shetland, Antarctica. *Brazilian Journal of Microbiology*, *49*(4), 695. https://doi.org/10.1016/J.BJM.2018.02.005

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, *40*(1), 1–29.

Wickham, H., Navarro, D., & Pederson, T. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org/

Wilkins, L. G. E., Ettinger, C. L., Jospin, G., & Eisen, J. A. (2019). Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Scientific Reports 2019 9:1*, *9*(1), 1–15. https://doi.org/10.1038/s41598-019-39576-6

Woodcroft, B. J., Boyd, J. A., & Tyson, G. W. (2016). OrfM: a fast open reading frame predictor for metagenomic data. *Bioinformatics*, *32*(17), 2702–2703. https://doi.org/10.1093/BIOINFORMATICS/BTW241

Yoon, S.-H., Ha, S., Lim, J., Kwon, S., & Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek 2017 110:10*, *110*(10), 1281–1286. https://doi.org/10.1007/S10482-017-0844-4

Zaman, M., Saggar, S., Blennerhassett, J. D., & Singh, J. (2009). Effect of urease and nitrification inhibitors on N transformation, gaseous emissions of ammonia and nitrous oxide, pasture yield and N uptake in grazed pasture system. *Soil Biology and Biochemistry*, *41*(6), 1270–1280. https://doi.org/10.1016/J.SOILBIO.2009.03.011

Zeng, Y., Baumbach, J., Barbosa, E. G. V., Azevedo, V., Zhang, C., & Koblížek, M. (2016). Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. *Environmental Microbiology Reports*, *8*(1), 139–149. https://doi.org/https://doi.org/10.1111/1758-2229.12363

Zeng, Y., Nupur, Wu, N., Madsen, A. M., Chen, X., Gardiner, A. T., & Koblížek, M. (2021). Gemmatimonas groenlandica sp. nov. Is an Aerobic Anoxygenic Phototroph in the Phylum Gemmatimonadetes. *Frontiers in Microbiology*, *11*, 3395. https://www.frontiersin.org/article/10.3389/fmicb.2020.606612

Zhang, H., Sekiguchi, Y., Hanada, S., Hugenholtz, P., Kim, H., Kamagata, Y., & Nakamura, K. (2003). Gemmatimonas aurantiaca gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. *International Journal of Systematic and Evolutionary Microbiology*, *53*(4), 1155–1163. https://doi.org/10.1099/ijs.0.02520-0