Coursework Declaration and Feedback Form

The Student should complete and sign this part

| Student Number: 2603933 | Student Name: Kaung Sett | | |
|---|---|--|--|
| Programme of Study (e.g. MSc in Electronics and Electrical Engineering): MSc in Civil Engineering | | | |
| Course Code: ENG5059P | Course Name: MSc Project | | |
| Name of <u>First</u> Supervisor: Dr Umer Zeeshan Ijaz | Name of <u>Second</u> Supervisor: Dr Ciara Keating | | |
| Title of Project: Whole Genome Analysis of Nitrifying Bact | eria in Construction and the Built Environment | | |
| Declaration of Originality and Sub | mission Information | | |
| I affirm that this submission is all my own work in accordance with the University of Glasgow Regulations and the School of Engineering requirements Signed (Student) : | E N G 5 0 5 9 P | | |
| Date of Submission : 20 Aug 2021 | | | |
| Feedback from Lecturer to Student – to be comple | ted by Lecturer or Demonstrator | | |
| Grade Awarded: Feedback (as appropriate to the coursework which was assessed): | | | |
| Lecturer/Demonstrator: | Date returned to the Teaching Office: | | |



Whole Genome Analysis of Nitrifying Bacteria in Construction and the Built Environment

Student Name: Kaung Sett Student ID: 2603933 Supervisor Name: Dr Umer Zeeshan Ijaz Co-supervisor Name: Dr Ciara Keating

August 2021

A thesis submitted in partial fulfilment of the requirements for the degree of MASTER OF SCIENCE IN CIVIL ENGINEERING

ACKNOWLEDGEMENT

First and foremost, the author would like to acknowledge his supervisors, Dr. Umer Zeeshan Ijaz and Dr Ciara Keating, for giving him this opportunity and their invaluable technological ideas, thoroughly bioinformatics and environmental concepts for this project during its preparation, analysis and giving their supervision on his project.

Next, the author desires to express his deepest gratitude to University of Glasgow for providing this opportunity to improve his knowledge and skills.

Then, the author deeply appreciated to his parents for their continuous love, financial support, and encouragement throughout his entire life. With the strength they give, the author finished this project without any difficulty.

Last but not least, greatly thanks to teammates of the author who are in the same project team with him through the semester.

ABSTRACT

Nitrifying bacteria, organisms that take part in the process of nitrification, are found in various environmental and engineering systems, such as wastewater treatment plants, soil, and freshwater supply. These nitrifying bacteria are mainly divided into ammonia-oxidizing bacteria (AOB), ammonia-oxidizing archaea (AOA) and nitrite-oxidizing bacteria (NOB). AOB and AOA serve as ammonia oxidizers in the nitrification process whilst NOB are the nitrite oxidizers.

The aim of this research project is to analyse the genome sequences of nitrifying bacteria using whole genome analysis (pangenome analysis). The analysis is performed for the respective genera included in the nitrifying bacteria. The whole analysis process is implemented within a Linux environment on the Orion server hosted by Dr Umer Ijaz. While implementing the whole genome analysis technique, different software tools with respective functions are used. Prokka is used for genome annotation, Roary is used for generating pangenome and METABOLIC is used for metabolic analysis in genomes. The pan-genome plots and the presence and absence of the genes, which are output from the analysis, are studied, and investigated with pan-genome visualisation tools. The functional genes from the nitrifying bacteria. According to the presence and absence of the functional genes, the effectiveness and functional process of the nitrifying bacterial species are studied.

In my study, I observed that most of the AOB genomes contain amoA, amoB, amoC, nirK, norB and norC so they are functional only in ammonia oxidation, nitrite reduction to ammonia and nitric oxide reduction. *Nitrosomonas europaea* is considered the most efficient AOB due to the genes presenting in its genomes. Only a few functional genes involved in nitrification, nirK, nosZ, nrfA and nxrA are present in the AOA genomes. Therefore, I conclude that AOB have increased capacity as ammonia oxidizers than AOA, according to the genes existing in their genome sequencing. NOB species present several functional genes such as narG, narH, nirB, nirD, nirK, nxrA, nxrB, nrfH, nrfA and norC, so they are efficient in operating reduction of nitrates, nitrites, or nitrous oxide as well as nitrite oxidation. *Nitrospira defluvii, Nitrobacter hamburgensis* and *Nitrobacter vulgaris* are the most effective nitrite oxidizers among the NOB species.

Key words: whole genome analysis, nitrifying bacteria, AOB, AOA, NOB

TABLE OF CONTENTS

| ACKNOW | LEDGEMENTi |
|-----------|---|
| ABSTRAC | |
| TABLE OF | F CONTENTS iii |
| LIST OF A | BBREVIATIONSv |
| LIST OF F | IGURESvii |
| LIST OF T | ABLES viii |
| CHAPTER | . 1 |
| INTRODU | CTION |
| 1.1 E | Background1 |
| 1.2 A | Ammonia-oxidizing bacteria (AOB) |
| 1.3 A | Ammonia-oxidizing archaea (AOA) |
| 1.4 N | Vitrite-oxidizing bacteria (NOB) |
| 1.5 R | Related research on nitrifying bacteria in construction and built environment |
| 1.6 A | Aims and Objectives |
| CHAPTER | |
| METHOD | OLOGY10 |
| 2.1 In | nitial preparation on genomes11 |
| 2.2 P | Primary analysis of genomes |
| 2.2.1 | Prokka workflow13 |
| 2.2.2 | Roary workflow |
| 2.2.3 | METABOLIC workflow14 |
| 2.3 V | /isualisation of the results15 |
| 2.3.1 | Coinfinder |
| 2.3.2 | RStudio15 |
| 2.4 D | Description of the functional genes in the results16 |

| CHAPTER | . 3 |
|----------|---|
| RESULTS | |
| 3.1 R | Results of ammonia-oxidizing bacteria (AOB)17 |
| 3.1.1 | Results of Nitrosomonas genus17 |
| 3.1.2 | Results of Nitrosospira genus19 |
| 3.1.3 | Results of Nitrosococcus genus |
| 3.1.4 | Pan-genome plot of AOB22 |
| 3.2 R | Results of ammonia-oxidizing archaea (AOA)23 |
| 3.2.1 | Results of Thaumarchaea23 |
| 3.2.2 | Results of Nitrosopumilus genus25 |
| 3.2.3 | Pan-genome plot of AOA |
| 3.3 R | Results of nitrite-oxidizing bacteria (NOB)27 |
| 3.3.1 | Results of Nitrobacter genus |
| 3.3.2 | Results of Nitrospira genus |
| 3.3.3 | Pan-genome plot of NOB |
| 3.4 C | Coinfinder output |
| CHAPTER | .4 |
| DISCUSSI | ON |
| CHAPTER | . 5 |
| CONCLUS | SION |
| REFEREN | CES |
| APPENDE | X |

LIST OF ABBREVIATIONS

| Abbreviation | Explanation |
|-----------------|--|
| amo | Ammonia monooxygenase |
| anf | Associative nitrogen fixation |
| AOA | Ammonia-oxidizing archaea |
| AOB | Ammonia-oxidizing bacteria |
| ATP | Adenosine triphosphate |
| COD | Chemical oxygen demand |
| CPU | Central processing unit |
| csp | Cold shock-like protein |
| DNA | Deoxyribonucleic acid |
| DO | Dissolved oxygen |
| FASTA | Fast-all |
| FISH | Fluorescence in situ hybridization |
| FNA | Fasta nucleic acid |
| GB | Gigabyte |
| GBK | Genbank |
| GFF3 | General feature format |
| НАО | Hydroxylamine oxidoreductase |
| ICM | Intracellular multiplication |
| Km | Michaelis constant |
| METABOLIC | Metabolic and Biogeochemistry Analyses in Microbes |
| NADH | Nicotinamide adenine dinucleotide |
| nap | Periplasmic nitrate reductase |
| nar | Nitrate reductase |
| NCBI | National Centre for Biotechnology Information |
| NH ₄ | Ammonia |
| nif | Nitrogen fixation |
| nir | Dissimilatory nitrite reductase |
| NOB | Nitrite-oxidizing bacteria |
| NO ₂ | Nitrogen dioxide |
| NO ₃ | Nitrate |

| nor | Nitric oxide reductase |
|------|--|
| nos | Nitrous oxide reductase |
| nrf | Ammonium producing nitrite reductase |
| nxr | Nitrite oxidoreductase |
| OHOs | Ordinary heterotrophic organisms |
| рН | Potential of hydrogen |
| PN | Partial nitrification |
| qPCR | Quantitative polymerase chain reaction |
| RAM | Random access memory |
| rRNA | Ribosomal ribonucleic acid |
| rsp | Ribosomal protein |
| SBR | Sequencing batch reactor |
| SNP | Single nucleotide polymorphisms |
| SSH | Secure shell |
| vnf | Vanadium nitrogen fixation |
| WGS | Whole genome sequencing |

LIST OF FIGURES

| Figure | | Page |
|--------|---|------|
| 1 | Nitrification | 1 |
| 2 | Pan-genome frequency and pan-genome pie of Nitrosomonas genus | 18 |
| 3 | Pan-genome matrix of Nitrosomonas genus | 18 |
| 4 | Pan-genome frequency and pan-genome pie of Nitrosospira genus | 19 |
| 5 | Pan-genome matrix of Nitrosospira genus | 20 |
| 6 | Pan-genome frequency and pan-genome pie of Nitrosococcus genus | 21 |
| 7 | Pan-genome matrix of Nitrosococcus genus | 21 |
| 8 | Pan-genome plot of AOB | 22 |
| 9 | Pan-genome frequency and pan-genome pie of Thaumarchaea | 24 |
| 10 | Pan-genome matrix of Nitrosococcus genus | 24 |
| 11 | Pan-genome frequency and pan-genome pie of Nitrosopumilus genus | 25 |
| 12 | Pan-genome matrix of Nitrosopumilus genus | 26 |
| 13 | Pan-genome plot of AOA | 27 |
| 14 | Pan-genome frequency and pan-genome pie of Nitrobacter genus | 28 |
| 15 | Pan-genome matrix of Nitrosopumilus genus | 29 |
| 16 | Pan-genome frequency and pan-genome pie of Nitrospira genus | 30 |
| 17 | Pan-genome matrix of Nitrosopumilus genus | 30 |
| 18 | Pan-genome plot of NOB | 31 |
| 19 | Output heatmap of pan-genomes for nitrifying bacteria | 32 |

LIST OF TABLES

| Table | | Page |
|-------|--|------|
| 1 | Selected species of the nitrifying bacteria | 12 |
| 2 | Functional genes in the nitrogen cycle | 16 |
| 3 | Summary statistics of genes for Nitrosomonas genus | 17 |
| 4 | Summary statistics of genes for Nitrosospira genus | 19 |
| 5 | Summary statistics of genes for Nitrosococcus genus | 20 |
| 6 | Summary statistics of genes for Thaumarchaea | 23 |
| 7 | Summary statistics of genes for Nitrosopumilus genus | 25 |
| 8 | Summary statistics of genes for Nitrobacter genus | 28 |
| 9 | Summary statistics of genes for Nitrospira genus | 30 |
| 10 | Functional genes found in each nitrifying bacteria genus | 33 |

CHAPTER 1

INTRODUCTION

1.1 Background

Nitrogen is present widely in the environment and the nitrogen cycle provides the organisms in the ecosystem with extensive nutrients and energy. Conversely, environment modification has dramatically changed the nitrogen cycle, which has a great impact on soil, water and atmosphere, the fundamental resources in civil engineering. In the case of water, additional nitrogenous compounds released in water sources from wastewater consequences overstimulation of the growth of algae, lowering the concentration of dissolved oxygen and creating the environmental pollution. Therefore, the nutrient levels in wastewater treatment plants must be decreased by nitrification to prevent that pollution and maintain the ecosystem. Nitrification is an important two-step biological oxidation process in the nitrogen cycle as shown in figure 1. It has an essential role in governing the proportions of inorganic nitrogen species in various environments, not only wastewater treatment plants but also drinking water and aquatic engineering systems (Peng and Zhu, 2006).



Figure 1: Nitrification (Stein, 2015)

Nitrifying bacteria are chemolithotrophic organisms that take part in the process of nitrification. They are found in a variety of environmental groups and are most abundant where

there is a high concentration of ammonia. Due to high ammonia concentration, nitrifying bacteria grow in bodies of water, with large inputs and outputs of sewage and wastewater, as well as freshwater (Belser, 1979). Types include the ammonia-oxidizing bacteria (AOB) and ammonia-oxidizing archaea (AOA), which perform the first stage of nitrification, converting ammonium (NH₄) to nitrite (NO₂), and the nitrite-oxidizing bacteria (NOB), which accomplishes the second stage, converting nitrite (NO₂) to nitrate (NO₃).

1.2 Ammonia-oxidizing bacteria (AOB)

The ammonia-oxidizing bacteria (AOB) are microorganisms known as ammonia oxidizers that produce energy primarily by converting ammonia to nitrite. A limited quantity of energy is produced in this stage, leading to low microbe rates of growth, and making isolation challenging. Winogradsky identified the first AOB in the late nineteenth century Dworkin and Gutnick, 2012). Various microbiological groups cultivated and classified AOB from a range of settings in the mid-twentieth century, including marine waters, coastal systems, soils, and wastewater treatment systems. At ammonium concentrations that were usually much higher than those found in the environment, most of these isolations were carried out. DNA-based technologies are used to define the isolated AOB in a phylogenetic way (Bollmann, French and Laanbroek, 2011).

The ammonia-oxidizing bacteria (AOB) are represented by the members of Betaproteobacteria and Gammaproteobacteria. The majority of AOB belongs to Betaproteobacteria, on the other hand, only a few marine AOB are members of the Gammaproteobacteria. AOB are classified into three categories, each with its own set of eco-physiological characteristics and favoured environments. The first is the Nitrosomonas genus, which is divided into six lineages. The first lineage includes four species: *Nitrosomonas europaea*, *Nitrosomonas eutropha*, *Nitrosomonas halophila*, and *Nitrosomonas mobilis*, which are distinguished by their moderate salt necessity, negative oxygen diffusion, and can be separated from wastewater treatment plants, eutrophic freshwater, and activated sludge. The second Nitrosomonas lineage contains *Nitrosomonas communis*, which has no salt requirements, no water solubility and can be found in non-acidic soils. Additionally, the third lineage, which includes *Nitrosomonas nitrosa* species, has no salt need but does have positive water solubility and is often found in polluted freshwater resources. Furthermore, the fourth Nitrosomonas lineage, that includes two species: *Nitrosomonas ureae* and *Nitrosomonas*

oligotropha, with no salt necessity, positive water solubility and prefers anaerobic fresh water and organic salts as its favoured environments. Finally, the fifth and sixth lineages, which include *Nitrosomonas marina*, *Nitrosomonas aestuarii*, and *Nitrosomonas cryotolerans*, are both obligate halophilic, have positive water solubility, and are commonly found in marine environments (Soliman and Eldyasti, 2018).

Nitrosospira genera: *Nitrosospira briensis*, *Nitrosospira lacus* and *Nitrosospira multiformis* are among the second group of AOBs. They are mainly found in soils, rocks, and freshwater, and have the similar eco-physiological features of no salt requirement, positive or negative oxygen diffusion. The last category of AOBs includes Nitrosococcus genus: *Nitrosococcus oceani*, *Nitrosococcus halophilus*, *Nitrosococcus wardiae* and *Nitrosococcus watsonii*, all are obligate halophilic and found in marine environments, with the former having positive water solubility but the latter having negative water solubility (Soliman and Eldyasti, 2018).

1.3 Ammonia-oxidizing archaea (AOA)

The ammonia-oxidizing archaea (AOA) were found by the finding of ammoniaoxidizing genes in the Archaea domain at the beginning of the twenty-first century, when previously the existence of amoA genes were confined to bacteria only, according to a microbiome investigation of soil and marine samples. *Nitrosopumilus maritimus*, which is the first AOA, was obtained from a marine environment in Seattle (Bollmann, French and Laanbroek, 2011). AOAs have been found to contribute considerably to the global nitrogen cycle as the major nitrifiers in the marine and different soils. Since they can oxidise ammonia at considerably lower substrate concentrations than AOBs, they are likely to prevail in oligotrophic environments.

The Thaumarchaeota (or) Thaumarchaea are the ammonia-oxidizing archaea (AOA), which are the most common prokaryotes in nature, with a wide range of habitats including aquatic, land and geothermal (Konneke et al., 2014). This distinct and deep-branching phylum within the Archaea, which was formerly categorised as "mesophilic Crenarchaeota", was constituted from genetic testing. This new phylum contains not only all identified archaeal ammonia oxidizers, but also many groups of environmental genomes indicating bacteria with unknown metabolic processes, as observed from DNA sequencing of the 16S rRNA gene. Ecophysiological investigations of ammonia-oxidizing Thaumarchaeota indicate that they have

adapted to low ammonia levels and have an autotrophic or perhaps mixotrophic existence (Pester, Schleper and Wagner, 2011). Their high abundance in the ocean (up to 20% of all bacteria and archaea) and relatively low total ammonium substrate threshold focus on providing conclusive proof for their role as powerful ammonia oxidizers in the environment, in which their autotrophic (or potentially partially heterotrophic) lifestyle also contributes to primary production. Low K_m values (an index of the affinity of enzyme) for ammonia oxidation in unfertilized soils and wastewater might indicate that some AOA ecotypes contribute to nitrogen removal, especially when ammonia is limited. Nonetheless, it's feasible that certain Thaumarchaeota utilise additional substrates for energy production, or that they can just convert to ammonia oxidation only under specific environmental conditions (Pester, Schleper and Wagner, 2011). Additionally, the Nitrosopumilus genus is a significant AOA consists of *Nitrosopumilus maritimus, Nitrosopumilus adriaticus, Nitrosopumilus cobalaminigenes, Nitrosopumilus oxyclinae, Nitrosopumilus piranensis and Nitrosopumilus ureiphilus.*

Despite extensive cultivation efforts and knowledge of complete genome sequences, the role of AOB and AOA in the environment remains restricted. It is still hard to differentiate between AOB and AOA as the bacteria responsible for ammonia oxidation in the environment. More research must be mandatory in the future for the microbiologists to examine the reactions of well-characterized pure and enrichment behaviours of AOA and AOB to environmental circumstances to fully understand the task of these nitrifying bacteria in their natural habitats (Bollmann, French and Laanbroek, 2011).

1.4 Nitrite-oxidizing bacteria (NOB)

The nitrite-oxidizing bacteria (NOB), which are gram-negative bacteria, acquire their energy requirements by means of converting nitrite to nitrate. NOB are represented by the members of Alphaproteobacteria, Gammaproteobacteria and Deltaproteobacteria. Regarding their physiology, two main genera is considered for the NOB, which is the Nitrobacter genus and the Nitrospira genus. There are also other rare NOB genera, such as Nitrococcus and Nitrospina. Apart from the main source of nitrite, all members of the Nitrobacter genus, which include *Nitrobacter hamburgensis*, *Nitrobacter vulgaris* and *Nitrobacter winogradskyi*, may utilise organic energy sources. Nitrospira defluvii, *Nitrospira japonica*, *Nitrospira lenta* and *Nitrospira moscoviensis* are included in Nitrospira genus. Even though certain strains were

discovered from coastal areas or lakes, no obligatory salt necessity has been identified in Nitrobacter isolates. While members of the Nitrobacter genus appear to be free-living cells in their native habitats, Nitrospira has been found connected to flocs or biofilms on several occasions (Koops and Pommerening-Röser, 2001).

Usually, NOB are relatively easy to handle due to the fact that they can adapt to high substrate concentrations. The cultures of Nitrobacter and Nitrococcus reach turbidity visible to the naked eye, which is hard to figure out for another group, i.e., Nitrospira and Nitrospina. These groups of nitrifying bacteria grow solely under nitrite limitation. On top of that, as the members of the Nitrospira genus possess a high tendency for aggregation and growth, they are also accompanied by the formation of microcolonies which are noticeable as flocs (Spieck and Lipski, 2011). The grouping correlates with the fact that Nitrobacter consists of intracytoplasmic membranes which anchor the key catabolic enzyme, namely nitrite oxidoreductase (NXR), on the cytoplasmic side of the ICM. In accordance, a putative nitrite/nitrate antiport system is encoded in all sequenced Nitrobacter genomes. As in comparison, the NXR of the genomes Nitrospira and Nitrospina, which are more nitrite sensitive genera of NOB, which does not have ICM, can be found on the periplasmic side of the cytoplasmic membrane (Watson et al., 1989). With that being the case, the necessity of nitrite transport systems is not required. Consequently, the periplasmic orientation of NXR helps the latterly described genera to grow effectively in measures like lower concentrations of nitrite, comparatively better than Nitrobacter and Nitrococcus. In some situations, the inhibitory effects, caused by high nitrite concentrations on the growing process of those three genera, are sometimes a result of the external orientations. Due to the basis of the recently obtained genome sequence, the presence of two gene copies encoding a sub-unit of the NXR in candidate Nitrospira defluvii which are regulated could explain the high nitrite tolerance of the Nitrospira genus (Daims et al., 2010). When NOB cultures start to consume nitrite, they must replenish on a regular basis for the purpose of maintaining a sufficient number of cells for further analyses. For some specific strains like Nitrospira marina, when organic compounds are introduced to inorganic nutrient media, biomass development rises even more. In the case of Nitrobacter, it has the ability for heterogenic trophic growth, when the majority of other NOB do not assimilate carbon from organic substances (Spieck and Lipski, 2011).

NOB thrive in a variety of environments (terrestrial, marine, acidic) and have a variety of existences (autotrophic, mixotrophic, and heterotrophic) therefore different media compositions are needed to equivalent their specific growth needs in the lab. Consumption of high nitrite concentrations can generate a significant number of cells from Nitrobacter and Nitrococcus, but the accumulation of cells from Nitrospira and Nitrospina needs extended nutrition processes. Planktonic cells can be isolated using dilution series or plating procedures, however microcolony-forming strains, such as Nitrospira, are more difficult to isolate. Physiological experiments, such as determining the temperature or pH-optimum, is carried out using active laboratory cultures of NOB, although achieving reference values such as cell protein content or cell counts may be difficult due to formations of floc and poor cell density (Spieck and Lipski, 2011).

Another significant observation is that NOB can survive several years even though they're starved from energy and reductant. On the other hand, it takes weeks to months for NOB to reactivate their metabolic processes and resume growth. During the exponential phase of growth, some NOB grow relatively swiftly with minimal generation times of 8h for *Nitrobacter winogradskyi* or 12h for *Nitrospira moscoviensis* (Watson et al., 1989).

1.5 Related research on nitrifying bacteria in construction and built environment

According to related resources, the studies relevant to the nitrifying bacteria, construction, and the built environment that they are found mainly include wastewater treatment plants, soil, freshwater as well as other engineering systems. Engineers and microbiologists have been studying the relationship between nitrification efficiency and the organisation of the nitrifying bacteria species for over a decade.

Numerous AOB 16S rRNA gene sequences are being documented throughout multiple configurations, allowing the development of rRNA-based methods for the characterisation of AOB species composition in wastewater treatment plants. Additionally, the whole genome sequencing of the *Nitrosomonas europaea* species, coding for a subunit of ammonia monooxygenase (amoA), provides a new way to identify the AOB (McTavish et al., 1993). The development of the Nitrosospira genus can be enhanced under certain circumstances (usually low pH and low temperature) and/or in industrial wastewater treatment facilities (Siripong and Rittmann, 2007). According to the research, just one or two AOB species are generally detected in a wastewater treatment plant, however other systems might have a greater diversity of AOB species. Fluorescence in situ hybridization (FISH) has also been used to accurately measure AOB in activated sludge flocs and biofilms. Although FISH is a useful technique, it might be difficult to see stained cells in materials with a lot of autofluorescence (e.g., industrial wastewater). Since accurate counts necessitate a 103–104 cells/ml minimum

concentration, FISH also provides a very high detection limit. The abundance of AOB has recently been measured utilising quantitative polymerase chain reaction (qPCR) to quantify 16S rRNA or amoA gene copies (Bellucci and Curtis, 2011). This technique is a reliable, accurate, and quick quantitative technique that is a viable alternative to FISH, however, it is restricted by DNA extraction rate and PCR biases (Martin-Laurent et al., 2001).

In some situations, the members of the Nitrospira genus, which are retrieved from highly nitrogen loaded environments such as wastewater treatment plants or a marine biofilter, are observed with specifically high tolerance towards nitrite (Off et al., 2010). A new bacteria known as the Anammox (anaerobic ammonium oxidation) bacterium has been found in activated sludge, which is a prevalent biological operation for wastewater treatment (Whang, Chien, Yuan and Wu, 2009).

Pinto et al. (2016) present metagenomic evidence for the existence of a Nitrospira-like bacterium in a drinking water system that has the metabolic capacity to serve as an ammonia oxidizer. This genome was shown to be linked to Nitrospira-like NOB based on ribosomal proteins, 16S rRNA, and the nxrA gene. The presence of the whole ammonia oxidation gene family, including amoA and HAO, on a single structure inside this metagenome group, implies the existence of the previously identified comammox capability. The establishment of this structure inside the Nitrospira metagenome group is supported by studies focusing on coverage usage of two distinct genome binning methods, and nucleic acid and protein similarity studies. The amoA gene discovered in this metagenome group differs from traditional ammonia oxidizers and its clusters with the unique amoA gene of comammox Nitrospira. This discovery, which has a big impact on nitrogen conversions in both built and natural systems, implies that previously observed differences in the abundances of nitrifying bacteria might be explained by Nitrospira-like species' ability to fully oxidise ammonia.

Whang et al. (2009) stated that the physiological characteristics of nitrifying bacteria are reflected in the distribution of various nitrifying bacteria species in their surroundings. Ammonium and nitrite concentrations particularly are thought to be key determinants in the selection of different Nitrifying bacteria species. Species of the Nitrosospira genus and *Nitrosomonas oligotropha* species are the dominating AOB in low-ammonium environments, but perhaps the *Nitrosomonas europaea* cluster is prevalent in high-ammonium environments. Recent findings, supported by genetic analysis tools, show that Nitrosospira coexists alongside well-known Nitrosomonas and that Nitrospira is frequently the dominating NOB in activated sludge systems.

A recirculating aquaculture system is developed based on the utilization of wastewater using nitrifying bacteria as a technique for decreasing nitrogenous wastes in the production system. The formation of hazardous nitrogenous compounds resulting from excess and excrement is the major problem in intensive aquaculture. The use of AOB in the biofiltration process allows these nitrogen metabolites to be removed (Khangembam, Sharma and Chakrabarti, 2017). The Shedd Aquarium's nitrifying filtration unit yielded Nitrosopumilus maritimus, a marine AOA strain (Könneke et al., 2005). The biofilters must function consistently and reliably for a recirculating system to be successful.

Rather than correlation studies, recent investigations of grown ammonia-oxidizers and analytical designs give a more reliable method for assessing niche specialisation. The concept of niche specialisation makes an implicit assumption regarding environmental adaptation and the resulting connection among physiological and environmental variables. Microbial cultures of AOB and AOA are notoriously difficult to obtain. According to the implemented biological studies in soil, the cultivation is successful on various AOB species and only four AOA species. *Nitrososphaera viennensis* is the only purified species (Tourna, M. et al., 2011).

If the isolates are indicative of their belonging phylogenetic groups, as well as ecologically relevant communities, cultures provide quantitative kinetics study and give solid evidence of function and connections between function and particular genes. Genome sequencing becomes easier if cultivation is implemented and Prosser proved that by sequencing the genomes of two soil AOB (*Nitrosomonas europaea* and *Nitrosospira multiformis*) and one soil AOA (*Nitrosoarchaeum koreensis*) (Prosser and Nicol, 2012).

Anammox applications, partial nitrification, and long-term reliability are all important factors in minimizing energy consumption in conventional wastewater. This is especially important for household wastewater with a low COD ratio. Most quick start-up PN methods now have a negative impact on NOB and only a minor negative impact on AOB. "Selective elimination" refers to the technique that results in variations in particular growth rates between the different nitrifying bacteria groups (Li et al., 2021).

In most investigations, it is therefore unavoidable to lower both AOB and NOB apparent growth rates, however, this would result in more significant reductions in NOB growth rates (Larriba et al., 2020). The method of "simultaneous elimination and selective enrichment" for achieving or recovering PN has been developed. According to the study, establishing the optimal FNA levels is difficult, and treatment with a moderate FNA level inhibited nitrifying bacteria. Then, in order to attain PN, higher DO levels and a longer aeration time was employed (Wang et al., 2020).

Another research by Liu et al. (2017) states that during the first 7 days of aerobic deprivation, nitrifying bacteria gradually degraded, consequently, a slower selective recovery potential of NOB than AOB, allowing PN to recover more quickly. The bioactivities of heterotrophic organisms (OHOs) were reduced at the same time in the aerobic phase when DO was limited (Li et al., 2021). At the early aerobic phase, OHOs require a surge of oxygen, while concurrently inhibiting nitrifying bacteria. This restriction, however, is only temporary and can be removed. OHOs are hindered in the absence of COD, resulting in flourishing nitrification with enough oxygen (Li et al., 2017).

In a study in China, related to nitrifying bacteria in wastewater treatment, a 285-day run of a lab-scale SBR filled with residential wastewater was used to examine nitrogen removal and nitrite formation, as well as the potential of attaining PN quickly by adjusting aeration time. To further validate the primary processes for attaining PN, the abundance and distribution of nitrifying bacteria were evaluated. The alterations of core bacteria in the genomic structure were evaluated using moderate genome sequencing analysis, and the potential in the residential wastewater anammox treatment process was investigated (Li et al., 2021).

1.6 Aims and Objectives

The main aim of the study is to investigate the genomes and the presence of genes in nitrifying bacteria using whole genome analysis. The analysis will be accomplished for each genus of the nitrifying bacteria individually. From determining the presence of the genes in the nitrifying bacteria, the genes in the species affecting the nitrification process that occurred in engineering systems such as soil, wastewater and freshwater will be indicated.

The objectives of this study are as follows:

- To annotate the genomes using Prokka.
- To conduct pangenome analysis using Roary for
 - a. AOB: Nitrosomonas genus, Nitrosospira genus, Nitrosococcus genus,
 - b. AOA: Thaumarchaea, Nitrosopumilus genus,
 - c. NOB: Nitrobacter genus, Nitrospira genus.
- To generate metabolic profiles and biogeochemical cycling diagrams using METABOLIC.
- To compare the functional capabilities within each genus.

CHAPTER 2

METHODOLOGY

The whole genome analysis, which is also called whole genome sequencing (WGS) is a technique being used to determine the complete deoxyribonucleic acid (DNA) sequence of an organism's genome at once. This requires sequencing all the individual's chromosomal DNA, as well as mitochondrial DNA. Full genome sequencing, complete genome sequencing, and entire genome sequencing are all approaches for sequencing the entire genome. WGS has mostly been utilised for scientific purposes. WGS data will be an essential tool to assist medical intervention in the future of new treatments. The method of single nucleotide polymorphisms (SNP) level gene sequencing is also used to locate operational variations from prospective studies and enhance the knowledge accessible to evolutionary biologists and environmental biotechnologists, perhaps laying the groundwork for bioinformatics studies. WGS is not the same as DNA profiling, since it identifies not only the possibility that genetic material comes from a specific individual or group but also the supplementary information of genetic connections. Furthermore, thousands of genomes have been wholly or partially sequenced in the present (Kamran, 2018).

The research and analysis procedure of this project is carried out by means of the whole genome sequencing technique. The following methodology is applied for this project.

- 1) The determination of the names of the prokaryotes (bacteria and archaea) from the literature review of related studies concerning the nitrifying bacteria
- 2) The acquisition of the genomes from the selected species of the genus for the analysis
- The annotation process of the downloaded genomes with the use of relevant software and database
- 4) The analysis of the annotated genomes to obtain pan-genome datasets using the appropriate pipeline
- 5) The analysis of the metabolic competencies of the genomes utilizing the related software
- The investigation and discussion of the results obtained from the whole genome analysis workflow

All related workflows and tutorials were provided by Dr Umer Zeeshan Ijaz (<u>http://userweb.eng.gla.ac.uk/umer.ijaz/</u>), at the University of Glasgow.

The whole genome analysis workflow was carried out on the **Orion Cluster**, which is a bioinformatics high-performance computing facility hosted and developed by Dr Umer's Environmental'Omics lab. Orion is a bioinformatics cluster with over 500 tools and workflows for metagenomics innovations including those developed by Dr Umer's lab and is specially designed for projects with big data analytics. (http://userweb.eng.gla.ac.uk/umer.ijaz/#orion) To be able to get access to the Orion Cluster, the 'terminal' program from MacOS is used as the SSH client.

2.1 Initial preparation on genomes

The genomes needed for the research were downloaded by each species from the NCBI database, which is maintained by the National Centre for Biotechnology Information. The remote cluster was used to store all the downloaded data. During this procedure, relevant genomic data files including details of gene sequences, which were retrieved from a database, were stored in directories corresponding to their species names under their genus names.

Even though all the genomes of the species under the entire AOB, NOB and AOA groups were downloaded initially, before doing the continuing analysis, the genus which only has one species, and one genome were eliminated due to the requirement of the research. The following table 1 is the selected species under the corresponding genus with the number of genomes.

| Group | Genus | Total genomes | Species | Number of genomes |
|------------------|-------|--------------------------|-----------------------|----------------------|
| | | | Nitrosomonas europaea | 5 |
| | | | Nitrosomonas eutropha | 8 |
| AOB Nitrosomonas | 43 | Nitrosomonas halophila | 1 | |
| | | Nitrosomonas mobilis | 1 | |
| | | Nitrosomonas communis | 4 | |
| | | Nitrosomonas nitrosa | 5 | |
| | | Nitrosomonas oligotropha | 5 | |
| | | Nitrosomonas ureae | 8 | |

Table 1: Selected species of the nitrifying bacteria

| | | | Nitrosomonas marina | 2 |
|-----|----------------|-----|--------------------------------|----|
| | | | Nitrosomonas aestuarii | 2 |
| | | | Nitrosomonas cryotolerans | 2 |
| | | | Nitrosospira briensis | 3 |
| | Nitrosospira | 18 | Nitrosospira lacus | 1 |
| | | | Nitrosospira multiformis | 14 |
| | | | Nitrosococcus halophilus | 1 |
| | Nitrosococcus | 8 | Nitrosococcus oceani | 5 |
| | Millosococcus | 0 | Nitrosococcus wardiae | 1 |
| | | | Nitrosococcus watsonii | 1 |
| | | | Nitrobacter hamburgensis | 1 |
| | Nitrobacter | 4 | Nitrobacter vulgaris | 1 |
| | | | Nitrobacter winogradskyi | 2 |
| NOB | Nitrospira | 10 | Nitrospira defluvii | 7 |
| | | | Nitrospira japonica | 1 |
| | | | Nitrospira lenta | 1 |
| | | | Nitrospira moscoviensis | 1 |
| | | | Nitrososphaeria archaeon | 46 |
| | Thaumarchaea | 116 | Nitrososphaerales archaeon | 11 |
| | | | Nitrososphaeraceae archaeon | 13 |
| | | | Nitrosopumilaceae archaeon | 46 |
| AOA | | 6 | Nitrosopumilus maritimus | 1 |
| AUA | Nitrosopumilus | | Nitrosopumilus adriaticus | 1 |
| | | | Nitrosopumilus cobalaminigenes | 1 |
| | | | Nitrosopumilus oxyclinae | 1 |
| | | | Nitrosopumilus piranensis | 1 |
| | | | Nitrosopumilus ureiphilus | 1 |

Note: Thaumarchaea is not a genus, it is a phylum and classifications under Thaumarchaea are archaeon.

2.2 Primary analysis of genomes

Following the downloading and systematic categorization process of genomic data, as described above, three software programmes were used to complete the genome analysis process: the Prokka workflow, the Roary workflow, and the METABOLIC workflow. The detailed command lines that are used for the workflows of Prokka, Roary and METABOLIC are shown in the <u>APPENDIX</u> section.

2.2.1 Prokka workflow

Prokka is a software tool used for rapid prokaryotic genome annotation, the process of determining characteristics of interest in a set of genomic DNA sequences and identifying those with relevant information. This programme allows users to annotate the genomes of bacteria, archaea, and viruses efficiently in a short period of time, then generate standards-compliant output files containing sequences and annotations (Seemann, 2014).

In this Prokka workflow, the downloaded genomic FASTA files with FNA extensions are used as input files and twelve separate files are produced for each genome as output files, which include a master annotation in GFF3 format, standard Genbank file in GBK format and multiple FASTA files.

2.2.2 Roary workflow

Roary is a high-speed pan-genome pipeline, that generates the pan-genome using annotated assemblies in GFF3 format. It can implement the analysis of data statistics with thousands of samples, something that is practically impossible with conventional approaches, without degrading the quality of the results, by using only a typical computer. Roary can perform the analysis containing 128 samples in less than an hour with 1 GB of RAM and a single CPU, while other techniques would require weeks and hundreds of GB of RAM to complete that analysis (Page et al., 2015).

The GFF3 files obtained from the Prokka workflow was used as input files for this analysis. The GFF3 files of an entire genus were put into a new directory since Roary had to be executed for each genus separately. A spreadsheet list that shows the presence and absence of the genes in the genomes is obtained as the main output for further results visualisations.

The output files include three figures as Roary plots. The first plot shows a tree that is compared to a matrix with the presence and absence of core and accessory genes. The second one is a graph illustrating the proportion of genes vs the number of genomes. Finally, a pie chart depicting the distribution of genes and the number of isolates in which they are found is included.

2.2.3 METABOLIC workflow

METABOLIC (Metabolic and Biogeochemistry Analyses in Microbes) is a scalable programme that analyses genomes at the level of individual organisms and/or microbial groups to enhance microbial ecology and biogeochemistry. Annotation of microbial genomes, sequence identification of biochemically verified invariant protein molecules, determination of metabolism indicators, metabolic pathway studies, and contribution to specific biogeochemical alterations and cycles are all part of the genome-scale procedure. The input genomes for METABOLIC are normally isolates, metagenome-assembled genomes, or single-cell genomes. The community-scale process adds to genetic code analytics by calculating microbial community contributions to biogeochemical cycles, determining genome abundance in the group, and identifying possible microbial metabolic handoffs and exchange. METABOLIC perform better than any other competing software and servers in terms of accuracy, resilience, and consistency, according to the evaluations. A variety of metagenomic datasets from the terrestrial subsurface, marine subsurface, soil, wastewater plants, freshwater lakes, and even human gut are capable to be analysed on the METABOLIC workflow. METABOLIC is available in two versions: METABOLIC-G and METABOLIC-C. METABOLIC-G.pl does not require the input of sequencing data to generate metabolic profiles and biogeochemical cycle diagrams for input genomes. METABOLIC-C.pl produces the same results as METABOLIC-G.pl, but it also creates assistance in terms of metabolism since it accepts metagenomic read files and figures out how much of the genome is covered (Zhou et al., 2019).

METABOLIC-G.pl is used for the workflow of this project. Since the input genomic data for the METABOLIC workflow must be in FASTA file format, the format of FNA files is transformed to FASTA file format. The output results include the METABOLIC result in the form of spreadsheets for metabolism as well as a range of visualisations such as biogeochemical cycle potential, presentation of consecutive metabolic transformations, and nutrient cycling diagrams.

2.3 Visualisation of the results

The output data can be interactively visualised using additional software tools. For this whole genome analysis project, the results are visualised by running Coinfinder on Orion Cluster and producing pan-genome plots via RStudio.

2.3.1 Coinfinder

Coinfinder is a software tool used to determine the association or dissociation condition of the groups of genes (gene families) in pan-genomes from one another more frequently than would be anticipated by coincidental (Whelan, Rusilowicz and McInerney, 2020). Coinfinder is implemented with Python programming language on Orion Cluster provided by Dr Umer. The Roary output data: gene_presence_absence.csv and core-gps_fasttree.newick are used as input files together with the Coinfinder manuscript, which is downloaded using the git clone command into the respective directory on the cluster. The output data generated from Coinfinder are coincidence heatmap and genes network.

2.3.2 RStudio

Another result visualisation process was operated on the RStudio platform, an integrated development environment for the R programming language that includes a terminal, a syntax-highlighting editor and the features for visualising plots, mapping, troubleshooting and workspace management (RStudio | Open source & professional software for data science teams, 2021).

The pan-genome plots are produced via RStudio using R packages: vegan and ggplot2. Vegan is a package for environmental scientists, which includes tools for diversity analysis, coordination techniques, and dissimilarity analysis (Oksanen, 2020). ggplot2 is a free data visualisation tool for the R statistical programming language based on 'the Grammar of Graphics' (Wickham and Grolemund, 2010). It may be used to declare a graphic's source data frame and to establish a set of plot aesthetics that will be shared by all successive layers.

The detailed command lines that are used for the workflows of Coinfinder and RStudio are shown in the <u>APPENDIX</u> section.

2.4 Description of the functional genes in the results

While focusing on the presence and absence of the genes in the genomes or concerning the chemical reactions in the genomes, there are important genes to be emphasised. Since all the analyses of this research are about nitrifying bacteria, the genes that are focused on will be the genes that are functional in nitrogen cycling. The important genes to look for in the results are shown in table 2 with their corresponding functions and abbreviations.

| Genes | Gene Names | Functions | |
|------------------|-------------------------------------|-------------------------|--|
| amoA, amoB, amoC | ammonia monooxygenase | Ammonia oxidation | |
| anfD, anfK, anfG | nitrogenase iron-iron protein | | |
| nifD, nifK | nitrogenase molybdenum-iron protein | Nitrogen fixation | |
| vnfD, vnfK, vnfG | nitrogenase vanadium-iron protein | | |
| nifH | nitrogenase iron protein | | |
| nxrA, nxrB | nitrite oxidoreductase | Nitrite oxidation | |
| napA, napB | periplasmic nitrate reductase | Nitrate reduction | |
| narG, narH | nitrite oxidoreductase | Initiale reduction | |
| nrfH, nrfA, nrfD | cytochrome c nitrite reductase | Nitrite reduction to | |
| nirB, nirD, nirK | nitrite reductase | Ammonia | |
| nirS | hydroxylamine reductase | Nitrite reduction | |
| norB, norC | nitric oxide reductase | Nitric oxide reduction | |
| nosD, nosZ | nitrous-oxide reductase | Nitrous oxide reduction | |

Table 2: Functional genes in the nitrogen cycle

Note: The alphabets at the end of the gene abbreviations belong to the sub-unit of that gene, e.g., subunit A, subunit B, alpha subunit, beta subunit, etc.

CHAPTER 3

RESULTS

The findings of this project's whole genome analysis are primarily split into three parts for AOB, NOB and AOA respectively. In each part, the results are described for an individual genus using visualisations such as figures and charts.

The presence and absence of genes in the genomes are discussed widely in this chapter. For each genus, a summary of statistics such as the number of total genes found in the whole genus, the core genes which are found in all the genomes in the genus, the soft-core genes which are found in almost every genome in the genus, is presented. The metabolic chemical reactions discovered in the genomes are also discussed with relevant justifications.

3.1 Results of ammonia-oxidizing bacteria (AOB)

While discussing the outcomes obtained from whole genome analysis in AOB, results for each genus is presented initially. Then, the overall genomes for AOB are discussed with the use of a pan-genome plot.

3.1.1 Results of Nitrosomonas genus

The whole genome analysis of the Nitrosomonas genus includes 43 genomes in total. The statistical result of genes from the Roary workflow is shown in table 3.

| Core genes | $(99\% \le \text{strains} \le 100\%)$ | 1 |
|-----------------|---------------------------------------|-------|
| Soft-core genes | $(95\% \le \text{strains} < 99\%)$ | 3 |
| Shell genes | $(15\% \le \text{strains} < 95\%)$ | 2328 |
| Cloud genes | $(0\% \le \text{strains} < 15\%)$ | 51231 |
| Total genes | $(0\% \le \text{strains} \le 100\%)$ | 53563 |

Table 3: Summary statistics of genes for Nitrosomonas genus

The only core gene that is found in all Nitrosomonas genus is atpH, an ATP sub-unit. Soft-core genes are a cold shock-like protein named cspD, the ribosomal proteins S10 and S12, rpsJ and rspL respectively.



Figure 2: Pan-genome frequency and pan-genome pie of Nitrosomonas genus

The Roary plots are shown in figures 2 and 3 describing the pan-genome frequency, pan-genome pie, and pan-genome matrix respectively.



Figure 3: Pan-genome matrix of Nitrosomonas genus

From the METABOLIC results, amoA, amoB, amoC, nirK, norB, norC are the only significant genes that are present in the Nitrosomonas genus. So, only ammonia oxidation, nitrite reduction to ammonia and nitric oxide reduction occurs in this genus. The genomes of *Nitrosomonas europaea*, *Nitrosomonas cryotolerans* and *Nitrosomonas mobilis* have all the above genes so they are considered as the best ammonia oxidizers in this genus. *Nitrosomonas marina* and *Nitrosomonas aestuarii* absent amoC gene in their genomic sequence.

Nitrosomonas eutropha has alternative gene absence in their genomes, whereas most *Nitrosomonas ureae* absent nor genes. *Nitrosomonas nitrosa* is considered to have the least performance in nitrification since they are missing almost every gene in the nitrogen cycle and two of its genomes present no genes from that cycle.

3.1.2 Results of Nitrosospira genus

In total, 18 genomes from the Nitrosospira genus have been analysed. The genes result from the Roary workflow is shown in table 4.

| Core genes | $(99\% \le \text{strains} \le 100\%)$ | 31 |
|-----------------|---------------------------------------|-------|
| Soft-core genes | $(95\% \le \text{strains} < 99\%)$ | 0 |
| Shell genes | $(15\% \le \text{strains} < 95\%)$ | 4133 |
| Cloud genes | $(0\% \le \text{strains} < 15\%)$ | 19734 |
| Total genes | $(0\% \le \text{strains} \le 100\%)$ | 23898 |

 Table 4: Summary statistics of genes for Nitrosospira genus

The core genes commonly presented in the Nitrosospira genus include subunits of ribosomal proteins, ATP synthases and proteases, NADH-quinone oxidoreductases and other genes. Anaerobic nitric oxide reductase is a significant gene that is found in almost every genome.



Figure 4: Pan-genome frequency and pan-genome pie of Nitrosospira genus

Figures 4 and 5 illustrate the pan-genome frequency, pan-genome pie, and pan-genome matrix, which are output plots from Roary.



Figure 5: Pan-genome matrix of Nitrosospira genus

The presence of important genes in *Nitrosospira multiformis* is nearly the same as with Nitrosomonas genus since all subunits of amo, nirK, norB and norC are present in all of the genomes. The genomes of *Nitrosospira briensis* and *Nitrosospira lacus* present all the amo subunits and nirK as other Nitrosospira genomes but nitric oxide reductase (nor) genes are absent in them. All the remaining genes from the nitrogen cycle are absent in the Nitrosospira genus.

3.1.3 Results of Nitrosococcus genus

The whole genome analysis of the Nitrosococcus genus includes 8 genomes. The following table 5 is the statistical outcome of genes from the Roary process.

| Core genes | $(99\% \le \text{strains} \le 100\%)$ | 26 |
|-----------------|---------------------------------------|-------|
| Soft-core genes | $(95\% \le \text{strains} < 99\%)$ | 0 |
| Shell genes | $(15\% \le \text{strains} < 95\%)$ | 4245 |
| Cloud genes | $(0\% \le \text{strains} < 15\%)$ | 9054 |
| Total genes | $(0\% \le \text{strains} \le 100\%)$ | 13325 |

Table 5: Summary statistics of genes for Nitrosococcus genus

Generally, the core genes presented in the Nitrosococcus genus are like those in the Nitrosospira genus since they are also subunits of ribosomal proteins, ATP synthases and proteases and NADH-quinone oxidoreductases.



Figure 6: Pan-genome frequency and pan-genome pie of Nitrosococcus genus

Figures 6 and 7 represent the pan-genome frequency, pan-genome pie, and pan-genome matrix, correspondingly, using Roary plots.



Figure 7: Pan-genome matrix of Nitrosococcus genus

As common in other AOB species, amoA, amoB, amoC, nirK, norB and norC are the significant genes found and other remaining genes from the nitrogen cycle are absent in the Nitrosococcus genus. However, amoA and amoC do not exist in all the genomes of *Nitrosococcus halophilus*, *Nitrosococcus wardiae* and some genomes of *Nitrosococcus oceani*.

The genomic sequence of *Nitrosococcus watsonii* presents all the above genes and additionally, it also presents nitrite reductase, nrfA.

3.1.4 Pan-genome plot of AOB

The differentiation of AOB genomes depending on the genes' presence and absence is plotted using the RStudio. As shown in figure 8, the AOB genomes have similar sequences since they all are at the same side of the plot.



Figure 8: Pan-genome plot of AOB

Moreover, Nitrosospira genomes and Nitrococcus genomes are more similar in their genomic sequences. GCA 003051045, a genome of Nitrosomonas nitrosa is the only genome that is a lot more alternative from the other AOB genomes.

3.2 Results of ammonia-oxidizing archaea (AOA)

In the genome analysis of AOA, instead of analysing each archaeon including in the AOA of nitrifying bacteria, the whole phylum Thaumarchaea is considered as a single analysis. However, the Nitrosopumilus genus, the most significant genus in Thaumarchaeota, has been analysed separately.

3.2.1 Results of Thaumarchaea

The whole genome analysis of the Thaumarchaea phylum includes 116 genomes in total. The statistical result of genes from the Roary workflow is presented in table 6.

| Core genes | $(99\% \le \text{strains} \le 100\%)$ | 0 |
|-----------------|---------------------------------------|--------|
| Soft-core genes | $(95\% \le \text{strains} < 99\%)$ | 0 |
| Shell genes | $(15\% \le \text{strains} < 95\%)$ | 25 |
| Cloud genes | $(0\% \le \text{strains} < 15\%)$ | 101709 |
| Total genes | $(0\% \le \text{strains} \le 100\%)$ | 101734 |

Table 6: Summary statistics of genes for Thaumarchaea

The significance of this phylum Thaumarchaea is that there are neither core genes nor soft-core genes. Therefore, the shell genes of the genomes of this phylum are investigated and most of them are hypothetical proteins, ribosomal proteins, and DNA-directed RNA polymerase subunits.



Figure 9: Pan-genome frequency and pan-genome pie of Thaumarchaea

The Roary plots are shown in figures 9 and 10 describing the pan-genome frequency, pan-genome pie, and pan-genome matrix respectively.



Figure 10: Pan-genome matrix of Thaumarchaea

Only a few genes from the nitrogen cycle, nirK, nosZ and nrfA are present in the genomes of *Nitrososphaeria archaeon*. Only one genome from *Nitrososphaerales archaeon* presents nitrite oxidoreductase subunit (nxrA), a gene from the nitrogen cycle. *Nitrososphaeraceae archaeon* and *Nitrosopumilaceae archaeon* present none of the genes from the nitrogen cycle. Generally, Thaumarchaea genomes do not involve many nitrogen cycling processes in their metabolism. Apart from that, they have some chemical reactions such as amino acid utilization, metabolism of organic sulfur and sulfur oxidation.

3.2.2 Results of Nitrosopumilus genus

Totally, the whole genome analysis of the Nitrosopumilus genus includes 6 genomes. The following table 7 is the statistical outcome of genes from the Roary process.

| Core genes | $(99\% \le \text{strains} \le 100\%)$ | 63 |
|-----------------|---------------------------------------|-------|
| Soft-core genes | $(95\% \le \text{strains} < 99\%)$ | 0 |
| Shell genes | $(15\% \le \text{strains} < 95\%)$ | 11175 |
| Cloud genes | $(0\% \le \text{strains} < 15\%)$ | 0 |
| Total genes | $(0\% \le \text{strains} \le 100\%)$ | 11238 |

Table 7: Summary statistics of Nitrosopumilus genus

The core genes that are mostly found in the Nitrosopumilus genus are hypothetical protein groups. Other core genes existed in Nitrosopumilus genomes are DNA-directed RNA polymerase subunits, NADH quinone oxidoreductase subunits, hydroxy pyruvate reductases, cold shock-like proteins, ribonucleases etc. The following figures 11 and 12 demonstrate the pan-genome frequency, pan-genome pie, and pan-genome matrix, correspondingly, using Roary plots.



Figure 11: Pan-genome frequency and pan-genome pie of Nitrosopumilus genus

Similar to most of the Thaumarchaea genomes, every genome in the Nitrosopumilus genus lacks nitrogen cycling genes in their genomic sequences.



Figure 12: Pan-genome matrix of Nitrosopumilus genus

3.2.3 Pan-genome plot of AOA

The pan-genome plot shows the variation between AOA genomes upon the genes' presence and absence. Nitrosopumilus genomes have similar genomic sequences to the Nitrosopumilaceae genomes according to figure 13. Nitrosophaeria genomes are considered to have the most common sequences since they are spread all over the plot like other different archaeon genomes. Most of the Nitrosophaerales genomes are comparable to the Nitrosophaeria genomes. Nitrosophaeraceae genomes and Nitrosopumilaceae genomes are quite similar in their genomic sequences according to the pan-genome plot.



Figure 13: Pan-genome plot of AOA

3.3 Results of nitrite-oxidizing bacteria (NOB)

The whole genome analysis results for each genus of NOB are discussed with respective visualisations. After that, the evaluation of overall genome plot for NOB is developed.

3.3.1 Results of Nitrobacter genus

In total, 4 genomes from the Nitrobacter genus have been analysed. The genes result from the Roary workflow is shown in table 8.

| Core genes | $(99\% \le \text{strains} \le 100\%)$ | 217 |
|-----------------|---------------------------------------|-------|
| Soft-core genes | $(95\% \le \text{strains} < 99\%)$ | 0 |
| Shell genes | $(15\% \le \text{strains} < 95\%)$ | 12907 |
| Cloud genes | $(0\% \le \text{strains} < 15\%)$ | 0 |
| Total genes | $(0\% \le \text{strains} \le 100\%)$ | 13124 |

Table 8: Summary statistics of genes for Nitrobacter genus

From the statistics, the fact that the Nitrobacter genus has a large number of core genes compared to other nitrifying bacteria genera is observed.



Figure 14: Pan-genome frequency and pan-genome pie of Nitrobacter genus

The above figure 14 illustrates the pan-genome frequency and pan-genome pie, whereas the following figure 15 shows the pan-genome matrix, which are output plots from Roary.



Figure 15: Pan-genome matrix of Nitrobacter genus

Since the Nitrobacter genus contains the species which are nitrite oxidizers, none of the amo genes will be present in its genomes. The genomes of Nitrobacter hamburgensis and Nitrobacter vulgaris involve narG, narH, nirB, nirD and nirK genes in their metabolism. All the genes mentioned above except narG are present in the genomes of Nitrobacter winogradskyi. Therefore, these genomes have efficiency in reducing nitrates or nitrites.

3.3.2 Results of Nitrospira genus

The whole genome analysis of the Nitrospira genus includes 10 genomes in total. The statistical result of genes from the Roary workflow is shown in table 9.

| Core genes | $(99\% \le \text{strains} \le 100\%)$ | 0 |
|-----------------|---------------------------------------|-------|
| Soft-core genes | $(95\% \le \text{strains} < 99\%)$ | 0 |
| Shell genes | $(15\% \le \text{strains} < 95\%)$ | 2100 |
| Cloud genes | $(0\% \le \text{strains} < 15\%)$ | 28055 |
| Total genes | $(0\% \le \text{strains} \le 100\%)$ | 30155 |

Table 9: Summary statistics of genes for Nitrospira genus

In the Nitrospira genus, there are no core genes and soft-core genes existed in the genomes. The genes that are present in most genomes of this genus are transcriptional regulatory protein, an ATP synthase subunit, an oxalate oxidoreductase subunit hypothetical protein and a 30S ribosomal protein.



Figure 16: Pan-genome frequency and pan-genome pie of Nitrospira genus

The Roary plots are shown in figures 16 and 17 describing the pan-genome frequency, pan-genome pie, and pan-genome matrix respectively.



Figure 17: Pan-genome matrix of Nitrospira genus

From the METABOLIC results, Nitrospira lenta and Nitrospira japonica contain the nxrA, nxrB, nirD and nirK genes from the nitrogen cycle. Nitrospira moscoviensis present nxrA, nxrB, nrfH and nirK genes in its genome. All the genomes of Nitrospira defluvii have the genes nxrA, nxrB, nirD and nirK in their sequence where one genome of Nitrospira defluvii presents additional nrfH, nrfA and norC. So, these Nitrospira genomes have chemical reactions for nitrite oxidation, nitrite reduction and nitrous oxide reduction in their metabolism.

3.3.3 Pan-genome plot of NOB

When observing the pan-genome plot of NOB genomes illustrated in figure 18, it is significant that the genomes of two NOB genera have different genomic sequences. All the Nitrobacter genomes are similar to each other whereas, the Nitrospira genomes have some alterations between their sequences.



Figure 18: Pan-genome plot of NOB

3.4 Coinfinder output

The heatmap, which is an output from Coinfinder, for this whole genome analysis of nitrifying bacteria is shown in figure 19.



Figure 19: Output heatmap of pan-genomes for nitrifying bacteria

CHAPTER 4

DISCUSSION

In this project, the whole genome sequences of nitrifying bacteria are analysed diversely by several workflows. Generally, from observing the genomes sequences of different nitrifying bacteria genera, the differences of genes in these bacteria and archaea, metabolism occurring in the genomes, metabolic chemical reactions are discovered more clearly. The functional genes found in each nitrifying bacteria genus are listed in table 10.

| Genus | Genes found |
|----------------|--|
| Nitrosomonas | amoA, amoB, amoC, nirK, norB, norC |
| Nitrosospira | amoA, amoB, amoC, nirK, norB, norC |
| Nitrosococcus | amoA, amoB, amoC, nirK, norB, norC |
| Thaumarchaea | nirK, nosZ, nxrA, nrfA |
| Nitrosopumilus | - |
| Nitrobacter | narG, narH, nirB, nirD, nirK |
| Nitrospira | nxrA, nxrB, nirD, nirK, nrfH, nrfA, norC |

Table 10: Functional genes found in each nitrifying bacteria genus

Among the functional genes responsible for nitrogen transformations, most of the AOB genomes contain amoA, amoB, amoC, nirK, norB and norC. Therefore, most of these AOB species are functional only in ammonia oxidation, nitrite reduction to ammonia and nitric oxide reduction. Only a few species from AOB are able to perform other nitrogen cycling processes. Among the AOB species, *Nitrosomonas europaea* is considered to be the most efficient ammonia-oxidizing bacteria due to the genes presenting in its genomes.

A remarkable finding that is found through whole genome analysis of AOA is a lot of Thaumarchaea genomes absent functional genes for nitrification in their genomic sequences. Only a few functional genes, nirK, nosZ, nrfA and nxrA are present in AOA genomes. The function of most AOA are nitrite reduction to ammonia and nitrous oxide reduction. According to the genes existing in their genome sequencing, AOB are the better option as ammonia oxidizers than AOA. The NOB species present several functional genes in the nitrogen cycle such as narG, narH, nirB, nirD, nirK, nxrA, nxrB, nrfH, nrfA and norC. Thus, these NOB species are efficient in operating reduction of nitrates, nitrites, or nitrous oxide, as well as nitrite oxidation. *Nitrospira defluvii*, *Nitrobacter hamburgensis* and *Nitrobacter vulgaris* are the most effective nitrite oxidizers among the NOB species.

CHAPTER 5

CONCLUSION

Based on the analysis of whole genome sequencing, this project demonstrates the important role of genes and genome sequences in nitrifying bacteria in construction and the built environment. Bacteria and archaea species of Nitrosomonas genus, found in wastewater treatment systems, activated sludge, soil, and freshwater, Nitrosospira genus and Nitrospira genus, found in soil and water, Nitrosococcus, found in freshwater systems Nitrobacter genus, found in soil and freshwater systems, and Thaumarchaea found in several engineering systems are analysed respectively in this project.

In this project, I have successfully annotated the genomes of nitrifying bacteria are using Prokka and conducted pan-genome analysis using annotated genomes in Roary tool. The metabolic profiles and biogeochemical cycles are also generated using METABOLIC. The number of genes present in each genus and the pan-genome plots for each genus is studied. Moreover, the functional genes from the nitrogen cycle are discovered in these nitrifying bacteria. By investigating the genome sequences, strong ammonia oxidizers and nitrite oxidizers are also considered. This genomic study of nitrifying bacteria advance the knowledge on importance of genes in the metabolic activities of genomes which have direct impact on the corresponding system. For instance, the stronger ammonia oxidizer with more functional genes might develop the process of wastewater treatment plants.

However, there are still some genus and individual species that must be left out due to the limitations of the analysis of this study, especially for ammonia-oxidizing archaea. More detailed analysis might be required to understand more on the genome sequences of the nitrifying bacteria. Further analysis on more nitrifying bacteria genera for the engineering systems and more specific details of the species relating with genes and their functions impacting on the corresponding environment might advantage the role of microbial communities in the engineering systems.

REFERENCES

- Belser, L.W., 1979. Population ecology of nitrifying bacteria. Annual reviews in microbiology, 33(1), pp.309-333.
- Bellucci, M. and Curtis, T.P., 2011. Ammonia-oxidizing bacteria in wastewater. *Methods in enzymology*, 496, pp.269-286.
- Bollmann, A., French, E. and Laanbroek, H.J., 2011. Isolation, cultivation, and characterization of ammonia-oxidizing bacteria and archaea adapted to low ammonium concentrations. *Methods in enzymology*, 486, pp.55-88.
- Coskuner, G., Ballinger, S.J., Davenport, R.J., Pickering, R.L., Solera, R., Head, I.M. and Curtis, T.P., 2005. Agreement between theory and measurement in quantification of ammonia-oxidizing bacteria. *Applied and Environmental Microbiology*, 71(10), pp.6325-6334.
- 5. Daims, H., Lücker, S., Paslier, D.L. and Wagner, M., 2011. Diversity, environmental genomics, and ecophysiology of nitrite-oxidizing bacteria. *Nitrification*, pp.295-322.
- Dworkin, M. and Gutnick, D., 2012. Sergei Winogradsky: a founder of modern microbiology and the first microbial ecologist. *FEMS microbiology reviews*, 36(2), pp.364-379.
- Kamran, A., Whole genome sequencing and analysis: Molecular plant biodiversity and DNA analysis, Department of Botany, University of the Punjab (2018)
- 8. Khangembam, C.D., Sharma, J.G. and Chakrabarti, R., 2017. Diversity and abundance of ammonia-oxidizing bacteria and archaea in a freshwater recirculating aquaculture system. *HAYATI Journal of Biosciences*, *24*(4), pp.215-220.
- Könneke, M., Bernhard, A.E., José, R., Walker, C.B., Waterbury, J.B. and Stahl, D.A., 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, 437(7058), pp.543-546.
- Könneke, M., Schubert, D.M., Brown, P.C., Hügler, M., Standfest, S., Schwander, T., von Borzyskowski, L.S., Erb, T.J., Stahl, D.A., and Berg, I.A., 2014. Ammoniaoxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proceedings of the National Academy of Sciences*, 111(22), pp.8239-8244.
- Koops, H.P. and Pommerening-Röser, A., 2001. Distribution and ecophysiology of the nitrifying bacteria emphasizing cultured species. *FEMS Microbiology ecology*, 37(1), pp.1-9.

- 12. Li, J., Zhang, L., Peng, Y. and Zhang, Q., 2017. Effect of low COD/N ratios on stability of single-stage partial nitritation/anammox (SPN/A) process in a long-term operation. *Bioresource technology*, *244*, pp.192-197.
- 13. Li, S., Li, J., Yang, S., Zhang, Q., Li, X., Zhang, L. and Peng, Y., 2021. Rapid achieving partial nitrification in domestic wastewater: Controlling aeration time to selectively enrich ammonium oxidizing bacteria (AOB) after simultaneously eliminating AOB and nitrite oxidizing bacteria (NOB). *Bioresource Technology*, 328, p.124810.
- Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J.C., Soulas, G. and Catroux, G., 2001. DNA extraction from soils: old bias for new microbial diversity analysis methods. *Applied and environmental microbiology*, 67(5), pp.2354-2359.
- McTavish, H.J.A.F., Fuchs, J.A., and Hooper, A.B., 1993. Sequence of the gene coding for ammonia monooxygenase in Nitrosomonas europaea. *Journal of bacteriology*, 175(8), pp.2436-2444.
- Oksanen, J., 2018. Vegan: An Introduction to Ordination: Processed with Vegan 2.5-3 in R Under Development (Unstable) (2018-10-23 r75481) on October 24, 2018.
- Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane, J.A. and Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), pp.3691-3693.
- Peng, Y. and Zhu, G., 2006. Biological nitrogen removal with nitrification and denitrification via nitrite pathway. *Applied microbiology and biotechnology*, 73(1), pp.15-26.
- Pester, M., Schleper, C. and Wagner, M., 2011. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Current opinion in microbiology*, 14(3), pp.300-306.
- 20. Pinto, A.J., Marcus, D.N., Ijaz, U.Z., Bautista-de Lose Santos, Q.M., Dick, G.J. and Raskin, L., 2016. Metagenomic evidence for the presence of comammox Nitrospiralike bacteria in a drinking water system. *Msphere*, 1(1), pp. e00054-15.
- Prosser, J.I. and Nicol, G.W., 2012. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. *Trends in microbiology*, 20(11), pp.523-531.
- 22. RStudio | Open source & professional software for data science teams. [online] Available at: https://www.rstudio.com>. (2021)
- 23. Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), pp.2068-2069.

- 24. Siripong, S. and Rittmann, B.E., 2007. Diversity study of nitrifying bacteria in full-scale municipal wastewater treatment plants. *Water research*, *41*(5), pp.1110-1120.
- 25. Soliman, M. and Eldyasti, A., 2018. Ammonia-Oxidizing Bacteria (AOB): opportunities and applications—a review. *Reviews in Environmental Science and Bio/Technology*, 17(2), pp.285-321.
- 26. Spieck, E. and Lipski, A., 2011. Cultivation, growth physiology, and chemotaxonomy of nitrite-oxidizing bacteria. *Methods in enzymology*, *486*, pp.109-130.
- 27. Stein, L.Y., 2015. Cyanate fuels the nitrogen cycle. Nature, 524(7563), pp.43-44.
- 28. Tourna, M., Stieglmeier, M., Spang, A., Könneke, M., Schintlmeister, A., Urich, T., Engel, M., Schloter, M., Wagner, M., Richter, A. and Schleper, C., 2011. Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. *Proceedings of the National Academy of Sciences*, 108(20), pp.8420-8425.
- 29. Whelan, F.J., Rusilowicz, M. and McInerney, J.O., 2020. Coinfinder: detecting significant associations and dissociations in pangenomes. *Microbial genomics*, *6*(3).
- Wickham, H. and Grolemund, G., 2010. Graphics for Statistics and Data Analysis with R. *Journal of Statistical Software*, *36*(1), pp.1-2.
- 31. Zhou, Z., Tran, P.Q., Breister, A.M., Liu, Y., Kieft, K., Cowley, E.S., Karaoz, U. and Anantharaman, K., 2020. METABOLIC: High-throughput profiling of microbial genomes for functional traits, biogeochemistry, and community-scale metabolic networks.
- 32. Wang, B., Wang, Z., Wang, S., Qiao, X., Gong, X., Gong, Q., Liu, X. and Peng, Y., 2020. Recovering partial nitritation in a PN/A system during mainstream wastewater treatment by reviving AOB activity after thoroughly inhibiting AOB and NOB with free nitrous acid. *Environment international*, 139, p.105684.
- 33. Bergey, D.H., Hendricks, D., Holt, J.G. and Sneath, P.H., 1984. *Bergey's Manual of systematic bacteriology. Vol. 2.* Williams & Wilkins.
- 34. Whang, L.M., Chien, I.C., Yuan, S.L., and Wu, Y.J., 2009. Nitrifying community structures and nitrification performance of full-scale municipal and swine wastewater treatment plants. *Chemosphere*, *75*(2), pp.234-242.

APPENDIX

ACCESSING ORION CLUSTER

Step 1: Install CISCO AnyConnect Secure Mobility Client and turn on the VPN on University of Glasgow, next, use the Terminal software that is internally installed within MacOS to connect:

ssh studentprojects@howe.eng.gla.ac.uk

Step 2: Open own local directory for the project in the cluster to store the sequences and analysed data

cd /shared5/studentprojects/Kaung

DATA COLLECTION

Step 1: On Orion cluster, enable NCBI software

ftp ftp.ncbi.nlm.nih.gov

Step 2: Access to the location of genomes data on NCBI

cd genomes/genbank/bacteria cd genomes/genbank/archaea

Step 3: Download the genomes by species name to own local directory on cluster

wgetftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Nitrosomonas_eutropha/ assembly_summary.txt genomeNum=\$(grep -c "." assembly_summary.txt) if [\$genomeNum gt 700]; then grep "Complete\|Chromosome" assembly_summary.txt | cut -f20 > var.txt; else cut -f20 assembly_summary.txt | sed '1,2d' > var.txt; fi for f in `cat var.txt`; do name=\$(grep -w "\$f" assembly_summary.txt | cut -f9 | cut -f2 -d'=' | sed 's/ /_/g' | sed 's/\//_/g' | sed 's/\:/_g' | sed 's/)/_/g' | sed 's/(/_g'); xx=\$(grep -w "\$f" assembly_summary.txt | cut -f10 -d'/'); wget --tries=75 -c \$f/\$xx_genomic.fna.gz; done

gzip -d *.gz

PROKKA WORKFLOW

Step 1: Enable minconda2 environment where all the tools are available

export PATH=/home/opt/miniconda2/bin:\$PATH

Step 2: Activate pangenome

source activate pangenome

Step 3: Run Prokka for the downloaded genomes

for i in \$(ls *.fna); do echo "Processing \$i"; prokka \$i --locustag \${i%_genomic.fna}
--outdir \${i%_genomic.fna}; done

ROARY WORKFLOW

Step 1: Open a new directory called Roary and copy the gff files to that directory

mkdir roary
for i in \$(ls */*.gff); do cp \$i roary/\$(echo \$i | sed 's!/.*!!').gff; done

Step 2: Enable minconda2 environment

export PATH=/home/opt/miniconda2/bin:\$PATH

Step 3. Enable Roary

source activate pangenome
export PERL5LIB=/usr/local/lib/perl5/site_perl/5.22.0/

Step 4. Apply Roary and put the Roary output to the Roary tree directory

roary -f ./roary_tree -e -n -v ./roary/*.gff

Step 5. Generate Roary plots

python /home/opt/roary_scripts/roary_plots.py
roary tree/accessory binary genes.fa.newick roary tree/gene presence_absence.csv

METABOLIC WORKFLOW

Step 1. Enable minconda2 environment

export PATH=/home/opt/miniconda2/bin:\$PATH

Step 2: Enable METABOLIC software on Orion Cluster and go to your directory:

source activate metabolic

Step 3: Set the path to METABOLIC repository to use the software export PATH=/home/opt/METABOLIC:\$PATH

Step 4. Open a new directory for METABOLIC

mkdir METABOLIC

Step 5: Copy the test genomes in own local directory

cp *.fna /shared5/studentprojects/Kaung/Nitrosomonas_GENUS/METABOLIC

Step 6: Change the extension to fasta format

for i in \$(ls *.fna); do mv \$i \${i%.fna}.fasta; done

Step 7: Run METABOLIC and put the output files to an output directory

METABOLIC-G.pl -in-gn METABOLIC -0 METABOLIC_OUTPUT

COINFINDER WORKFLOW

Step 1. Enable minconda2 environment

export PATH=/home/opt/miniconda2/bin:\$PATH

Step 2: Enable coinfinder on cluster

source activate coinfinder-env

Step 3: Open a new directory for coinfinder

mkdir coinfinder-test cd coinfinder-test

Step 4: Download the associated data to the directory

git clone https://github.com/fwhelan/coinfinder-manuscript.git

Step 5: Copy the required genetic files to the directory

cp coinfinder-manuscript/gene_presence_absence.csv
cp coinfinder-manuscript/core-gps_fasttree.newick

Step 6: Run Coinfinder

coinfinder -i gene_presence_absence.csv -I -p core-gps_fasttree.newick -o
output

RSTUDIO WORKFLOW

abund_table<-read.csv("presence_absence_table.csv",header=TRUE,row.name=1)
library(vegan)
library(ggplot2)
library(ggrepel)</pre>

abund table.dist<-vegdist(abund table, method="jaccard")</pre>

ord<-capscale(abund_table ~ 1,distance="jaccard df<-as.data.frame(scores(ord, display = "sites"))</pre>

df\$Colours=1

df["Genome3", "Colours"]<-2
df["Genome4", "Colours"]<-2
df["Genome5", "Colours"]<-3</pre>

pdf("myplot.pdf",height=10,width=10)
p <- ggplot(df, aes(MDS1, MDS2))
p<- p+geom_point(color = 'red')
p<-p + geom_label_repel(aes(label = rownames(df),fill=factor(Colours)),size
= 3.5) + theme_bw()
p<-p+guides(fill=FALSE)
print(p)
dev.off()</pre>