Control of aquifer geology on the spatio-temporal distribution of groundwater prokaryotes in England

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Lancaster Environment Centre

Lancaster University

Archita Bhattacharyya

June 2025

Abstract

Groundwater is a source of drinking water for over two billion people globally. Groundwater also harbours vast and under-explored prokaryotic communities essential to subsurface biogeochemical cycling and maintaining drinking water quality. Despite their importance, large-scale national studies to understand the spatio-temporal variation of groundwater prokaryotes are scarce. This is the first systematic study of prokaryotic ecosystems in the groundwater of three major aquifers of England, which is the source of one-third of the public water supply. This research employed modern microbial monitoring technologies, such as flow cytometry and eDNA metabarcoding. Groundwater samples were collected from public supply sources of three major aquifers of England: Permo-Triassic sandstone, Cretaceous chalk, and Jurassic limestone. The research aimed to assess the collective influence of aquifer geology, groundwater recharge and chemistry in shaping the groundwater prokaryotic ecosystems on a national spatial scale.

The aquifer geology and surface connectivity were observed to be significant drivers of total bacterioplankton cell concentration (TCC). The karstic limestone aquifers showed almost two times higher TCC than intergranular sandstone and dual porosity chalk aquifers due to more frequent allochthonous prokaryotic input, evident from the higher abundance of animal parasitic DNA in the limestone aquifer. The seasonal recharge seemed to have impacted only the chalk aquifer, where higher groundwater levels were related to a reduction in TCC. The TCC reduction in the chalk aquifer could be a result of dilution by the recharge water from the unsaturated zone of the aquifer, containing a lower bacterial concentration. The sandstone aquifer was enriched in prokaryotic classes such as *Gammaproteobacteria, Bacteroidia* and *OM190*, and a higher proportion of chemoheterotrophic and autotrophic functions. In contrast, chalk and limestone aquifers exhibited compositional and functional similarity, characterised by a greater presence of *Omnitrophia, Nanoarchaeia, Dehalococcoidia, Gracilibacteria, a*nd *Saccharimonadia*, and many unknown functions indicating the presence of cryptic functional potentials. The results suggested that the community composition of the two carbonate aquifers could be different from the ferro-silicate

sandstone aquifer due to differences in groundwater chemistry, controlled by mineral dissolution from the aquifer matrix. While on the national scale, the prokaryotic communities varied significantly between aquifer types, within individual aquifers, spatial variation seemed to be controlled strongly by total dissolved nitrogen concentrations and overlying strata thickness. A focused study in the sandstone communities revealed that while the communities did not shift seasonally, the recharge age of groundwater was related to community composition. The groundwater with a younger recharge age was higher in dissolved nitrogen and oxygen and originated from shallower zones of the aquifer, hosted heterotrophic and parasitic families, including *Omnitrophaceae* and *Nanoarchaeia.* In contrast, the old recharge age groundwater with lower dissolved nitrogen and oxygen concentration originated from deeper parts of the aquifer and hosted autotrophic families such as *Gallionellaceae, Rhodocyclaceae, Hydrogenophilaceae* and *Comamonadaceae*. Thus, within individual aquifers, groundwater nutrient chemistry controlled by recharge age had a substantial impact on the community composition.

This thesis is a significant contribution to the growing number of national studies on groundwater prokaryotic ecosystems. Unlike many regional studies, this research established that on a large spatial scale, the proximity of sampling sites did not possess prokaryotic ecosystem similarities. Instead, similar prokaryotes were selected collectively by similar geologies and similar levels of surface connectivity. Thus, different aquifer geologies can be used in the future for classifying prokaryotic ecosystem management zones. The datasets from relatively clean drinking water sources can be used as a reference microbial community structure for England’s major aquifers for future monitoring and groundwater management strategies.

List of contents

Table of Contents

List of figures and tables

Acronyms

16S rRNA	16S ribosomal RNA gene
Al^{3+}	Aluminium ion
ANOSIM	Analysis of Similarity
ANOVA	Analysis of Variance
AOC	Assimilable Organic Carbon
ASV	Amplicon Sequence Variants
ATP	Adenosine Triphosphate
BIX	Biological index
BL1	Blue Laser Detector Channel 1
BL3	Blue Laser Detector Channel 3
BTEX	Benzene, Toluene, Ethylbenzene, And Xylenes
Ca	Calcium
CCl_3F (CFC11)	Trichlorofluoromethane
CCl_2F_2 (CFC12)	Dichlorodifluoromethane
$C_2Cl_3F_3$ (CFC113)	1,1,2-trichloro-1,2,2-trifluoroethane
Cd^{2+}	Cadmium ion
CH_4	Methane
Cl	Chlorine
CO_2	Carbon dioxide
Co^{2+}	Cobalt ion
CFC	Chlorofluorocarbon
CPR	Candidate Phylum Radiation
dbRDA	distance-based Redundancy Analysis
DIC	Dissolved Inorganic Carbon
DNA	Deoxyribonucleic Acid
DNRA	Dissimilatory nitrate reduction to ammonium
DO	Dissolved Oxygen
DOC	Dissolved Organic Carbon
EC	Electrical Conductivity
eDNA	Environmental DNA
EEM	Excitation-Emission Matrix
FAPROTAX	Functional Annotation of Prokaryotic Taxa
FCM	Flow Cytometry

Fe^{2+}	Ferrous ion
FI	Fluorescence Index
fOM	Fluorescent Organic Matter
FSC	Forward Scatter
GHI	Groundwater Health Index
GWL	Groundwater Levels
H_2	Hydrogen molecule
H_2O	Water
H_2S	Hydrogen sulphide
HC	Hierarchical Cluster
HCO_3^-	Bicarbonate ion
HIX	Humification Index
HLF	Humic-like Fluorescence
HNA	High Nucleic Acid (bacterial cells)
ICC	Intact Cell Concentration
IQR	Interquartile Range
LCM	Land Cover Map
LEfSe	Linear Discriminant Analysis Effect Size
LNA	Low Nucleic Acid (bacterial cells)
LOCAR	Lowland Catchment Research Programme
Mn^{2+}	Manganous ion
N_2	Dinitrogen gas
N_2O	Nitrous oxide gas
Na	Sodium
NH_4^+	Ammonium ion
Ni^{2+}	Nickel ion
NO_3^-	Nitrate ion
NO_2^-	Nitrite ion
NO	Nitric oxide gas
O_2	Oxygen molecule
PARAFAC	Parallel Factor Analysis
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PES	Polyethersulfone
PI	Propidium Iodide

RNA	Ribonucleic Acid
RSD	Relative standard deviation
RSU	Raman Standard Unit
S^0	Sulphur
SO_3^{2-}	Sulphite ion
SO_4^{2-}	Sulphate ion
$S_2O_3^{2-}$	Thiosulphate ion
SGPI	SYBR Green I and Propidium Iodide mix
SPZ	Source Protection Zone
SSC	Side Scatter
TCC	Total Cell Concentration
TDN	Total Dissolved Nitrogen
TLF	Tryptophan-like Fluorescence
TyLF	Tyrosine-like Fluorescence
UO_2^{2+}	Uranyl ion
UK	United Kingdom
Zn^{2+}	Zinc ion

Glossary

Aquifer	A saturated, permeable geologic formation (or part of one) that can store and transmit enough water to wells and springs to be useful as a water supply.
Unconfined aquifer	An aquifer whose upper surface is the water table and thus in direct contact with the atmosphere; its water levels rise and fall freely with recharge and discharge.
Confined aquifer	A water-bearing stratum that is completely overlain by a comparatively impermeable confining layer; the water is under pressure greater than atmospheric and will rise in a well above the top of the aquifer.
Unsaturated zone	Subsurface material above the water table that contains both air and water; it controls the transmission of water from the land surface to the saturated zone.
Saturated zone	This is the part of an aquifer, below the water table, in which almost all pores and fractures are saturated with water.
Water table	The water table is the upper surface of the saturated zone.
Borehole	A deep, narrow hole drilled into the ground to access water stored in underground aquifers and commonly used to pump water for supply.
Spring	A natural discharge point where groundwater flows out of the ground to the surface, typically occurring when the water table intersects the land surface or through fractures in the rock.
Groundwater recharge	The natural (or artificial) process by which water is added to the saturated zone, most commonly by downward percolation of rainwater through soil and the unsaturated zone, but also by lateral entry of water from a surface water body, such as streams.
Hydraulic conductivity	The volume of water that will move in unit time through a unit cross-section of an aquifer under a unit hydraulic gradient; it depends on both the fluid and the porous medium.
Karst	A landscape developed on soluble rocks (typically limestone or dolomite) and characterised by sinkholes, caves, sinking streams and underground drainage produced by chemical dissolution.
Pore throat	In an intergranular rock, the small pore space at the point where two grains meet, which connects two larger pore volumes.
Groundwater flow path	The three-dimensional trajectory that groundwater follows from its point of recharge to a point of discharge

Groundwater residence time	The elapsed time required for a parcel of groundwater to move from its point of recharge at the water table to a specified point in the aquifer.
Groundwater recharge age	The time when the water particle entered the saturated zone of the aquifer after exiting the atmospheric contact.
Flow cytometry	A laser-based technique for rapid, multi-parameter analysis and counting of individual cells in suspension.
Environmental DNA	DNA released by organisms into the environment (water, soil, air) enables biodiversity assessment without direct sampling of the organisms.
Planktonic bacteria	Free-floating bacterial cells suspended in water rather than attached to surfaces or sediment.
Biofilm	A surface-attached community of microorganisms embedded in a self-produced extracellular polymeric matrix.
Ecosystem structure	Organisation of an ecosystem, which includes both biotic and abiotic components.
Community composition	Abundance and diversity of species within a biological community.
Electron acceptor	Ions or molecules which are reduced by taking up electrons and act as oxidising agents in chemical reactions
Electron donor	Ions or molecules which are oxidised by losing electrons and act as reducing agents in chemical reactions
Aerobic	Metabolic processes require molecular oxygen as the terminal electron acceptor.
Anaerobic	Metabolic processes where the electron acceptor can be anything but oxygen, and take place strictly in the absence of molecular oxygen.
Facultative anaerobe	Facultative anaerobic organisms can grow either with or without oxygen and can change their metabolic processes depending on the presence of oxygen.
Chemoheterotroph	An organism that derives both energy and carbon from the oxidation of pre-formed organic compounds.
Lithoautotroph	A microbe that oxidises inorganic compounds (e.g., H_2S, NH_4^+) for energy and fixes CO_2 as its carbon source.

Acknowledgement

A PhD is often seen as a lonely venture. However, it is incredible how much help, both academic and personal, of various degrees, goes into finishing a massive piece of research. First and foremost, I thank my supervisors, Dr. James Sorensen, Prof. Daren Gooddy, Dr. Ben Surridge and Dr. Daniel Read for their support and supervision over the last four years. Their constant guidance has led me to see bigger pictures of my work. I was able to work with some incredible collaborators at Thames Water and Severn Trent. Thank you, Mike, Simon, Liam, Karen and Amanda, for providing support with my sampling, helping in the field, and your valuable insight into my research results. My sample analysis would not be successful without the active help and lab training by colleagues and friends in the UKCEH labs. Also, thank you, Barry, Pete, Liz and Ankita, for helping with fieldwork. And of course, everyone in the Envision team, thank you not only for your constant support, but also for your fantastic effort to create a community of PhD students. Special thanks to Liz and Melissa for helping to proofread my thesis and for leaving those positive, encouraging comments throughout the draft, which gave me a moral boost. Moving from one's home country to pursue a career is hard, but the people I met and the life-long connections I made were worth the leap. Shout out to everyone in BGS Wallingford for making me feel included in your community as soon as I moved here. Cheers to every friend I made in Wallingford and Lancaster, with whom I made some precious memories, which I will cherish for life. Finally, some debts cannot be acknowledged with a simple thank you, and yet, we try. Thank you to my mum, Bhaswati, who has been the role model of my life, and my little brother, Archisman, for whom I try to be a better role model every day. Finally, behind every successful person is an incredibly supportive partner, and for me, it is Aman. I guess now it is my turn to help with your doctoral thesis, as you did for me, and afterwards, we can continue to be the first critics of each other's manuscripts, forever.

The research was supported by the NERC Envision Doctoral Training Programme (Grant reference number NE/S007423/1).

Declaration

I declare that this thesis is entirely my work and has not been presented for any other degree. Throughout the research, I regularly discussed the overall direction of the research with my supervisors Dr. James Sorensen, Prof. Daren Gooddy, Dr. Ben Surridge and Dr. Daniel Read.

Archita Bhattacharyya

30th June 2025.

1. Introduction

1.1 Background

Around 71% of the Earth's surface is covered by water, but only 2.5% of this is fresh water, with the rest being saline water in the oceans. Within the total freshwater resources, 69% of water is stored in the ice caps, surface water, including rivers and lakes, constitutes less than 1%, and the remaining 30% is stored in the subsurface as groundwater (U.S. Geological Survey, 2019). Groundwater is conceived as the water present in the pore spaces and fractures of subsurface geological units, including sediments and rocks. A highly porous and permeable geological unit, capable of storing and yielding groundwater, is termed an aquifer, such as sandstone or limestone, among other lithologies (Figure 1-1). This groundwater serves as a drinking water source for over 2 billion people worldwide (Hiscock, 2011). Groundwater systems also host a unique and complex microbial ecosystem, specifically adapted to thrive in a dark and nutrient-poor environment (Chapelle, 2000; Malard *et al.*, 2023). The groundwater microbiome is dominated by prokaryotes (higher abundance of bacteria and less abundant archaea), which can be either attached to the aquifer matrix or suspended in the groundwater (Griebler and Lueders, 2009). These groundwater prokaryotes contribute to biogeochemical cycling of elements and provide essential ecosystem services (Griebler and Avramov, 2015). Water supply industries regularly monitor planktonic (suspended) microbes to detect harmful pathogens and ensure the safety of public health (John and Rose, 2005; Willis *et al.*, 2013). This thesis also focuses on only the suspended subset of the groundwater prokaryotes, rather than the prokaryotic population attached to the aquifer matrix.

Microorganisms form the most diverse component of the biosphere and, despite their microscopic size, drive global cycles of matter and energy transformations (Goldscheider *et al.*, 2006). Their habitats span the atmosphere, soils, oceans, surface waters, the human body and, crucially for this thesis, groundwater systems, where research only gained momentum in the latter half of the twentieth century. An early assumption was that life could not persist below the topsoil because microbial abundance and organic nutrient concentrations decline sharply with depth (Ghiorse and Wilson, 1988; Wilson *et al*., 1983). However, research began to focus more strongly on the groundwater microbes in the second half of the twentieth century, mainly due to their association with outbreaks of waterborne diseases and evidence of microbial capacity for contaminant biodegradation (Bitton *et al*., 1983; Dunlap *et al*., 1972; Gibert, 1994; Yates *et al*., 1985). Among some of these early works, Dunlap *et al*. (1972) detected microbes in a floodplain aquifer, Whitelaw and Rees (1980) identified nitrogen-transforming bacteria in the English Chalk, and Dockins *et al*. (1980) found sulphate reducers within the groundwater in the USA. By the end of the twentieth century, it was evident that indigenous microbial communities in groundwater are abundant, diverse and ecologically significant (Danielopol *et al*., 2000; Ghiorse and Wilson, 1988; Gibert, 1994). The structure of these communities is affected by aquifer mineralogy, pore size, electron-acceptor supply, nutrient availability and physicochemical factors such as pH and redox potential (Balkwill *et al*., 1989; Pedersen, 1997; Sinclair and Ghiorse, 1989). Despite these early examples of research efforts, at the beginning of the twenty-first century, the microbial realm of the groundwater system was described as an “unseen ocean” by Danielopol *et al*. (2000), possibly due to the unusual goal of merging expertise in microbiology, which is primarily a laboratory-based science and hydrogeology, which primarily has been a field-based science (Chapelle, 2000). However, since the second decade of the 21st century, an increasing number of studies on groundwater microbes have been reported, strongly supported by the emergence of affordable and quick microbial detection technologies.

Groundwater planktonic prokaryotes mediate the biogeochemical transformation of carbon, nitrogen, phosphorus, sulphur, and other nutrients by driving redox cycles as they utilise various organic and inorganic chemicals as electron donors and acceptors for respiration (Falkowski *et al*., 2008; Goldscheider *et al*., 2006; Griebler and Lueders, 2009). These cycles are often interconnected, as different prokaryotes with unique functional capabilities participate in various aspects of redox cycling, such as denitrification and nitrification (Anantharaman *et al*., 2016) (Section 2.2). However, groundwater prokaryotic functions are complex, with several microbes capable of switching between functions at differing nutrient concentrations (Anantharaman *et al.*, 2016). Furthermore, prokaryotes can evolve and develop new metabolic potentials to degrade emerging contaminants (Kolvenbach *et al.*, 2014). By transforming these chemicals, groundwater prokaryotes can support a stable and high-quality groundwater resource. These functions represent their primary ecosystem services, i.e., goods and services provided by the environment that benefit human societies (Griebler and Avramov, 2015). Korbel and Hose (2011) defined a healthy groundwater ecosystem as one that maintains its ecosystem structure under stress and sustainably provides ecosystem goods and services. These services were politically recognised in the 2006 European Groundwater Directive (European Union, 2006), which stated that groundwaters are important both as resources for human consumption and as unique habitats that need to be sustainably managed. To manage the groundwater ecosystems effectively, two critical steps are to first classify different management units based on their ecosystem similarities and second, to define a reference groundwater community composition to assess changes under stress (Hose *et al.*, 2023). However, the spatio-temporal variation of groundwater prokaryotic communities is not yet well understood due to the unpredictability of physical, chemical and biological processes in the subsurface environments and the lack of prokaryotic community data from highly heterogeneous aquifer systems.

Following the development of modern technologies such as environmental DNA (eDNA) sequencing and flow cytometry, the understanding of groundwater prokaryotic ecosystems has gained momentum. An increasing volume of data from extensive spatial surveys is attempting to provide valuable information regarding the distribution of groundwater prokaryotes and their contribution to biogeochemical cycles (Abraham and Close, 2024; Harris *et al.*, 2025; Korbel *et al.*, 2024; Merino *et al.*, 2022; Sirisena *et al.*, 2018; Smith *et al.*, 2018; Zhong *et al.*, 2023). These studies have yielded valuable information about the types and functions of prokaryotes present in groundwater and were considerable developments compared to earlier regional studies using T-RF biomarkers (terminal restriction fragments of bacterial 16S rDNA), which provided limited information about prokaryotic taxonomy and function (Griebler *et al.*, 2010; Sirisena *et al.*, 2013; Stein *et al.*, 2010; Steube *et al.*, 2009). Along with regional-scale surveys, only two examples of national-scale surveys of groundwater prokaryotes using eDNA techniques can be found in New Zealand and China (Sirisena *et al.*, 2018; Zhong *et al.*, 2023). Similar national-scale studies are essential to establish the determining factors of the variation of groundwater prokaryotic communities, as well as to classify ecosystem management areas and define reference groundwater microbiology (Hose *et al.*, 2023).

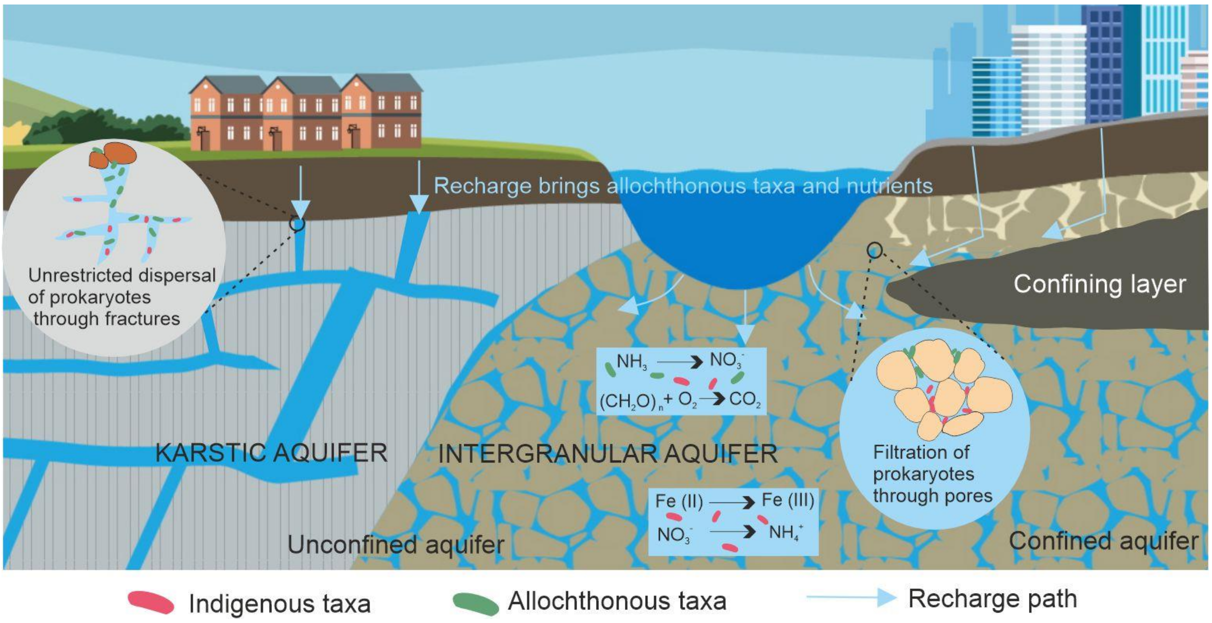


Figure 1-1. Conceptual diagram showing groundwater stored in karstic and intergranular aquifers, allochthonous taxa import with recharge water in different aquifer types, and differences in metabolic activities depending on surface nutrient influence at various depths.

Although aquifer geology is known to shape the physical and chemical properties of groundwater, its role in controlling the spatio-temporal variation of groundwater prokaryotic communities remains underexplored. Emerging evidence suggests that aquifer geology provides the physical space for prokaryotic habitat and transportation, and controls groundwater chemistry, which controls prokaryotic metabolism (Griebler and Lueders, 2009). The size difference between aquifer pore-throats and microbial cells can result in physical filtration of prokaryotic cells transported from the surface, as well as within-aquifer migration of these prokaryotes (Bloomfield *et al.*, 2001; Taylor *et al.*, 2004; van Driezum *et al.*, 2018; Voisin *et al.*, 2018) (Figure 1-1). For instance, the filtration capacity of the European karstic aquifers was found to control bacterioplankton cell concentration by determining how frequently surface bacteria can migrate into the groundwater (Farnleitner *et al*., 2005; Savio *et al*., 2018; Sinreich and Pochon, 2023; Sinreich *et al*., 2014). Aquifer mineralogy further controls groundwater chemistry through the dissolution-precipitation dynamics of minerals (Elango and Kannan, 2007). The groundwater ionic chemistry, which was potentially regulated via rock-water interaction, was found to be one of the significant factors contributing to the spatial differences of planktonic prokaryote community structures (Abraham and Close, 2024; Amalfitano *et al*., 2014; Couton *et al*., 2023; Zhong *et al*., 2023). However, to date, very few regional studies have assessed the role of aquifer geology as a criterion for the spatial variation of prokaryotes (Abraham and Close, 2024; Amalfitano *et al*., 2014; Couton *et al*., 2023).

Studies on the temporal dynamics of the planktonic prokaryotes by the groundwater recharge process further suggest a critical role of aquifer geology. Nutrients and allochthonous prokaryotes from the surface enter the groundwater during the recharge process, resulting in perturbation of the local prokaryotic ecosystems by selection of dominant prokaryotes capable of using the newly arrived nutrients, and/ or competition and adaptation of allochthonous prokaryotes in the groundwater environment (Cooper *et al*., 2016; Fiedler *et al*., 2018; Fillinger *et al*., 2021; Stegen *et al*., 2018; van Driezum *et al*., 2018; Villeneuve *et al*., 2022; Voisin *et al*., 2018; Yan *et al*., 2021; Zhou *et al*., 2002; Zhou *et al*., 2012). The ease of groundwater recharge depends on the size and interconnectedness of the pore spaces and fractures in the aquifer matrix, i.e., groundwater movement is faster and farther through karstic fractures than intergranular pore spaces. Depending on how easily recharge water can enter the aquifer, the pulse of nutrients and allochthonous prokaryotes can be transported into the various depths of different aquifers (Figure 1-1). For instance, the recharge-related perturbations were witnessed in the karstic feature-dominated aquifer further from the recharge area, whereas in the intergranular space-dominated aquifer, the impact of recharge was rapidly attenuated within a shorter distance from the recharge area (Villeneuve *et al*., 2022). The frequency of long-term disturbances by recharge was found to be an essential factor impacting the spatial variation of groundwater communities (Ben Maamar *et al*., 2015; Farnleitner *et al*., 2005; Smith *et al*., 2012;

Yan *et al*., 2021). However, the temporal dynamics of the prokaryotes in different geological settings are rarely compared.

Despite a body of past research suggesting that aquifer geologies may exert significant control over the groundwater prokaryotic ecosystems in both space and time, only a small number of studies have compared prokaryotes in geologically contrasting aquifers on a large regional scale to characterise the distribution of prokaryotic ecosystems (Abraham and Close, 2024; Amalfitano *et al*., 2014; Couton *et al*., 2023; Griebler *et al*., 2010; Harris *et al*., 2025; Stein *et al*., 2010). Moreover, most of these spatial studies have not been temporally repeated, preventing the conceptualisation of whether community dynamics upon groundwater recharge are different in different aquifer types. Of these, the studies by Stein *et al*. (2010) and Griebler *et al*. (2010) compared bacterial biomarkers in different aquifer types in two seasons using bacterial eDNA biomarkers (terminal restriction fragments or T-RFs), which only help in grouping the prokaryotes based on the biomarker similarities, but do not support the identification of the prokaryotic types. The studies of Abraham and Close (2024), Couton *et al*. (2023) and Amalfitano *et al*. (2014) were performed using eDNA sequencing, allowing the detection of different prokaryotes in contrasting geologies, but these studies were performed on a small regional scale without a national-scale coverage and without seasonal repetitions. Amalfitano *et al*. (2014) compared two hydraulically connected aquifers, and although they observed different prokaryotic communities in alluvial and volcanic aquifers, the hydraulic connection can allow prokaryotic dispersal and homogenisation, as was observed by Stein *et al*. (2010) and Griebler *et al*. (2010). This implies that the exploration of prokaryotic ecosystems exclusive to an aquifer type should be performed between hydraulically disconnected aquifers. In the two national surveys found to date, aquifer geologies were not considered as a potential controlling factor of the spatial variation of prokaryotes, and no temporal assessment was performed. Instead of comparing different aquifers with different geologies, the study areas were chosen based on climatic regions and “geo-environments” by Zhong *et al*. (2023) in China. Although lithologies were compared in the national survey of New Zealand, aquifer types were not found to be a significant contributor of prokaryotic community differences, probably due to a lack of enough representative samples from lithologically different aquifers (Sirisena *et al*., 2018). Thus, a significant research gap exists about whether aquifer geology plays a substantial role in the large-scale spatial and temporal variation of groundwater prokaryotes and whether aquifers can be used as a potential classifier of groundwater ecosystem management zones.

1.2 Thesis aims, objectives and scopes

This thesis consists of national-scale research, focused on characterising the spatial variation of planktonic prokaryotic ecosystems and their temporal changes in response to recharge in geologically contrasting aquifer types. This is the first systematic study of groundwater prokaryotes in England, United Kingdom, where one-third of the public water supply relies on groundwater from major aquifers (British Geological Survey, 2019). Due to the high dependency on groundwater for public supply, the groundwater resources are under constant stress. However, only the groundwater ecosystems of macrofauna have been systematically studied recently (Weitowitz *et al*., 2017). In a review of groundwater microbiology of the UK, Gregory *et al*. (2014) highlighted the lack of systematic exploration of groundwater microbial ecosystems, and to date, such a study is lacking. This study was performed on three major English aquifers with unique geological characteristics (Figure 1-2), which are also globally extensive groundwater sources. These aquifers are: the Permo-Triassic sandstone aquifer, representative of red sandstone aquifers commonly found across Europe, North Africa, the Middle East, and North America (Celle-Jeanton *et al*., 2009); the Cretaceous chalk aquifer, known as northwestern Europe's principal source of potable groundwater (Gunn *et al*., 1995); and the Jurassic limestone aquifer, regarded as one of the most significant carbonate aquifers globally (Worthington and Ford, 2009).

The main aim of this thesis is to determine the primary controlling factors on the spatial and temporal variation of planktonic prokaryotes in the major English aquifers through a national-scale study (Figure 1-2). The thesis objectives were:

1. To optimise prokaryotic sample collection and analysis methods for groundwater systems with low prokaryotic concentration.
2. To assess the controls on the spatial and seasonal variation of bacterioplankton concentrations in the three different aquifers.
3. To assess the impact of aquifer geology on the spatial variation of groundwater planktonic prokaryotic community composition.
4. To assess the impact of recharge age on the spatial variation of prokaryotic community composition within a single major aquifer type.
5. To identify knowledge gaps, study limitations and future directions to characterise groundwater prokaryotic ecosystems in other parts of the world.

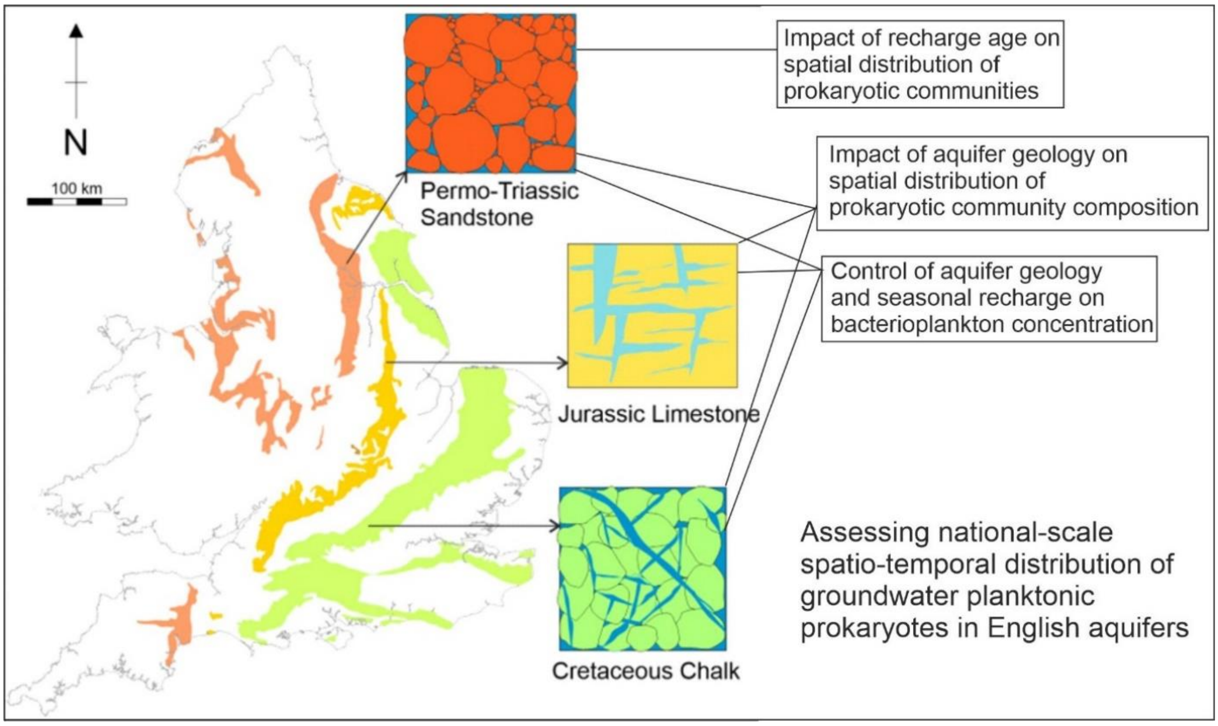


Figure 1-2. The map of England shows outcrops of the three major aquifers where groundwater flows through intergranular spaces in sandstone, through fractures in limestone and through both intergranular spaces and fractures in chalk. The main research objectives are mentioned along with lines indicating which aquifers were studied to address each of the objectives. (Aquifer outlines from BGS © UKRI (2023), map outline contains OS data © Crown copyright and database rights 2025)

Sampling was performed in collaboration with the regional water supply companies, which permitted access to untreated groundwater at drinking water sources. These aquifers were hydraulically not connected (Figure 1-1), allowing exploration of intrinsic groundwater prokaryotes, without the potential of mixing by dispersal between the aquifers. The samples from near-continuously pumping boreholes also represented true aquifer communities instead of borehole communities, which may be growing separately from the communities within the aquifers (Korbel *et al*., 2017; Sorensen *et al*., 2013). Contemporary analytical techniques, including flow cytometry (Van Nevel *et al*., 2017) and eDNA metabarcoding (Saccò *et al*., 2022), were used to characterise the communities. The ecosystem indicators measured in this study are bacterioplankton concentration, prokaryotic biodiversity, taxonomic composition and functional potentials, all parameters commonly recommended for characterising groundwater ecosystem health (Hose *et al*., 2023; Korbel and Hose, 2011). Due to the high dependency on groundwater as a public supply source, understanding its microbiology can help to design sustainable groundwater management strategies that consider the need for maintaining ecosystem services.

1.3 Thesis structure

The thesis is structured as follows:

Objective no.	Chapter no.	Chapter title	Chapter content
	2	Current understanding of the controls of groundwater ecosystem variation	Literature review of existing research on groundwater microbiology
1	3	Methods for microbial community analysis	Experimental results of optimising flow cytometric sample preparation and eDNA sample collection protocols
2	4	Aquifer geology controls the bacterioplankton concentration and dynamics in groundwater-derived public water sources.	Investigates the impacts of aquifer geology and seasonal recharge on the flow cytometric bacterioplankton concentration and its association with nutrients in groundwater
3	5	Aquifer geology plays a significant role in the spatial variation of the groundwater prokaryotic communities	Investigates how aquifer geology, nutrients and other environmental variables impact prokaryotic community composition assessed using eDNA metabarcoding
4	6	Impact of groundwater recharge age on spatial variation of planktonic prokaryotic communities in Permo-Triassic sandstone aquifer	Investigates the impact of seasonal recharge and decadal recharge age of groundwater on nutrient concentration and spatial variation of prokaryotic community composition assessed using eDNA metabarcoding.
5	7	Concluding discussion	Summarises key findings, highlights the contributions of the thesis to groundwater microbiology, discusses limitations, and outlines directions for future research.

Table 1-1. Table mapping thesis objectives to chapter numbers, the title of the chapters and the key content addressed in each chapter.

2. Current understanding of the controls of groundwater prokaryotic distribution

2.1 Basic concepts of groundwater microbiology and prokaryotic microbial ecosystems

Groundwater microbes are microscopic organisms with a size range between 0.2 and 100 micrometres (μm). These organisms are adapted to thrive in the nutrient-poor and dark environment of aquifers (Goldscheider *et al.*, 2006; Griebler and Lueders, 2009). Based on the cell structure of the microbes, they can be either prokaryotes or eukaryotes. The prokaryotic organisms are unicellular and have a primitive cell structure, while eukaryotes can be both unicellular and multicellular and have a more complex cell structure. Some characteristic features of prokaryotic cell structures are the absence of mitochondria, the absence of membrane-bound cell organelles, the presence of unorganised and twisted DNA and RNA molecules in a membrane-less nucleoid structure and their incapability of carrying out endocytosis (the process of engulfing and assimilating extracellular substances) (Vellai and Vida, 1999). Unlike prokaryotes, more complex eukaryotes have membrane-bound genetic material called a nucleus and membrane-bound cell organelles controlling different aspects of cellular functions.

All microorganisms can be classified into the three domains: Bacteria, Archaea and Eukaryotes (Figure 2-1). The domain system of classification, proposed by Woese *et al*. (1990), revolutionised the taxonomic classification system based on external morphological traits (phenotypes) and accounted for more fundamental differences of organisms, that is, differences in molecular structure of genes (genotype). Genes or DNAs are made up of four nucleotides: adenine (A), uracil (U) (thymine or T for eukaryotes), guanine (G) and cytosine (C). The genotypic classification is based on the theory that the sequences of the four nucleotides change during evolution. By comparing the genetic sequences, one can determine how far down the ancestral lineage two organisms shared a common ancestor. The smaller-subset ribosomal-

RNA (SSU rRNA) which is the 16S rRNA for bacteria and archaea, was functionally and structurally conserved, and thus were inferred to have common ancestry (Woese *et al.*, 1990). Therefore, based on the SSU rRNA data, it was proposed that two species can be grouped into a taxonomic class or taxa based on their proximity in the evolutionary tree (Woese, 1987). The taxonomic classification system involves arranging two organisms into a group based on their degree of genetic similarity. Within the domains, organisms are grouped into smaller subsets according to their similarities and their distance from a common ancestor. The highest rank within each domain is the kingdom. Then, as the distance of an animal from the common ancestor increases, the following ranks are respectively phylum, class, order, family, genus, and species (Woese, 1987; Woese *et al.*, 1990). The classification of organisms in different ranks is a branch of science known as taxonomy, and the organisms classified into one of the ranks are a taxon, or taxa in a group. According to the new tree of life there are 92 bacterial and 26 archaeal phyla (Hug *et al.*, 2016) (Figure 2-1). In the NCBI database, currently the bacterial and archaeal sequences are classified into 167 and 39 phyla respectively (NCBI Insights, 2021).

This thesis will only focus on the prokaryotic microbes in groundwater since they make up the majority of the groundwater microbial population (Griebler and Lueders, 2009). The two prokaryotic domains are bacteria and archaea. Bacteria are the most abundant group of prokaryotic organisms in groundwater. Their concentration in groundwater can range between 10^2 and 10^7 cells/mL (Griebler and Lueders, 2009). Bacteria have unique bacterial-type ribosomal RNA (rRNA) genes, used for their taxonomic classification, the genetic material is twisted into a membrane less nucleoid, and bacterial cell membrane is composed of 40% diacyl glycerol diether lipid and 60% protein (Woese, 1987). Microorganisms belonging to the domain archaea have a different molecular structure from bacteria, but also different from eukaryotes, leading to addition of this new domain. In groundwater, archaeal abundance is about 20% of total prokaryotic abundance (Griebler and Lueders, 2009). Archaea have archaeal-type rRNA genes distinct from bacterial rRNA genes, genetic material is wrapped around in histone protein resembling a eukaryotic nucleus, and archaeal cell membrane forming lipid is predominantly diacyl glycerol diester (Woese, 1987; Woese *et al.*, 1990).

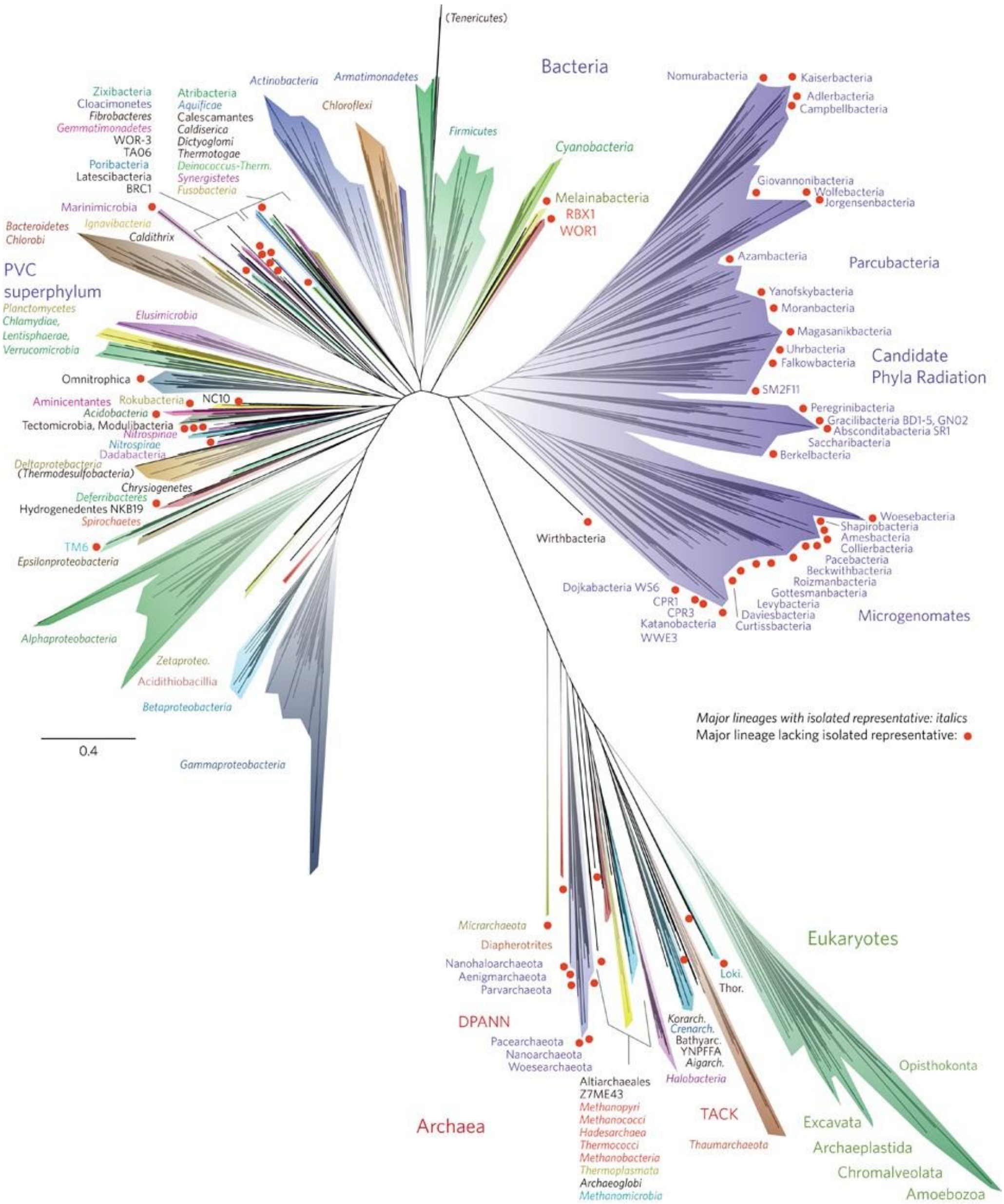


Figure 2-1. Updated tree of life figure showing three domains, (Source: Hug *et al*. (2016))

Genomic sequencing techniques has shown that cellular biochemical machineries like translation (mechanism of reading genetic codes to build protein) and DNA replication (copying genetic material) are different in archaea and bacteria domains (Cavicchioli, 2011). Studies comparing archaeal and bacterial replicative systems show that bacteria employ different enzymes and sigma-factor driven transcription whereas, archaea use DNA polymerases, transcription factors (controls gene expression of on or off) and translational factors (proteins that assist in translation) are homologous (genes sharing common origin) to those in eukaryotes (Bobbo *et al.*, 2024). Although eukaryotes and archaea share the homologous genes reflecting a common ancestry, morphologically, archaea also lack many of the hallmark traits of eukaryotic cells (e.g. a nucleus, membrane-bound organelles, and complex cytoskeletal systems), leading them to be grouped separately from the Eukaryotes (Cavicchioli, 2011). Among the four major subgroups of Archaea, namely, Euryarchaeota, TACK, Asgard and DPANN, TACK subgroup of archaea is the closest to Eukaryotes (Hug *et al.*, 2016) (Figure 2-1), although recently new protein signatures in Asgard archaeal genome suggested that Asgard are potentially sister group to eukaryotes (Eme *et al.*, 2023).

2.2 Prokaryotic functions and ecosystem health of groundwater

The groundwater ecosystem, like the rest of the Earth's biosphere, is a complex pool of oxidised and reduced material in disequilibrium, which maintains the flow of matter and energy to sustain life (Falkowski *et al.*, 2008). The addition and removal of chemicals from habitats by atmospheric and tectonic processes maintains this disequilibrium (Falkowski *et al.*, 2008). The constant flow of matter and energy from the Earth to the biosphere and back to Earth is controlled by several chemical reactions occurring in the cells of organisms, and the process of storage and release of these chemicals is referred to as biogeochemical cycling (Chapelle, 2000). The microorganisms in the subsurface of the Earth play a substantial role in controlling the biogeochemical cycling of major elements, which are essential building stones of biomolecules, like carbon, oxygen, hydrogen, nitrogen, phosphorus, sulphur, and iron (Falkowski *et al.*, 2008). In the subsurface groundwater environment, in which sunlight is absent, biogeochemical processes are the key source of energy for the lives of subsurface microorganisms. In simple terms, microbes exploit the oxidised and reduced substances by transferring electrons from an electron donor to an electron acceptor, and in the process, generating ATP, which is the unit of energy in organisms (Chapelle, 2000; Falkowski *et al.*, 2008). These microbially mediated redox transformation process in groundwater prevents the accumulation of any chemical species in impermissible concentrations for human consumption. Thus, biogeochemical processes are vital microbial ecosystem services which help to maintain permissible chemical quality of groundwater.

Metabolically, groundwater prokaryotes are chemotrophs, which obtain energy by transferring electrons between redox species of chemicals (dissolved or precipitated) available in their surroundings. Based on the electron-donor and acceptor (nutrients) source, groundwater bacteria can be either autotrophs or heterotrophs. The autotrophs found in dark groundwater environments are generally chemoautotrophs, which derive energy by facilitating redox processes of chemicals. The bacteria that use inorganic chemicals as nutrients are called lithotrophs. The chemoheterotrophs use organic carbon as the energy source (Chapelle, 2000). If the bacteria require oxygen as the terminal electron acceptor for respiration, then they are called obligate aerobes (e.g., *Mycobacterium, Methylomonas*, etc.). If the bacteria can only survive in the absence of oxygen, by using other electron acceptors (e.g., ferric iron, halogenated organic compounds), they are called obligate anaerobes (e.g., *Anaerolina, Dehalococcoides,* etc.). Some bacteria use oxygen as an electron acceptor, but in anaerobic conditions, they can also use other electron acceptors, and they are called facultative anaerobes (e.g., *E coli, Salmonella, Enterobacter*, etc.). Earlier understanding about archaeal metabolism was that among the two major archaeal clades, *Crenarchaeota* comprised of mainly thermoacidophiles and *Euryarchaeota* comprised of mainly halophiles and methanogens (Spang *et al*., 2017). In groundwater systems, methanogenic archaea are found in the absence of oxygen and the presence of carbon dioxide and hydrogen (Chapelle, 2000). However, with sophisticated gene sequencing techniques, many archaeal lineages found in non-extreme environments such as groundwater were discovered (Spang *et al*., 2017). Archaeal genes flexible to various oxic conditions with carbon and nitrogen redox transformation capacities have been detected in groundwater (Gios *et al*., 2023; Kumar *et al*., 2017; Lazar *et al*., 2017). The examples of prokaryotic functions utilising various reactants and producing multiple metabolic products are provided in Table 2-1. Since these prokaryotes are adapted to low-nutrient conditions, they are sensitive to disturbances in the chemical concentration of their ambient environment (Griebler and Lueders, 2009).

Element	***Process***	***Reactant(s)***	***Product (s)***	***Bacteria example***
C-H-O	Carbon fixation	CO_2, H_2O	$(CH_2O)_n$	*Sulfurimonas, Planctomycetes, Nitrospira, Thaumarchaeota*
	Aerobic Respiration	$(CH_2O)_n + O_2$	$CO_2 + H_2O +$ ATP	*Deltaproteobacteria, Chloroflexi, Burkholderiales, Rokubacteria*
	Anaerobic respiration	$(CH_2O)_n$	CO_2	*Rokubacteria, Deltaproteobacteria*
	Fermentation	$(CH_2O)_n$	Acetate /ethanol + H_2	*Candidatus, Burkholderia, Planctomycetes, Chloroflexi*
	Methanogenesis	CO_2	CH_4	*Burkholderiales, methanogenic archaea*
	Methane oxidation	CH_4	CO_2	
N	Nitrogen fixation	N_2	NH_4^+	*Azotobacter, Rhizobium*
	Nitrification	NH_3	NO_3^-	*Nitrosomonas, Nitrobacter*
	Denitrification	NO_3^-	NO_{2-}	*Deltaproteobacteria, Sulfurimonas, Ignavibacteria*
		NO_{2-}	NO	*Planctomycetes, Sulfurimonas, Ignavibacteria*
		NO	N_2O	*Bulkholderiale, Sulfurimonas, Ignavibacteria*
		N_2O	N_2	*Rhodoferax, Sulfurimonas, Ignavibacteria*
	Dissimilatory nitrate reduction to ammonia (DNRA)	NO_3^-	NH_4^+	*Burkholderia, Gammaproteobacteria*
	Anaerobic ammonia oxidation (anammox)	NH_4^+	N_2	*Planctomycetes, Rhodoferax*
S	Sulphur reduction	S^0	H_2S	*Candidatus, Deltaproteobacteria*
	Sulphur oxidation	S^0	SO_3^{2-}	*Candidatus, Burkholderia*
	Sulphite oxidation	SO_3^{2-}	SO_4^{2-}	*Deltaproteobacteria, Sulfurimonas*
	Sulphide oxidation	H_2S	S^0	*Sulfurimonas, Candidatus*
	Sulphate reduction	SO_4^{2-}	SO_3^{2-}	*Deltaproteobacteria, Desulfobacula*
	Thiosulphate oxidation	$S_2O_3^{2-}$	S	*Planctomycetes, Gammaproteobacteria*
			H_2S, SO_3^{2-}	*Deltaproteobacteria, Candidatus*
			SO_4^{2-}	*Candidatus, Burkholderia, Sulfurimonas*
Fe	Iron oxidation	Fe (II)	Fe (III)	*Thiobacillus, Gallionella*
	Iron reduction	Fe (III)	Fe (II)	*Shewanella, GS-15*

Table 2-1. Table of elemental transformation reactions and typical bacterial taxa capable of using the redox transformation of the chemical species of main nutrients, that are, carbon, hydrogen, oxygen, nitrogen, sulphur and iron.

There are four characteristics of prokaryotic biogeochemical functions in groundwater. First, the transformation of elements from a completely reduced to a completely oxidised state can be either a single reaction or an integration of a series of stepwise reactions (Anantharaman *et al*., 2016). For example, nitrate reduction or denitrification reactions follow stepwise transformation of nitrate (NO_3^-) to nitrite (NO_2), nitric oxide (NO), nitrous oxide (N_2O) and finally molecular nitrogen (N_2). More examples of the stepwise redox transformations can be found in Table 2-1. Second, often one species cannot carry out all the steps of a stepwise reaction. The number of Bacteria which can carry out two respective steps of a multistep reaction is far less than the number of microbes which can process only one step of a reaction (Anantharaman *et al*., 2016). An example of known bacterial taxa performing different steps of denitrification is in Table 2-1. Third, in a microbial population, there will be many species that have the genes to carry out the same biogeochemical reactions. Thus, if one species is eliminated under any environmental perturbation, the community will still retain its functional stability. This phenomenon is described as the "portfolio effect" (Konopka *et al*., 2015). Table 2-1 shows an example of multiple organisms carrying out one function. Fourth, the point in time and space where a biogeochemical reaction takes place is determined by the chemical condition of the ambient environment and the active bacteria. For example, if plenty of oxygen and ammonia are present, then the aerobic ammonia-oxidising bacteria will be active, and they will convert ammonium to nitrate. In short, the prokaryotic species that will survive in a specific groundwater environment depends on the compatibility of the available resources and the ability of the prokaryotic species to utilise the resources (Anantharaman *et al*., 2016).

The function of a prokaryotic ecosystem is related to its ecosystem services. Healthy groundwater prokaryotic ecosystems provide important ecosystem services by maintaining their functional integrity despite transient disturbances to their environment and by transforming nutrients and preventing their accumulation in groundwater at an impermissible limit for human consumption (Griebler and Avramov, 2015). Saccò *et al.* (2024) described groundwater as a keystone ecosystem to promote its effective conservation strategies that can be implemented to protect groundwater biodiversity and maintain essential ecosystem services. Korbel and Hose (2011) and Hose *et al.* (2023) outlines definitions of the terms related to ecosystem health and indicators, which are described in Table 2-2.

Term	Definition
Ecosystem Health	The expression of an aquifer's ability to sustain its ecological functioning in accordance with its organisation while providing the ecosystem goods and services.
Ecosystem services	Goods and services provided by the environment that are of value or benefit to humans, e.g., maintaining water quality
Functional indicators	Reflect and or quantify the processes undertaken within an ecosystem and which may serve as surrogates for the provision of ecosystem services, e.g., functional potential of the community.
Organisational indicators	Reflect the structure and composition of the biota and physicochemical attributes of the ecosystem, e.g., Shannon diversity.
Stressors	Factors or pressures that disrupt the natural state of an ecosystem, e.g., heavy metal pollution.

Table 2-2. Definitions of terms related to groundwater ecosystems, quoted from Hose *et al.* (2023) and Korbel and Hose (2011).

Korbel and Hose (2011) proposed the use of the Groundwater Health index (GHI) framework using the indicators defined in Table 2-3, where assessing the GHI involves a tiered approach, starting with generic indicators and progressing to more detailed assessments based on reference sites, ultimately providing a measure of groundwater ecosystem health based on the number of failed indicators. Microbial abundance, biodiversity and functional potential can be used as ecosystem health indicators (Fillinger *et al.*, 2019b; Griebler *et al.*, 2010; Korbel and Hose, 2011; Stein *et al.*, 2010). The definition of the measurable variables to determine ecosystem health, their units for measurement and their relevance as an ecosystem health indicator are given in Table 2-3. To use the ecosystem health indicators for groundwater ecosystem management, it is essential to classify ecosystem management zones in different spatial and temporal scales and then define the "reference" values from minimally disturbed communities to detect community changes upon disturbances (Hose *et al.*, 2023; Korbel and Hose, 2011). Previously, groundwater ecosystems have been classified on a large spatial scale based on macro-fauna presence in Germany (Stein *et al.*, 2012) and England (Weitowitz *et al.*, 2017). However, there is still debate about what the appropriate criteria are for microbial ecosystem management classification zones at a large spatial scale. Moreover, the reference structure of the groundwater ecosystems is lacking in most aquifers of the world due to the scarcity of data from minimally disturbed groundwater systems.

Parameter	Definition
Ecological niche	It is the position of a particular species in an ecosystem where it has all the required resources and can actively survive.
Cell density	The number of total microbial cells in each volume of sample is measured as counts/mL for groundwater and is higher in groundwater prone to chemical and microbial contamination.
Cell viability	The proportion of live and healthy cells in a population capable of metabolic activity and growth.
Species richness	The number of different species in a community.
Species evenness	The measure of how equally abundant the individuals are for each species. If there is an almost equal abundance of individuals belonging to each species, the community is said to have high evenness.
Species diversity	Diversity combines both species richness and evenness. If a community has high species richness with equal abundance of individuals of each species, the community has high diversity.
Alpha-diversity	Mean species diversity in a community at a specific point in space and/or time.
Beta-diversity	Difference in microbial diversity in two communities from different points in space and/or time.
Diversity indices	It is the measurement of diversity. There are different diversity indices, e.g., Simpson's diversity index, Shannon's diversity index, etc. A higher diversity index helps a community to withstand external perturbations.
Shannon diversity index (H)	$H = -\Sigma(p_i * \ln(p_i))$, where, p_i= proportion of total population represented by species i. Higher the value of H, higher the species diversity in a community.
Functional potential	The range of metabolic activities a microbial community can perform, determined based on known capabilities of known taxa using the FAPROTAX function (Louca *et al.*, 2016).

Table 2-3. Microbial ecosystem terms used for ecosystem health determination.

2.3 Factors controlling the spatio-temporal variation of the prokaryotic ecosystem

2.3.1 Aquifer properties

The geological properties of an aquifer can play a vital role in controlling the prokaryotic ecosystem in groundwater in two ways. First is by providing physical space to the microbes, and the second is by controlling the physicochemical properties of groundwater. The subsurface environment has limited physical space, and thus the habitats of microorganisms are constrained to the pore spaces, fractures, and fissures (Gregory *et al*., 2014). The subsurface mobility of microbes is dependent on the size of pore throats or fracture apertures, matrix mineralogy, and groundwater flow velocity. The size range for the majority of bacteria is 0.1 to 100 µm, for archaea it is 0.1 to 10 µm (Gregory *et al*., 2014; Taylor *et al*., 2004). If the pore-throat of porous aquifer matrix, or the aperture size of fractured aquifer matrix, is smaller than the prokaryotic size, it can physically filter them out and therefore restrict microbial transportation. Microbial filtration by the aquifer matrix has critical implications for the attenuation of pathogens in the groundwater system (Bloomfield *et al*., 2001; Hunt and Johnson, 2017; Taylor *et al*., 2004). Bloomfield *et al*. (2001) showed that in unconsolidated Permo-triassic sandstone of England, the median pore throat size is between 0.1 and 90 µm, which may allow easier transportation of smaller pathogenic microbes like *Clostridia, E. coli*, and *Salmonella* through aquifer media, but tighter pore-throats may filter out larger pathogenic microbes like *Cryptosporidium*. Due to the larger apertures of dissolutionally enlarged karstic fractures (0.5-2 cm diameter) and small conduits (5-30 cm diameter), karstic aquifers are more vulnerable to faecal microbial contamination from the surface (Maurice *et al*., 2023a; Maurice *et al*., 2023b).

Another critical factor is the ionic strength of the aquifer matrix-forming minerals, which can allow attachment of mobile microbes. Hunt and Johnson (2017) explained that since most of the pathogenic microbes and aquifer matrix mineral grains have negative surface charge, larger microbes at a lower physical gap (<10 nm) from the sediment surface are under dominant repulsive force and are more likely to be more mobile. In contrast, the smaller microbes at higher gaps (10-100 nm) from the sediment surface face a weak van der Waals attraction force and are less likely to be mobile. This can result in greater mobility of larger microbes than smaller microbes in the aquifer. Among different aquifer matrices, sand grains are found to have higher ionic strength and higher bacterial retention if coated with positively charged Aluminium (Al) and Iron (Fe) metal-oxyhydroxides, which can impede bacterial mobility and cause clogging in a sandy aquifer (Bolster *et al.*, 2001). With time, the sorption rate on sediment gets reduced as biofilms cover the matrix surface, allowing more microbes to stay in suspended form and move freely (Bolster *et al.*, 2001).

Groundwater flow velocity is another important determinant of prokaryotic movement in groundwater. The flow velocity of groundwater is proportional to the cross-sectional area of gaps within the flow media and the hydraulic gradient. Unconsolidated sediments and fractured aquifers with larger openings of groundwater flow media allow faster groundwater flow along with limited filtration of the suspended microbes, which also allows higher mobilisation rates of suspended microbes (Taylor *et al.*, 2004). In the fractured and fissured rocks with high flow velocity, groundwater is vulnerable to pathogenic contamination due to their easy and rapid mobilisation from the surface to large distances within the aquifers along groundwater flow paths (Taylor *et al.*, 2004). The hydraulic gradient can increase flow velocity during both recharge and discharge periods. During groundwater recharge events, as the hydraulic gradient increases flow velocity, biofilm-forming microbes can be detached and get transported as suspended load (Yan *et al.*, 2021). Savio *et al.* (2018) documented detached biofilms in suspension during high discharge events in karstic aquifers, which add to the suspended bacterial populations. In groundwater pumping boreholes, a steep rise in hydraulic gradient can disturb the biofilms and resuspend them (Roudnew *et al.*, 2014). Because topography shapes hydraulic gradients and subsurface connectivity (e.g. steeper slopes induce stronger gradients and flow pathways), it can modulate microbial community assembly processes. In steeper, more dissected topography where hydraulic gradients are stronger, dispersal is more likely to dominate community assembly (i.e. microbes are transported more freely), whereas in flatter or more stable terrain, where flow is slower and environmental niches are more constrained, selection becomes more dominant, filtering communities according to local geochemistry and redox conditions (Retter *et al.*, 2023).

Aquifer matrix properties not only affect the physical properties of microbial movement but also control the chemical properties of groundwater in two ways. First, physical factors like water flow rate and flow gradient control the movement of ions in groundwater, and second, the water chemistry is controlled by the reaction of water and rock. Ben Maamar *et al*. (2015) reported that hydrological flow-paths can control surface water input and associated oxygen and nitrate transportation in a fractured aquifer. As a result, the groundwater chemistry and subsequently the microbiology were different in groundwater pockets with regular surface-water input compared to isolated groundwater pockets with minimal surface water input. Smith *et al*. (2012) found that the presence of a confining layer resulted in higher chemical concentrations of Fe, sulphur and organic carbon in unconfined groundwater than confined groundwater, and different prokaryotic communities. A recent large-scale survey of a French limestone aquifer demonstrated that microbial richness declines in deeper or more reducing zones and that redox-sensitive factors (e.g. Fe/Mn, dissolved O_2) are strong predictors of community turnover (Harris *et al*., 2025). Amalfitano *et al*. (2014) compared the prokaryotic communities as groundwater flows from volcanic to alluvial aquifer and found that the “hydrogeochemical facies” shifts from silicate, phosphate, potassium, DO rich with higher redox potential groundwater in volcanic aquifer to calcium, bicarbonate ion, Fe and Mn rich, low redox potential, low DO groundwater in alluvial aquifer, controlled by both the surface influence and rock-water interaction. Consequently, the microbiology of the two aquifers was also different. Mehrshad *et al*. (2021) observed that in granitic aquifers of Sweden and Finland, despite a large spatial distance, similar ecological niches provided by similar lithologies hosted similar core microbiome. Similarly, Zhong *et al*. (2023) found from a national prokaryotic survey in China that Na, K, Cl and bicarbonate ionic concentrations, controlled by the aquifer geology, were strong determinants of prokaryotic community differences in pristine groundwater. Recently, Abraham and Close (2024) conducted a comparative study of volcanic and sandy aquifers of New Zealand and found similar results, where the distinct hydrogeochemical conditions within fractured basalt and coarse sand aquifers are the primary drivers of their observed prokaryotic community differences.

Although there is strong evidence that aquifer properties can have a significant effect on the prokaryotic community structures, and hence, support distinct biogeochemical processes, comparative accounts of prokaryotic communities in different geologies are rare. For example, Stein *et al*. (2010), Griebler *et al*. (2010) and Amalfitano *et al*. (2014) compared two different aquifer types with fractured and porous geologies, but due to the hydraulic connectivity between them, there is a chance of microbial dispersal between aquifers, which was observed by Stein *et al*. (2010) and Griebler *et al*. (2010). In the recent example of Abraham and Close (2024), the geographic difference of the fractured and porous aquifer sites indicates that the microbial data may have been from aquifers without hydraulic connections. Another study by Couton *et al*. (2023) found that spring sources from shallow unconsolidated porous and fissured aquifers showed distinct micro and macro organisms. However, more studies from different aquifer geologies should be conducted in national-scale surveys. There is a potential that different aquifer types may be the basis of classification of groundwater ecosystem management categories on a national scale. Additionally, in regional studies, the scarcity of studies accounting for geologically unique and hydraulically disconnected aquifers prevents accounting for the intrinsic prokaryotic communities of groundwater in different aquifers.

2.3.2 Groundwater recharge

A major controlling factor of the subsurface ecosystems is the replenishment of groundwater nutrients during recharge events. During natural recharge by rainwater or river water infiltration, or artificial recharge, as surface water infiltrates through the soil and percolates into groundwater via the overlying strata, it washes down essential macronutrients, among which the effects of replenishment of organic carbon, oxygen and nitrate are widely studied. The replenished DOC and DO concentrations were found to trigger heterotrophic respiration rates and growth in bacterial numbers (Cooper *et al*., 2016; Pedersen *et al*., 2008; Reiss *et al*., 2019; Stegen *et al*., 2016;

Voisin *et al*., 2018; Zhou *et al*., 2012). After the occurrence of groundwater recharge, the increased pH, temperature, EC, and increased concentration of DOC, DO and nitrate, among other essential compounds, changes the groundwater physiochemistry, and the new physiochemistry selects the dominant prokaryotes which can survive in the post-recharge ambient conditions (Fillinger *et al*., 2021; Smith *et al*., 2015). Several studies about the impacts of the groundwater chemistry on the prokaryotic communities are outlined in Section 2.3.3. Additionally, recharge from different land-use types has been shown to impact the microbiome in shallow groundwater due to differences in chemical and anthropogenic stressor input. Couton *et al*. (2023) documented that groundwater ecosystems under a forested area were more diverse and species-rich than agricultural areas, potentially due to stressor input from agricultural areas. Similarly, Korbel *et al*. (2013) found that shallow (< 30 m deep) groundwater microbiome compositions were different under different agrarian land management types.

Besides replenishing nutrients, recharge water introduces various allochthonous microorganisms into the groundwater. Although there are agreeable answers to "what" changes in the community upon the arrival of allochthonous taxa, uncertainties exist about "how" the changes occur. In the aquifer, the sediment-attached communities are more or less stable communities showing little to no change under surface water influence, while significant changes occur within the planktonic prokaryotic communities (Fillinger *et al*., 2021; Fillinger *et al*., 2019c; Yan *et al*., 2021). Upon allochthonous microbial loading by recharge water intrusion via large apertures of fractures or conduits (preferential flow paths), or short flow path from surface water source to groundwater, bacterial cell concentrations get elevated (Besmer *et al*., 2016; Fiedler *et al*., 2018; Fillinger *et al*., 2021; Sorensen *et al*., 2018). During groundwater recharge through preferential flow paths, pathogenic microbes can be introduced, making the groundwater susceptible to contamination (Chik *et al*., 2020; Sorensen *et al*., 2018; Vucinic *et al*., 2022). Mainly, if the catchment land use consists of farmlands, the groundwater community was found to harbour pathogens and antimicrobial resistant genes due to allochthonous input (Smith *et al*., 2012). On the other hand, filtration of allochthonous Bacteria can result in dilution of groundwater cell concentrations upon recharge (Karwautz *et al*., 2022). Conflicting evidence exists about the contribution of the allochthonous taxa among the dominant groundwater prokaryotes. Fiedler *et al*. (2018) recorded that river water intrusion can create new ecological niches, fostering the proliferation of either previously rare prokaryotic groups within the groundwater community or newly arrived groups from the river water. Fillinger *et al*. (2019c) found that in alluvial aquifers, community shifts may result from temporal succession, where early-arriving allochthonous colonisers outcompete late-arriving ones post-recharge. Yan *et al*. (2021) found that during recharge into a fractured aquifer, allochthonous soil-derived taxa (e.g., *Saccharimonadales* and *Ca. Peribacteria*) became dominant over indigenous groundwater taxa (e.g., *Nitrospira* and *Thermodesulfovibrionia*), and new prokaryotic migration into the aquifer environment led to an elevated alpha-diversity of the community. In a recent study including 10-years long dataset revealed that in a limestone aquifer, groundwater isolated from frequent surface recharge by mudstone layers showed a dominant selection pressure on microbial communities, while groundwater with more surface recharge showed dominance of stochastic processes shaping the communities (Wang *et al*., 2025b). In contrast, Fillinger *et al*. (2021) found that changes in an alluvial groundwater community upon groundwater recharge were due to the sorting within the indigenous community due to the selection process by the new groundwater physicochemical condition, rather than the proliferation of allochthonous taxa. Some studies suggested that the extent to which allochthonous prokaryotic intrusion may alter the indigenous groundwater prokaryotes depends on the local geological conditions. Chik *et al*. (2020) and Villeneuve *et al*. (2022) reported that allochthonous bacteria can disperse to larger distances in fractured karstic geologies with faster groundwater flow, compared to those in porous geologies with slow-flowing groundwater. This evidence suggested that a knowledge gap remains about how the groundwater prokaryotic community changes in response to the groundwater recharge in different aquifer types.

2.3.3 Groundwater chemistry

The microorganisms residing in the oligotrophic aquifer environment are metabolically either chemoheterotrophic or chemolithoautotrophic (Danielopol *et al.*, 2000; Goldscheider *et al.*, 2006), adapted to a resource-poor environment, where even a slight change in the physicochemical environment can drive a significant shift in the microbial community parameters (Gregory *et al.*, 2014; Griebler and Lueders, 2009). Numerous studies have documented how the presence of certain chemical entities controls the groundwater microbial community structure and biogeochemical transformation of other chemicals.

One of the most important controlling factors is DOC, which is the fundamental component for chemoheterotrophic respiration. The source of DOC in groundwater can be either related to allochthonous carbon input during recharge or in situ DOC within the aquifer matrix (Shen *et al.*, 2015). Additionally, microbial necromass is a substantial source of DOC in deeper groundwater (Geesink *et al.*, 2022; Parkes *et al.*, 2014). Recent findings suggest that the groundwater bacterial community can be self-sufficient and produce DOC by fixing dissolved CO_2. In a limestone aquifer, under strong limiting abundance of labile DOC and oxygen, anammox bacteria were found to couple the anaerobic ammonium oxidation process with autotrophic CO_2 fixation (Kumar *et al.*, 2017). In both oxic and anoxic groundwater, bacterial CO_2 fixation occurs along with coupled nitrogen and/or sulphur cycling (Overholt *et al.*, 2022; Wegner *et al.*, 2019). In carbonate aquifers, the carbon fixation rate by groundwater bacteria can be almost equal to CO_2 fixation rates in oligotrophic marine surface water (Overholt *et al.*, 2022). These findings highlighted the importance of groundwater ecosystems in regulating the global carbon budget.

In the saturated zones of aquifers, DOC supports a non-competitive, highly diverse and evenly distributed heterotrophic microbial community structure (Zhou *et al.*, 2002). The DOC input with surface water during groundwater recharge can raise bacterial concentration (Hofmann *et al.*, 2020; Reiss *et al.*, 2019; Zhou *et al.*, 2012). However, the DOC surge in the hyporheic zone was related to a reduction in prokaryotic diversity due to the dominance of selective species capable of carbon metabolism (Li *et al*., 2012; Stegen *et al*., 2016; Zhou *et al*., 2012). A surge of thermodynamically bioavailable DOC from surface water sources increases the heterotrophic respiration rate in the groundwater, which results in a decreased DOC concentration and increased CO_2 (dissolved inorganic carbon or DIC) as the respiration product and also stimulates microbial dissolution of the carbonate in the aquifer matrix (Cooper *et al*., 2016; Stegen *et al*., 2018). The concentration of DOC not only changes the community structure but also promotes other metabolic functions of the prokaryotes and controls biogeochemical processes. Most notably, DOC acts as an electron donor in microbial DNRA (Dissimilatory nitrate reduction to ammonium) and denitrification reactions (Liu *et al*., 2017). Low DOC (<1.3 mg/L) groundwater environment was found to prevent heterotrophic denitrification (Ben Maamar *et al*., 2015). Moreover, the concentration of DOC can reduce the toxicity of heavy metals like UO_2^{2+}, Zn^{2+}, Co^{2+}, Cd^{2+}, and Ni^{2+} by forming organic acid complexes with these cations (Carlson *et al*., 2019). On the other hand, DOC facilitates microbially mediated geogenic arsenic release in most arsenic-polluted groundwater (Islam *et al*., 2004; Wang *et al*., 2014).

Another critical controlling factor of groundwater microbiology and biogeochemistry is DO concentration. Aquifer environment is generally oxygen poor. In the subsurface environment, DO is supplied into the aquifers during the recharge process. As the replenished surface nutrients, such as DOC and DO concentrations decreased with depth, the ecosystem was found to change from an aerobic-activity dominated to an anaerobic-activity dominated ecosystem (Pedersen *et al*., 2008). However, recent evidence suggests that in-situ oxygen production within the aquifer by anaerobic processes like denitrification and sulphate reduction is also possible (Ruff *et al*., 2023; Ruff *et al*., 2024). This indicated that the oxygen gradient in groundwater is not necessarily related to spatial distance from or time lag after recharge. When oxygen-rich recharge water infiltrates into an unsaturated aquifer, most of the DO gets consumed by the aerobic microbes in the overlying soil layer (Voisin *et al*., 2018). With time and distance from the recharge source, oxygen consumption by aerobic respiration depletes the DO pool, and the groundwater gradually becomes suboxic to anoxic, and the prokaryotic communities change along the oxygen concentration gradient. In oxic conditions, aerobic taxa are found to grow as dominant taxa in the community, while suboxic to anoxic groundwater hosts obligate anaerobes and anaerobic taxa (Kumar *et al.*, 2017). Smith *et al.* (2015) found that in a typical oligotrophic oxygen-poor porous aquifer environment, the bacterial community shifted from anaerobic taxa dominated (e.g., *Sphingomonadales* and *Rhodospirilales)* to aerobic taxa dominated (e.g., *Burkholderiales, Flavobacteriales, Pseudomonadales*) upon oxygen-rich wastewater recharge.

The biogeochemical processes shift from anaerobic processes to aerobic processes along the gradient of increasing DO concentration. For instance, Kumar *et al.* (2017) reported a shift of aerobic ammonium oxidation to anaerobic ammonium oxidation with decreasing DO concentration in carbonate aquifer groundwater. Due to the presence of different active microbial chemoautotrophic pathways in different oxygen concentrations, Overholt *et al.* (2022) reported that carbon fixation processes in anoxic groundwater were associated with sulphur oxidation, but in oxic groundwater, C-fixation was associated with the nitrification process. Under oxic conditions, groundwater can become nitrate-rich as the NO_3^- reduction processes are suppressed by NH_4^+ oxidation (Liu *et al.*, 2017). Wegner *et al.* (2019) noted that percolation of oxygen-rich water into karstic groundwater was positively correlated with nitrifying bacteria *Nitrospira* abundance and associated nitrification process, whereas anoxic conditions were positively correlated with anammox bacteria like *Planctomycetes*, and associated anammox process. Kim *et al.* (2018) observed that in a hyporheic zone, anoxic groundwater upwelling stimulated microbial denitrification. Danczak *et al.* (2016a) found that in a hyporheic zone, depending on the relative contribution of oxic river water and oxygen-poor groundwater, the metal-reducing bacteria families like *Geobacteraceae* and *Desulfuromonadaceae* control the annual iron redox cycling. Therefore, oxygen concentration has an important influence on biogeochemical transformation of other elements, such as carbon, nitrogen, sulphur, and iron.

Since groundwater prokaryotes can survive under a narrow range of low chemical concentrations, severe disturbance of water quality by chemical contamination can shift the community structure. Hemme *et al.* (2015) reported that uncontaminated groundwater harboured a diverse prokaryotic community with *Burkholderia* and *Pseudomonas* dominance. In contrast, groundwater heavily contaminated with nitrate, sulphate, uranium, and radionuclides hosted a stressed community with a low diversity and the dominance of *Rhodanobacter.* The dominance shift from the *Pseudomonas* population to the *Rhodanobacter* population in a heavy metal-contaminated low pH groundwater was also documented by Carlson *et al.* (2019), who explained that the shift of dominant bacterial taxa is due to deterministic selection imposed by low pH and high UO_2^{2+}, Mn^{2+}, Al^{3+}, Cd^{2+}, Co^{2+}, Zn^{2+}, and Ni^{2+} concentrations. Similarly, in a nuclear waste contaminated site, highly contaminated zones had strong community selection pressure, whereas moderate to low contamination sites with lower chemical stress showed stochastic processes dominated community assembly (Ning *et al.*, 2024). Due to loss of prokaryotic diversity under chemical stress, the biogeochemical transformation of chemicals can be suppressed, for instance, a change in carbon and nitrogen turnover rates resulting in quality degradation of the bulk groundwater (Hemme *et al.*, 2015).

Pristine aquifer communities have a wide range of phylogenetic species which have the same metabolic functions, making the system more stable, as loss of a particular species will not collapse the biogeochemical system, and due to the presence of genes programmed to perform multiple metabolic functions, the community can rapidly modify and withstand environmental stresses (Anantharaman *et al.*, 2016; Danczak *et al.*, 2018; Hemme *et al.*, 2015; Konopka *et al.*, 2015; Korbel and Hose, 2011; Zhong *et al.*, 2023). However, despite limited studies on the effects of groundwater chemistry in uncontaminated aquifer microbial communities, numerous studies were performed on either chemically contaminated aquifers (Carlson *et al.*, 2019; Fahy *et al.*, 2005; Hemme *et al.*, 2015; Mattsson *et al.*, 2015; Pickup *et al.*, 2001), or shallow (generally under 50 m deep) groundwater systems with frequent surface water input (Danczak *et al.*, 2016a; Kim *et al.*, 2018; Liu *et al.*, 2017; Stegen *et al.*, 2018; Voisin *et al.*, 2018; Wegner *et al.*, 2019). There is a crucial knowledge gap about the prokaryotic ecosystems in groundwater with limited contamination and limited influence from recharge water. A lack of data from cleaner groundwater used by the drinking water supply industries leads to problems while deducing a baseline community structure and thus devising effective management plans to maintain a pristine groundwater ecosystem.

2.3.4 Groundwater residence time

The residence time of groundwater is also a crucial governing factor of the prokaryotic communities. The young groundwater, which has recently mixed with surface water, can contain surface-derived nutrients, like DO, DOC, nitrate, etc., with neutral to alkaline pH, lower conductivity, and a higher temperature. As recharge water moves within the aquifer, the DOC, DO, and nitrogen species are transformed with time, and the water chemistry changes with time and distance from the point of recharge (Cooper *et al*., 2016; Stegen *et al*., 2016; Voisin *et al*., 2018). Older groundwater also had a longer contact time with the aquifer matrix, allowing rock-water interaction to dissolve and precipitate chemicals (Elango and Kannan, 2007). Such groundwater can be rich with Fe, Mn, sulphur compounds, etc., with low pH and high conductivity. Ancient groundwater stored as formation water can often have unique chemistry, such as saline water stored in deep British chalk aquifers (Shand *et al*., 2007). In a national survey of groundwater bacterial distribution in New Zealand, communities were different in the older water with more than 100 years of residence time than their younger counterparts, primarily related to redox conditions (Sirisena *et al*., 2013). However, this study did not include the taxonomic classification of the prokaryotes. Ben Maamar *et al*. (2015) studied in detail the effect of groundwater age on the geochemistry and microbiology at a fractured hard rock aquifer in Western France. They distinguished the groundwater according to the residence time as old (>40 years to millennium scale) and recent (<25 years). They found that the recent groundwater has high DO due to recharge, high nitrate concentration from agricultural land, as well as high redox potential and consequently harboured bacterial families like *Comamonadaceae, Oxalobacteraceae,* and *Rhodocyclaceae,* and plenty of denitrifiers, although denitrification was rare in their study area due to a lack of DOC. Also, the recent water shows vulnerability to faecal contamination from livestock waste, as evident from the presence of *Clostridium*. On the other hand, the older groundwater has high iron, manganese, high pH and more anoxic conditions and a low nitrate concentration as a result of reduced interaction with surface water. Due to low mixing with surface water, older groundwater showed complete redox cycling of Fe and S as evident from the presence of Fe-oxidising *Galionella* and *Sideroxydans,* Fe-reducing *Anaeromyxobacter*, sulphur-oxidising *Desulfobulbus, Desulfatirhabdium* and sulphur-reducing *Desulfosalsimonas* together. However, the large-scale spatial distribution of prokaryotic ecosystems in groundwater and how it depends on groundwater residence times is not widely studied.

2.4 Current understanding about groundwater microbiology of England

In England, groundwater is the source of ~30% of the industrial public water supply (British Geological Survey, 2019). Three aquifer types, Permo-triassic sandstone, Cretaceous chalk and Jurassic limestone, form the three major aquifers of England (Allen *et al*., 1997). Yet the groundwater microbiology of these aquifers is largely unexplored. The review paper by Gregory *et al*. (2014) recorded the contemporary state of knowledge about the English groundwater microbes. Most work before that time has focused on groundwater pathogen movement (Bloomfield *et al*., 2001; Cronin *et al*., 2003; Powell *et al*., 2003) or biodegradation of contaminants in contaminated groundwater (Aburto and Ball, 2009; Fahy *et al*., 2005; Fahy *et al*., 2006). Although information about uncontaminated groundwater was available from reference sites of the contaminated sites, systematic studies of groundwater microbiology are scarce. Some studies have included using microbial isolates based on their functional importance. For instance, Whitelaw and Rees (1980) detected nitrate-reducing and ammonium-oxidising bacteria in the unsaturated zone of the English chalk aquifer, Bartlett *et al*. (2010) reported the presence of sulphate-reducing bacteria in the sandstone aquifer, Fahy *et al*. (2005) reported the dominance of *Betaproteobacteria* and *Firmicutes* in uncontaminated sandstone groundwater sites. Utilising molecular techniques to gain a holistic view of the groundwater microbiology was rare, except for one example by Sorensen *et al*. (2013), where the T-RFLP molecular fingerprinting technique was used in the chalk aquifer. Gregory *et al*. (2014) suggested that a systematic survey of groundwater microbiology of England should be undertaken to understand their spatial and temporal variation and their contribution to the subsurface biogeochemical cycles.

Since then, minimal advancement has been made to understand the groundwater microbiology of England. A few studies have utilised advanced microbial detection technologies to explore microbial communities. For example, flow cytometry was used to detect a rise in cell concentration and faecal pathogen contamination in chalk and limestone groundwater sources (Sorensen *et al.*, 2018; Sorensen *et al.*, 2020). In another study, flow cytometry was used to study changes in groundwater cell concentrations in response to groundwater flooding (Reiss *et al.*, 2019). By contrast, groundwater macrofauna have been systematically studied, and different groundwater systems have been classified based on the macrofauna habitat quality (Weitowitz *et al.*, 2017). In contrast, state-of-the-art eDNA-sequencing techniques have been used for systematic studies of groundwater microbial ecosystems in many European countries (Griebler *et al.*, 2010; Retter *et al.*, 2023; Stein *et al.*, 2010; Steube *et al.*, 2009), the USA (Merino *et al.*, 2022), China (Zhong *et al.*, 2023), Australia (Korbel *et al.*, 2024; Smith *et al.*, 2018) and New Zealand (Abraham and Close, 2024; Sirisena *et al.*, 2018). This type of systematic regional study is necessary in England for the sustainable management of groundwater resources.

2.5 Research gaps and thesis relevance

Sustainable management of the limited groundwater resources requires comprehensive data about the microbial ecosystems, targeted towards maintaining their ecosystem health, thus their ecosystem services (Hose *et al.*, 2023; Korbel and Hose, 2011; Saccò *et al.*, 2024). As indicated in Section 2.3, since the 2000s, an increasing number of studies have investigated controlling factors on groundwater prokaryotic community composition. However, the spatio-temporal variation of groundwater prokaryotic ecosystems has been scarcely explored at a national scale (Sirisena *et al.*, 2018; Zhong *et al.*, 2023). As mentioned in Section 2.3.1, in large-scale spatial surveys, the role of aquifer geology in controlling spatio-temporal variation has not been explored despite substantial research pointing towards the fact that aquifer geology can be a potential classification parameter of groundwater prokaryotic ecosystems. Of particular note, large-scale studies are often not temporally repeated, preventing the conceptualisation of how groundwater prokaryotes can change in response to groundwater recharge in different aquifer types. This thesis makes a novel contribution to this area of science in the following ways:

First, this is the first study of the groundwater microbes in English aquifers at a national level. The study was systematically performed to explore the prokaryotic microbes suspended in groundwater from three major aquifers of England, and modern techniques such as flow cytometry and eDNA amplicon sequencing were used to understand bacterial concentration, prokaryotic taxonomic composition, biodiversity, and functional potentials.

Second, a national-scale survey of groundwater microbiology using modern eDNA sequencing techniques is rare and a relatively recent advancement (Sirisena *et al*., 2018; Zhong *et al*., 2023). However, in both studies, aquifer geology was not considered a major controlling factor of spatial variation. This thesis is relevant because in this national-scale survey, samples from different aquifer types were collected and their prokaryotic ecosystems were compared. Due to no hydraulic connections between the aquifers, the community compositions were intrinsic to each aquifer type.

Third, a knowledge gap remains about how the groundwater prokaryotic community changes in response to the recharge process in different aquifer types (Section 2.3.2). This thesis is relevant to this knowledge gap because prokaryotic data, specifically cell concentration data, were collected from the three aquifers in two different seasons. Seasonal community composition data were collected focusing on a particular aquifer. Thus, an attempt was made to understand temporal patterns in different aquifer geologies.

Fourth, the prokaryotic community data at different groundwater residence times have been rarely documented but are a critical determining factor of the prokaryotic community structure (Section 2.3.4). This thesis is relevant to this knowledge gap because in a particular aquifer, an attempt was made to establish the relationship between groundwater residence time and prokaryotic communities.

Lastly, most groundwater prokaryotic community data are collected from such shallow, near-surface (generally < 50 m deep), and vulnerable aquifers. However, the groundwater microbial community data from uncontaminated aquifers should be collected to define reference microbiology for monitoring disturbances in microbial groundwater quality (Hemme *et al*., 2015; Hose *et al*., 2023; Zhong *et al*., 2023). This thesis is relevant to this knowledge gap because groundwater was collected from drinking water pumping boreholes of various depths (18 m to 391 m) and springs. Both the borehole and spring catchments used as drinking water sources are generally protected from contamination (Environment Agency, 2019; Foster and Chilton, 2003). Since the data is from groundwater without severe contamination, this can be used as a groundwater monitoring reference for respective aquifers in England.

3. Methods of microbial community analysis

This chapter addresses thesis **Objective 2**: To optimise prokaryotic sample collection and analysis methods for groundwater systems with low prokaryotic concentration.

3.1 Flow cytometry

Flow cytometry (FCM) has become an integral tool in analysing and quantifying microbial communities within groundwater and other freshwater systems, owing to its ability to provide rapid, multiparametric insights at the single-cell level. (Hammes and Egli, 2010). The working principle of FCM relies on hydrodynamic focusing of fluorescent-stained microbial cells, which are individually passed through a laser beam for detection of forward scatter (FSC), side scatter (SSC), and fluorescence from nucleic acid stains (Figure 3-1). As it is independent of cultivation, this cost-effective and rapid technique is increasingly utilised in the UK groundwater supply industry to enumerate bacterial cells (Van Nevel *et al*., 2017). Flow cytometry is typically used to measure the bacterioplankton concentration in groundwater. This analysis may or may not include archaea population within the bacterial gates. However, the archaea concentration in groundwater is only 20% of the prokaryotic population (Griebler and Lueders, 2009), and the protocol for flow cytometric archaeal cell count is not well established yet. Therefore, only bacterioplankton cell concentration was used when reporting flow cytometric measurement results.

The total cell concentration (TCC) of bacterioplankton, a primary parameter measured by flow cytometry, is being increasingly used to detect short-term changes in microbial water quality and to classify long-term vulnerability. Typical TCC of uncontaminated groundwater falls within 10^2 to 10^6 cells/mL range, and in contaminated groundwater, it can range from 10^3 to 10^7 cells/mL (Griebler and Lueders, 2009). In vulnerable groundwater with frequent pathogenic contamination events, the TCC was found to be higher than the TCC in low-vulnerability springs in Alpine karstic aquifers (Farnleitner *et al*., 2005; Sinreich *et al*., 2014). In time-series analysis, pathogen indicator organism spikes were related to a rise in TCC, followed by a drop in TCC as the pathogen indicator number declined (Sorensen *et al.*, 2018; Sorensen *et al.*, 2020; Vucinic *et al.*, 2022). A key development in the last decade has been the deployment of automated online FCM platforms for real-time, high-frequency microbial surveillance (Besmer *et al.*, 2014). Bacterial TCC was proposed as a biomonitoring tool, used in tandem with measurements of assimilatory carbon and bacterial activity, to differentiate between contaminated and uncontaminated groundwater (Fillinger *et al.*, 2019b). Besides total cell concentration, flow cytometry also allows differentiation between high versus low nucleic acid content cells and intact versus damaged cells. Differentiation of these fingerprints is crucial for ecological assessments as the HNA bacteria are assumed to be taxonomically different from the LNA populations (Proctor *et al.*, 2018), and cellular intactness informs about the bacterial cell viability (Davey and Guyot, 2020). Intact cell counts and HNA bacteria counts were found to act as indicators of microbial contamination and anthropogenic disturbances in groundwater (Amalfitano *et al.*, 2014; Vucinic *et al.*, 2022).

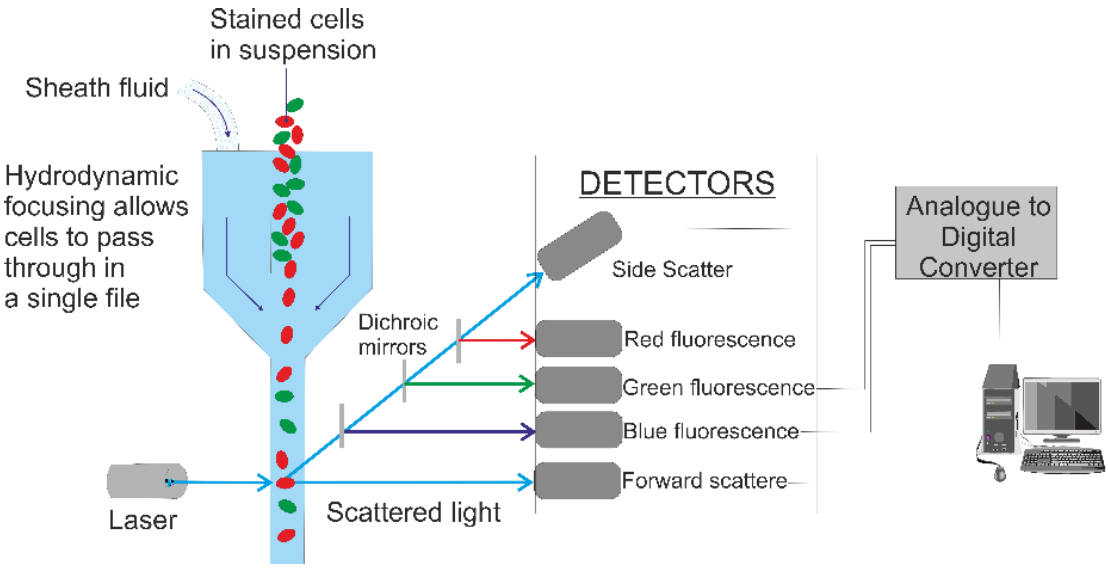


Figure 3-1. Schematic diagram of the working principle of flow cytometry showing hydrodynamic focusing of stained bacterial cells and different scattered light and fluorescence detector channels.

3.1.1 Optimising bacterioplankton staining protocol for flow cytometric measurement

For flow cytometric measurement, the bacterial cells suspended in water samples are stained with fluorescent dyes so that they can be identified using the scattered and fluorescent light detectors. Prest *et al*. (2013) showed that sample preparation, i.e., staining temperature and incubation time, impacted the reproducibility of flow cytometric bacterial enumeration results from different water sources. While this experiment used multiple samples, groundwater samples were not used for optimising the sample preparation conditions. Currently, efforts are being made to standardise these protocols for large-scale application in major water supply companies and research groups (Safford and Bischel, 2019). The following experiment was performed to assess the effect of incubation temperature and time on groundwater samples and to establish standardised conditions for flow cytometric sample preparation.

For the experiment, three groundwater samples were collected from one Chalk (Bh-N) and two alluvial gravel boreholes (Bh-P & Bh-A) situated at the Boxford farm LOCAR experimental site in the Lambourne River valley in England. Boreholes were pumped to remove three volumes of water before sampling so that the microbiome of the aquifer could be intercepted instead of the modified microbes in the open borehole. Samples were collected in sterile Falcon™ tubes and stored at 4°C for 24 hours to mimic water company sample storage conditions.

The cytometer used in this experiment was a ThermoFisher™ Attune CytPix flow cytometer paired with an Attune CytKick autosampler. For bacterial total cell concentration (TCC), SYBR Green I stain was diluted from a 10000X stock to a 100X concentration using Miliq™ water. This green-fluorescent stain labels all the bacterial cell nucleic acids, enabling their detection on a side scatter (SSC) versus green fluorescence (BL1) intensity plot. Additionally, this staining protocol allows differentiation of high nucleic acid bacteria (HNA), showing higher green fluorescence intensity, from low nucleic acid bacteria (LNA), showing lower intensities. To differentiate intact cells (ICC) from damaged cells, SGPI stain was prepared by mixing SYBR Green I (100X) with 1 mg/mL propidium iodide in a 5:1 ratio. Propidium iodide stains nucleic acids of bacteria with damaged cell walls with red fluorescence, while SYBR Green I label all nucleic acids with green fluorescence. This enabled differentiation of red-fluorescent damaged or dead cells from green-fluorescent intact or live cells.

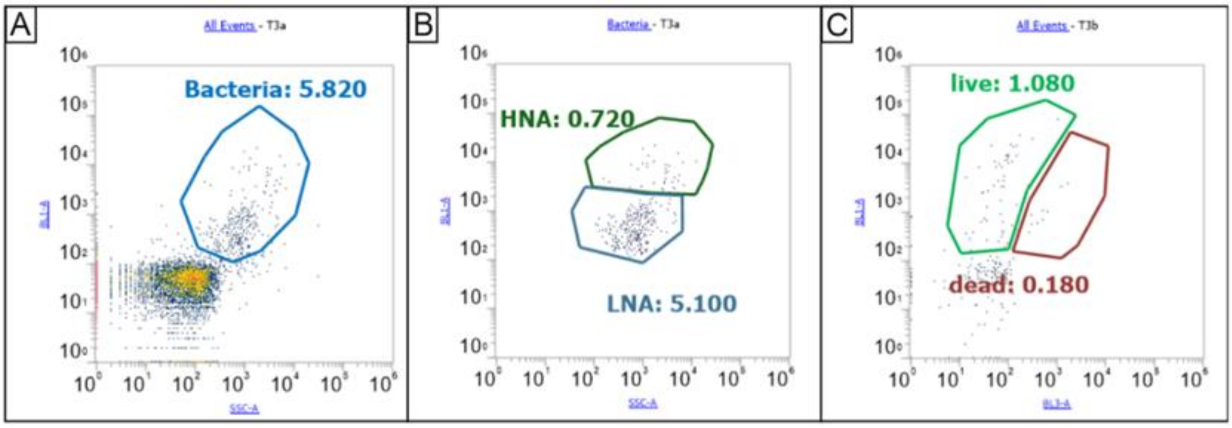


Figure 3-2. Gates used for A. total bacteria, B. HNA and LNA bacteria, and C. live bacteria enumeration, with an example of bacterial signatures within the gates in a water company sample.

For TCC measurement, 200 µL of groundwater was pipetted into six 2 mL Eppendorf™ microcentrifuge tubes. Each tube was stained with 2 µL of 100X SYBR Green I and mixed by a vortex mixer. For ICC analysis, 200 µL of each sample was stained with 2.4 µL SGPI and vortexed. Three tubes from each batch were incubated at room temperature (22°C), and three at 35°C using an Eppendorf™ Thermomixer C. Tubes were taken out of incubation at 10, 15, and 20-minute intervals. These staining conditions were similar to those used in the experiment of Prest *et al.* (2013). After measurement, bacterial TCC, HNA and ICC concentrations were measured using manually prepared gates (Figure 3-2). These gates were prepared based on a previous experiment using a pseudomonas stock to delineate the field where stained bacterial signatures were detected from the field of background noise, as well as some pond water samples, where live-dead conditions were simulated by using 70% ethanol to damage the cells (not shown here).

The experiment showed that during TCC measurement, samples from Bh-A and Bh-P showed no considerable change upon changing the incubation time. But increasing incubation temperature showed a maximum of 2 cells/µL difference. In the case of Bh-N at 22°C incubation temperature, 10 minutes incubation showed higher TCC values than 20 minutes incubation. The relative standard deviation (RSD) of TCC of 6 replicates of each sample incubated under different conditions was 6.1%, 14.8% and 8.7% for Bh-A, Bh-N and Bh-P, respectively. A two-way ANOVA revealed no significant effect (p > 0.05) of incubation temperature, time, or their interaction on TCC (Figure 3-3. A). Similar results for HNA were also found (not shown).

The incubation experiment for ICC measurement showed that samples from Bh-A reacted minimally to changing incubation time or temperature. Samples from Bh-N showed higher ICC at incubation times of 15 and 20 minutes at 22°C, but at incubation temperature 35°C, 10 minutes incubation time yielded higher ICC values. Bh-P samples showed similar ICC values in all the conditions except for almost 50% higher ICC upon 15-minute incubation at 22°C. ICC variation was prominent in Bh-N, where ICC value was higher at 22°C if incubated for 10 minutes, but at 35°C, 15- and 20-minute incubation yielded higher ICC values. For ICC, Bh-P exhibited the highest RSD (30%), followed by Bh-N (19%) and Bh-A (6.3%), and a two-way ANOVA found no significant effect (p > 0.05) of incubation conditions on ICC (Figure 3-3. B).

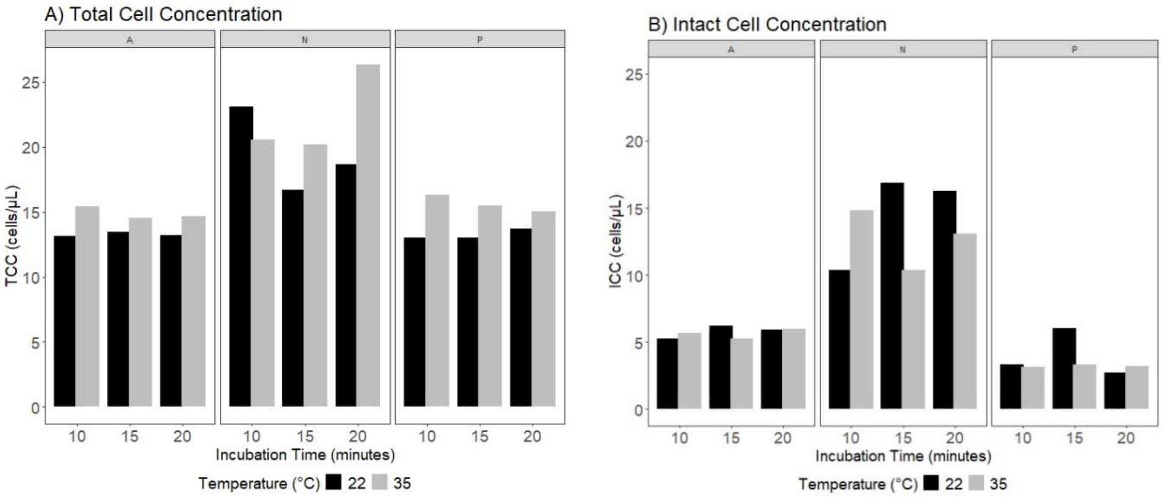


Figure 3-3. Bar plots of Bacterial A. TCC and B. ICC at different incubation temperatures and times.

Although incubation conditions did not significantly affect flow cytometric variables, sample Bh-A consistently showed the lowest variation in TCC and ICC. In contrast, Bh-N showed the highest variations upon changing incubation conditions. Such variability may arise from the intrinsic properties of samples, given that they were from different aquifer types. However, this was a speculation, and no particular reason could be given from the above experiment. However, based on the suggestion of Prest *et al.*

(2013), a constant incubation time and temperature were used for all sample analysis. This constant condition was chosen to be 22°C for 10 minutes for all staining protocols for the convenience of incubation.

3.2 Environmental DNA sequencing

The environmental DNA amplicon sequencing technique has been a groundbreaking development in groundwater microbial studies as it allows detection of hundreds of species from a single sample, enabling comprehensive biodiversity assessments of entire communities (Pawlowski *et al*., 2020; Saccò *et al*., 2022). Compared to traditional culture-based detection methods, where targeted microbes can be grown in a laboratory culture for their detection, this eDNA sequencing allows detection of all microbes in a community cost-effectively and rapidly, and enables discovery of novel sequences (Castelle and Banfield, 2018). The process begins with collecting environmental DNA from aquatic environments, including groundwater samples. The environmental DNA, or eDNA, serves as a forensic proof of the presence of any organism in a sample, and it can be either intracellular or extracellular genetic material (Pawlowski *et al*., 2020). In the case of classification of prokaryotic organisms, the 16S ribosomal RNA or 16S rRNA encoding genes are sequenced since these gene is universally conserved across bacteria and archaea (Janda and Abbott, 2007). The V3-V4 region of the 16S rRNA gene is the most variable region of the prokaryotic population. Thus, sequencing this particular area allows differentiation of most of the prokaryotic microorganisms (Abellan-Schneyder *et al*., 2021). This gene segment of the eDNA fragments is selected by universal primers and amplified by polymerase chain reaction. Due to the development of next-generation sequencing technology, like Illumina Miseq, high-throughput and rapid production of thousands of sequences is possible at a lower cost than traditional Sanger sequencing (Abellan-Schneyder *et al*., 2021). The sequences of the base-pairs, i.e., A, U, G and C, are unique for each organism. The unique sequences, with the exact same nucleotide sequences are classed as an Amplicon Sequence Variant or ASV. The ASVs are matched with a global database such as Silva, GreenGenes, the Ribosomal Database Project, etc., which helps to assign a taxonomy to the ASVs (Abellan-Schneyder *et al*., 2021).

However, reference genetic sequences from poorly explored systems like groundwater are not very robust, which leads to the discovery of many novel sequences (Castelle and Banfield, 2018).

The result of the sequencing is a table of read numbers of each ASV in the samples, and a table of ASV assigned to taxonomic ranks. The ASV read counts are converted to relative abundance of individual ASVs in each sample, or community compositional data (Gloor *et al*., 2017), and the relative abundance is used for downstream analysis depending on research question. Calculation of Bray-Curtis distance, statistical methods such as analysing community differences using ANOSIM (Analysis of Similarity), or modelling community differences and dependencies on environmental variations using ReDundancy analysis (RDA) among many other ecological statistics are analysed based on the relative abundance of ASVs. This is because relative abundance of ASV reads do not represent true counts of a taxa in a sample since multiple technical factors prevent treating ASV reads as absolute abundance measures. For example, DNA extraction and library preparation can remove genetic material, PCR amplification bias can skew sequence counts dramatically, different 16S sequences can amplify with varying efficiency due to primer mismatches, GC content, etc. (Gloor *et al*., 2017). Schloss (2024) argued that rarefying the ASV reads by randomly subsampling (without replacement) to a uniform sequencing depth eliminated some biases related to variable library sizes. Unlike macro-organism eDNA surveys where larger animals may shed more DNA, in microbial communities all individuals are single cells of roughly similar size. Although individual bacterial taxa differ in genome size or rRNA gene copy number, recent work shows that these differences have limited impact on between-sample community composition metrics (e.g., beta-diversity, PCoA, NMDS) in microbial studies (Gao and Wu, 2023). Recent advances are being made to obtain absolute microbial abundance by measuring the total 16S gene copies via qPCR, total DNA mass or by adding internal spike-in standards and then combining that with relative abundances (Tettamanti Boshier *et al*., 2020; Wang *et al*., 2025a), although these methods still does not completely account for non-uniform amplification biases due to GC content, primer mismatch etc. (Wang *et al*., 2025a). With additional analytical capabilities required to do such analysis, representing community composition using ASV relative abundances is a common practice, even in recent groundwater microbiome studies, for example, Abraham and Close (2024); Korbel *et al.* (2024); Sirisena *et al.* (2018); Wang *et al.* (2025b); Yan *et al.* (2021); Zhong *et al.* (2023) etc among others.

Louca *et al.* (2016) established a tool called FAPROTAX, which uses experimentally derived evidence to link specific taxa to their known metabolic capabilities, creating functional profiles of microbial communities. However, for microbial communities such as groundwater, where there is a lack of reference genes, the functional annotation has poor coverage and is inadequate (Sansupa *et al*., 2021). Despite the limitation, this tool can be used as an exploratory tool for functions in an unexplored microbiome, like the groundwater. The 16S rRNA gene sequencing has been successfully applied for exploring the distribution and dynamics of groundwater prokaryotic ecosystems.

3.2.1 Optimising eDNA collection method for groundwater samples

For sampling eDNA for prokaryotic sample analysis, generally a filter of 0.2 µm pore size is used. For marine samples, it has been demonstrated that the combination of filter membrane type and eDNA extraction kit used for eluting the DNA from the filter significantly affects eDNA yield (Djurhuus *et al*., 2017; Hinlo *et al*., 2017). However, no such experiment for groundwater samples with low prokaryotic abundance has been performed. Thus, no standardised method for groundwater eDNA collection and extraction currently exists. Therefore, the impact of filtration protocol should be tested to optimise eDNA collection from groundwater samples. Additionally, the volume of filtered water also influences eDNA yield. High-biomass surface water samples, such as marine or riverine water, generally require only 0.2 L to 2 L to obtain sufficient DNA for PCR amplification (Kumar *et al*., 2022). In contrast, low-biomass samples, like groundwater, often require significantly larger filtration volumes. Literature on groundwater bacterial eDNA sequencing on the Illumina platform reported filtered volumes ranging anywhere from 1.3 L (Stegen *et al*., 2016) to 38 L (Danczak *et al*., 2018), yet no consensus on an optimal filtration volume exists to capture enough eDNA for amplification.

The following experiment aimed to answer 1. Which filter and DNA extraction kit combination yields the highest DNA concentration from groundwater samples? 2. What filtration volume is needed to yield DNA concentrations above a 1 µg/mL threshold (commonly asked by commercial sequencing services)? The experiment design is provided in Figure 3-4.

This experiment was conducted using two chalk boreholes, N15BH and BGSBH, with 24 m and 53 m depth, respectively. Both boreholes were flushed for three borehole volumes before sampling to receive representative samples from the aquifer and not the modified borehole sample. Groundwater was collected in six 70 L drums and subsequently aliquoted into twenty-four 10 L drums. Each drum was cleaned with 1% Virkon™ solution, rinsed thoroughly with sample water, and filled simultaneously using Y-splitters to minimise bacterial concentration (TCC) variation during water pumping.

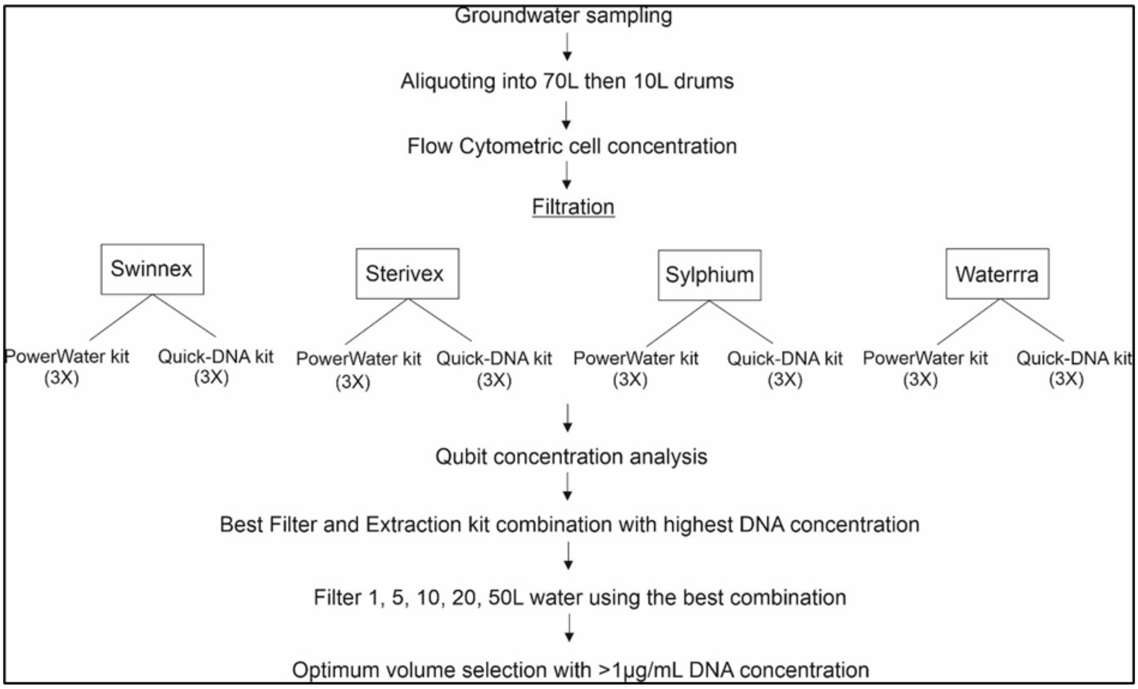


Figure 3-4. eDNA collection method optimisation experimental design.

Four commercially available sterile 0.2 µm filters were tested: 1. Whatman® mixed cellulose ester membrane filters (Cytvia, USA) held within 47mm Swinnex® filtration units (EMD Milipore Corp, USA) 2. Sterivex™ filters with Polyethersulfone (PES) membrane (Merck KGaA, Darmstadt, Germany), 3. Sylphium eDNA Dual Filter™ with

PES membrane (Sylphium molecular ecology, Groningen) and 4. Waterra eDNA Filter™ with Polyethersulfone (PES) membrane (Waterra, USA). For each filter type, 8 L of groundwater was filtered in six replicates. DNA extraction was performed on three replicates using the Qiagen DNeasy® PowerWater® Kit and on the remaining three using the Zymo Quick-DNA™ Miniprep Plus Kit (Figure 3-4). The PowerWater kit employs bead-beating mechanical lysis, whereas the Quick-DNA kit uses Proteinase K-based chemical lysis to elute the DNA from the filter membranes. Filters were preserved with 1 mL DNA/RNA Shield™ reagent.

Extraction protocols were followed per manufacturer instructions, with adaptations for specific filter designs where necessary. The filter membranes from the Swinnex filter units kept in the DNA/RNA shield were extracted using two respective extraction kits following the manufacturer's protocol. For the Sterivex Filters, the shield liquid was expelled with a syringe, and then the filter units were cracked open, and the filter was physically removed from its plastic casing to use in the following extraction steps. For both the Sylphium and Waterra filters, cracking the filter units was not possible. Therefore, only the shield liquid was mixed vigorously within the filter units and pushed out using syringes onto Whatman® membrane filters. These membrane filters were used for extraction following the usual protocols. In the case of three Sylphium filters used for Quick-DNA kit extraction, the proteinase K and Solid tissue buffer of the kit were added directly to the filter unit, mixed thoroughly, and digested at 55ºC for 3 hours. Then the liquid from the filter was pushed out into sterile tubes and further processed following the rest of the protocol. The concentration of total DNA in eluted samples after the extraction steps was measured on a Qubit™ fluorometer by staining the DNA with a Qubit™ dsDNA Quantification Assay Kit.

The average TCC of N15BH was 16.8 cells/µL. The highest DNA yield from this sample was obtained using Sterivex filter - Quick-DNA kit combination (4.7±0.6 µg/mL) and Sterivex filter - PowerWater kit combination (3.4±0.3 µg/mL). The average TCC of BGSBH was 3.2 cells/µL. The highest DNA yield from this sample was obtained using Sterivex filter - PowerWater kit (3.5±0.8 µg/mL). The Swinnex filter produced lower yields, while the Sylphium and Waterra filters yielded the least DNA. Therefore, Sterivex filter - PowerWater kit combination yielded consistently more DNA from samples with various bacterial concentrations, and this was chosen to be the optimal DNA collection and extraction method.

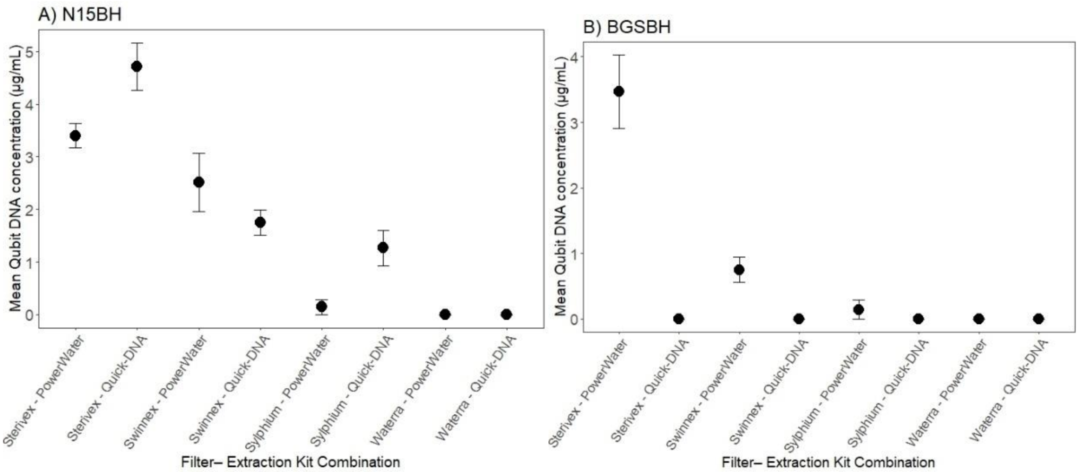


Figure 3-5. Qubit eDNA concentrations (µg/mL) using different combinations of filter and extraction kits, for groundwater samples collected from A. N15BH and B. BGSBH. Plots showing mean value and error bars.

Another experiment was performed to find the optimal filtration volume to yield DNA concentrations above the 1 µg/mL threshold. Using the Sterivex filter and PowerWater kit combination, we assessed the optimal filtration volume for obtaining DNA concentrations exceeding 1 µg/mL. This experiment was conducted on the BGSBH groundwater sample, representing a system with low bacterial abundance and thus allowing determination of the minimum filtration volume. Five filtration volumes (1 L, 5 L, 10 L, 20 L, and 50 L) were tested. Single Sterivex filters were used for 1 L and 5 L samples, while 3, 4, and 6 filters in a series were used for 10 L, 20 L, and 50 L samples, respectively, to reduce filtration time and avoid potential clogging of filters. Filters were preserved in DNA/RNA Shield™ reagent, and membranes and liquid from all the filters in series used for each volume were pooled for extraction.

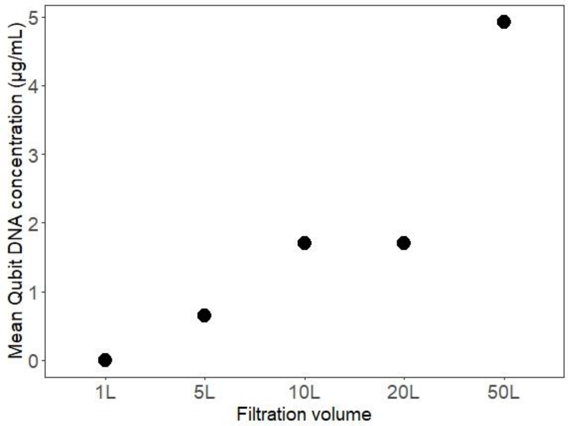


Figure 3-6. Qubit concentration of DNA at different filtration volumes of BGSBH samples, DNA collected using Sterivex filter, PowerWater kit combination.

DNA concentrations above 1 µg/mL were obtained at 10 L and 20 L filtration volumes (average 1.7 µg/mL) (Figure 3-6). The highest yield was at 50 L (4.9 µg/mL). However, 1 L filtration volume yielded a DNA concentration below the detection level. 5 L filtration volume still yielded 0.6 µg/mL DNA concentration, which was lower than the threshold, but still detectable. Although 50 L yielded the highest DNA concentration, logistical constraints made it impractical for routine sampling at water company sites due to the cost and time (3 hours for 50 L filtration). Conversely, 10 L and 20 L volumes were manageable, requiring 30 minutes and 1 hour, respectively. We considered that in case of low water pressure at sample taps of groundwater companies, filtering water samples as high as 50 L would be logistically challenging. To balance obtaining enough DNA and the practicality of sampling, the standard filtration protocol was chosen to be 15 L filtration volume or 45 minutes filtration time, whichever finishes earlier.

4. Aquifer geology controls the bacterioplankton concentration and seasonal changes in groundwater-derived public water sources

This chapter addresses thesis **Objective 3**: To assess the controls on the spatial and seasonal variation of bacterioplankton concentrations in the three different aquifers. A version of this chapter is under preparation for submission.

The dataset is submitted to the Environmental Information Data Centre (EIDC) under the title “Groundwater bacterioplankton, fluorescent organic matter and nutrient concentration of three major aquifers in England, September-October 2022 and January-February 2023”

4.1 Introduction

Groundwater is a daily source of drinking water for about half the global population (Hiscock, 2011). Groundwater is also the host of a rich microbial ecosystem that comprises around 15% of the total global biomass and 80% of the prokaryotic biomass (Bar-On *et al.*, 2018). These ecosystems, dominated by bacteria, offer a unique contribution to global biodiversity due to the adaptation of species to an environment devoid of sunlight and typically limited in chemical energy resources (Chapelle, 2000; Goldscheider *et al.*, 2006; Griebler and Lueders, 2009). Groundwater bacterial ecosystems also provide numerous services, including biodegradation of anthropogenic contaminants, pathogen inactivation, and nutrient transformation, with groundwater considered a large carbon sink (Griebler and Avramov, 2015; Tomlinson and Boulton, 2010). The ecosystem comprises bacteria in benthic (attached) and planktonic (suspended) communities (Griebler and Lueders, 2009). The bacterioplankton fraction is most relevant for the water industry, which routinely analyses groundwater for faecal indicator organisms to assess enteric pathogen risks (John and Rose, 2005; Willis *et al.*, 2013).

Bacterioplankton total cell concentration (TCC) is an important groundwater quality and groundwater ecosystem health indicator (Fillinger *et al*., 2019b; Hose *et al*., 2023). In uncontaminated aquifers, TCC usually ranges from 10^2 to 10^6 cells/mL (Griebler and Lueders, 2009). High-frequency TCC monitoring can detect short-term microbial water quality disturbances, and elevated TCC often coincides with Escherichia coli occurrences (Besmer *et al*., 2016; Sorensen *et al*., 2018; Vucinic *et al*., 2022). Persistent intrusion of surface-derived bacteria can lead to consistently high TCC and can be used to classify long-term groundwater vulnerability to microbial contamination (Farnleitner *et al*., 2005; Sinreich *et al*., 2014). Flow cytometry measures TCC rapidly and cheaply and can simultaneously report community fingerprints such as the proportions of high- and low-nucleic-acid cells (HNA/LNA) that reflect microbial phylogeny (Proctor *et al*., 2018) and intact cell concentration (ICC) indicating cellular viability (Davey and Guyot, 2020). The European and UK water utilities are now incorporating these metrics in their laboratory-based and online monitoring framework as an early warning system for microbial water quality (Safford and Bischel, 2019; Van Nevel *et al*., 2017). Due to the growing use of flow cytometry in groundwater monitoring frameworks of England, it is becoming increasingly important to characterise the spatio-temporal variation of bacterioplankton concentrations in groundwater supply sources. Additionally, research has recommended defining monitoring reference values which can be used to detect disturbances in groundwater microbial quality (Fillinger *et al*., 2019b; Hose *et al*., 2023).

There are limited studies investigating how aquifer geology controls the spatio-temporal variation of bacterioplankton in different aquifers. Groundwater moves through different aquifers on a spectrum between slow intergranular flow through intergranular pores to rapid conduit flow in karstic systems. It is well known that these flow paths through both the unsaturated and saturated zones can physically control allochthonous bacterial migration into the aquifer by filtration (Bloomfield *et al*., 2001; Sinreich *et al*., 2014; van Driezum *et al*., 2018). However, research linking spatio-temporal variation of total bacterioplankton concentration to different aquifer geologies is restricted to a handful of studies in Central Europe. In Switzerland, highly karstified parts of an aquifer showed around 10 times higher TCC than less karstified areas or alluvial aquifers, which the authors attributed to the varied aquifer filtration capacity controlling the long-term frequency of allochthonous bacterioplankton intrusion (Farnleitner *et al.*, 2005; Sinreich and Pochon, 2023; Sinreich *et al.*, 2014). A national study in Austria observed similar patterns where higher surface connectivity of karstic aquifer was linked with both allochthonous bacterial intrusion and soil-derived humic organic matter intrusion (Harjung *et al.*, 2023). Due to the bacterioplankton migration attenuation, deeper Alpine karstic aquifers under impermeable confining strata were found to possess lower TCC than shallower unconfined karst aquifers (Sinreich and Pochon, 2023). However, others have reported no TCC difference between shallow unconfined porous and deeper confined fractured aquifers in Australia (Smith *et al.*, 2012). Temporally, the extent of filtration capacity can control the TCC response to groundwater recharge. In karst aquifers, groundwater recharge can trigger increases in TCC (Sorensen *et al.*, 2018) as well as %HNA and %ICC (Vucinic *et al.*, 2022), potentially because of allochthonous bacteria arriving from the surface, given the association with E. coli detections. In alluvial porous aquifers, due to filtration, the response of river water recharge decreases farther away from the point of recharge (van Driezum *et al.*, 2018). There is also evidence that groundwater recharge can dilute TCC in alluvial porous aquifers (Karwautz *et al.*, 2022). However, there is a scarcity of spatially extensive (such as Harjung *et al.* (2023)) and temporally repeated studies, which tried to characterise the spatio-temporal variation of bacterioplankton concentrations in different geologies.

Additionally, nutrient availability can control the metabolism and growth of bacterioplankton in the nutrient-poor groundwater environment. The replenishment of dissolved organic carbon (DOC) and dissolved organic matter (DOM) with groundwater recharge frequently correlates positively with TCC (Fillinger *et al.*, 2019b; Hofmann *et al.*, 2020; Reiss *et al.*, 2019; Sorensen *et al.*, 2018; Zhou *et al.*, 2012). With increasing depth from 4-450 meters, as the surface-originated organic matter and oxygen decrease sharply, the bacterial numbers were also found to decrease about 10-fold (Pedersen *et al.*, 2008). Groundwater recharge is also related to different inorganic nitrogen species replenishment, which can be associated with the proliferation of selective bacterial taxa (Liu *et al.*, 2017; Wegner *et al.*, 2019), although relationship of nitrogen with TCC is not well understood. In a study comparing the prokaryotes of volcanic and alluvial aquifers, Amalfitano *et al*. (2014) related the bacterioplankton TCC to the organic and inorganic nutrient availability, instead of differential allochthonous prokaryotic intrusion due to the filtration capacity of the aquifer matrix. Some studies have also related TCC spatial variation to land-use category due to differences in nutrient input from different land-use types (Korbel *et al*., 2013) although others find no such effects (Fillinger *et al*., 2019b).

This study is a novel investigation of how different aquifer geology controls the spatio-temporal variation of groundwater bacterioplankton concentration (TCC) and community (using HNA/LNA and ICC metrics) across a national spatial scale. This is also the first study conducted on the drinking water sources of the three major aquifers of England. Groundwater bacterioplankton TCC, HNA and ICC were compared in three distinct aquifers and during typical pre-recharge and peak-recharge seasons to understand geological controls on bacterioplankton concentration variation. The three aquifers had different geologies and hydrological characteristics and were globally important aquifers. These aquifers were the Permo-Triassic sandstone, typical of red sandstone aquifers in Europe, North Africa, the Middle East, and North America (Celle-Jeanton *et al*., 2009); the Cretaceous chalk aquifer, which is the most important freshwater-reserve in north-western Europe (Gunn *et al*., 1995); and the Jurassic limestone aquifer, which is globally one of the most important carbonate aquifers (Worthington and Ford, 2009). The study objectives were to assess the impact of aquifer geology, seasonal recharge, organic matter and nutrients on the spatio-temporal variation of TCC, HNA, and ICC in the major aquifers of England.

4.2 Study area and methodology

4.2.1 Study area and aquifer characteristics

The study area comprises the three major aquifers used for public water supply in England (Figure 1), which each have different hydrogeological flow conceptualisations (intergranular, dual porosity and karstic). In the Permo-Triassic sandstone aquifer, groundwater flow through both the unsaturated and saturated zones is primarily through intergranular spaces and consequently, water movement is slow, although locally, some degree of preferential flow paths along fractures exist (Allen *et al*., 1997). Given that the majority of the flow occurs through pore spaces, it is denoted as an "intergranular aquifer" in this chapter.

The Cretaceous chalk aquifer is a white limestone consisting of a low-permeability, high-porosity matrix that is intersected by high-permeability, low-porosity horizontal fractures and vertical joints (Price, 1987; Worthington and Ford, 2009). Groundwater recharge predominantly occurs under piston pressure through the unsaturated zone matrix, typically at a slow (~1 m/year) rate, although more rapid movement of water through vertical joints can occur during notably wet periods (Maurice *et al*., 2023b; Sorensen *et al*., 2015). Groundwater movement in the saturated zone occurs mainly through the high permeability fracture network, which can be dissolutionally enlarged (Maurice *et al*., 2023b; Price, 1987; Worthington and Ford, 2009). The chalk was denoted as a "dual porosity aquifer" in this chapter.

The Jurassic limestone is a moderately karstified aquifer (Worthington and Ford, 2009). Groundwater recharge can occur via multiple small channels or through stream sinks, common in some areas, and rivers can experience substantial losses and gains in flow while crossing the limestone outcrop (Maurice *et al*., 2023a). Groundwater flow primarily occurs rapidly and locally through secondary dissolution features like fractures, fissures and some conduits, (Maurice *et al*., 2023a; Worthington and Ford, 2009) and therefore the aquifer is denoted as a "karstic aquifer" in this chapter.

A total of 144 raw (untreated) groundwater samples were collected from 101 different public water sources (Figure 4-1). The samples from near-continuously pumping boreholes represent the aquifer bacterioplankton communities and not the modified borehole communities (Korbel *et al*., 2017; Sorensen *et al*., 2013). From the intergranular, dual porosity and karstic sites, respectively, 29, 32 and 8 samples were collected in the pre-recharge (September-October in 2022) season, and respectively, 34, 33 and 8 were collected in the peak-recharge season (January-February in 2023).

The peak-recharge sampling was planned ahead of the recharge season, without knowledge of the antecedent conditions.

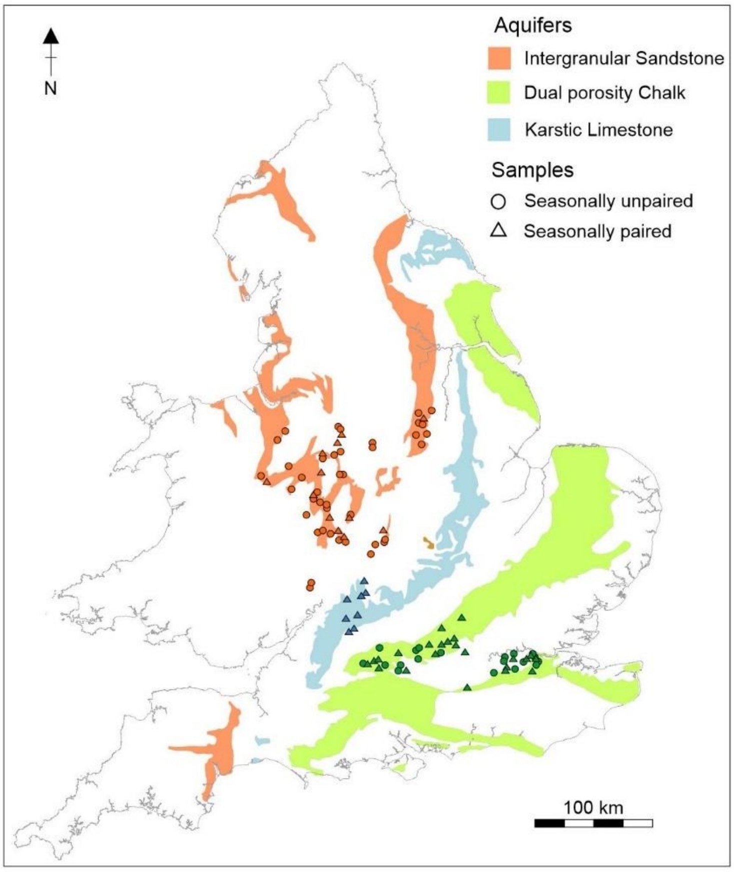


Figure 4-1. Outcrop of the three sampled aquifers in England and Wales displaying sampled boreholes with shapes indicating sites sampled in one (seasonally unpaired) or both seasons (seasonally paired). (Boreholes location retrieved from GeoIndex data centre's (NGDC) scanned borehole collection BGS © UKRI (2023), map outline contains OS data © Crown copyright and database rights 2025)

Every effort was made to collect samples evenly spread across areas of the aquifers utilised by the collaborating water companies and to sample identical sources during both seasons, but that was not possible due to operational constraints of the water companies. A total of 12, 23 and 8 of these samples were seasonally paired in the intergranular, dual porosity, and karstic aquifers, respectively and hereafter referred to as seasonally paired, with the remaining sites referred to as unpaired sites (Figure 4-1)

In almost every aquifer, the sites had an even spread of confined and unconfined sites, sites with rapid surface water intrusion vulnerability, except for mostly unconfined sites from the dual porosity aquifer and mostly non-vulnerable sites from the karstic aquifer (Appendix 1). Likewise, land-use patterns in the borehole source protection zone-1 (SPZ1) of every aquifer were mixed, and no aquifer was associated with any particular land-use category (Appendix 1). Intergranular aquifer sites had deeper groundwater sources, followed by dual porosity sites, and karstic aquifers had the shallowest total borehole depth and top of the borehole perforation depth (Appendix 1). Four of the karstic sites had a spring source.

4.2.2 Sampling procedure

Sampling adhered to standard microbiological sampling protocols for the English water industry (Willis *et al.*, 2013) and was conducted by water company personnel. Sample taps were sterilised by flaming and were flushed for the recommended time at each source to produce representative groundwater samples. Samples were collected in either untreated or dosed 500 mL sterile plastic bottles, depending on the water company. The dosed bottles contained 1 mL of 18 mg/L sodium thiosulphate solution for neutralising excess chlorine. Analysis of blank bottles showed the dosed bottles had no bacterial cells or fluorescence signature (Appendix 2). All sample bottles were transported at 3-7°C for under 8 hours following collection, then transferred to 4°C storage. Flow cytometry and fluorescence/absorbance analysis were typically conducted within 24 hours of sample collection, but eight samples arrived later and were analysed within 48 hours.

4.2.3 Flow Cytometry

Bacterioplankton enumeration was performed using an Attune™ CytPix™ Flow Cytometer (Thermo Fisher Scientific, USA) following a modified version of the protocol described in Prest *et al.* (2013) to enumerate total cell count (TCC), and fingerprinting of high/low nucleic acid content (HNA/LNA), and intact cells (ICC). For TCC and HNA/LNA, water samples (200 µL) were stained with 2 µL of 100X SYBR Green I stain (diluted with molecular grade water from 10000X stock). To determine ICC, water samples (200 μL) were stained with 2.4 μL of SGPI stain that was prepared by combining 100X SYBR Green I and 1 mg/mL Propidium iodide in a 1:5 ratio. Both sets of samples were incubated in the dark at room temperature (22ºC) for 10 minutes before analysis (incubation condition optimised in Section 3.1). All samples were prepared and analysed in triplicate. For each sample replicate, 50 μL of stained volume was measured at a flow rate of 100 μL/min. Ultrapure water was used as blank samples at the beginning of each plate.

Using the Attune Cytpix built-in software, TCC and HNA/LNA were determined on a density plot of side-scatter light (SSC) versus green fluorescence (BL1) channel intensities. ICC was measured using a density plot of BL1 versus red fluorescence (BL3) channel intensities. A set of manually pre-made gates was used to differentiate and enumerate bacterioplankton cells (Figure 3-2). Each scatter plot was manually examined for noise and errors, and a total of 15 TCC replicates and 34 ICC replicates from different samples, along with 2 ICC replicates of one sample were removed for being erroneous (background noise within bacteria gates). For the TCC measurements, the relative standard deviation (RSD) for 95% of samples was within 12.5%, and the maximum RSD was 15%. For the ICC measurements, the RSD for 95% of the samples was within 16.6%, and the maximum RSD was 29%. The mean cell concentration of the replicates was used for onward analysis.

Since 8 samples were analysed between 24 and 48 hours, the stabilities of TCC, HNA concentration and ICC between 24 and 48 hours were checked. A subset of 9 samples from both water companies were analysed after 24 hours and then after 48 hours. Over the period, TCC values remained within an RSD of 14%, HNA concentration values remained within an RSD of 16.7%, but one sample ICC had an RSD of 21%, and the rest had RSDs under 16.7% (Appendix 3). Since the RSDs remained within the RSDs of triplicates, the old samples were not removed from the dataset.

4.2.4 Escherichia coli analysis

E. coli analysis was undertaken by the plate cultivation method to evaluate evidence for recent faecal contamination. CHROMagar™ E. coli nutrient broth was prepared by dissolving 18.65 g of the powder base in 500 mL of de-ionised water and autoclaving at 121°C for 15 minutes. The broth (20 mL) was dispensed into 90 mm Sterilin™ polystyrene Petri dishes. Each groundwater sample (100 mL) was filtered through a 0.45 µm sterile Whatman™ Polyethersulfone (PES) filter. The filter was placed aseptically in the Petri dish on the agar surface and incubated at 37°C for 24 hours in aerobic conditions. There was an absence of blue colonies in any Petri dish, confirming no *E. coli* in any sample.

4.2.5 Fluorescence Spectroscopy

Fluorescence and absorbance analysis of organic matter (OM) was performed simultaneously using an Aqualog® fluorimeter (Horiba Scientific, Japan). Unfiltered samples were analysed in a quartz cuvette of 1 cm path length after being left for an hour to equilibrate to room temperature (20°C). The scan was performed between excitation wavelengths of 240-600 nm with a 1 nm step interval and emission wavelengths from 200-800 nm with a 1.16 nm step interval. Scan integration time was 0.5 seconds. Molecular-grade water was used as a blank. The instrument correction is applied by default on the Aqualog® built-in software.

4.2.6 Dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) analysis

Groundwater samples (50 mL) were filtered through 0.45 µm Whatman™ PES filter membranes into HDPE plastic bottles. Dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) analyses were performed using an Elementar Vario Cube (Elementar Ananlysensysteme GmbH; Langenselbold, Germany). Undiluted samples were acidified using 0.05 mL of 30% HCl and sparged with 99% O2 at 850°C for C and N combustion. The resulting CO2 gas was detected by a precision gas analyser, and an electrochemical detector detected NO.

4.2.7 Data analysis

4.2.7.1 Fluorescence and absorbance data

The blank corrected excitation-emission matrix (EEM) scans and absorbance files were exported for PARAFAC analysis (Stedmon and Bro, 2008), to identify fluorescent organic matter peaks (fOM), using the staRdom package (Pucher *et al.*, 2019) in R v.4.3.2. The excitation-emission matrices were cut at excitation/emissions wavelengths of 240-400/250-550 nm, before correction for the inner filter effect using the absorbance data, removal of the Raman and Rayleigh scatter lines, and normalisation into Raman standard Unit (RSU) using the blank sample scans done before sample analysis each day. Thereafter, the fluorescence indices (Fluorescence Index (FI), Biological Index (BIX) and Humification Index (HIX)) (Gabor *et al.*, 2014) were quantified. Followed by this, a PARAFAC model was run with four components, 10000 iterations, 40 random starts, 10-8 tolerance levels, and non-negative constraints. Excluding eight high-leverage points detected during the model performance check, the final model was run with 136 samples. The final model had a 96.8% fit, with a Tucker's congruency value of 0.98 and core consistency value of 51.3%, demonstrating the model was valid and not over-fitted. The fOM components from peaks 1 to 4 corresponded with Coble-peaks peak-C, peak-M, peak-T and peak-B, respectively (Coble, 1996) (Appendix 4). The first two components were humic-like, showed a strong correlation ($r = 0.95$, $p<0.001$), and were aggregated to represent a single humic-like fluorescent component (HLF). Component 3 was consistent with tryptophan-like fluorescent OM (TLF), and component 4 was consistent with tyrosine-like fluorescent OM (TyLF) (Coble, 1996; Gabor *et al.*, 2014). All the fOM loadings were expressed in Raman standard units (RSU).

4.2.7.2 Statistical analysis

Statistical analysis was performed using R v.4.3.2, and a p-value <0.05 was considered significant. Non-parametric statistical tests were used following examination of QQ-normal plots and Shapiro-Wilks tests ($p<0.05$ threshold for non-normal distribution) of variables (Shapiro and Wilk, 1965) (Appendix 5), which confirmed their non-normal distribution. To test if there was statistically significant differences in the three bacterioplankton concentration variables (TCC, %HNA and %ICC), and organic and inorganic chemicals (fOMs, DOC, TDN) among the three aquifers, Kruskal-Wallis rank sum test (Kruskal and Wallis, 1952), and post-hoc Dunn's multiple comparison tests (Dunn, 1964) were performed and results were reported with $\chi 2$ and Z-values, respectively. To test the seasonal differences of the variables in each aquifer, first, the statistical significance of differences of medians between seasons was tested using the Wilcoxon rank-sum exact test on all samples from each season (paired and unpaired), and the results were reported with W-value (Wilcoxon, 1992). Next, seasonal change at individual sites was assessed using the Wilcoxon signed rank exact test for seasonally paired samples and test results were reported with a V-value (Wilcoxon, 1992). Boxplots were prepared for visualisation with box hinges representing the interquartile range (IQR) and the median, the whiskers representing points up to 1.5 times the IQR and any point beyond that is deemed to be shown as an outlier (McGill *et al.*, 1978). To correlate bacterioplankton TCC with fOM, DOC and TDN at each aquifer during any particular season, Spearman's correlation test was performed on unpaired samples from each aquifer and season category only when n>10 (Bonett and Wright, 2000). Significance thresholds for all analyses were $p<0.001$ for strong significance, $p<0.01$ for moderate significance and $p<0.05$ for weak significance.

4.3 Results

4.3.1 Bacterioplankton concentration variables across aquifer types

Flow cytometric analysis showed that bacterioplankton total cell concentration (TCC) was significantly different between the three aquifer types (Kruskal-Wallis $\chi 2 = 11.15$, $p<0.01$). Post-hoc Dunn's pairwise comparison showed that the median TCC in the karstic aquifer was significantly higher than in the intergranular (Dunn's Z=3.09, $p<0.01$) and the dual porosity (Z=3.23, $p<0.01$) aquifers (Figure 4- 2. A). Specifically, the median TCC in the karstic aquifer (1.9×10^4 cells/mL) was approximately twice that of the intergranular (1×10^4 cells/mL) and dual porosity aquifer (8.7×10^3 cells/mL) (Appendix 6).

The %HNA of bacterioplankton cells also differed significantly between the three aquifer types ($\chi2 = 23.87$, $p<0.001$). Dunn's test revealed significantly higher median %HNA in intergranular aquifer samples (20%) compared to both the karstic (16%, $Z = 2.15$, $p<0.05$) and dual porosity aquifers (15%, $Z = 4.84$, $p<0.001$) (Figure 4-2. B, Appendix 6). Overall, the bacterioplankton population predominantly comprised low nucleic acid bacteria (LNA) across all aquifer types (median=80- 84%). The median %ICC of bacterioplankton cells was 21-24% across all aquifer types with no significant differences among the medians (Figure 4-2. C, Appendix 6).

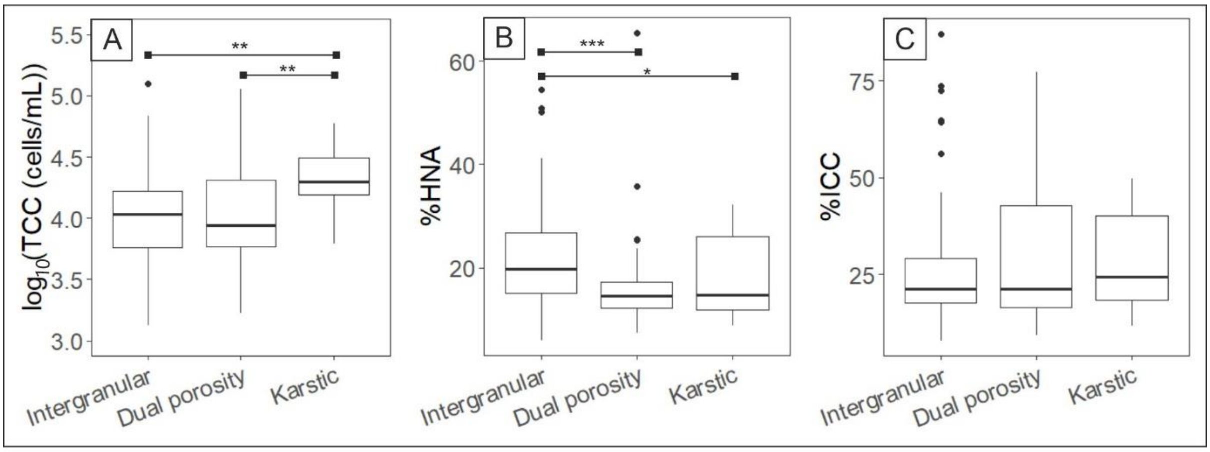


Figure 4-2. Boxplots of bacterioplankton populations in the three aquifer types showing the A) total cell concentration, B) high nuclear acid percentage (%HNA), and C) intact cell percentage (%ICC), have significant differences where significance values are *** for $p<0.001$, ** for $p<0.01$ and * for $p<0.05$.

4.3.2 Seasonal trend of bacterioplankton concentration variables across aquifer types

Only the dual porosity aquifer exhibited significant changes in median bacterioplankton TCC between pre- and peak-recharge seasons (Figure 4-3. A) (Wilcoxon rank sum test, W=686). The peak-recharge median TCC (7.1×10^3 cells/mL) was almost half that of the pre-recharge median TCC (1.3×10^4 cells/mL) (Appendix 6). This seasonal change is corroborated by a significant change in the paired data (Wilcoxon signed rank test, V=250, p<0.001) where peak-recharge TCC was lower at 87% of the dual porosity sites (Figure 4-3. D). In contrast, no significant seasonal shifts in median TCC were observed in unpaired or paired data from the intergranular or karstic aquifers.

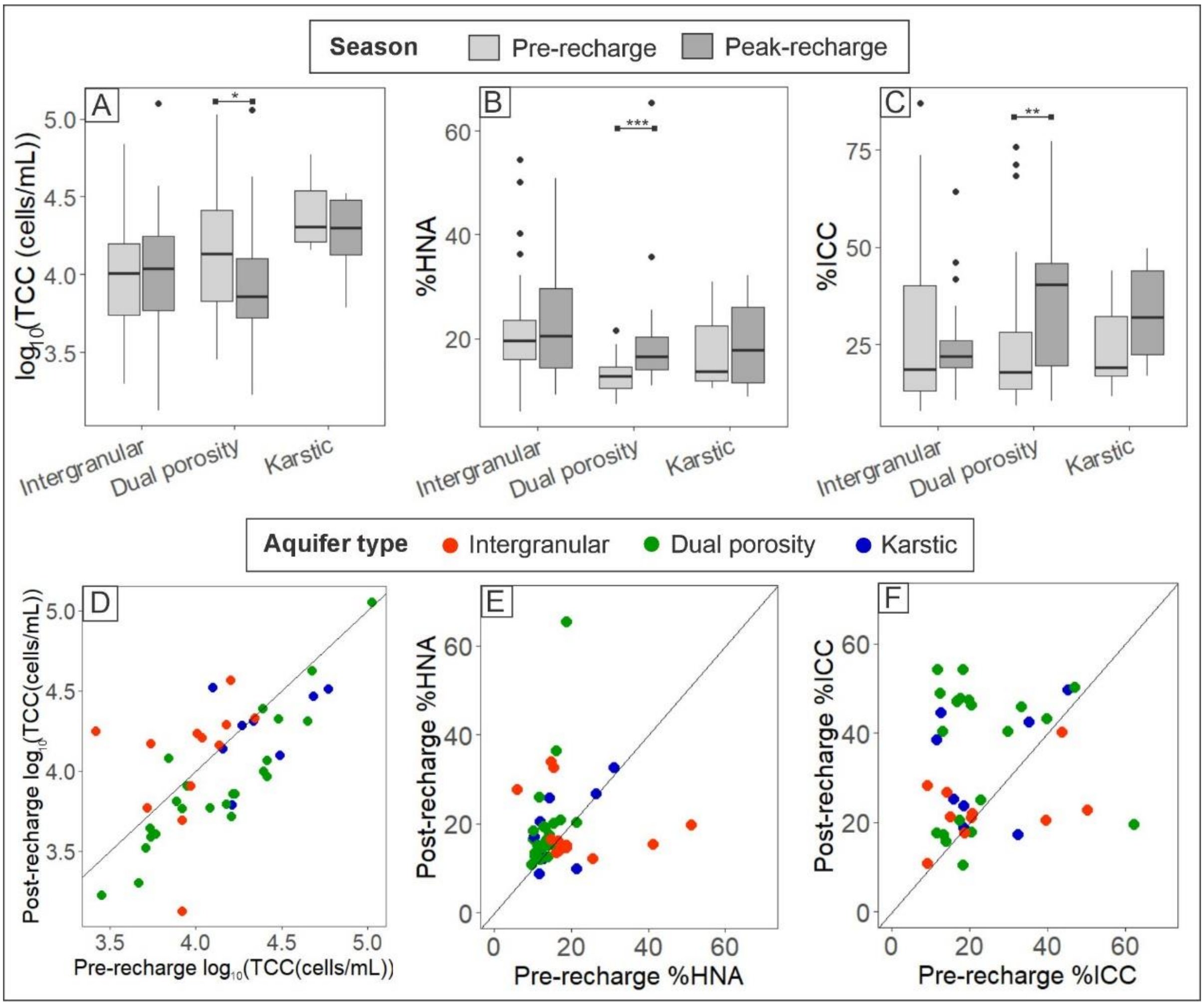


Figure 4-3. Boxplots of pre- and peak-recharge changes in bacterioplankton characteristics across all sites in the three aquifer types: A) total cell concentration, B) high nucleic acid percentage, C) intact cell percentage, showing significant differences where significance values are *** for p<0.001, ** for p<0.01 and * for p<0.05, D-F) Dot plots of bacterioplankton population abundances for all seasonally paired sites, with a 1:1 line shown.

The unpaired data indicated a significant increase in median %HNA (pre-recharge 13%, peak-recharge 16%; W=202, p<0.001) and median %ICC (pre-recharge 18%, peak-recharge 41%; W=286, p<0.01) between pre- and peak-recharge seasons in the dual porosity aquifer (Figure 4-3. B, C, Appendix 6). This finding was also supported by observations at dual porosity paired sites, where for 87% of the sites, %HNA was higher ($V=11$, $p<0.001$) and for 82% of the sites, %ICC was higher ($V=69$, $p<0.05$) in the peak-recharge season. No statistically significant changes in %HNA were observed in the intergranular or karstic aquifers (Figure 4-3. B), which was also supported by observations from the paired sites (Figure 4-3. E). Similarly, the %ICC in the intergranular aquifer remained unchanged, but in the karstic aquifer, peak-recharge %ICC (32%) was higher than pre-recharge %ICC (19%).

4.3.3 Chemical characteristics variable differences across aquifers and seasons

HLF, DOC concentrations and HIX and BIX indices were the only variables that showed statistically significant differences between the aquifer types (HLF: $\chi2 = 20.7$, $p<0.001$; DOC: $\chi2 = 8.07$, $p<0.05$) (Figure 4-4. A, Appendix 6). Among the fOMs, HLF was the most abundant organic matter compared to TLF and TyLF in all the aquifers. Median HLF in both the karstic (0.25 RSU) and the dual porosity (0.21 RSU) aquifers was significantly higher than the intergranular aquifer (0.14 RSU) ($Z = 3.7$, $p<0.001$ and 3.6, $p<0.001$ respectively). Median DOC was significantly lower in the intergranular aquifer (0.85 mg/L) than in the dual porosity aquifer (0.98 mg/L; $Z = 2.7$, $p<0.01$), but not in the karstic aquifer (0.96 mg/L). Among the fluorescence indices, HIX was significantly higher in the karstic aquifer (0.78) than in the intergranular aquifer (0.67; $Z = 2.4$, $p<0.05$), although the magnitude of the difference was minimal. BIX was significantly higher in the dual porosity aquifer (0.76) than in the intergranular aquifer (0.68; $Z=3.6$, $p<0.001$), with minimal magnitude difference. TLF, TyLF, TDN and FI did not differ significantly between the aquifer types.

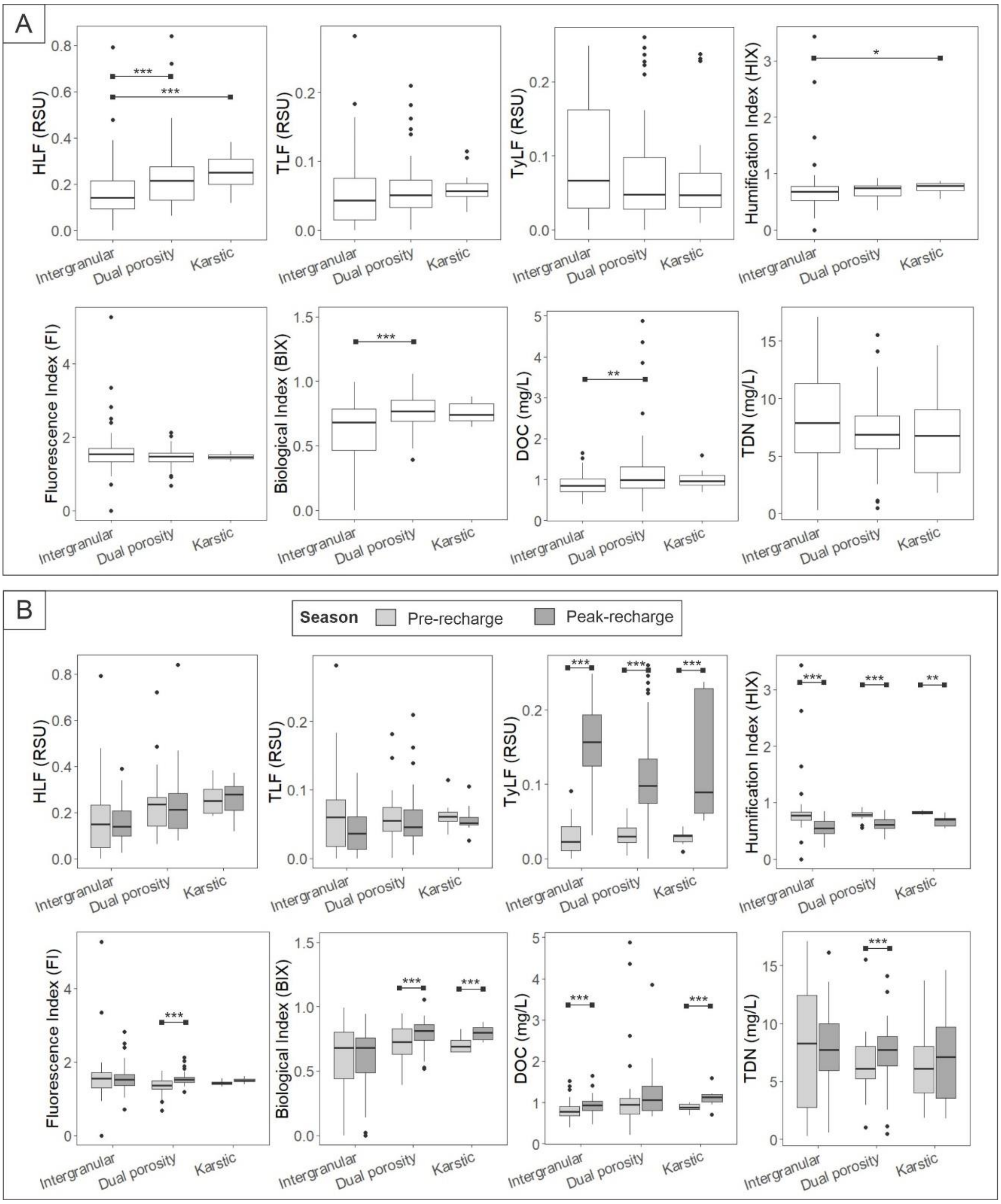


Figure 4-4. Boxplots of fluorescent organic matter loading, fluorescence indices and nutrient concentrations, across A) aquifers and B) seasons. Significance differences are denoted by *** for p<0.001, ** for p<0.01 and * for p<0.05.

Significant seasonal fluctuations were found in the concentration of DOC, TDN and tyrosine-like fluorescence along with HIX, BIX and FI indices in both the unpaired and paired datasets (Figure 4-4. B; Appendix 6 and 7). Median DOC concentrations increased in all the aquifers, though significantly in the intergranular (0.77 mg/L to 0.93 mg/L; W = 344, $p<0.05$) and the karstic aquifers (0.87 mg/L to 1.1 mg/L; W= 9.5, $p<0.05$), but not the dual porosity aquifer (0.94 mg/L to 1.05 mg/L). Significant DOC increases were also reflected in the paired dataset (Intergranular: V=1, $p<0.01$; Karstic: V=2, $p<0.05$). A significant seasonal increase of median TDN was restricted to the dual porosity aquifer (6.1 mg/L to 7.7 mg/L) and was observed in both the unpaired (W = 375, $p<0.05$) and paired data (V = 2, $p<0.001$). Median TLF decreased between 20% in the karstic and 50% in the intergranular aquifers following recharge, but these trends were insignificant, except for the paired dual porosity aquifer, where significant TLF reduction was observed in 87% of the paired sites (V=252, $p<0.001$) and the median TLF of the unpaired sites reduced from 0.06 RSU to 0.04 RSU, which, however, was statistically not significant. Tyrosine-like fluorescence showed a significant increase in the peak-recharge season across all aquifers, with the median doubling from 0.06 to 0.13 RSU, along with a notable rise across all the intergranular and karstic paired sites and 78% of the paired dual porosity sites. No significant changes in HLF were observed in any aquifer type. Among the fluorescence indices, HIX dropped by 0.2 in all three aquifers. BIX increased in both the dual porosity (0.72 to 0.81) and the karstic aquifer (0.7 to 0.8). FI significantly increased from 1.35 to 1.51 only in the dual porosity aquifer.

4.3.3 Correlation of bacterioplankton TCC and chemical concentrations

HLF, TLF and DOC were typically positively correlated with TCC in the dual porosity and intergranular aquifers in both seasons (Figure 4-5). The HLF-TCC and TLF-TCC relationships were stronger in the dual porosity aquifer (mean $\rho=0.61$) than in the intergranular aquifer (mean $\rho=0.45$). The only non-significant correlation (DOC-TCC, pre-recharge, dual porosity aquifer) becomes positive and significant ($\rho=0.6$) when the four outliers (DOC > 2.5 mg/L) are excluded. The HLF and TLF-TCC relationships were stronger than the DOC-TCC on all occasions in the dual porosity aquifer. There were no consistent trends in HLF-, TLF- and DOC-TCC correlation coefficients seasonally. TDN did not show any significant correlation with TCC on any occasion.

Additionally, TyLF was significantly negatively correlated with TCC in the intergranular aquifer in both seasons.

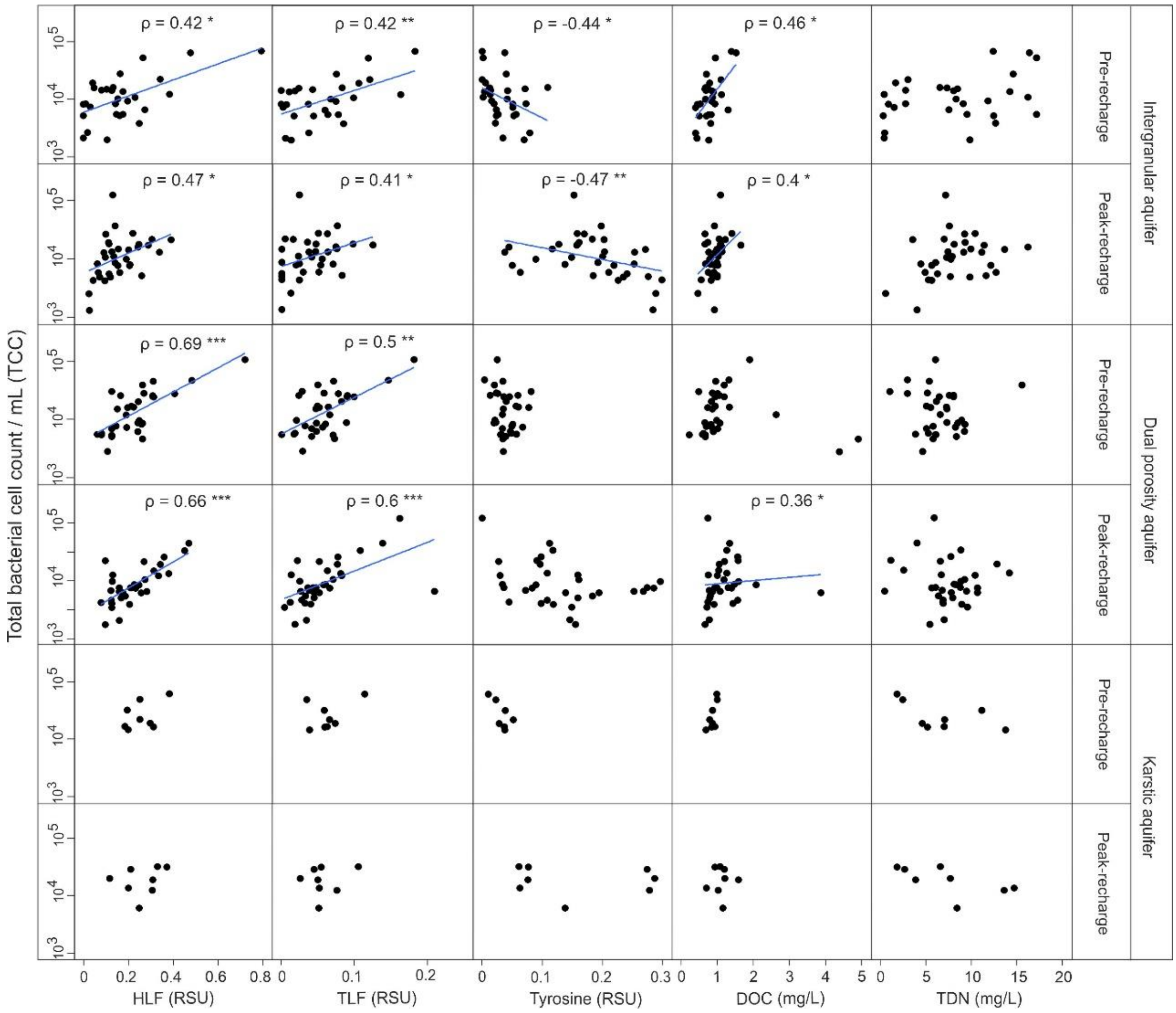


Figure 4-5. Correlation plots of bacterioplankton TCC with nutrient and fluorescent organic matters in different aquifer types in pre- and peak-recharge seasons. Only significant ($p<0.05$) Spearman's correlation coefficients are shown where n>10.

There were significant mutual correlations amongst HLF, TLF and DOC in the intergranular and dual porosity aquifers, which showed aquifer-wise and seasonally between pre- and peak recharge periods variability (Appendix 8). HLF and TLF were always significantly correlated (ρ=0.47-0.92), yet HLF was more related to DOC (mean ρ=0.60) than TLF was (mean ρ=0.38). HLF-DOC and TLF-DOC relationships were stronger in the intergranular than in the dual porosity aquifer. HLF-DOC relationships were weaker in peak-recharge (Intergranular: ρ=0.7, Dual porosity non-significant), becoming non-significant for TLF-DOC.

4.4 Discussion

4.4.1 Differences in bacterioplanktons are related to aquifer geology

Median bacterioplankton TCC in the karstic aquifer was approximately twice that of the intergranular and dual porosity aquifers, both of which had similar TCC (Figure 4-2. A). This difference could arise from chemical and physical differences in the three aquifers. Almost an even number of samples were collected from different aquifer confinement categories, catchment land use patterns and rapid surface water recharge potentials (Appendix 1). Thus, the TCC differences in the three aquifers could not be explained by either of the aquifers being covered by a particular land-use type, or being a confined aquifer, or all sites being impacted regularly by rapid recharge from the surface.

Bacterioplankton in situ growth in the groundwater can be facilitated by the presence of organic carbon (Hofmann *et al*., 2020; Reiss *et al*., 2019; Zhou *et al*., 2012). However, some studies claim that the bioavailability of organic matter is a more important facilitator of bacterial proliferation than the total organic carbon concentration (Cooper *et al*., 2016). Our study also reflected that despite higher DOC concentration in dual porosity aquifer (Figure 4-4. A), the TCC was not the highest (Figure 4-2. A), indicating DOC concentration was not the main reason driving TCC difference among the aquifers. The typically bioavailable protein-like TLF (Coble, 1996) concentration in all the aquifers was minimal, whereas the typically non-labile humic-like HLF (Coble, 1996) was the dominant organic matter fraction (Figure 4-4. A). The concentration of HLF was highest in the karstic aquifer, followed by the dual porosity and then the intergranular aquifer (Figure 4-4. A), due to more soil-derived fOM in karstic than dual porosity and intergranular aquifers. This was further supported by the higher HIX in karstic and dual porosity aquifers (Figure 4-4. A), which indicates complex microbially reworked and potentially allochthonous fOM (Serène *et al*., 2025).

Finally, the stronger TCC-HLF correlation in all the aquifers in every season indicated that both organic matter and bacterial cells could have originated in the soil layer and were transported together into the groundwater (Harjung *et al*., 2023).

An allochthonous bacterioplankton origin can explain the observed difference in TCC between the three aquifers. The intergranular aquifer exhibits slow groundwater recharge through tight pore throats with a median aperture of 0.1-90 μm, which is sufficiently small to filter out some large surface bacteria (bacteria size range 0.1-10 μm) allowing a diminished fraction of bacterioplankton intrusion into the groundwater from the surface during recharge (Bloomfield *et al*., 2001) (Figure 4-6. A). Due to filtration, bacterioplankton intrusion will be diminished in the deeper zones of the aquifer (Fiedler *et al*., 2018; van Driezum *et al*., 2018), which is reflected in the negative correlation between TCC and borehole perforation depth (top of borehole perforation depth) (ρ = -0.4). In case of the dual porosity chalk aquifer, water trapped in the very narrow pore-throats (median 0.5-1 μm) of the unsaturated zone matrix is displaced downward toward the saturated zone under piston pressure from subsequent recharge events (Maurice *et al*., 2023b; Price, 1987). The filtration capacity of narrow pore-throats may exert an even stronger filtration on the allochthonous bacterioplankton intrusion than the intergranular aquifer, allowing only a small fraction of bacteria to enter the aquifer (Gooddy *et al*., 2001; Gooddy, 2002) (Figure 4-6. B). Conversely, the limestone karstic aquifer has relatively lower filtration capacity than the other two aquifer types, allowing a larger proportion of surface bacterioplankton to enter groundwater via networks of dissolutionally enlarged fractures with larger apertures (0.5-2 cm) over long timescales (Maurice *et al*., 2023b). Additionally, the high groundwater flow velocity within the karstic aquifer may scour biofilms and resuspend the bacteria, adding to the planktonic load (Savio *et al*., 2018) (Figure 4-6. C). Over a long period of time, easier intrusion of surface bacteria and detachment of biofilm could explain the characteristic higher TCC of the karstic aquifer.

The TCC difference in dual porosity and karstic aquifer fits very well with the karst vulnerability classification presented by Sinreich *et al*. (2014) in Swiss Alpine karst springs. In non-vulnerable, less karstified groundwater, TCC was in the order of 10^3

cells/mL, a value seen in our dual porosity chalk samples. Moderately vulnerable and moderately karstic springs showed a median TCC of 1.4×10^4 cells/mL during microbially uncontaminated periods, which matches the value of our karstic limestone sample. Thus, the TCC in English carbonate aquifers reflected the degree of karstification and, thus, the filtration capacity towards allochthonous bacterioplankton intrusion. The TCC values also reflected that the karstic limestone aquifer is more vulnerable towards pathogen contamination than a dual porosity aquifer. The TCC of uncontaminated sandstone in other parts of the world was found to be within a wide range of 10^3 to 10^6 cells/mL, with the higher range of values common from shallow unconsolidated sandstone aquifers, underscoring the highly heterogeneous habitat of these aquifers (Amalfitano *et al.*, 2014; Hazen *et al.*, 1991; Pellizzari *et al.*, 2016).

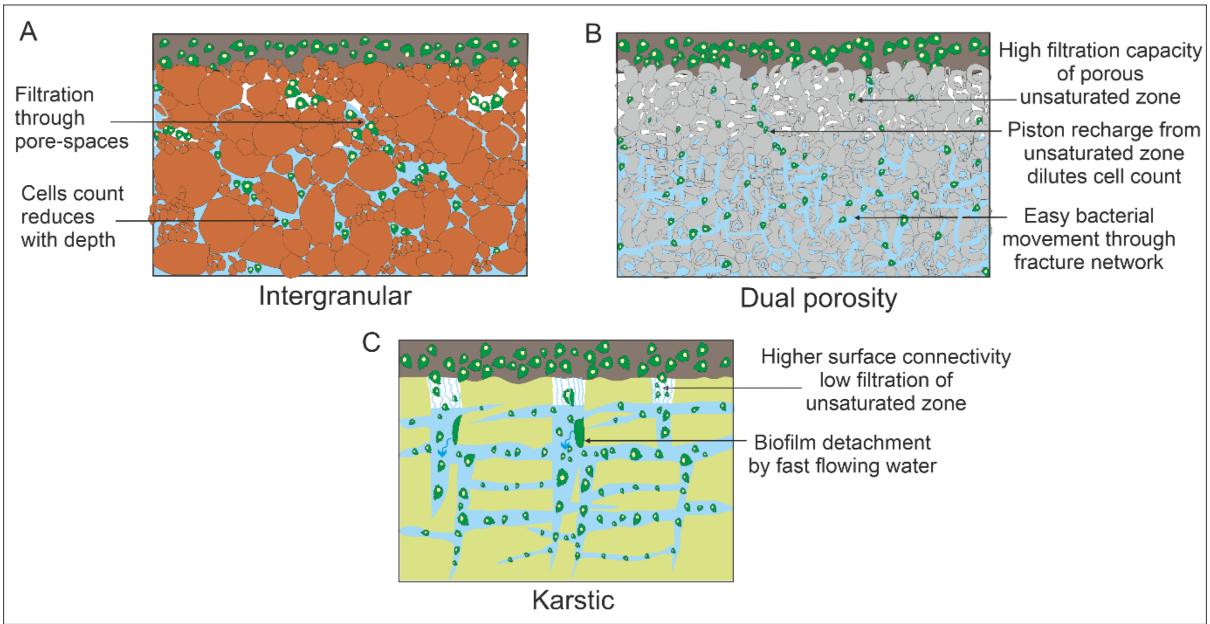


Figure 4-6. Conceptual figure of bacterial intrusion from the brown soil layer through the white unsaturated zone to the blue saturated zone of the A. intergranular aquifer where bacterial intrusion is filtered by highly heterogeneous pore-throat sizes; B. dual porosity aquifer where bacterial intrusion is filtered by pore-throat size of median 0.45 μm in the unsaturated zone; C. karstic aquifer where 0.2-5 cm fractures allow easier and unfiltered bacterial intrusion, (figure not to scale).

The interpretation of HNA and LNA bacterioplankton from flow cytometric measurements is still debatable. While some studies suggest that HNA bacteria have larger cells than LNA bacteria (Wang *et al.*, 2009), others indicate that HNA is the active bacterial proportion (Liu *et al.*, 2016). The latest interpretation is that they are phylogenetically different, i.e., HNA bacteria represent larger bacterial cells with larger nucleic acid, and the LNA bacteria represent the ultrasmall bacteria, containing reduced nucleic acid with streamlined functions (Proctor *et al*., 2018). Therefore, the higher %HNA in the intergranular aquifer sites (Figure 4-2. B) may indicate that taxonomically different bacterioplankton were present in this aquifer, as compared to those from the karstic and dual porosity aquifers. The %ICC was around 21-24% in all aquifers (Figure 4-2. C), indicating that the communities were dominated by damaged and non-viable cells (Davey and Guyot, 2020). Generally, the %ICC was found to be very high in samples with pathogen indicators (Vucinic *et al*., 2022) which were absent in samples of this study.

4.4.2 Seasonal change depended on aquifer geology and groundwater level rise

Bacterioplankton TCC reduced by almost 50% at the peak of the typical recharge season in the dual porosity aquifer, alongside increases in %HNA and %ICC, whereas in the other aquifers, there were no significant changes in bacterioplankton variables. The contrasting seasonal changes of the microbial communities between aquifers could reflect differences in recharge mechanisms and/or geographical differences in recharge during the study period.

Within the geographical extent of the dual porosity and karstic aquifers, there was more rainfall between the pre- and peak-recharge sampling periods than in the intergranular area. Within the geographical extents of the dual porosity and karstic aquifers, there was more rainfall between the pre- and peak-recharge sampling periods compared with the intergranular area (Water Situation Report, 2023). Between November and January, when groundwater levels were typically rising in the dual porosity and karstic aquifers, there was 50% and 25% more rainfall across those areas, respectively, than across the intergranular aquifer extent, where there was no consistent increase in groundwater levels (Appendix 9).

Additionally, the seasonal rise of TDN, FI and BIX in dual porosity and karstic aquifers was consistent with evidence of groundwater recharge (Wilson *et al*., 2025), but no such change was observed across the intergranular aquifer. The rise in DOC and HIX in any aquifer cannot be considered a valid proxy of recharge, as HIX relates to the organic matter complexity, which can increase by microbial reworking, and DOC can be both allochthonous or produced in situ (Shen *et al*., 2015). In short, despite the sampling being undertaken during the typical peak-recharge season, there was no consistent evidence for recharge in the intergranular aquifer, but there was clear evidence of recharge in the dual porosity and karstic aquifers.

The reduction in TCC in the dual porosity aquifer following recharge could be due to the "dilution effect", which was noted in a porous aquifer in the alpine region (Karwautz *et al*., 2022). Previously, Reiss *et al*. (2019) also found lower TCC following recharge in the English chalk aquifer. In the dual porosity aquifer, recharge typically occurs by piston displacement, where older water in the unsaturated zone is pushed down by newer recharge water arriving from the surface (Price, 1987). This water stored in the matrix of the unsaturated zone could be lower in TCC than water moving through the fractures of the saturated zone due to filtration of relatively larger bacteria (average bacterial size 0.1 to 10 μm) by the small pore throats (median 0.5 μm; Price (1987)). Hence, during piston displacement recharge, groundwater TCC can be diluted due to the arrival of low TCC water. The seasonal reduction in TLF during the recharge season is also supportive of piston displacement of recharge water from the unsaturated zone. Additionally, the increase of TDN after recharge aligns with the fact that the chalk unsaturated zone is a large reserve of nitrate (Sorensen *et al*., 2015), which can leach into the aquifer with recharge. The TCC changes were accompanied by changes in the bacterial community, evidenced through increases in %HNA and %ICC. The seasonal increase in %HNA and %ICC may reflect the selective enrichment of bacterioplankton with larger genomes capable of actively metabolising newly available nutrients, such as TDN. This trend could also indicate a distinct microbial community in the unsaturated zone matrix, with preferential migration of high-nucleic-acid, viable cells into the saturated zone. The depressed TCC in groundwater can either recover after consuming available organic matter, as was observed by Reiss *et al*. (2019) in the British chalk aquifer, or during flashy recharge, some preferential flow paths can be activated, allowing migration of allochthonous bacteria into the saturated aquifer.

Despite the observable GWL rise in the karstic aquifer, the median TCC remained seasonally unchanged, and TCC values in the paired data indicated four out of eight sites had lower peak-recharge season TCC, three sites with unchanged TCC and one site with higher peak-recharge TCC (Figure 4-3. A, D). The hydrographs of karstic borehole and spring-flow data indicated that the peak groundwater level was achieved about two months before the peak-recharge sampling period, and the groundwater levels had already started declining during the sampling period (Appendix 9). The TCC change in the karstic aquifer in response to recharge is generally short-lived (Vucinic *et al*., 2022). As a result, the effect of recharge on TCC may have subsided in the time-lag between aquifer recharge and the post-recharge sampling season. However, a sign of change in bacterioplankton population persisted in the form of higher median %ICC following recharge (Figure 4-3. C).

4.4.3 Study implications, limitations and future directions

Flow cytometry is gaining widespread acceptance as a rapid monitoring tool for microbial water quality in water supply industries (Safford and Bischel, 2019; Van Nevel *et al*., 2017). However, for the meaningful interpretation of groundwater quality disturbances using flow cytometry, establishing a baseline dataset from pristine or less contaminated groundwater sources is imperative from a regulatory standpoint (Fillinger *et al*., 2019b; Hose *et al*., 2023). The spatio-temporal variation of bacterioplankton concentrations of microbial contamination-free drinking water sources makes the study a strong basis for defining reference conditions for different aquifer types in the study area. Moreover, due to random sampling, the TCC, %HNA and %ICC values in this study were not biased to any specific part of the aquifers, but instead presented a national-scale view of each of the major aquifers of England.

Rainfall during preceding days can result in rapid increases in bacterial abundance (Vucinic *et al*., 2022), but there was no evidence of differences in rainfall immediately prior to, or during, sampling influencing the results. Rainfall did occur prior and during the pre-recharge season sampling; however, there was no evidence for shallow summer soil moisture deficits being overcome and, consequently, of groundwater recharge, including in the lower storage karstic aquifer hydrographs (Appendix 9). There was no rainfall during the anticipated peak-recharge sampling, and, prior to sampling, there had been no appreciable rainfall for 18-24 days across each aquifer extent (Appendix 9). The absence of *E. coli* in all samples also confirms that there was no evidence of rapid water movement from the surface shortly prior to sampling.

The key limitation of the reference values generated in this study is the limited number of karstic samples. But the interquartile range of TCC values in karstic aquifer was also narrower but still higher than the other aquifers. A further limitation was that, due to pre-planned sampling, it was not possible to fully capture the recharge response of bacterioplankton in all the aquifer types covered in this study. In the future, more samples from the karstic aquifers are recommended. Additionally, the seasonal changes in the bacterial abundance should be observed over multiple pre- and peak-recharge seasons (example Yan *et al*. (2021) to understand the abundance dynamics in individual aquifers. In addition to the multi-season repetition, a temporally frequent sampling at each aquifer is suggested so that the moment of recharge is not missed and the response of bacterioplankton to recharge can be adequately recorded. Currently, industrial efforts are being made to collect weekly flow cytometric data of regulatory samples. Utilising these datasets could be helpful for future studies to observe the safe range of TCC in each aquifer type, and introduce reference TCC models for individual aquifers to use as monitoring references and rapidly detect contamination. Additionally, in the future, the taxonomic assemblages of the three microbiologically distinct aquifers should be studied using an eDNA amplicon sequencing approach to test whether the community structure of the intergranular aquifer is different from the dual porosity and karstic aquifer, as was anticipated from the different %HNA value in the intergranular aquifer.

4.5 Conclusion

This study contributed to the current knowledge to better conceptualise the groundwater bacterioplankton spatio-temporal variation pattern. Bacterioplankton concentration at drinking water sources of three major English aquifers was found to vary at different aquifer geologies and was potentially controlled by filtration capacities of the aquifer towards bacterioplankton transportation from the surface. Over the long term, the karstic aquifer with higher fracture aperture allows more bacterioplankton intrusion than the intergranular and the dual porosity aquifers with smaller pore-throats and higher filtration capacities. The dominance of soil-originated HLF and robust HLF-TCC correlation corroborated the surface origin of bacteria and organic matter. Low %ICC and absence of *E. coli* in all the aquifers indicated no recent surface-bacterial intrusion event, and %HNA indicated possible phylogenetic differences in the intergranular aquifer compared to dual porosity and karstic aquifer. The only notable seasonal reduction in dual porosity aquifer was driven by TCC dilution by cell-poor unsaturated zone water entering the aquifer. Overall, this study highlights the importance of considering local hydrogeological features as well as recharge variability to understand bacterioplankton variation patterns in both space and time. Enhanced understanding of this topic and conceptualisation of controls on bacterioplankton can be expected by temporally frequent monitoring, to capture the recharge responses of the bacterioplankton communities in various hydrogeological settings. The bacterioplankton concentration values generated in the study originated from microbially uncontaminated drinking water pumping sources, making them an ideal basis for defining a national reference for the three aquifers for future groundwater monitoring.

5. Aquifer geology plays a significant role in the spatial variation of the groundwater prokaryotic communities

This chapter addresses thesis **Objective 3**: To assess the impact of aquifer geology on the spatial variation of groundwater planktonic prokaryotic community composition. A version of this chapter is under preparation for submission.

The dataset is published at the National Centre for Biotechnology and Information, titled "Amplicon sequences (16S) from samples collected from groundwater survey of UK aquifers" with BioProject accession number PRJNA1268368 (ID 1268368 - BioProject - NCBI)

5.1 Introduction

Groundwater serves as a drinking water resource for over 2 billion people worldwide (Hiscock, 2011). However, the EU Groundwater Directive also recognised groundwater as an important ecosystem and suggested managing it not only as a resource but also as a habitat (European Union, 2006). In these largely unexplored groundwater ecosystems, the prokaryotic communities, consisting of bacteria and archaea, play a vital role in the biogeochemical transformation of dissolved chemicals in groundwater and thereby provide a crucial ecosystem service by maintaining a safe groundwater quality for human consumption (Falkowski *et al.*, 2008; Griebler and Avramov, 2015). The planktonic prokaryotic communities are also spatio-temporally variable and more dynamic than their attached counterpart (Griebler and Lueders, 2009). Therefore, understanding the spatial differences of these planktonic prokaryotic communities is a necessary step towards interpreting their variable ecosystem services and effectively managing groundwater resources while also prioritising ecosystem health. Advances in flow cytometry and DNA sequencing technologies have enabled the rapid and cost-effective exploration of bacterial concentration, prokaryotic diversity, taxonomy, and functions, which are commonly used as indicators of groundwater microbial ecosystem health (Hose *et al.*, 2023; Korbel and Hose, 2011). In the future, these technologies can be incorporated into a groundwater health monitoring framework (Safford and Bischel, 2019; Watson *et al*., 2024). However, for the successful implementation of groundwater management strategies, it was suggested that different ecosystem management zones should be classified and monitoring references of each management zone should be assessed (Hose *et al*., 2023).

Although there is a clear understanding of how aquifer geology controls physical space in the subsurface and the chemical compositions of groundwater, both of which can impact the prokaryotic communities, the control of aquifer geology on the spatial variation of these communities is sparsely studied. Physical constraints, including pore-throat size and aperture size of fractures, filter the allochthonous bacteria from entering the aquifer, as well as from movement within the aquifer (Bloomfield *et al*., 2001; Taylor *et al*., 2004). Higher frequency of allochthonous bacterial migration in highly karstic alpine aquifers was found to have at least 10 times the cell concentration than less karstic aquifers with less frequent bacterial migration (Farnleitner *et al*., 2005; Sinreich and Pochon, 2023; Sinreich *et al*., 2014). The allochthonous prokaryotic input was found to permanently change the groundwater community diversity and composition over multiple years of prokaryotic intrusion with recharge and their adaptation in groundwater (Yan *et al*., 2021). But the spatial variation of allochthonous prokaryotic intrusion was found to be controlled by aquifer geology, as the allochthonous prokaryotes travelled a larger distance from the recharge point in a karstic system due to easy dispersal along large fracture apertures, and travelled a smaller distance from the recharge point in an intergranular system with a higher filtration capacity towards dispersal (Villeneuve *et al*., 2022).

The aquifer geology also controls the groundwater chemistry by rock-water interaction and by regulating the replenishment of nutrients from the surface, and the groundwater chemistry selects the prokaryotes capable of utilising available nutrients for metabolism (Falkowski *et al*., 2008). The aquifer matrix mineralogy controls groundwater chemistry through the dissolution-precipitation dynamics of the minerals (Elango and Kannan, 2007). The ionic chemistry of groundwater controlled by aquifer mineralogy was found to be an important factor influencing the spatial differences of prokaryotic communities (Abraham and Close, 2024; Amalfitano *et al.*, 2014; Zhong *et al.*, 2023). During groundwater recharge, along with allochthonous prokaryotes, nutrients such as dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) are also replenished, and the distance to which these nutrients can move before getting microbially transformed can depend on the ease of groundwater movement, controlled by matrix pore size and permeability. The DOC and TDN concentrations typically support microbial diversity by providing essential nutrients for growth and metabolism, and by selecting dominant groundwater taxa that can metabolise these nutrients (Cooper *et al.*, 2016; Danczak *et al.*, 2016b; Kumar *et al.*, 2017; Stegen *et al.*, 2016; Wegner *et al.*, 2019). Other factors, such as groundwater depth and land use, also significantly impact the spatial distribution of nutrients available for the prokaryotes. Deeper and confined aquifers tend to harbour stable, oligotrophic microbial communities adapted to low-nutrient environments in contrast to shallower and unconfined aquifers influenced by greater nutrient influx from the surface (Ben Maamar *et al.*, 2015; Pedersen *et al.*, 2008; Smith *et al.*, 2012). Within a shallow depth range of up to 30 m, the spatial variation of prokaryotic communities was also found to be influenced by the type of nutrient input from different land use practices of the recharge area (Couton *et al.*, 2023; Korbel *et al.*, 2013).

Despite the evidence suggesting that the aquifer geology can influence the spatial variation of groundwater prokaryotic communities, limited extensive regional-scale studies have investigated the differences in groundwater prokaryotic community compositions at contrasting aquifer types. Earlier regional studies have provided limited information about the impact of aquifer geology on the prokaryotes due to the use of only eDNA biomarkers, providing limited information about prokaryotic taxonomy, as well as due to the dispersal of prokaryotes at geographically proximal sites (Griebler *et al.*, 2010; Sirisena *et al.*, 2013; Stein *et al.*, 2010). In other regional studies where eDNA sequencing was used for taxonomic identification, chemical differences in the groundwater from different aquifers due to the integrated effect of surface connectivity and rock-water interactions were found to be a major driver of community differences (Abraham and Close, 2024; Couton *et al.*, 2023; Harris *et al.*,

2025) even if the aquifers were hydraulically connected (Amalfitano *et al.*, 2014). While this evidence could indicate that different aquifer geologies can be used as groundwater ecosystem management zones (Hose *et al.*, 2023), the only two known national studies have not adequately documented prokaryotes from different aquifer types. In New Zealand, Sirisena *et al.* (2018) observed no impact of aquifer lithology on the prokaryotic communities, probably due to the small number of representative samples from different aquifer types for reliable comparison. Zhong *et al.* (2023) found that the groundwater major ion chemistry had a significant impact on the prokaryotic communities, implying a potential effect of different aquifer geologies. But there is a substantial gap in current research where a national-scale assessment of groundwater prokaryotic ecosystems has looked at the differences of communities at different aquifer types.

In this research, a national-level spatial study was undertaken primarily to investigate the differences in groundwater planktonic prokaryotic communities across different aquifer geologies. The study area was in England, where groundwater from three major aquifers is the source of one-third of the public supply water. These included the Permo-Triassic sandstone aquifer, representative of red sandstone aquifers commonly found across Europe, North Africa, the Middle East, and North America (Celle-Jeanton *et al.*, 2009); the Cretaceous chalk aquifer, one of the important potable groundwater sources of northwestern Europe (Gunn *et al.*, 1995); and the Jurassic limestone aquifer, regarded as one of the most significant carbonate aquifers globally (Worthington and Ford, 2009). Despite the national importance of these aquifers in England, as well as their international relevance, the differences of groundwater prokaryotes have not been systematically explored using eDNA sequencing techniques to date, making this study novel. The prokaryotic communities were compared across different aquifers, and the impacts of other potential environmental variables that can influence community structure were also assessed to decompose the relative effects of geology and other environmental factors on the prokaryotic communities. In Chapter 4, by using the flow cytometric high and low nucleic acid cell fingerprinting, it was speculated that potentially different taxonomic compositions were present in sandstone as compared to chalk and limestone aquifer, which was tested in this chapter using the evidence from eDNA sequencing results.

5.2 Study area and methodology

5.2.1 Study area

This study focused on England's three most extensive and strategically important aquifer groups supplying water to the public and industry (Figure 5-1). The hydrogeological description of the three aquifer types is provided in Section 4.2.1. The Permo-Triassic sandstone aquifer units of Worcestershire, Shropshire, Staffordshire, west Cheshire, Nottinghamshire, and Yorkshire were collectively referred to as the sandstone aquifer in this study. Sandstone aquifer is characterised by slow groundwater movement, as flow predominantly occurs through the pore spaces within the rock matrix (Allen *et al*., 1997). The Chalk aquifers of Hampshire, the Colne and Lee catchment, the North Downs, and Yorkshire were collectively referred to as the chalk aquifer in this study. The chalk aquifer is composed of highly porous but low-permeability microcrystalline white limestone. It exhibits a complex hydrogeological system dominated by low permeability matrix-flow in the aquifer unsaturated zone and by karstic dissolution flow in the saturated zone (Maurice *et al*., 2023b; Price, 1987). Although karstic features are present, this aquifer is described as micro-karstic, as large conduits and caves are rare (Maurice *et al*., 2023b; Worthington and Ford, 2009). The Jurassic Limestone aquifers of the Cotswolds and Yorkshire were collectively referred to as the limestone aquifer in this study. The limestone aquifer is a moderately karstified system (Worthington and Ford, 2009) where groundwater flow primarily occurs along secondary dissolution features such as fractures (aperture size 0.2-5 cm), and in some areas, along small conduits (aperture size 5-30 cm) (Mathewson *et al*., 2022).

The groundwater major ion chemistry of the sandstone aquifer is highly variable. It is controlled by rock-water interaction, carbonate cement dissolution, locally mixing with older water, and it falls within a Ca-Mg-HCO_3^--SO_4^{2-} type water (Griffiths *et al*., 2002;

Shand *et al*., 2002; Smedley *et al*., 2005; Tyler-Whittle *et al*., 2002). The groundwater chemistry of the chalk aquifer is largely Ca-HCO_3^- type due to carbonate mineral dissolution, but in confined and reducing parts of the aquifer, it ranges from Ca-HCO3 to Ca-Na-HCO3-Cl type (Shand *et al*., 2003; Smedley *et al*., 2004; Stuart and Smedley, 2009). The groundwater chemistry of the limestone aquifer is Ca-HCO_3^- type due to carbonate dissolution (Bearcock *et al*., 2015; Neumann *et al*., 2003).

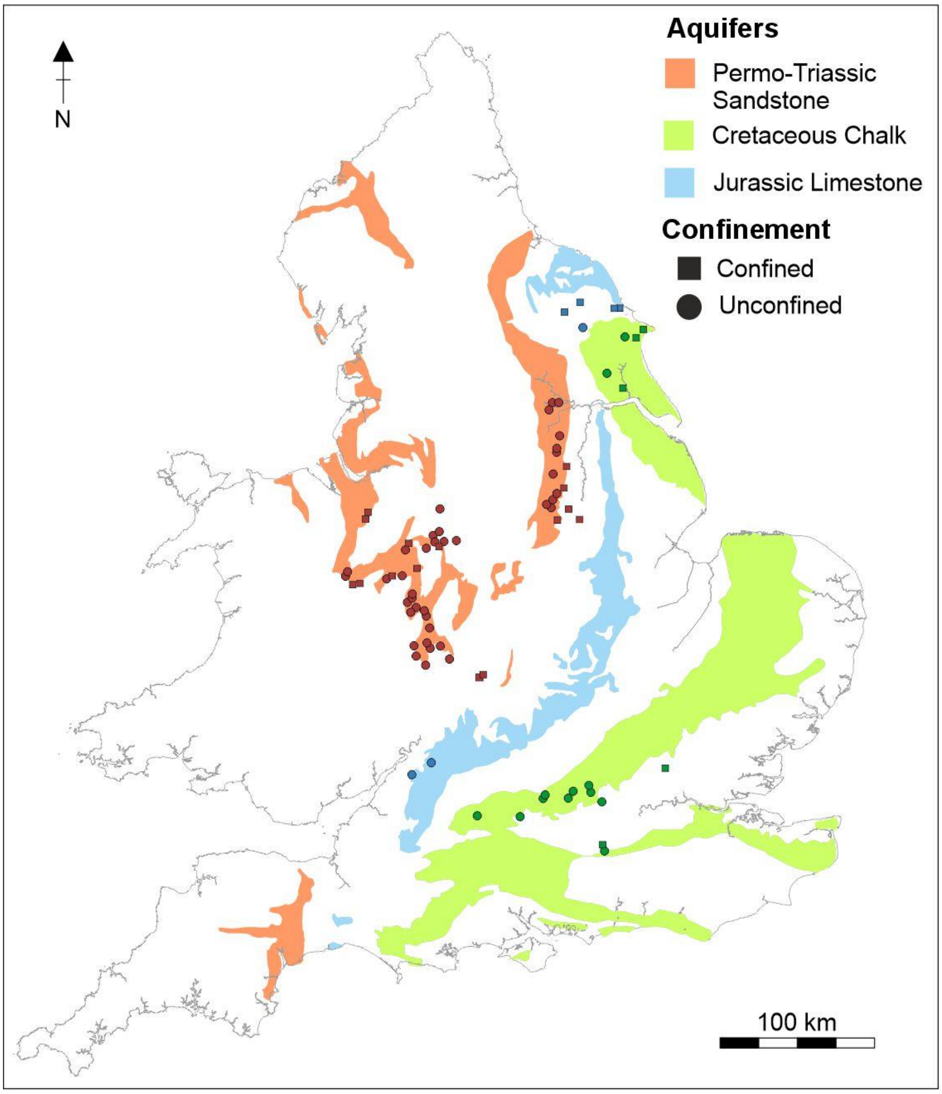


Figure 5-1. Study area in England showing the three aquifer outcrops and sample sites in each aquifer, with shapes indicating confined and unconfined sites. (Boreholes location retrieved from GeoIndex data centre's (NGDC) scanned borehole collection BGS © UKRI (2023), map outline contains OS data © Crown copyright and database rights 2025)

Groundwater samples were collected from 77 public water supply pumping stations. From Sandstone, Chalk and Limestone aquifers, 53, 17 and 7 sites were sampled, respectively. Most of the sites were unconfined, in sandstone and chalk aquifers, but 15, 5 and 4 sites from sandstone, chalk and limestone aquifers, respectively, were confined. The sandstone aquifer had deeper groundwater sources with the total borehole depth ranging from 76-391 m, the borehole perforation depth (the top of the borehole screen) ranging from 9-147 m, and the thickness of overlying strata composed of drift and/ or confining layer was ranging from 1-142m. The chalk aquifer groundwater sources were shallower, with a total borehole depth of 18-110 m, a borehole perforation depth ranging from 10-50 m, and an overlying strata thickness of 1-47m. One of the Limestone aquifer sites in Cotswold was a confined spring source with 30 m of overlying strata thickness. The other limestone aquifer groundwater sources were also shallower than the sandstone aquifer, with a total borehole depth ranging from 29-120 m, the borehole perforation depth ranging from 3-50 m, and an overlying strata thickness ranging from 0.6-61 m (Appendix 10).

5.2.2 Sample collection

Groundwater sampling was performed between December 2022 and March 2023. Raw sample taps were sterilised by applying a flame and then flushed for 10 minutes or the time specified at each sample tap to collect representative groundwater samples. The sampling rig was disinfected with a 10% chlorine solution (Instachlor™ tablets, Palintest, UK) and rinsed with sample water. Raw groundwater was filtered through 0.22 µm Sterivex™ filters with PES membranes (Merck, UK). Out of 53 sandstone samples, 47 samples and out of the 7 limestone samples, 2 samples were collected using three filters attached in parallel and the rest of the samples were collected using four filters attached in parallel. Filtration volume for sandstone samples was between 2.5 and 25 L, with two samples of unknown filtration volume. Filtration volume for chalk samples was between 4 and 20 L, with three samples of unknown filtration volume. Filtration volume for limestone samples was between 6 and 20 L, with one sample of unknown filtration volume. A filtration volume versus diversity plot was generated to ensure that filtration volume did not impact the prokaryotic communities (Appendix 12). All filters were preserved with 1 mL Zymo™ DNA/RNA Shield, transported to the laboratory on ice within 72 hours, stored at -20°C, and processed within 13 months. Raw water samples were collected in 50 mL sterile centrifuge bottles and stored at 3–7 °C for flow cytometric analysis. 50 mL of the raw groundwater samples were filtered through 0.45 µm Whatman™ PES filter membranes into clean HDPE plastic bottles for chemical analysis.

5.2.3 eDNA analysis

The filters were defrosted at 4°C, and the filter membrane and DNA/RNA shield liquid of all the filters used for each sample were extracted in a PowerWater Bead Pro Tube™. The remaining DNA extraction was performed using the Qiagen DNeasy PowerWater ™ DNA extraction kit, following the manufacturer's protocol. The final elution volume was 100 µL, with DNA concentrations ranging from 0–9 µg/mL (Qubit™ dsDNA BR assay). The V4 region of the 16S rRNA gene was amplified by polymerase chain reaction (PCR) following the PCR cycling protocol described by Newbold *et al.* (2023) in a Bio-Rad PCR C100 thermal cycler (Bio-Rad, UK), using 515F (Forward: GTGYCAGCMGCCGCGGTAA) and 806R (Reverse: GGACTACNVGGGTWTCTAAT) PCR primers (Abellan-Schneyder *et al.*, 2021). PCR products were purified using the Zymo ZR-96 DNA clean-up kit before performing a second round of indexing PCR. Final PCR products were normalised with the NGS Normalisation kit (Norgen, UK), pooled, vacuum-concentrated, and run on a 2% agarose gel. DNA bands were excised, purified using the Qiagen MinElute gel extraction kit, and quantified at 1.1 µg/mL (Qubit dsDNA HS assay). The pooled library was sequenced using the MiSeq reagent kit (v2) on the Illumina MiSeq2 platform.

5.2.4 Flow cytometry

Raw groundwater samples were analysed within 48 hours of collection and storage at 3–7 °C, and the storage time had minimal effect on bacterioplankton concentration (Appendix 3). Bacterioplankton concentration was measured using an Attune™ CytPix™ Flow Cytometer. For total cell concentration (TCC) and high/low nucleic acid

(HNA/LNA) cell fingerprinting, 200 µL of groundwater was mixed with 2 µL of 100X SYBR Green I stain and incubated in the dark at room temperature (22 °C) for 10 min. The analysis protocol and bacterial gates used were the same as mentioned in Section 3.1. The relative standard deviation (RSD) for the TCC analysis remained below 25%. %HNA was calculated as the percentage of TCC.

5.2.5 Dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) analysis

The analysis was performed on 50 mL of filtered groundwater samples and followed the same procedure as mentioned in Section 4.2.6. Dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) were analysed using an Elementar Vario Cube (Elementar Analysensysteme GmbH, Langenselbold, Germany).

5.2.6 Land use analysis

The 2021 UKCEH Land Cover Map (LCM) was accessed via the UKCEH data repository (Marston *et al.*, 2022). The Environment Agency's 2019 Source Protection Zone 1 (SPZ-1) dataset was downloaded from the open source (Environment Agency, 2019). Spatial analysis was conducted in ArcGIS Pro 3.2.2 (Esri) using the Zonal Statistics tool to determine the distribution of land-use types within each SPZ-1 surrounding the groundwater pumping stations. This involved counting 10 m^2 pixels classified as woodland, grassland, arable land and urban areas within each SPZ-1. The relative proportion of each land-use category was then calculated as a percentage of the total SPZ-1 area for each site.

5.2.7 Data analysis

5.2.7.1 Bioinformatics

All data were analysed using R v4.3. Sequence data were processed using the DADA2 pipeline (Callahan *et al.*, 2016). Low-quality reads with Q-scores below 30, forward and reverse adapter sequences, and PhiX genome matches were removed, and reads with a maximum expected error of 10 were filtered before merging the forward and reverse reads. Amplicon sequence variants (ASVs) were identified, chimeric sequences were removed, and taxonomy was assigned using the SILVA database (accessed April 2024) at the kingdom to genus levels.

A rarefaction curve of the samples was presented to show adequate sequencing depth was achieved. Sequences were rarefied to a depth of 16,277 reads (Appendix 12). retaining only bacterial and archaeal ASVs for downstream analyses. While rarefaction is the best practice (Schloss, 2024), there is a chance of underestimating diversity. Therefore, downstream analysis result of unrarefied sequences is presented in Appendix 12 to ensure consistency with rarefied data.

The clean eDNA sequences were analysed using the Microeco package in R (Liu *et al.*, 2021). The relative abundances, alpha and beta diversity were calculated using the ASV counts in each sample. The Bray-Curtis dissimilarity matrix was estimated based on the differences in relative abundances of the ASVs in each sample and Sørensen-Dice dissimilarity matrix was calculated based on presence/absence pattern of the ASVs in each sample. The relative abundance of functional potentials was calculated using FAPROTAX, which assigns known functions to specific prokaryotic genera or species with proven metabolic roles (Louca *et al.*, 2016). The donut plots presented the relative abundance of prokaryotic taxa and functional potentials, with the percentages representing the mean relative abundance of respective taxa in a particular aquifer. To test whether the Bray-Curtis dissimilarity among the samples was explained by the environmental variables (geology, confinement, groundwater depth, land use, DOC and TDN concentrations), a distance-based ReDundancy (dbRDA) analysis was performed (Legendre and Anderson, 1999). To ensure robustness of interpretation, another dbRDA model was ran on the Sørensen-Dice dissimilarity matrix. The marginal effect size of each environmental variable (explanatory power or R^2 of each variable after constraining the other terms) and the statistical significance of each explanatory power was assessed with 999 permutation test iterations using the capscale () function in the Vegan package in R. A post-hoc Mantel test of Spearman Correlations between Euclidean distance matrix of environmental variables and the Bray–Curtis dissimilarity matrix of prokaryotic communities was performed to test the association of differences in the environmental variables and the differences in the communities (Smouse *et al*., 1986). To test which taxa at the class and family levels, and which functional potentials had significantly different abundance among the aquifers, a LEfSe analysis (Linear discriminant analysis Effect Size) was performed (Segata *et al*., 2011), and box plots presented the results.

5.2.7.2 Statistical analysis

The flow cytometric variables (TCC and %HNA), DOC and TDN concentrations, were non parametrically distributed (Shapiro-Wilk test $p<0.05$) (Shapiro and Wilk, 1965). These variables and prokaryotic Shannon diversity indices in the three aquifer types were visualised using boxplots, with the box hinges representing the interquartile range (IQR) and the median, and the whiskers representing 1.5 times the IQR, with any data point beyond this range considered an outlier (McGill *et al*., 1978). To test if the differences of these variables were statistically significantly different in the three aquifers, the Kruskal-Wallis rank sum test (Kruskal and Wallis, 1952) was performed, and results were presented by χ^2 value. For the variables with significant differences among aquifers, post-hoc Dunn's multiple comparison tests were performed to test for the pair-wise differences and results were presented with Z-value (Dunn, 1964). Spearman correlation tests were performed to test whether TCC and Shannon diversity indices were related to any environmental variables. For all statistical tests, the significance thresholds were $p<0.05$ for weak significance, $p<0.01$ for moderate significance and $p<0.001$ for strong significance levels.

5.3 Results

5.3.1 Relationship of bacterioplankton concentration and environmental variables

The bacterioplankton total cell concentration (TCC) was significantly ($\chi^2 = 6.3$, $p<0.05$) different across the three aquifer types. Dunn's pairwise test indicated significantly higher TCC in the limestone aquifer than in the chalk ($Z=2.5$, $p<0.05$) and the sandstone (Z=1.9, $p<0.05$) aquifers. The median TCC value of the limestone (2.3×10^4 cells/mL) was almost 2 times higher than that of the sandstone (1×10^4 cells/mL) and chalk (8.7×10^3 cells/mL) aquifers (Figure 5-2. A). The percentage of high nucleic acid cells (%HNA) significantly varied across the three aquifer types ($\chi^2 = 4.5$, $p<0.05$). The %HNA was significantly higher in the sandstone aquifer (median = 25.3%) than that in the chalk (Z=2, $p<0.05$; median = 12.4%) and limestone aquifers (Z=1, $p<0.05$; median = 14.5%) (Figure 5-2. B).

The DOC concentration varied significantly across the three aquifers ($\chi^2 = 15.4$, $p<0.001$) and was highest in the chalk aquifer (median 1.3 mg/L), lower in the limestone (0.73 mg/L), and lowest in the sandstone (0.71 mg/L) aquifer (Figure 5-2. C). TDN concentrations did not vary significantly across the aquifers, but the sandstone aquifer had a larger range of TDN concentrations (IQR=3.3-10.8 mg/L) than the chalk (IQR=6.5-10 mg/L) and limestone (IQR=4.1-8.6 mg/L) aquifers (Figure 5-2. D). DOC concentration and TCC showed significant positive correlations. The sandstone aquifer showed a weak positive DOC-TCC correlation ($\rho=0.36$, $p<0.05$), the chalk aquifer had a strong positive correlation ($\rho=0.71$, $p<0.01$) (Figure 5-2. E).

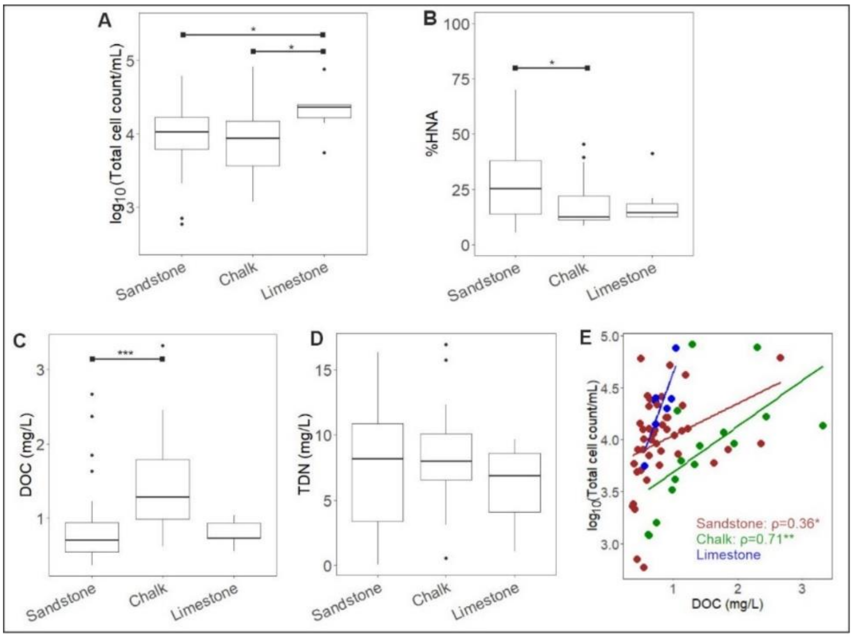


Figure 5-2. Boxplots representing differences in A. bacterial total cell concentration (TCC), B. bacterial %HNA cells, C. Dissolved organic carbon (DOC) concentration, D. Total dissolved nitrogen (TDN) concentration, and E. Correlation plot between TCC and DOC with lines

showing linear regression and ρ-values representing Spearman correlation coefficients. All plots are divided by the three aquifer types, and the significance levels in the Dunn's test and Spearman's correlation test were *** $p<0.001$, ** $p<0.01$, and * $p<0.05$.

5.3.2 Relationship of prokaryotic community structure and environmental variables

A total of 39,032 unique ASVs were identified and matched to the SILVA database for taxonomic assignment. A rarefaction curve showed that sufficient sequencing depth was achieved to cover most of the taxa (Appendix 12). Additionally, the sample filtration volume did not impact the species richness and Shannon diversity (Appendix 12). The Shannon diversity indices were significantly different among the three aquifers (χ^2 = 34.9). Sandstone aquifer had significantly lower Shannon diversity index (median=5.4) than the chalk (Z=5, $p<0.001$; median=6.5) and limestone aquifers (Z=3.8, $p<0.001$; median=6.5) (Figure 5-3. A). The Shannon diversity index had a significant negative correlation with borehole depth (ρ= -0.53, $p<0.001$) and a significant positive correlation with DOC concentration (ρ= 0.54, $p<0.001$).

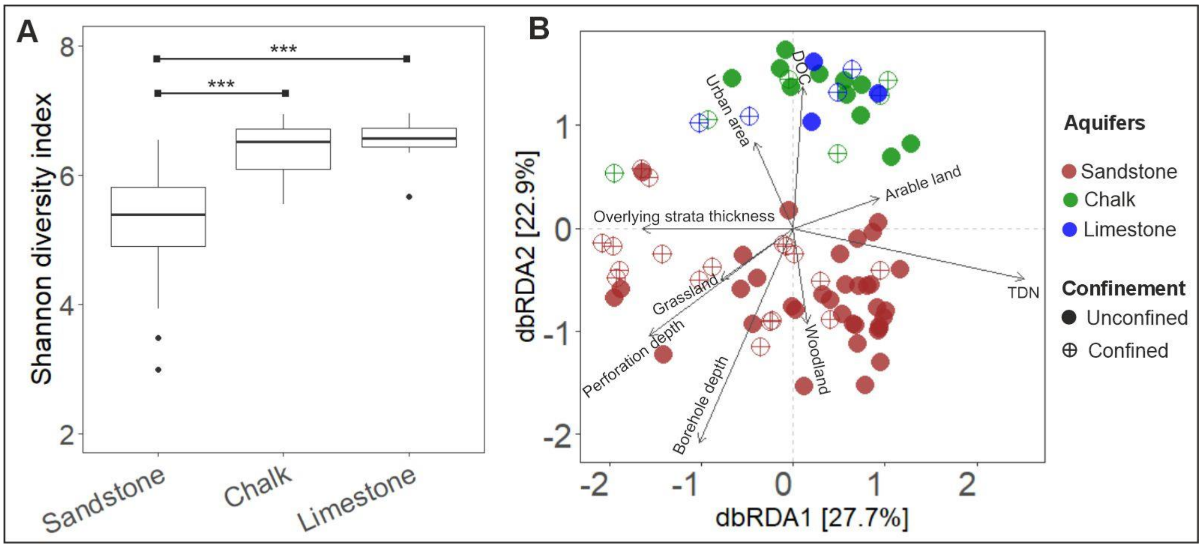


Figure 5-3. A. Differences in Shannon diversity index among the three aquifers based on prokaryotic ASVs; B. dbRDA biplot of Bray-Curtis distances of ASVs and arrows showing loading of environmental variables along the two dbRDA axes. Sites were coloured according to aquifer type and shaped by aquifer confinement.

	dbRDA1	dbRDA2	Capscale- R^2 (Marginal effect)	Mantel correlation coefficient (ρ-value)
Eigen values (constrained variables only)	1.9	1.7		
Aquifer			5.5% ***	
Confinement			1.3%	
Borehole Depth	-0.33	-0.94	1.6% *	0.22 **
Perforation depth	-0.75	-0.65	1.2%	0.27 **
Overlying strata thickness	-0.99	-0.06	1.7% *	0.34 **
DOC	0.1	0.99	1.6% *	0.24 **
TDN	0.98	-0.15	4.4% ***	0.32 **
Woodlands in SPZ1	0.25	-0.96	1.4%	-0.03
Grassland in SPZ1	-0.8	-0.6	1.4%	-0.06
Arable land in SPZ1	0.92	0.4	1.4%	0.07
Urban area in SPZ1	-0.47	0.87	1.3%	0.01

Table 5-1. explanatory power (R^2) of each variable and mantel test ρ values of Spearman correlation between Bray-Curtis dissimilarity and Euclidean distance of environmental variables across all the aquifers, significance levels were *** $p<0.001$, **$p<0.01$ and * $p<0.05$.

The dbRDA analysis indicated that environmental parameters such as aquifer geology, confinement, borehole depth, perforation depth, overlying strata thickness, DOC concentration, TDN concentration and the coverage of four land-use categories in the SPZ1, constrained a total of 26.3% of the Bray-Curtis dissimilarity. The dbRDA model was statistically significant ($F=1.9$, $p<0.001$). The permutation test with marginal effects showed that aquifer geology had the highest and most significant ($p<0.001$) explanatory power ($R^2=5.5$), followed by TDN concentration ($R^2=4.4$). The DOC concentration ($R^2=1.6$), borehole depth ($R^2=1.6$), and the overlying strata thickness ($R^2=1.7$) had lower and weakly significant ($p<0.05$) explanatory powers (Figure 5-3. B; Table 5-1). The dbRDA model based on the Sørensen-Dice dissimilarity matrix showed that 24.9% of the community dissimilarity was constrained by the environmental variables, and the strongest explanatory powers were of the aquifer geology ($R^2=8.2\%$) and TDN concentrations ($R^2=4.1\%$) (Appendix 14).

The Mantel correlation test of Bray-Curtis dissimilarity and Euclidean distance of environmental variables (Table 5-1) showed that community difference had a moderate but significant correlation with the differences in TDN concentration (ρ=0.32) and overlying strata thickness (ρ=0.34), and a weak but significant correlation with DOC concentration (ρ=0.24), total borehole depth (ρ=0.22) and perforation depth (ρ=0.27). Additionally, a Distance-decay analysis showed an even weaker correlation with geographic distance among the sites (ρ=0.18, $p<0.05$) (Appendix 15). The chalk and limestone aquifer prokaryotes were associated with higher DOC concentration and shallower borehole and perforation depth. In contrast, the sandstone communities were associated with lower DOC and deeper groundwater sources (Figure 5-3. B). The effects of aquifer confinement, perforation depth, and coverage of land use categories did not have significant explanatory power on the community variance. The dbRDA biplot showed that only in the sandstone aquifer, the confined sites had apparently different communities than the unconfined sites and were associated with low TDN concentration and thicker overlying strata (Figure 5-3. B). However, the model parameters suggested no significant effect of confinement across all the aquifer types. These results were similar in unrarefied sequences (Appendix 12).

5.3.3 Prokaryotic taxonomic composition of three aquifers

Upon matching the 39,032 unique ASVs to the SILVA database for taxonomic assignment, it was observed that taxonomic assignment became poorer moving down the taxonomic ranks. The aquifers had higher proportion of ASVs belonging to bacteria kingdom (mean=81%) than ASVs belonging to archaeal kingdom (mean=19%). 76.8% of the ASVs were assigned to 75 unique phyla, and 71.6% of the ASVs were classified at the class level. In contrast, only 55.7% of ASVs were assigned at the order level, only 27.6% at the family level and only 16.5% at the genus level. Among the 385 dominant ASVs with relative abundance greater than 1% in at least one sample, 22 bacterial ASVs were unclassified even at the phylum level (Appendix 13).

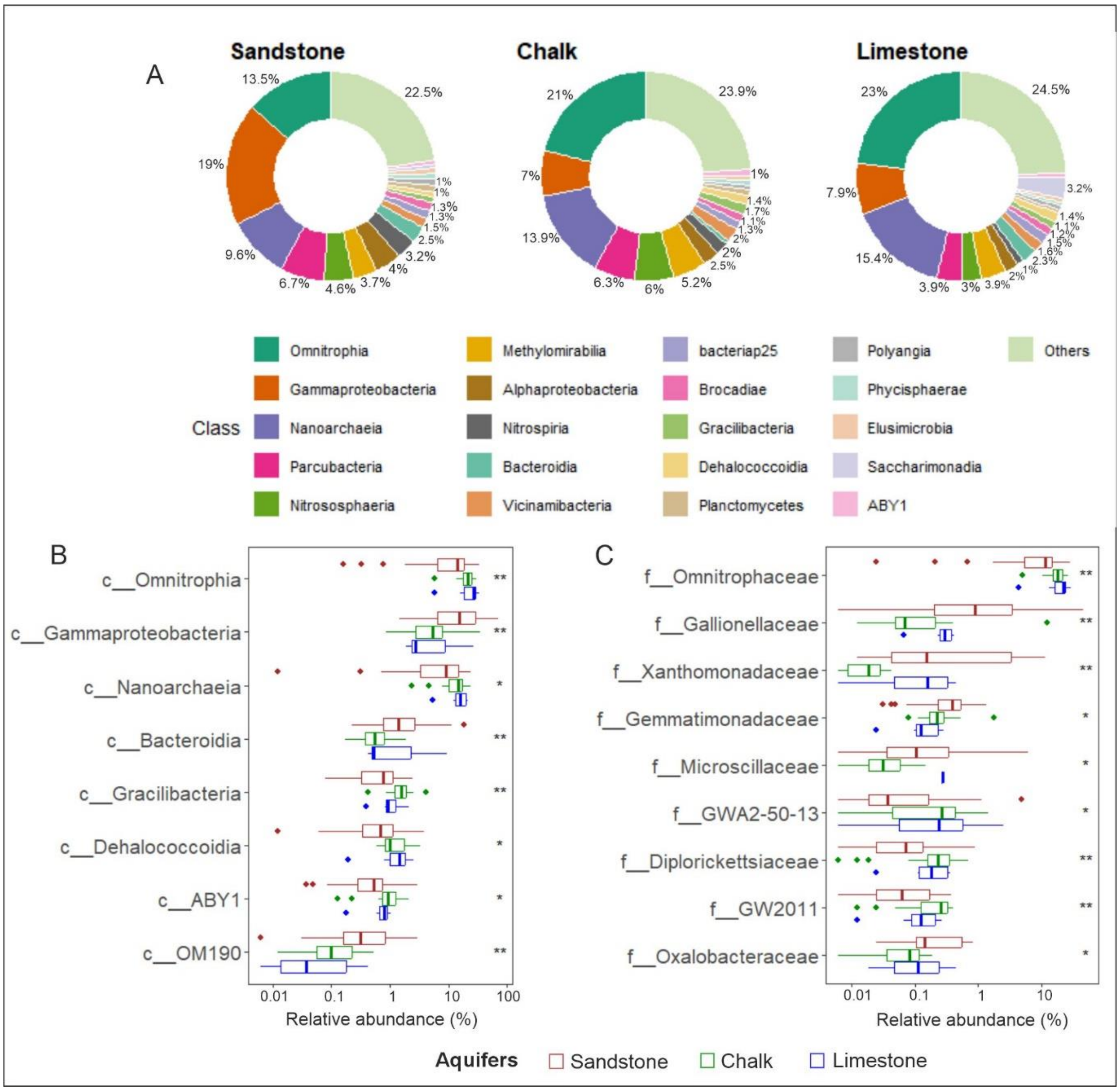


Figure 5-4. A. Donut plots of dominant prokaryotic classes with more than 1% mean relative abundance in each aquifer type showing the mean relative abundance of each class as percentages of the total community; B. Boxplots of prokaryotic taxa with significantly different relative abundances of prokaryotic classes and C. families, and according to LEfSe analysis, significance levels of the differences were ** $p<0.01$ and * $p<0.05$.

The donut plots of the mean relative abundances of taxonomic classes in each aquifer showed apparent differences among the three aquifers (Figure 5-4. A). The dominant prokaryotic classes were the same in both rarefied and unrarefied data (Appendix 12). The LEfSe analysis (Figure 5-4. B, C) revealed that among the three aquifers, the sandstone aquifer had almost three times higher relative abundance of Gammaproteobacteria (19%) than the other two aquifers (7 to 8%). Within the class

Gammaproteobacteria, families *Gallionellaceae, Xanthomonadaceae*, and *Oxalobacteraceae* were dominant in sandstone, followed by limestone aquifer. Family *Diplorickettsiaceae* within class *Gammaproteobacteria* was dominant in the chalk and limestone aquifers than in the sandstone aquifer. Class *Bacteroidia* was dominant in the sandstone (2.5%) and limestone (2.3%) aquifers, and family *Microscillaceae* within this class was dominant in the sandstone aquifer. The sandstone aquifer also had a higher relative abundance of the *OM190* class. The chalk and limestone aquifers possessed similar relative abundances of classes *Omnitrophia* (21-23%) (family *Omnitrophaceae*), *Nanoarchaeia* (13.9-15.4%) (family *GW2011*), and *Dehalococcoidia* (1.4%), which were relatively higher than those in the sandstone aquifer. Only the chalk aquifer had dominance of classes *Gracilibacteria* (1.7%) and *ABY1* (1%) than other aquifers. The limestone aquifer also showed a high abundance of class *Saccharimonadia* (3.2%) (Figure 5-4. A).

5.3.4 Prokaryotic functional potentials of three aquifers

Due to the presence of many unclassified sequences, and only 16.5% of the ASVs classified to a genus level, there were more rare and unknown functions in three aquifers than there were known functions (Figure 5-5). The nutrients used and the redox condition requirements of each function are provided in Appendix 16. Among the known functions, chemoheterotrophy was dominant in the sandstone aquifer (15.9%) than in the chalk (9.1%) and limestone aquifer (7.3%). The sandstone aquifer had a higher abundance of aerobic chemoheterotrophy (13%) than chalk (5.4%) and limestone aquifers (4.8%). Anaerobic chemoheterotrophy potential was similar in the three aquifers (2.5-3.9%). The sandstone and chalk aquifers had similar abundances of hydrocarbon degradation, methanotrophy, methylotrophy, nitrite respiration and nitrogen respiration functional potentials (between 1.6-1%), which were absent in the limestone aquifer. The sandstone aquifer also had dark iron oxidation potential (1.2%), nitrification (1%) and non-photosynthetic cyanobacteria (1%). Only the limestone aquifer had a dominance of animal parasites and symbiont functional potential (2.4%).

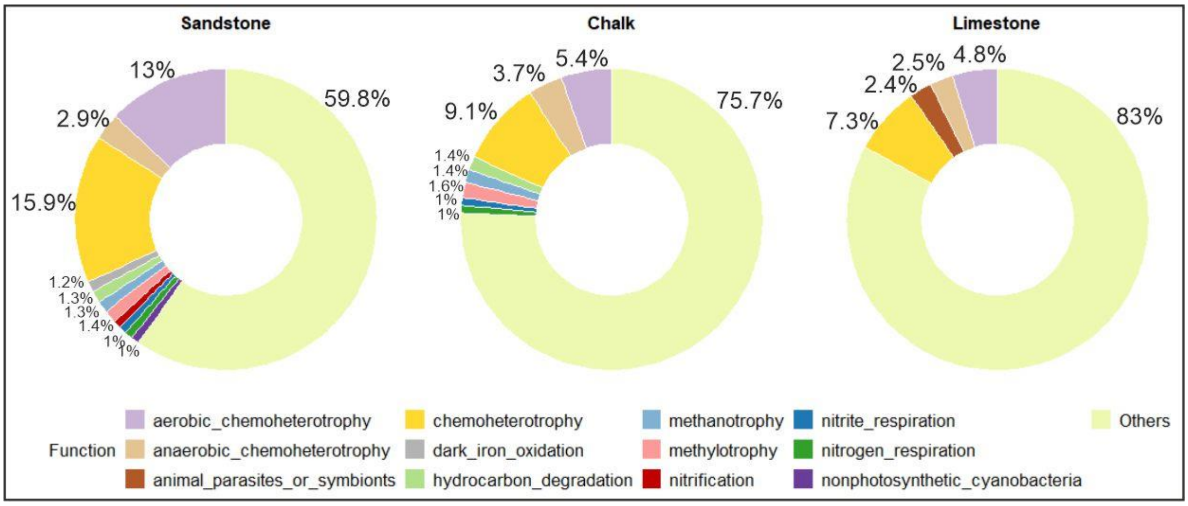


Figure 5-5. Donut plots showing mean relative abundances (%) of functional potentials with more than 1% mean relative abundance in three aquifers.

5.4 Discussion

The observed bacterioplankton concentration in the limestone aquifer was over twice that of both the sandstone and chalk aquifers. However, the taxonomic composition of the prokaryotic ecosystem in the sandstone aquifer exhibited notable differences when compared to the chalk and limestone ecosystems. This proved that the hypothesis that the sandstone aquifer has different prokaryotic types than the chalk and limestone aquifers was true.

5.4.1 Controlling factor of bacterioplankton concentrations

Groundwater bacterioplankton TCC variation across the three aquifers (Figure 5-2. A) was consistent with Chapter 4 (Figure 4-2. A). In Chapter 4, it was suggested that over the long term, differences in allochthonous bacterioplankton input, controlled by each aquifer's filtration capacity, explained TCC variation. This chapter provided further evidence that over the long term, more frequent allochthonous bacterioplankton input into the limestone aquifer resulted in higher TCC in this aquifer compared to the sandstone and chalk aquifers. The functional potentials related to animal parasites and symbionts, which are generally allochthonous pathogens found in groundwater systems, were detected only in the limestone aquifer. The presence of the eDNA from known microorganisms with animal parasitic functions did not prove that the pathogens were active. Still, it only indicated that the water encountered known animal parasites in the recent past, and samples were collected before the eDNA could degrade in the environment (Harrison *et al.*, 2019). Although the animal parasite signature was linked with TCC in limestone, the limestone aquifer catchment was not necessarily dominated by only arable land or grassland with grazing animals from where the animal symbionts could have originated (Appendix 11). This indicated that the TCC input was not related to a particular land use type but only related to how easily and frequently the allochthonous bacteria can enter the aquifer. The absence of animal-associated functions in chalk and sandstone aquifers indicates less easy surface bacterial loading due to filtration of those organisms and potential degradation and loss of any pathogenic extracellular eDNA.

The absence of animal parasitic signature in the chalk aquifer, but their presence in the limestone aquifer, also reflected the differences in the degree of karstification of these two aquifers. Chalk is a micro-karstic aquifer (Worthington and Ford, 2009) where around 70-90% of the groundwater recharge occurs through a matrix-dominated unsaturated zone (Maurice *et al.*, 2023b) that imposes a filtration effect on the allochthonous bacterioplanktons through the narrow pore-throats (median 0.5 µm; (Price, 1987)). In contrast, the limestone aquifer is moderately karstic (Worthington and Ford, 2009), with less filtration capacity imposed by larger apertures of karstic dissolution fractures (0.2-5 cm). Sinreich *et al.* (2014) classified alpine karstic springs and found that springs with a low degree of karstification had less frequent spikes of pathogenic indicators and had a TCC in the order of 10^3 cells/mL, like what was observed in the chalk aquifer in this study (8.7×10^3 cells/mL). The moderately karstified springs exhibited more frequent spikes in pathogen indicator bacteria and had TCC in the 10^4 cells/mL range, like the observed TCC in the limestone aquifer in this study (2.3×10^4 cells/mL). Thus, the TCC variation in English karstic aquifers reflected the degree of karstification and microbial contamination vulnerability, similar to the alpine karstic aquifers.

5.4.2 Controlling factor of prokaryotic community structure

Flow cytometry data provided further insights into the prokaryotic community structure. Although some earlier studies have proposed that the HNA bacteria are larger than the LNA bacteria (Wang *et al*., 2009), others suggested that HNA bacteria reflect the metabolically active portion of the community (Liu *et al*., 2016). However, the current perspective is that the HNA bacteria are phylogenetically distinct and possess a larger genome size, capable of a variety of metabolic functions, whereas LNA bacteria are different types of ultrasmall organisms with a reduced genome size (Proctor *et al*., 2018). This study showed that the prokaryotic communities of the sandstone aquifer were dominated by classes such as *Gammaproteobacteria, Alphaproteobacteria, Planctomycetota* and *Bacteroidia* (Figure 5-4. A), which are known to possess larger genetic material capable of performing various metabolic functions in a wide range of environmental conditions (Boussau *et al*., 2004; Hahnke *et al*., 2016; Kaboré *et al*., 2020). In the chalk and limestone aquifers, ultrasmall prokaryotic classes, such as *Omnitrophia* and *Nanoarchaeia*, dominated, which are known to possess reduced genetic material for energy-efficient, streamlined metabolic functional potentials (Gios *et al*., 2023; Seymour *et al*., 2023). These observations support the possibility that flow cytometric HNA/LNA fingerprinting can be used in the future to quantify the cell count of taxonomically distinct bacterioplankton with smaller or larger genetic material.

The high Shannon diversity indices (5.4-6.5) of the aquifers indicated the presence of a rich groundwater prokaryotic habitat in the English aquifers. Among them, the sandstone aquifer showed a lower median diversity than the chalk and limestone aquifers (Figure 5-3. A). Rarefaction and filtration volume analyses (Appendix 12) confirmed that sequencing depth and sample volume did not influence diversity estimates. Thus, the differences in the diversity indices in the three aquifers can arise due to one of two reasons. Similar taxonomic composition of the limestone and chalk aquifers may reflect that similar groundwater chemistries, shaped by rock–water interactions (Shand *et al*., 2007), selected the growth of similar taxa of more diverse types. The groundwater chemistry of the sandstone aquifer may have selected the proliferation of different and less diverse types of taxa. The dbRDA biplot also supported that the taxonomic composition of the chalk and limestone aquifers was similar and distinct from the sandstone aquifer. Alternatively, Shannon diversity was consistently higher in shallower boreholes and samples with higher DOC concentrations, likely due to easier infiltration and adaptation of allochthonous prokaryotes or the sustained production of diverse taxa under stable DOC supply in low-nutrient conditions (Benk *et al*., 2019; Yan *et al*., 2021). The limestone and chalk aquifer samples were obtained from shallower borehole depths than sandstone, potentially supporting the idea that easier allochthonous prokaryotic input and their adaptation led to higher diversity in shallow aquifers (Benk *et al*., 2019; Yan *et al*., 2021). Whether the reason for the similar alpha-diversity indices was the selection of communities by water chemistry controlled by the aquifer matrix or the allochthonous intrusion in shallower groundwater warrants further investigation.

In sandstone aquifers, the dominant prokaryotic classes like *Gammaproteobacteria (Gallionellaceae, Xanthomonadacea, Oxalobacteraceae), OM190, Bacteroidia*, etc., are reportedly facultative aerobes and are known to perform complex carbon degradation (Cardman *et al*., 2014; Gülay *et al*., 2016; Gutierrez, 2019). The abundance of chemoheterotrophic functional potentials supported this observation. In the deeper sandstone groundwater sources (Appendix 10), the presence of aerobic chemoheterotrophic taxa and functional potentials may indicate the presence of an oxic environment in deep groundwater (median dissolved oxygen= 5.2 mg/L in Chapter 6). The oxic environment in deeper groundwater may not be due to oxygen replenishment from the surface during recharge. Instead, in situ oxygen production by processes such as denitrification or sulphate reduction is possible (Ruff *et al*., 2024). Additionally, the presence of the family *Gallionellaceae,* which is known to perform dark iron oxidation, indicated the impact of rock-water interaction in this aquifer, as the pyrite in the sandstone matrix can act as the source of ferrous iron for a known function of this family (Jakus *et al*., 2021).

In the chalk and limestone aquifers, higher abundance of ultrasmall classes like *Omnitrophia* and *Nanoarchaeia* may indicate the presence of cryptic or hidden carbon and nitrogen transformation potentials by these ultrasmall prokaryotes (Beaver and

Neufeld, 2024; Gios *et al.*, 2023; Seymour *et al.*, 2023). The cryptic functions are not well understood (Beaver and Neufeld, 2024), and therefore, there was a large proportion of unconstrained functional potentials in these two aquifers. Additionally, the host-associated classes like *ABY1, Gracilibacteria,* and *Omnitrophia* indicated a potential in-situ source of organic carbon, as these microbes help in degrading microbial necromass and release DOC (Geesink *et al.*, 2022). An abundance of dehalogenesis-performing *Dehalococcoidia* (Saiyari *et al.*, 2018) indicated the presence of anoxic pockets with high chloride concentrations in chalk and limestone aquifers, which are reported at areas with mixing with old formation water (Shand *et al.*, 2007).

The dbRDA analysis (Figure 5-3. B, Table 5-1) showed a clear distinction between the carbonate (chalk and limestone) and the ferro-silicate (sandstone) aquifers. However, the environmental variables used to constrain the Bray-Curtis dissimilarity explained only 26.3% of the variance in the communities, indicating a large proportion of unmeasured variables in this study may impact the community structures. It is possible that analysing the concentrations of major ions in the groundwater samples can improve the explanatory power of the dbRDA model. The groundwater major ion chemistry of most of the chalk and limestone aquifers in England is Ca-HCO_3^- type due to bicarbonate dissolution from the carbonate matrix (Bearcock *et al.*, 2015; Neumann *et al.*, 2003; Shand *et al.*, 2003; Smedley *et al.*, 2004; Stuart and Smedley, 2009), but sandstone aquifer groundwater chemistry ranges from Ca-HCO_3^- to Ca-Mg-HCO_3^--SO_4^{2-} type (Griffiths *et al.*, 2002; Shand *et al.*, 2002; Smedley *et al.*, 2005; Tyler-Whittle *et al.*, 2002). The groundwater major ion chemistry, controlled by rock-water interaction, was found to be a significant determinant of community differences in groundwater in other regional studies (Abraham and Close, 2024; Amalfitano *et al.*, 2014; Couton *et al.*, 2023; Zhong *et al.*, 2023). Although groundwater source depth could play a part in community differences among the aquifers, the explanatory power of borehole depth and perforation depth was lower and less significant than the explanatory power of different geologies (Table 5-1).

The groundwater TDN in England is primarily composed of agricultural nitrate (Stuart and Lapworth, 2016), abundance of which is the biggest water quality issue in this country (Foster and Bjerre, 2023; Stuart *et al*., 2016; Wang *et al*., 2012). The dbRDA analysis also revealed a strong influence of TDN concentration on the variation of prokaryotic community differences within each aquifer type. The TDN concentration was higher in sites of all the aquifers with a thinner overlying stratum (Figure 5-3. B). In England, nitrate is stored in the unsaturated zones of sandstone and chalk aquifers and slowly leaches into the saturated zone of the aquifer over an extended period, replenishing the groundwater nitrogen (Sorensen *et al*., 2015; Wang *et al*., 2012). Thus, the TDN concentration in the groundwater of any site with thinner overlying strata indicates how easily recharge water can leach TDN from the unsaturated zone into the saturated zone. Despite the higher nitrate concentration in English aquifers, denitrification is limited due to low DOC concentration (Rivett *et al*., 2007). Aligning with previous literature, in this study, denitrifying functions were not observed in any of the aquifers. Instead, the presence of classes such as *Nitrospira* and *Nitrisosphaeria* indicated that inorganic nitrogen oxidation pathways such as nitrification (Wegner *et al*., 2019), were more dominant nitrogen transformation pathways in the sandstone and chalk aquifers. However, as indicated previously, the presence of the eDNA of inorganic nitrogen-oxidising organisms does not necessarily mean their active presence in the groundwater. Autotrophic denitrification in sandstone aquifer may be possible, where the abundance of the family *Gallionellaceae* can perform dark iron oxidation coupled with nitrate reduction (Ben Maamar *et al*., 2015; Jakus *et al*., 2021). The critical role of TDN in controlling community composition within each aquifer aligns with the strong influence of inorganic nitrogen on ecosystem structures in other regional surveys (Abraham and Close, 2024; Korbel *et al*., 2024; Sirisena *et al*., 2018). Moreover, the impact of TDN on groundwater ecosystems suggests that over time, slow leaching of anthropogenic nitrate into the aquifers, especially those with shallow overlying strata, can alter the groundwater prokaryotic communities and their biogeochemical cycles. Despite nitrate concentration being the most critical water quality issue in England, the lack of known denitrifiers in groundwater can pose concerns regarding nitrate attenuation.

5.4.3 Study implications, limitations and future directions

By utilising modern ecosystem assessment techniques, including flow cytometry and eDNA metabarcoding, this study produced the first comprehensive dataset of prokaryotic ecosystems in groundwater from major aquifers in England. This study makes a substantial contribution to the growing number of global studies attempting to characterise the regional and national-scale spatial variation of groundwater microbial communities, incorporating ecosystem health assessment for groundwater resource management (Abraham and Close, 2024; Korbel *et al.*, 2024; Sirisena *et al.*, 2018; Smith *et al.*, 2018; Zhong *et al.*, 2023). Although different aquifer geologies were not adequately addressed in previous national-level surveys (Sirisena *et al.*, 2018; Zhong *et al.*, 2023), in this study, the comparison of hydraulically disconnected and distinct aquifer geologies revealed a strong influence of aquifer geology on both the bacterioplankton concentration and the prokaryotic community structure.

The study utilised groundwater from drinking water pumping stations in England, thus producing the first dataset from drinking water sources. As eDNA sequencing technology becomes more accessible, this technique has been proposed for incorporation into water supply industries for groundwater ecosystem management and monitoring (Watson *et al.*, 2024). Similar national-scale groundwater ecosystem monitoring of drinking water resources is becoming increasingly necessary to establish a monitoring reference. This dataset from groundwater sources with minimal contamination can serve as the first dataset to build a modern reference for future groundwater ecosystem monitoring in England. It was also clear that the reference ecosystem of groundwater will be spatially variable and is primarily impacted by aquifer geology. Moreover, despite the geographic separation of sandstone, limestone and chalk sites, the communities in different geologies were more different than the ecosystems in geographically proximal sites, for example, proximal sites of three different geologies in the North-East of England (Figure-5-1). The correlation between geographic separation and community differences was weak (Appendix 15) and in the dbRDA plot, the communities from same geologies were clustered together rather than overlapping with other geologies (Figure 5-3. B). This aligns with a previous observation where Mehrshad *et al*. (2021) found geographically separated granitic aquifer had similar core-microbiome due to similar ecological niches of granitic lithologies. In any future spatially extensive studies, different geologies should be considered as a potential factor influencing groundwater prokaryotic variation.

A notable observation was the low percentage of ASVs classified at a taxonomic level. Only 76.8% of ASVs were assigned to a phylum level, and only 16.5% to a genus level. Couton *et al*. (2023) also reported similar findings in the groundwater of Austria, where many ASVs were not classified into a taxonomic rank. This can be attributed to the presence of numerous novel taxa in groundwater that have not yet been sequenced. The scarcity of whole-genome sequences in groundwater samples prevents the classification of many species, as well as understanding their metabolic functions and their contribution to the biogeochemical cycling of elements. Functional assignment using FAPROTAX is a preliminary attempt at exploring the groundwater prokaryotic functions, which was also reported in recent groundwater studies (Korbel *et al*., 2024). However, this approach relies on assigning known functions to known genera. Sansupa *et al*. (2021) reported that the functional assignment becomes less reliable with a lower proportion of taxonomic assignment of ASVs at the genus level. Therefore, more metagenomics studies (e.g., Anantharaman *et al*. (2016)) are necessary for characterising the taxonomic and functional variation of groundwater prokaryotes.

The key limitations of this study were the unbalanced sample sizes and the lack of sampling campaign repetition. The number of samples from sandstone sites was much higher than that from chalk and limestone sites. With higher and a more balanced sample count from each aquifer, it would also be possible to find the core-taxa of each aquifer, although the threshold of core-taxa determination is arbitrary. When large volume of samples (130 to 733 samples) was available, (Zhong *et al*., 2023) set the threshold of core-taxa as taxa available in at least 50% samples with at least 0.01% relative abundance, whereas with a smaller sample number (5 to 20) from a water treatment work, (Gülay *et al*., 2016) set the threshold as taxa available in more than 1% relative abundance with no prevalence threshold. Due to the sample-number dependent threshold of core-taxa calculation, it was not possible to determine the core-taxa of the three aquifers with unbalanced sample number. Besides, in order to define the reference ecosystems, multiple sampling campaigns with sample replicates should be undertaken in the future to differentiate the range of "natural ecosystem variation" from ecosystem variation under environmental stresses. With more sampling repetitions it would also be possible to reveal the natural ecosystem of each aquifer type.

Although this study has revealed a clear distinction between the communities of different aquifers, there was also an indication of spatial variation of community structure within each aquifer. In sandstone and chalk aquifers, there was an apparent variation of ecosystems along a gradient of TDN, but this was not so clear in the limestone aquifer. In a recent study of a French limestone aquifer, the authors suggested that the baseline ecosystem varies within a range depending on the local chemical signatures (Harris *et al*., 2025). This suggests that more intensive sampling of chalk and limestone may reveal the environmental variables that control the spatial differences of communities within individual aquifer types.

In the future, more work should be done to understand the factors which control the ecosystem variations and improve the explanatory power of the dbRDA model. While the environmental variables used in this study explained only 26.3% of the community variation, it is common is groundwater studies. Current literature suggests that selection pressure by chemical variables can have a wide variety of explanatory power on the microbial community variation, such as ~10% (Villeneuve *et al*., 2022), ~35% (Yan *et al*., 2020), ~45% (Wang *et al*., 2025b) of the variation in groundwater microbial communities. Other environmental variables such as recharge can explain 12-30% (Wang *et al*., 2025b; Yan *et al*., 2021), spatial distance can explain ~18% (Yan *et al*., 2020) of the community variation. In a national-scale study, (Zhong *et al*., 2023) found that ~75-92% of the community variation remained unexplained by both geographic separation and environmental variables, and the unexplained community variation increased with depth of the aquifers as surface disturbances got weaker. In fact, selection pressure is typically stronger near contaminated sites compared to uncontaminated sites, where stochastic processes, such as dispersal and random drift dominate (>60%) as community assembly processes (Ning *et al*., 2024). Since the groundwater samples in this study were free of heavy contamination, the lower explained community variation by known environmental variables in expected. It was also observed that dispersal process in areas with higher hydraulic gradient has less impact of environmental selection on the microbial communities (Retter *et al*., 2023), which is a suitable explanation for lower impact of selection by environmental variables near the abstraction boreholes with typically large hydraulic gradients. In the future, the dominance patterns stochastic community assembly processes could be explored. Additionally, to improve the dbRDA model constrains, more comprehensive physicochemical data can be collected. Important physicochemical parameters, such as dissolved oxygen (DO), pH, temperature, conductivity, sulphate, nitrogen species, major ions, and concentrations of Fe and Mn species, have been shown to impose large constrains on ecosystem variations (Amalfitano *et al*., 2014; Ben Maamar *et al*., 2015; Harris *et al*., 2025; Sirisena *et al*., 2018; Zhong *et al*., 2023).

Our Capscale model showed that the overlying strata thickness and TDN concentration play a crucial role in the community composition of individual aquifers, it is possible that groundwater recharge also plays a key role in community composition. The frequency of recharge in each aquifer can be assessed using the groundwater recharge age. Previously, it was found that frequent input of recharge water from the surface can be a major controlling factor of communities (Ben Maamar *et al*., 2015). Similarly, Sirisena *et al*. (2013) observed that bacterial biomarker clusters showed differences in groundwater with newer and older (>100 years) residence times. In the future, studies focused on each aquifer should be performed to assess the hypothesis that recharge age will play a crucial role in controlling within-aquifer prokaryotic community variation.

5.5 Conclusion

The three major aquifers in England host different prokaryotic communities, primarily influenced by the geology of the aquifers. The concentrations of bacterioplankton are controlled by the frequency of allochthonous bacterial input into the groundwater, which is influenced by the filtration capacity of the respective aquifer matrix. However, the taxonomic compositions of prokaryotic communities were different in three aquifers, and similar in aquifers with comparable groundwater major ion chemistry. Therefore, the hypothesis that the sandstone aquifer will have different prokaryotic ecosystems from the chalk and limestone aquifers was true. This indicated that the prokaryotic community compositions depended on the chemical composition of the aquifers, specifically whether they were silicate or carbonate aquifers. The community differences within each aquifer depended on the gradient of TDN concentration and the thickness of overlying strata. Thus, besides aquifer mineralogy, the amount of chemicals accompanying surface recharge appears to impact the communities within individual aquifer types. The study revealed the presence of numerous novel ASVs in groundwater that may not have been previously classified. The prokaryotic ecosystem data from drinking water sources can also serve as a reference for future groundwater monitoring.

6. Impact of groundwater recharge age on spatial variation of planktonic prokaryotic communities in Permo-Triassic sandstone aquifer

This chapter addresses thesis **Objective 4**: To assess the impact of recharge age on the spatial variation of prokaryotic community composition within a single major aquifer type. A version of this chapter is under preparation for submission.

This is part of the same dataset mentioned in Chapter 5 and published at the National Centre for Biotechnology and Information, titled “Amplicon sequences (16S) from samples collected from groundwater survey of UK aquifers” with BioProject accession number PRJNA1268368 (ID 1268368 - BioProject - NCBI).

6.1 Introduction

The groundwater systems harbour specialised prokaryotes that play essential roles in global biogeochemical cycles (Falkowski *et al.*, 2008) and contribute to ecosystem services, such as maintaining safe quality of drinking water (Griebler and Avramov, 2015). Consequently, monitoring and protecting groundwater ecosystems have become a significant priority within environmental policy frameworks (Hose *et al.*, 2023; Watson *et al.*, 2024). Research has indicated that groundwater recharge by either natural processes (Danczak *et al.*, 2016a; Villeneuve *et al.*, 2022; Yan *et al.*, 2021; Zhou *et al.*, 2012) or artificial processes (Fiedler *et al.*, 2018; Sidhu *et al.*, 2015; Voisin *et al.*, 2018) controls the temporal variation of groundwater prokaryotes. With recharge water, nutrients like oxygen, carbon and nitrogen, as well as allochthonous taxa, are transported into the groundwater. The chemical disturbances can impose selection pressure on groundwater indigenous taxa and cause proliferation of dominant taxa, which can utilise the modified chemical constituents (Cooper *et al.*, 2016; Fillinger *et al.*, 2019a; Stegen *et al.*, 2016). Due to unrestricted and frequent replenishment of nutrients and oxygen by recharge, shallow unconfined aquifers harbour aerobic taxa, while confined deeper aquifers, separated from recharge, can harbour anaerobic and autotrophic taxa (Smith *et al*., 2012). As a result of the differences in chemical replenishments from different land use practices in aquifer catchments, in shallow aquifers (1-30m deep), land use was found to impact the groundwater prokaryotic communities (Couton *et al*., 2023; Korbel *et al*., 2013), although in deeper groundwater, such an impact was not evident (Sirisena *et al*., 2018). Some of the allochthonous taxa, migrated with recharge, can compete with the in situ microbial community for limited resources and may adapt to the groundwater environment (Benk *et al*., 2019; Fiedler *et al*., 2018; Yan *et al*., 2021). In shallow aquifers, seasonal recharge changes community structure, which again returns to the near-original state during the post-recharge recession period, resulting in an oscillatory temporal dynamic (Wang *et al*., 2025b; Yan *et al*., 2021). In deeper and isolated parts of aquifers, with limited surface connectivity, such disturbances in response to recharge are less pronounced in the short term (Danczak *et al*., 2016a; Villeneuve *et al*., 2022; Wang *et al*., 2025b). Due to the spatially variable recharge response of groundwater prokaryotes, large-scale spatial studies should be temporally repeated to record the spatial differences of recharge responses in a large region, which is lacking from current literature.

It has been sparsely studied how the long-term groundwater recharge age can impact the groundwater prokaryotes. A national survey of New Zealand groundwater using bacterioplankton molecular profiles showed that different microbial clusters existed in old (>100 years) and reducing versus young and oxidising groundwater, although taxonomic identification was not performed in this study (Sirisena *et al*., 2013). A study of fractured hard rock aquifers in France has shown that the distribution of younger groundwater and older groundwater, controlled by local hydrological flow paths, selected the proliferation of different aerobic and anaerobic taxa (Ben Maamar *et al*., 2015). However, no known large-scale spatial studies have assessed the effect of groundwater recharge age on the spatial variation of groundwater prokaryotic communities.

This study aimed to find the relationship between groundwater recharge and the spatial variation of groundwater prokaryotic communities by repeating seasonal samples and by measuring the long-term recharge age of groundwater. For this, regionally extensive monitoring was performed for the first time on the Permo-Triassic sandstone aquifer units in England, which is one of the most important groundwater sources in the country (Allen *et al*., 1997). This aquifer is also a typical example of red sandstone aquifers in Europe, North Africa, the Middle East, and North America (Celle-Jeanton *et al*., 2009). In England, these aquifer units are heavily exploited by intensive groundwater abstraction and have been affected by legacy nitrate contamination, organic pollutant inputs, and applications related to nuclear waste storage and carbon injection (Holloway and Savage, 1993; Medici and West, 2022; Rivett *et al*., 2007). However, no research was found to explore the prokaryotic spatial distribution in this aquifer by using eDNA sequencing techniques. This study not only aimed to develop a crucial understanding of the spatial variation of prokaryotic communities and groundwater recharge age, but the dataset was also an essential first step to explore prokaryotic communities of the sandstone aquifer, which will inform national groundwater ecosystem health management strategies and predict changes under newer pressures, such as emerging pollutants and increasing abstraction stress.

Groundwater samples were collected from actively abstracting boreholes used for the drinking water supply. Seasonal sampling was performed to test whether the community of the groundwater recharge season changes during the recession season. Also, groundwater recharge age was measured to test if the community composition can be related to the length of time since recharge water entered the subsurface from atmospheric contact. Among the environmental variables, DOC, TDN and DO concentrations, groundwater physicochemistry were measured as these can vary depending on recharge age, catchment land use pattern was assessed as it can control the nutrient concentration, aquifer confinement and groundwater source depth were measured as these can impact how long it takes for recharge water to travel to the sample source. In Chapter 4, since the bacterioplankton cell concentration and flow cytometric community metrices (high/low nucleic acid cells and intact cells) of the aquifer did not show substantial changes, it was hypothesised that seasonal recharge would also have minimal impact on the community structure. In Chapter 5, results indicated a considerable variation of sandstone aquifer groundwater communities along a gradient of decreasing total dissolved nitrogen (TDN) and increasing thickness of overlying strata, indicating a potential impact of long-term nutrient input from the surface on the spatial variation of communities. From the evidence in Chapter 5, it was hypothesised that groundwater recharge age would be related to the spatial variation of prokaryotic communities.

6.2 Study area and methodology

6.2.1 Study area

Groundwater samples were collected from untreated water taps at 47 drinking water pumping boreholes managed by Severn-Trent Water across the English Midlands and East Midlands (Figure 6-1). All sites abstract water from the Permo-Triassic red sandstone aquifer units, composed of hydrologically connected Sherwood and Bridgnorth sandstone groups. The sandstone aquifer units of west Cheshire, Staffordshire, Shropshire, Worcestershire, and Nottinghamshire used in this study are collectively termed the sandstone aquifer hereafter. Groundwater is stored in interconnected pore spaces within the sandstone matrix, which provides high storage capacity given aquifer porosity of 25–30%, but which limits flow due to low permeability. Horizontal hydraulic conductivity (Interquartile range (IQR): 0.19–2.04 m/day) is approximately twice as high as vertical conductivity (IQR: 0.07–1.18 m/day), leading to slow recharge rates (~1 m/year) (Allen *et al.*, 1997). The heterogeneity of the aquifer further influences groundwater flow, with variability arising from the grain size distribution, cementation degree, and fracture networks (Allen *et al.*, 1997). Microbial studies of sandstone have encompassed pathogens only and found that the small pore-throat size (0.1-90 μm) of the aquifer matrix often prevents pathogenic intrusions, although preferential flow paths such as fractures can allow rapid intrusion from the surface (Bloomfield *et al.*, 2001; Powell *et al.*, 2003).

The 47 sites were sampled twice a year, once in winter (February-March in 2023) and again in summer (August-September 2023). Winter is the typical peak-recharge season, and summer is the typical recession or post-recharge season, as anticipated from historical British weather trends and groundwater level data. Among them, 16 boreholes were in the confined aquifer and 31 boreholes were in the unconfined aquifer. Borehole depths ranged from 76 m to 391 m, and the perforation depth (top of the borehole screen) was from 9 m to 147 m deep. The thickness of overlying strata on the aquifer, including drift and/or confining layer, ranged between 1 m and 142 m. Groundwater levels (GWL) from Environment Agency monitoring boreholes (Environment Agency, 2024) were used to visualise GWL changes from recharge to recession season. Most of the unconfined sites showed a decline in GWL from peak-recharge to post-recharge season, while confined sites exhibited no consistent seasonal change (Appendix 17).

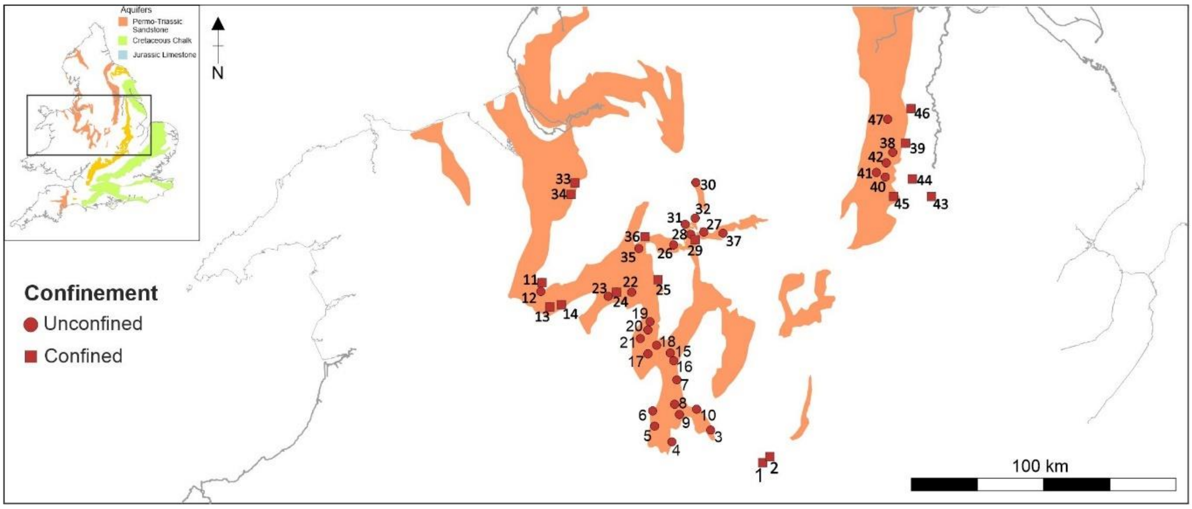


Figure 6-1. Study area showing locations of sample pumping stations, the shaded area is the outcrop of the sandstone aquifer, and numbers indicate site identification numbers and shapes indicate aquifer confinement at the sites. (Boreholes location retrieved from GeoIndex data centre's (NGDC) scanned borehole collection BGS © UKRI (2023), map outline contains OS data © Crown copyright and database rights 2025)

6.2.2 eDNA Sampling and analysis

Raw groundwater was filtered using three 0.22 µm Sterivex™ filters with PES membranes (Merck, UK) attached in parallel. Filtration was performed for 45 min, or a maximum of 15 litres, whichever was reached first. Filtration volume ranged from 5.5 L to 15 L. The rest of the eDNA collection and analysis protocol was the same as mentioned in Sections 5.2.2 and 5.2.3. The eDNA sequencing was performed using the MiSeq reagent kit (v2) on an Illumina MiSeq2 platform (Illumina Inc., USA).

6.2.3 Physico-chemical analysis

After eDNA sampling, field measurements of pH, conductivity, temperature, and dissolved oxygen (DO) were conducted using Mettler Toledo Seven2Go Pro™ probes connected to a flow cell to avoid atmospheric interference. The probes were calibrated using appropriate standards: pH (4, 7), conductivity (718 and 1413 µs/cm), and DO (100% atmospheric saturation).

6.2.4 Flow cytometry

Raw water samples were collected in 50 mL sterile centrifuge bottles and stored at 4ºC for a maximum of 48 h before analysis. For the measurements of total cell concentration (TCC), high and low nucleic acid cell (HNA/LNA) fingerprinting and bacterial intact cell concentration (ICC) analysis using an Attune™ CytPix™ Flow Cytometer, the same protocol mentioned in Section 3.1 was used.

6.2.5 Dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) analysis

A 50 mL sample was filtered through 0.45 µm Whatman™ PES membrane filters into dry HDPE bottles, and dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) concentrations were measured using an Elementar Vario Cube (Elementar Analysetechnik GmbH; Langenselbold, Germany) using the same protocol mentioned in Section 4.2.

6.2.6 Groundwater recharge age measurement using Chlorofluorocarbon (CFC)

Groundwater recharge age, or the apparent age was determined using analysis of three Chlorofluorocarbon (CFC) compounds, i.e., CCl3F (CFC11), CCl2F2 (CFC12), C2Cl3F3 (CFC113). Measuring these anthropogenic tracers in groundwater yields the residence time of the water, that is, the time since the water was in equilibrium with the atmosphere. The apparent groundwater age, or recharge age, depends on Henry's law constant calculated at the recharge temperature, which was considered 11 °C (Gooddy *et al.*, 2006).

CFC sampling and analysis were performed following the methods described by Oster *et al.* (1996) and Gooddy *et al.* (2006). CFC samples were collected in glass bottles under a water jacket to ensure no atmospheric contact. CFCs were analysed by gas chromatography with an electron capture detector (GC-ECD) after pre-concentration using cryogenic methods. The detection limit for CFC concentration was 0.01 pmol/L and the concentrations were determined after resolving the effects of recharge temperature, pressure, salinity, excess air, and degassing. The analysis was calibrated to the bulk air standards collected at an atmospheric monitoring station in the AGAGE network. CFC-12 has been shown to be the CFC least susceptible to contamination in UK studies (Darling *et al.*, 2012) and hence was used for recharge age calculation hereafter in this paper.

6.2.7 Land use pattern analysis

The UKCEH 2021 Land Cover Map (LCM) was obtained from the UKCEH data repository (Marston *et al.*, 2022). The Environment Agency source protection zone 1 (SPZ-1) map was downloaded from the open source (Environment Agency, 2019). The percentage coverage of each land use category in the SPZ-1 was calculated using the ArcGIS Pro 3.2.2 (Esri) zonal statistics tool, as mentioned in Section 5.2.6.

6.2.8 Data analysis

All data was analysed in R v.4.3.2.

6.2.8.1 Bioinformatics

The eDNA sequence data were processed using the DADA2 pipeline (Callahan *et al.*, 2016). Sequences below a quality score of Q30 were trimmed, forward and reverse adapters were removed, reads with a maximum expected error of 10 were filtered, reads matching against the PhiX genome were removed, and finally, forward and reverse reads were merged. The sequences were aligned to find the amplicon sequence variants (ASV), the read frequency of each ASV in each sample was calculated, and chimeric sequences were removed. The ASVs were compared against the SILVA database (Silva) accessed April, 2024. and the taxonomic assignment of ASVs was done in Kingdom, Phylum, Class, Order, Family and Genus level.

The sequences were rarified to a uniform sequencing depth of 14,696 reads and filtered to include only bacteria and archaea kingdoms for downstream analysis. While rarefaction is the best practice (Schloss, 2024) there is a chance of underestimated diversity. Therefore, downstream analysis result of unrarefied sequences is presented in Appendix 21 to ensure consistency with rarefied data.

Processed sequencing data were analysed using the Microeco package in R (Liu *et al*., 2021). The relative abundance of the ASVs was calculated, followed by Shannon diversity index and Bray-Curtis dissimilarity matrix (beta diversity) estimations based on the relative abundances of ASVs and Sørensen-Dice dissimilarity matrix was calculated based on presence/absence pattern of the ASVs in each sample. Analysis of similarity (ANOSIM) was performed to check if the relative abundances of ASVs in the samples changed significantly between the recharge sampling season and the recession sampling season. To find sampling sites with similar prokaryotic communities, hierarchical clustering was performed based on the Bray-Curtis dissimilarity matrix, and a dendrogram was constructed using the Ward.D2 method (Murtagh and Legendre, 2014). A Silhouette plot (Appendix 20) showed that there seem to be an optimum five clusters. Donut plots of relative abundances of the dominant taxa (mean relative abundance more than 1% in a hierarchical cluster) in each hierarchical cluster was used to visualise the differences among the clusters. To identify the dominant prokaryotic taxa which varied significantly ($p<0.001$) in abundance among the five HCs, a linear discriminant analysis effect size (LEfSe) (Segata *et al*., 2011) was performed. To assess the relationship between environmental variables, including the recharge age and the prokaryotic community dissimilarities, a distance-based ReDundancy Analysis (dbRDA) (Legendre and Anderson, 1999) was performed, and r^2-value, or environment fit values of each environmental variable, calculated using envfit() function was reported. To ensure robustness of interpretation, another dbRDA model was ran on the Sørensen-Dice dissimilarity matrix. A post-hoc Mantel test of Spearman Correlations between the Euclidean distance matrix of environmental variables and the Bray–Curtis dissimilarity matrix of prokaryotic communities was performed to test the association of differences in the environmental variables and the differences in the communities (Smouse *et al.*, 1986). Finally, boxplots of environmental variables, including recharge age, were prepared to check which variables significantly differed across different prokaryotic community clusters.

6.2.8.2 Statistical analysis

Shapiro-Wilks test value of $p<0.05$ indicated all environmental variables (Flow cytometric TCC, %HNA, %ICC, DOC, TDN, DO, pH, conductivity, temperature) were non-normally distributed (Shapiro and Wilk, 1965). Among the descriptive statistics, the interquartile range (IQR) and medians of the variables were reported. Winter (typical recharge-season) data were plotted against summer (typical recession-season) data to observe seasonal changes in environmental variables, and the Wilcoxon signed-rank exact test was performed to test for significant differences, and results were reported using V-value (Wilcoxon, 1992). A principal component analysis (PCA) was performed to find the associations among environmental variables (Wold *et al.*, 1987). Boxplots were prepared for visualisation of the range of environmental variables across different hierarchical clusters. The box hinges represent the interquartile range (IQR) and the median, the whiskers represent points up to 1.5 times the IQR and any point beyond that is deemed to be an outlier (McGill *et al.*, 1978). For all statistical tests, the significance thresholds were $p<0.05$ for weak significance, $p<0.01$ for moderate significance and $p<0.001$ for strong significance levels.

6.3 Results

6.3.1 Relation of seasonal recharge and long-term recharge age with groundwater bacterioplankton concentration and physico-chemistry

Groundwater samples exhibited low bacterial concentrations with an extensive range (IQR= 6.5×10^{3}-2.7×10^{4} cells/mL) with a low proportion of high nucleic acid cells (15-33%) and intact cells (12-22%). Dissolved organic carbon (DOC) concentrations were low (0.53-0.92 mg/L). In contrast, total dissolved nitrogen (TDN) concentrations (3.3–10.7 mg/L) were high due to legacy nitrate in UK groundwater (Rivett *et al.*, 2007). Groundwater was suboxic (DO= 3.2-7.3 mg/L). Groundwater pH was slightly alkaline (7.2-7.6), with low conductivity (439-700 µS/cm). CFC-12 recharge age data indicated variable groundwater recharge age, ranging from the detection limit of 1949 to 2023, with an IQR spanning 1968 to 1988. A distinction between unconfined and confined sites was observed (Figure 6-2. A), with unconfined sites having higher TDN (7-11.7 mg/L) and DOC (0.6-0.9 mg/L) concentrations compared to the confined sites (TDN= 0.4-7, DOC= 0.4-0.6 mg/L). The confined sites exhibited higher temperature (11-14°C) compared to the unconfined sites (10.6-12°C).

The median values and interquartile ranges of the physico-chemical variables did not change significantly in the sandstone aquifer. But there were minor differences at individual sites, which were reflected in the pairwise test. Wilcoxon test showed that among the flow cytometric parameters, only %HNA showed a significant drop from winter (median=26%) to summer (18%) ($V=257$, $p<0.05$). From winter to summer, there was a drop in TDN ($V=71$, $p<0.001$) and DO concentrations ($V=45$, $p<0.001$), along with an increase in DOC concentration ($V=587$, $p<0.05$), pH ($V=951$, $p<0.001$), and temperature ($V=1173$, $p<0.001$). However, the direction and magnitude of seasonal change of any variable showed no dependence on aquifer confinement (Figure 6-2. A)

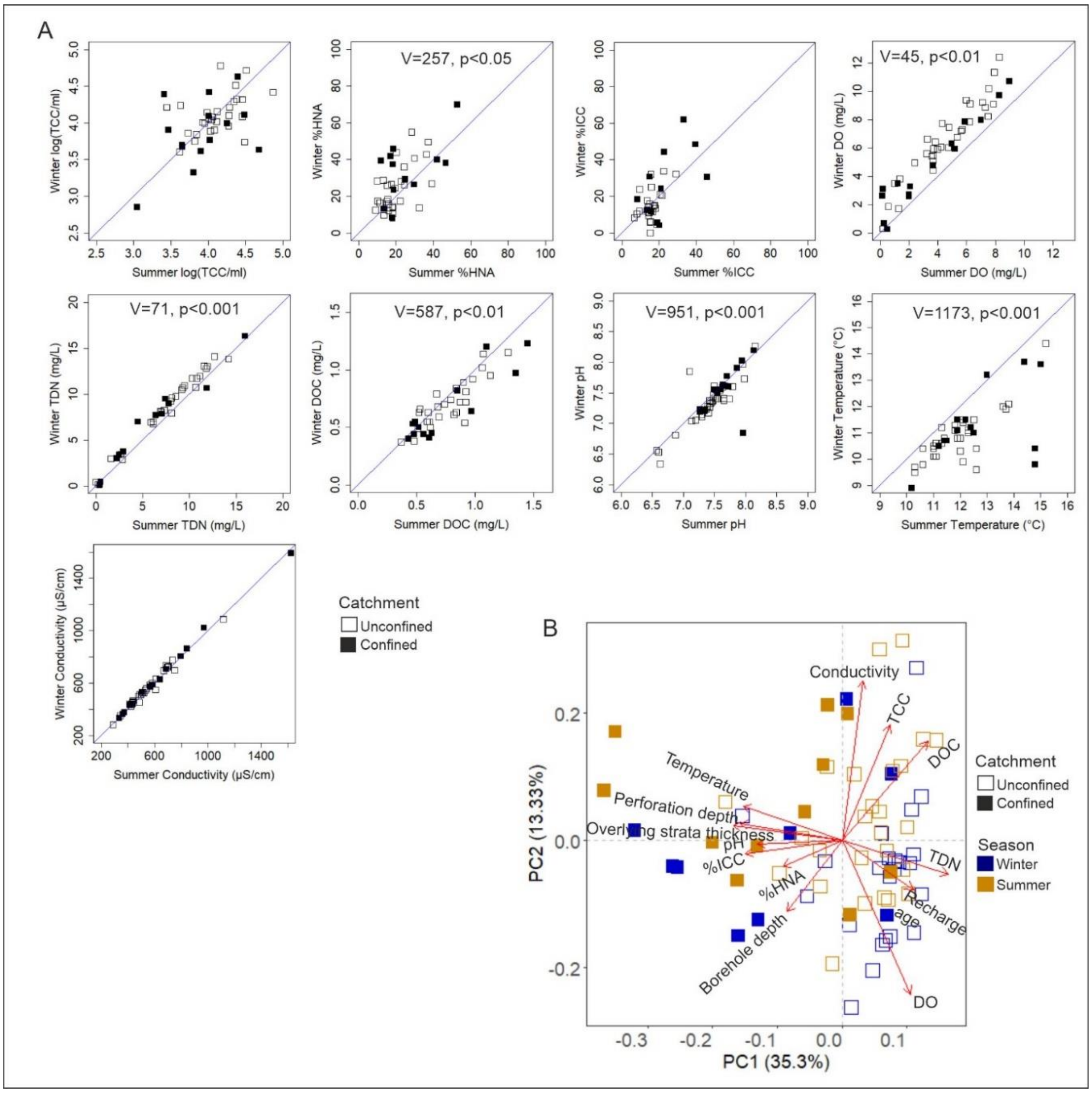


Figure 6-2. A. Groundwater bacterioplankton concentration, nutrient concentrations and physico-chemical variables showing summer versus winter values along with results of Wilcoxon paired test results for seasonally significantly different parameters and x=y line for reference; B. Principal component plot of the environmental variables with arrow direction and length indicating loading direction and amount of loading along PC1 and PC2 axes, shapes indicating confinement and colours indicating sample season.

A principal component analysis (PCA) showed that the first four principal components (PCs) with eigenvalues greater than 1 explained 63.5% of the total inertia in the data, and the loadings of each variable along the first four PCs are indicated in Appendix 18. The PCA biplot (Figure 6-2. B) showed that the unconfined groundwater from shallower depths under thinner overlying strata zone had a younger recharge age, a strong positive association with DO and TDN concentrations, and a weaker association with DOC concentrations. The confined groundwater from deeper perforation depths and thicker overlying strata was positively associated with warmer and alkaline groundwater with low TDN, DOC and DOC concentrations. Moreover, bacterioplankton TCC had a strong positive association with DOC (Spearman correlation $\rho=0.5$, $p<0.001$) and a negative association with perforation depth ($\rho=-0.46$, $p<0.001$), but higher %ICC and %HNA bacteria were present in deeper and older groundwater. PC1 explained 35.3% of the inertia and had positive loadings of recharge age (0.24) along with TDN (0.36), DOC (0.3), DO (0.23) and negative loadings of pH (-0.28), temperature (-0.33), overlying strata thickness (-0.37), perforation depth (-0.34), %HNA (-0.2) and %ICC (-0.33). PC2 explained 13.3% of the variance and had a strong positive loading of bacterial TCC (0.4) along with DOC (0.34), and conductivity (0.55), and negative loadings of DO (-0.53) and borehole depth (-0.24) (Appendix 18).

Another PCA was performed to assess the effects of land-use patterns in source protection zone-1 (SPZ-1) on groundwater chemistry (Appendix 19). The PCA plot showed that most samples with an unconfined catchment and a higher proportion of arable land coverage in SPZ-1 showed a higher TDN concentration. No further relationships between chemistry and other land-use categories were observed.

6.3.2 Spatial variation of prokaryote communities of the sandstone aquifer

Amplicon sequencing identified 28,691 unique amplicon sequence variants (ASVs), comprising 85% bacteria and 15% archaea. Shannon diversity indices had an IQR from 4.8 to 5.8. The hierarchical clustering based on the Bray-Curtis distance of the communities showed that the prokaryotic communities can potentially be classified into five clusters (Figure 6-3. A), which was found to be the optimum number of clusters using a Silhouette plot (Appendix 20). Analysis of similarity (ANOSIM) showed that the seasonal shifts in relative abundance were not significant ($R=-0.16$, $p=0.98$) in any of the sites in the five clusters. The negative R value also indicates that the spatial variation within a season was larger than the seasonal variation at each site (Chapman and Underwood, 1999). The dendrogram shows that seasonal samples from a single site had the lowest sum-squared distance (Figure 6-3. A). These results were also similar in unrarefied sequences (Appendix 21).

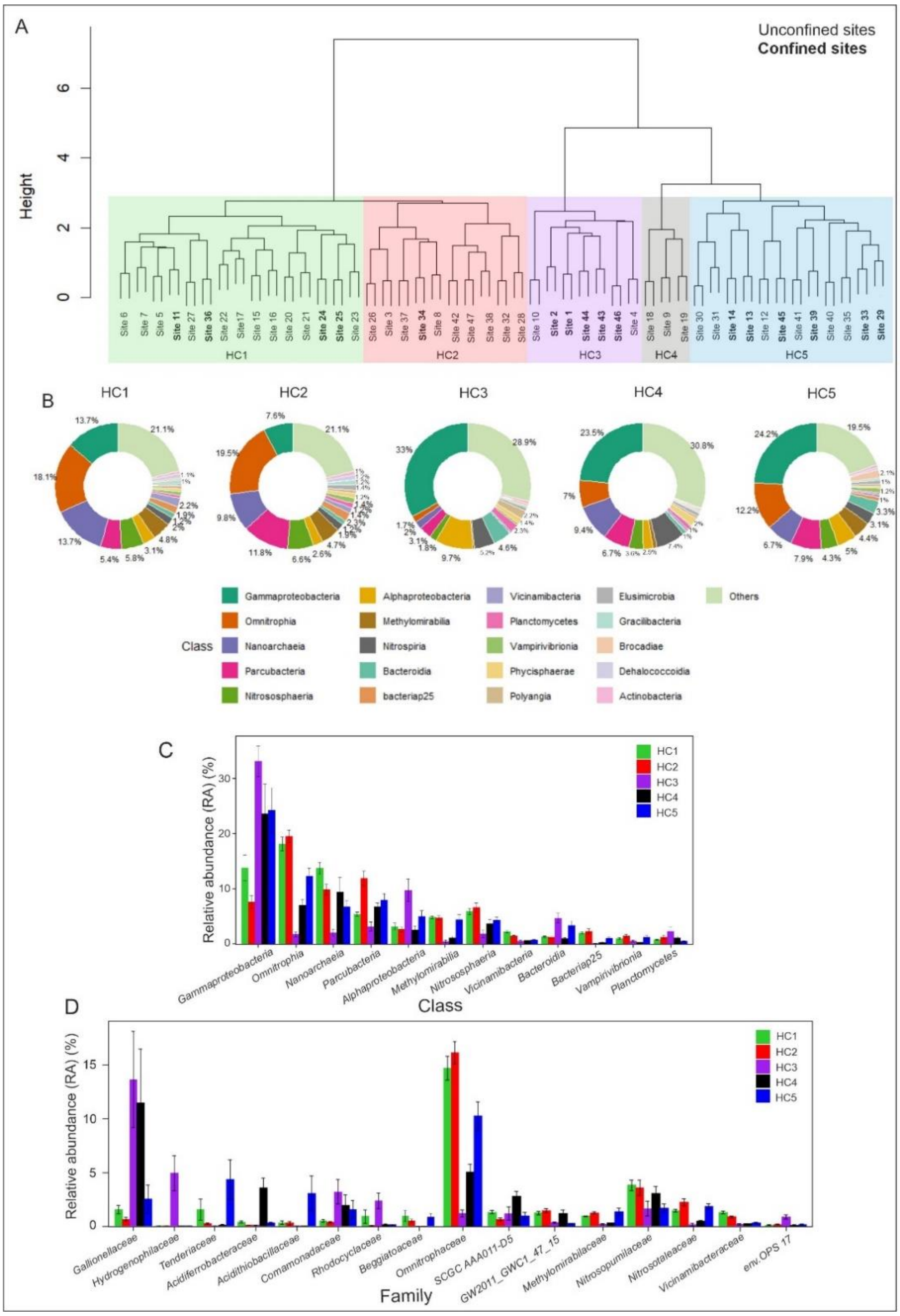


Figure 6-3. A. Hierarchical clusters (HCs) of Bray-Curtis distance of the ASV abundances, created by the Ward.D2 method, show five clusters and the number of confined and unconfined sites in each cluster. B. Donut plots of dominant prokaryotic phyla with mean relative abundance (>1%) in the five HCs. C. Bar plot shows significantly different prokaryotic classes and D. prokaryotic families across the five HCs according to LEfSe analysis, with the significance threshold being p<0.001.

The percentage of ASVs classified into a taxonomic rank was lower at lower ranks. 76.2% ASVs were classified into a phylum level, 70.8% into a class level, 54.1% into an order level, 30.1% into a family level and only 16.2% into a genus level. Functional potentials of the ASVs were assessed using FAPROTAX (Louca *et al*., 2016). However, due to low levels of genus-level assignment of ASVs, the functional assignments were deemed not reliable (Sansupa *et al*., 2021). Therefore, a reliable comparison of functional potentials in the five HCs was not possible and not reported in the results (Appendix 23).

Significant class-level and family-level differences ($p<0.001$) among clusters were identified using LEfSe analysis (Figure 6-3. C, D). Higher abundances of *Omnitrophia, Nanoarchaeia, Methylomirabilia, Nitrososphaeria, Vicinamibacteria, Bacteriap25*, and *Vampivibrionia* than other clusters characterised clusters HC1 and HC2. The main differences between HC1 and HC2 were the higher *Parcubacteria* abundance in HC2 and the higher *Gammaproteobacteria* families *Tenderiaceae* and *Rhodocyclaceae* in HC1. HC3 had the highest abundance of *Gammaproteobacteria* families *Gallionellaceae, Hydrogenophilaceae, Comamonadaceae, Rhodocyclaceae,* and *Bacteroidota* families *env. OPS 17*, *Chitinophagaceae, Alphaproteobacteria*, and *Planctomycetes* classes. The HC4 and HC5 clusters had similar abundances of *Gammaproteobacteria, Nanoarchaeia, Parcubacteria*, and *Nitrososphaeria* classes. The differences between these clusters were that the HC4 had higher abundances of *Gammaproteobacteria* families *Gallionellaceae, Acidiferrobacteraceae, Commamonadaceae*, and higher abundance of *Nanoarchaeia* families *SCGC AAA011-D5, GW2011_GWC1_47_15* and *Nitrososphaeria* family *Nitrosopumilaceae*. In contrast, HC5 had higher *Gammaproteobacteria* families *Tenderiaceae, Acidithiobacillaceae*, and families *Omnitrophaceae, Methylomirabilaceae, Nitrosotalaeceae, Chitinophagacea*.

6.3.3 Controls of environmental variables and recharge age on prokaryotic community differences

The dbRDA model parameters showed that 32% of the total variance of ASV abundance was constrained by the explanatory variables used in the analysis, with dbRDA1 and dbRDA2 axes representing 35.9% of the constrained variance. A biplot of the dbRDA1 and dbRDA2 ordinations indicated that the hierarchical clusters had separate fields of distribution and were associated with certain environmental variables (Figure 6-4. A). The dbRDA model based on the Sørensen-Dice dissimilarity matrix showed a similar result with 34.3% of the community dissimilarity constrained by the environmental variables (Appendix 22). Eleven out of fifteen sites belonged to HC1 and HC5 clusters and the ASVs had a deeper, older groundwater source with DOC, DO and TDN poor water. Except the four sites in HC4 and HC1 clusters, most of the unconfined sites belonged to HC1 and HC2 clusters, and six unconfined sites belonged to HC5 cluster and the ASVs had a shallower source, with younger groundwater and higher DOC, DO and TDN concentrations.

Among the physico-chemical variables, significant ($p<0.001$) and strong environmental fit of the constrained portion of the prokaryotic community differences were made by TDN, DO, DOC levels, along with weaker but significant ($p<0.001$) contributions of CFC-12 recharge age, overlying strata thickness, perforation depth, and temperature (Table 6-2). The Mantel test of Spearman Correlation indicated that the Bray-Curtis distance of prokaryotic communities had moderate but significant ($p<0.01$) correlation with differences in overlying strata thickness, perforation depth, TDN, DO, temperature, pH, and weaker correlation with the recharge age and borehole depth (Table 6-2). Land use coverage in SPZ-1 did not show any significant explanatory power or Mantel correlation with the differences in the prokaryotic communities. Additionally, the geographic distance between the sites did not show significant correlation with the Bray-Curtis community differences (Appendix 24).

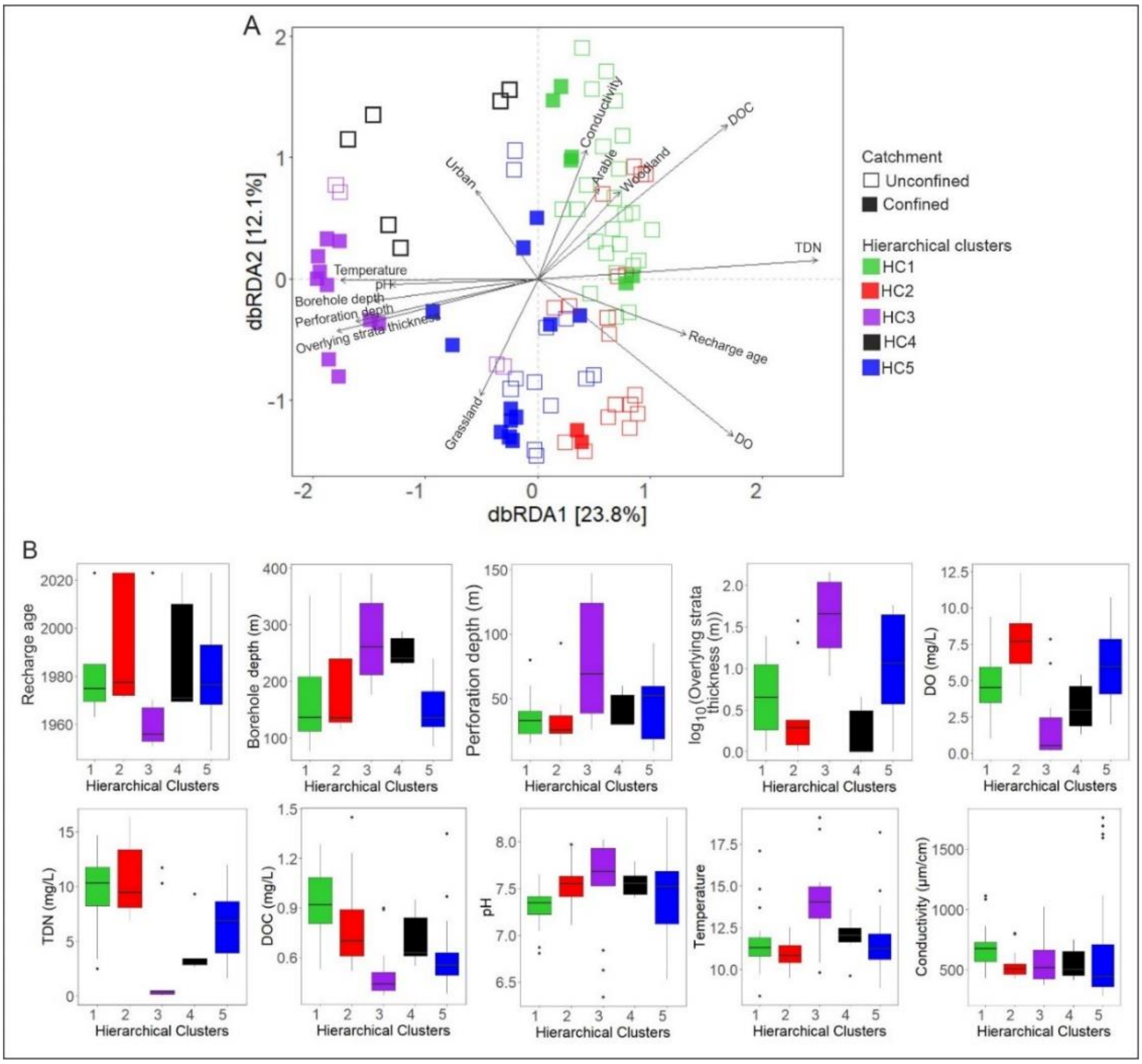


Figure 6-4. A. dbRDA plot of Bray-Curtis distances of ASVs and arrows showing loading of environmental variables along the two dbRDA axes, sites were coloured by HCs and shaped by aquifer confinement. B. Boxplots of all environmental variables across the different hierarchical clusters.

	dbRDA1	dbRDA2	Environment fit (r^2) and p-value indicators	Mantel correlation coefficient and p-value indicator
Eigenvalues	2.8	1.5		
Recharge age	0.9	-0.43	**0.16** ***	0.17
Overlying strata	-0.95	-0.3	**0.33** ***	**0.43** **
Perforation	-0.96	-0.27	**0.25** ***	**0.32** **
Well depth	-0.98	-0.15	**0.17** ***	0.17
TDN	0.99	0.07	**0.67** ***	**0.42** **
DO	0.7	0.71	**0.64** ***	**0.31** **
DOC	0.7	0.71	**0.6** ***	0.13
pH	-0.99	-0.04	0.11 **	**0.33** **
Temperature	-1.0	0.00	**0.28** ***	**0.4** **
Conductivity	0.27	0.96	0.12 **	0.11
Arable land	0.47	0.88	0.01	-0.02
Grassland	-0.35	-0.93	0.08	0.05
Woodland	0.59	0.8	0.05	-0.02
Urban area	-0.48	0.87	0.01	-0.05

Table 6-2. Table of loading of environmental variables along dbRDA1 and dbRDA2, environment fit (r^2) of each variable and Spearman correlation coefficient (ρ Value) of the Mantel test of correlation between Bray-Curtis distance and Euclidean difference matrix of environmental variables. Significance levels were *** for $p<0.001$, and ** for $p<0.01$. Significant and large r^2 and ρ values are in bold letters.

Boxplots of environmental variables in the five clusters (Figure 6-4. B) showed that the prokaryotic communities associated with younger groundwater recharge ages in clusters HC1 (IQR= 1970-1985) and HC2 (1972- 2023) were primarily present in the unconfined aquifer with shallow groundwater sources (IQR: perforation depth= 20-52 m; borehole depth= 112-276 m) and high TDN concentration (5.2-13.4 mg/L). The main differences between HC1 and HC2 clusters were that the HC2 cluster had higher DO (6.1-9 mg/L) and lower DOC (0.6-0.9 mg/L) concentration with thinner overlying strata (1-2.4m), and the HC1 cluster had higher DOC (0.8-1 mg/L) but lower DO (3.5-6 mg/L) concentration with thicker overlying strata (2-11m). Prokaryotic communities of HC5 (1967-1983) were associated with groundwater recharge ages slightly older than the HC1 and HC2 clusters, and 50% of the sites were in the confined aquifer. However, all the sites in HC5 originated from shallower depths (perforation depth = 20-57m; borehole depth= 120-168m, overlying strata thickness= 3-44 m) and the ranges of TDN (4-8.6 mg/L), DO (4-7.8 mg/L) and DOC (0.5-0.6 mg/L) concentrations were lower than HC1 and HC2 sites. The prokaryotic communities in three HC4 sites originated from the unconfined aquifer with thin overlying strata (1-4 m) but deeper groundwater sources (perforation depth = 20-52m; borehole depth = 232-276m). However, the recharge ages of two sites were older than 1972, but one of the sites showed a modern (2023) recharge age. The HC4 sites also had low TDN (2.8-3.4 mg/L), DO (1.8-4.6 mg/L) and moderate DOC (0.6-0.8 mg/L) concentrations. Oldest groundwater recharge ages were associated with the communities of HC3 cluster (1953- 1967), which were mainly composed of confined sites with deep groundwater sources (perforation depth = 38-124m; borehole depth 212-338m) with the highest overlying strata thickness (18- 116 m). The groundwater was warmer (13-14.9) and more alkaline (7.5-7.9) with the lowest TDN (0.1-0.5 mg/L), DOC (0.3-0.5 mg/L) and DO (0.2-2.4 mg/L) concentrations.

6.4 Discussion

As hypothesised, this study found that in the sandstone aquifer, bacterioplankton concentration and relative abundance of prokaryotic ASVs did not change between winter recharge and summer recession seasons. The lack of seasonal change was true for both the shallow and deeper sites, irrespective of the catchment confinement. But agreeing with the second hypothesis, it was observed that the prokaryotic communities in hierarchical clusters from deeper and confined aquifer sources were associated with nutrient-poor older groundwater recharge age, which were different from the communities in hierarchical clusters from shallower and unconfined aquifer, which were associated with younger recharge age and relatively higher nutrient concentrations.

6.4.1 Reason for the lack of seasonal change in sandstone communities

There was an absence of significant shifts in bacterioplankton concentration (Figure 6-2. A) and prokaryotic community composition (Figure 6-3. A) between recharge and recession seasons suggesting slow recharge (recharge rate ~1 m/year)) through sandstone intergranular spaces (Allen *et al.*, 1997) prevent seasonal changes in the communities. The effect of recharge depends on the distance of the groundwater from the surface, as well as local geology. Fractured aquifers (up to 60 m) with more surface connectivity revealed recharge introduced soil-derived bacteria, which started to adapt to the groundwater environment, leading to a unidirectional change in community structure (Wang *et al.*, 2025b; Yan *et al.*, 2021). Conversely, in alluvial aquifers, the shallower aquifer communities residing at about 4 m depth showed more temporal changes than deeper communities present at about 6 m depth (Danczak *et al.*, 2016b), indicating that with increasing depth, recharge effects drastically decrease in the intergranular aquifer. Additionally, recharge can impact communities of fractured aquifers more than porous aquifers due to unrestricted intrusion of allochthonous prokaryotes through larger fractures than small porous spaces (Villeneuve *et al.*, 2022). The slow recharge into the sandstone aquifer through 1 m to 142 m thick overlying strata at ~1m /year rate may not lead to changes in water quality or import allochthonous prokaryotes into the aquifer within one season. In fact, 75% of the groundwater had recharge ages of 1994 or older, and only 11 samples had recent groundwater recharge age, indicating that most of the groundwater recharge occurs very slowly over many years.

Notably, 11 sites showed recent recharge ages (2023) despite showing no signs of seasonal shifts in prokaryotes, nutrient concentrations, or physiochemical properties. Recent recharge ages can be due to the presence of rapid recharge pathways along fractures (Cronin *et al.*, 2003), or due to sampling or analytical error, although these possibilities were not tested and assessing the role of preferential recharge paths in seasonal sampling was not within the scope of this study.

The lack of seasonal sampling repetition was a key limitation of the study. While the peak to post-recharge changes in groundwater prokaryotes were not statistically significant, it is possible that with the slow recharge in sandstone aquifer, the changes in groundwater prokaryotes may occur slowly. It is essential to perform annual samples in recharge and recession period over multiple years (example Yan *et al.* (2021) and Wang *et al.* (2025b)) to observe the prokaryotic dynamics.

6.4.2 Relation of prokaryotes and recharge age

The spatial variation of prokaryotic communities revealed clear patterns linked to the groundwater recharge age and hydrogeological conditions. The dominant classes, such as *Gammaproteobacteria, Alphaproteobacteria, Nitrospirota*, and *Planctomycetes*, especially in the HC3, HC4 and HC5 sites with deeper and older water, highlighted the reliance of the ecosystem at these sites on lithoautotrophic processes and specialised nutrient cycling, including iron and nitrogen turnover (Ben Maamar *et al*., 2015; Mosley *et al*., 2022; Wegner *et al*., 2019) (Appendix 23). A large fraction of the communities in the shallower and younger groundwater of the HC1, HC2, HC4 and HC5 sites were composed of classes *Omnitrophia, Parcubacteria,* belonging to Candidate phylum radiation (CPR) bacterial superphylum and *Nanoarchaeia,* belonging to the and DPANN archaeal superphylum (Castelle *et al*., 2017; Gios *et al*., 2023; Seymour *et al*., 2023). Members of these classes are known to perform cryptic carbon and nitrogen cycling in very low-nutrient groundwater environments and are capable of heterotrophic metabolism by living symbiotic or parasitic lifestyles (Castelle and Banfield, 2018; Chaudhari *et al*., 2024; Gios *et al*., 2023; Mehrshad *et al*., 2021). A conceptual model to represent the differences of the prokaryotes is given in Figure 6-5. These prokaryotic community similarities were not related to their spatial proximity, nor did the recharge age or chemistry follow a spatial gradient (Appendix 24).

The sites belonging to the HC2 (Figure 6-5) had the youngest groundwater recharge age (IQR=1972- 2023) and the highest TDN and DO concentration, indicating a higher influence of the recharge process on the communities. The easier migration of surface water could be due to thinner overlying strata (1-2.4m) and shallow groundwater source via perforation depth (23-36m) (Figure 6-4. B). The higher DO concentration in the HC2 sites could be due to the lack of DOC levels for heterotrophic uptake of oxygen. The typical oligotrophic groundwater supported the prokaryotic assemblage dominated by ultrasmall heterotroph classes *Omnitrophia, Parcubacteria, Nanoarchaeia* with cryptic carbon and nitrogen cycling potentials and potential of parasitic lifestyle (Beaver and Neufeld, 2024; Gios *et al*., 2023; Seymour *et al*., 2023;

Tian *et al.*, 2020). Dominance of the typical nitrifying family *Nitrosopumilaceae* and the ammonia-oxidising family *Nitrosotalaeceae* highlights the importance of nitrogen oxidation and nitrogen cycling in the shallow sandstone aquifer (Wegner *et al.*, 2019). Interestingly, soil variety *Parcubacteria* undergoes genetic streamlining and loses almost half the size of genetic material while travelling into an oligotrophic groundwater environment resulting in a symbiotic lifestyle (Chaudhari *et al.*, 2024). The abundance of *Parcubacteria* therefore may further strengthen the potential of comparatively higher surface influence and younger age in this cluster.

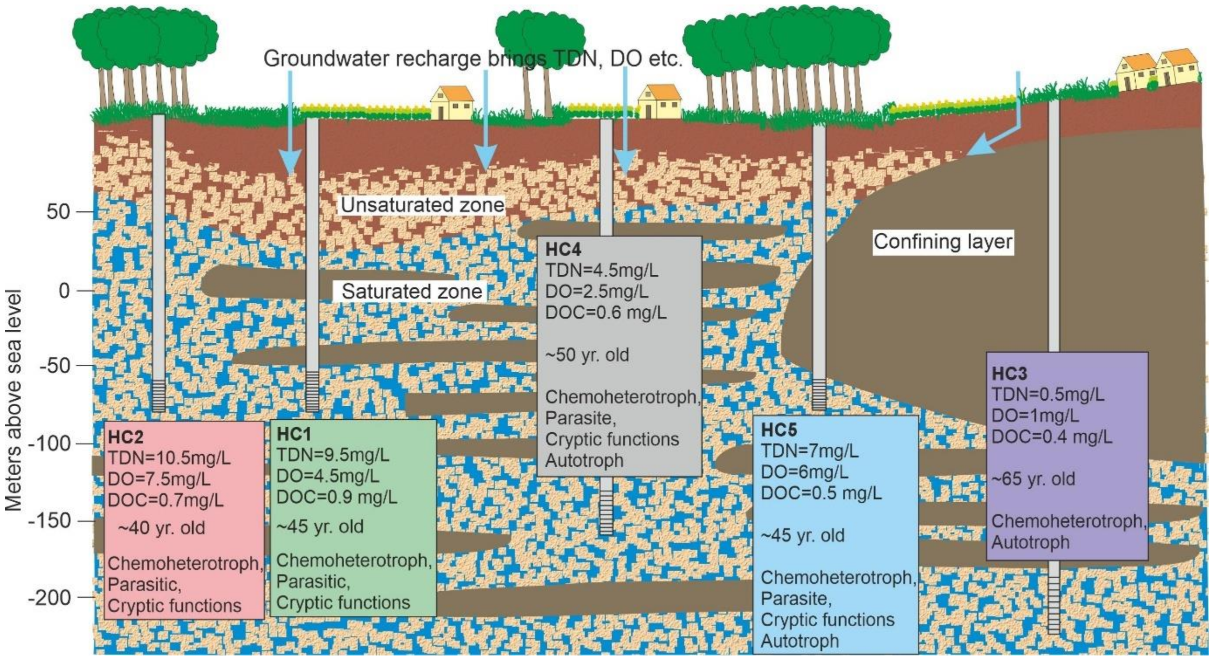


Figure 6-5. Conceptual model illustrating groundwater prokaryotic assemblages in groundwater of Permo-triassic sandstone aquifer where each borehole is a representative of the hierarchical clusters and median concentrations of groundwater chemistry, median recharge age and dominant prokaryote types are given in boxes, maroon top layer is the soil, orange layer is intergranular aquifer with groundwater in intergranular spaces and brown layers are impermeable layers.

Sites belonging to the HC1 cluster (recharge age IQR=1970- 1985) represented the communities which had the closest resemblance to the HC2 communities but also had commonalities with older groundwater assemblages (Figure 6-5). These sites, dominated by unconfined aquifer sites, possessed thicker overlying strata (2-11m) but similar perforation depths compared to the HC2 sites, which may explain the older recharge age due to the longer travel distance of groundwater from the surface to the saturated zone of the aquifer. However, DOC concentration was the highest in this cluster, which could explain the lower DO concentration owing to heterotrophic uptake. Despite lower surface connectivity, the higher DOC concentration in the HC1 sites may indicate the presence of an in situ DOC source (Geesink *et al.*, 2022; Shen *et al.*, 2015). Like the HC2 assemblages, these sites had dominant ultrasmall heterotrophs, but like older water, also possessed abundant *Gallionellaceae* and *Rhodocyclaceae* (Ben Maamar *et al.*, 2015; Jakus *et al.*, 2021).

The HC3 cluster had the oldest groundwater recharge age (IQR=1953- 1967). The deep confined sites with the unique HC3 prokaryotic assemblage (Figure 6-5) were dominated by iron-oxidising autotroph *Gallionellaceae* and sulphur-oxidising mixotroph *Hydrogenophilaceae*, along with the presence of denitrifying *Rhodocyclaceae, Comamonadaceae*, ammonia-oxidising *Nitrosopumilaceae* and anaerobic carbon processing *env.OPS 17* (Ben Maamar *et al.*, 2015; Jakus *et al.*, 2021; Wegner *et al.*, 2019). *Commamonadaceae* and *Planctomycetes* are known for their ability to degrade complex carbon sources as chemohetetroptrophs (Ben Maamar *et al.*, 2015). In the oldest part of the aquifer, which has been cut off from DO and TDN replenishment from the surface, chemolithoautotrophs can potentially sustain complete redox cycles of nutrients along with their attached counterparts, as expected in the deep isolated groundwater in the fractured aquifer and deep confined pristine aquifers (Ben Maamar *et al.*, 2015; Flynn *et al.*, 2013). As a result of minimal recharging-related perturbations over a long period of time, the community may have adapted to efficiently cycle the available nutrients and make available new redox species for each other.

The HC4 and HC5 sites had many similarities and represented a transition between young and older groundwater ages (Figure 6-5). These two clusters possessed intermediate groundwater ages (IQR=1967- 2010), indicating lower influence of surface water than the HC1 and HC2 sites but more influence than the HC3 sites. The entry of recharge water into these sites can be slower due to the marl intercalations at the deep unconfined HC4 sites and the thick overlying strata (3.7-44m) overlying both confined and unconfined HC5 sites. Lower DO and higher DOC in the HC4 sites compared to higher DO and lower DOC in the HC5 sites (Figure 6-4. B) might be related to the fact that the presence of DOC limited heterotrophic utilisation of DO in the HC5 sites. Moreover, higher DOC concentrations in the deeper HC4 sites with limited surface connectivity further strengthened the speculation of an in situ organic carbon source within these sites, but not in the sites belonging to HC5. In the TDN and DO poor and DOC-rich HC4 sites, the dominant autotrophic taxa indicated the community's functional potential to perform redox cycling of Fe, S, and N by *Gallionellaceae*, *Acidiferrobacteraceae*, ammonia oxidation by *Nitrosopumilaceae,* anaerobic denitrification and complex carbon degradation by *Comamonadaceae* (Jakus *et al*., 2021; Mosley *et al*., 2022; Umezawa *et al*., 2016; Wegner *et al*., 2019). The higher TDN and DO-rich and DOC-poor HC5 sites also possessed autotrophic Fe and S cycling, with methane oxidation potential indicated by the abundance of *Acidithiobacillaceae and Methylomirabilis* (Herrmann *et al*., 2015; Ludington *et al*., 2017; Mosley *et al*., 2022). Both the HC4 and HC5 sites possessed similarities to younger water assemblages belonging to the HC1 and HC2. Like younger water, ultrasmall prokaryotes within classes *Omnitrophia, Parcubacteria, Nanoarchaeia* were also present in the HC4 and HC5 sites.

6.4.3 Study implications, limitations and future work

An interesting point prevailed in the study was the variation of prokaryotic communities along the TDN concentration gradient (Figure 6-4. A), which was hypothesised to be an artefact of recharge in Chapter 5. In the groundwater of England, majority of nitrogen is nitrate, with trace amounts of other inorganic nitrogen species (Stuart and Lapworth, 2016). In England, nitrate contamination of groundwater from both diffuse and point sources has been a major water quality concern (Rivett *et al*., 2007). The nitrate in sandstone aquifers is stored in the unsaturated zone and gets slowly released into the groundwater with downward recharge water movement (Wang *et al*., 2012). The nitrate arrival into the water table started increasing around 1945, peaked and reached a plateau in the 1990s (Wang *et al*., 2012). In this study, the recharge age and TDN concentration were found to be related. In the HC3 sites with a mean groundwater recharge age of more than 65 years, TDN concentration was the lowest, followed by increasing TDN concentration as the groundwater recharge age decreased from 50-45 years in HC4, HC5 and less than 45 years in HC2 and HC1 (Figure 6-5). The gradual change in prokaryotic communities along the TDN gradient indicated that anthropogenic nitrate contamination had a strong influence. The presence of some autotrophic taxa may suggest that at some sites, the potential for nitrate removal is higher than at others. For example, the gradual increase in relative abundance of chemoautotrophs such as *Gallionellaceae* and *Rhodocyclaceae* in HC2, HC4 and HC5 could be related to denitrification coupled with Fe and S oxidation functions of these prokaryotes (Jakus *et al*., 2021). Thus, nitrate concentration and prokaryotic community structure in groundwater should be more closely monitored to test the impact of anthropogenic nitrate pollution on the groundwater ecosystem and the nitrate remediation potential of these ecosystems. Further, Harris *et al*. (2025) demonstrated that even within minimally polluted groundwater in a limestone aquifer, the ‘baseline’ bacterial communities show significant spatial variation depending on water chemistry, demonstrating that ‘baseline’ is not a single fixed state but a constrained envelope of “natural variability”. If the groundwater ecosystem and nitrate in the minimally polluted groundwater sources of this study are continuously monitored over a long time, potentially this range of “natural variation” can be differentiated from nitrate contamination related variations.

This study also showed that prokaryotic community variation depends on the hydrological conditions of the consolidated sandstone aquifer. Ben Maamar *et al*. (2015) showed that the depth of the sample is less important than the hydrogeological flow loops and how far those loops can bring surface water into a fractured rock. In the case of sandstone aquifers, some conclusive patterns of increasing recharge age with a thicker overlying strata and deeper groundwater sources were observed. This is possible because of the low vertical hydraulic conductivity (0.07-1.18 m/d) in this aquifer. However, outliers in each HCs indicated that we simplified the local hydrogeology to explain regionally extensive patterns. Local fractures may provide pathways for surface water input in the deeper parts of the aquifer (Sorensen *et al*., 2021). Similarly, fractures in deeper sandstone can allow the lateral dispersal of microbes and nutrients. These scenarios can cause deeper confined sites to possess prokaryotes, similar to shallower unconfined sites. These local variables could not be wholly accounted for in this study. Additionally, it should be emphasized that the hierarchical clustering (HC) approach applied here is inherently exploratory. It helped reveal groups of sites with similar community which in turn had chemical and age similarities. But it does not define an absolute five of clusters for this sandstone aquifer. With larger sample sets or under different hydrogeological settings, sites may not always be classified into five clusters.

The prokaryotes present at different sandstone sites had unique metabolic functions, which can control the local biogeochemical transformation of elements. However, these functions were not well described using FAPROTAX. The FAPROTAX algorithm uses known prokaryotic genus to assign functional potentials (Louca *et al*., 2016). However, given that the genus-level assignment of the ASVs was very poor (only 16.2%), the functional assignment was not reliable (Appendix 23), something that was also observed in Chapter 5 (Section 5.3.4). This resonates with the observation of Couton *et al*. (2023), which emphasised the lack of groundwater reference sequences in global databases. More meta-omics approaches should be applied to analyse the groundwater prokaryotic genome sequences and the databases should be updated with the prokaryotic taxonomic classification and functions.

6.5 Conclusion

Groundwater eDNA metabarcoding of the sandstone aquifer revealed that groundwater recharge age impacted the spatial variation of the planktonic prokaryotic communities. While seasonal recharge showed no impact, the long-term recharge age of groundwater controlled the nutrient concentrations, which in turn controlled the prokaryotic communities. By using a clustering approach on the prokaryotic ASVs, similar community structures were related to similar groundwater recharge ages and similar chemistry. The strong impact of total dissolved nitrogen (TDN) concentration on the communities was related to nitrate input, and thus, the sign of anthropogenic pressure was observed on the prokaryotic communities. The shallow zones of the aquifer were generally rich in nutrients, harbouring heterotrophic and parasitic organisms, and as the groundwater gets older and nutrient concentrations get depleted in the deeper zones of the aquifer, the communities transition towards autotrophy-dominated. The application of amplicon sequencing has expanded our understanding of microbial diversity in sandstone aquifers, but remains limited in assessing functional traits. As the number of national-level studies on groundwater microbiomes is increasing globally, the groundwater recharge age should be considered as an important controlling factor of the spatial variation of the prokaryotes.

7. Concluding discussion

This chapter addresses thesis **Objective 5:** to identify knowledge gaps, study limitations and future directions to characterise groundwater prokaryotic ecosystems in other parts of the world.

7.1 Overview of prokaryotic microbiology of different aquifers of England

This thesis investigated planktonic prokaryotic ecosystems within three hydrogeologically distinct aquifers (Permo-Triassic sandstone, Cretaceous chalk, and Jurassic limestone) using a combination of flow cytometry and eDNA metabarcoding (Figure 7-1). The results revealed that, on the national-level, different aquifer types possess different bacterioplankton abundance, prokaryotic diversity, taxonomy and functional potentials as well as different controls on the spatio-temporal variation of prokaryotic communities (Figure 7-2). The key findings obtained by addressing the thesis objectives were:

Objective 1: **To optimise prokaryotic sample collection and analysis methods for groundwater systems with low prokaryotic concentration.**

Key outcomes:

- For flow cytometric analysis of bacterioplankton concentration, the sample staining condition was chosen to be incubation at 22ºC for 10 minutes, and this condition was applied for all the samples analysed for this thesis.
- For eDNA filtration, Sterivex™ filters with Polyethersulfone (PES) membrane and 0.22 µm pore size (Merck KGaA, Darmstadt, Germany) were selected, and for eDNA extraction from the filters, Qiagen DNeasy® PowerWater® Kit was selected and used for all the eDNA sample collection and extraction. The optimum filtration condition was selected to be 15 L filtered water or 45 minutes of filtration, whichever was earliest.

Objective 2: **To assess the controls on the spatial and seasonal variation of bacterioplankton concentrations in the three different aquifers.**

Key outcomes:

- Spatially, the bacterioplankton total cell concentration (TCC) was variable in different geologies. The median TCC were about two times higher in the karstic limestone than in the dual-porosity chalk or intergranular sandstone. The karstic aquifer with higher surface connectivity and frequent allochthonous bacterial input had higher TCC than the intergranular and dual porosity aquifers, which are more likely to filter out a substantial proportion of allochthonous prokaryotic input.
- Seasonally, only the chalk aquifer showed ~50% dilution of TCC upon groundwater recharge, potentially due to water containing low cell counts being flushed down from the unsaturated zone under piston-pressure. TCC at the karst and sandstone aquifer sites remained stable due to missing recharge signatures and a dry winter, warranting the collection of more frequent data over a more extended period of time.

Objective 3: To assess the impact of aquifer geology on the spatial variation of groundwater planktonic prokaryotic community composition.

Key outcomes:

- The spatial variation in prokaryotic community compositions appeared to be strongly related to aquifer geology. The ferro-silicate sandstone aquifer had a groundwater prokaryotic community with a higher mean relative abundance of *Gammaproteobacteria, Bacteroidia* and *OM190.* The carbonate chalk and limestone aquifers had similar prokaryotic communities, with a higher abundance of *Omnitrophia*, *Nanoarchaeia* and *Dehalococcoidia,* and the main difference between the two aquifers was a higher abundance of *Gracilibacteria* and *ABY1* in chalk*, a*nd a higher abundance of *Saccharimonadia* in limestone.

- The prokaryotes in the sandstone aquifer are known to have large genetic material capable of diverse metabolic functions. The carbonate chalk and limestone aquifers hosted ultrasmall prokaryotic communities, with smaller genetic material and reduced streamlined metabolic functions. This aligned with the flow cytometric results showing a higher proportion of high nucleic acid bacteria (HNA) in the sandstone aquifer and a higher proportion of low nucleic acid bacteria (LNA) in the chalk and limestone aquifer.
- Inside chalk and sandstone aquifers, concentration gradients in total dissolved nitrogen (dominated by nitrate) and thickness gradients of overlying strata controlled the spatial variation of the prokaryotic communities. However, known prokaryotes with denitrification potentials were not found, indicating little potential for anthropogenic nitrate attenuation in the aquifers.
- Large numbers of unclassified amplicon sequence variants (ASVs) and unknown functional potentials indicated a substantial gap in reference datasets of prokaryotic genetic sequences.

Objective 4: To assess the impact of recharge age on the spatial variation of prokaryotic community composition within a single major aquifer type.

Key outcomes:

- Low groundwater recharge rates in the sandstone aquifer may have resulted in a seasonally unchanged prokaryotic community composition.
- Over a decadal period, groundwater recharge in shallower aquifers with thinner overlying strata resulted in higher TDN and DO concentrations, and in deeper parts of aquifers covered by thicker overlying strata, as the recharge age increased, the TDN and DO concentrations declined. In deepest and oldest groundwater, TDN and DO concentrations were minimal. The prokaryotic communities varied along the gradient of nutrient concentrations, which was controlled by groundwater recharge age.
- The prokaryotes in young groundwater clusters (~40 years) were dominated by ultrasmall chemoheterotrophs and parasites, followed by a transition between

chemoheterotrophic and parasitic-dominated communities to autotrophic-dominated communities in older groundwater (45-50 years), and an autotroph-dominated community in the oldest (~65 years) groundwater samples. Prokaryotes known for chemoheterotrophic metabolism were present in samples with all ages.

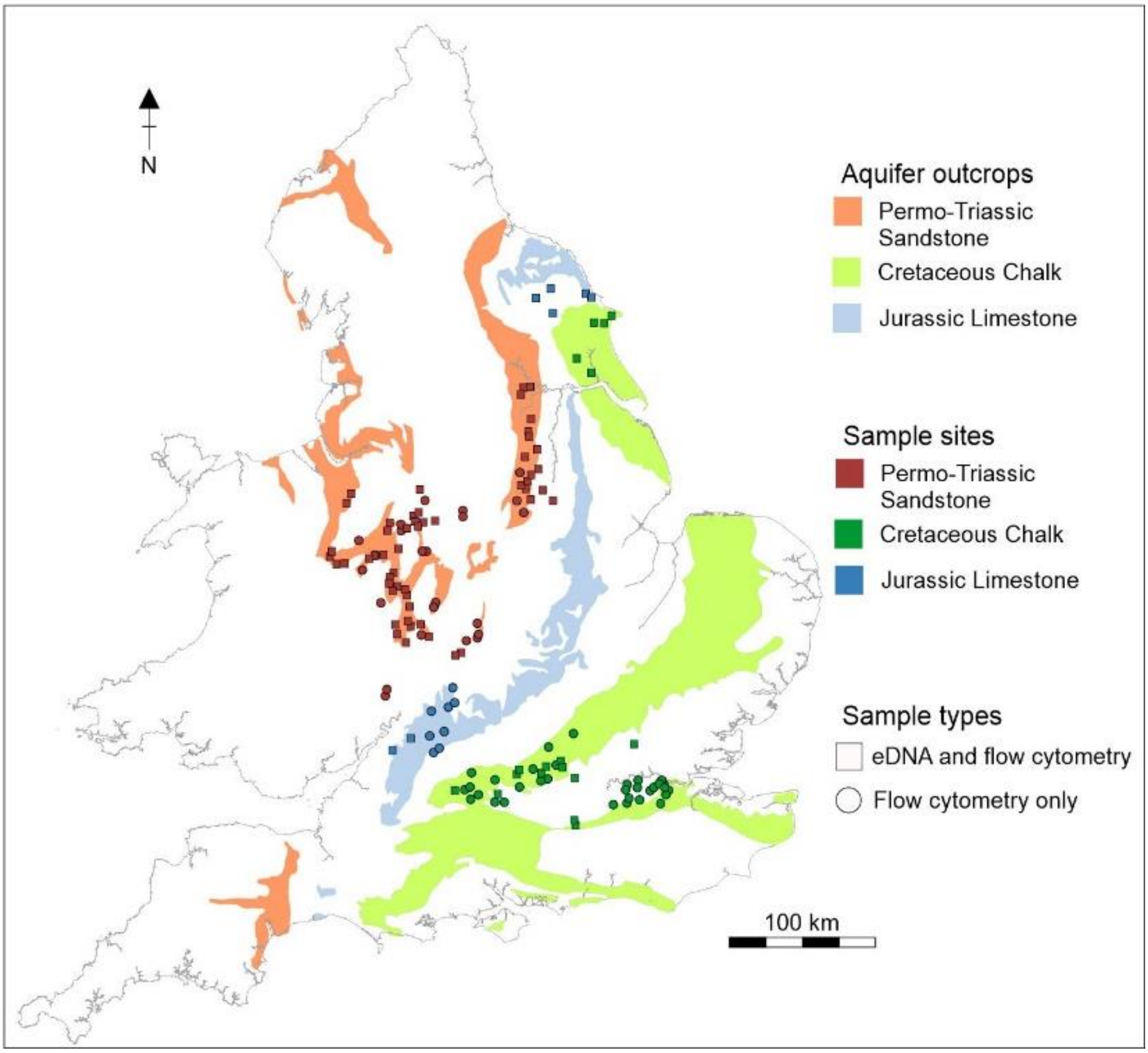


Figure 7-1. Study area in England showing the outcrops of three major aquifers and points showing the sample sites, point colours showing sampled aquifer and point shapes showing sample types, i.e., both eDNA and flow cytometry sample collection, or only flow cytometry sample collection.

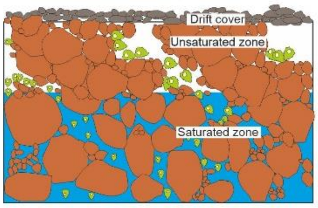
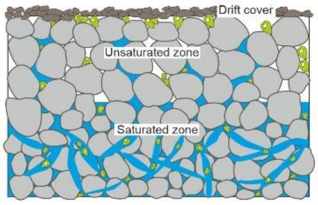
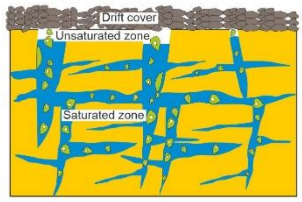
	Permo-triassic Sandstone	Cretaceous Chalk	Jurassic Limestone
Conceptual Figures			
TCC	5.7×10^3 to 1.8×10^4 cells/mL	5.8×10^3 to 2×10^4 cells/mL	1.5×10^4 to 3.1×10^4 cells/mL
%HNA	15.8 - 26.7	12.2 - 17.2	11.8 - 25.9
%ICC	17.6 - 29.1	16.5 - 42.7	18 - 40
Shannon diversity index	4.8 - 5.8	6.1 - 6.7	6.5 - 6.7
Dominant taxonomic classes (in order of decreasing mean abundance)	*Gammaproteobacteria, Omnitrophia, Nanoarchaeia, Parcubacteria, Nitrososphaeria, Methylomirabillia, Alphaproteobacteria, Nitrospira, Bacteroidia, Vicinamibacteria, Bacteriap25, Brocadiae, Planctomycetes, Polyangia.*	*Omnitrophia, Nanoarchaeia, Gammaproteobacteria, Parcubacteria, Nitrososphaeira, Methylomirabillia, Alphaproteobacteria, Nitrospiria, Vicinamibacteira, Gracilibacteria, Dehalococcoidia, Bacteriap25, Brocadiae, ABY1.*	*Omnitrophia, Nanoarchaeia, Gammaproteobacteria, Parcubacteria, Methylomirabillia, Saccharimonadia, Nitrososphaeira, Bacteroidia, Alphaproteobacteria, Vicinamibacteira, Bacteriap25, Dehalococcoidia, Brocadiae, Gracilibacteria, Nitrospiria,*
Functional potentials (in order of decreasing mean abundance)	Chemoheterotrophy, Aerobic chemoheterotrophy, Anaerobic chemoheterotrophy, Methylotrophy, Hydrocarbon degradation, Methanotrophy, Dark iron oxidation, Nitrification, Nitrite respiration, Nitrogen respiration, Nonphotosynthetic cyanobacteria.	Chemoheterotrophy, Aerobic chemoheterotrophy, Anaerobic chemoheterotrophy, Methylotrophy,Hydrocarbon degradation, Methanotrophy, Methanotrophy, Nitrite respiration, Nitrogen respiration.	Chemoheterotrophy, Aerobic chemoheterotrophy, Anaerobic chemoheterotrophy Animal parasites or symbionts.
Environmental controls	TDN, DOC, DO concentrations, water depth, Overlying strata thickness, Confinement, Recharge age	DOC, TDN concentration, Overlying strata thickness, seasonal recharge	(More samples needed)

Figure 7-2. Planktonic prokaryotic microbial community composition in three distinct aquifer types of England, showing conceptual figures of the aquifer geologies, and ecosystem parameters showing interquartile ranges of flow cytometric TCC, %HNA, %ICC, Shannon diversity index, decreasing order of dominant taxa and dominant functional potentials with mean abundance over 1% in an aquifer and major environmental factors that control the spatio-temporal variation of the groundwater prokaryotes.

In all three aquifers, samples were collected for either only bacterioplankton concentration analysis or for both concentration and eDNA-based analysis (Figure 7-1). In some previous regional studies, the geographic proximal sample locations revealed similar community composition (Griebler *et al*., 2010; Stein *et al*., 2010), which was not the case in the research reported in this thesis. Even if the sampling sites within sandstone, chalk and limestone aquifers were geographically relatively close in Northeast England, the communities were still different in the different geologies. Moreover, the sample locations from similar geologies, but separated by large distances, still showed the same community compositions. For example, in the Permo-triassic sandstone units of the Midlands and East Midlands had similar bacterioplankton TCC, %HNA and %ICC, as well as the hierarchical clusters of prokaryotic communities were not always in spatial proximity and Distance-decay plot also suggested community differences did not possess a significant correlation with spatial difference (Appendix 24). Similarly, chalk aquifers of the Southeast and Yorkshire, and limestone aquifers of the Cotswolds and Yorkshire had similar TCC, %HNA and %ICC range. In the dbRDA plot, despite their geographic separation, chalk and limestone samples always had different community structures compared to the sandstone aquifer (Figure 5-3. B). Although the Distance-decay plot showed a weak correlation between geographic distance and community dissimilarity in a national scale (Appendix 15), the correlation was weaker than that with the chemical and borehole depth variations (Table 5-1). This observation has important implications for groundwater ecosystem health management. Hose *et al.* (2023) suggested that effective groundwater ecosystem health management strategies require classification of the management areas. Since this study established that aquifer geology exerts significant control over the prokaryotic ecosystems, on a national scale, aquifer geologies should be considered as different management areas for future management strategies. Within a single aquifer type in a region, groundwater bodies with similar recharge age and similar nutrient concentrations can be used as ecosystem management zones. This strategy can be implemented instead of classifying the management areas based on the geographic proximity of groundwater bodies.

7.2 Key contributions to the current state of research

Through addressing Objectives 2 to 4 as described in Section 1.2, the thesis has made the following contributions to the current state of knowledge regarding groundwater microbiology.

First, the thesis generated novel national-scale groundwater prokaryotic community data in England, UK. Only two national-scale studies were reported from New Zealand (Sirisena *et al*., 2018) and China (Zhong *et al*., 2023), where eDNA sequencing was used to explore groundwater prokaryotes. This scarcity of national datasets can be an obstacle towards effective groundwater source management strategies, where the ecosystem services provided by the spatio-temporally variable prokaryotes are considered for decision making. This thesis has made a significant contribution to the growing number of large-scale studies that are currently taking place in Australia (Korbel *et al*., 2024; Smith *et al*., 2018), New Zealand (Abraham and Close, 2024; Sirisena *et al*., 2018), China (Zhong *et al*., 2023), the US (Merino *et al*., 2022) and many European countries (Amalfitano *et al*., 2014; Benk *et al*., 2019; Couton *et al*., 2023). By optimising modern technology for low-biomass groundwater systems, important ecosystem health indicators were measured.

Second, in national surveys, aquifer geology is not generally considered a potential ecosystem classification parameter for groundwater management. However, for effective groundwater ecosystem management, classifying different groundwater ecosystems is crucial (Hose *et al*., 2023). This thesis established that on a large spatial scale, such as a national scale, different aquifers can be treated as unique microbiomes with unique environmental drivers. The prokaryotic concentration difference in the three aquifers depended on the degree of surface connectivity. The prokaryotic taxonomic and functional assemblages were different in the three aquifers depending on the respective aquifer mineralogy, as well as nutrient replenishment controlled by the surface connectivity of the aquifer. In the future, national studies should include aquifer types based on their mineralogy and surface connectivity during survey design.

Third, depending on the recharge mechanism within each aquifer, the temporal dynamics of prokaryotes were found to be different. In the chalk aquifer, where recharge predominantly occurs by piston pressure, bacterial concentration underwent dilution after recharge. In the sandstone aquifer, where slow recharge occurs through intergranular spaces, seasonal change in prokaryotes was not noticeable, but the communities shifted over multiple decades, as was observed by the relation of communities with the recharge age. This temporal shift has not been studied in large-regional studies previously. Besides, the effect of recharge age on the prokaryotic community structure has been studied once in a fractured aquifer before (Ben Maamar *et al.*, 2015). The critical message revealed in the current thesis is that the temporal dynamics of groundwater prokaryotes depended on aquifer type.

Fourth, in this study, the groundwater samples were free of microbial and heavy chemical contaminations that would make the water unusable as drinking water, making this an ideal first step to define reference microbiology for future groundwater monitoring and management. Most of the microbial studies in the English aquifers have been performed in heavily contaminated sites, with chemical and microbial contaminants, while contaminant-free groundwater microbiology has been largely ignored (Gregory *et al.*, 2014). Assessment of a "baseline ecosystem" would ideally require pristine groundwater samples during new borehole construction, assuming no anthropogenic signature (Zhong *et al.*, 2023). Since the data in this thesis originates from boreholes constructed decades ago, and most of the groundwater resources of England are contaminated by agricultural nitrate (Rivett *et al.*, 2007), the community compositions cannot be truly considered a "baseline community" from a "pristine" source. The current dataset on bacterioplankton concentration and community composition from less contaminated groundwater can be used as a modern monitoring reference for the drinking water sources. Using a modern reference from less contaminated sources was proposed for groundwater chemistry assessment by Edmunds *et al.* (2003). The reference was different depending on the aquifer type and can be used for future detection of changes in microbial water quality due to heavy abstraction, emerging contaminants or climate change.

Fifth, this study produced important microbial community composition datasets from a number of globally understudied aquifers, specifically, chalk and sandstone aquifers. The chalk aquifer is one of the most productive aquifers in the UK and northwestern Europe (Gunn *et al.*, 1995), yet microbial data from this aquifer are scarce. Sandstone aquifers are globally essential and the most productive aquifers (Van der Gun, 2022).

Previous research on eDNA-based studies of prokaryotic communities of the chalk aquifer was not found, except for some flow cytometric studies. Although shallow, unconsolidated sandstone aquifers are commonly studied (Abraham and Close, 2024; Couton *et al.*, 2023; Korbel *et al.*, 2024), deeper consolidated sandstone aquifer communities have been rarely studied, especially in places without high levels contamination (Hazen *et al.*, 1991; Korbel *et al.*, 2024; Pellizzari *et al.*, 2016). This thesis created large-scale prokaryotic community datasets from these important aquifer types. The datasets stored in public databases can be used for comparison of the prokaryotes in similar aquifers globally.

Finally, background research showed that groundwater prokaryotic studies are often performed on shallow aquifers (commonly less than 50 m) or spring sources (Couton *et al.*, 2023; Fiedler *et al.*, 2018; Griebler *et al.*, 2010; Korbel *et al.*, 2024; Reiss *et al.*, 2019; Stegen *et al.*, 2018; Stein *et al.*, 2010). These are easy to sample using open boreholes or active springs in the field. However, two limitations can arise from sampling only shallow sources. The shallow aquifers are close to the surface and thus are rich in recharge-related nutrients. Hence, they may not represent communities in aquifers deeper than 50 m, which may not be enriched in recharge-related nutrients (Pedersen *et al.*, 2008). Second, if the boreholes are not flushed properly, the community studied can be prokaryotes growing near the boreholes, rather than prokaryotes within the aquifer far from the boreholes (Korbel *et al.*, 2017; Sorensen *et al.*, 2013). However, in this thesis, through industrial collaboration and using regularly pumping boreholes, a new possibility of sourcing groundwater samples was explored. This allowed for the study of groundwater sources from both shallow (springs) and deeper (up to 391 m) sources, and regularly pumping boreholes provided samples from within the aquifer. In future studies, such collaborative opportunities can be explored.

7.3 Study limitations and future directions

The research is not without its limitations, and acknowledging them should pave the path for further research.

A balanced number of samples from all three aquifers under consideration would be ideal for comparison of groundwater prokaryotic ecosystems. However, the sandstone aquifer had more samples than chalk and limestone aquifers, especially for eDNA amplicon analysis. To mitigate erroneous interpretations arising from imbalanced sample sizes, boxplots were used to ensure that the interquartile ranges of flow cytometric variables were small in limestone samples in Chapter 4 (Figure 4-2) and both chalk and limestone samples (Figure 5-2). In Chapter 5, the limestone samples also showed a low Bray-Curtis difference from each other, and the dbRDA showed that the limestone samples had a tight field of distribution (Figure 5-3. B). Both chalk and limestone samples had no overlap with the field where most of the sandstone samples were plotted (Figure 5-3. B). Therefore, despite the unbalanced sample size, the differences between aquifer categories were considered reliable.

In the future, more flow cytometric and eDNA amplicon data from limestone and chalk aquifer samples should be collected to find appropriate reference conditions of groundwater ecosystems in these two aquifers. With higher sample counts from chalk and limestone aquifers, it will also be possible to use a similar clustering approach as in Chapter 6 to reveal ecosystem classes within the aquifers.

Although the sampling in Chapters 4 and 6 was performed in typical groundwater recharge and recession seasons to analyse temporal differences in the aquifer communities, in each case, the sampling period did not match with actual groundwater recharge events. The sampling campaigns were pre-planned based on the historical data on the typical groundwater recharge season. Groundwater levels in monitoring boreholes showed that only in the chalk aquifer, groundwater recharge caused a reduction in bacterioplankton TCC. In Chapter 6, the groundwater level dropped from recharge to recession period in the unconfined sandstone sites, although this did not change the TCC or taxonomy. In the limestone aquifer, the groundwater level rise seemed to occur two months before the sampling season, so possibly the recharge response may have been missed (Chapter 4). Additionally, the seasonal samples were not repeated, restricting our understanding of whether the same seasonal patterns arise in every season.

A future study in both sandstone and limestone aquifers is required by integrating data from atmospheric rainfall, groundwater level change and prokaryotic community change to detect these temporal shifts. A high-frequency online flow cytometric study in chalk aquifer (e.g., Sorensen *et al.* (2018)) may reveal the reason for the temporal TCC drop observed in Chapter 4, and may also show the process of TCC recovery. Similar high frequency study in sandstone and limestone aquifers can also reveal the temporal fluctuation of TCC in these aquifers. Moreover, monthly eDNA analysis of the sites could be undertaken to appropriately observe how the microbiology of each aquifer site responds to recharge and recession periods. The temporal studies should also be repeated over multiple years, (example Yan *et al.* (2021)) to observe which seasonal changes persist over multiple seasons.

Past studies have shown that groundwater chemistry has a significant role in selecting the prokaryotic organisms based on their compatibility with metabolic activity and availability of nutrients (Anantharaman *et al.*, 2016; Wegner *et al.*, 2019). However, all the nutrients that may impact the community differences could not be analysed, mainly due to a limited research budget. By making a dbRDA redundancy model, an attempt was made to find the chemistry, groundwater depth and land-use dependencies on the taxonomic assemblage differences. But due to limited chemical data analysed, the dbRDA model explained only 26% of the community variabilities between the three aquifers (Chapter 5) and explained only 32% of the community variation within the sandstone aquifer.

In the future, more chemical data from the three aquifers may help better explain community differences between and within the aquifers and strengthen the redundancy model. The additional chemicals which should be considered in future work, based on existing research (Fillinger *et al.*, 2019b; Sirisena *et al.*, 2018; Zhong *et al.*, 2023; Zhou *et al.*, 2012), include concentrations of nitrate, ammonia, nitrite, oxygen, methane, assimilable organic carbon (AOC), potassium, calcium, magnesium, iron, sulphate, and chloride. A stronger model of environmental drivers of community compositions will be essential to predict community structure based only on groundwater chemistry and the local aquifer architecture, where collecting eDNA data is not possible.

Using eDNA metabarcoding is a cutting-edge tool for assessing the presence of a species in an environment, although the interpretation of the resulting data remains very challenging. The eDNA is environmental DNA, which can be either intracellular DNA of an organism present in the sample or extracellular DNA shed by any organism previously present in a sample (Pawlowski *et al*., 2020). Thus, the presence of eDNA does not necessarily indicate the active presence of an organism but may suggest, at some point in time, the organism was in contact with the sample. To assess the viability of a microbe in a sample, targeted taxa could be grown in a laboratory culture. However, the groundwater in the three aquifers reported in the thesis had a subordinate community of intact prokaryotes (~20%), making most of the microbes non-viable to grow in a culture. Further, most of the groundwater microbes have not been sequenced, and novel taxa are still being discovered, making the protocol to grow and identify viable microbial cells impossible to design at this stage. This leads to two serious questions in the groundwater context. At what stage was the microbe in contact with the groundwater, and whether the microbe was in attached form or suspended form. These questions were beyond the scope of this thesis.

Fundamental research about eDNA degradation rate and transportation rate in different environmental samples is required to interpret the relationship between eDNA presence and active taxa presence. Differentiation between attached and planktonic taxa should be done during the coring of a new borehole in the aquifers by collecting eDNA samples from the sediment matrix and pore water.

Functional assignment using FAPROTAX (Louca *et al*., 2016) has its own limitations. Sansupa *et al*. (2021) reported that in soil samples, if the percentage of ASVs assigned to a genus level is lower, the functional potential assigned to the community becomes less reliable. This is probably true because the FAPROTAX algorithm matches known functions to the known genus or species of microbes. In the case of the samples analysed for the thesis, the taxonomic assignment of the ASVs was very low at the family (26%) and genus levels (16%). Many groundwater taxa lack whole genome sequences, making their functional assignment data unreliable. The uncertainty regarding eDNA presence and taxa presence presents additional difficulty about whether the functional potential was due to the presence of currently active microbes or microbes that encountered the sample in the past. Functional potential using FAPROTAX was also performed in a recent regional study in Australia (Korbel *et al.*, 2024). While this approach is a strong exploratory effort for first-hand interpretation of microbial functions, lots of unknown microbial functions exist in groundwater.

More robust and expensive alternatives, such as eDNA metagenomics and meta-transcriptomics, should be used to determine the active microbial functions and their contribution to biogeochemical cycles. Complete interpretation of biogeochemical activity may require a multivariate approach. A meta-omics study can indicate active prokaryotic genes in a sample (Anantharaman *et al.*, 2016; Wegner *et al.*, 2019). Using the isotopes of O, C, N, and S can help to delineate microbially and chemically reworked fractions of chemicals. Finally, the chemical concentrations can be used to find whether the abundances of active genes correlate with the presence or absence of electron donor or acceptor concentrations used for their functions (Kumar *et al.*, 2017; Liu *et al.*, 2022).

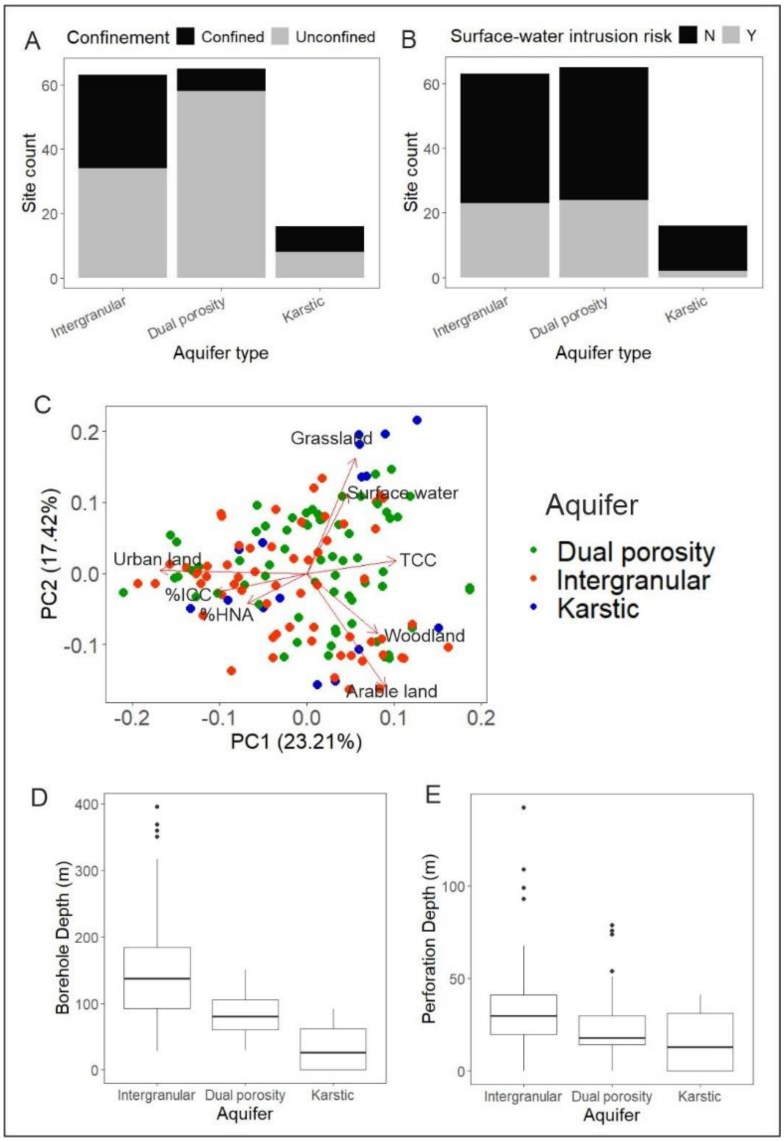
Groundwater microbiome is also composed of eukaryotes, such as protists and fungi, although their numbers are far lower than prokaryotes (Griebler and Lueders, 2009). Despite their lower abundance, these eukaryotes play an important role in the ecosystem. For example, Risse-Buhl *et al.* (2013) and Lin *et al.* (2012) found that the protistan number depends on the number of bacteria, and by actively grazing on the prokaryotes, they help to control their numbers and release new organic carbon in the groundwater. The focus of this thesis was on the prokaryotic community, but that does not provide a holistic view of the microbial community.

A study of holistic groundwater microbiome using 16S, 18S and ITS metabarcoding is required in the future. The protists and fungi microorganisms can be identified using 18S and ITS metabarcoding approaches, respectively. Moreover, groundwater macrofauna data can be integrated in the future for a complete picture of the groundwater ecosystems.

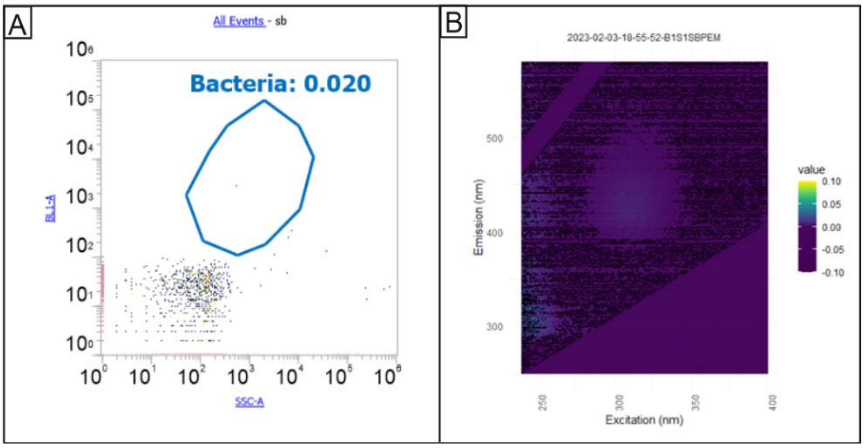
Research dissemination

International conferences		
Poster	Relating dissolved organic matter and bacterial biomass in the English aquifer systems	European Geosciences Union, Vienna, 2024
Presentation	Flow cytometric analysis reveals aquifer properties control baseline bacterial abundance	Internation association of Hydrogeologists conference, Davos, 2024
Presentation	Decadal evolution of groundwater planktonic prokaryotes of deep sandstone aquifer	European Geosciences Union, Vienna, 2025
Datasets		
eDNA sequences	Amplicon sequences (16S) from samples collected from groundwater survey of UK aquifers	NCBI Accession: PRJNA1268368 ID: 1268368
Flow cytometry data	Groundwater bacterioplankton, fluorescent organic matter and nutrient concentration of three major aquifers in England, September-October 2022 and January-February 2023	Submitted to EIDC

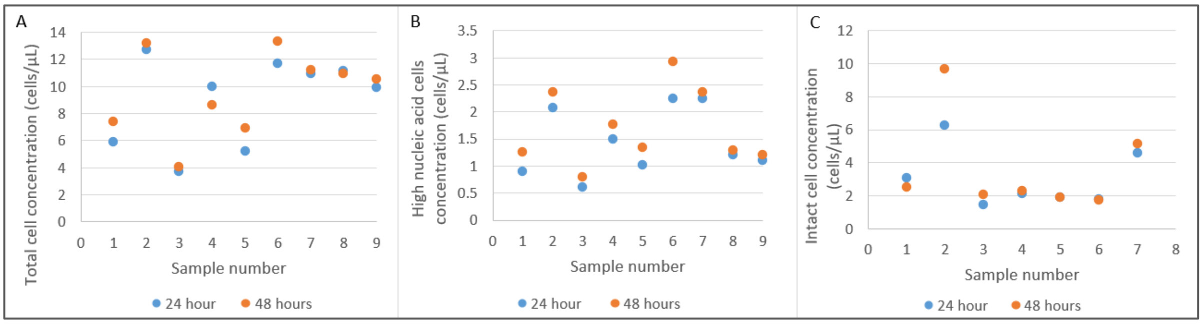
Appendix



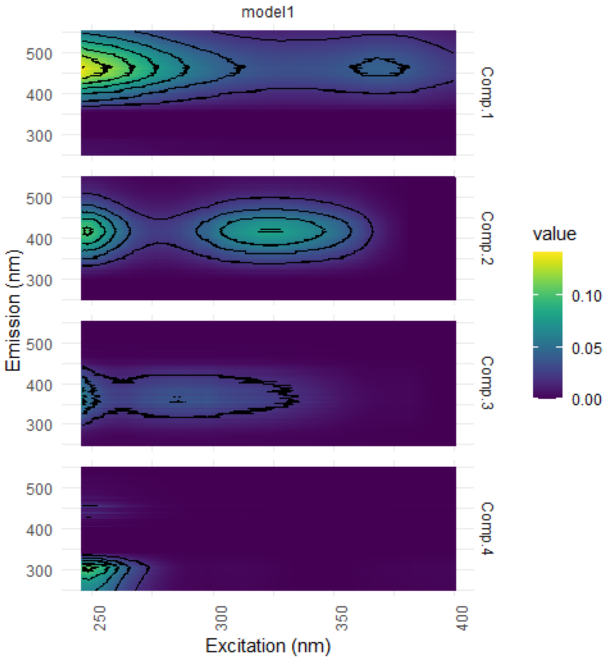
Appendix 1. Site properties of three aquifer types showing A. Distribution of confined and unconfined sites; B. Distribution of sites with and without risk of surface water intrusion; C. Principal component analysis of flow cytometric bacterioplankton concentration variables (TCC, %HNA, %ICC) and land-use categories (Based upon LCM2021 © UKCEH 2022) in source protection zone-1 (SPZ1) of the sites and points coloured by aquifer categories. D, E. Borehole total and groundwater inflow depths; for karstic springs borehole and inflow depths are considered 0m. Borehole data retrieved from GeoIndex data centre's (NGDC) scanned borehole collection BGS © UKRI (2023)



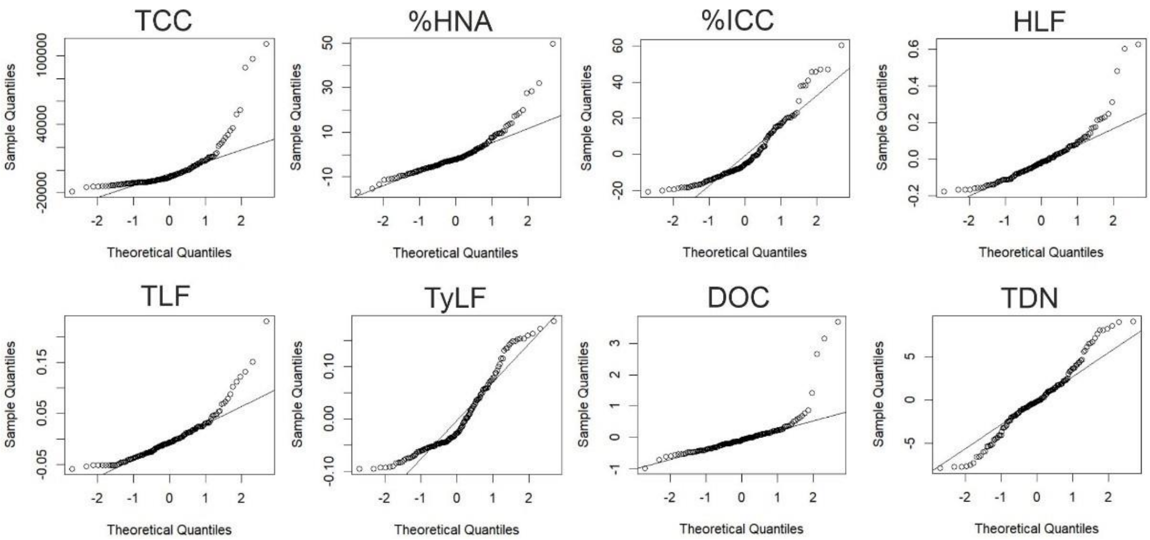
Appendix 2. 18 mg/L sodium thiosulphate solution dosed bottles filled with Miliq water showing dosing had no impact on bacterial count and fluorescence signatures. A. Bacterial total cell concentration (TCC) gate, B. Fluorescence signature of dosed blank bottle.



Appendix 3. Stability checks on random groundwater samples after 24- and 48-hours storage time and analysing for A. Total cell concentration, B. High nucleic acid cell concentration, and C. Intact cell concentration values using flow cytometry.



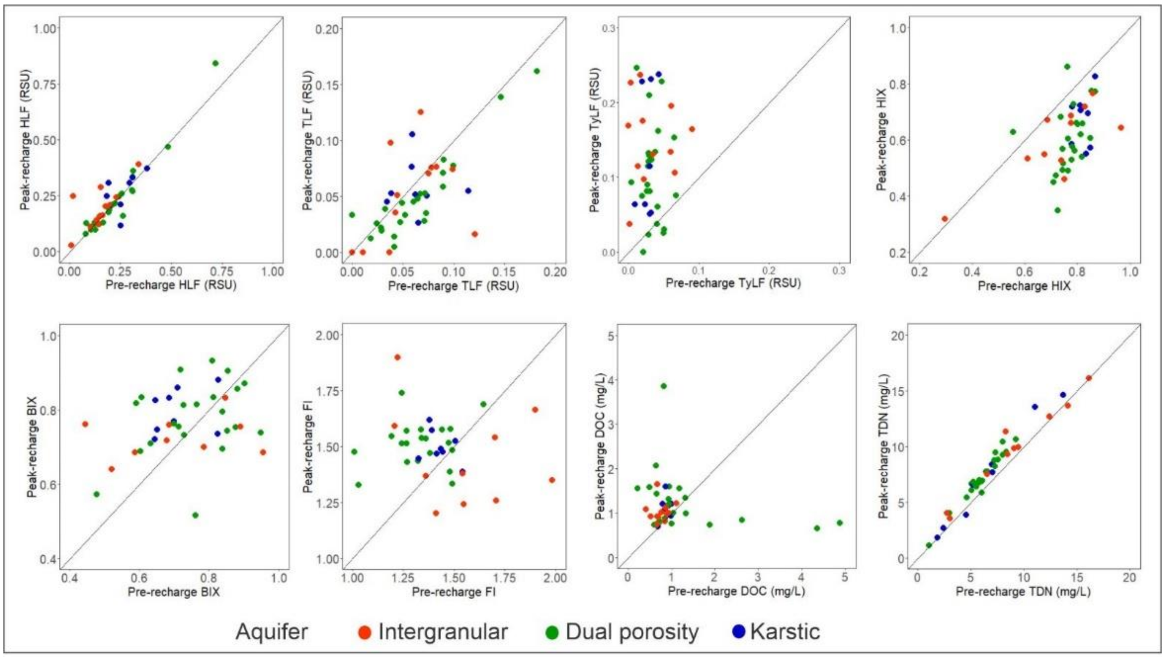
Appendix 4. Components 1-4 obtained from PARAFAC analysis of the Excitation-Emission matrix of fluorescent organic matter.



Appendix 5. QQ-normal plots of ANOVA model residuals of A. TCC, B. %HNA, C. %ICC, D. HLF, E. TLF, F. TyLF, G. DOC and H. TDN showing non-normal distribution of parameters.

Aquifers	Intergranular			Dual porosity			Karstic		
Seasons	All (n=63)	Pre-recharge (n=29)	Peak-recharge (n=34)	All (n=65)	Pre-recharge (n=32)	Peak-recharge (n=33)	All (n=16)	Pre-recharge (n=8)	Peak-recharge (n=8)
TCC (cells/mL)	**1×10^4** (5.7×10^3–1.6×10^4)	**1×10^4** (5.4×10^3–1.5×10^4)	**1×10^4** (5.9×10^3–1.7×10^4)	**8.7×10^3** (5.8×10^3–2×10^4)	**1.3×10^4** (6.7×10^3–2.6×10^4)	**7.1×10^3** (5.2×10^3–1.2×10^4)	**1.9×10^4** (1.5×10^4-3.1×10^4)	**2×10^4** (1.6×10^4-3.5×10^4)	**1.9×10^4** (1.3×10^4-3×10^4)
%HNA	**20** (15.8-26.7)	**19.6** (15.9-23.5)	**20** (14.3-29.6)	**15.8** (12.2-17.2)	**12.8** (10.4-14.5)	**16.4** (12.2-17.2)	**14.6** (11.8-25.9)	**16.4** (14-20.2)	**17.7** (11.5-26.1)
%ICC	**21.3** (17.6-29.1)	**18.6** (13.2-40.2)	**22** (19.2-26)	**21.3** (16.5-42.7)	**18** (13.7-28.3)	**40.5** (19.5-46)	**24.3** (18-40)	**19** (17-32.3)	**32** (22.4-44)
HLF (RSU)	**0.14** (0.09-0.21)	**0.15** (0.04-0.23)	**0.14** (0.1-0.2)	**0.21** (0.13-0.27)	**0.23** (0.14-0.26)	**0.21** (0.13-0.3)	**0.25** (0.2-0.3)	**0.25** (0.2-0.3)	**0.27** (0.2-0.3)
TLF (RSU)	**0.04** (0.01-0.07)	**0.06** (0.02-0.08)	**0.03** (0.01-0.06)	**0.05** (0.03-0.07)	**0.05** (0.04-0.07)	**0.04** (0.03-0.07)	**0.05** (0.05-0.06)	**0.06** (0.05-0.06)	**0.05** (0.05-0.06)
TyLF (RSU)	**0.06** (0.03-0.16)	**0.02** (0.01-0.04)	**0.15** (0.12-0.2)	**0.04** (0.03-0.1)	**0.03** (0.02-0.04)	**0.1** (0.07-0.13)	**0.04** (0.03-0.07)	**0.03** (0.02-0.03)	**0.1** (0.06-0.23)
DOC (mg/L)	**0.85** (0.7-1.01)	**0.77** (0.68-0.9)	**0.93** (0.8-1.02)	**1** (0.8-1.31)	**0.94** (0.72-1.1)	**1.05** (0.81-1.4)	**0.96** (0.86-1.1)	**0.87** (0.83-0.95)	**1.1** (1-0.12)
TDN (mg/L)	**7.8** (5.3-11.3)	**8.2** (2.7-12.4)	**7.7** (5.9-10)	**6.8** (5.6-8.5)	**6.1** (5.2-8)	**7.7** (6.3-8.9)	**6.7** (3.5-9)	**6.1** (4-8)	**7.1** (3.5-9.6)
HIX	**0.67** (0.52-0.77)	**0.77** (0.69-0.84)	**0.54** (0.45-0.66)	**0.74** (0.6-0.78)	**0.78** (0.74-0.82)	**0.61** (0.54-0.7)	**0.78** (0.7-0.82)	**0.82** (0.8-0.84)	**0.7** (0.58-0.72)
BIX	**0.68** (0.46-0.78)	**0.68** (0.44-0.83)	**0.68** (0.48-0.75)	**0.76** (0.68-0.85)	**0.72** (0.63-0.83)	**0.81** (0.73-0.86)	**0.74** (0.7-0.82)	**0.7** (0.65-0.73)	**0.8** (0.74-0.83)
FI	**1.52** (1.32-1.7)	**1.54** (1.3-1.7)	**1.51** (1.35-1.65)	**1.45** (1.33-1.56)	**1.35** (1.26-1.48)	**1.51** (1.44-1.57)	**1.45** (1.41-1.51)	**1.42** (1.38-1.45)	**1.48** (1.46-1.53)

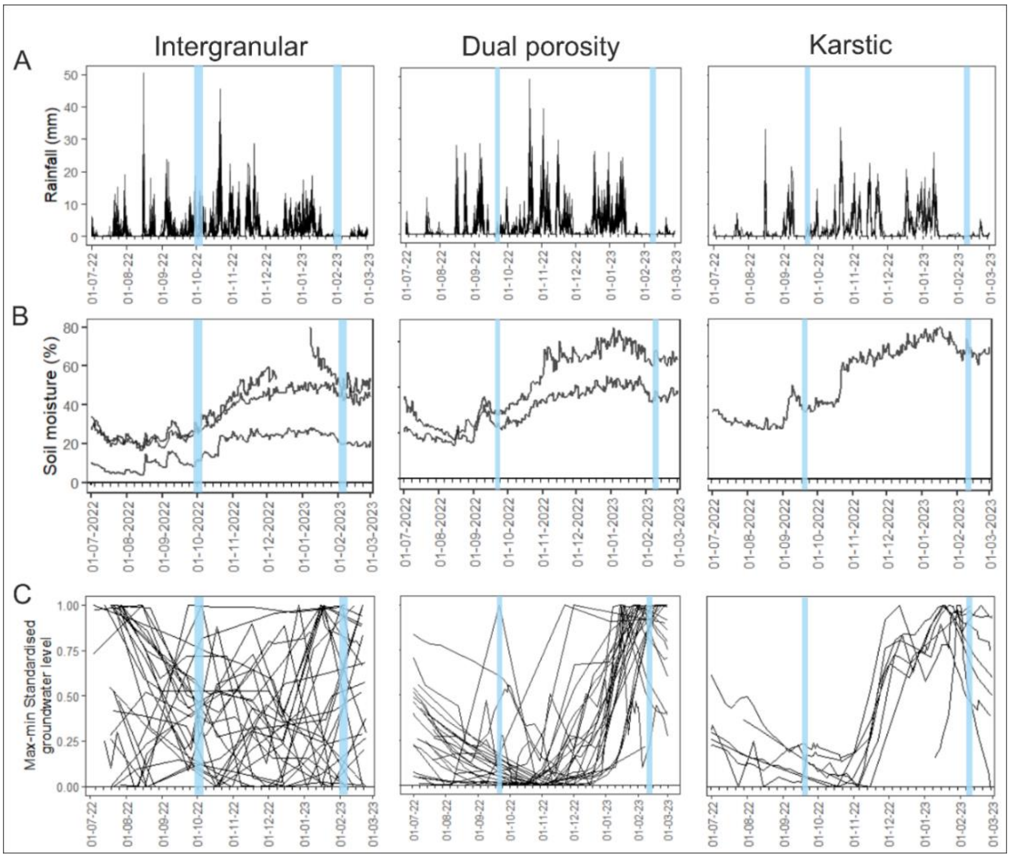
Appendix 6. Median and (Interquartile range or IQR) of bacterioplankton TCC, %HNA and %ICC and fOM, DOC and TDN concentrations in three aquifer types, pre-recharge and peak-recharge seasons of all samples including paired and unpaired data. The characters in bold are median and inside brackets are interquartile ranges.



Appendix 7: Paired sites for nutrients corroborating the observations from the unpaired data on seasonal changes observed in Section 4.3.3.

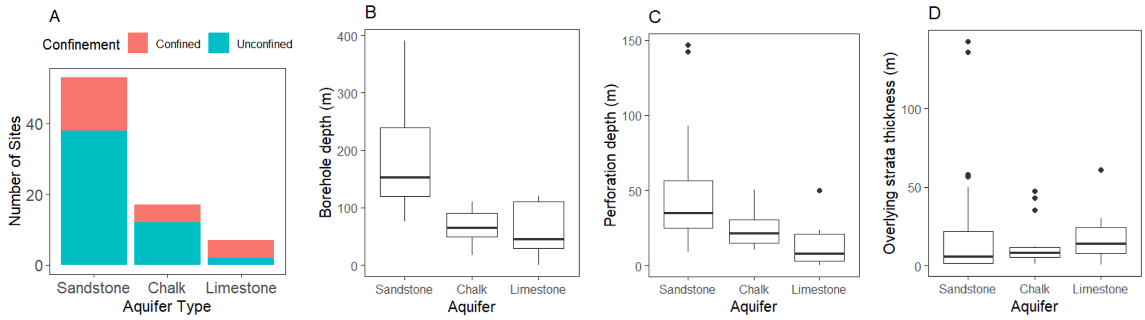
Aquifer	Intergranular		Dual porosity	
Season	Pre-recharge	Peak-recharge	Pre-recharge	Peak-recharge
DOC vs HLF	ρ=0.87, p<0.05	ρ=0.7, p<0.05	ρ=0.55, p<0.05	ρ=0.27, p>0.05
DOC vs TLF	ρ=0.66, p<0.05	ρ=0.16, p>0.05	ρ=0.42, p<0.05	ρ=0.28, p>0.05
HLF vs TLF	ρ=0.77, p<0.05	ρ=0.47, p<0.05	ρ=0.78, p<0.05	ρ=0.92, p<0.05

Appendix 8. Spearman correlation coefficient (ρ) and p-value among HLF and TLF with DOC in different aquifers at pre and peak-recharge seasons showing stronger DOC-HLF correlation and overall DOC-fOM correlation stronger in intergranular than dual porosity aquifer.

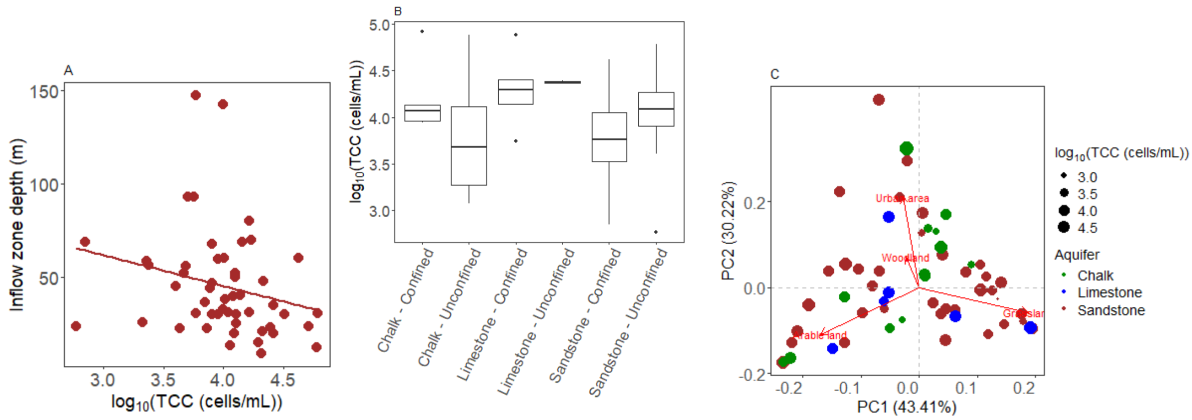


Month	Average rainfall (mm) in intergranular aquifer	Average rainfall (mm) in dual porosity aquifer	Average rainfall (mm) in karstic aquifer
July 2022	25.6	7.4	13
August 2022	27.1	27.4	27.5
September 2022	59.9	71.3	66
October 2022	99.2	91	105.7
November 2022	110.3	134.1	149.2
December 2022	52.8	83.5	89.3
January 2023	81.8	53.9	86.1

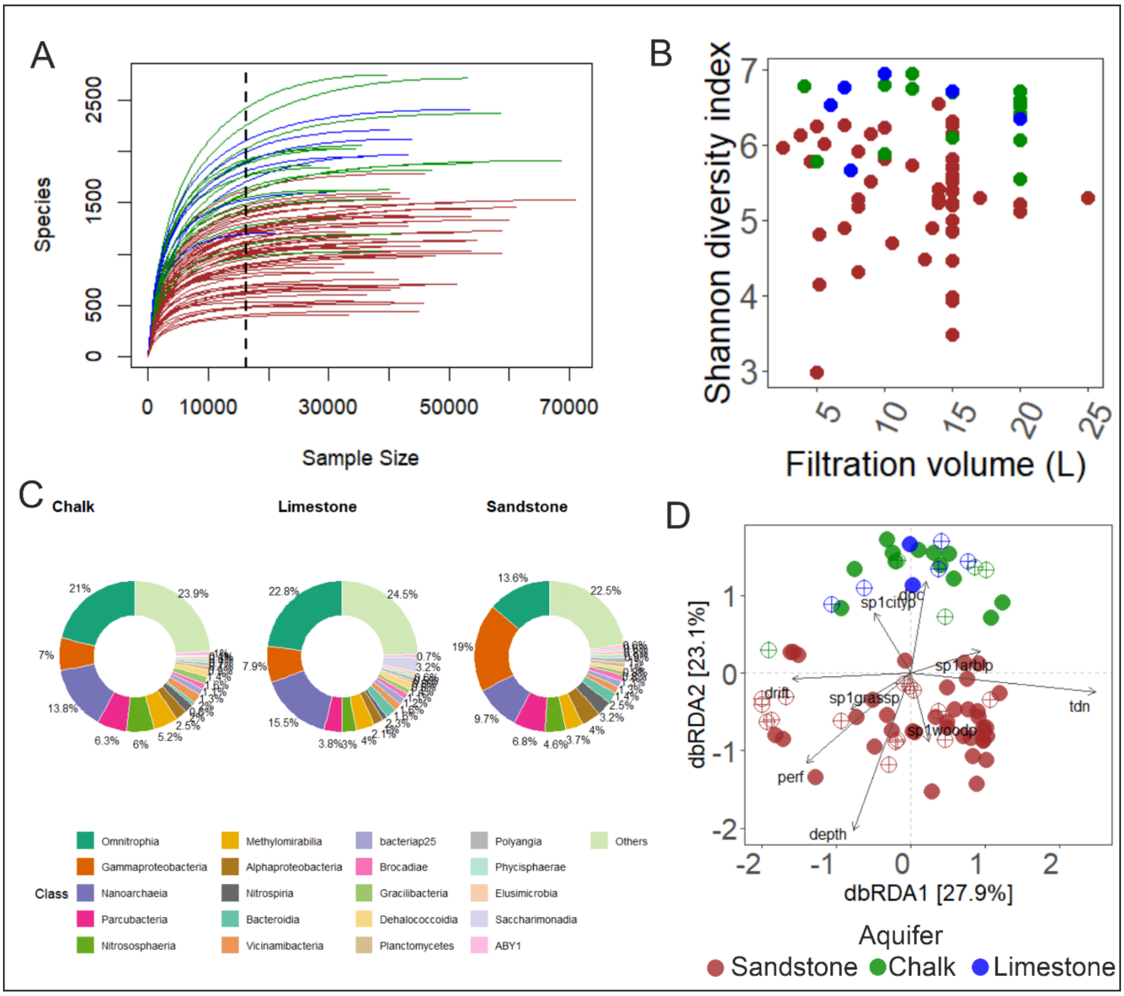
Appendix 9. Time-series of (A) daily rainfall (mm), (B) soil moisture (%), and (C) standardised groundwater level across intergranular, dual-porosity, and karstic aquifer types, between July 2022 and March 2023. Blue vertical bands indicate pre and peak-recharge sampling periods. The daily rainfall data and groundwater level data was retrieved from the (Environment Agency, 2024) (provided under **Open Government Licence 3.0**) and soil moisture data was retrieved from (Smith, 2024) (retrieved from Cosmic-ray soil moisture monitoring network | COSMOS). The table provides the average monthly rainfall three aquifer types.



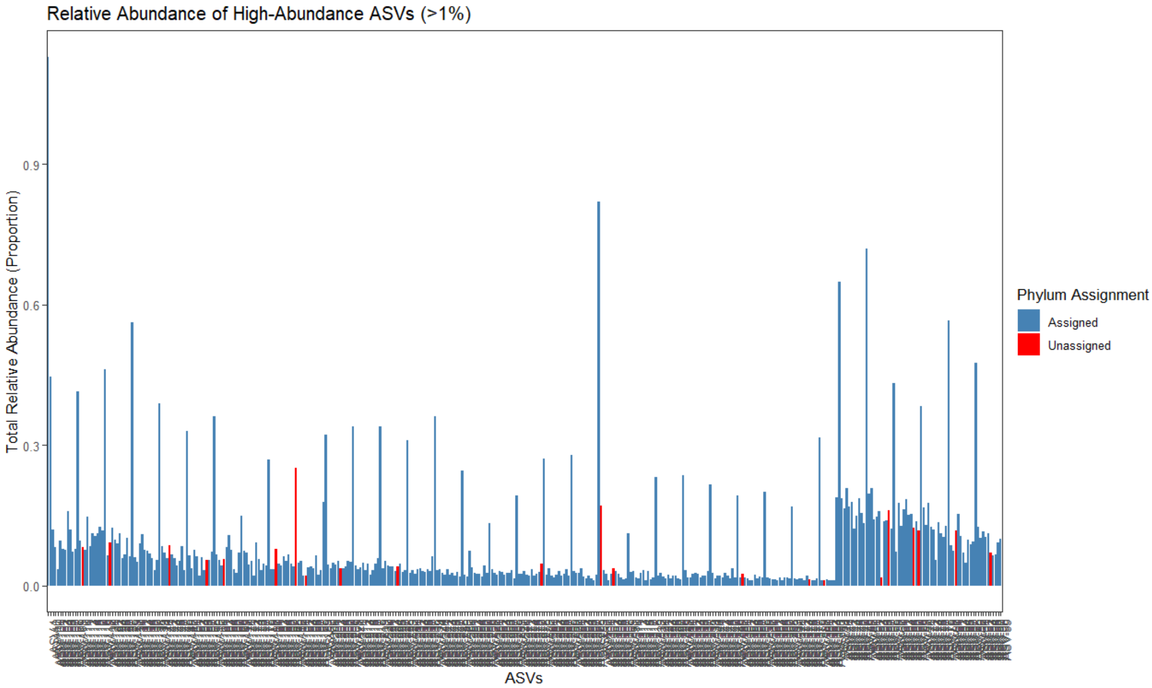
Appendix 10. A. number of confined and unconfined sites in three aquifers; Boxplots showing B. Total borehole depths, C. top of the borehole perforation depths, and D. Overlying strata thicknesses of the three aquifers, with the box hinges represent the interquartile range (IQR) and the median, the whiskers represent points up to 1.5 times the IQR and any point beyond that is deemed to be an outlier.



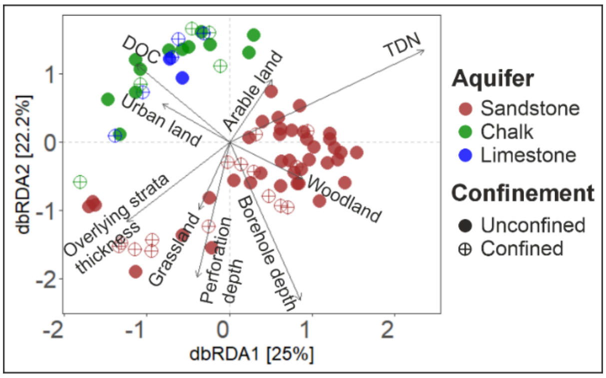
Appendix 11. A. Negative correlation between groundwater inflow depth and TCC in sandstone aquifer, B. Boxplots of TCC in unconfined and confined sites of each aquifer, C. Land use pattern (Based upon LCM2021 © UKCEH 2022) at different geological settings in the Source protection zone-1 of each site, point sizes indicate value of $\log_{10}$(cells/mL).



Appendix 12. A. Rarefaction curve of three aquifers showing sufficient sequencing depth covering majority of the taxa and dotted line shows rarefaction depth of 16277 reads; B. Filtration volume plotted against Shannon diversity index indicated no impact of filtration volume in any aquifer on the community variables; C. dominant taxonomic assemblage of three aquifers, without rarefaction and D. dbRDA plot of Bray-curtis dissimilarity (unrarefied) between sites of the three different aquifers, with arrows indicating direction and loading of strength of the environmental variables.

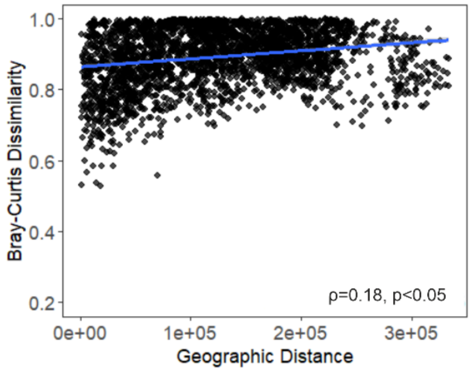


Appendix 13. Assigned and unassigned ASVs among the dominant ASVs, i.e., ASVs in more than 1% relative abundance in at least one sample.



Variable	dbRDA1	dbRDA2	Capscale- R^2 (Marginal effect)
Eigen values (constrained variables only)	1.42	1.26	
Aquifer			8.2% ***
Confinement			1.7% **
Borehole Depth	0.32	-0.94	1.5% *
Perforation depth	-0.21	-0.97	1.2%
Overlying strata thickness	-0.72	-0.7	1.8% *
DOC	-0.69	0.71	1.8% **
TDN	0.85	0.51	4.1% ***
Woodlands in SPZ1	0.86	-0.51	1.3%
Grassland in SPZ1	-0.37	-0.93	1.1%
Arable land in SPZ1	0.5	0.86	1.3%
Urban area in SPZ1	-0.82	0.56	1.3%

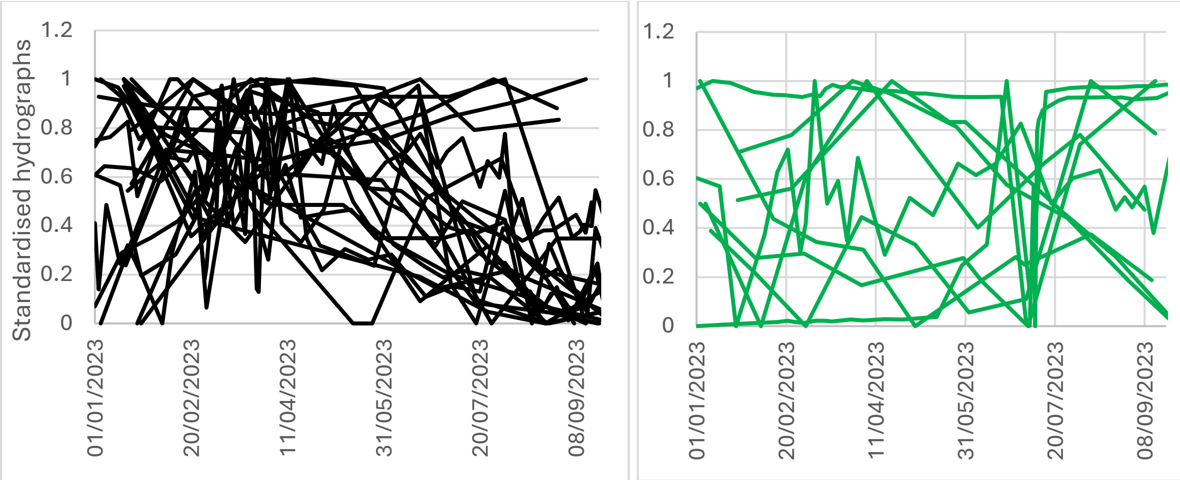
Appendix 14. dbRDA plot of Sørensen-Dice distances of ASVs and arrows showing loading of environmental variables along the two dbRDA axes, sites were coloured by aquifers and shaped by aquifer confinement. The table of loading of environmental variables along dbRDA1 and dbRDA2, and explanatory power of environmental variables (R^2) and the significance of each term is indicated as ***p<0.001, **p<0.01 and *p<0.05.



Appendix 15. Distance-decay plot of Geographic distance between the sites and Bray-Curtis distance between the prokaryotic communities, with ρ and p-values of Mantel correlation test result.

Function	elements	main_element	electron_donor	electron_acceptor	aerobic
human_pathogens_all	C	C	C	variable	variable
animal_parasites_or_symbionts	C	C	C	variable	variable
aromatic_compound_degradation	C	C	C	variable	variable
intracellular_parasites	C	C	C	variable	variable
predatory_or_exoparasitic	C	C	C	variable	variable
nonphotosynthetic_cyanobacteria	C	C	C	NA	variable
chemoheterotrophy	C	C	C	variable	variable
methanotrophy	C,H	C	C	variable	variable
hydrogenotrophic_methanogenesis	C,H	C	H	C	no
methanogenesis	C,H	C	variable	variable	no
methylotrophy	C,H	C	variable	variable	variable
aromatic_hydrocarbon_degradation	C,H	C	C	variable	variable
hydrocarbon_degradation	C,H	C	C	variable	variable
aerobic_chemoheterotrophy	C,O	C	C	O	yes
anaerobic_chemoheterotrophy	C,O	C	C	variable	no
dark_iron_oxidation	Fe	Fe	Fe	variable	variable
dark_hydrogen_oxidation	H	H	H	variable	variable
anammox	N	N	N	N	no
nitrogen_fixation	N	N	variable	N	variable
nitrite_respiration	N	N	variable	N	no
nitrate_respiration	N	N	variable	N	no
nitrate_reduction	N	N	variable	N	variable
nitrogen_respiration	N	N	variable	N	no
ureolysis	N,C	N	none	none	variable
aerobic_ammonia_oxidation	N,O	N	N	O	yes
aerobic_nitrite_oxidation	N,O	N	N	O	yes
nitrification	N,O	N	N	O	yes
sulfate_respiration	S	S	variable	S	no
respiration_of_sulfur_compounds	S	S	variable	S	no
dark_sulfide_oxidation	S	S	S	variable	variable
dark_thiosulfate_oxidation	S	S	S	variable	variable
dark_oxidation_of_sulfur_compounds	S	S	S	variable	variable
human_associated	variable	variable	variable	variable	variable

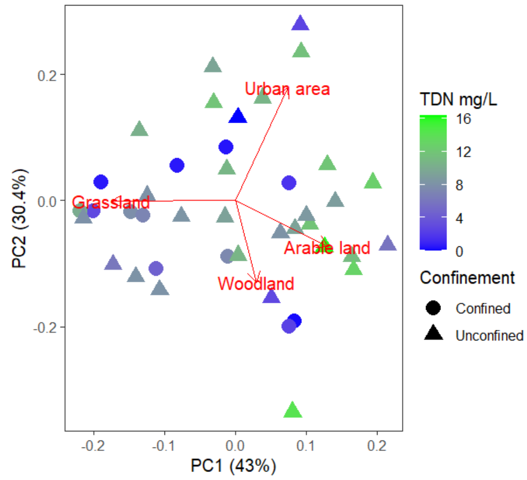
Appendix 16. Description of functional potentials used in FAPROTAX



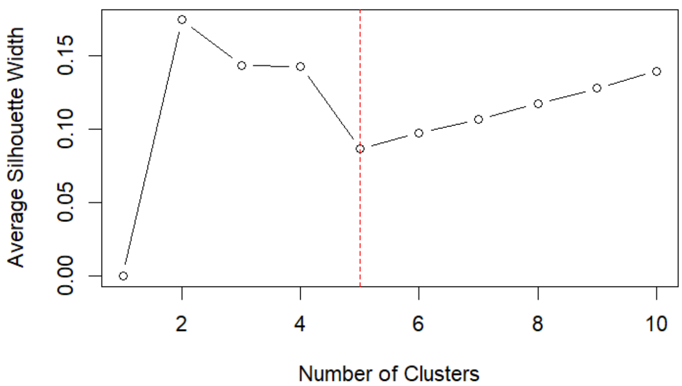
Appendix 17. Groundwater levels trend between winter and summer sampling seasons in 2023, separately for A. unconfined and B. confined sites. (includes data from the Environment Agency, 2024, provided under [Open Government Licence 3.0](#)).

Principal components	PC1	PC2	PC3	PC4
Eigenvalues	2.14	1.3	1.1	1.05
Variance explained	35.3%	13.3%	9.3%	8.5%
TCC	0.16	**0.4**	0.18	**-0.4**
%HNA	**-0.2**	-0.09	**0.68**	0.05
%ICC	**-0.33**	-0.04	**0.4**	0.06
CFC-12 age	**0.24**	-0.17	0.17	**0.56**
Drift thickness	**-0.37**	0.05	0.08	0.02
Perforation	**-0.34**	0.05	0.05	0.15
Well depth	-0.18	**-0.24**	**-0.33**	**0.48**
TDN	**0.36**	-0.11	0.01	0.02
DO	**0.23**	**-0.53**	0.06	-0.17
DOC	**0.3**	**0.34**	-0.02	**0.23**
pH	**-0.28**	-0.01	**-0.36**	-0.19
Temperature	**-0.33**	0.11	-0.19	0.0
Conductivity	0.06	**0.55**	0.02	**0.36**

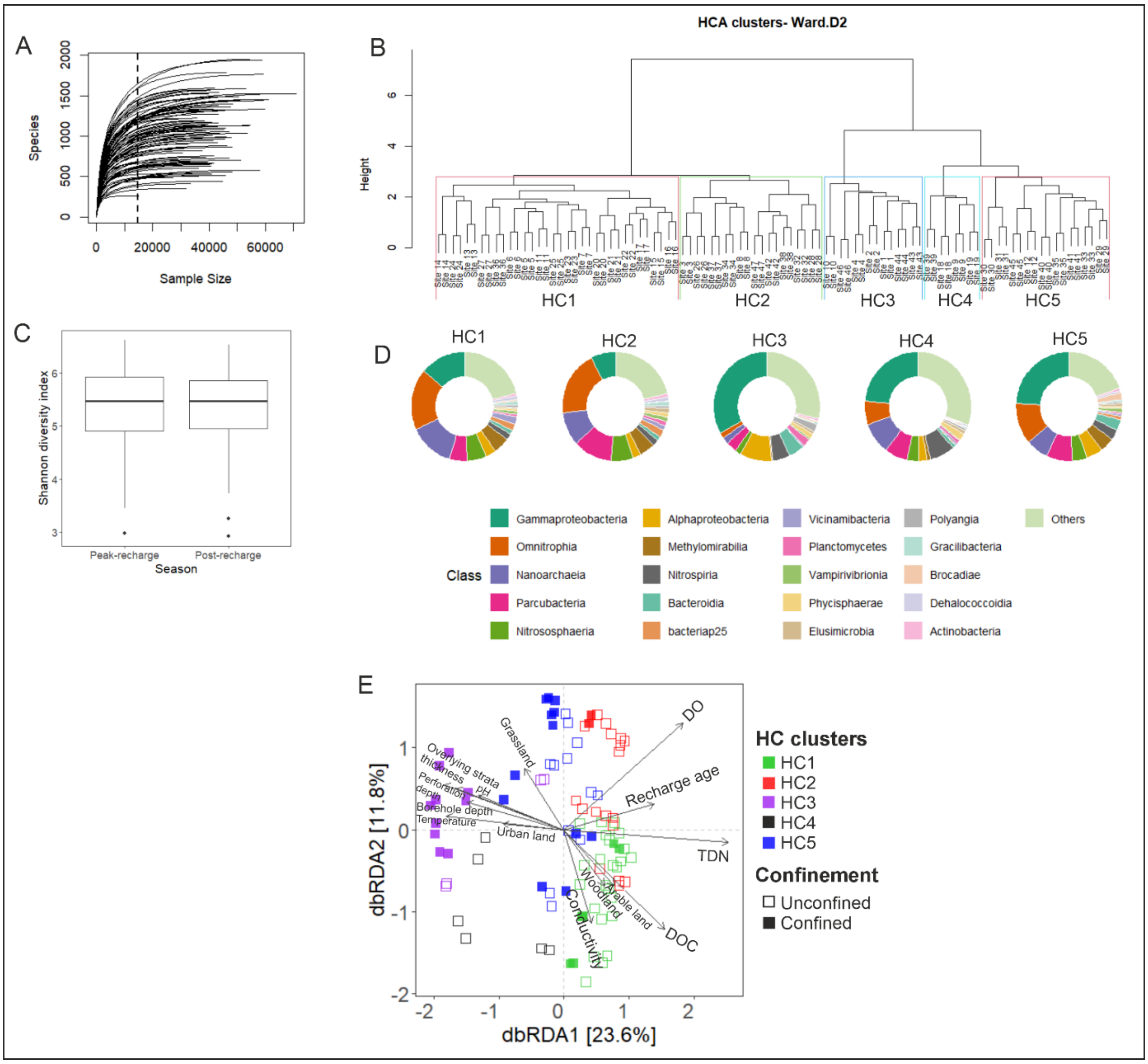
Appendix 18. Loading values of environmental variables along 1st four principal components, with big loading values shown in bold.



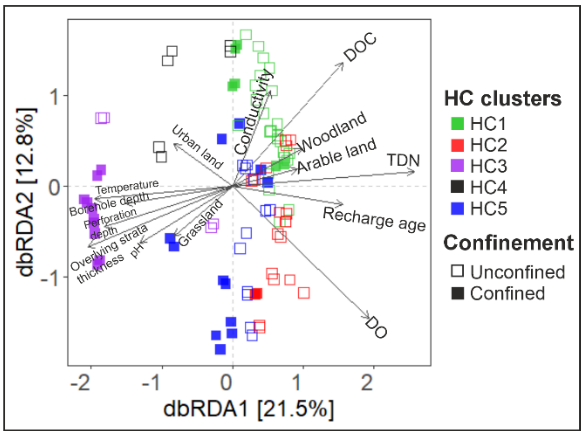
Appendix 19. PCA of land use pattern, points are coloured by TDN concentration, shaped by catchment confinement, showing that most samples with an unconfined catchment and a higher proportion of arable land coverage in SPZ-1 showed a higher TDN concentration.



Appendix 20. Silhouette plot showing 5-cluster level is the optimum cluster level where Sum-Squared distance dips the most.

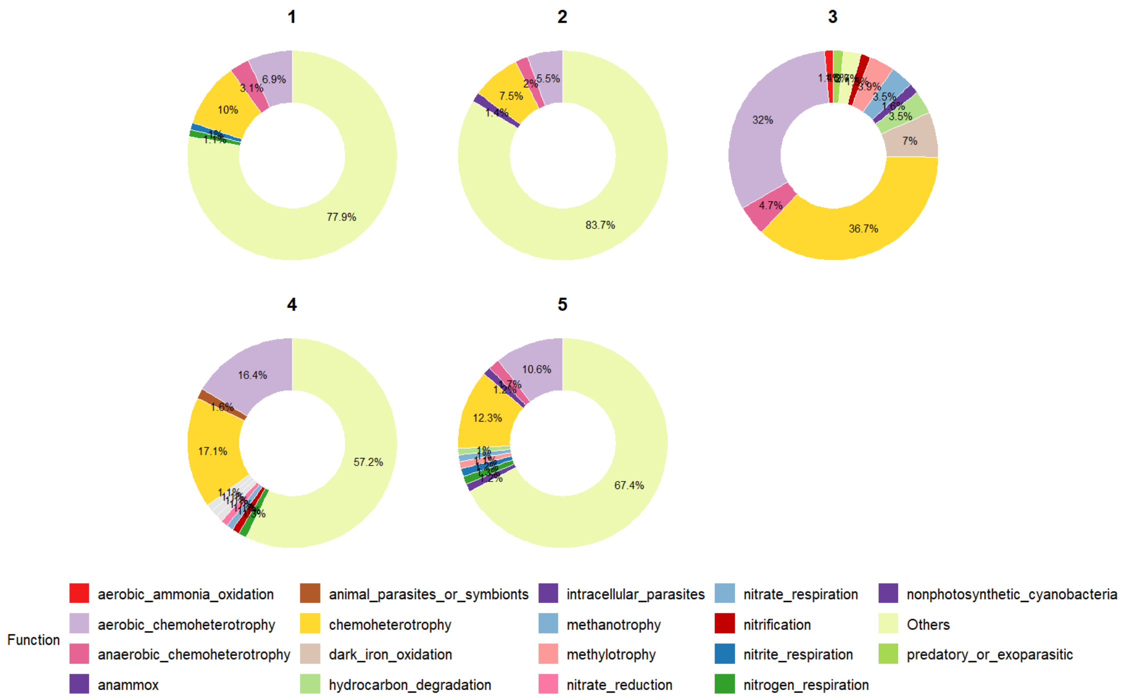


Appendix 21. A. Rarefaction curve of three aquifers showing sufficient sequencing depth covering majority of the taxa and dotted line shows rarefaction depth of 14696 reads; B. Hierarchical clusters shown in a dendrogram prepared using WardD2 method; C. Boxplots showing Shannon diversity index of the samples in peak and post-recharge seasons; D. dominant taxonomic assemblage of 5 clusters and E. dbRDA plot of Bray-Curtis dissimilarity between sites of the 5 different clusters, with arrows indicating direction and loading of strength of the environmental variables.

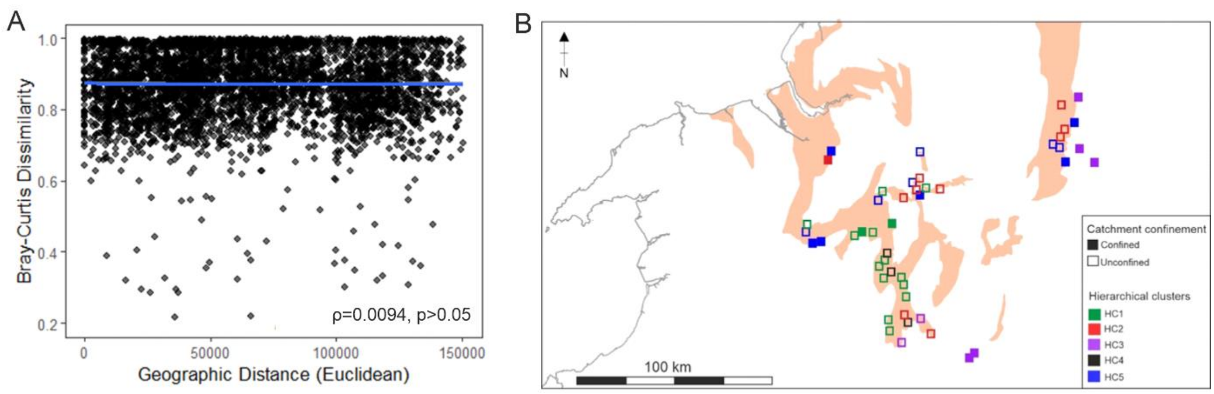


	dbRDA1	dbRDA2	Environment fit (r^2) and p-value indicators
Eigenvalues	2.2	1.3	
Recharge age	0.98	-0.18	0.19 *
Overlying strata	-0.94	-0.33	0.44 ***
Perforation	-0.96	-0.25	0.31 ***
Borehole depth	-0.99	-0.11	0.17 *
TDN	0.99	0.03	0.66 ***
DO	0.74	-0.66	0.7 ***
DOC	0.69	0.71	0.46 ***
pH	-0.88	-0.47	0.15
Temperature	-0.99	-0.05	0.34 **
Conductivity	0.36	0.93	0.08
Arable land	0.97	0.2	0.01
Grassland	-0.8	-0.6	0.02
Woodland	0.9	0.42	0.03
Urban area	-0.82	0.56	0.01

Appendix 22. dbRDA plot of Sørensen-Dice distances of ASVs and arrows showing loading of environmental variables along the two dbRDA axes, sites were coloured by HCs and shaped by aquifer confinement. The table of loading of environmental variables along dbRDA1 and dbRDA2, and environment fit of environmental variables (r^2) and the significance of each term is indicated as ***p<0.001, **p<0.01 and *p<0.05.



Appendix 23. Donut plot of mean relative abundance of functional potentials at each hierarchical cluster (HC1 to HC5) identified using FAPROTAX, and showing HC3 had more known functions than HC1 and HC2



Appendix 24. Geographic distribution of the five hierarchical clusters, coloured by the cluster each site belonged to and shaped by aquifer confinement, showing that the HCs were not distributed based on spatial proximity.

References

Abellan-Schneyder, I., Matchado, M.S., Reitmeier, S., Sommer, A., Sewald, Z., Baumbach, J., List, M. and Neuhaus, K. 2021. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. Msphere 6(1), 10.1128/msphere. 01202-01220.

Abraham, H.M. and Close, M. 2024. GROUNDWATER ECOSYSTEMS PRELIMINARY BIODIVERSITY OF UNCOMMON AQUIFERS: FRACTURED BASALT AND COARSE SAND.

Aburto, A. and Ball, A.S. 2009. Bacterial population dynamics and separation of active degraders by stable isotope probing during benzene degradation in a BTEX-impacted aquifer. Revista internacional de contaminación ambiental 25(3), 147-156.

Allen, D., Brewerton, L., Coleby, L., Gibbs, B., Lewis, M., MacDonald, A., Wagstaff, S. and Williams, A. 1997. The physical properties of major aquifers in England and Wales.

Amalfitano, S., Del Bon, A., Zoppini, A., Ghergo, S., Fazi, S., Parrone, D., Casella, P., Stano, F. and Preziosi, E. 2014. Groundwater geochemistry and microbial community structure in the aquifer transition from volcanic to alluvial areas. Water research 65, 384-394.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J. and Karaoz, U. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nature communications 7(1), 13219.

Balkwill, D., Fredrickson, J. and Thomas, J. 1989. Vertical and horizontal variations in the physiological diversity of the aerobic chemoheterotrophic bacterial microflora in deep southeast coastal plain subsurface sediments. Applied and Environmental Microbiology 55(5), 1058-1065.

Bar-On, Y.M., Phillips, R. and Milo, R. 2018. The biomass distribution on Earth. Proceedings of the National Academy of Sciences 115(25), 6506-6511.

Bartlett, R., Bottrell, S.H., Sinclair, K., Thornton, S., Fielding, I.D. and Hatfield, D. 2010. Lithological controls on biological activity and groundwater chemistry in Quaternary sediments. Hydrological Processes: An International Journal 24(6), 726-735.

Bearcock, J., Smedley, P. and Milne, C. 2015. Baseline groundwater chemistry: the Corallian of the Vale of Pickering, Yorkshire.

Beaver, R.C. and Neufeld, J.D. 2024. Microbial ecology of the deep terrestrial subsurface. The ISME Journal 18(1), wrae091.

Ben Maamar, S., Aquilina, L., Quaiser, A., Pauwels, H., Michon-Coudouel, S., Vergnaud-Ayraud, V., Labasque, T., Roques, C., Abbott, B.W. and Dufresne, A. 2015. Groundwater isolation governs chemistry and microbial community structure along hydrologic flowpaths. Frontiers in Microbiology 6, 1457.

Benk, S.A., Yan, L., Lehmann, R., Roth, V.-N., Schwab, V.F., Totsche, K.U., Küsel, K. and Gleixner, G. 2019. Fueling diversity in the subsurface: composition and age of dissolved organic matter in the critical zone. Frontiers in Earth Science 7, 296.

Besmer, M.D., Epting, J., Page, R.M., Sigrist, J.A., Huggenberger, P. and Hammes, F. 2016. Online flow cytometry reveals microbial dynamics influenced by concurrent natural and operational events in groundwater used for drinking water treatment. Scientific reports 6(1), 38462.

Besmer, M.D., Weissbrodt, D.G., Kratochvil, B.E., Sigrist, J.A., Weyland, M.S. and Hammes, F. 2014. The feasibility of automated online flow cytometry for in-situ monitoring of microbial dynamics in aquatic ecosystems. Frontiers in microbiology 5, 265.

Bitton, G., Farrah, S., Ruskin, R., Butner, J. and Chou, Y. 1983. Survival of pathogenic and indicator organisms in ground water. Groundwater 21(4), 405-410.

Bloomfield, J., Gooddy, D., Bright, M. and Williams, P. 2001. Pore-throat size distributions in Permo-Triassic sandstones from the United Kingdom and some implications for contaminant hydrogeology. Hydrogeology Journal 9, 219-230.

Bobbo, T., Biscarini, F., Yaddehige, S.K., Alberghini, L., Rigoni, D., Bianchi, N. and Taccioli, C. 2024. Machine learning classification of archaea and bacteria identifies novel predictive genomic features. BMC genomics 25(1), 955.

Bolster, C.H., Mills, A.L., Hornberger, G.M. and Herman, J.S. 2001. Effect of surface coatings, grain size, and ionic strength on the maximum attainable coverage of bacteria on sand surfaces. Journal of contaminant hydrology 50(3-4), 287-305.

Bonett, D.G. and Wright, T.A. 2000. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. Psychometrika 65, 23-28.

Boussau, B., Karlberg, E.O., Frank, A.C., Legault, B.-A. and Andersson, S.G. 2004. Computational inference of scenarios for α-proteobacterial genome evolution. Proceedings of the National Academy of Sciences 101(26), 9722-9727.

British Geological Survey 2019 Groundwater resources in the UK.

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

Cardman, Z., Arnosti, C., Durbin, A., Ziervogel, K., Cox, C., Steen, A. and Teske, A. 2014. Verrucomicrobia are candidates for polysaccharide-degrading bacterioplankton in an arctic fjord of Svalbard. Applied and Environmental Microbiology 80(12), 3749-3756.

Carlson, H.K., Price, M.N., Callaghan, M., Aaring, A., Chakraborty, R., Liu, H., Kuehl, J.V., Arkin, A.P. and Deutschbauer, A.M. 2019. The selective pressures on the microbial community in a metal-contaminated aquifer. The ISME journal 13(4), 937-949.

Castelle, C.J. and Banfield, J.F. 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell 172(6), 1181-1197.

Castelle, C.J., Brown, C.T., Thomas, B.C., Williams, K.H. and Banfield, J.F. 2017. Unusual respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the Candidate Phyla Radiation. Scientific reports 7(1), 40101.

Cavicchioli, R. 2011. Archaea—timeline of the third domain. Nature Reviews Microbiology 9(1), 51-61.

Celle-Jeanton, H., Huneau, F., Travi, Y. and Edmunds, W.M. 2009. Twenty years of groundwater evolution in the Triassic sandstone aquifer of Lorraine: impacts on baseline water quality. Applied Geochemistry 24(7), 1198-1213.

Chapelle, F.H. (2000) Ground-water microbiology and geochemistry, John Wiley & Sons.

Chapman, M. and Underwood, A. 1999. Ecological patterns in multivariate assemblages: information and interpretation of negative values in ANOSIM tests. Marine ecology progress series 180, 257-265.

Chaudhari, N.M., Pérez-Carrascal, O.M., Overholt, W.A., Totsche, K.U. and Küsel, K. 2024. Genome streamlining in Parcubacteria transitioning from soil to groundwater. Environmental Microbiome 19(1), 41.

Chik, A.H., Emelko, M.B., Anderson, W.B., O'Sullivan, K.E., Savio, D., Farnleitner, A.H., Blaschke, A.P. and Schijven, J.F. 2020. Evaluation of groundwater bacterial community composition to inform waterborne pathogen vulnerability assessments. Science of the Total Environment 743, 140472.

Coble, P.G. 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Marine chemistry 51(4), 325-346.

Cooper, K.J., Whitaker, F.F., Anesio, A.M., Naish, M., Reynolds, D.M. and Evans, E.L. 2016. Dissolved organic carbon transformations and microbial community response to variations in recharge waters in a shallow carbonate aquifer. Biogeochemistry 129(1), 215-234.

Couton, M., Hürlemann, S., Studer, A., Alther, R. and Altermatt, F. 2023. Groundwater environmental DNA metabarcoding reveals hidden diversity and reflects land-use and geology. Molecular Ecology 32(13), 3497-3512.

Cronin, A., Taylor, R., Powell, K., Barrett, M., Trowsdale, S. and Lerner, D. 2003. Temporal variations in the depth-specific hydrochemistry and sewage-related microbiology of an urban sandstone aquifer, Nottingham, United Kingdom. Hydrogeology Journal 11(2), 205-216.

Danczak, R.E., Johnston, M.D., Kenah, C., Slattery, M. and Wilkins, M.J. 2018. Microbial community cohesion mediates community turnover in unperturbed aquifers. Msystems 3(4), 10.1128/msystems. 00066-00018.

Danczak, R.E., Sawyer, A.H., Williams, K.H., Stegen, J.C., Hobson, C. and Wilkins, M.J. 2016a. Seasonal hyporheic dynamics control coupled microbiology and geochemistry in Colorado River sediments. Journal of Geophysical Research: Biogeosciences 121(12), 2976-2987.

Danczak, R.E., Yabusaki, S.B., Williams, K.H., Fang, Y., Hobson, C. and Wilkins, M.J. 2016b. Snowmelt induced hydrologic perturbations drive dynamic microbiological and geochemical behaviors across a shallow riparian aquifer. Frontiers in Earth Science 4, 57.

Danielopol, D.L., Pospisil, P. and Rouch, R. 2000. Biodiversity in groundwater: a large-scale view. Trends in Ecology & Evolution 15(6), 223-224.

Darling, W., Gooddy, D., MacDonald, A. and Morris, B. 2012. The practicalities of using CFCs and SF6 for groundwater dating and tracing. Applied Geochemistry 27(9), 1688-1697.

Davey, H. and Guyot, S. 2020. Estimation of microbial viability using flow cytometry. Current Protocols in Cytometry 93(1), e72.

Dockins, W.S., Olson, G.J., McFeters, G.A. and Turbak, S.C. 1980. Dissimilatory bacterial sulfate reduction in Montana groundwaters. Geomicrobiology Journal 2(1), 83-98.

Dunlap, W.J., Cosby, R.L., McNabb, J.F., Bledsoe, B.E. and Scalfb, M.R. 1972 Probable impact of NTA on ground water, p. 201, US Government Printing Office.

Dunn, O.J. 1964. Multiple comparisons using rank sums. Technometrics 6(3), 241-252.

Edmunds, W., Shand, P., Hart, P. and Ward, R. 2003. The natural (baseline) quality of groundwater: a UK pilot study. Science of the Total Environment 310(1-3), 25-35.

Elango, L. and Kannan, R. 2007. Rock–water interaction and its control on chemical composition of groundwater. Developments in environmental science 5, 229-243.

Eme, L., Tamarit, D., Caceres, E.F., Stairs, C.W., De Anda, V., Schön, M.E., Seitz, K.W., Dombrowski, N., Lewis, W.H. and Homa, F. 2023. Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. Nature 618(7967), 992-999.

Environment Agency 2019 Manual for the production of Groundwater Source Protection Zones, Environment Agency Bristol.

Environment Agency 2024 Hydrology Data Explorer (Retrieved June 2024)

European Union 2006 EU Groundwater Directive 2006/118/EC, Official Journal of the European Union.

Fahy, A., Lethbridge, G., Earle, R., Ball, A.S., Timmis, K.N. and McGenity, T.J. 2005. Effects of long-term benzene pollution on bacterial diversity and community structure in groundwater. Environmental microbiology 7(8), 1192-1199.

Fahy, A., McGenity, T.J., Timmis, K.N. and Ball, A.S. 2006. Heterogeneous aerobic benzene-degrading communities in oxygen-depleted groundwaters. FEMS microbiology ecology 58(2), 260-270.

Falkowski, P.G., Fenchel, T. and Delong, E.F. 2008. The microbial engines that drive Earth's biogeochemical cycles. science 320(5879), 1034-1039.

Farnleitner, A.H., Wilhartitz, I., Ryzinska, G., Kirschner, A.K., Stadler, H., Burtscher, M.M., Hornek, R., Szewzyk, U., Herndl, G. and Mach, R.L. 2005. Bacterial dynamics in spring water of alpine karst aquifers indicates the presence of stable autochthonous microbial endokarst communities. Environmental Microbiology 7(8), 1248-1259.

Fiedler, C.J., Schönher, C., Proksch, P., Kerschbaumer, D.J., Mayr, E., Zunabovic-Pichler, M., Domig, K.J. and Perfler, R. 2018. Assessment of microbial community dynamics in river bank filtrate using high-throughput sequencing and flow cytometry. Frontiers in Microbiology 9, 419356.

Fillinger, L., Hug, K. and Griebler, C. 2019a. Selection imposed by local environmental conditions drives differences in microbial community composition across geographically distinct groundwater aquifers. FEMS microbiology ecology 95(11), fiz160.

Fillinger, L., Hug, K. and Griebler, C. 2021. Aquifer recharge viewed through the lens of microbial community ecology: Initial disturbance response, and impacts of species sorting versus mass effects on microbial community assembly in groundwater during riverbank filtration. Water research 189, 116631.

Fillinger, L., Hug, K., Trimbach, A.M., Wang, H., Kellermann, C., Meyer, A., Bendinger, B. and Griebler, C. 2019b. The DA-(C) index: A practical approach towards the microbiological-ecological monitoring of groundwater ecosystems. Water research 163, 114902.

Fillinger, L., Zhou, Y., Kellermann, C. and Griebler, C. 2019c. Non-random processes determine the colonization of groundwater sediments by microbial communities in a pristine porous aquifer. Environmental microbiology 21(1), 327-342.

Flynn, T.M., Sanford, R.A., Ryu, H., Bethke, C.M., Levine, A.D., Ashbolt, N.J. and Santo Domingo, J.W. 2013. Functional microbial diversity explains groundwater chemistry in a pristine aquifer. BMC microbiology 13, 1-15.

Foster, S. and Bjerre, T.K. 2023. Diffuse agricultural pollution of groundwater: addressing impacts in Denmark and Eastern England. Water Quality Research Journal 58(1), 14-21.

Foster, S. and Chilton, P. 2003. Groundwater: the processes and global significance of aquifer degradation. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 358(1440), 1957-1972.

Gabor, R.S., Baker, A., McKnight, D.M. and Miller, M.P. 2014. Fluorescence indices and their interpretation. Aquatic organic matter fluorescence 303.

Gao, Y. and Wu, M. 2023. Accounting for 16S rRNA copy number prediction uncertainty and its implications in bacterial diversity analyses. ISME communications 3(1), 59.

Geesink, P., Taubert, M., Jehmlich, N., von Bergen, M. and Küsel, K. 2022. Bacterial necromass is rapidly metabolized by heterotrophic bacteria and supports multiple trophic levels of the groundwater microbiome. Microbiology Spectrum 10(4), e00437-00422.

Ghiorse, W.C. and Wilson, J.T. 1988. Microbial ecology of the terrestrial subsurface. Advances in applied microbiology 33, 107-172.

Gibert, J. 1994. Basic attributes of groundwater ecosystems and prospects for research. Groundwater ecology.

Gios, E., Mosley, O.E., Weaver, L., Close, M., Daughney, C. and Handley, K.M. 2023. Ultra-small bacteria and archaea exhibit genetic flexibility towards groundwater oxygen content, and adaptations for attached or planktonic lifestyles. ISME communications 3(1), 13.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V. and Egozcue, J.J. 2017. Microbiome datasets are compositional: and this is not optional. Frontiers in microbiology 8, 2224.

Goldscheider, N., Hunkeler, D. and Rossi, P. 2006. Microbial biocenoses in pristine aquifers and an assessment of investigative methods. Hydrogeology Journal 14(6), 926-941.

Gooddy, D., Hughes, A., Williams, A., Armstrong, A., Nicholson, R. and Williams, J. 2001. Field and modelling studies to assess the risk to UK groundwater from earth-based stores for livestock manure. Soil use and management 17(2), 128-137.

Gooddy, D.C. 2002. Movement of leachate from beneath turkey litter sited over chalk in Southern England. Journal of Environmental Science and Health, Part B 37(1), 81-91.

Gooddy, D.C., Darling, W.G., Abesser, C. and Lapworth, D.J. 2006. Using chlorofluorocarbons (CFCs) and sulphur hexafluoride (SF6) to characterise groundwater movement and residence time in a lowland Chalk catchment. Journal of hydrology 330(1-2), 44-52.

Gregory, S., Maurice, L., West, J. and Gooddy, D. 2014. Microbial communities in UK aquifers: current understanding and future research needs. Quarterly Journal of Engineering Geology and Hydrogeology 47(2), 145-157.

Griebler, C. and Avramov, M. 2015. Groundwater ecosystem services: a review. Freshwater Science 34(1), 355-367.

Griebler, C. and Lueders, T. 2009. Microbial biodiversity in groundwater ecosystems. Freshwater Biology 54(4), 649-677.

Griebler, C., Stein, H., Kellermann, C., Berkhoff, S., Brielmann, H., Schmidt, S., Selesi, D., Steube, C., Fuchs, A. and Hahn, H.J. 2010. Ecological assessment of groundwater ecosystems–vision or illusion? Ecological Engineering 36(9), 1174-1190.

Griffiths, K., Shand, P. and Ingram, J. 2002. Baseline report series. 2, the Permo-Triassic sandstones of west Cheshire and the Wirral.

Gülay, A., Musovic, S., Albrechtsen, H.-J., Al-Soud, W.A., Sørensen, S.J. and Smets, B.F. 2016. Ecological patterns, diversity and core taxa of microbial communities in groundwater-fed rapid gravity filters. The ISME journal 10(9), 2209-2222.

Gunn, J., Downing, R., Price, M. and Jones, G. 1995. The Hydrogeology of the Chalk of North-West Europe. Transactions of the Institute of British Geographers 20(3), 397.

Gutierrez, T. 2019. Aerobic hydrocarbon-degrading Gammaproteobacteria: xanthomonadales. Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes, 191-205.

Hahnke, R.L., Meier-Kolthoff, J.P., García-López, M., Mukherjee, S., Huntemann, M., Ivanova, N.N., Woyke, T., Kyrpides, N.C., Klenk, H.-P. and Göker, M. 2016. Genome-based taxonomic classification of Bacteroidetes. Frontiers in microbiology 7, 2003.

Hammes, F. and Egli, T. 2010. Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications. Analytical and bioanalytical chemistry 397, 1083-1095.

Harjung, A., Schweichhart, J., Rasch, G. and Griebler, C. 2023. Large-scale study on groundwater dissolved organic matter reveals a strong heterogeneity and a complex microbial footprint. Science of The Total Environment 854, 158542.

Harris, J., Thouin, H., Joulian, C., Pinson, S., Charron, M. and Devau, N. 2025. Exploring microbial diversity in relation to hydrological signatures and anthropic pressures in the Beauce Aquifer, France. Science of the Total Environment 977, 179402.

Harrison, J.B., Sunday, J.M. and Rogers, S.M. 2019. Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proceedings of the Royal Society B 286(1915), 20191409.

Hazen, T.C., Jiménez, L., López de Victoria, G. and Fliermans, C.B. 1991. Comparison of bacteria from deep subsurface sediment and adjacent groundwater. Microbial Ecology 22, 293-304.

Hemme, C.L., Tu, Q., Shi, Z., Qin, Y., Gao, W., Deng, Y., Nostrand, J.D.V., Wu, L., He, Z. and Chain, P.S. 2015. Comparative metagenomics reveals impact of contaminants on groundwater microbiomes. Frontiers in microbiology 6, 1205.

Herrmann, M., Rusznyák, A., Akob, D.M., Schulze, I., Opitz, S., Totsche, K.U. and Küsel, K. 2015. Large fractions of CO2-fixing microorganisms in pristine limestone aquifers appear to be involved in the oxidation of reduced sulfur and nitrogen compounds. Applied and Environmental Microbiology 81(7), 2384-2394.

Hiscock, K.M. 2011. Groundwater in the 21st century–meeting the challenges. Sustaining Groundwater Resources: A Critical Element in the Global Water Crisis, 207-225.

Hofmann, R., Uhl, J., Hertkorn, N. and Griebler, C. 2020. Linkage between dissolved organic matter transformation, bacterial carbon production, and diversity in a shallow oligotrophic aquifer: results from flow-through sediment microcosm experiments. Frontiers in microbiology 11, 543567.

Holloway, S. and Savage, D. 1993. The potential for aquifer disposal of carbon dioxide in the UK. Energy Conversion and Management 34(9-11), 925-932.

Hose, G.C., Di Lorenzo, T., Fillinger, L., Galassi, D.M.P., Griebler, C., Hahn, H.J., Handley, K.M., Korbel, K., Reboleira, A.S. and Siemensmeyer, T. 2023. Assessing groundwater ecosystem health, status, and services. Groundwater ecology and evolution, 501-524.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y. and Ise, K. 2016. A new view of the tree of life. Nature microbiology 1(5), 1-6.

Hunt, R.J. and Johnson, W.P. 2017. Pathogen transport in groundwater systems: contrasts with traditional solute transport. Hydrogeology Journal 25(4), 921.

Islam, F.S., Gault, A.G., Boothman, C., Polya, D.A., Charnock, J.M., Chatterjee, D. and Lloyd, J.R. 2004. Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. Nature 430(6995), 68-71.

Jakus, N., Blackwell, N., Straub, D., Kappler, A. and Kleindienst, S. 2021. Presence of Fe (II) and nitrate shapes aquifer-originating communities leading to an autotrophic enrichment dominated by an Fe (II)-oxidizing Gallionellaceae sp. FEMS Microbiology Ecology 97(11), fiab145.

Janda, J.M. and Abbott, S.L. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. Journal of clinical microbiology 45(9), 2761-2764.

John, D.E. and Rose, J.B. 2005. Review of factors affecting microbial survival in groundwater. Environmental science & technology 39(19), 7345-7356.

Kaboré, O.D., Godreuil, S. and Drancourt, M. 2020. Planctomycetes as host-associated bacteria: a perspective that holds promise for their future isolations, by mimicking their native environmental niches in clinical microbiology laboratories. Frontiers in Cellular and Infection Microbiology 10, 519301.

Karwautz, C., Zhou, Y., Kerros, M.-E., Weinbauer, M.G. and Griebler, C. 2022. Bottom-up control of the groundwater microbial food-web in an alpine aquifer. Frontiers in Ecology and Evolution 10, 854228.

Kim, H., Kaown, D., Mayer, B., Lee, J.-Y. and Lee, K.-K. 2018. Combining pyrosequencing and isotopic approaches to assess denitrification in a hyporheic zone. Science of the Total Environment 631, 755-764.

Kolvenbach, B.A., Helbling, D.E., Kohler, H.-P.E. and Corvini, P.F. 2014. Emerging chemicals and the evolution of biodegradation capacities and pathways in bacteria. Current opinion in biotechnology 27, 8-14.

Konopka, A., Lindemann, S. and Fredrickson, J. 2015. Dynamics in microbial communities: unraveling mechanisms to identify principles. The ISME journal 9(7), 1488-1495.

Korbel, K., Chariton, A., Stephenson, S., Greenfield, P. and Hose, G.C. 2017. Wells provide a distorted view of life in the aquifer: implications for sampling, monitoring and assessment of groundwater ecosystems. Scientific reports 7(1), 40702.

Korbel, K., Hancock, P., Serov, P., Lim, R. and Hose, G. 2013. Groundwater ecosystems vary with land use across a mixed agricultural landscape. Journal of Environmental Quality 42(2), 380-390.

Korbel, K. and Hose, G. 2011. A tiered framework for assessing groundwater ecosystem health. Hydrobiologia 661, 329-349.

Korbel, K., McKnight, K., Greenfield, P., Angel, B., Adams, M., Chariton, A. and Hose, G. 2024. Bioassessment of groundwater ecosystems II. Sampling methods and analysis of eDNA for microbes and stygofauna in shallow sandstone aquifers. Report prepared for the Independent Expert Scientific Committee on Unconventional Gas Development and Large Coal Mining Development through the Department of Climate Change. Energy, the Environment and Water. Commonwealth of Australia.

Kruskal, W.H. and Wallis, W.A. 1952. Use of ranks in one-criterion variance analysis. Journal of the American statistical Association 47(260), 583-621.

Kumar, G., Farrell, E., Reaume, A.M., Eble, J.A. and Gaither, M.R. 2022. One size does not fit all: Tuning eDNA protocols for high-and low-turbidity water sampling. Environmental DNA 4(1), 167-180.

Kumar, S., Herrmann, M., Thamdrup, B., Schwab, V.F., Geesink, P., Trumbore, S.E., Totsche, K.-U. and Küsel, K. 2017. Nitrogen Loss from Pristine Carbonate-Rock Aquifers of the Hainich Critical Zone Exploratory (Germany) Is Primarily Driven by Chemolithoautotrophic Anammox Processes. Frontiers in Microbiology 8(1951).

Lazar, C.S., Stoll, W., Lehmann, R., Herrmann, M., Schwab, V.F., Akob, D.M., Nawaz, A., Wubet, T., Buscot, F. and Totsche, K.-U. 2017. Archaeal diversity and CO2 fixers in carbonate-/siliciclastic-rock groundwater ecosystems. Archaea 2017(1), 2136287.

Legendre, P. and Anderson, M.J. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecological monographs 69(1), 1-24.

Li, D., Sharp, J.O., Saikaly, P.E., Ali, S., Alidina, M., Alarawi, M.S., Keller, S., Hoppe-Jones, C. and Drewes, J.r.E. 2012. Dissolved organic carbon influences microbial community composition and diversity in managed aquifer recharge systems. Applied and environmental microbiology 78(19), 6819-6828.

Lin, X., McKinley, J., Resch, C.T., Kaluzny, R., Lauber, C.L., Fredrickson, J., Knight, R. and Konopka, A. 2012. Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. The ISME journal 6(9), 1665-1676.

Liu, C., Cui, Y., Li, X. and Yao, M. 2021. microeco: an R package for data mining in microbial community ecology. FEMS microbiology ecology 97(2), fiaa255.

Liu, T., Kong, W., Chen, N., Zhu, J., Wang, J., He, X. and Jin, Y. 2016. Bacterial characterization of B eijing drinking water by flow cytometry and M i S eq sequencing of the 16S rRNA gene. Ecology and evolution 6(4), 923-934.

Liu, Y., Liu, C., Nelson, W.C., Shi, L., Xu, F., Liu, Y., Yan, A., Zhong, L., Thompson, C. and Fredrickson, J.K. 2017. Effect of water chemistry and hydrodynamics on nitrogen transformation activity and microbial community functional potential in hyporheic zone sediment columns. Environmental science & technology 51(9), 4877-4886.

Liu, Y., Xin, J., Wang, Y., Yang, Z., Liu, S. and Zheng, X. 2022. Dual roles of dissolved organic nitrogen in groundwater nitrogen cycling: Nitrate precursor and denitrification promoter. Science of the Total Environment 811, 151375.

Louca, S., Parfrey, L.W. and Doebeli, M. 2016. Decoupling function and taxonomy in the global ocean microbiome. Science 353(6305), 1272-1277.

Ludington, W.B., Seher, T.D., Applegate, O., Li, X., Kliegman, J.I., Langelier, C., Atwill, E.R., Harter, T. and DeRisi, J.L. 2017. Assessing biosynthetic potential of agricultural groundwater through metagenomic sequencing: A diverse anammox community dominates nitrate-rich groundwater. PLoS One 12(4), e0174930.

Malard, F., Griebler, C. and Retaux, S. (2023) Groundwater Ecology and Evolution, Elsevier Science.

Marston, C., Rowland, C.S., O'Neil, A.W. and Morton, R.D. 2022 Land Cover Map 2021 (1km summary rasters, GB and N. Ireland), NERC EDS Environmental Information Data Centre.

Mathewson, E., Maurice, L., Farrant, A. and Atkinson, T. 2022. BGS karst report series: J2. Karst in the Jurassic limestones of Central England.

Mattsson, M.K., Liu, X., Yu, D. and Kontro, M.H. 2015. Depth, soil type, water table, and site effects on microbial community composition in sediments of pesticide-contaminated aquifer. Environmental Science and Pollution Research 22(13), 10263-10279.

Maurice, L., Farrant, A. and Mathewson, E. 2023a. BGS karst report series: J3. Karst in the Jurassic Great and Inferior Oolite Groups of southern England.

Maurice, L., Farrant, A.R., Mathewson, E. and Atkinson, T. 2023b. Karst hydrogeology of the Chalk and implications for groundwater protection. Geological Society, London, Special Publications 517(1), SP517-2020-2267.

McGill, R., Tukey, J.W. and Larsen, W.A. 1978. Variations of box plots. The american statistician 32(1), 12-16.

Medici, G. and West, L. 2022. Review of groundwater flow and contaminant transport modelling approaches for the Sherwood Sandstone aquifer, UK; insights from analogous successions worldwide. Quarterly Journal of Engineering Geology and Hydrogeology 55(4), qjegh2021-2176.

Mehrshad, M., Lopez-Fernandez, M., Sundh, J., Bell, E., Simone, D., Buck, M., Bernier-Latmani, R., Bertilsson, S. and Dopson, M. 2021. Energy efficiency and biological interactions define the core microbiome of deep oligotrophic groundwater. Nature Communications 12(1), 4253.

Merino, N., Jackson, T.R., Campbell, J.H., Kersting, A.B., Sackett, J., Fisher, J.C., Bruckner, J.C., Zavarin, M., Hamilton-Brehm, S.D. and Moser, D.P. 2022. Subsurface microbial communities as a tool for characterizing regional-scale groundwater flow. Science of the Total Environment 842, 156768.

Mosley, O.E., Gios, E., Weaver, L., Close, M., Daughney, C., van der Raaij, R., Martindale, H. and Handley, K.M. 2022. Metabolic diversity and aero-tolerance in anammox bacteria from geochemically distinct aquifers. Msystems 7(1), e01255-01221.

Murtagh, F. and Legendre, P. 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? Journal of classification 31, 274-295.

NCBI Insights 2021 NCBI Taxonomy to include phylum rank in taxonomic names.

Neumann, I., Brown, S., Smedley, P. and Besien, T. 2003. Baseline report series. 7, the great and inferior oolite of the Cotswolds district.

Newbold, L.K., Taylor, J., Warren, J., Read, D.S., Walsh, K. and Thorpe, A. 2023 Illumina Amplicon Sequencing using two step PCR V.1 UK Centre for Ecology & Hydrology.

Ning, D., Wang, Y., Fan, Y., Wang, J., Van Nostrand, J.D., Wu, L., Zhang, P., Curtis, D.J., Tian, R. and Lui, L. 2024. Environmental stress mediates groundwater microbial community assembly. Nature Microbiology 9(2), 490-501.

Oster, H., Sonntag, C. and Münnich, K. 1996. Groundwater age dating with chlorofluorocarbons. Water Resources Research 32(10), 2989-3001.

Overholt, W.A., Trumbore, S., Xu, X., Bornemann, T.L., Probst, A.J., Krüger, M., Herrmann, M., Thamdrup, B., Bristow, L.A. and Taubert, M. 2022. Carbon fixation rates in groundwater similar to those in oligotrophic marine systems. Nature Geoscience 15(7), 561-567.

Parkes, R.J., Cragg, B., Roussel, E., Webster, G., Weightman, A. and Sass, H. 2014. A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere: geosphere interactions. Marine Geology 352, 409-425.

Pawlowski, J., Apothéloz-Perret-Gentil, L. and Altermatt, F. 2020. Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. Molecular Ecology 29(22), 4258-4264.

Pedersen, K. 1997. Microbial life in deep granitic rock. FEMS microbiology reviews 20(3-4), 399-414.

Pedersen, K., Arlinger, J., Eriksson, S., Hallbeck, A., Hallbeck, L. and Johansson, J. 2008. Numbers, biomass and cultivable diversity of microbial populations relate to depth and borehole-specific conditions in groundwater from depths of 4–450 m in Olkiluoto, Finland. The ISME journal 2(7), 760-775.

Pellizzari, L., Morozova, D., Neumann, D., Kasina, M., Klapperer, S., Zettlitzer, M. and Würdemann, H. 2016. Comparison of the microbial community composition of pristine rock cores and technical influenced well fluids from the Ketzin pilot site for CO 2 storage. Environmental Earth Sciences 75, 1-17.

Pickup, R., Rhodes, G., Alamillo, M., Mallinson, H., Thornton, S. and Lerner, D. 2001. Microbiological analysis of multi-level borehole samples from a contaminated groundwater system. Journal of contaminant hydrology 53(3-4), 269-284.

Powell, K.L., Taylor, R.G., Cronin, A.A., Barrett, M.H., Pedley, S., Sellwood, J., Trowsdale, S.A. and Lerner, D.N. 2003. Microbial contamination of two urban sandstone aquifers in the UK. Water Research 37(2), 339-352.

Prest, E., Hammes, F., Kötzsch, S., van Loosdrecht, M.C. and Vrouwenvelder, J.S. 2013. Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. Water research 47(19), 7131-7142.

Price, M. 1987. Fluid flow in the Chalk of England. Geological Society, London, Special Publications 34(1), 141-156.

Proctor, C.R., Besmer, M.D., Langenegger, T., Beck, K., Walser, J.-C., Ackermann, M., Bürgmann, H. and Hammes, F. 2018. Phylogenetic clustering of small low nucleic acid-content bacteria across diverse freshwater ecosystems. The ISME journal 12(5), 1344-1359.

Pucher, M., Wünsch, U., Weigelhofer, G., Murphy, K., Hein, T. and Graeber, D. 2019. staRdom: versatile software for analyzing spectroscopic data of dissolved organic matter in R. Water 11(11), 2366.

Reiss, J., Perkins, D.M., Fussmann, K.E., Krause, S., Canhoto, C., Romeijn, P. and Robertson, A.L. 2019. Groundwater flooding: Ecosystem structure following an extreme recharge event. Science of The Total Environment 652, 1252-1260.

Retter, A., Haas, J.C., Birk, S., Stumpp, C., Hausmann, B., Griebler, C. and Karwautz, C. 2023. From the mountain to the valley: drivers of groundwater prokaryotic communities along an Alpine River corridor. Microorganisms 11(3), 779.

Risse-Buhl, U., Herrmann, M., Lange, P., Akob, D.M., Pizani, N., Schönborn, W., Totsche, K.U. and Küsel, K. 2013. Phagotrophic protist diversity in the groundwater of a karstified aquifer–morphological and molecular analysis. Journal of Eukaryotic Microbiology 60(5), 467-479.

Rivett, M., Smith, J., Buss, S. and Morgan, P. 2007. Nitrate occurrence and attenuation in the major aquifers of England and Wales. Quarterly Journal of Engineering Geology and Hydrogeology 40(4), 335-352.

Roudnew, B., Lavery, T.J., Seymour, J.R., Jeffries, T.C. and Mitchell, J.G. 2014. Variability in bacteria and virus-like particle abundances during purging of unconfined aquifers. Groundwater 52(1), 118-124.

Ruff, S.E., Humez, P., de Angelis, I.H., Diao, M., Nightingale, M., Cho, S., Connors, L., Kuloyo, O.O., Seltzer, A. and Bowman, S. 2023. Hydrogen and dark oxygen drive microbial productivity in diverse groundwater ecosystems. Nature Communications 14(1), 3194.

Ruff, S.E., Schwab, L., Vidal, E., Hemingway, J.D., Kraft, B. and Murali, R. 2024. Widespread occurrence of dissolved oxygen anomalies, aerobic microbes, and oxygen-producing metabolic pathways in apparently anoxic environments. FEMS Microbiology Ecology 100(11), fiae132.

Saccò, M., Guzik, M.T., van der Heyde, M., Nevill, P., Cooper, S.J., Austin, A.D., Coates, P.J., Allentoft, M.E. and White, N.E. 2022. eDNA in subterranean ecosystems: Applications, technical aspects, and future prospects. Science of the Total Environment 820, 153223.

Saccò, M., Mammola, S., Altermatt, F., Alther, R., Bolpagni, R., Brancelj, A., Brankovits, D., Fišer, C., Gerovasileiou, V. and Griebler, C. 2024. Groundwater is a hidden global keystone ecosystem. Global change biology 30(1), e17066.

Safford, H.R. and Bischel, H.N. 2019. Flow cytometry applications in water treatment, distribution, and reuse: A review. Water research 151, 110-133.

Saiyari, D.M., Chuang, H.-P., Senoro, D.B., Lin, T.-F., Whang, L.-M., Chiu, Y.-T. and Chen, Y.-H. 2018. A review in the current developments of genus Dehalococcoides, its consortia and kinetics for bioremediation options of contaminated groundwater. Sustainable Environment Research 28(4), 149-157.

Sansupa, C., Wahdan, S.F.M., Hossen, S., Disayathanoowat, T., Wubet, T. and Purahong, W. 2021. Can we use functional annotation of prokaryotic taxa (FAPROTAX) to assign the ecological functions of soil bacteria? Applied Sciences 11(2), 688.

Savio, D., Stadler, P., Reischer, G.H., Kirschner, A.K., Demeter, K., Linke, R., Blaschke, A.P., Sommer, R., Szewzyk, U. and Wilhartitz, I.C. 2018. Opening the black box of spring water microbiology from alpine karst aquifers to support proactive drinking water resource management. Wiley Interdisciplinary Reviews: Water 5(3), e1282.

Schloss, P.D. 2024. Rarefaction is currently the best approach to control for uneven sequencing effort in amplicon sequence analyses. Msphere 9(2), e00354-00323.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. and Huttenhower, C. 2011. Metagenomic biomarker discovery and explanation. Genome biology 12, 1-18.

Serène, L., Mazzilli, N., Batiot-Guilhe, C., Emblanch, C., Gillon, M., Babic, M., Dupont, J., Simler, R. and Blanc, M. 2025. Questioning calculation and interpretation of fluorescence indices in natural waters studies. Journal of Hydrology 650, 132524.

Seymour, C.O., Palmer, M., Becraft, E.D., Stepanauskas, R., Friel, A.D., Schulz, F., Woyke, T., Eloe-Fadrosh, E., Lai, D. and Jiao, J.-Y. 2023. Hyperactive nanobacteria with host-dependent traits pervade Omnitrophota. Nature microbiology 8(4), 727-744.

Shand, P., Edmunds, W., Lawrence, A., Smedley, P. and Burke, S. (2007) The natural (baseline) quality of groundwater in England and Wales, Environment Agency.

Shand, P., Tyler-Whittle, R., Besien, T., Peach, D., Lawrence, A. and Lewis, H. 2003. Baseline report series. 6, the Chalk of the Colne and Lee River catchments.

Shand, P., Tyler-Whittle, R., Morton, M., Simpson, E., Lawrence, A., Pacey, J. and Hargreaves, R. (2002) British Geological Survey Commissioned Report No. CR/02/102N, British Geological Survey Keyworth, UK.

Shapiro, S.S. and Wilk, M.B. 1965. An analysis of variance test for normality (complete samples). Biometrika 52(3-4), 591-611.

Shen, Y., Chapelle, F.H., Strom, E.W. and Benner, R. 2015. Origins and bioavailability of dissolved organic matter in groundwater. Biogeochemistry 122, 61-78.

Sidhu, J., Toze, S., Hodgers, L., Barry, K., Page, D., Li, Y. and Dillon, P. 2015. Pathogen decay during managed aquifer recharge at four sites with different geochemical characteristics and recharge water sources. Journal of environmental quality 44(5), 1402-1412.

Sinclair, J. and Ghiorse, W. 1989. Distribution of aerobic bacteria, protozoa, algae, and fungi in deep subsurface sediments. Geomicrobiology Journal 7(1-2), 15-31.

Sinreich, M. and Pochon, A. (2023) EuroKarst 2022, Málaga: Advances in the Hydrogeology of Karst and Carbonate Reservoirs, pp. 85-92, Springer.

Sinreich, M., Pronk, M. and Kozel, R. 2014. Microbiological monitoring and classification of karst springs. Environmental Earth Sciences 71, 563-572.

Sirisena, K.A., Daughney, C.J., Moreau-Fournier, M., Ryan, K.G. and Chambers, G.K. 2013. National survey of molecular bacterial diversity of New Zealand groundwater: relationships between biodiversity, groundwater chemistry and aquifer characteristics. FEMS microbiology ecology 86(3), 490-504.

Sirisena, K.A., Daughney, C.J., Moreau, M., Sim, D.A., Lee, C.K., Cary, S.C., Ryan, K.G. and Chambers, G.K. 2018. Bacterial bioclusters relate to hydrochemistry in New Zealand groundwater. FEMS Microbiology Ecology 94(11), fiy170.

Smedley, P., Neumann, I. and Brown, S. 2005. Baseline report series. 20, the Permo-Triassic sandstone aquifer of Shropshire.

Smedley, P., Neumann, I. and Farrell, R. 2004. Baseline report series. 10, the Chalk aquifer of Yorkshire and North Humberside.

Smith, R., Antoniou, V., Askquith-Ellis, A., Ball, L.A., Bennett, E.S., Blake, J.R., Boorman, D.B., Brooks, M., Clarke, M., Cooper, H.M., Cowan, N., Cumming, A., Evans, J.G., Farrand, P., Fry, M., Harvey, D., Houghton-Carr, H., Howson, T., Jiménez-Arranz, G., Keen, Y., Khamis, D., Leeson, S., Lord, W.D., Morrison, R., Nash, G.V., O'Callaghan, F., Rylett, D., Scarlett, P.M., St Quintin, P., Stanley, S., Swain, O.D., Szczykulska, M., Teagle, S., Thornton, J.L., Trill, E.J., Vincent, P., Ward, H.C., Warwick, A.C., Winterbourn, J.B. 2024 Daily and sub-daily hydrometeorological and soil data (2013-2023) [COSMOS-UK]. . Centre., N.E.E.I.D. (ed).

Smith, R.J., Jeffries, T.C., Roudnew, B., Fitch, A.J., Seymour, J.R., Delpin, M.W., Newton, K., Brown, M.H. and Mitchell, J.G. 2012. Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems. Environmental Microbiology 14(1), 240-253.

Smith, R.J., Paterson, J.S., Sibley, C.A., Hutson, J.L. and Mitchell, J.G. 2015. Putative effect of aquifer recharge on the abundance and taxonomic composition of endemic microbial communities. PloS one 10(6), e0129004.

Smith, R.J., Paterson, J.S., Wallis, I., Launer, E., Banks, E.W., Bresciani, E., Cranswick, R.H., Tobe, S.S., Marri, S. and Goonan, P. 2018. Southern South Australian groundwater microbe diversity. FEMS Microbiology Ecology 94(10), fiy158.

Smouse, P.E., Long, J.C. and Sokal, R.R. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic zoology 35(4), 627-632.

Sorensen, J., Vivanco, A., Ascott, M., Gooddy, D., Lapworth, D., Read, D., Rushworth, C., Bucknall, J., Herbert, K. and Karapanos, I. 2018. Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water. Water Research 137, 301-309.

Sorensen, J.P., Aldous, P., Bunting, S.Y., McNally, S., Townsend, B.R., Barnett, M.J., Harding, T., La Ragione, R.M., Stuart, M.E. and Tipper, H.J. 2021. Seasonality of enteric viruses in groundwater-derived public water sources. Water Research 207, 117813.

Sorensen, J.P., Butcher, A.S., Stuart, M.E. and Townsend, B.R. 2015. Nitrate fluctuations at the water table: implications for recharge processes and solute transport in the Chalk aquifer. Hydrological Processes 29(15), 3355-3367.

Sorensen, J.P., Diaw, M.T., Pouye, A., Roffo, R., Diongue, D.M., Faye, S.C., Gaye, C.B., Fox, B.G., Goodall, T. and Lapworth, D.J. 2020. In-situ fluorescence spectroscopy indicates total bacterial abundance and dissolved organic carbon. Science of the Total Environment 738, 139419.

Sorensen, J.P., Maurice, L., Edwards, F.K., Lapworth, D.J., Read, D.S., Allen, D., Butcher, A.S., Newbold, L.K., Townsend, B.R. and Williams, P.J. 2013. Using boreholes as windows into groundwater ecosystems. PLoS One 8(7), e70264.

Spang, A., Caceres, E.F. and Ettema, T.J. 2017. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science 357(6351), eaaf3883.

Stedmon, C.A. and Bro, R. 2008. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. Limnology and Oceanography: Methods 6(11), 572-579.

Stegen, J.C., Fredrickson, J.K., Wilkins, M.J., Konopka, A.E., Nelson, W.C., Arntzen, E.V., Chrisler, W.B., Chu, R.K., Danczak, R.E. and Fansler, S.J. 2016. Groundwater–surface water mixing shifts ecological assembly processes and stimulates organic carbon turnover. Nature communications 7(1), 11237.

Stegen, J.C., Johnson, T., Fredrickson, J.K., Wilkins, M.J., Konopka, A.E., Nelson, W.C., Arntzen, E.V., Chrisler, W.B., Chu, R.K. and Fansler, S.J. 2018. Influences of organic carbon speciation on hyporheic corridor biogeochemistry and microbial ecology. Nature communications 9(1), 1-11.

Stein, H., Griebler, C., Berkhoff, S., Matzke, D., Fuchs, A. and Hahn, H.J. 2012. Stygoregions–a promising approach to a bioregional classification of groundwater systems. Scientific Reports 2(1), 673.

Stein, H., Kellermann, C., Schmidt, S.I., Brielmann, H., Steube, C., Berkhoff, S.E., Fuchs, A., Hahn, H.J., Thulin, B. and Griebler, C. 2010. The potential use of fauna and bacteria as ecological indicators for the assessment of groundwater quality. Journal of Environmental Monitoring 12(1), 242-254.

Steube, C., Richter, S. and Griebler, C. 2009. First attempts towards an integrative concept for the ecological assessment of groundwater ecosystems. Hydrogeology Journal 17(1), 23.

Stuart, M. and Lapworth, D. 2016. Macronutrient status of UK groundwater: Nitrogen, phosphorus and organic carbon. Science of the total environment 572, 1543-1560.

Stuart, M. and Smedley, P. 2009. Baseline groundwater chemistry: the Chalk aquifer of Hampshire.

Stuart, M., Ward, R., Ascott, M. and Hart, A. 2016. Regulatory practice and transport modelling for nitrate pollution in groundwater.

Taylor, R., Cronin, A., Pedley, S., Barker, J. and Atkinson, T. 2004. The implications of groundwater velocity variations on microbial transport and wellhead protection–review of field evidence. FEMS Microbiology Ecology 49(1), 17-26.

Tettamanti Boshier, F.A., Srinivasan, S., Lopez, A., Hoffman, N.G., Proll, S., Fredricks, D.N. and Schiffer, J.T. 2020. Complementing 16S rRNA gene amplicon sequencing with total bacterial load to infer absolute species concentrations in the vaginal microbiome. Msystems 5(2), 10.1128/msystems. 00777-00719.

Tian, R., Ning, D., He, Z., Zhang, P., Spencer, S.J., Gao, S., Shi, W., Wu, L., Zhang, Y. and Yang, Y. 2020. Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. Microbiome 8, 1-15.

Tomlinson, M. and Boulton, A.J. 2010. Ecology and management of subsurface groundwater dependent ecosystems in Australia–a review. Marine and Freshwater Research 61(8), 936-949.

Tyler-Whittle, R., Brown, S. and Shand, P. 2002. Baseline report series. 3, the Permo-Triassic sandstones of south Staffordshire and north Worcestershire.

U.S. Geological Survey 2019 How Much Water is There on Earth?

Umezawa, K., Watanabe, T., Miura, A., Kojima, H. and Fukui, M. 2016. The complete genome sequences of sulfur-oxidizing Gammaproteobacteria Sulfurifustis variabilis skN76 T and Sulfuricaulis limicola HA5 T. Standards in genomic sciences 11, 1-8.

Van der Gun, J. 2022. Large aquifer systems around the world. The Groundwater Project, Guelph, ON.

van Driezum, I.H., Chik, A.H., Jakwerth, S., Lindner, G., Farnleitner, A.H., Sommer, R., Blaschke, A.P. and Kirschner, A.K. 2018. Spatiotemporal analysis of bacterial biomass and activity to understand surface and groundwater interactions in a highly dynamic riverbank filtration system. Science of the Total Environment 627, 450-461.

Van Nevel, S., Koetzsch, S., Proctor, C.R., Besmer, M.D., Prest, E.I., Vrouwenvelder, J.S., Knezev, A., Boon, N. and Hammes, F. 2017. Flow cytometric bacterial cell counts challenge conventional heterotrophic plate counts for routine microbiological drinking water monitoring. Water Research 113, 191-206.

Vellai, T. and Vida, G. 1999. The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. Proceedings of the Royal Society of London. Series B: Biological Sciences 266(1428), 1571-1577.

Villeneuve, K., Violette, M. and Lazar, C.S. 2022. From recharge, to groundwater, to discharge areas in aquifer systems in Quebec (Canada): shaping of microbial diversity and community structure by environmental factors. Genes 14(1), 1.

Voisin, J., Cournoyer, B., Vienney, A. and Mermillod-Blondin, F. 2018. Aquifer recharge with stormwater runoff in urban areas: Influence of vadose zone thickness on nutrient and bacterial transfers from the surface of infiltration basins to groundwater. Science of the Total Environment 637, 1496-1507.

Vucinic, L., O’Connell, D., Teixeira, R., Coxon, C. and Gill, L. 2022. Flow cytometry and fecal indicator bacteria analyses for fingerprinting microbial pollution in karst aquifer systems. Water Resources Research 58(5), e2021WR029840.

Wang, C., Yang, Y., Xu, X., Wang, D., Shi, X., Liu, L., Deng, Y., Li, L. and Zhang, T. 2025a. The quest for environmental analytical microbiology: absolute quantitative microbiome using cellular internal standards. Microbiome 13(1), 26.

Wang, H., Herrmann, M., Schroeter, S.A., Zerfaß, C., Lehmann, R., Lehmann, K., Ivanova, A., Pohnert, G., Gleixner, G. and Trumbore, S.E. 2025b. Groundwater microbiomes balance resilience and vulnerability to hydroclimatic extremes. Communications Earth & Environment 6(1), 683.

Wang, L., Stuart, M., Bloomfield, J., Butcher, A., Gooddy, D., McKenzie, A., Lewis, M. and Williams, A. 2012. Prediction of the arrival of peak nitrate concentrations at the water table at the regional scale in Great Britain. Hydrological Processes 26(2), 226-239.

Wang, Y., Hammes, F., Boon, N., Chami, M. and Egli, T. 2009. Isolation and characterization of low nucleic acid (LNA)-content bacteria. The ISME Journal 3(8), 889-902.

Wang, Y., Li, P., Li, B., Webster, G., Weightman, A.J., Jiang, Z., Jiang, D., Deng, Y. and Wang, Y. 2014. Bacterial diversity and community structure in high arsenic aquifers in Hetao Plain of Inner Mongolia, China. Geomicrobiology Journal 31(4), 338-349.

Water Situation Report 2023 Monthly water situation report: England, Environment Agency.

Watson, S.E., Taylor, C.H., Bell, V., Hooper, A.S., Bellamy, T.R., Kille, P. and Perkins, R.G. 2024. Utilising eDNA methods and interactive data dashboards for managing sustainable drinking water. Sustainability 16(5), 2043.

Wegner, C.-E., Gaspar, M., Geesink, P., Herrmann, M., Marz, M. and Küsel, K. 2019. Biogeochemical regimes in shallow aquifers reflect the metabolic coupling of the elements nitrogen, sulfur, and carbon. Applied and environmental microbiology 85(5), e02346-02318.

Weitowitz, D., Maurice, L., Lewis, M., Bloomfield, J., Reiss, J. and Robertson, A. 2017. Defining geo-habitats for groundwater ecosystem assessments: an example from England and Wales (UK). Hydrogeology Journal, 1-14.

Whitelaw, K. and Rees, J. 1980. Nitrate-reducing and ammonium-oxidizing bacteria in the vadose zone of the chalk aquifer of England. Geomicrobiology Journal 2(2), 179-187.

Wilcoxon, F. (1992) Breakthroughs in statistics: Methodology and distribution, pp. 196-202, Springer.

Willis, C., Nye, K., Aird, H., Lamph, D. and Fox, A. 2013. Examining food, water and environmental samples from healthcare environments. Microbiological Guideline. Pub Health England pp, 1-48.

Wilson, G.J., Aind, D.A., Mukherjee, A., Polya, D.A., Gooddy, D.C. and Richards, L.A. 2025. Monsoonal rainfall initiates autochthonous alteration of dissolved organic matter composition in Indian groundwaters. Applied Geochemistry, 106298.

Wilson, J.T., McNabb, J., Balkwill, D. and Ghiorse, W. 1983. Enumeration and characterization of bacteria indigenous to a shallow water-table aquifer. Groundwater 21(2), 134-142.

Woese, C.R. 1987. Bacterial evolution. Microbiological reviews 51(2), 221-271.

Woese, C.R., Kandler, O. and Wheelis, M.L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences 87(12), 4576-4579.

Wold, S., Esbensen, K. and Geladi, P. 1987. Principal component analysis. Chemometrics and intelligent laboratory systems 2(1-3), 37-52.

Worthington, S.R. and Ford, D. 2009. Self-organized permeability in carbonate aquifers. Groundwater 47(3), 326-336.

Yan, L., Hermans, S.M., Totsche, K.U., Lehmann, R., Herrmann, M. and Küsel, K. 2021. Groundwater bacterial communities evolve over time in response to recharge. Water Research 201, 117290.

Yan, L., Herrmann, M., Kampe, B., Lehmann, R., Totsche, K.U. and Küsel, K. 2020. Environmental selection shapes the formation of near-surface groundwater microbiomes. Water Research 170, 115341.

Yates, M.V., Gerba, C.P. and Kelley, L.M. 1985. Virus persistence in groundwater. Applied and environmental microbiology 49(4), 778-781.

Zhong, S., Zhou, S., Liu, S., Wang, J., Dang, C., Chen, Q., Hu, J., Yang, S., Deng, C. and Li, W. 2023. May microbial ecological baseline exist in continental groundwater? Microbiome 11(1), 152.

Zhou, J., Xia, B., Treves, D.S., Wu, L.-Y., Marsh, T.L., O'Neill, R.V., Palumbo, A.V. and Tiedje, J.M. 2002. Spatial and resource factors influencing high microbial diversity in soil. Applied and Environmental Microbiology 68(1), 326-334.

Zhou, Y., Kellermann, C. and Griebler, C. 2012. Spatio-temporal patterns of microbial communities in a hydrologically dynamic pristine aquifer. FEMS Microbiology Ecology 81(1), 230-242.