Imperial College London

Characterizing freshwater macroinvertebrates of Bangladesh using metagenetic techniques

Md Mizanur Rahman

A thesis submitted to Imperial College London for the degree of Doctor of Philosophy

March 2022

Department of Life Sciences
Imperial College London

Abstract

The degradation of freshwater ecosystems has become a global concern, in particular, the critical conditions of rivers in Bangladesh demand a monitoring programme through the assessment of bioindicator organisms. Macroinvertebrates as prominent bioindicators are widely used for assessing the health of aquatic ecosystems. Recent technological advances have enabled routine assessment with the genomic characterization of macroinvertebrates using different metagenetic techniques such as DNA barcoding for individual specimen identification, metabarcoding for multi-species identification of bulk samples and mitochondrial metagenomics for extraction of mitogenomes from mixed samples. In this thesis, I commence by generating Cytochrome Oxidase subunit (COI) barcodes for Bangladeshi freshwater macroinvertebrates belonging to the Ephemeroptera, Plecoptera, Trichoptera, Coleoptera, Hemiptera, Odonata, Diptera, Gastropoda and Bivalvia. These barcodes can be used as a DNA reference library for species identification in metabarcoding of macroinvertebrates. I also aim for exploring complete mitogenomes from selected macroinvertebrates using a mitochondrial metagenomic pipeline. I carry out phylogenetic analysis with protein-coding genes that reveals the evolutionary relationship of Bangladeshi macroinvertebrate lineages and also support deeper level identification of barcodes placing them into the phylogenetic tree (chapter 2). In chapter 3, I assess some methodological aspects of the metabarcoding pipeline required for diversity estimation from complex bulk samples of macroinvertebrates in large-scale biomonitoring programmes. These include preparation of bulk macroinvertebrate samples, optimization of the procedure of homogenization of samples required for DNA extraction, strategies for DNA pooling from these extracts, choice of robust universal primers, and viable OTU clustering for reliable diversity estimation. The results have implications for the optimization and standardization of these steps in metabarcoding of freshwater macroinvertebrates. In chapter 4, I apply the metabarcoding technique to establish the macroinvertebrate diversity and impact of various types of anthropogenic disturbances on the freshwater macroinvertebrates in highland and lowland rivers. The results document high diversity, local endemicity and pronounced responses to disturbance in largely unexplored but threatened habitats of Bangladesh. My investigations manifest the viability of metagenetic techniques for applied conservation management as a step towards building a biomonitoring system in freshwater ecosystems globally.

Declaration of Originality

I, Md. Mizanur Rahman hereby declare that the work presented in this thesis is the result of my own research and has not been submitted in any form for the award of a higher education degree elsewhere. Where information has been derived from other sources, I confirm that this has been duly identified and acknowledged in the thesis.

Copyright Declaration

Acknowledgements

First and foremost, all praise to Almighty Allah, the Most Gracious, and the Most Merciful for giving me the blessings, the strength, the chance and endurance to complete this study.

I would like to express my sincere gratitude to my supervisor, Prof. Alfried Vogler, for his time, generous guidance, continual support and encouragement to carry out my PhD study.

I am particularly appreciating Commonwealth Scholarship Commission (CSC) who has awarded me the prestigious Commonwealth Scholarship and funded this project at Imperial College London and Natural History Museum London (NHM). I also gratefully acknowledge the study grants provided by The Charles Wallace Bangladesh Trust and Gilchrist Educational Trust. These grants helped me a lot at the ending stage of my studies.

I would like to thank all members of the Vogler lab for their help and companionship throughout the project. In particular, Thomas Creedy for his relentless efforts for solving bioinformatics problems. I am also grateful to Alfred Burian at the University of Derby, for his suggestions and instructions for statistical analysis. Many other people have provided essential assistance throughout this study: my field assistants Abdullah Helal, Suman Mandal, Kazi Mamun-ur Rashid and Zatlong Vom; Vladimir Blagoderov at the Sackler Biodiversity Imaging Lab, NHM, for access to microscopes and training; Stephen Russell and Elina Lugli, at the Molecular Lab, NHM for advice and assistance for wet-lab activities. I appreciate Steve Brook and Robert Angus for their help to identify some morphospecies. Special thanks to Peter Foster who maintained NHM's computing resources servers that I used herein for bioinformatic analyses.

Yet all of this wouldn't have been possible without the support of my family. I dedicate this thesis to the memories of my beloved father and father in law, who left me in the period of this study. They always believed in my ability to be successful in every sphere of life. You are no more but still, I feel your inspiration and best wishes to move forward. I always express my deep sense of gratitude and indebtedness to my mother who has been with me at every turn. Last but not least, I owe thanks to a very special person, my wife, Mahmuda for her unfailing love, support and understanding during my pursuit of PhD study that made the completion of my thesis possible.

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Chapter 1: General Introduction

1.1 Freshwater biodiversity and biomonitoring

Freshwater ecosystems play a fundamental ecological role, including economically important services to humans (e.g., drinking water, flood control, food production) and provide a vast range of habitats for many aquatic plants and animals. Unfortunately, the sustainability of freshwater biodiversity has been greatly affected across the planet by a wide range of interacting stressors including pollution, habitat loss, over-exploitation, invasion of alien species and climate change. The increasing trend of these stressors, particularly the growing pollutants and habitat degradation in freshwater bodies, is causing biodiversity loss and limiting ecosystem services (Arthington et al., 2010; Chapin et al., 2000; Loreau et al., 2001), which has become a global ecological and environmental concern (Cai, Varis, and Yin, 2017; Nyenje et al., 2010; Xiong and Zhan, 2018; Vörösmarty et al., 2010). These intense problems have escalated the pressing need for comprehensive monitoring of ecological alternations, causes and consequences of rapid changes of community structure to restore impaired ecosystems in management programmes (Geist, 2011).

There are several ways to assess the water quality of lotic and lentic freshwater ecosystems; traditionally, river or stream health assessments are based on the analysis of physical and chemical data of water including temperatures, turbidity, pH, total dissolved solids (TDS), nitrate, phosphate, biological oxygen demand (BOD), dissolved oxygen (DO) and heavy metals. Basically, these parameters inadequately assess the status of impaired waters with a snapshot of the condition of a water body at the moment of measurement and these can change over time, which lacks the integrative measure of the overall health of a stream (United States Environmental Protection Agency, 2005). Instead, biological monitoring or 'biomonitoring' provide an integrated, comprehensive assessment of the health of a water body over time using biological indicator organisms. Generally, biomonitoring of freshwaters entails the surveillance of aquatic organisms (bioindicators) through the detection of their changes in diversity, abundance, and behaviours due to perturbation or pollution (Li, Zheng and Liu, 2010). Bioindicator organisms respond to any morphological, chemical and biological degradation in rivers or streams because of their diversity, specific habitat preference, significant ecological role in the food chain (Karr, 1999; Kenney et al., 2009), sedentary habit and a relatively long life-cycle in water (Stark and Maxted, 2007). They live

with the changes that occurred in the aquatic environment and their positive or negative responses are reflected in taxonomic and functional diversity with the intensity of stresses (Rosenberg and Resh, 1993). Therefore, bioindicator organisms are used for monitoring freshwater ecosystems rather than physical and chemical elements (Stark and Maxted, 2007). In aquatic ecosystems, macroinvertebrates are the most frequently utilized bioindicators though there are several alternatives (e.g., algae, periphytons, diatom and fish) used as indicators in streams and rivers (Hodkinson and Jackson, 2005; Li, Zheng and Liu, 2010; Neville and Yen, 2007).

1.2 Macroinvertebrates as potential bioindicators in freshwater ecosystems

As potential ecological indicators, the sensitivity of macroinvertebrates to environmental disturbances has been well established relative to other indicator organisms in freshwater ecosystems (Lamoureaux et al., 2004; Rosenberg and Resh, 1993). In a stream, the hydro-morphology, physical and chemical properties, nutrient availability, and ecological processes are the key drivers of macroinvertebrate community composition (Heino, 2014). They are the crucial components of aquatic food chains for cycling organic matter and nutrient resources from lower to upper trophic levels (Wallace and Webster, 1996). Having sedentary habits of larvae or nymphs and a relatively long-life cycle, most benthic macroinvertebrate organisms are representative of site-specific ecological conditions (Bonada et al., 2006; Hutchinson et al., 1998; Rosenberg and Resh, 1993). They respond to a range of environmental stressors (e.g., organic and inorganic pollutants, habitat destruction and climate change etc.) by their presence/absence, abundance and functional behaviour (Li, Zheng and Liu, 2010).

Freshwater macroinvertebrates (e.g., insects, molluscs, and annelids) have been the most commonly used focal groups for assessing the ecological quality of freshwater ecosystems worldwide (Bady et al., 2005; Barbour et al., 1999; Bonada et al., 2006; Carew et al., 2013; Clews et al., 2014; Clarke et al., 2003; Hajibabaei et al., 2012; Menezes et al., 2010; Lau and Lauer, 2015; Lakew and Moog, 2015; Nichols and Dyer, 2013; Serrana et al., 2019). Among them, mayflies (Ephemeroptera) and stoneflies (Plecoptera) are highly sensitive to most of the stressors while some species of caddisflies (Trichoptera) are somewhat tolerant to some perturbations. Insects in the orders Ephemeroptera, Plecoptera and Trichoptera are collectively called EPTs, who rely on freshwater bodies to complete their lifecycles and have

specific habitat requirements for their functional diversity such as scrapers or grazers, shredders, collectors or gatherers, filterers, and predators. Therefore, they show varied sensitivity to different disturbance regimes in their habitats and respond to distinct states of environmental gradients (Aagaard et al., 2004; Ekrem, Willassen and Stur, 2007; Verneaux and Verneaux, 2002; Webb et al., 2012).

Odonata (dragonflies) are dependent on water for the development of their pre-adult stages (nymphs or larvae) including food and shelters (Casas et al., 2018; Kalkman et al., 2008). Having a unique predatory behaviour and different tolerance limits to contaminants, odonates are considered to be good indicators of environmental health and water quality (Casas et al., 2018; Hart et al., 2014; Quisil et al., 2014). Water beetles (aquatic Coleoptera) hold significant ecological importance providing nutrients from allochthonous organic matters for primary production in freshwater ecosystems (Heino et al., 2008; Kagalou et al., 2006). The composition of the aquatic beetle community can be affected by various environmental factors, including altered hydro-morphology, land use, vegetation cover and water chemistry (Bloechl et al., 2010). Because of their high species diversity and narrow tolerance to ecological conditions, they have been considered suitable indicator species in Europe and the USA (Dong et al., 2014; Miserendino and Archangelsky, 2006).

The role of water bugs as predators, prey and scavengers makes them ecologically important taxa in any freshwater ecosystem. For instance, water bugs in the family Corixidae are important food items in the diet of many aquatic invertebrates and vertebrates. The majority of water bugs are known to be highly pollution tolerant, and some are limited to similarly restricted habitats and certainly indicative of such limited ecotypes (Epler, 2006). The non-biting midges, Chironomidae, are well known to show moderate to high tolerance to pollution and indicate the ecological status of environmental gradients. The species composition of chironomids is frequently used to assess and monitor the health of rivers and streams (e.g., Aagaard et al., 2004; Ekrem, Willassen and Stur, 2007; Verneaux and Verneaux, 2002).

Therefore, the characterization and identification of these macroinvertebrate taxa is central to the biomonitoring programmes for measuring diversity metrics or indices (e.g., richness, evenness, biotic indices, multimetric indices), and determining functional feeding groups (FFGs) as well as potential indicators (Buss et al., 2015; Li, Zheng and Liu, 2010). However, the key challenge of these approaches is the identification of diverse macroinvertebrate

groups to explore comprehensive diversity information using reliable, rapid, and cost-effective techniques.

1.3 Characterization of aquatic macroinvertebrates using traditional methods

Characterization and identification of macroinvertebrate communities is a pivotal step for developing metrics, indices, or other measures in any biomonitoring scheme of freshwater ecosystems. As the identification in traditional biomonitoring mainly relies on morphological features of mature organisms, taxonomic assignment of immature and damaged specimens to a genus or species is a longstanding impediment in biological elements assessments (Carew et al., 2013; Pfrender et al., 2010). For example, small organisms especially the nymph and larval stages of mayflies, stoneflies, caddisflies, and dragonflies are often difficult or impossible to identify at finer taxonomic resolution (Sweeney et al., 2011). Thereby, in spite of substantial effort devoted to characterizing macroinvertebrate communities with morphological features, it produces only low diversity coverage with a coarse taxonomic assignment and limits to observations on highly restricted sets of invertebrates (Bonada et al., 2006; Hajibabaei et al., 2011). A coarse level identification might mislead the ecological assessments as species of the same genus or family sometimes responds to different stressors (Lenat and Resh, 2001; Pilgrim et al., 2011; Sweeney et al., 2011). The morphological investigation is also not effective for cryptic species identification even for adults with distinctive characters (Pfrender et al., 2010; Shen et al., 2016). All of these limitations in identification and characterization ultimately lead to errors and imprecision in assessments of habitat and water quality (Lenat and Resh, 2001; Stribling et al., 2008). In addition to these, morphology-based species identification is a time-consuming, expensive and laborious approach that requires extensive taxonomic expertise (Aylagas et al., 2014; Hajibabaei et al., 2011; Yu et al., 2012). Therefore, morphological taxonomic processing of macroinvertebrates is a significant bottleneck in the development and operation of large-scale bioassessment programs. Although macroinvertebrates identification and characterization still largely rely on traditional morphological examination, over the past decade this has been supplemented by a range of DNA-based genomic techniques that allow standardized identification of a wide range of taxa without specialist taxonomic expertise.

1.4 DNA based genomic approach for large scale assessment of macroinvertebrates

Genomic approaches are perhaps the most powerful innovation in biodiversity studies, using DNA or RNA sequences of a gene or a part of a gene (called barcode markers) or of whole or partial (mitochondrial and nuclear genomes to analyse biological systems. The development of genomic techniques holds great promises for identification, characterization, and monitoring of the biodiversity in different ecosystems (e.g., Andújar et al., 2015; Arribas et al., 2016; Bourlat et al., 2013; Carew and Hoffmann., 2015; Elbrecht et al., 2017; Kuntke et al., 2020; Stein et al., 2014). With the rapid advancement in sequencing technologies the amount of genetic data on organisms, communities, and habitats has been enormously increased since the beginning of the 2000s (Bik et al., 2012; Bourlat et al., 2013; Hajibabaei et al., 2011; Mardis, 2008; Ratnasingham and Hebert, 2007). As a result of this development, the application of DNA data has gained scientific acceptance to be included for routine methodology in most biological disciplines, including freshwater biodiversity (Blackman et al., 2019; Elbrecht and Steinke, 2019).

Among the DNA-based approaches, both genetic and genomic techniques are being used in biodiversity studies where genetics generally deals with the structure, composition and role of a single gene whereas genomics includes all genes and their combined effects on different attributes of the organism. For instance, DNA barcoding as one of the genetic techniques, initially offered an opportunity to identify species avoiding the bottleneck of traditional taxonomy. DNA barcoding involves the exploration of standard marker genes known as DNA barcodes (e.g., the mitochondrial gene cytochrome c oxidase I, COI in animals) for identifying taxa with greater accuracy compared to morphological methods (Hebert et al., 2003). It provides a platform to promote the cataloguing of biodiversity with species delineation, identification of cryptic species and discovery of new species (Ball et al., 2005; Joly et al., 2014; Hebert et al., 2003; Zhou et al., 2011). However, large scale assessment of biodiversity with Sanger barcodes is inefficient to assign taxonomies to hundreds or thousands of samples (Yu et al., 2012). Although DNA barcoding requires an individual PCR amplification for each specimen, thousands of these amplified specimens can be combinedly sequenced using the NGS sequencing platform bypassing Sanger sequencing. Therefore, DNA barcoding is still an effective tool, and its application is being advanced to reliably identify aquatic macroinvertebrates (Webb et al., 2012) for building barcodes reference databases coupled with their morphological identification. These identified barcodes are also

indispensable for evaluating the metabarcodes and mitogenome sequences (Arribas et al., 2016; Bista et al., 2016) explored from other metagenetic and metagenomic approaches (i.e., metabarcoding and mitochondrial metagenomics). However, to overcome the time-consuming process of conducting single PCR and limitations for simultaneous identification of large numbers of specimens, high-throughput sequencing (HTS) technologies are being used to allow large-scale identification in a massively parallel manner. In this study, COI barcoding was used for developing a reference database of Bangladeshi freshwater macroinvertebrates and for their subsequent use in metabarcoding and mitochondrial metagenomic techniques.

Generally, metagenetics refers to wide-ranging analyses of biodiversity through the amplification and sequencing of homologous genes (Creer et al., 2010) of a community. While Sanger sequencing-based standard barcoding is not an ideal tool for investigating highly abundant and diverse community macroinvertebrate samples, DNA metabarcoding promotes the characterization of species composition in bulk samples or environmental DNA samples (Alberdi et al., 2018; Aylagas et al., 2016; Cristescu, 2014; Deagle et al., 2014) through mass amplification and sequencing of a standard marker gene of a community. Metagenomics in particular mitochondrial metagenomics is an approach (MMG) where PCR free shotgun sequencing produces mitogenomes from mixed or environmental samples (Arribas et al., 2016). MMG explores the mitogenomes of entire communities of organisms (Thomas et al., 2012) and those mitogenomes have greater potential for taxonomic and phylogenetic characterization of a community because of their greater power in phylogenetic analysis.

The availability of HTS platforms to generate millions of DNA sequences simultaneously (Gibson et al., 2015: Shokralla et al., 2015) and the ecologists' need for high-throughput taxa identification have facilitated the application of DNA metabarcoding and MMG in biomonitoring programs (Aylagas et al., 2014; Bourlat et al., 2013; Dowle et al., 2015). HTS permits massive parallel multiplex sequencing to explore the taxonomic composition of bulk or environmental samples at a very low cost (e.g., Gill et al., 2013; Mardis, 2008; Sweeney et al., 2011). Such a dramatic leap in sequencing capacity has revolutionized many areas of scientific inquiry in particularly for biodiversity assessment in freshwater ecosystems (Taberlet et al., 2012). Therefore, to address the challenge of traditional assessment of ecosystems, barcoding (Cordero et al., 2017; Morinière et al., 2017; Sweeney et al., 2011;

Stein et al., 2013; Trebitz, et al., 2015; Weigand et al., 2019), metabarcoding (Andújar et al., 2018; Beng et al., 2016; Braukmann, et al., 2019; Bush et al., 2019; Elbrecht et al., 2017; Oliverio et al., 2018; Yu et al., 2012) and metagenomics (Andújar et al., 2015; Arribas et al., 2016; Crampton-Platt et al., 2016; Gomez-Rodriguez et al., 2015) are widely employed in the characterization of biodiversity and observations of ecosystem structure and function (Alberdi et al., 2018; Baird and Hajibabaei, 2012; Ji et al., 2013; Porter and Hajibabaei, 2018). These genomic applications in biodiversity assessment provide rapidly greater resolution, depth and consistency in the identification and characterization of organisms at a lower cost than morphological approaches (Bohan et al., 2017; Cristescu et al., 2014; Evans et al., 2016; Gibson et al., 2015; Serrana et al., 2019).

Despite the potentiality of these metagenetic approaches for comprehensive biodiversity study in freshwater ecosystems, the performance of these novel methods fluctuates due to various issues arising from sample preparation to downstream analysis. These major challenging issues in the metabarcoding pipeline include maximizing the inclusion of organisms' DNA from bulk samples through viable sample preparation techniques, primers choice for target markers, minimizing PCR and sequencing error, and the optimization of clustering of reads for OTUs recovery. In metabarcoding, the preparation of bulk samples is a vital step for retaining all life stages (eggs, larvae, pupae, adults) of organisms and removing unwanted debris that may affect DNA extraction and PCR. Hence, specimen sorting, exclusion of unwanted matters and sample preparation using colloidal solution (e.g., Ludox) have been employed in several studies to ensure the maximum inclusion of organisms from a complex bulk sample (Arribas et al., 2016; Creedy, Ng and Vogler, 2019; Elbrecht, Peinert and Leese, 2017). However, the findings of these studies are still inconclusive for choosing the most effective approach for the processing of complex bulk samples in a cost and time-saving manner.

Primer choice is one of the initial critical steps in a metabarcoding study to amplify the gene marker of target organisms which is often guided by using degenerate primers. Due to biases from the PCR primers (Clarke et al., 2014), DNA from bulk samples tends to be often differentially amplified (Elbrecht and Leese, 2015; Piñol, et al., 2019). It is virtually impossible to design such primers without mismatches to some of the target species. Therefore, the efficacy of selected primers should be tested prior to their application to amplify community organisms in large-scale biodiversity assessment. Clustering of

sequences into Operational Taxonomic Units (OTUs), based on a similarity threshold is another challenge to approximate species numbers. Although this approach can remove sequence artefacts, it may cause over or underestimation of OTU counts by over splitting and lumping of sequences in the metabarcoding pipeline (Clare et al., 2016). In MMG, the major challenges include the pooling of equimolar concentration of genomic DNA from taxonomically distant groups during library preparation, retrieving the expected number of reads for each specimen from the sequencing platform, annotating mitochondrial genes and finally extracting complete mitogenomes for the target taxa.

Therefore, the improvement of these emerging genomic approaches is still a great priority for the scientific community which necessitates their standardization and validation for large-scale biomonitoring in freshwater ecosystems. In this study, we aimed for applying these metagenetic techniques to characterize Bangladeshi macroinvertebrate communities focusing on some methodological issues, diversity assessment and environmental degradation in river ecosystems.

1.5 Freshwater ecosystems monitoring and macroinvertebrate studies in Bangladesh

Bangladesh is exceptionally endowed with a vast variety of flora and fauna due to its unique geophysical location. About seven hundred rivers and numerous open water bodies (floodplains, smaller creeks, ponds and lakes) seasonally cover more than 50 per cent of the country's land surface, which are known to be rich in aquatic biodiversity. IEDS (2003) estimated over 3000 species of plants and 300 species of fish and other aquatic fauna depend on freshwater for the whole or part of their life cycle. Unfortunately, with the deteriorating trend of global freshwater ecosystems, the rivers in Bangladesh are experiencing critical conditions due to huge population pressure, rapid uncontrolled industrialization, discharge of chemical pollutants, destruction of natural water bodies and ultimately global climate change (Ahmed et al., 2011b; Akter, Kurisu and Hanaki, 2017; Hasan, Shahriar and Jim, 2019; Kamal, Malmgren-Hansen and Badruzzaman, 1999; Majumder, 2009). The status of inland water ecosystems showed that the globally most threatened river catchments are to be found in the Indian subcontinent (UN World Conservation Monitoring Centre) including Bangladesh. Freshwater bodies both in highland and lowland rivers impacted by a wide range of anthropogenic perturbations or stressors have been a serious environmental concern.

However, a bio-surveillance system is not adequately established in the country to monitor the health of freshwater ecosystems.

Until 1990, water management in Bangladesh was mainly focused on controlling floods and improvement of drainage and irrigation systems that severely impacted other sectors, in particular aquatic biodiversity and ecosystems (NWPo 1999). In the past several decades, the Government has been working towards Integrated Water Resources Management (IWRM), which is a comprehensive water management concept integrating water, land, and related sectors to gain maximum economic and social welfare with a sustainable environment. Consequently, Bangladesh has prepared a National Water Policy (NWPo 1999) and a National Water Management Plan (NWMP 2001) that are the principal frameworks for applying integrated water resource practices in Bangladesh (Alam and Quevauviller, 2013). The NWPo sets out various provisions and highlights the importance of the protection, restoration, and preservation of aquatic ecosystems and biodiversity. The policy states that the Government will strengthen appropriate monitoring organisations for tracking groundwater recharge, surface and groundwater use, and changes in surface and groundwater quality (Article 4.7, NWPo). The NWPo triggered the formulation of the National Water Management Plan (NWMP) which is a rolling framework and consists of immediate (short-term), indicative (medium-term), and perspective (longer-term) plans. This government-approved framework outlines a series of programmes under eight major clusters, of which the cluster Environment and Aquatic Resources concerns the assessment, monitoring and preservation of the aquatic and water-depended ecosystems through the implementation of different sub-programmes. Among others, the National Water Quality Monitoring (NWQM) programme is associated with the assessment and monitoring of surface waters (e.g., rivers, streams, floodplains, coastal waters etc.). This programme advocates for bioindicator species to assess the ecological status of water bodies of the country, albeit there is no clear indication for comprehensive assessment methods. The Department of Environment (DoE) is the government-assigned organization for monitoring water quality and their activities are apparently limited to the physical and chemical assessment of inland water bodies (DoE, 2016). WARPO (Water Resources and Planning Organization) is mandated for macro-level planning for water resources and implementing of different projects and is also responsible to maintain, update and disseminate the National Water Resources Database (NWRD). NWRD is relatively enriched with geospatial and physicochemical data but still lacks the information for bioindicators data except for fishes. This scarcity of bioindicator data necessitates an

expanded bioassessment program with characterization of macroinvertebrates to evaluate the ecological status in order to prioritise conservation and devise management plans.

Besides government initiatives, there is a dearth of studies for aquatic biodiversity and biomonitoring in the country. Many studies were performed for detecting the physico-chemical properties of river water, heavy metal concentrations in water, sediments, fish and shellfish of different rivers (e.g., Ahmed et al., 2009; Ahmed et al., 2013; Ahmed et al., 2007; Bhuiyan et al., 2015; Bhuyan et al., 2019; Chakraborty et al., 2013; Fatema et al., 2018; Islam et al., 2018; Kamal et al., 2007; Mokaddes et al., 2013; Sarkar et al., 2015). Conversely, the country has no adequate updated aquatic macroinvertebrate diversity data except for some sporadic studies on the diversity of benthic fauna, aquatic insects, mosquitoes, odonates, hemipterans, oligochaetes and molluscs (e.g., Ahad et al., 2012; Bashar et al., 2014; Ali and Issaque, 1975; Ali et al., 1978; Begum, Ismail and Ali, 1989; Chowdhuri and Aktaruzzaman,1981; Khan, Rahman and Islam, 1997; Mustafa et al., 2013; Nasiruddin et al., 2014; Price et al., 2016; Sana and Ali, 2011). Biomonitoring studies in the country include only a recently developed biotic index for lakes using traditional biodiversity assessment techniques (Chowdhury et al., 2016).

In contrast, many well defined biotic indices have been employed globally in biomonitoring of rivers, streams and lakes such as the Biological Monitoring Working Party Score System (BMWP; Armitage et al., 1983; Blakely et al., 2014; Ghetti, 1997; Hilsenhoff, 1987; Stark and Maxted, 2007), Average Score Per Taxon (ASPT), the River InVertebrate Prediction And Classification System (RIVPACS), and AUSRIVAS (Clarke et al., 2003; Hawkes, 1997; Simpson and Norris, 2000; Wright et al., 2000). Different multimetric indices (Baptista et al., 2007; Cho et al., 2011; Jun et al., 2012; Hering et al., 2006; Klemm et al., 2003; Moya et al., 2011; Stoddard et al., 2008) are also in place though most of these are based on traditional methods (Clews et al., 2014; Flotemersch et al., 2006; Li and Liu, 2010; Phen et al., 2014; Subramanian and Sivaramakrishnan, 2007). In addition, specific bioassessment methods or protocols have been developed in many countries e.g., in China (Wu et al., 2015; Wang et al., 2014), Africa (Lakew and Moog, 2015), Singapore (Blakely et al., 2014), New Zealand (Gray and Harding, 2012), Thailand (Boonsoong, Sangpradub and Barbour, 2009), Vietnam (Nguyena et al., 2014) and in the Hindu Kush-Himalayan region (Ofenbock, et al., 2010). It is noteworthy that the EU Water Framework Directive (WFD, 2000/60/EC) is a key legislative framework for basin-wide integrated water resources management aiming at

achieving the good ecological and chemical status of all surface waters in Europe (Leese et al., 2016; Alam and Quevauviller, 2013). In the context of this framework, ecological status is assessed through the analyses of biological quality elements (BQEs) and then deteriorated water bodies are identified and restored. Furthermore, DNAqua-net, an international EU research network, has been working to adapt DNA-based methods to the WFD and identified the challenges, impacts and potentials of metagenetic approaches in ecological assessment of aquatic ecosystems (Leese et al., 2016). A comparative study (Alam and Quevauviller, 2013) of Bangladeshi water management practices with WFD supported this framework to adapt for the effective implementation of ecological assessment programmes. All of these developments could be used for IWRM implementation, in particular for large scale river monitoring in developing countries like Bangladesh, taking into account the socio-economic and other relevant issues. Therefore, in the context of the country's poorly studied invertebrate fauna, immense anthropogenic pressures on aquatic ecosystems, the present study aimed for the DNA-based assessment of macroinvertebrates to establish their diversity and responses to current environmental degradation in upland and lowland rivers of Bangladesh. Bioassessment of aquatic invertebrates through metagenetic approaches, targeted by this study, is the first step toward future use in the ecological status assessment of rivers and lakes across the country.

1.6 General aims and structure of the thesis

The mainstay of the project was: a) characterization of macroinvertebrates through barcoding and mitochondrial metagenomics (MMG); b) standardization of some methodological aspects of macroinvertebrate metabarcoding for large-scale monitoring of freshwater ecosystems and c) the application of metabarcoding to assess macroinvertebrates community structures and environmental degradation in river ecosystems. Firstly, COI barcodes and mitogenomes of morphologically identified morphospecies were explored through DNA barcoding and MMG to develop reference databases for the upland and lowland rivers. This database, in particular the mitogenomes will contribute to the identification and taxonomic assignment of metabarcoding led OTUs in bulk community samples. The mitogenome based phylogeny of macroinvertebrates was studied to provide an insight into the evolutionary relatedness of local fauna and the identification of anonymous barcodes/OTUs through placement within mitogenome based phylogenetic trees. Bulk sample processing, replication of homogenate

samples, testing of primer efficacy and OTU clustering were investigated to produce maximum outputs in the metabarcoding of macroinvertebrates. Finally, metabarcoding led biodiversity data were analysed to explore macroinvertebrate diversity and the ecological status of streams and rivers along environmental gradients in Bangladesh. The thesis is divided into the following four chapters:

Chapter 1: General Introduction

A general overview of freshwater biomonitoring using macroinvertebrates is provided in this chapter. The current scenario of traditional monitoring approaches and DNA-based developments and their applications in ecological assessment and biodiversity monitoring are discussed. The limitation areas of current genomic methods for biomonitoring were outlined and possible improvement areas were also identified. A brief description of Bangladeshi freshwater bodies including currently explored macroinvertebrate diversity, monitoring status contrasting global aspects are given. The rationale, aims, objectives and chapter-wise thesis structure of the present study were placed in this chapter.

Chapter 2: DNA barcode database for Bangladeshi freshwater macroinvertebrates and their mitogenome based phylogeny

The main aspect of this chapter was the construction of a DNA barcode reference database for Bangladeshi freshwater macroinvertebrates with an emphasis on ecological indicator species in highland and lowland rivers. Seven main groups of macroinvertebrates including the orders Ephemeroptera, Plecoptera, Trichoptera, Hemiptera, Odonata, Coleoptera and Diptera were targeted from highland and lowland rivers to build this barcode library. Attempts have also been made to explore DNA barcodes from the phylum Annelida and Mollusca. The effectiveness of different species delimitation methods was shown with extracted barcodes from major taxa. The levels of divergence of these OTUs/species within major groups of macroinvertebrates and their species richness and composition in river ecosystems were explored. The potential of the MMG pipeline for mitogenomes exploration from bulk genomic DNA of macroinvertebrates was investigated. The mitogenomes based phylogenetic study of macroinvertebrates was performed and the identification of barcodes was also checked placing them in phylogenetic trees.

Chapter 3: Metabarcoding for high-throughput freshwater bioassessment: prospects and methodological challenges

In this chapter, I explored various aspects of a standard metabarcoding method, considering the field-based sampling, laboratory-based extraction of specimens and DNA, and bioinformatics protocols, considering of challenges underlying the metabarcoding pipelines. This chapter mainly aims for the application of metabarcoding for macroinvertebrate community assessment investigating the effects of potential sample preparation methods on OTUs/species explorations, the optimization of the number of homogenized samples for DNA extraction and the effect of DNA pooling (before PCR) on final outputs. In addition, primers' efficacy on different macroinvertebrate groups and the impact of different OTU clustering techniques on community diversity were also tested.

Chapter 4: Metabarcoding of macroinvertebrates to assess diversity and environmental degradation in river ecosystems of Bangladesh

This chapter mainly aimed at establishing the diversity (alpha and beta) of freshwater invertebrates and assessing the impact of anthropogenic activities on aquatic macroinvertebrates in the highland and lowland rivers of Bangladesh. Metabarcoding was used to evaluate the degree to which total species diversity of macroinvertebrates and the diversity of the disturbance-sensitive and pollution tolerant macroinvertebrate taxa were impacted by human-induced stressors. Secondly, I screened for potential indicator species associated either with poor or good ecological status by correlation with a set of environmental variables evaluated for each sampling site.

1.7 References

Aagaard, K., Solem, J.O., Bongard, T. and Hanssen, O., 2004. Studies of aquatic insects in the Atna River 1987–2002. *Hydrobiologia*, *521*(1), pp.87-105.

Ahad, M. A., Raja, S. I., Khatun M. M. and Hossain, M. A., 2012. Aquatic Insect Biodiversity of Northern Region of Bangladesh. *International Journal of Bio-resource and Stress Management*, *3*(3):387-393

Ahmed, M. K., Ahmed, S., Rahman, S., Haque, M.R., and Islam, M.M., 2009. Heavy metals concentration in water, sediments and their bioaccumulation in some freshwater fishes and mussel in Dhaleshwari River, Bangladesh. *Terr. Aquat. Environ. Toxicol*., *3* (1), pp. 33–41.

Ahmed, M.K., Das, M., Islam, M.M., Akter, M.S., Islam, S. and Al-Mansur, M.A., 2011. Physico-chemical properties of tannery and textile effluents and surface water of River Buriganga and Karnatoli, Bangladesh. *World Applied Sciences Journal*, *12*(2), pp.152-159.

Ahmed, S. M., Chowdhury, M. M. K, and Nahar, L., 2019. Molecular characterization of small indigenous fish species (SIS) of Bangladesh through DNA barcodes. *Gene, 684*, pp.53-57.

Akter, K.S., Kurisu, K. and Hanaki, K., 2017. Water use and pollution recognition from the viewpoint of residents in Dhaka, Bangladesh. *Water*, *9*(5), p.331.

Alam, M. and Quevauviller, P., 2013. A Comparative Analysis of Bangladesh Water Management Practice with European Water Framework Directives. *Engineering International*, *1*(2).

Alberdi, A., Aizpurua, O., Gilbert, M.T.P. and Bohmann, K., 2018. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), pp.134-147.

Ali, S. M. And Issaque, A.Q.M., 1975. A systematic study of freshwater Oligochaeta from Dacca city, Bangladesh. Bangladesh. *Bangladesh J. Zool*., 3(1), pp.55-61.

Ali, S., Haque, A.S.M. , Oppenheimer, J.R. and Aziz, K.M.S., 1978. Studies on the bottom fauna of three fish ponds in Dacca city, Bangladesh. *Bangladesh J. Zool. 6*(1), pp.43-55.

Aylagas, E., Borja, Á., Irigoien, X. and Rodríguez-Ezpeleta, N., 2016. Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers in Marine Science*, *3*, p.96.

Andújar, C., Arribas, P., Gray, C., Bruce, C., Woodward, G., Yu, D.W. and Vogler, A.P., 2018. Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular Ecology*, *27*(1), pp.146-166.

Andújar, C., Arribas, P., Ruzicka, F., Crampton-Platt, A., Timmermans, M.J. and Vogler, A.P., 2015. Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology*, *24*(14), pp.3603-3617.

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M. and Vogler, A.P., 2016. Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology and Evolution*, *7*(9), pp.1071-1081.

Armitage, P. D., Moss, D., Wright, J. F., and Furse, M. T., 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research*, *17*, pp.333–347.

Arthington, Á. H., Naiman, R. J., Mcclain, M. E., and Nilsson, C., 2010. Preserving the biodiversity and ecological services of rivers: new challenges and research opportunities. *Freshwater Biology*, *55*, pp.1–16.

Aylagas, E., Borja, A. and Rodríguez-Ezpeleta, N., 2014. Environmental status assessment using DNA metabarcoding: towards a genetics-based marine biotic index (gAMBI). *PloS one*, 9(3), p.e90529.

Bady, P., Doledec, S., Fesl, C., Gayraud, S., Bacchi, M. and Scholl, F., 2005. Use of invertebrate traits for the biomonitoring of European large rivers: the effects of sampling effort on genus richness and functional diversity. *Freshwater Biology*, *50*, pp.159–173.

Baird, D. J. and Hajibabaei, M., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, *21*, pp. 2039–2044.

Ball S. L., Hebert P. D. N., Burian S. K., Webb J. M., 2005. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society*, *24*, pp.508–524.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling., 1999. *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish*, Second Edition. U.S. Environmental Protection Agency; Washington, D.C.

Bashar, K. Reza, M. S., Razzak, M. A., Rahman, K.M. Z., Goda, P. and Howlader A. J., 2014. Faunistic study of Odonata (Dragonfly & Damselfly) in some selected regions of Bangladesh. *Journal of Entomology and Zoology Studies*, *2* (4) pp. 1-6.

Begum, M., Ismail, M. And Ali, S., 1989. A note on Freshwater mussels of Bangladesh. *Bangladesh J. Zool., 17*(2), pp.203-204.

Beng, K.C., Tomlinson, K.W., Shen, X.H., Surget-Groba, Y., Hughes, A.C., Corlett, R.T. and Slik, J.F., 2016. The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Scientific Reports*, *6*, p.24965.

Bhuiyan, M. A. H., Dampare, S. L. B., Islam, M. A. and Suzuki, S., 2015. Source apportionment and pollution evaluation of heavy metals in water and sediments of Buriganga

River, Bangladesh, using multivariate analysis and pollution evaluation indices. *Environ Monit Assess*, *187*, 4075.

Bhuyan, M.S., Bakar, M.A., Rashed-Un-Nabi, M., Senapathi, V., Chung, S.Y. and Islam, M.S., 2019. Monitoring and assessment of heavy metal contamination in surface water and sediment of the Old Brahmaputra River, Bangladesh. *Applied Water Science*, *9*(5), pp.1-13.

Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R. and Thomas, W.K., 2012. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, *27*(4), pp. 233-243.

Bista, I., Carvalho, G.R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M. and Creer, S., 2017. Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, *8*(1), pp.1-11.

Blakely, Tanya J., Eikaas, Hans S. and Harding, J.S., 2014. The Singscore: a macroinvertebrate biotic index for assessing the health of Singapore's streams and canals. *Raffles Bulletin of Zoology,* 62: 540–548.

Bloechl, A., Koenemann, S., Philippi, B. and Melber, A., 2010. Abundance, diversity and succession of aquatic Coleoptera and Heteroptera in a cluster of artificial ponds in the North German Lowlands. *Limnologica-Ecology and Management of Inland Waters*, 40(3), pp.215-225.

Blackman, R.C., Mächler, E., Altermatt, F., Arnold, A., Beja, P., Boets, P., Egeter, B., Elbrecht, V., Filipe, A.F., Jones, J.I. and Macher, J., 2019. Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring–the need for calibration experiments. *Metabarcoding and Metagenomics*, *3*, pp.49-57.

Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. and Woodward, G., 2017. Next-generation global biomonitoring: large-scale, automated reconstruction of ecological networks. *Trends in Ecology & Evolution*, *32*(7), pp.477-487.

Bonada, N., Prat,N., Resh,V. H. and Statzner, B., 2006. Developments in Aquatic Insect Biomonitoring: A comparative analysis of recent approaches. *Annu. Rev. Entomol.*, 51, pp. 495–523.

Boonsoong, B. Sangpradub N. and Barbour, M. T., 2009. Development of rapid bioassessment approaches using benthic macroinvertebrates for Thai streams. *Environ. Monit. Assess*., *155*, pp.129–147.

Bourlat, S.J., Borja, A., Gilbert, J., Taylor, M.I., Davies, N., Weisberg, S.B., Griffith, J.F., Lettieri, T., Field, D., Benzie, J. and Glöckner, F.O., 2013. Genomics in marine monitoring: new opportunities for assessing marine health status. *Marine Pollution Bulletin*, *74*(1), pp.19-31.

Braukmann, T.W., Ivanova, N.V., Prosser, S.W., Elbrecht, V., Steinke, D., Ratnasingham, S., de Waard, J.R., Sones, J.E., Zakharov, E.V. and Hebert, P.D., 2019. Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources*, *19*(3), pp.711-727.

Bush, A., Compson, Z.G., Monk, W.A., Porter, T.M., Steeves, R., Emilson, E., Gagne, N., Hajibabaei, M., Roy, M. and Baird, D.J., 2019. Studying ecosystems with DNA metabarcoding: lessons from biomonitoring of aquatic macroinvertebrates. *Frontiers in Ecology and Evolution*, 7, p.434.

Buss, D.F., Carlisle, D.M., Chon, T.S., Culp, J., Harding, J.S., Keizer-Vlek, H.E., Robinson, W.A., Strachan, S., Thirion, C. and Hughes, R.M., 2015. Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. *Environmental Monitoring and Assessment, 187*(1), pp.1-21.

Cai, J., Varis, O. and Yin, H., 2017. China's water resources vulnerability: a Spatio-temporal analysis during 2003–2013. *Journal of cleaner production*, *142*, pp.2901-2910.

Carew, M.E., Pettigrove, V.J., Metzeling, L. and Hoffmann, A.A., 2013. Environmental monitoring using next-generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in Zoology*, *10*(1), pp.1-15.

Carew, M.E. and Hoffmann, A.A., 2015. Delineating closely related species with DNA barcodes for routine biological monitoring. *Freshwater Biology*, 60(8), pp.1545-1560.

Casas, P.A.S., Sing, K.W., Lee, P.S., Nuñeza, O.M., Villanueva, R.J.T. and Wilson, J.J., 2018. DNA barcodes for dragonflies and damselflies (Odonata) of Mindanao, Philippines. *Mitochondrial DNA Part A*, 29(2), pp.206-211.

Chakraborty, C., Huq, M. M., Ahmed, S., Tabassum, T. and Miah, M. R., 2013. Analysis of the causes and impacts of water pollution of Buriganga River: A Critical Study. *International Journal of Scientific & Technology Research*, *2* (9), pp.245-252.

Chapin III, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E. and Mack, M.C., 2000. Consequences of changing biodiversity. *Nature*, *405*(6783), p.234.

Chowdhuri, S. H. And Aktaruzzaman, M.,1981. Dragonfly (Odonata: Anisoptera) Larvae from Chittagong. *Bangladesh J. Zool*., 9(2), pp.131-144.

Chowdhury, G. W., Gallardo, B. and Aldridge, D. C., 2015. Development and testing of a biotic index to assess the ecological quality of lakes in Bangladesh. *Hydrobiologia*, 765, pp.55–69.

Cho, W.S., Park, Y.S., Park, H.K., Kong, H.-Y. and Chon, T.S., 2011. Ecological informatics approach to screening of integrity metrics based on benthic macroinvertebrates in streams. *International Journal of Limnology*, *47*, S51–S62.

Clare, E.L., Chain, F.J., Littlefair, J.E. and Cristescu, M.E., 2016. The effects of parameter choice on defining molecular operational taxonomic units and resulting ecological analyses of metabarcoding data. *Genome*, *59*(11), pp.981-990.

Clarke R. T., Wright J. F. and Furse M. T., 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling*, 160, pp. 219-233.

Clews, E., Low, E., Belle, C. C., Todd P. A., Eikaas, H. S. and Ng, P.K.L., 2014. A pilot macroinvertebrate index of the water quality of Singapore's reservoirs. *Ecological Indicators, 38*, pp.90–103.

Cordero, R.D., Sánchez-Ramírez, S. and Currie, D.C., 2017. DNA barcoding of aquatic insects reveals unforeseen diversity and recurrent population divergence patterns through broad-scale sampling in northern Canada. *Polar Biology*, *40*(8), pp.1687-1695.

Crampton-Platt, A., Yu, D.W., Zhou, X. and Vogler, A.P., 2016. Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience*, 5(1), pp.s13742-016.

Creer, S., Fonseca, V.G., Porazinska, D.L., Giblin-Davis, R.M., Sung, W., Power, D.M., Packer, M., Carvalho, G.R., Blaxter, M.L., Lambshead, P.J.D. and Thomas, W.K., 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, *19*, pp.4-20.

Cristescu, M.E., 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution*, *29*(10), pp.566-571.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. and Taberlet, P., 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, 10(9), pp.20140562.

Department of Environment (DoE), 2016. Surface and groundwater quality report. Ministry of Environment and Forest, Government of the People's Republic of Bangladesh. Pp.95

Dong, B., Geng, C., Cai, Y. and Ji, L., 2014. Aquatic Coleoptera response to environmental factors of freshwater ecosystems in Changbai Mountain, northeast China. *Aquatic Ecosystem Health & Management*, *17*(2), pp.171-178.

Dowle, E.J., Pochon, X., C. Banks, J., Shearer, K. and Wood, S.A., 2016. Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, *16*(5), pp.1240-1254.

EC (European Commission), 2003a. Common Implementation Strategy for the Water Framework Directive (2000/60/EC): Guidance Document No. 11: Planning Processes. Office for Official Publications of the European Communities, Luxembourg http:// ec.europa.eu/environment/water/water-framework/facts_figures/guidance_docs_ en.htm

Ekrem, T., Willassen, E. and Stur, E., 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution*, 43(2), pp.530-542.

Elbrecht, V., Vamos, E.E., Meissner, K., Aroviita, J. and Leese, F., 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, 8(10):1265-1275.

Epler, J.H., 2006. Identification manual for the aquatic and semi-aquatic Heteroptera of Florida. Division of Water Resource Management. Florida Department of Environmental Protection, Tallahassee, USA, pp.195.

Evans, D.M., Kitson, J.J., Lunt, D.H., Straw, N.A. and Pocock, M.J., 2016. Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems. *Functional Ecology*, *30*(12), pp.1904-1916.

Fatema, K., M. Begum, M. Al Zahid and M. E. Hossain., 2018. Water quality assessment of the river Buriganga, Bangladesh. *J. Biodivers. Conserv. Bioresour. Manag. 4*(1):47-53.

Flotemersch, J.E., Stribling, J.B. and Paul, M.J., 2006. Concepts and approaches for the bioassessment of nonwadeable streams and rivers. US Environmental Protection Agency. Report 600/R-06/127, Washington, DC.

Geist, J., 2011. Integrative freshwater ecology and biodiversity conservation. *Ecological Indicators*, *11*(6), pp.1507-1516.

Ghetti, PF, 1997. Extended biotic index (IBE): macroinvertebrates in the quality control of running water environments: application manual. Autonomous Province of Trento, Provincial Agency for Environmental Protection.

Gibson J. F., Shokralla S., Curry C., Baird D. J., Monk W. A., King I. and Hajibabaei M. 2015. Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. *PLoS One*, *10* (10), pp.e0138432.

Gill, B. A., Harrington, R. A., Kondratieff, B. C., Zamudiok, R., Poff, N. L. and Funk, W. C. 2013.Morphological Taxonomy, DNA Barcoding, and Species Diversity in Southern Rocky Mountain Headwater Streams. *Freshwater Science*, 33(1), pp.288-301.

Gomez-Rodrıguez, C., Crampton-Platt, A., Timmermans, M.J. T. N., Baselga, A. and Vogler, A.P. 2015. Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex Assemblages. *Methods in Ecology and Evolution*, 6, pp.883–894.

Gray D. P. and Harding, J.S., 2012. Acid Mine Drainage Index (AMDI): a benthic invertebrate biotic index for assessing coal mining impacts in New Zealand streams, *New Zealand Journal of Marine and Freshwater Research*, 46(3), pp.335-352.

Gregory J. P., Jeffrey E. B., & Benjamin M. L., and Michael J. W., 2013. Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams. *Environ Monit Assess,* 185, pp.1515–1540.

Global Water Partnership (GWP), 2000. Integrated Water Resources Management. Global Water Partnership (GWP) Technical Advisory Committee, Background Paper No.4.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C. and Baird, D.J., 2011. Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PLoS One*, 6(4), pp. e17497.

Hajibabaei, M., Spall, J.L., Shokralla, S. and van Konynenburg, S., 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, *12*(1), pp.1-10.

Hart, L.A., Bowker, M.B., Tarboton, W. and Downs, C.T., 2014. Species composition, distribution and habitat types of Odonata in the iSimangaliso Wetland Park, KwaZulu-Natal, South Africa and the associated conservation implications. *PLoS One*, *9*(3), pp.e92588.

Hasan, M.K., Shahriar, A. and Jim, K.U., 2019. Water pollution in Bangladesh and its impact on public health. *Heliyon, 5*(8), p.e02145.

Hawkes, H.A. (1997). Origin and development of the biological monitoring working party score system. *Water Ressearch*, *32* (3).

Heino, J., 2014. Taxonomic surrogacy, numerical resolution and responses of stream macroinvertebrate communities to ecological variables: are the inferences transferable among regions? *Ecological Indicatotrs*, 36, pp.186–194.

Heino, J., Mykrä, H. and Kotanen, J., 2008. Weak relationships between landscape characteristics and multiple facets of stream macroinvertebrate biodiversity in a boreal drainage basin. *Landscape Ecology*, 23(4), pp.417-426.

Hebert, P.D.N., Cywinska, A., BALL, S.L. and Dewaard, J.R., 2003. Biological identification through DNA barcodes. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 270, pp.313–321.

Hering, D., Feld, C. K., Moog, O. and Ofenböck, T., 2006. Cookbook for the development of a multimetric index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. *Hydrobiologia*, *566*, pp.311–342.

Hilsenhoff, W.L., 1987. An improved biotic index of organic stream pollution. *Great Lakes Entomologist*, 20, pp.31-39.

Hodkinson, I.D. and Jackson, J.K., 2005. Terrestrial and aquatic invertebrates as bioindicators for environmental monitoring, with reference to mountain ecosystems. *Environmental Management*, 35(5), pp.649-666.

Hutchinson, T.H., Solbe, J. and Kloepper-Sams, P.J., 1998. Analysis of the ecetoc aquatic toxicity (EAT) database III—comparative toxicity of chemical substances to different life stages of aquatic organisms. *Chemosphere*, 36(1), pp.129-142.

IEDS (Institute for Environment and Development Studies) report 2003. Aquatic ecology and dangerous substances: Bangladesh perspective**.** Diffuse Pollution Conference Dublin.

Irish, S.R., Al-Amin, H.M., Alam, M.S. and Harbach, R.E., 2016. A review of the mosquito species (Diptera: Culicidae) of Bangladesh. *Parasites & Vectors*, 9(1), pp.1-9.

Islam, M. M. Karim, M. R., Zheng X., and Xiaofang L., 2018. Heavy Metal and Metalloid Pollution of Soil,Water and Foods in Bangladesh: A Critical Review. *Int. J. Environ. Res. Public Health* ,15, p.2825.

Jannickemoe, S., Schartau, A., Baekken, T., Mcfarland, A. ,2010. Assessing macroinvertebrate metrics for classifying acidified rivers across northern Europe. *Freshwater Biology.* 55, pp.1382–1404.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A. and Larsen, T.H., 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), pp.1245-1257.

Joly, S., Davies, T.J., Archambault, A., Bruneau, A., Derry, A., Kembel, S.W., Peres-Neto, P., Vamosi, J. and Wheeler, T.A., 2014. Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Molecular Ecology Resources*, *14*(2), pp.221-232.

Jun, Y.C., Won, D.H., Lee, S.H., Kong, D.S. and Hwang, S.J., 2012. A multimetric benthic macroinvertebrate index for the assessment of stream biotic integrity in Korea. *International Journal of Environmental Research and Public Health*, *9*, pp.3599–3628.

Kagalou, I., Economidis, G., Leonardos, I. and Papaloukas, C., 2006. Assessment of a Mediterranean shallow lentic ecosystem (Lake Pamvotis, Greece) using benthic community diversity: Response to environmental parameters. *Limnologica, 36*(4), pp.269-278.

Kalkman V., J, Clausnitzer V, Dijskstra K.D.B., Paulson, D.R, van Tol J., 2008. Global diversity of dragonflies (Odonata) in freshwater. *Hydrobiology*, 6, pp.14–18.

Kamal, D., Khan, A.N., Rahman, M.A. and Ahamed, F., 2007. Study on the physico chemical properties of water of Mouri River, Khulna, Bangladesh. *Pakistan Journal of Biological Sciences*,10(5), pp.710-717.

Kamal, M.M., Malmgren-Hansen, A. and Badruzzaman, A.B.M., 1999. Assessment of pollution of the River Buriganga, Bangladesh, using a water quality model. *Water Science and Technology*, 40(2), pp.129-136.

Karr, J.R., 1999. Defining and measuring river health. *Freshwater Biology*, *41*(2), pp.221-234.

Kenney, M.A., Sutton-Grier, A.E., Smith, R.F. and Gresens, S.E., 2009. Benthic macroinvertebrates as indicators of water quality: The intersection of science and policy. *Terrestrial Arthropod Reviews*, *2*(2), p.99.

Khan, M. A. G., Rahman, M. and Islam S., 1997. Chironomid larvae of central-east and south-east Bangladesh – Keys and diagnoses with notes on ecology and distribution. *Bangladesh J. Zool*. 25(1), pp.29-46.

Klemm, D. J., Blocksom, K. A., Fulk, F. A., Herlihy, A. T., Hughes, R. M., Kaufmann, P. R., Peck, D. V., Stoddard, J. L., Thoeny, W. T., Griffith, M. B., and Davis, W. S. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. *Environmental Management*, *31*, pp.656–669.

Kuntke, F., de Jonge, N., Hesselsøe, M. and Nielsen, J.L., 2020. Stream water quality assessment by metabarcoding of invertebrates. *Ecological Indicators*, *111*, p.105982.

Lakew, A. and Moog, O., 2015. A multimetric index based on benthic macroinvertebrates for assessing the ecological status of streams and rivers in central and southeast highlands of Ethiopia. *Hydrobiologia,* 751, pp.229–242.

Lamouroux, N., Doledec, S. and Gayraud, S. 2004. Biological traits of stream macroinvertebrate communities: Effects of microhabitat, reach, and basin filters. *Journal of the North American Benthological Society*, 23, pp. 449 - 466.

Lau, J. K. and Lauer, T. E., 2015. Nationwide Benthic Macroinvertebrate Assemblage Multimetric Indices: Identifying Inconsistencies and Limitations in Reporting Stream Impairment Status, USA. *Environmental Management,* 56, pp.11–23.

Lenat, D.R. and Resh, V.H. ,2001. Taxonomy and stream ecology - The benefits of genus- and species-level identifications. *Journal of the North American Benthological Society*, 20, pp. 287–298.

Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., Mergen, P., Pawlowski, J., Piggott, J., Rimet, F., Steinke, D., Taberlet, P., Weigand, A., Abarenkov, K., Beja, P., Bervoets, L., Björnsdóttir, S., Boets, P., Boggero, A., … Zimmermann, J. (2016). DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Research Ideas and Outcomes*, *2*, e11321. https://doi.org/10.3897/rio.2.e11321

Li, L., Zheng, B. and Liu, L., 2010. Biomonitoring and bioindicators used for river ecosystems: definitions, approaches and trends. *Procedia Environmental Sciences*, 2, pp.1510-1524.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, A., Hooper, D.U., Huston, M.A., Raffaelli, D., Schmid, B. and Tilman, D., 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, 294 (5543), pp.804-808.

Majumder, A., 2009. Bangladesh River Pollution Threatens Millions. Retrieved from Reuters: https://www.reuters.com/article/us-bangladesh-rivers/bangladesh -river-pollution-threatens-millions-idUSTRE54I04G20090519.

Mardis, E.R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24, pp.133–141.

Menezes, S., Baird, D.J., Soares, A.M., 2010. Beyond taxonomy: a review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. *J. Appl. Ecol.* 47 (4), pp.711–719.

Miserendino, L.M. and Pizzolón, L.A., 2000. Macroinvertebrates of a fluvial system in Patagonia: altitudinal zonation and functional structure. *Archiv für Hydrobiologie*, pp.55-83.

Mokaddes, M. A. A., Nahar, B. S. and Baten, M.A. 2013. Status of Heavy Metal Contaminations of River Water of Dhaka Metropolitan City. *J. Environ. Sci. & Natural Resources*, *5*(2), pp.349 – 353.

Morinière, J., Hendrich, L., Balke, M., Beermann, A.J., König, T., Hess, M., Koch, S., Müller, R., Leese, F., Hebert, P.D. and Hausmann, A., 2017. A DNA barcode library for Germany′ s mayflies, stoneflies and caddisflies (Ephemeroptera, Plecoptera and Trichoptera). *Molecular Ecology Resources*, *17*(6), pp.1293-1307.

Mustafa, T., Alam, S. M. I., Jamal, M. S., Johora, F.T., Likhon, M. A. K., Hoq, T. and Naser, M. N., 2013. Abundance of benthic fauna in winter and summer seasons at three water bodies of Dhaka, Bangladesh. *Bangladesh J. Zool.*, *41*(1), pp.79-86.

Nasiruddin, M., Azadi, M. A. and Reza, M. S., 2014. Abundance and diversity of aquatic insects in two water bodies of Chittagong university campus. *Bangladesh J. Zool*., *42*(1), pp.19-33.

Neville, P.J. and Yen, A.L., 2007. Standardising terrestrial invertebrate biomonitoring techniques across natural and agricultural systems. *Australian Journal of Experimental Agriculture*, *47*(4), pp.384-391.

Nguyena, H. H., Everaertb, G., Gabrielsc, W., Hoangd T. H. And Goethals, P. L.M., 2014. A multimetric macroinvertebrate index for assessing the waterquality of the Cau river basin in Vietnam. *Limnologica,* 45, pp.16– 23.

Nichols. S. J and Dyer, F. J. 2013. Contribution of national bioassessment approaches for assessing ecological water security: an AUSRIVAS case study. *Front. Environ. Sci. Eng.*, 7(5), pp.669–687.

Nyenje, P.M., Foppen, J.W., Uhlenbrook, S., Kulabako, R. and Muwanga, A., 2010. Eutrophication and nutrient release in urban areas of sub-Saharan Africa—a review. *Science of the total environment*, *408*(3), pp.447-455.

Ofenbock, T., Moog, O. Sharma, S. and Korte T., 2010. Development of the HKHbios: a new biotic score to assess the river quality in the Hindu Kush-Himalaya. *Hydrobiologia*, 651, pp. 39–58.
Oliverio, A.M., Gan, H., Wickings, K. and Fierer, N., 2018. A DNA metabarcoding approach to characterize soil arthropod communities. *Soil Biology and Biochemistry*, *125*, pp.37-43.

Pfrender, M., Hawkins, C., Bagley, M., Courtney, G., Creutzburg, B., Epler, J., Fend, S., Ferrington Jr, L., Hartzell, P., Jackson, S. and Larsen, D., 2010. Assessing macroinvertebrate

biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. *The Quarterly Review of Biology*, 85(3), pp.319-340.

Phen, C., Loc, D. T. B., Kunpradid, T., Larsen, H., Duyen, N. V., Sivongxay, N. and Vidthayanon, C., 2014. Report on the 2011 Biomonitoring Survey of the Lower Mekong River and selected tributaries, MRC Technical Paper No. 43. Mekong River Commission, Vientiane, pp.78.

Pilgrim, E.M., Jackson, S., Swenson, S., Turcsanyi, I., Friedman, E., Weigt, L. and Bagley, M.J., 2011. Incorporation of DNA barcoding into a large-scale biomonitoring program: opportunities and pitfalls. *Journal of the North American Benthological Society*, 30, pp.217–231.

Porter, T.M. and Hajibabaei, M., 2018. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology*, *27*(2), pp.313-338.

Price, B. W., Allan E. L., Marathe, K., Sarkar, V., Simon, C. and Kunte, K., 2016. The cicadas (*Hemiptera*: *Cicadidae*) of India, Bangladesh, Bhutan, Myanmar, Nepal and Sri Lanka: an annotated provisional catalogue, regional checklist and bibliography. *Biodivers Data J.*, (4), p.e8051

Quisil, J.C, Nuneza, O.M., Villanueva, R.J.T., 2014. Impact of mine tailings on the species diversity of Odonata fauna in Surigao del Sur, Philippines. *J Biodiver Environ Sci.*, 1, pp.465–476.

Ratnasingham S, and Hebert P.D.N., 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol Ecol Notes*,7, 355–364.

Rosenberg, D.M. and Resh, V.H. (Eds.), 1993. Freshwater Biomonitoring and Benthic Macroinvertebrates. Chapman and Hall, New York.

Sana, B. B. and Ali, M. S. 2011. A preliminary list of aquatic Coleoptera (Arthropoda: Insecta) in Natore and Rajshahi districts of Bangladesh. *J. Taxon. Biodiv. Res*., 5: 1-4.

Sarkar, M., Rahman, A. K. M. L., Islam, J. B., Ahmed, K. S., Uddin, M. N. and Bhoumik, N. C., 2015. Study of hydrochemistry and pollution status of the Buriganga river, Bangladesh. *Bangladesh J. Sci. Ind. Res. 50*(2), pp. 123-134.

Serrana, J.M., Miyake, Y., Gamboa, M. and Watanabe, K., 2019. Comparison of DNA metabarcoding and morphological identification for stream macroinvertebrate biodiversity assessment and monitoring. *Ecological Indicators*, *101*, pp.963-972.

Shen Y., Guan, L., Wang D. and Gan X., 2016. DNA barcoding and evaluation of genetic diversity in Cyprinidae fish in the midstream of the Yangtze River. *Ecol Evol.,* 6(9): 2702–2713.

Shokralla, S., Porter, T.M., Gibson, J.F., Dobosz, R., Janzen, D.H., Hallwachs, W., Golding, G.B. and Hajibabaei, M., 2015. Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports*, 5, p.9687.

Simpson, J. and Norris, R. H. 2000. Biological assessment of water quality: development of AUSRIVAS models and outputs. In J. F. Wright, D. W. Sutcliffe, and M. T. Furse (Eds.), RIVPACS and similar techniques for assessing the biological quality of freshwaters (pp. 125–142). Ableside: Freshwater Biological Association and Environment Agency.

Stark, J. D. and Maxted, J. R., 2007. A biotic index for New Zealand's soft-bottomed streams, *New Zealand Journal of Marine and Freshwater Research*, 41(1), pp.43-61.

Stein, E.D., White, B.P., Mazor, R.D., Jackson, J.K., Battle, J.M., Miller, P.E., Pilgrim, E.M. and Sweeney, B.W., 2014. Does DNA barcoding improve performance of traditional stream bioassessment metrics?. *Freshwater Science*, 33(1), pp.302-311.

Stribling, J.B., Pavlik, K.L., Holdsworth, S.M. and Leppo, E.W., 2008. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *Journal of the North American Benthological Society*, 27(4), pp.906-919.

Subramanian, K.A. and Sivaramakrishnan, K.G., 2007. Aquatic Insects for Biomonitoring Freshwater Ecosystems-A Methodology Manual.Ashoka Trust for Ecology and Environment (ATREE), Bangalore, India. Pp.31.

Sweeney, B.W., Battle, J.M., Jackson, J.K. and Dapkey, T., 2011. Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality?. *Journal of the North American Benthological Society*, *30*(1), pp.195-216.

Taberlet, P.,C. Eric,Pompanon F.,Brochmann C. and Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, (21), pp.2045–2050.

Thomas, T., Gilbert, J. and Meyer, F., 2012. Metagenomics-a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1), pp.1-12.

Trebitz, A.S., Hoffman, J.C., Grant, G.W., Billehus, T.M. and Pilgrim, E.M., 2015. Potential for DNA-based identification of Great Lakes fauna: match and mismatch between taxa inventories and DNA barcode libraries. *Scientific Reports*, *5*, p.12162.

Venter, H.J. and Bezuidenhout, C.C., 2016. DNA-based identification of aquatic invertebrates-useful in the South African context?. *South African Journal of Science*, *112* (5-6), pp.1-4.

Verneaux, V. and Verneaux, J., 2002. Assessing lake functioning using the macrobenthic community with special reference to Chironomidae (Diptera). A subalpine lake (Lake Annecy) as an example. *Archiv für Hydrobiologie*, pp.61-78.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R. and Davies, P.M., 2010. Global threats to human water security and river biodiversity. *Nature*, *467*(7315), p.555.

Wallace, J.B. and Webster, J.R., 1996. The role of macroinvertebrates in stream ecosystem function. *Annual Review of Entomology*, 41: 115-139.

Wang, Y., Teng, E., Liu T., Lv, Y., Ji, X., Giesy, J. P. And Hollert, H., 2014. A national pilot scheme for monitoring and assessment ofecological integrity of surface waters in China. *Environmental Development*.

Webb J.M., Jacobus L.M., Funk D.H., Zhou X. and Kondratieff, B., et al., 2012. A DNA Barcode Library for North American Ephemeroptera: Progress and Prospects. *PLoS ONE* 7(5), p.e38063.

Weigand, H., Beermann, A.J., Čiampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M.F., Grabowski, M., Rimet, F., Rulik, B. and Strand, M., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of the Total Environment*, *678*, pp.499-524.

Wright, J.F., Sutcliffe, D.W. and Furse, M.T., 2000. Assessing the biological quality of fresh waters: RIVPACS and other techniques. In Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association. Wright, J.F., 2000.An introduction to RIVPACS. Ref. 149a, pp. 1-24.

Wu, W., Xu, Z., Zhan, C., Yin, X. and Yu, S., 2015. A new framework to evaluate ecosystem health: a case study in the Wei River basin, China. Environmental Monitoring and Assessment, *187*(7), pp.1-15.

Xiong, W. and Zhan, A., 2018. Testing clustering strategies for metabarcoding-based investigation of community–environment interactions. *Molecular Ecology Resources*, *18*(6), pp.1326-1338.

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C. and Ding, Z., 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*(4), pp.613-623.

Zhou, X., Robinson, J. L., Geraci, C. J., Parker, C. R., Flint, O. S. Jr, Etnier, D. A., and Hebert, P. D. N.,2011. Accelerated construction of a regional DNA-barcode reference library: Caddisflies (Trichoptera) in the Great Smoky Mountains National Park. *Journal of the North American Benthological Society*, 30(1), pp.131–162.

Chapter 2

DNA barcode database for Bangladeshi freshwater macroinvertebrates and their mitogenome based phylogeny

Chapter 2: DNA barcode database for Bangladeshi freshwater macroinvertebrates and their mitogenome based phylogeny

2.1 Abstract

Macroinvertebrates are commonly used as indicator organisms for water quality and ecosystem assessments. However, reliable morphological identification of macroinvertebrate species is a challenging task for any biomonitoring programme. DNA barcoding has been a promising tool for the identification of aquatic macroinvertebrates. A reliable DNA reference dataset is a prerequisite for identification using barcodes. This study represents a first-time step towards building a DNA barcode library for Bangladeshi freshwater macroinvertebrates that generated 812 barcodes of the mitochondrial cytochrome oxidase subunit I gene representing 320 species from lowland and highland rivers. Among them, three species delimitation methods combinedly produced 34 Ephemeroptera, 7 Plecoptera, 26 Trichoptera, 47 Coleoptera, 29 Hemiptera, 37 Odonata, 65 Diptera, and 31 Mollusca species. Additionally, 20 Decapoda species and 12 Annelida species were also delimited from 59 and 24 barcodes respectively. Genetic distances followed the general rule of species boundary with barcodes where the interspecific K2P distance estimated for most of the families was 15 to 30-fold higher than that of intraspecific distance. Interspecific and intraspecific genetic distances measurement of target groups could be used for prioritising the conservation of bioindicator macroinvertebrates. Barcodes of known taxa will be a valuable reference dataset for metabarcoding of macroinvertebrates. I also aimed to explore complete mitogenomes from selected morphospecies of macroinvertebrates using a mitochondrial metagenomics pipeline. In total, 108 complete mitogenomes (>15000 bp) and 89 partial contigs (3000 to <15000 bp) were produced from 287 different morphospecies. The phylogenetic tree constructed with protein-coding genes of mitogenomes placed evolutionary studies of Bangladeshi macroinvertebrates in the context of existing data for taxa from elsewhere in the world. The sound placement of barcode sequences in the mitogenome-based phylogenetic framework was also effective for their deeper identification.

2.2 Introduction

2.2.1 Biodiversity patterns of bioindicator macroinvertebrates in freshwater ecosystems

In freshwater ecosystems, macroinvertebrates are a highly diverse group and well known for their wide-ranging sensitivity to environmental alterations that makes them potential bioindicators in ecological assessment of rivers and streams (Kenney et al., 2009; Lenat, 1993; Sweeney et al., 2011; Resh et al., 1995). They are widely distributed in different regions (bottom sediments, bank vegetation or floating and suspended substratum) of lakes, ponds, streams, and rivers across geographic regions from lowland water bodies to high altitude mountain streams. Macroinvertebrates, especially aquatic insects (e.g., mayflies, stoneflies, caddisflies, water beetles, water bugs, dipteran, dragonflies and damselflies), annelids and molluscs respond to organic pollution (Armitage et al., 1983; Zamora-Muñoz and Alba-Tercedor, 1996), heavy metals (Poulton et al., 1995; Smolders et al., 2003) and habitat degradation and biological invasion (Barbour et al., 1999; Karr and Chu, 1999; Li, Zheng, and Liu, 2010). Hence, freshwater macroinvertebrates have a long history of use in studies of natural and human-induced changes in water bodies and the effect of these changes on community structure and function of ecosystems (Daily and Ellison, 2012).

The most sensitive bioindicator insect orders in freshwaters are Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddisflies) (Baird and Sweeney, 2011; Cordero et al., 2017; Hering et al., 2004; Moriniere et al., 2017; Zhou et al., 2011). So far, more than 3,000 described species are representing the order Ephemeroptera belonging to 400 genera and 42 families (Barber-James et al., 2007). Studies of their taxonomy and discovery of new species are still quite incomplete and numerous unknown species and genera await description, especially in Southeast Asia. Plecoptera, or stoneflies, is a small order of insects containing more than 3,497 described species so far. The fauna and diversity of stoneflies in North America, Europe, Australia and New Zealand are relatively well-known but much less so in South Asia (Fochetti and De Figueroa, 2007). The order Trichoptera is represented by around 12,627 species in 610 genera and 46 families, but the rate of new species descriptions from the Neotropics, Madagascar, Africa, south-east Asia, China and the Philippines hints that there are more than 50,000 species in total, which suggests that only around 20–25% of the world species of Trichoptera have been described (De Moor and Ivanov, 2007).

Odonata (dragonfly and damselfly) larvae are found in almost every freshwater environment. This order is relatively well studied and includes 5,680 known species of which 2,739 belong to the suborder Zygoptera (19 families) and 2,941 to the suborder Anisoptera (12 families). It is estimated that between 1,000 and 1,500 species of dragonflies await descriptions that predict the actual number of odonate species may be close to 7,000. The highest diversity is found in flowing waters in tropical rain forests and the Oriental, Australasian and especially the Neotropical regions which hold the highest number of undescribed species (Kalkman et al., 2007).

The aquatic and semi-aquatic Hemiptera (water bugs) are a common component of the aquatic insects consisting of 4,810 species, of which 4,656 species in 326 genera of 20 families inhabit freshwaters. It is also estimated that more than 1,100 species remain to be described. Overall water bugs are most numerous in the tropical regions and species richness is highest in the Neotropical and Oriental regions harbouring 1,289 and 1,103 species, respectively (Polhemus and Polhemus, 2007). Aquatic beetles represent one of the largest groups of aquatic animals with an estimated 18,000 species of which 70% are already described and 30% still await description. Although about 30 beetle families have aquatic representatives, six families (Dytiscidae, Hydraenidae, Hydrophilidae, Elmidae, Scirtidae and Gyrinidae) are dominating in freshwater systems including around 15000 estimated species (Jäch and Balke, 2008).

Dipterans are commonly the most diverse and abundant macroinvertebrates which rely on various freshwater bodies for the completion of early life cycles (larval and pupal stages). Taxonomists assume that most of the extant species of Diptera are still undescribed (Bickel et al., 2009), with a global estimate of 400,000 to 800,000 species compared with ~160,000 described species (Pape, Blagoderov and Mostovski, 2011). Among 19 families of aquatic Diptera, Chironomidae often dominates in aquatic communities and largely occur in the Palaearctic and Nearctic Regions although they are distributed in all continents. So far, a total of 4,147 species in 339 genera and 11 subfamilies are unambiguously aquatic in their immature stages (Ferrington, 2007).

In contrast to this global estimation, in Bangladesh, there is no comprehensive biodiversity information (morphology or DNA-based) for the macroinvertebrate fauna though the country is known to support a vast variety of flora and fauna due to its unique geophysical location.

The available studies are very limited to sporadic records and incomplete diversity information of aquatic fauna conducted in selective parts of the country. A couple of studies revealed the biodiversity and abundance of benthic fauna in lakes (Ahad et al., 2012; Sharmin et al., 2018; Mustafa et al., 2013), ponds (Nasiruddin, Azadi and Reza, 2014) and polluted rivers (Hossain et al., 2015). Ephemeroptera, Trichoptera and Plecoptera are the most understudied macroinvertebrate groups in the country except for some family or genus-level records from pond and river ecosystems (Nasiruddin, Azadi and Reza, 2014; Hossain et al., 2015). Ali et al. (1978) recorded 58 species under the taxa Coleoptera, Hemiptera, Odonata, Chironomidae, Oligochaeta and Mollusca from three urban ponds in Dhaka city. A preliminary list of aquatic beetles from northern districts of Bangladesh estimated 27 species under 3 families and 14 genera (Sana and Ali, 2011). The odonates have been studied more than other insects though all are based on adult morphological characters. No studies are available on the aquatic forms of odonates except for taxonomic and distribution notes on some dragonflies fly larvae (Chowdhury and Akteruzzaman, 1981). A checklist of the Odonata from the eastern region of Bangladesh includes 49 species of Anisoptera in 32 genera, and 47 species of Zygoptera in 18 genera (Chowdhury and Mohiuddin, 2011). Another faunistic study of odonates conducted in the south-eastern, north-eastern, and central and south-west regions of the country reported 48 species (25 dragonflies and 23 damselflies) under 8 families: Libellulidae, Aeshnidae, Gomphidae, Coenagrionidae, Platycnemididae, Calopterygidae, Lestidae and Protoneuridae (Bashar et al., 2014). Aquatic and semi-aquatic hemipterans also remain understudied and only a single study recorded 17 species in 12 genera and 9 families (Hossain and Rahman, 2018). In this context, the scarcity of biodiversity data and the increasing degradation trend of freshwater ecosystems in Bangladesh demand large-scale macroinvertebrate diversity studies with a fast and effective method that could be the primary effort for their conservation and the basis for ecological assessment of freshwater ecosystems. In essence, DNA-based methods have proven to be vital tools for rapid, reliable and cost-effective identification in biodiversity estimation, ecosystem assessment and monitoring.

2.2.2 DNA barcoding: a potential tool for biodiversity estimation and promoting metagenetic approaches for large scale bioassessment.

The key challenge in traditional approaches for biodiversity estimation is the identification and characterization of macroinvertebrates to a lower taxonomic level (e.g., genus or species), where morphological variability, polymorphisms and immature life stages create major hurdles to obtaining accurate, precise, rapid, and cost-effective estimates in freshwater ecosystems (Ball et al., 2005; Carew et al., 2007; Pfenninger et al., 2007; Sinclair and Greens, 2008; Weigand et al., 2019). To address this challenge, DNA barcoding has become the primary tool as an alternative to morphology-based identification that identifies taxa based on a short DNA sequence from a standardized genetic marker, such as the mitochondrial gene cytochrome c oxidase I (COI) for most metazoans (Hebert et al., 2003, 2003b). This approach has proven useful particularly for species-level identification in insects, regardless of gender (Ekrem et al., 2007), developmental stage (Cordero et al., 2017; Zhou et al., 2010), size, or even damage to specimens (Ball et al., 2005; Carew and Hoffmann, 2015; Geraci, Al-Saffar and Zhou, 2011; Janzen et al., 2005).

DNA barcoding is being used in biodiversity estimation and environmental assessments to identify known species with genetic barcodes and to assign unknown specimens to putative species. Genetic distances or barcode gaps are generally used to designate species with an average distance of ≥2% among individuals in different putative species (Ball et al., 2005, Zhou et al., 2009) or on a level of 10× the intraspecific variation (Hebert et al., 2004). Hence, the power of DNA barcodes is subject to higher average interspecific genetic distances than the average intraspecific distances (Hebert et al., 2003, 2003b, 2004; Shen et al., 2016; Ward et al., 2005). Sequences for the same species generally fall into a monophyletic cluster on a phylogenetic tree with intraspecific distances (Srivathsan and Meier, 2012; Shen et al., 2016). However, barcodes with insufficient sequence divergence can be problematic for the separation of closely related taxa, particularly when levels of intra- and inter-specific variation overlap (e.g., Carew and Hoffmann, 2015; Kaila and Stahls, 2006; Van Velzen et al., 2012). For instance, the intra-specific variation in COI is high in some insects, gastropods and amphibians and usually overlaps with inter-specific variation (Davison et al., 2009; Meier et al., 2006). Due to high intraspecific variability in DNA barcodes, individuals from the same species might also be placed into multiple deeply divided monophyletic groups, but it can be difficult to determine a species-specific monophyletic node on a phylogenetic tree

(Elias et al., 2007). Therefore, no method for delimiting DNA barcodes into species-level entities is universally accepted. Alternatively, large scale sampling, using another gene or multiple genes and application of combined species delimitation methods e.g., barcode gap based ABGD (Puillandre et al., 2012) and phylogeny-based GMYC (Fujisawa and Barraclough, 2013; Monaghan et al., 2000; Pons et al., 2006), bPTP (Zhang et al., 2013) and mPTP (Kapli, et al., 2017) might minimize the constraints of DNA barcoding.

2.2.3 DNA barcodes and reference datasets for large scale bioassessment in freshwater ecosystem

Regardless of the approach to species delimitation, DNA barcodes of known species promote species identification in bulk or environmental samples processed with high throughput sequencing (HTS) (Carew et al., 2013; Ji et al., 2013; Yu et al., 2012). This strength has made this approach a potentially reliable tool in water quality assessment programs and enhances bioassessment capacity by reducing the time and cost necessary for taxonomic identification (Baird and Hajibabaei, 2012; Janzen et al., 2009; Stein et al., 2014; Sweeney et al., 2011; Webb et al., 2012; Carew and Hoffmann, 2015). Recently, implementation options and applicability of DNA-based identification into ecological monitoring has been assessed under WFD where the suitability of this procedure was rated as high though completing a barcode reference library is one of the key challenges identified for invertebrates (Hering et al., 2018). Like WFD, the Marine Strategy Framework Directive (MSFD) also requires reliable barcode reference libraries to implement molecular identification tools in aquatic biomonitoring (Weigand et al., 2019). An incomplete database or a reference set with a coarse level identification may mislead ecological assessments as species of the same genus or family responds to various stressors (Lenat and Resh, 2001; Pilgrim et al., 2011; Sweeney et al., 2011). In essence, a reliable reference library of taxonomically verified material is a prerequisite for identification using barcodes (Webb et al., 2012). This library will permit DNA-barcode sequences from macroinvertebrate samples to be identified accurately by comparing specimen barcodes against library barcodes (Baird and Hajibabaei, 2011; Stein et al., 2014) such as the Barcode of Life Data Systems (BOLD; www.boldsystems.org) (Ratnasingham and Hebert, 2007) or GenBank (www.ncbi.nlm.nih.gov/genbank). Therefore, DNA barcoding is probably the best option for establishing the reference sequence libraries required to rapidly identify specimens of known species (Gwiazdowski et al., 2015; Weigand et al., 2019) as well as to register unknown species (Morinière et al., 2019) that will expedite

the utility of metabarcoding and metagenomics for large-scale biomonitoring in freshwater ecosystems.

2.2.4 DNA barcoding initiatives and database for taxa used in biomonitoring: a global overview

DNA barcoding has been equally popular for the identification of animal and plant species across the world. The biggest initiative for DNA barcoding - iBOL (International Barcode of Life) completed a major project 'BARCODE 500K' under which 500,000 species has already been barcoded. iBOL's current project is 'BIOSCAN' which aims for barcoding of 2.5 million species by 2026 (https://ibol.org). Among major databases, DNA barcode reference libraries, such as the Barcode of Life Data System (BOLD (Ratnasingham and Hebert, 2007) and GenBank (Benson et al., 2013) are comprehensive platforms for preserving DNA barcodes essential for biodiversity monitoring. Besides these, several countries have developed barcode initiatives concentrating on a specific group of organisms such as SwissBOL in Switzerland, NorBOL in Norway, ABOL in Austria. A gap analysis on DNA barcodes available in BOLD and GenBank databases showed that barcodes for fish, true bugs, caddisflies and vascular plants are better represented than other groups (Weigand et al., 2019). Recent important works using the COI barcode approach for Ephemeroptera (Cardoni et al., 2015; Selvakumar et al., 2016; Suh et al., 2019; Webb et al., 2012), Trichoptera (Erasmus et al., 2018; Zhou et al., 2016), Plecoptera (Gattolliat et al., 2016), Diptera (Ekrem, et al., 2007; Kim et al., 2012; Morinière et al., 2019), Odonata (Casas et al., 2018; Karthika et al., 2012), Hemiptera (Gwiazdowski et al., 2015; Havemann et al., 2018; Raupach et al., 2014) and Coleoptera (Cordero et al., 2017; Hendrich et al., 2015; Jung et al., 2016; Raupach et al., 2016) have increased knowledge of macroinvertebrates biodiversity across the world. These barcode data are potentially operative for species identification at local and regional levels but in different biogeographic regions, they are sometimes found less effective for lower-level identification due to divergent community structures. This context necessitates the development of DNA-based barcode databases including all geographical regions or countries of the word.

2.2.5 DNA barcoding for biodiversity studies: Bangladesh perspective

Bangladesh is known to have one of the most productive and diverse freshwater fauna because of its unique geographical location at the Indo-Burma biodiversity hotspot. However, freshwater invertebrates, in particular, macroinvertebrates are remarkably understudied which necessitate fast and reliable measures to assess their diversity before the extinction of many species. Hence, DNA barcoding could be one of the potential options to characterize known and unknown species of the invertebrate fauna of the country. This method has already received global-level acceptance in many biodiversity assessments and ecological monitoring programmes. However, this is not the case in developing countries, especially in Bangladesh where applications of DNA based techniques for biodiversity studies remains in a rudimentary stage except for some institutional efforts for freshwater fishes (Rahman et al., 2019; Ahmed et al., 2019; Habib et al., 2021). Most of these studies concerned the building of DNA barcodes for known species with their phylogenetic implications and for describing new species of fish (Rahman et al., 2016; Kullander et al., 2015; Kullander et al., 2017). A comprehensive DNA barcode library for 243 species of freshwater fish (Rahman et al., 2019) and a partial barcode database of marine fishes are currently available in Bangladesh (Ahmed et al., 2021; Habib et al., 2021).

In contrast, DNA barcodes-based characterization of invertebrates is limited to few studies for beetles (Aslam et al., 2019), parasitic wasps (Mazumdar et al., 2019), butterflies (Ghosh et al., 2019), fruit flies (Leblanc et al., 2019), which indicates the paucity of DNA based information for freshwater macroinvertebrates. DNA sequences of the barcode marker for Bangladeshi aquatic beetles, mayflies, caddisflies, stoneflies, water bugs, dragonflies and damselflies are nearly completely missing from GenBank and BOLD. In light of these circumstances, the present study aimed to develop the DNA based characterization of freshwater macroinvertebrates in selected upland and lowland rivers of Bangladesh. Therefore, one of the aims of the present study was to build a DNA barcode library for Bangladeshi freshwater macroinvertebrates groups including all major aquatic insect orders.

2.2.6 Mitogenomics and phylogeny of macroinvertebrates

Mitochondrial genes are the most widely utilized molecular markers for systematic and phylogenetic studies. At present, the COI gene proposed by Hebert et al. (2003), is regarded as a paramount DNA barcode and is widely used for the identification and characterization of most animal groups. As described earlier DNA barcoding with only the COI gene has some limitations especially to deliver a robust phylogeny. Moreover, the universal primers (Folmer et al., 1994) are not always capable to amplify the fragment of COI barcode and other primers are needed (Chen et al., 2011; Hoareau and Boissin, 2010; Lohman et al., 2009; Zou et al., 2012; Yu et al., 2016). Generally, multiple-genes based phylogeny gives more reliable evolutionary information than single gene-based studies. However, it is not straightforward to extract multiple genes or complete mitogenomes of organisms. Owing to the rapid advances in DNA sequencing, a PCR free complete mitochondrial genome study (mitogenomics) has been possible to accumulate more reliable taxonomic, phylogenetic and biodiversity information avoiding an exclusive reliance on COI or any other single gene. HTS coupling with bioinformatics tools is promoting the generation of mitogenomes of individual organisms or from environmental samples sidestepping the PCR, which has become a productive approach for many taxonomic and ecological studies.

Among the available approaches, Mitochondrial Metagenomics (MMG) is a methodology for shotgun sequencing of total DNA from specimen mixtures and subsequent bioinformatic extraction of mitochondrial sequences. This method is a ‘metagenome skimming’ method (Linard et al., 2015), which extracts gene sequences through genome assembly of sequencing reads from shallow shotgun sequencing of mixed specimen samples. Shotgun metagenomic sequencing of bulk community samples generates numerous reads corresponding to mitochondrial DNA, from which contigs can be assembled into full or partial mitogenomes (Dettai et al., 2012; Zhou et al., 2013; Gillett et al., 2014; Tang et al., 2014; Crampton-Platt et al., 2015). MMG permits the exploration of mitochondrial genome sequences for entire species assemblages, facilitating the concurrent analysis of taxonomic and ecological questions (Andujar et al., 2015).

The approach can be applied to phylogenetic analysis of taxonomically selected taxa, as an economical alternative to mitogenome sequencing from individual species, or to

environmental samples of mixed specimens (Crampton-Platt et al., 2016). In case of soil beetles, shotgun sequencing of bulk samples and subsequent reconstruction of mitochondrial genomes provided a solid phylogenetic framework to estimate species diversity (Andujar et al., 2015). The MMG method not only improves the current standards of DNA-based biodiversity assessment but also permits the application of phylogenetic community ecology to hyper-diverse and poorly known biota (Papadopoulou et al., 2015). Gomez-Rodrıguez et al. (2015) validated the power of mitochondrial metagenomics for community ecology and phylogeny of complex assemblages by demonstrating that species occurrences estimated with MMG are similar to those from standard barcodes. Recent studies (Arribas et al., 2016) demonstrated that the combination of PCR-based and shotgun sequencing pipelines is a powerful, cost-efficient approach for characterising soil arthropods in a phylogenetic and community ecology context. MMG and metabarcoding make the burdensome task of taxonomic identification more straightforward even for cryptic species, encompassing the detection of changes in species richness and distributions. Mitogenomic data sets also facilitate estimates of species counts within samples and are also effective for tracking population trajectories (Tang et al., 2015).

Availability of mitogenomes is equally critical for studies of freshwater communities, where they are poorly known taxonomically, e.g., in the taxonomically neglected but presumed species-rich freshwater habitats of Bangladesh. However, despite the growing uses of mitogenomes for ecological, phylogenetic and biodiversity studies of terrestrial arthropods, a paucity of mitogenomic data is still obvious for bioindicator macroinvertebrate communities. While thousands of CO1 reference sequences of macroinvertebrates are available in public databases (BOLD, GenBank), data for complete mitogenomes or multiple genes are scanty. Therefore, a mitogenome based phylogeny is inevitably important for the country's macroinvertebrate fauna to understand evolutionary and biogeographic affinities with other macroinvertebrates encountered elsewhere.

Mitogenome based phylogenies of macroinvertebrates with known identifications can enhance the utility of DNA barcoding, metabarcoding and metagenomics for large-scale biomonitoring in freshwater ecosystems. It greatly increases the efficiency of assigning taxonomic information to OTUs generated from metabarcoding pipelines. A mitogenome data set primarily serves the purpose of grouping the local fauna into the existing phylogenetic framework for each of the major classes of macroinvertebrates, even where

their species identification is unclear, while more detailed identifications and possible species descriptions can follow later. Mitochondrial genomes are strong phylogenetic markers that establish sound placement of these lineages, unlike shorter sequences such as the COI barcode that generally fail to provide an accurate phylogenetic tree. In addition, by linking the local species into the wider phylogenetic framework they are placed into clades of known functional roles, i.e. the phylogeny produces a predictive system for their traits. To that end, the mitogenome phylogeny of macroinvertebrates has great potential to provide higher taxonomic, functional and evolutionary information. Given the power of the MMG approach, the deeper studies of local species assemblages can profit greatly from this phylogenetically informed reference library of macroinvertebrates. For example, by developing phylogenies for chironomids and mayflies, Carew et al. (2013) argued that there is a strong phylogenetic signal for pollution responses and the phylogenetic tree can provide insights into processes that produce sensitive and tolerant taxa. Phylogenetic community structure and composition at the local or global level holds promise for understanding the species relatedness, taxonomy, biogeography and ecosystem functioning with environmental conditions (Cavender-Bares et al., 2009). Thus, the increasing availability of mitogenome based phylogenies helps to reveal the multitude of processes driving community structure alongside the evolutionary relationship.

2.2.7 General aims and research questions

The main aim of the present work was to collect and sequence a range of Bangladeshi macroinvertebrate species in order to build a DNA barcode library and mitogenome based phylogenetic study. I constructed a DNA barcode reference database with members of the various insect orders to explore their diversity at OTU/species-level under major (order/family) taxa groups in Bangladeshi rivers. I evaluated the performance of DNA barcoding for using different species delimitation methods as well as the levels of divergence and phylogenetic relationships of the studied taxa. These barcodes can be used as bait sequences to identify the taxa used for mitogenome exploration with a PCR free shotgun sequencing approach. The key research questions were set as follows: a) What are the species richness and composition of macroinvertebrates under major taxa (order/family) groups extant in lowland and highland rivers? b) To what extent, are these species genetically distant and phylogenetically related under the major taxa/hierarchical level (e.g., family)? c) For mitogenomes extraction, to what extent, is the MMG pipeline effective? d) How do all

protein-coding genes contribute to resolving the phylogenetic placement of macroinvertebrates? and e) To what extent, does the mitogenome data help to place the barcodes of local fauna into the existing phylogenetic framework for each of the major orders of macroinvertebrates? Furthermore, our barcodes of known taxa will be valuable reference data for HTS characterised applications such as metabarcoding of bulk community samples (to be described in chapter 4) and shotgun sequencing of bulk samples which will advance the biomonitoring efforts in Bangladesh.

2.3 Methods

2.3.1 Study area

Study sites were selected strategically in four deep, non-wadable and interconnected lowland rivers (including highly polluted and least polluted) and sixteen upland rivers (experiencing a wide range of anthropogenic pressure) in Bangladesh (Fig. 2.1). These two sets represent the two main biotas of the country. The lowland rivers namely the Buriganga and the Turag partially surround the capital city Dhaka and the other two, the Dhaleshwari and Kaliganga River are about 10-20 km away to the west of the city (Table 1 in the appendix and Fig. 2.1). The upland rivers are in the hilly area located in the south-eastern parts of the country bordering India and Myanmar.

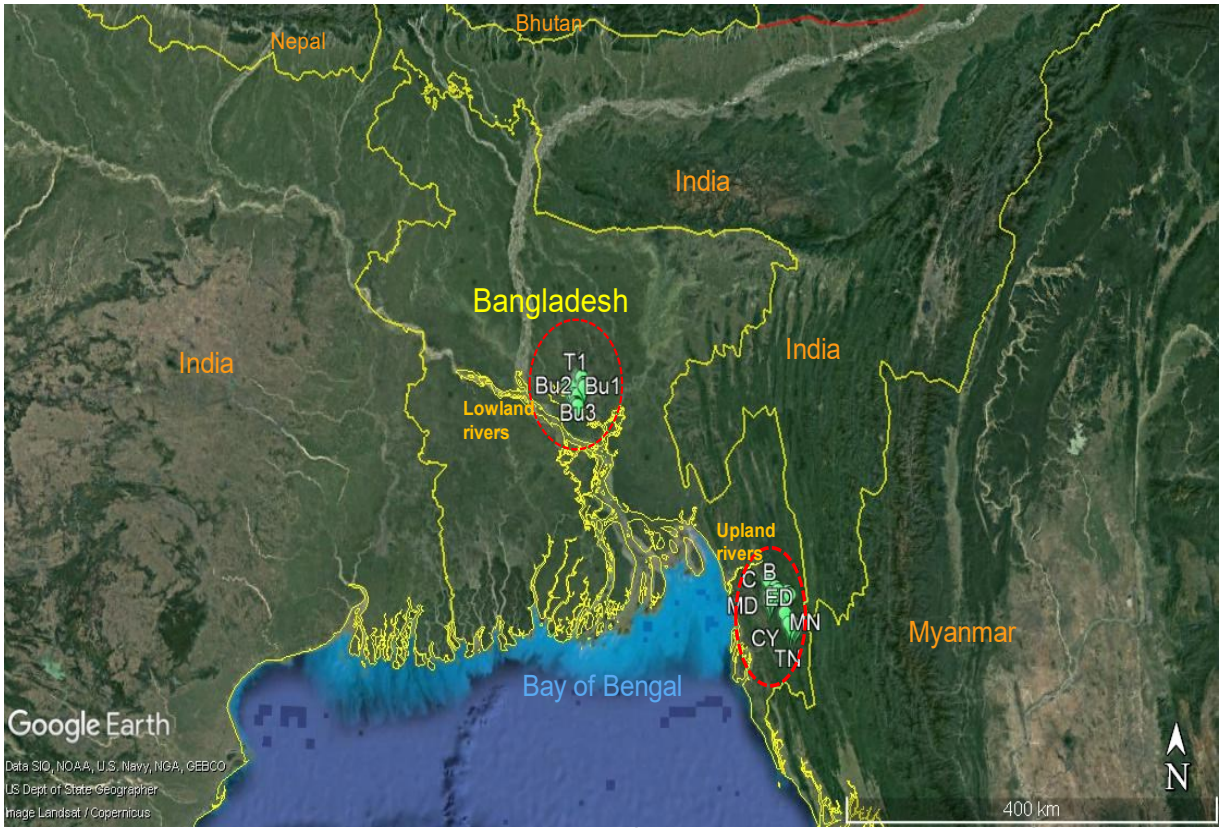

Figure 2. 1 Map of Bangladesh showing two major study areas red circled

2.3.1a Lowland Rivers

Buriganga River

The Buriganga River is a tidal river running from the western and southern parts of Dhaka City, the economic and political capital of Bangladesh. Originating from the Dhaleshwari River at Dharmaganj, it meets the Turag at Kholamora of Keraniganj flowing around Dhaka city. The average width and depth are 400 m and 10 m respectively. This is a commercially important and navigable river all year round and serves as a major transportation route and flood control and drainage outlet. It is also used for agricultural, sanitary, and industrial purposes (Alam et al., 2002). This river is known to have fish and other invertebrate diversity on which many local people depend for their livelihood. Rapid industrialization and unplanned urbanization along its banks, dumping of huge volumes of industrial effluents and sewerage toxic wastes have resulted in increased water pollution (Moniruzzaman et al., 2009; Islam et al., 2006; Ahmed et al., 2015).

The Turag River

The Turag River is an upper tributary of the Buriganga originating from the Bangshi River, flows through the north side of Dhaka and joins with the Buriganga River at Mirpur, Dhaka. The water of Turag is used for different purposes like drinking, bathing, washing, agriculture and irrigation. It is also used as the main navigation channel connecting the capital city and other parts of the country resulting in tremendous transport pressure by different river vessels. Currently, this river faces many problems and is extremely afflicted by water pollution. The chemical wastes of mills and factories, domestic waste, medical waste, tannery waste, sewage, dead animals, plastics, and oil are pollutants in this river (Ahmed et al., 2013). These pollutants interacting with the river system deteriorate the water quality and adversely affect the aquatic ecosystem as well as the livelihood of the local community (Meghla et al., 2013). The Department of Environment (DoE) has listed this river as one of the ecologically critical areas (ECA) in the country (DoE, 2009).

The Dhaleshwari River

The Dhaleshwari River is an important distributary of the Jamuna river in central Bangladesh with a total length of about 290 km having an average depth of 37.19 m and a maximum depth of 80.79 m respectively (Ahsan et al., 2018). It starts off the Jamuna near the north-

western tip of Tangail District and divides into two branches: the north branch retains the name Dhaleshwari and merges with the other branch, the Kaliganga River at the southern part of Manikganj District. It supports the habitat for aquatic organisms including a wide variety of fish, invertebrates and aquatic vegetations on the riverbank. The water of this river is generally used for irrigation purposes as the river is surrounded by agricultural lands though some industrial plants are active along the bank. Therefore, agricultural runoff with insecticides, pesticides and fertilizers is frequently released into the river, especially during the rainy seasons (Islam et al., 2012; Ahsan et al., 2018). The Doleshwari river is comparatively less polluted than the Buriganga and the Turag having low industrial pressure, but the major threats include the removal of riparian vegetation, chemical waste from farmlands and navigational transportation.

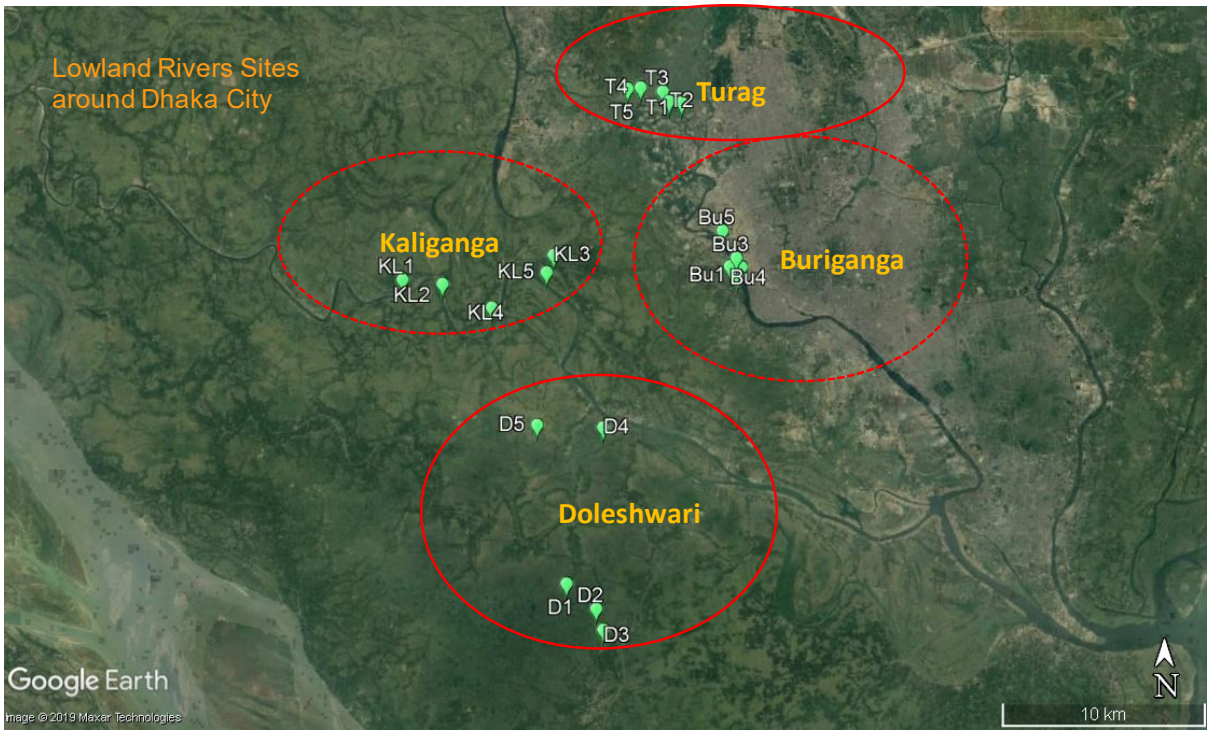

Figure 2. 2 Sampling sites in four lowland rivers where each site marked with river code

The Kaliganga River

The Kaliganga river is an upper tributary of the Dhaleshwari river. It cuts through the planes of Keraniganj and Nawabganj of Dhaka district and then emerges to the Jamuna River through Manikganj and Tangail districts. Like the Dhaleshwari river, the Kaliganga river is rich in diverse aquatic vertebrates and invertebrates and aquatic flora. Although industrial pollution is less evident, the discharge of chemical wastes from agricultural land is a key

concern to the biological health of the river. Household wastes from the local community may affect the water quality as people are found to use water for washing their clothes, bathing, washing their cattle etc.

2.3.1b Upland Rivers

Sampling sites at the upland rivers or streams were located in Bandarban, one of the hilly districts of the Chittagong Hill Tracts (CHT) in the southeast region of Bangladesh. The CHT includes three districts (Bandarban, Rangamati and Khagrachhari) which are enriched with many pristine hill streams. Most of these streams hold significant importance to the indigenous communities as their livelihood, social, cultural and religious affairs are inevitably linked with these pristine water bodies. Hill streams are highly variable and include a variety of habitats with boulders, pebbles, gravels, sands, cobbles and a relatively high proportion of leaf litter. These streams locally called *Chhora* flow through the tropical evergreen or semi-evergreen hill forest of the country. The riparian vegetation consists of shrubs and herbs. Besides terrestrial ecosystems of hilly areas, these streams are enriched with aquatic fauna in particular fish (Ahmed et al., 2013) and other invertebrate fauna including molluscs, beetles, water bugs, dragon and damselflies, mayflies, caddisflies and stoneflies etc. Along with global climate change, a variety of anthropogenic stressors including tourist visits, crop production, habitat fragmentation and household activities of tribal peoples are deteriorating the stream health that resulting in the loss of aquatic biodiversity. Removal of rocks for road construction is another significant stressor found during the present studies at several sampling sites. In total, 16 streams (Table 2.2 in the appendix and Fig. 2.3) were selected across the main river basin namely the Sangu River basin.

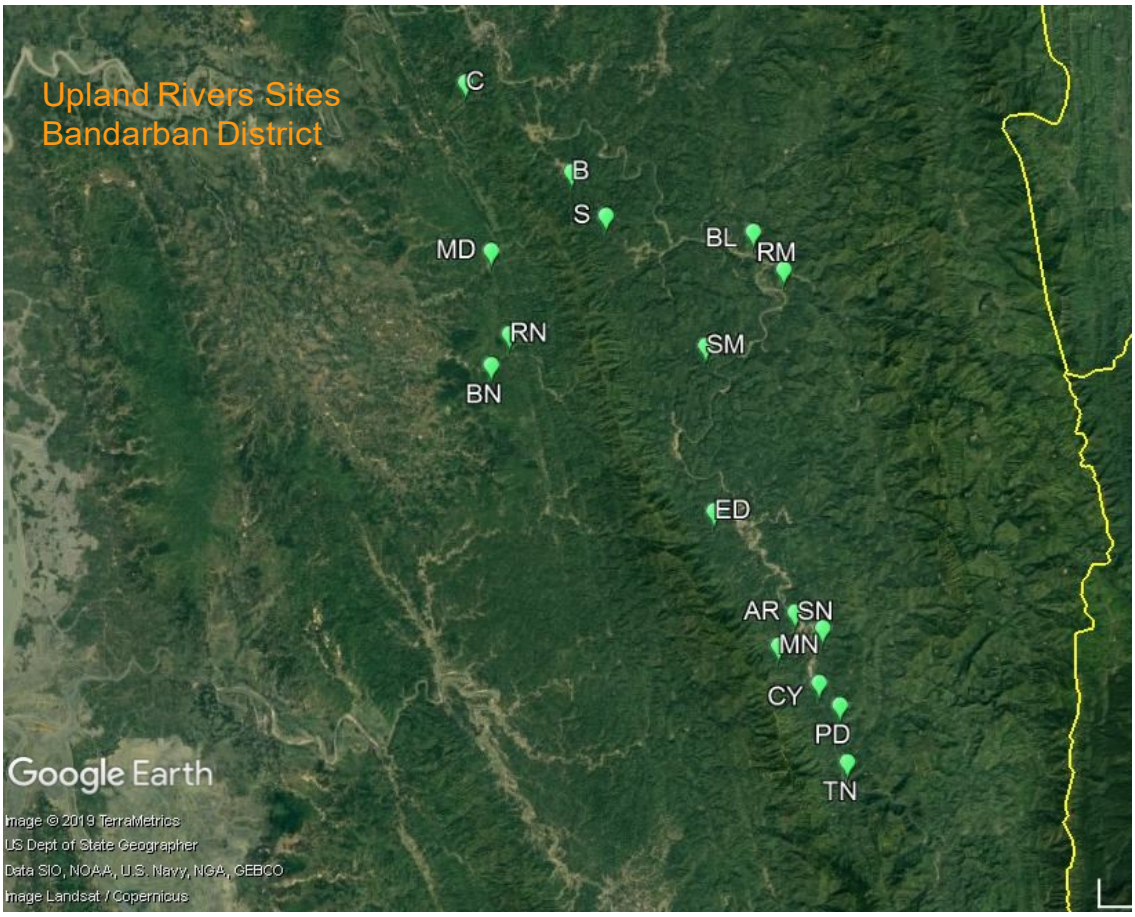

Figure 2. 3 Sampling sites in sixteen upland rivers where each river marked with river code

2.3.2 Field Sampling

Samples of macroinvertebrate communities were collected from the selected water bodies following standardized sampling techniques using two poles kick nets (mesh 500 microns). From 4 lowland rivers, five sampling sites were selected in each river covering the upstream to downstream channels (tributaries and creeks). The sampling reach for each site was around 500 m covering each bank of the river from where 2 samples were collected for barcoding and mitogenomic study of morphospecies to build a reference database. We also collected 2 samples from each of 16 upland rivers located at an environmental gradient area. The sampling reach for each stream was around 800-1000 m covering each bank of the river. Macroinvertebrate samples were processed in the field with consecutive sorting by sieving with different meshed (1 mm and 0.5 mm) sieve bucket that allows preliminary extraction of macroinvertebrate fauna and that were preserved in absolute ethanol.

2.3.3 Collection of morphospecies for building reference set

To develop reference data sets of macroinvertebrates across the sampling sites, morphospecies samples representing the main lineages of aquatic insects, molluscs and some crustaceans were taken from each sampling site prior to individual specimen extraction. A total of 960 morphospecies samples belonging to Coleoptera, Hemiptera, Odonata, Ephemeroptera, Plecoptera, Trichoptera, Diptera, Mollusca were sorted from lowland and upland samples. Morphological identification of the morphospecies specimens at the family level was undertaken *in situ* by visual inspection based on available taxonomic keys (e.g., Dobson et al., 2012) but more precise Linnaean identifications (e.g., genus or species) were made where possible. Specimens were then assigned to morphospecies within each lowest taxonomic level considering all observable features including body shape and proportions, total length, surface sculpturing, patterning and colouration.

2.3.4 DNA extraction from individual morphospecies samples

DNA was extracted from 960 morphospecies for mitochondrial COI and complete mitogenomes with the DNeasy 96 Blood and Tissue Kit following the manufacturer's protocols and eluted to a volume of 200 µl. Genomic DNA integrity for extracted DNA was assessed in some cases (for mitogenome targeted samples) by electrophoresis, migrating GelRed TM-stained DNA on an agarose 1.0% gel. Measurement of DNA quality and quantity with the Nanodrop ND-8000 (Thermo Scientific) system was done while 260/280 and 260/230 ratio and DNA concentration (ng/µl) were considered for the selection of samples for PCR amplification. DNA concentration of morphospecies samples (those are aimed for of MMG study) was also determined with the Quant-iT ds DNA HS assay kit using a Qubit R 2.0 Fluorometer (Life Technologies) to maintain an equilibrium concentration of MMG library preparation.

2.3.5 Amplification of COI barcode sequences

Standard PCR reactions to amplify mtDNA COI were undertaken for each of the morphospecies specimens. The 418 bp COI barcode was targeted using the redesigned degenerate primers. To maximise amplification of a diverse set of macroinvertebrates, these degenerate primers were designed with modifications of Folmer et al. (1994) and were

created from an alignment of complete mtDNA COI gene sequences for arthropods that were present in GenBank. The primers (fwd: CCNGAYATRGCNTTYCCNCG and rev: TANACYTCNGGRTGNCCRAARAAYCA) to which we attached the standard Illumina tails (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG for forward and reverse respectively) and 6-bp different tags to build Illumina ready PCR amplicon. The 25 µl polymerase chain reaction (PCR) mixtures contained 17.65 µl of sterilized ultrapure water, 2.50 µl of 10x PCR buffer, 0.75 µl of $MgCl_2$, 0.25 µl of dNTPs (10 mM each), 0.75 µl of each primer (10 mM), 0.1 µl of Taq DNA polymerase (5u/ µl, BIOTAQ^{TM} DNA Polymerase, BIOLINE), 0.25 µl of BSA (20 mg /ml, Thermo Scientific) and 2 µl of DNA template (20-150 ng/µl). DNA samples with high or low below this range were adjusted compromising the volume of water in the mixture. The PCR amplification conditions were as follows: 94^0C for 4 minutes (initial denaturation), 40 cycles at 94^0C for 30 seconds (denaturation), 48 ^{0}C for 30 seconds (annealing), 72^0C for 45 second (extension) and a final extension at 72^0C for 10 minutes. The PCR products were visualised on 1% agarose gel.

2.3.6 Sample pooling, library preparation and multiplex amplicon sequencing

PCR products of 960 morphospecies in 10 different plates were pooled together in single plates mainly based on the intensity of amplicon brightness on gel and also by averaging DNA concentration for the plates. The pooled PCR plates were then cleaned with Agencourt AMPure XP paramagnetic bead technology (Agencourt Bioscience Company, Massachusetts, USA) following the manufacturer's protocols with slight modification. After quality control, library preparation with secondary PCR and indexing with Nextera XT tags, amplicons are sequenced with the aim of 5000 reads per sample on an Illumina MiSeq platform (2x300 bp paired-end) at the sequencing facility of Earlham Institute, Norwich, UK.

2.3.7 Bioinformatics and amplicon data processing

Bioinformatic processing of paired-end raw sequences from Illumina MiSeq was carried out using the NAPtime pipeline (NGS Amplicon Pipeline), a set of Perl scripts developed by the wrapping of software for trimming, filtering, merging and clustering of NGS barcoding and metabarcoding sequences (Creedy et al., 2019; T. Creedy, pers.comm.). For analysis of barcode sequences, this pipeline includes several scripts namely NAPdemux, NAPtrim,

NAPmerge, NAPconvert and NAPselect, which perform step-by-step to produce finally processed barcode data. NAPdemux performs demultiplexing of paired-end read files in batch by acting as a wrapper for the excellent cutadapt program. The primary input to the script is a set of pairs of fastq files and tab or comma-delimited table(s) that specify the tag sequence(s) used in multiplexing each well. The outputs of this script are read files of each sample separated by samples' names and primer tags. NAPtrim carries out trimming of primers from a pair or set of paired files wrapping cutadapt or using fastx_trimmer (Hannon Lab, 2012). This script requires a set of pairs of fastq files as primary input and returns primer trimmed reads. This script allows primer trimming by either specifying the number of bases or by specifying a primer sequence. I specified the number of bases in forward and reverse primer. NAPmerge produces merged reads by merging of overlapping paired-end reads using PEAR (Zhang et al., 2014) with its quality control option –q as –pearquality. This quality control option determines the threshold for trimming low-quality parts of reads before merging. I used a PEAR-q value of 26 and fastq_filter expected error rate threshold of 1, also chosen by Arribas et al. (2016) and Creedy et al. (2019). NAPconvert performs the conversion of multiple fastq to fasta files using USEARCH110 fastq_filter with the maxee parameter using the -eemax option. I used the –eemax value of 1. NAPselect filters a set of fasta sequences for a single sample to select a putative barcode sequence for each individual organism using a barcode selection algorithm. This algorithm works on the assumption that the most frequent sequence (i.e. largest group) is likely to be the specimen barcode and is computed by two statistics: a bootstrap p-value and a BLAST score. Before selection, NAPselect dereplicates sequences and filters out any unique sequence specified by some parameters as -minsize (minimum number of sequences for a group), -minlength and –maxlength (minimum and a maximum length of the sequence to pass to selection). I set the values as 2 for minsize and 414 and 422 for -minlength and -maxlength respectively. I used –bootstraps of 10000 (number of times for reshuffling of sequences into groups), the –pvalue of 0 (maximum bootstrap pvalue at which a group is accepted as barcode) and a maximum threshold_pvalue of 0.5 (at which the entire sample should be discarded). Finally, NAPselect produces both high confidence and low confidence files for barcodes of all individual organisms.

2.3.8 Data analysis for building a barcode reference library of macroinvertebrates

NAPtime pipeline generated COI barcodes were transferred to Geneious software and after discarding ambiguous barcodes (which might have been created from sequencing errors) all other barcodes are grouped into respective orders and families. Generally, a threshold of sequence similarity of at least 97% is used to indicate potential species identification for animals. Therefore the retrieved COI barcodes were clustered using Usearch110 cluster_otus (Edgar, 2010) under a 97% similarity threshold for generating OTUs as a primary step for species delimitation under each family and order of macroinvertebrates. In addition, bPTP, a Bayesian PTP (Poisson Tree Processes) model was used for OTU or species delimitation. It is an updated version of the original maximum likelihood PTP with Bayesian support (BS) values to delimit species on the input tree (Zhang et al., 2013). This species delimitation was run via bPTP websites (https://species.h-its.org/). Sequences were aligned using the MUSCLE alignment method (Edgar, 2004) plugged in Geneious. Phylogenetic trees required for bPTP were constructed based on the maximum likelihood method (ML) in the RAxML (Randomized Accelerated Maximum Likelihood) software using default settings and GTRMIX-model (Stamatakis, 2006) on CIPRES (https://www.phylo.org/portal2/home.action). Furthermore, species delimitation was done through Refined Single Linkage (RESL) analysis on the BOLD platform which employs single linkage clustering as a tool for the preliminary assignment of an OTU and a subsequent finishing step that employs the Markov Clustering (MCL) approach (Ratnasingham and Hebert, 2013). For lower-level identification of all morphospecies barcodes, a BLAST search was performed against the NCBI database. The Sequence ID tool of the GBIF (Global Biodiversity Information Facility) platform was also used for the taxonomic assignment of barcodes where sequences are queried against a 99% clustered version of the International Barcode of Life project. The sequence comparisons, pairwise genetic distance (p-distance) and Kimura-2-Parameter (K2P) distance calculation (Kimura, 1980), with 1000 bootstrap replicates were performed using MEGA X (Kumar et al., 2018). Other statistical analyses were done using respective packages in R (R Core Team, 2018).

2.3.9 Sample Pooling, library preparation and sequencing of MMG samples

The MMG workflow starts with a pool of genomic DNA from multiple specimens which was shotgun sequenced using Illumina technology (Crampton-Platt et al., 2016). After measuring

the quantity and quality on Nanodrop and Qubit kits DNA from each selected morphospecies, DNAs were pooled in a library before shotgun sequencing on Illumina HiSeq platform. A total of 288 morphospecies samples were pooled in seven libraries and each library contained 40 to 60 samples. To minimize the effects of DNA concentration on assembly success across all samples, an approximate equimolar concentration of genomic DNA for each of the samples were maintained in a library aiming for a minimum of 20 ng of dsDNA in 50 µl required for the Illumina platform. About 200K paired-end reads were targeted from each of the morphospecies samples. Seven TruSeq libraries were generated, with an average insert size of 700 bp and were sequenced on a flow cell of Illumina HiSeq with 300 cycles and paired-end sequencing (2 x 300 bp reads) at the sequencing facility of Earlham Institute, Norwich, UK.

2.3.10 Bioinformatic process for MMG

After getting raw FASTQ files for each library from the high throughput sequencing pipeline, a quality control check of raw sequences was carried out using FastQC v0.10.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc). The raw sequences were trimmed of adapters using Trimmomatic (Bolger et al., 2014). This program is used to remove the adapter and index motifs associated with sequence reads coming from TrueSeq or Nextera libraries in the HTS pipeline. To simplify the de novo assembly of mitochondrial genomes, the complexity of the data sets was reduced by BLAST searching for similarity of the reads against database of known macroinvertebrate mitogenomes available at NCBI using BLASTn (E value 1 e-5; maximum target sequences 1; DUST filtering disabled) (Altschul et al., 1990). These putative mtDNA reads were assembled into full-length contigs using Celera (Myers et al., 2000), IDBA-UD (Peng et al., 2012) and Spades (Bankevich et al., 2012) assemblers (Andujar et al., 2015; Crampton-Platt et al., 2015). The resulting contigs from each assembler were blasted again against the reference database filtering for mtDNA hits for sequences that are at least 1 kb in length (E value =1e-5, maximum target sequences 1 with active DUST filtering). All mitochondrial contigs from three assemblies were imported to Geneious (version 8) and de novo assembled to combine overlapping sequences from all assemblers into longer scaffolds (Gillett et al., 2014; Crampton-Platt et al., 2015). To investigate the relationship between the number of generated sequencing reads and assembly success, all reads were mapped onto the obtained contigs using Geneious, allowing for 1% maximum mismatches, a maximum gap size of 2 bp and a minimum overlap of 1000 bp. From the

scaffolds, contigs were cleaned removing all mismatches of base-pairs and then consensus contigs were extracted as supercontigs. Taxonomic assignment of supercontigs was done through 'Map to Reference' with morphospecies barcodes as bait sequences in Geneious, as the identification of contigs facilitates their annotation with reference sequences. In each library, there were a few cases where meta-assembly resulted in chimeric copies of two full mitogenomes fused within the AT-rich regions. These were resolved by separating them manually in Geneious, searching for repeated regions that could be used to circularise each individual mitogenome. After chimera removal, the final gene annotation of each supercontig was carried out using the MitoZ toolkit (Meng et al., 2019). Annotated contigs from MitoZ were also checked again for start and stop codon mismatches and resolved by the alignment assessment (Muscle alignment) of each individual gene of all contigs. A flow diagram for bioinformatic processing of mitochondrial metagenomics is outlined in Fig. 2.4.

2.3.11 Phylogenetic study of macroinvertebrates' mitogenome

The available complete mitogenomes of Coleoptera, Odonata, Diptera, Ephemeroptera, Trichoptera and Hemiptera were downloaded from GenBank. Only "verified" (i.e. fully annotated) sequences were included in further analysis. Phylogenetic analyses were conducted for each order of the insects with the protein-coding sequences (PCGs) of explored mitogenomes combined with mitogenomes downloaded from GenBank/NCBI. The PCGs were aligned based on nucleotide sequences using MAFFT 7.402 (Katoh and Standley, 2013) and then concatenated with catfasta2phyml.pl (retrieved from https://raw.githubusercontent.com/ nylander/catfasta2phyml/master/). All morphospecies barcodes (COI) of each insect order were also aligned with the respective mitogenome alignment. Phylogenetic trees were then built for each order based on the maximum likelihood method (ML) in the RAxML (Randomized Accelerated Maximum Likelihood) programme using default settings and GTRMIX-model (Stamatakis, 2006) on CIPRES (https://www.phylo.org/portal2/home.action). Phylogenetic trees were visualized using an online tree display tool (iTOL: Interactive Tree Of Life) (Letunic and Bork, 2021).

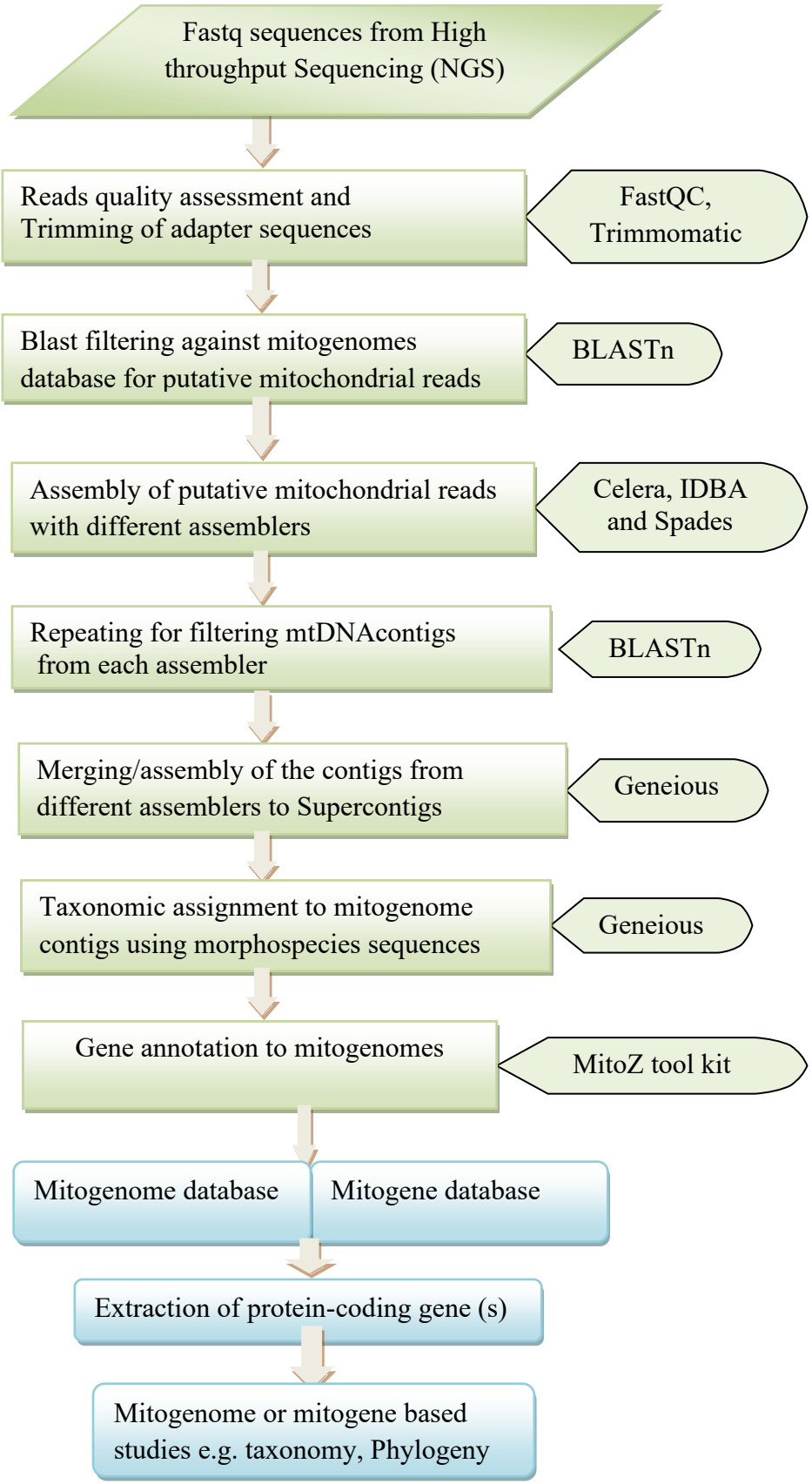

Figure 2. 4 Mitochondrial metagenomics (MMG) pipeline to extract mitochondrial genomes of macroinvertebrates

2.4. Results

2.4.1 DNA barcodes of freshwater macroinvertebrates of Bangladesh

A total of 812 COI barcodes for macroinvertebrates were obtained from 952 morphospecies collecting bulk samples from lowland and upland rivers of Bangladesh (Table 2.1). Most of these morphospecies were assigned to their respective families preliminarily based on morphological taxonomy. DNA barcodes represented a wide range of bioindicator macroinvertebrates including highly sensitive mayflies, stoneflies and caddisflies collectively called EPT (Ephemeroptera, Plecoptera and Trichoptera), moderately tolerant aquatic beetles (Coleoptera), water bugs (Hemiptera), dragon and damselflies (Odonata), shrimp and crabs (Decapoda), and also presumed highly disturbance-tolerant biting and non-biting midges (Diptera), freshwater snails and mussels (Mollusca), and annelids. The amplification frequency of COI fragments for arthropod taxa was similar across the orders and was higher than that of molluscan and annelid taxa.

Table 2. 1 DNA barcodes and the number of assigned species of Bangladeshi freshwater macroinvertebrates by thtree species delimitation processes. The number of morphospecies is the sum of morphologically distinct types obtained at each site.

Taxon	Morphospecies	COI fragment amplified	Barcodes obtained	OTUs/Species obtained		
				Usearch cluster	RESL	bPTP
Coleoptera	125	116	110	49	50	52 (49-62)
Ephemeroptera	152	145	136	36	37	41(41-47)
Plecoptera	20	17	16	7	8	8 (7-11)
Trichoptera	120	115	107	26	28	28 (27-37)
Hemiptera	80	76	74	31	34	35 (32-41)
Diptera	135	130	120	70	66	71(70-84)
Odonata	80	79	77	40	40	42 (40-48)
Decapoda	70	60	59	20	22	25 (24-28)
Annelida	40	25	24	12	14	15 (14-16)
Mollusca	130	92	89	31	33	45 (38-58)
Total	952	855	812	320	332	362

2.4.2 Species delimitation from COI barcodes of macroinvertebrates

Overall, DNA barcoding resulted in 320-262 species from successfully obtained 812 barcodes across target groups. The number of species delimited by clustering (at 97% similarity) algorithm (Usearch), phylogeny-based bPTP, and RESL varied across the families under each respective order. For most macroinvertebrates groups, bPTP detected more species than the Usearch clustering algorithm and RESL analysis (Table 2.1). Species that were confirmed by these three methods simultaneously, only those were finally accounted for species in this study. These included 36 Ephemeroptera species (5 families), 7 Plecoptera species under a single family Perlidae, 26 Trichoptera species (9 families), 47 Coleoptera species (10 families), 29 Hemiptera species (11 families), 38 Odonata species (6 families), 65 Diptera species (8 families), and 31 Mollusca species (9 families) (Fig. 2.5). Along with the family level identification during the sorting of morphospecies, a BLAST search of all barcodes against the NCBI database was carried out to reconfirm their family level taxonomic assignment. However, there were some species delimited barcodes (2 for Coleoptera, 2 for Hemiptera, 2 for Odonata and 5 for Diptera) that could not be assigned to their family and were mentioned as unspecified (Fig. 2.5). Additionally, 20 species of Decapoda and 12 species of Annelida were also delimited from 59 and 24 barcodes respectively.

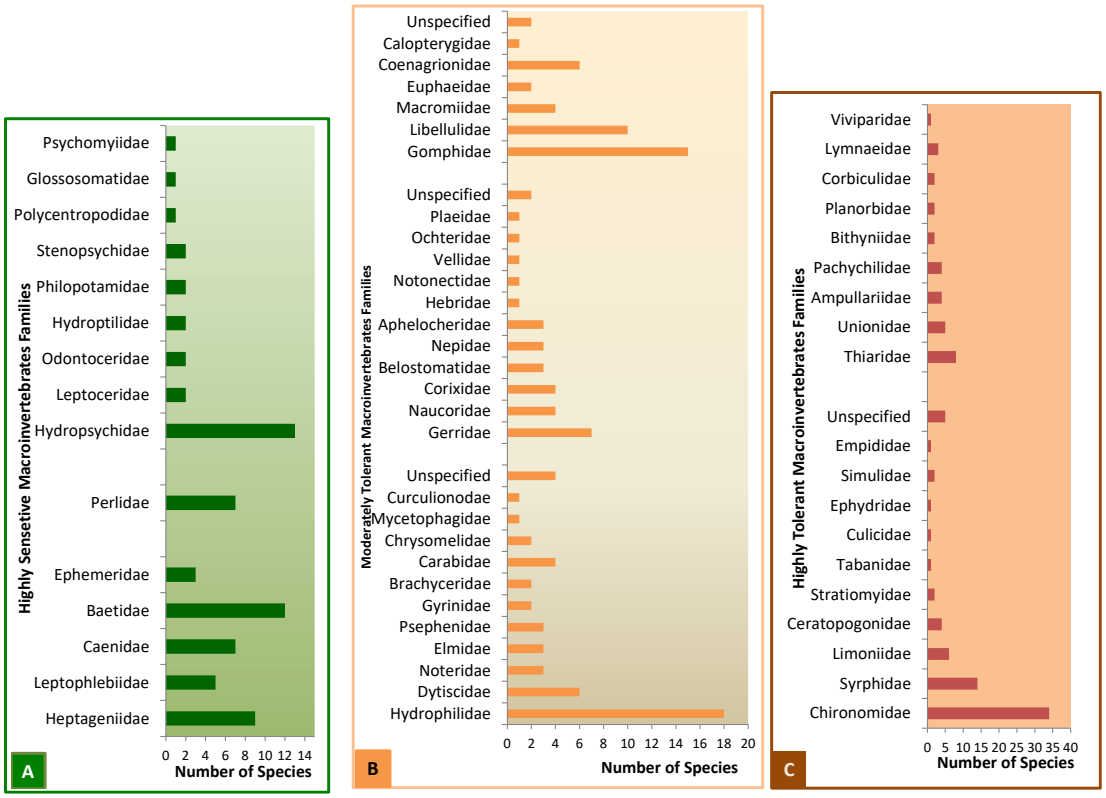

Figure 2. 5 Number of species identified using DNA barcodes under different families of Bangladeshi freshwater macroinvertebrates. A) Highly sensitive families of Ephemeroptera, Plecoptera and Trichoptera. B) Moderately tolerant families of Coleoptera, Hemiptera and Odonata. C) Highly tolerant families of Diptera and Mollusca.

2. 4.3. Species variation in macroinvertebrates under different families

2.4.3a EPT (Ephemeroptera, Trichoptera and Plecoptera)

Among the mayflies (Ephemeroptera), the number of delimited species varied across the families. Baetidae had the highest number of species groups followed by the family Heptageniidae for which bPTP and RESL delimited the same number of species as for Baetidae. Ephemeridae contained the lowest number of species that were equally detected by bPTP, Usearch and RESL. These methods produced the same result for the family Caenidae and Heptagenidae where each family contain 7 and 9 species respectively. In the case of Leptophlebiidae family, bPTP methods identified a higher number of species than the other two methods (Figs. 2.6, 2.7). For the caddisflies (Trichoptera) group, the highest numbers of species were identified under the family Hydropsychidae. The total number of species in 8 other families was almost equal to the number of hydropsychid species. For the families

Leptoceridae, Hydroptilidae, Philopotamidae, Odontoceridae and Stenopsychidae, all delimitation processes returned congruent results though 97% clustering method produced a slightly lower number of species for Hydropsychidae (Fig. 2.8). Under the stoneflies (Plecoptera) group, all morphospecies were assigned to a single-family Perlidae, which contained 7 to 8 species delimited by the three methods (Fig. 2.9).

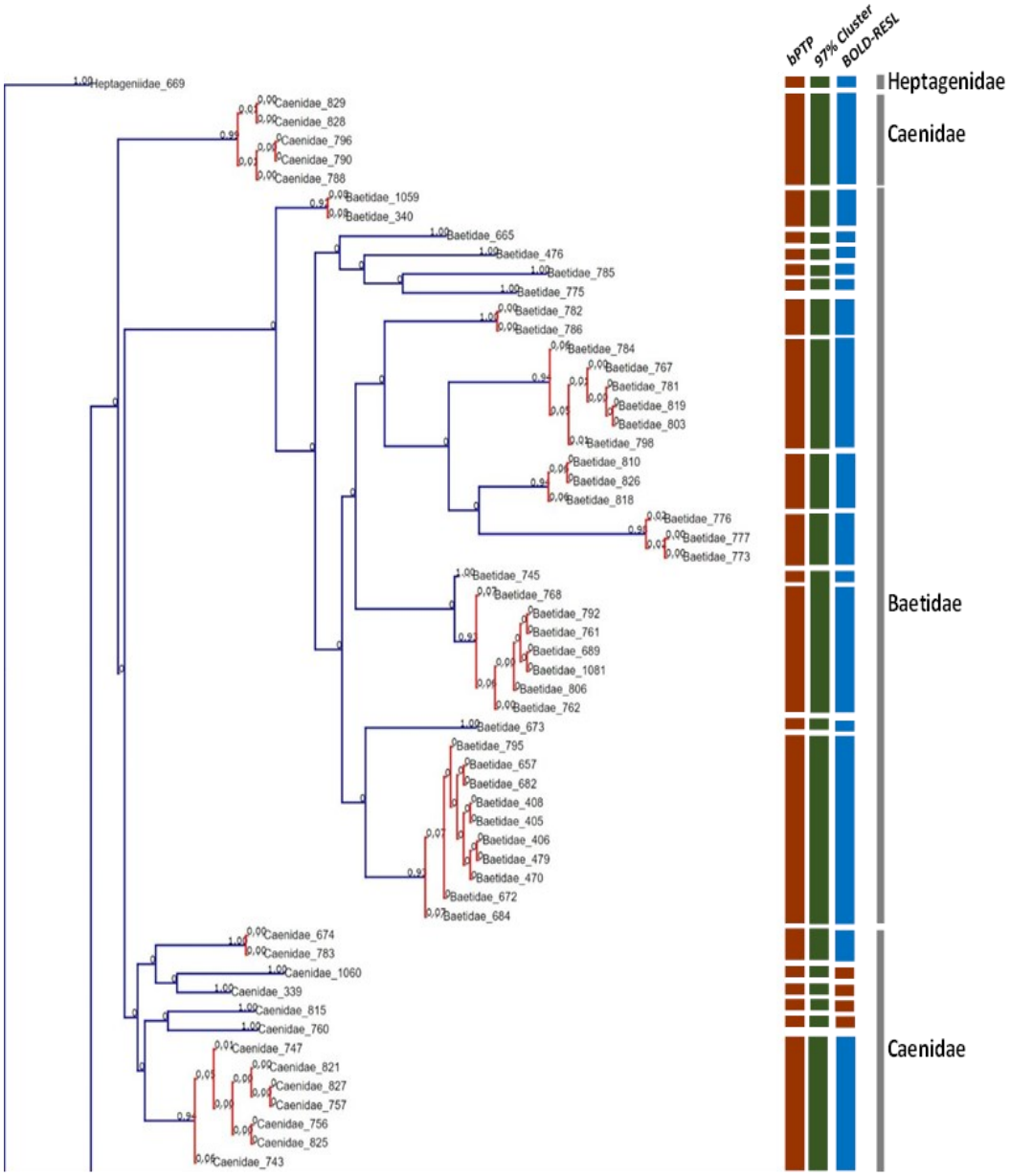

Figure 2. 6 Molecular species-delimitation analysis of the Ephemeroptera (Heptagenidae, Caenidae, Baetidae,) spp. by three methods: a Bayesian implementation of the Poisson tree processes (bPTP), Sequence clustering by 97% threshold clustering and Refined Single Linkage (RESL) analysis. Delimitation results are visualized as bars on a Bayesian maximum clade credibility tree of the cytochrome c oxidase subunit 1 gene. The node numbers indicate Bayesian posterior probabilities (Bayesian support values).

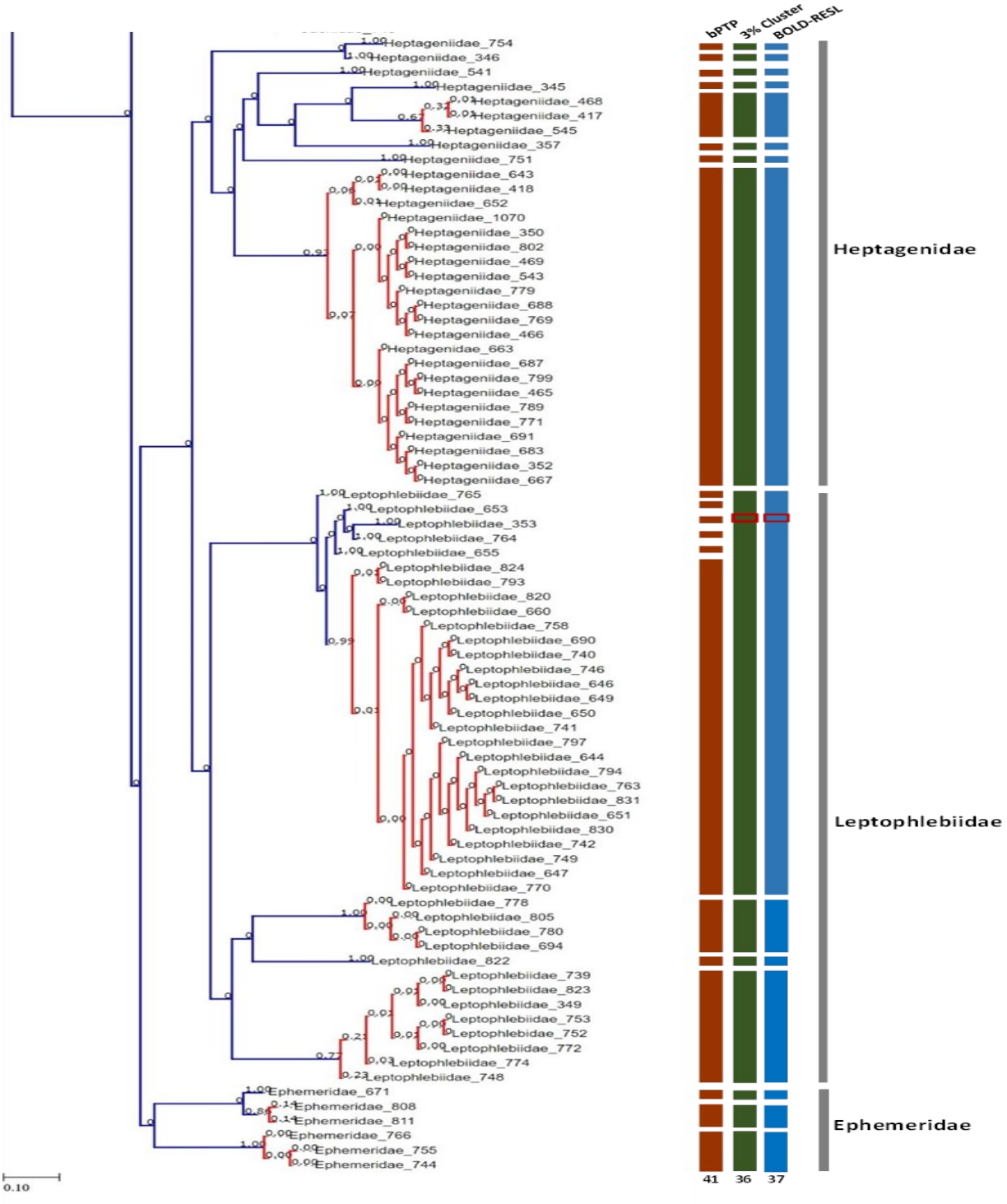

Figure 2. 7 Species-delimitation analysis of Ephemeroptera spp. The number below each column bar denotes the total species delimited by the respective method

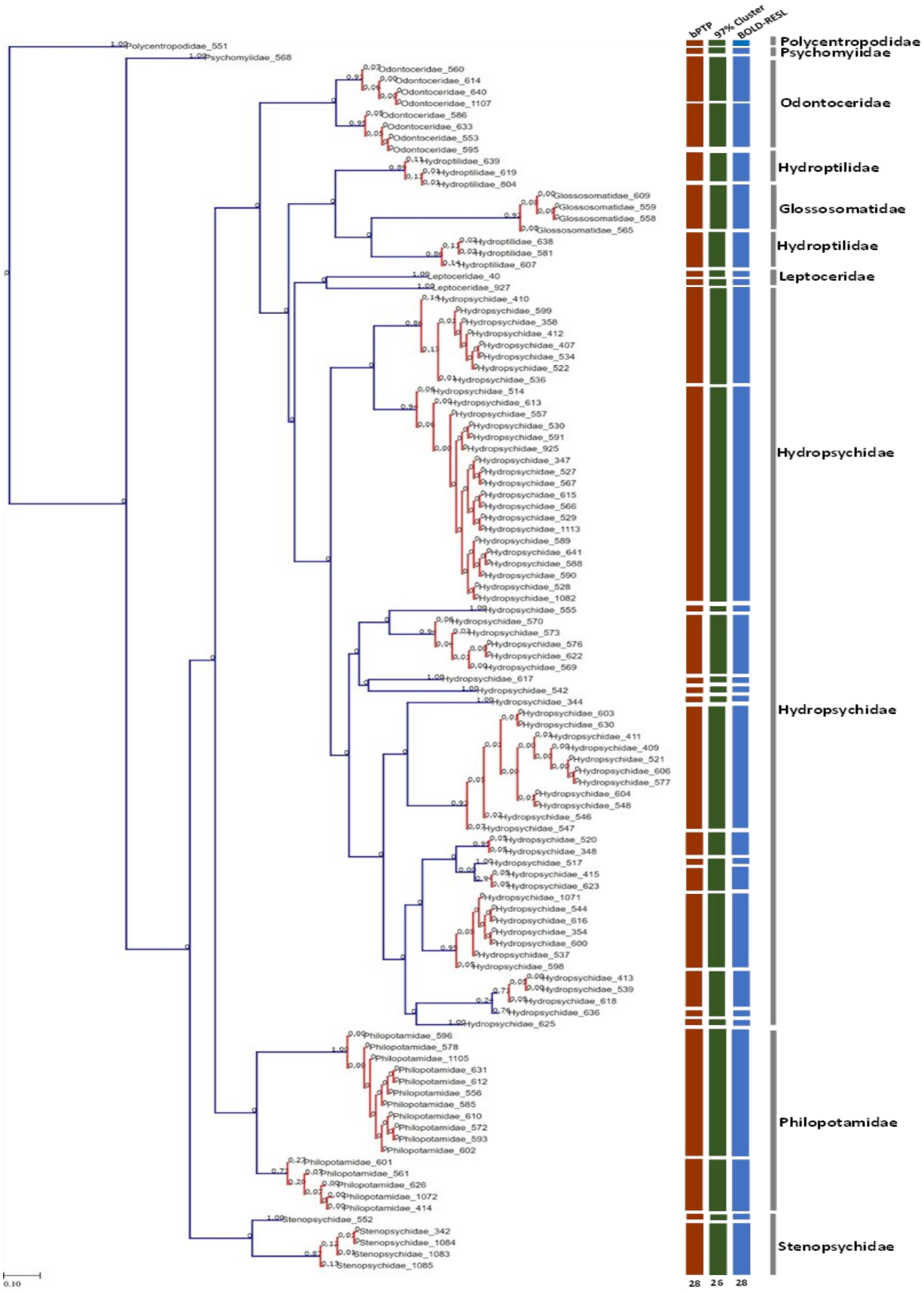

Figure 2. 8 Species-delimitation analysis of the Trichoptera.

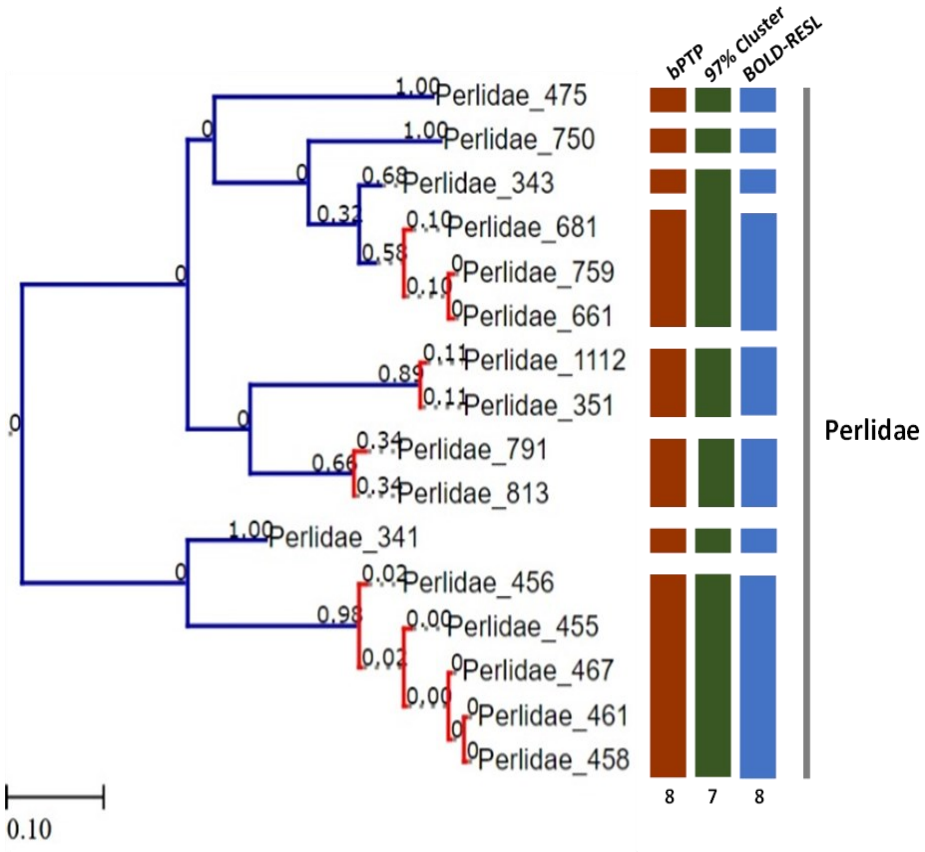

Figure 2. 9 Species-delimitation analysis of the Plecoptera.

2.4.3b Aquatic beetles (Coleoptera)

The family Hydrophilidae (water scavenger beetles) contained the highest number of species (18-19 spp.) and then diving beetles were the second-most species-rich group under the family Dytiscidae (6-7 spp.). Except for Hydrophilidae and Dytiscidae, for all other families of beetles, similar results were produced by bPTP, Usearch clustering and RESL delimitation procedures. Families such as Noteridae, Elmidae, and Psephenidae were detected as equally species-rich (3 spp.) families of water beetles. The whirligig beetles family, Gyrinidae was found to contain only two species. Several species of ground beetles (Carabidae), leaf beetles (Chrysomelidae) and weevils (Brachyceridae and Curculionidae) were also identified. Meanwhile, 4 barcodes mentioned here as unspecified could not be assigned to their families (Fig. 2.10).

2.4.3c Water bugs (Hemiptera)

Water bugs in the order Hemiptera consisted of divergent species including water striders (Gerridae), water scorpions and stick insects (Nepidae), creeping water bugs (Naucoridae), aphelocheirid bugs (Aphelocheridae), water cricket (Veliidae), lesser water boatman (Corixidae), greater water boatman or backswimmer (Notonectidae) and pigmy backswimmer (Pleidae). In addition, single species from each of two other families (Hebridae and Ochtheridae) were also detected. In the context of species delimitation, all three approaches detected an equal number of species for most of the families except for Corixidae where one more species was delimited by bPTP. The family Gerridae was the highest species (9 spp.) containing family followed by Corixidae (5 spp.) among the water bugs. Naucoridae and Belostomatidae were the third species-rich groups while Nepidae and Aphelocheridae placed in the fourth position containing an equal number of species (3 spp.). The species delimitation results for Hemiptera are visualized as bars on a Bayesian maximum clade credibility tree (Fig. 2.1 in the appendix).

2.4.3d Dragonflies and Damselflies (Odonata)

Dragonflies species belonging to three families (Gomphidae, Libellulidae, and Macromiidae) were more diversified than damselflies which also consisted of three families (Coenagrionidae, Calopterygidae Euphaeidae). As expected, Gomphidae and Libellulidae were found to be the first and second most diverse family respectively. bPTP, Usearch clustering and RESL detected the same number of species for all families except for Coenagrionidae where 2 more species were detected by bPTP. bPTP slightly inflated species numbers for the damselfly family, Coenagrionidae. Two unspecified species could not be given family-level taxonomic assignments. The species delimitation results for Odonata are visualized as bars on a Bayesian maximum clade credibility tree (Fig. 2.2 in the appendix).

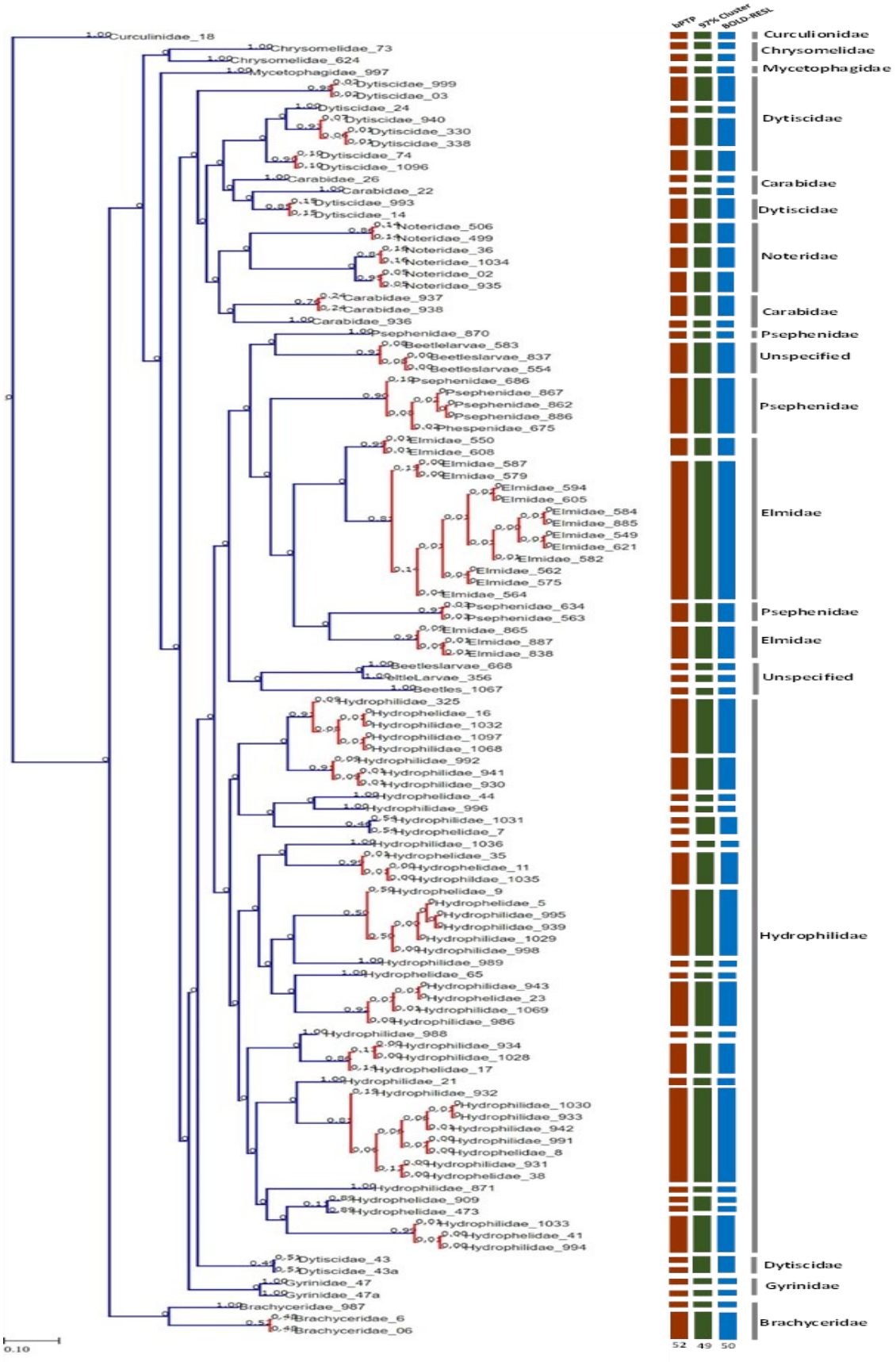

Figure 2. 10 Species-delimitation analysis of the Coleoptera.

2.4.3e Diptera and Mollusca

Highly tolerant dipteran and molluscan families contained nearly one-third of the identified species, of which the non-biting midges (Chironomidae) was the most species-rich family among all macroinvertebrates. Besides non-biting midges, other families combined constituted less than half of dipteran species. In particular, Culicidae, Tabanidae, Ephydridae and Empididae were the least species-rich families with only a single species in each family. All species delimitation methods produced similar results for all families except for Chironomidae where bPTP and RESL delimited one more species than the clustering method. Among molluscan macroinvertebrates, the number of species delimited under Gastropod snails was more than half of Bivalve mussels (Unionidae and Corbiculidae). As a family, Unionidae (freshwater mussels) was the highest species containing group among all molluscs and this number was inflated by the bPTP method. Among gastropods, Thiaridae (Trumpet snails) was the highest species containing group. The species delimitation results for Diptera and Mollusca are visualized as bars on a Bayesian maximum clade credibility tree (Fig. 2.3 and 2.4 in the appendix).

2.4.4 Genetic distance and barcoding gap among macroinvertebrates

The K2P distances and p-distances were compared at the inter-specific within each family and the intra-specific level of studied macroinvertebrates groups. It was quite evident that K2P distances were higher than p-distances for all families of macroinvertebrates.

2.4.4a EPT (Ephemeroptera, Plecoptera and Trichoptera)

The K2P and p-distances varied across the ephemeropteran families. The inter-specific distances for both models were lowest in the family Leptophlebiidae and averaged 21% and 19% respectively. The highest inter-specific K2P distances and p-distances were found in Baetidae which ranged from 20% to 31% and 18% to 25% respectively. For other families, the distances varied with a close range albeit the lowest inter-specific distances were more than 19% for both models (Fig. 2.11 left). The K2P and p-distances at the intra-specific level were nearly equal for all ephemeropteran species and ranged from 0.1 to 2.8 % across the species. For most of the baetid species, intra-specific distances were lower than other species groups that were limited to 0.50 %. The highest intra-specific distance was for ephemerid species and was estimated at around 1.8 %. The higher ranges of intra-specific difference

were recorded for a heptageniid (species 352, Fig. 2.11 right) and a leptophlebiid (species 644) species and for both of these species, the distance was nearly 1.50 % (Fig. 2.11 right).

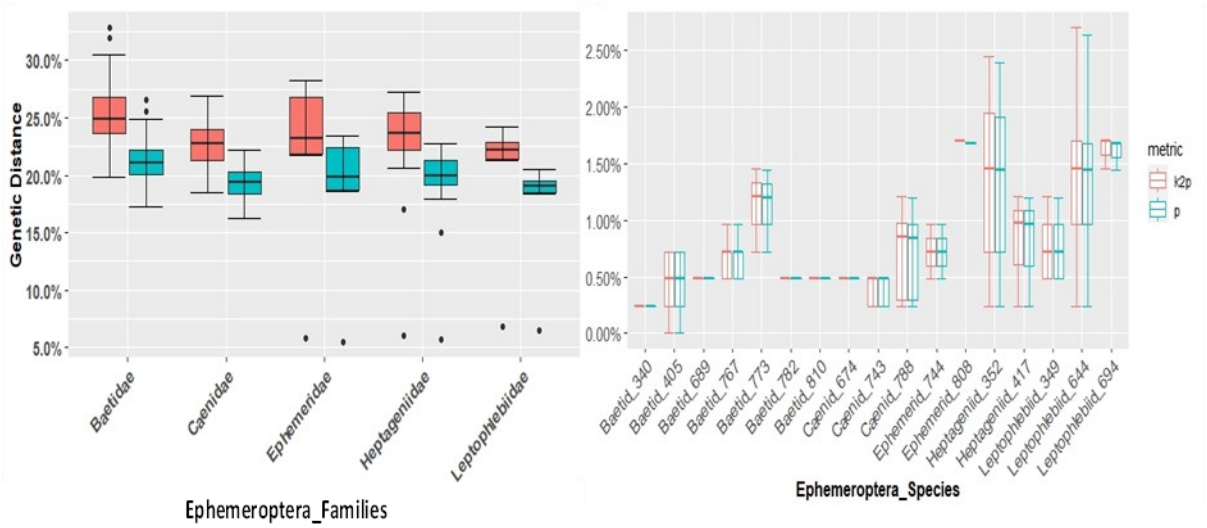

Figure 2. 11 Interspecific (within each family) (left) and intraspecific (right) genetic distance of Ephemeroptera based on K2P and p-distance models.

For trichopteran families, a significant variation in distances was measured for both distance models, where most of the families consisted of two species. Except for the family Odontoceridae (12 % for K2P and 11 % for p-distance), the lowest distances were estimated at over 15% for all other families. The highest inter-specific distance was found in Leptoceridae species which was more than 25 %. In the most species-rich family Hydropsychidae, the inter-specific distance was calculated as 25 % and 22 % for K2P and p-distance models respectively (Fig. 2.12 left). The intra-specific distances for most of the caddisfly species were found to be lower (0.5 %) like mayfly species. Notably, all hydropsychid, glossosomatid, hydroptilid, odontocerid, stenopsychid species contained the intra-specific distance within the range of 0.2 to 1.00 % except for one philopotamid species (species 414) where the average distance was calculated as around 1.50 % (Fig. 2.12 right).

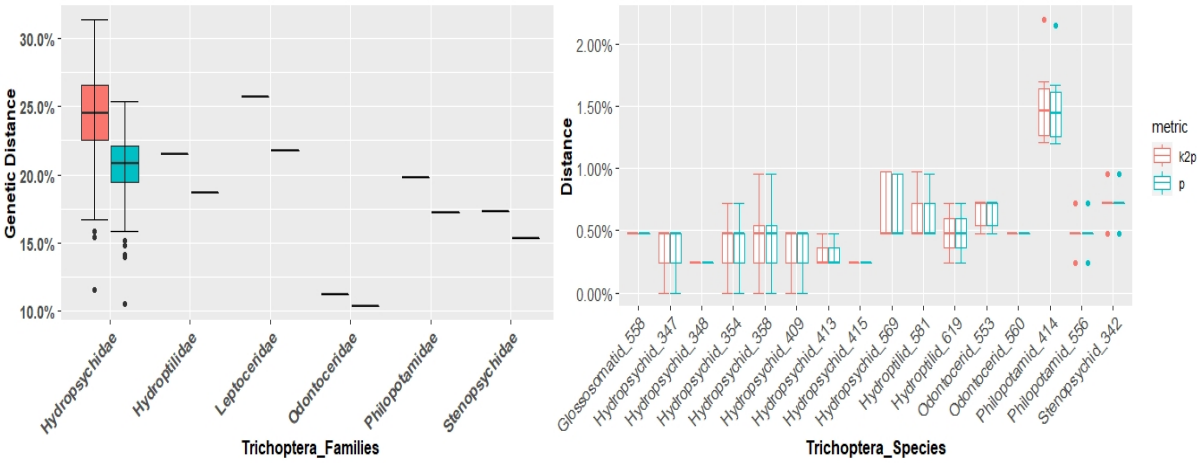

Figure 2. 12 Interspecific (within each family) (left) and intraspecific (right) genetic distance of Trichoptera based on K2P and p-distance models.

Under the order Plecoptera, a single-family was identified in which the K2P and p-distances inter-specific distances spanned from 18% to 29% and 16% to 24% respectively (Fig. 2.13 left). The estimation of intraspecific distances in four species of stoneflies showed a relatively higher intra-specific distance than mayflies and caddisflies. For example, for each of the perlid species, the average intra-specific distance was higher than 1.2 % except for one (species 351) where that was around 0.5 % (Fig. 2.13 right).

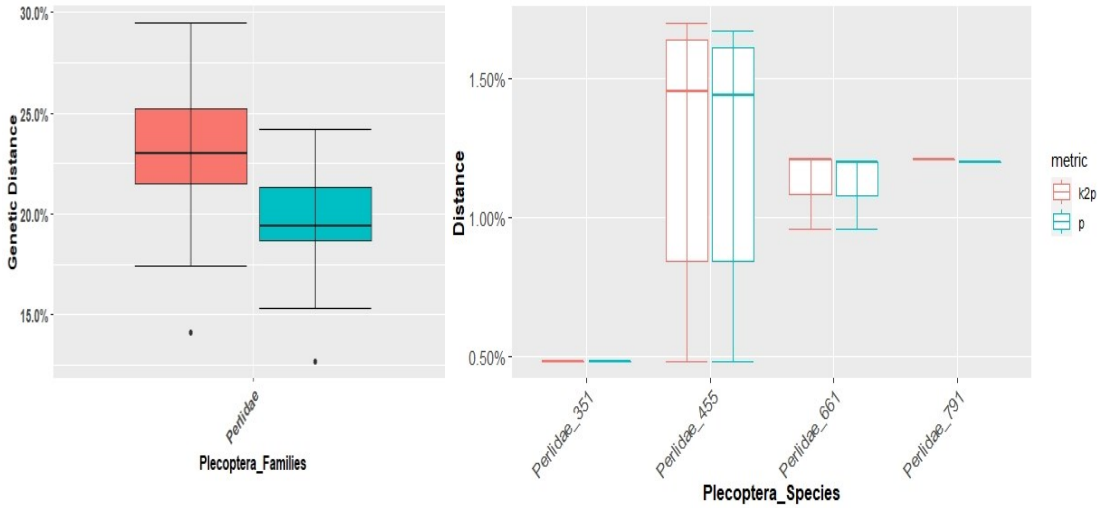

Figure 2. 13 Interspecific (within each family) (left) and intraspecific (right) genetic distance of Plecoptera based on K2P and p-distance models.

2.4.4b Coleoptera and Hemiptera

In aquatic beetles, the range of inter-specific distances within each family is comparatively higher than other macroinvertebrates families except for the family Gyrinidae where the distances for K2P and p-distances models were estimated as 4% and 3% respectively. The lowest K2P and p-distances were 18% and 16% in the family Brachyceridae and their highest values reached 28% and 23% respectively. The highest variation in both distance metrics was found in the family Noteridae which ranged from 4% to 24% and 3% to 22% respectively. The average interspecific distance in all families was more than 15% for both K2P and p-distances models (Fig. 2.14 left). As expected, the range of intra-specific distances (both K2P and p-distances models) for most of the water beetle species were also higher than other macroinvertebrates except for *Berosus* and *Laccophilus* species. The highest intra-specific distance was found in *Coelostoma* sp. and Noterid sp. that exceeded 1.2%. However, a higher variation in the intra-specific distance was recorded for *Sternolophus sp*, *Ragimbarti*a sp., Dytiscid sp. (species 330) and elmid species respectively (Fig. 2.14 right).

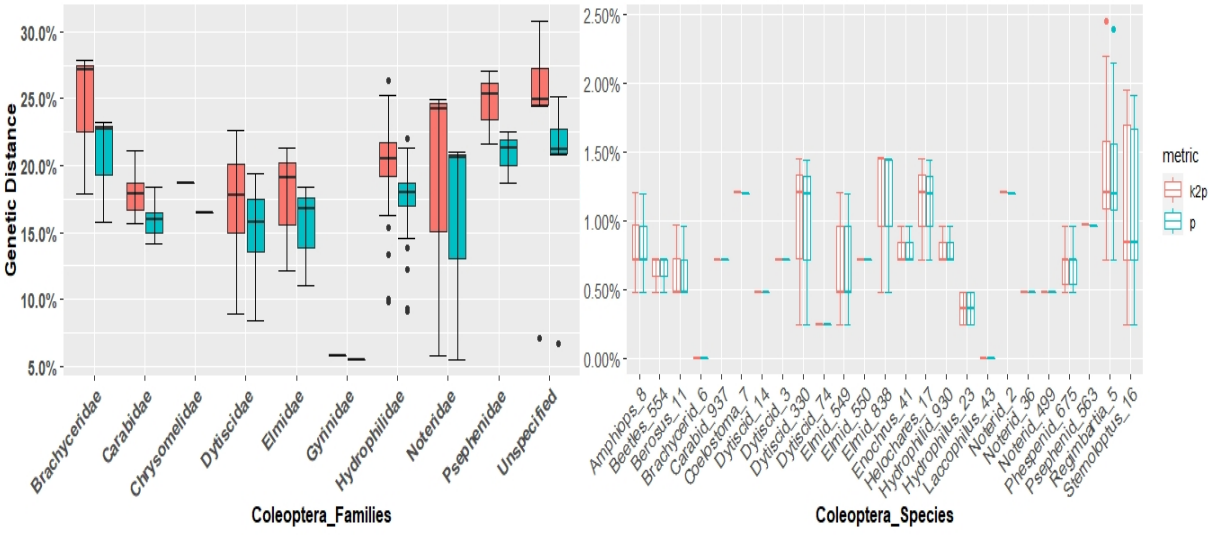

Figure 2. 14 Interspecific (within each family) (left) and intraspecific (right) genetic distance of Coleoptera based on K2P and p-distance models.

In water bugs (Hemiptera), overall distance metric values also reached over 15% in all families except the family Aphelocheridae in which the inter-specific distances were estimated as 12% and 11% for K2P and p-distance metrics respectively. The overall variation in all other hemipterans families was within a range of 15 to 25%. The highest inter-specific distances were observed for the lesser water boatman family (Corixidae) which were 23% for K2P and 20% for the p-distance metric (Fig. 2.5 in the appendix).

The significant differences in intra-specific were found in some species of aphelocherid bugs, for instance, the genetic distance among the individuals of one aphelocherid sp. (Species 849) was estimated as nearly 1.5 % whereas it was 0.5 % for other species of aphelocherid water bugs. The intra-species distance in all giant water bug species (Belostomatids) was almost the same as 0.5 %. The lowest intra-specific distance (0.25%) was found in a backswimmer species while the individuals of a creeping water bug species (Species 103) were genetically distant from other water bug species (Fig. 2.5 in the appendix).

2.4.4c Odonata and Diptera

In all odonate families, the overall lowest inter-specific distances for K2P and p-distance metrics was over 13% and the highest variation was measured as 14 to 28% (K2P) and 13% to 22% (p-distance) in Gomphidae. The inter-specific variations were more or less similar in the families Libellulidae and Macromiidae. The inter-specific distances for K2P and p-distance in damselfly family Coenagrionidae were measured as 18% to 22% and 17% to 18% respectively. Among dragonfly families, the lowest inter-specific distances were observed in the family Euphaeidae (Fig. 2.15 left). Among dragonflies and damselflies, the highest (1.1 %) and lowest (0.1 %) intra-specific distance was estimated in damselfly species. Notably, the intra-specific distances for all species of Euphaeidae and Gomphidae were equal to 0.5 %. Among dragonflies, the species belonging to Libellulidae and Macromiidae had higher intra-specific distances than others (Fig. 2.15 right).

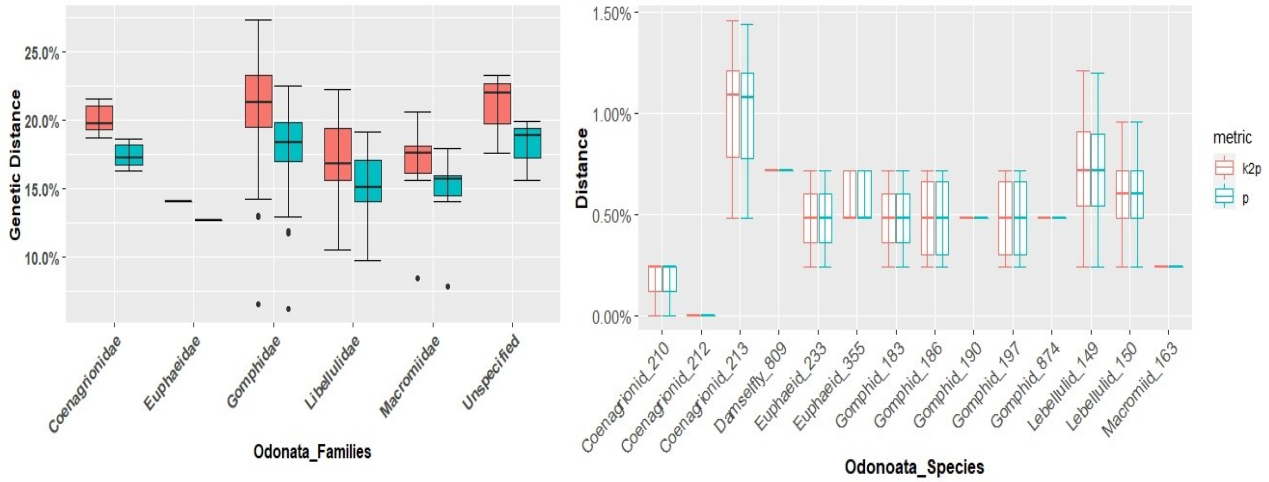

Figure 2. 15 Interspecific (within each family) (left) and intraspecific (right) genetic distance of Odonata based on K2P and p-distance models

Among the dipteran families, the non-biting midges family, Chironomidae contained the maximum inter-specific distances variation both for K2P and p-distance metrics. Although the overall inter-specific distances significantly varied for each family, the distance variation patterns were close in Limoniidae and Syrphidae families. The lowest distances were in the family Syrphidae measured as 3% and 5% for K2P and p-distance metrics respectively. The differences in intra-specific distances significantly varied within dipteran species. The lowest and highest intra-specific genetic distances were also observed in chironomid species. In case of non-chironomid dipterans like Limoniid, Simuliid and Syrphid had over or around 1% generic distances within the individuals of each species (Fig. 2.6 in appendix).

2.4.5. Taxonomic assignment of delimited species

Species identification against Genbank and GBIF entries using BLAST and sequence ID tool showed sequence similarity levels mostly outside of the widely applied 3% threshold of within-species diversity, indicating the lack of close relatives in the database. However, matches within the 3% interval were obtained in all target groups, except Plecoptera, as follows: Ephemeroptera-1 species by GenBank and GBIF of 36 delimited species, Trichoptera-10 species by GenBank and GBIF of 27 species, Coleoptera-10 species by GenBank and 14 species by GBIF of 30 species, Hemiptera-7 species by GenBank and 8 species by GBIF of 31 species, Odonata-20 species by GenBank and 22 species by GBIF of 40 species, Diptera-25 species by GenBank and 29 species by GBIF of 70 species, and

Mollusca -15 species by NCBI and 13 species by GBIF of 31 species. This estimation resulted in an overall proportion of hits at the 3% level (presumed species-level) of 25 to 55 %. The proportion of OTUs matched within the 3% level reflected known differences among major lineages in dispersal propensity and geographic ranges, which are generally highest in Odonata, followed by Trichoptera. The species of Coleoptera, Trichoptera and Ephemeroptera within the 3% level against Genbank and GBIF entries are given in Table 2.2. The identity of other taxa sequences is also given in the appendix (Table 2.3, 2.4 and 2.5).

Table 2. 2 Species identification of Coleoptera, Trichoptera and Ephemeroptera against Genbank and GBIF entries (within 3% level) using BLAST and sequence ID tool.

Major taxa group	Species_ID	NCBI		GBIF	
		Identity %	Nearest taxa matched	Identity %	Nearest taxa matched
Coleoptera	Brachyceridae_6	100	*Neochetina bruchi*	99.761	*Neochetina bruchi*
	Carabidae_22	100	*Bembidion xanthacrum*	100	*Bembidion xanthacrum*
	Dytiscidae_1096	99.761	*Cybister tripunctatus*	99.761	*Cybister tripunctatus*
	Hydrophilidae_992	99.761	*Sternolophus rufipes*	99.761	Hydrophilidae
	Brachyceridae_987	99.282	*Neochetina eichhorniae*	99.761	*Neochetina eichhorniae*
	Dytiscidae_338	99.043	*Coleoptera* sp.	99.043	*Dytiscus alaskanus*
	Dytiscidae_24	98.753	*Dytiscus* sp.	98.321	*Dytiscus marginalis*
	Hydrophilidae_1032	98.565	*Hydrophilus olivaceus*	99.522	*Hydrophilus triangularis*
	Mycetophagidae_997	98.565	*Mycetophagus* sp.	98.565	*Mycetophagus* sp.
	Chrysomelidae_73	97.368	*Galerucella nipponensis*	97.368	Galerucella sp.
	Hydrophilidae_934	---	---	98.804	Hydrophilidae
	Dytiscidae_993	---	---	98.982	Dytiscidae
	Hydrophildae_1035	---	---	99.282	Berosus sp.
	Carabidae_936	---	---	99.282	Carabidae
Trichoptera	Hydropsychidae_347	100	*Ceratopsyche guatitas*	100	*Hydropsyche* sp.
	Leptoceridae_927	99.761	*Setodes fluvialis*	99.761	*Setodes* sp.
	Stenopsychidae_552	99.761	*Stenopsyche benaventi*	99.761	*Stenopsyche benaventi*
	Hydroptilidae_639	99.522	*Hydroptila thuna*	99.522	*Hydroptila thuna*
	Hydroptilidae_581	99.043	*Orthotrichia lanna*	99.043	*Orthotrichia lanna*
	Hydropsychidae_555	98.804	*Potamyia phaidra*	98.804	*Potamyia phaidra*
	Philopotamidae_602	98.804	*Chimarra wiharawela*	99.034	*Chimarra* sp.
	Stenopsychidae_1083	98.783	*Stenopsyche sp.*	98.783	*Stenopsyche* sp.
	Odontoceridae_586	98.086	*Marilia* sp.	98.086	Marilia sp.
	Hydropsychidae_625	97.608	*Cheumatopsyche globosa*	97.608	*Cheumatopsyche globosa*
Ephemeroptera	Caenidae_1060	97.368	*Caenis* sp.	97.368	*Caenis* sp.

2.4.6 Mitogenome extraction from macroinvertebrates

In total, 287 morphospecies were selected from different groups of macroinvertebrates. After shotgun sequencing of these morphospecies, all obtained contigs were processed with different bioinformatics tools (described in Methods). I used three assembler programmes (Celera, IDBA-UD, Spades) separately to assemble the short contigs. Afterwards, outputs from three assemblers were combined to produce final mitogenomes. Barcodes for the same morphospecies were used as bait sequences to identify mitogenomes. A total of 282 contigs ranging from 3000 to ~ 17000 bp in length were extracted (Table 2.3). Duplicate, triplicate and in some cases multiple contigs were produced from the morphospecies which affected the total number of different mitogenomes. As a result, only 108 complete mitogenomes (>15000 bp) for different morphospecies were finally filtered from 127 complete mitogenomes. Moreover, 89 partial contigs (3000 to <15000 bp) were also produced from other morphospecies. In total, 197 different contigs were extracted from 287 morphospecies specimens (Table 2.3).

Mitogenome extraction success varied across the taxonomic groups (Fig. 2.16). The overall extraction rate was (partial and complete contigs) more than 80% for Coleoptera, Ephemeroptera, Trichoptera, Hemiptera, Diptera and Odonata. For Plecoptera and Annelida it was around 50% and for Mollusca and Crustacea, it was below 40%. In all cases, the proportion of complete mitogenome extraction was even lower, and no complete mitogenome was produced for any molluscan morphospecies.

Table 2. 3 Mitogenome extraction from different taxa using MMG pipeline

Taxa group	Number of morpho-species selected	Number of mitogenome contigs extracted				Morpho-species with partial mito-genome	Morpho-species with complete mito-genomes
		3k-10k bp	>10k bp	Complete mitogenome>15k bp	Total		
Coleoptera	33	18	27	22	45	10	18
Ephemeroptera	17	4	12	12	16	5	9
Plecoptera	6	0	3	2	3	1	2
Trichoptera	14	12	10	9	22	6	7
Hemiptera	30	14	12	12	26	15	10
Diptera	50	23	45	35	68	12	30
Odonata	40	5	49	25	54	11	25
Annelida	22	7	10	6	17	10	3
Mollusca	59	22	2	0	24	15	0
Shrimp+Crab	10	0	4	3	4	2	3
Others	6	2	1	1	3	2	1
Total	**287**	**107**	**175**	**127**	**282**	**89**	**108**

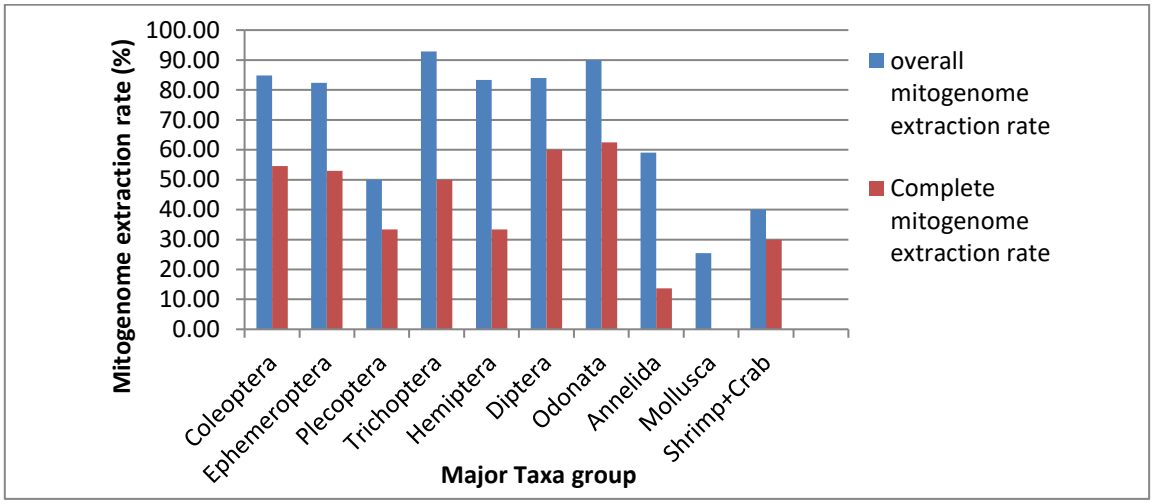

Figure 2. 16 MMG's success for extracting partial and complete mitogenomes in different taxa

2.4.7 Mitogenome Phylogeny of macroinvertebrates

Phylogenetic trees of major insect orders were built from the protein-coding genes based on the maximum likelihood method. Morphospecies barcodes of macroinvertebrates were also placed for grouping the local fauna into these phylogenetic frameworks for each of the major orders and establishing the sound placement of these lineages with more detailed identifications.

The phylogenetic analysis of the order Diptera using nucleotide sequences of protein-coding genes supported the monophyly of most of the families including Syrphidae, Stratiomyidae, Tabanidae, Limoniidae, Simuliidae and Tipulidae though the lineages formed different subclades under each respective family (Fig. 2.17). For instance, all species of Bangladeshi syrphids were clustered together under two major clades where each clade was subdivided into subclades, but all were placed in a monophyletic clade. The family Empididae was found paraphyletic with Dolichopodidae whereas the family Hybotidae represented a polyphyletic lineage. The non-biting midges (Chironomidae) were found as the most complex group with polyphyly where most of the Bangladeshi species clustered together forming several subclades. Some lineages of Chironomidae formed some distant clades closely related to Dixidae and Ceratopogonidae. Further, the family Ceratopogonidae was also found as a polyphyletic group. The phylogenetic tree also supported the sound placement of barcode sequences with reconfirmation of identified taxa (e.g., Syrphidae 335, Syrphidae 335, Tabanidae 320, Ephydridae 333, Stratiomyidae and Simuliidae, Fig. 2.17) and also provided further deep level identification of different lineages (e.g., Culicidae 338, Culicidae 337, Ceratopogonidae 367 and Empididae 654, Fig. 2.17) at subfamily, genus or even species level.

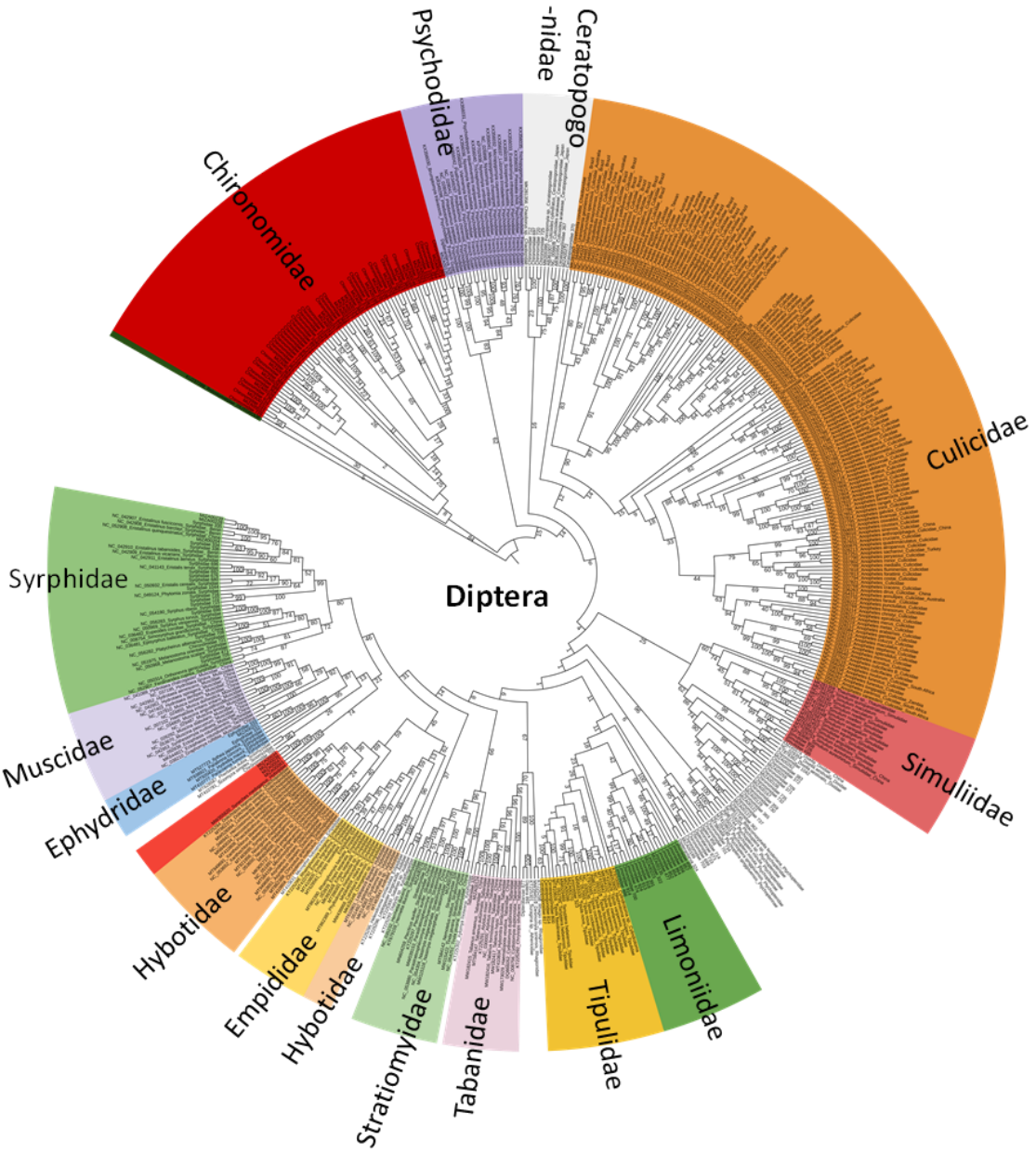

Figure 2. 17 Mitochondrial genomes (explored by this study and also downloaded from Genbank) based maximum likelihood tree for dipteran species.Tree was constructed with nucleotide sequences of protein-coding genes and COI barcodes of morphospecies of this study. The labels of explored mitogenomes are started with 'MIZA' and the barcode sequences are labelled with respective family name following a number.

The odonates tree clearly showed two large clades for Anisoptera and Zygoptera. Mitogenome (protein-coding genes) and barcode sequences used in this study supported the monophyly of all families (Fig. 2.18). Among the three distinct clades of damselfly families (Calopterygidae, Euphaeidae and Coenagrionidae and Euphaeidae), Calopterygidae and Euphaeidae lineages formed sister clades. The phylogenetic tree also revealed the monophyly of all families of dragonflies where Ashnidae and Gomphidae were found as sister clades and

Cordulidae and Macromiidae formed a pair of sister clades. All Bangladeshi libellulid dragonflies formed a large clade including several subclades representing different genera. The placement of barcode sequences of odonates in the phylogenetic tree reconfirmed the identification (e.g., the sequences Libellulidae 205, Coenagrionidae 237, Coenagrionidae 222, and Gomphidae 191) of many lineages and also provided additional information on deep-level identification for many lineages. For instance, the sequences (MIZA00216, Macromiidae 176, Macromiidae 163, Macromiidae 179 in Fig. 2.18) were confirmed at the family level while the phylogenetic tree established their genus (*Macromia*) level identification.

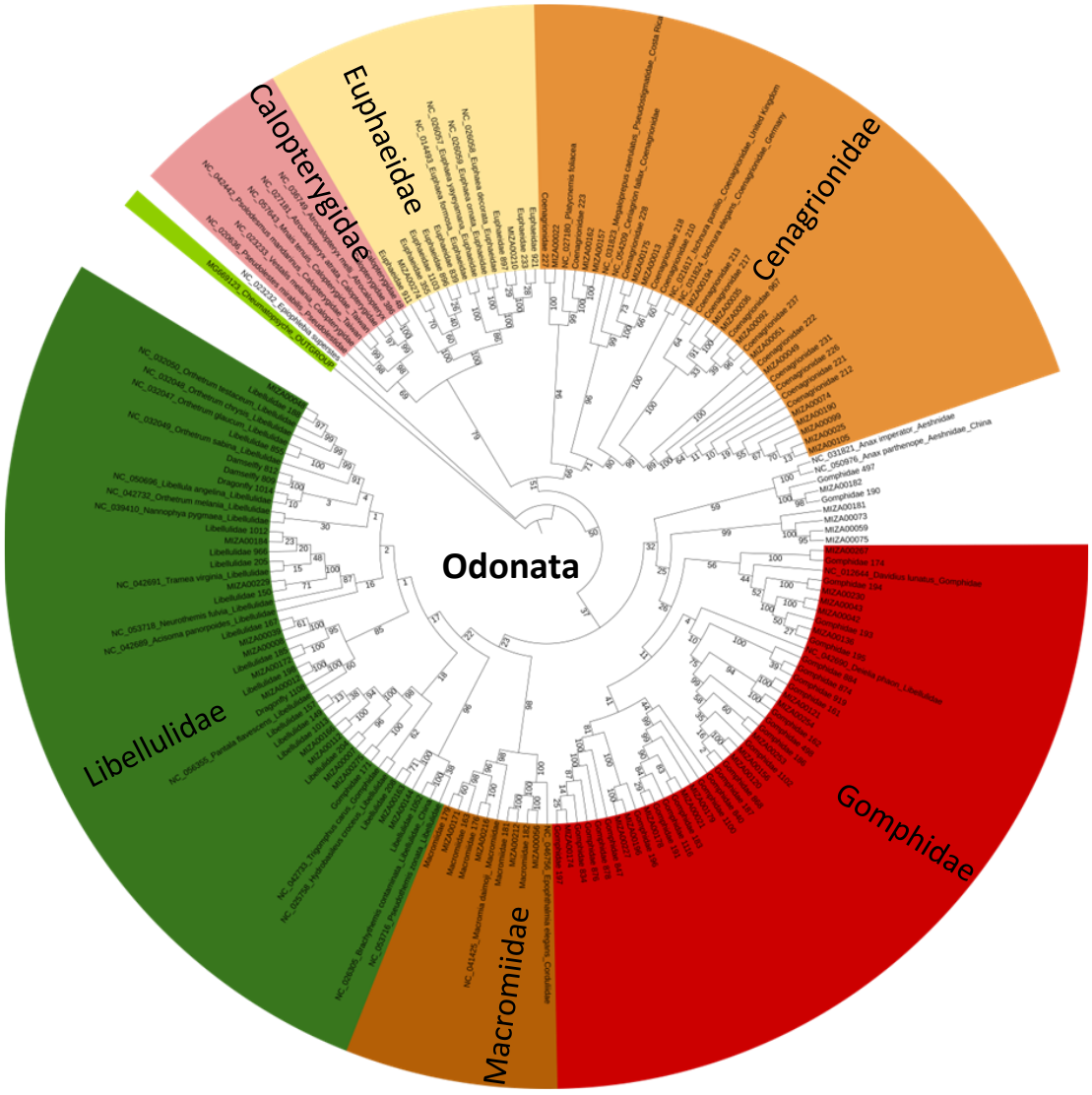

Figure 2. 18 Mitochondrial genomes-based maximum likelihood tree for odonate species. Tree was constructed with nucleotide sequences of protein-coding genes and COI barcodes of morphospecies explored by this study. . The labels of explored mitogenomes are started with 'MIZA' and the barcode sequences are labelled with respective family name following a number.

The phylogenetic tree of Ephemeroptera revealed the monophyly of 4 major families (Baetidae, Caenidae, Leptophlebiidae and Heptageniidae) (Fig. 2.19). The family Ephemeridae was found monophyletic, and a distant clade was formed by Bangladeshi ephemerid sequences that were closely related to Caenidae. In each family, a number of subclades also indicated phylogenetically divergent species of mayflies. The tree also showed the family Baetidae as a basal clade whereas Hepatgeniidae was recognised as the most terminal clade in this study. Furthermore, the tree assigned a good number of barcode sequences (e.g., Baetidae 476, Baetidae 1059, Caenidae 760, Caenidae 339, Leptophlebiidae 823, Leptophlebiidae 359, Heptageniidae 754 and Heptageniidae 751) to their lower/deep (genus) level identification that was unlikely by BLAST search.

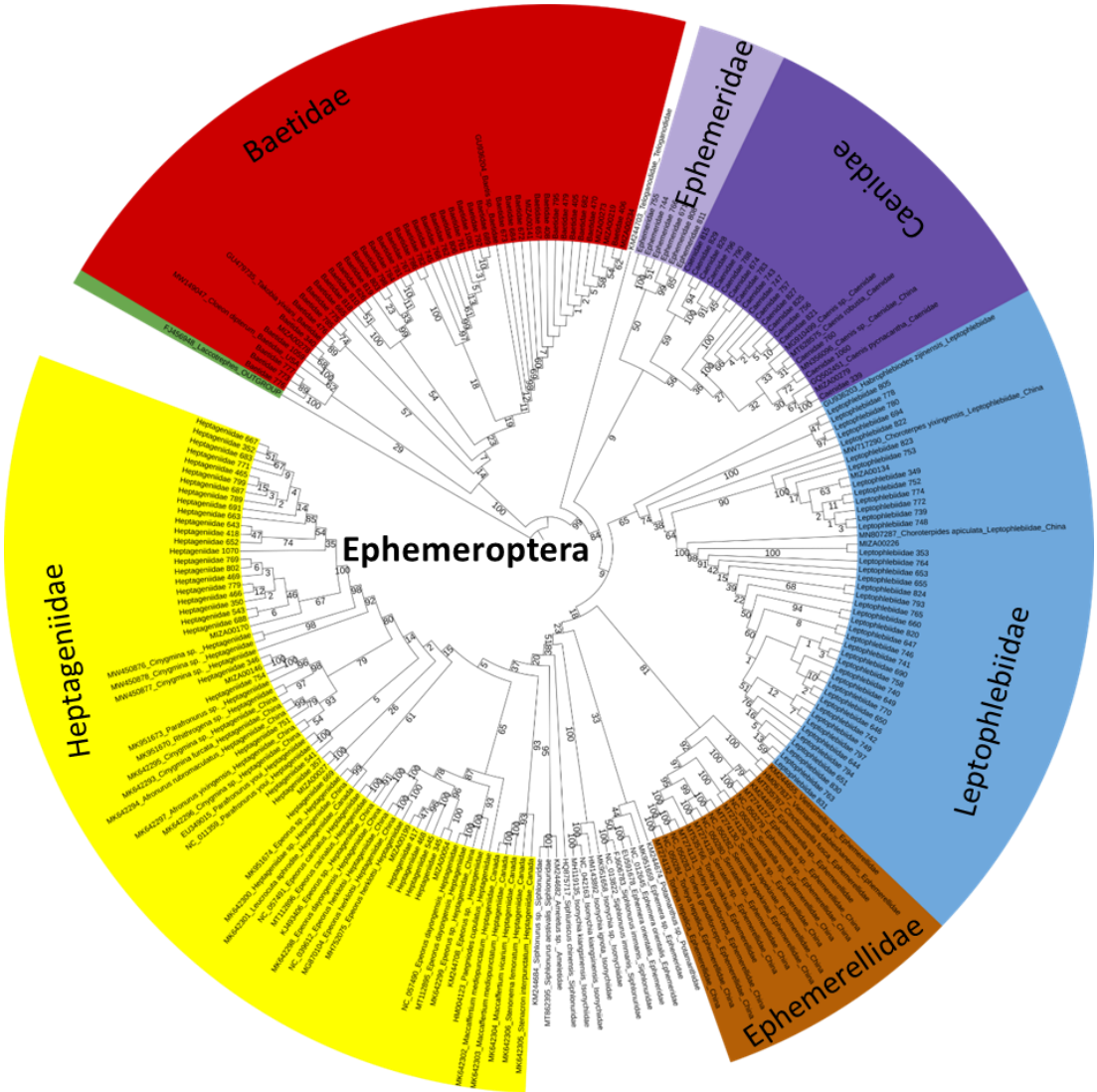

Figure 2. 19 Mitochondrial genomes-based maximum likelihood tree for ephemeropteran species. Tree was constructed with nucleotide sequences of protein-coding genes and COI barcodes of morphospecies explored by this study. The labels of explored mitogenomes are started with 'MIZA' and the barcode sequences are labelled with respective family name following a number.

The phylogenetic tree constructed with mitogenome sequences supported the monophyly of all families of Trichoptera where some families (e.g., Psychmyiidae, Polycentropodidae and Glossosomatidae) were placed in a single clade. Notably, most of the Bangladeshi families formed separate clades which were due to the absence of adequate sequences in the database or indicative of distantly related groups. The Hydropsychidae were found to be phylogenetically diverse and consisted of two large subclades.

The families Polycentropodidae, Psychomyiidae and Stenpsychidae formed a clade indicating their close evolutionary relationship. The phylogenetic tree confirmed the identification of the barcode sequences (e.g., Hydroptilidae 804, Hydropsychidae 529, Stenopsychidae 552, Philopotamidae 572, Stenopsychidae 342 etc. in Fig. 2.20). The phylogenetic framework also resolved the coarse identification of some barcode sequences (e.g., Hydropsychidae 617, Glossosomatidae 558, Glossosomatidae 559, Glossosomatidae 609, Polycentropodidae 551, and Psychomyiidae 558) placing them into lower rank (species or genus) (Fig. 2.20). Likewise, the phylogenetic tree constructed with mitogenomes and barcode sequences of Hemiptera was also found effective for revealing their evolutionary relationship and also for the lower-level taxonomic assignment of barcode sequences (Fig. 2.8 in appendix).

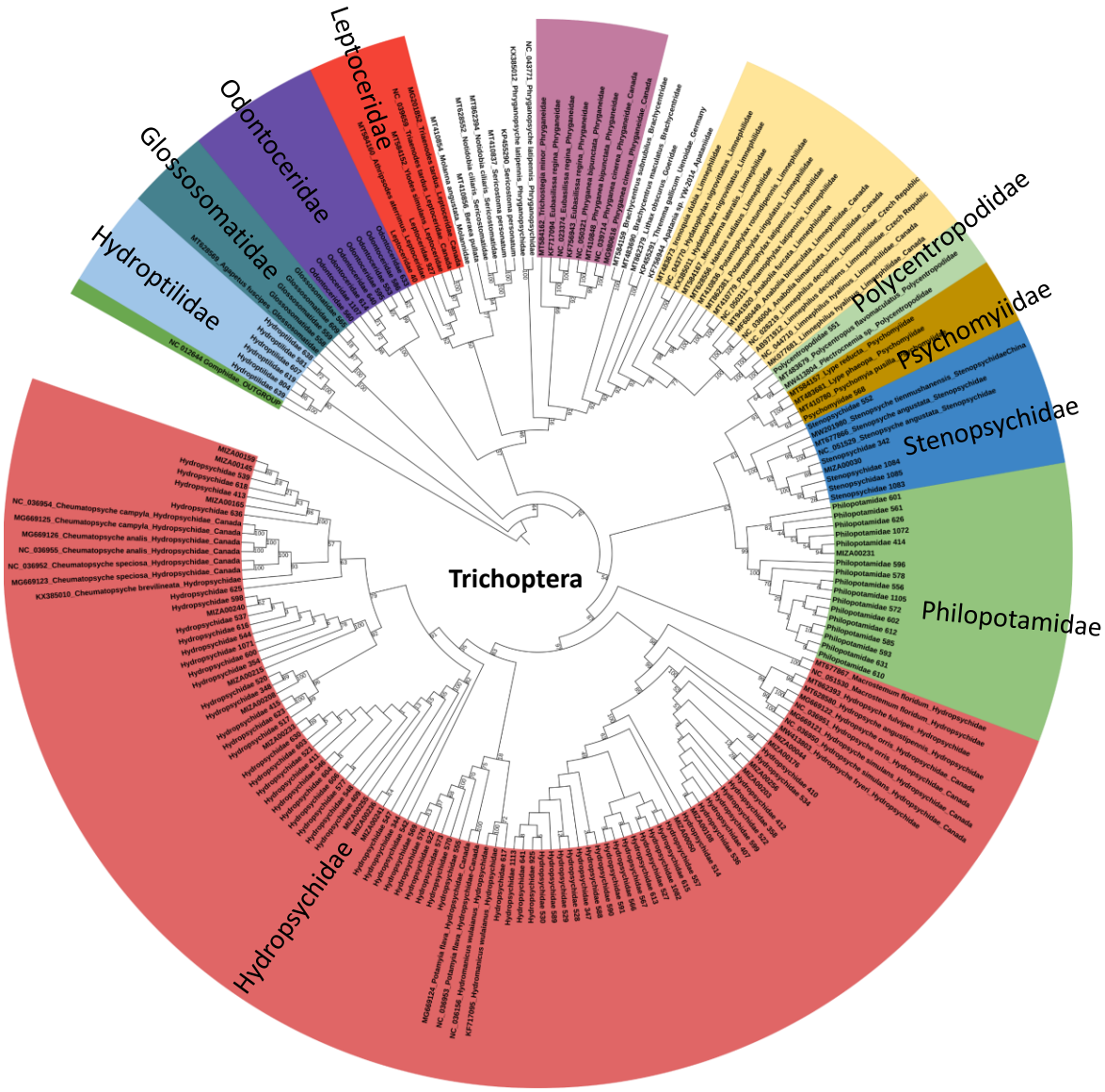

Figure 2. 20 Mitochondrial genomes-based maximum likelihood tree for trichopteran species. Tree was constructed with nucleotide sequences of protein-coding genes and COI barcodes of morphospecies explored by this study. The labels of explored mitogenomes are started with 'MIZA' and the barcode sequences are labelled with respective family name following a number.

2.5 Discussion

In this study, I reported the construction of a DNA barcode reference library with 812 barcodes for 320 Bangladeshi freshwater macroinvertebrates species covering different groups of bioindicators belonging to insects (Ephemeroptera, Plecoptera, Trichoptera, Coleoptera, Hemiptera, Odonata, Diptera), snails and mussels (Gastropoda and Bivalvia), and crabs and shrimps (Crustacea). So far this is a first-time initiative to build a database for macroinvertebrates in the country based on their molecular characterization with COI barcodes which is widely accepted for the identification of animal species. To date, DNA based methods have been hardly applied for the assessment of invertebrate diversity in Bangladesh which are still limited to DNA barcoding of selective insect species (Aslam et al., 2019; Mazumdar et al., 2019; Ghosh et al., 2019; Leblanc et al., 2019). In contrast, this study produced barcode sequences for the species of seven Orders of aquatic insects along with decapods and molluscs, which hinted at a diverse macroinvertebrate fauna, in particular, the diversity of sensitive and tolerant indicator organisms in freshwater ecosystems of Bangladesh. Notably, there was no record for mayflies, caddisflies and stonefly species before this study, whereas at least 70 species were included here for these three sensitive bioindicator groups. Although this study produced short barcodes of COI gene using the HTS platform, they can be used as a reference dataset in the identification of bioindicator fauna, particularly for large-scale biomonitoring of freshwater ecosystems.

This study included two main biotas sampling lowland and mountain rivers, but it did not estimate the coverage rate for the country's macroinvertebrate fauna. However, species accumulation curves (SACs) constructed with species richness indicated that SACs only just began to plateau (Fig. 2.9 in the appendix) given the sampling efforts made in this study. Although this finding provides a first estimate of the species richness of the study area, it also indicates the greater diversity of macroinvertebrates that remains to be discovered from other rivers, streams and wetlands of the country. The assignment of macroinvertebrate barcodes to lower-level taxa was also found challenging as more than half of barcodes did not match any NCBI/BOLD entries within a threshold level of 3%. This can be explained either by the endemism of the macroinvertebrate fauna or by the inadequacy of sequences in existing databases. As expected, species-level identified barcodes were found to match with adjacent countries. However, voucher specimens of most of the morphospecies have been retained to

facilitate the detailed Linnean classification. Barcode sequences of all taxa have also been submitted to BOLD (Barcode of Life Database: Project names-BDCOL, BDDIP, BDHEM, BDINV, BDMOL, BDODO, BDPLE, BDTRI).

DNA barcoding has been popular as a successful molecular identification tool for invertebrates such as insects (Kumar et al., 2018; Weigand et al., 2019; Ge et al., 2021). Also, our study shows that DNA barcoding with the COI gene has great potential for the identification of a wide range of macroinvertebrates though some barcodes of each taxonomic group (e.g., Coleoptera, Chironomidae and Mollusca) remain ambiguous for closely related species which is also supported by other studies (Versteirt et al., 2015; Ge et al., 2021). Even though the utility of the COI gene fragment in species delineation has now been proven in various applications, several potential pitfalls have been identified that cause the absence of a barcoding gap and misidentifications (Ermakov et al., 2015). These drawbacks include the introgression of mitochondrial DNA (mtDNA) due to hybridization and incomplete lineage sorting of mitochondrial haplotypes (Lukhtanov et al., 2009; Whitworth, et al., 2007). Moreover, non-functional copies of mtDNA within the nuclear DNA, NUMTs are sometimes co-amplified by universal primers (Hawlitschek et al., 2017; Hebert, et al 2004; Moulton, Song, and Whiting, 2010). Despite such complications of DNA barcoding, it has been a highly effective tool for species identification in many large-scale studies.

For species delimitation using barcodes of morphospecies, I used three methods including distance-based, Usearch (Edgar, 2010) clustering at a 3 % threshold, phylogenetic tree based bPTP (Zhang et al., 2013) and RESL (Refined Single Linkage) method on the BOLD platform. As these three methods are based on different principles, their delimited number of species was incongruent to some extent for some taxa. In particular, the bPTP (Poisson Tree Processes) model using Bayesian Support (BS) values generally suggested a higher number of delimited species those from Usearch clustering and RESL. However, the average number of bPTP processes was mostly similar to species delimited with Usearch. For instance, bPTP methods identified a higher number of species of Leptophlebiidae than the other two methods though these differences were limited to the delimitation of a few species. The “Poisson Tree Process” (PTP) identifies species status based on the distribution of branch lengths and the assumption of reciprocal monophyly in the gene tree (Zhang et al., 2013). The bPTP approach can process large datasets with thousands of species using a rooted non-ultrametric tree, unlike the GMYC method. The approach is expected to work best for identifying species

that are separated by long intervals between speciation events and that have small population sizes (Rannala and Yang, 2020). In contrast, the Usearch clustering method directly calculates the distances among nucleotide sequences strictly at a 3 % threshold that may trigger delimiting a lower number of species than bPTP. RESL is another clustering approach that generates initial OTU boundaries based on single linkage clustering and evaluates opportunities for refinement of OTU boundaries using Markov clustering. Finally, the optimal partitions for OTUs are made based on a cluster validation method using the Silhouette index (Ratnasingham and Hebert, 2013; Rousseeuw, 1987). In this study, all barcodes were produced from morphospecies identified under many families of macroinvertebrates, of which several families of insects had few barcodes, which was one of the limitations in this study using the bPTP species delimitation process. However, species delimitation of those families (containing a lower number of barcodes) was resolved by Cluster_otu of Usearch at a 3% threshold and RESL method. Finally, barcodes confirmed by both of these approaches were accounted for species in this study which might underestimate the total number of species to a lesser extent. However, there is no stand-alone comprehensive method for species delimitation using genomic data. Therefore, several methods of different principles used in the present study have provided a foundation of species delimitation process for Bangladeshi macroinvertebrate species.

Interspecific genetic distances were measured under each family by two models (p-distance and K2P distance) which showed variation among families of each macroinvertebrate order. For all families, K2P distances were higher than p-distance which resulted from their different calculation strategies. For instance, p-distance is calculated by dividing the number of nucleotide differences by the total number of nucleotides compared, which does not make any correction for multiple substitutions at the same site, while the K2P parameter model corrects for multiple hits, distinguishing transition and transversion substitution rates (Nei and Kumar, 2000). For the Ephemeroptera, the lowest distance was estimated for the family Leptophlebiidae identified with few species from a good number of barcodes. This result prioritises the earlier conservative measures to protect the Leptophlebiid species of mayflies. Likewise, for caddisflies, the species of Odontoceridae were more closely related than others. It was also remarkable that only the family Hydropsychidae contained a good number of species while other families included only a few species and their interspecific genetic distances were high. From the highly sensitive bioindicator group, Plecoptera, only a single family with few species was recorded which also hinted at the critical status of the river

ecosystems. Among aquatic beetles, the lowest interspecific distance was calculated for the Gyrinidae family containing only a few species. For other families of different orders the estimations of genetic distances are illustrated in the Results section which can be used in setting family-wise conservative measures to distinguish freshwater macroinvertebrates.

According to the principle of species identification using DNA barcodes, the interspecies genetic distance is much greater than that of intraspecies. Generally, 2% is the threshold value for species delimitation and, the interspecies genetic distance is over 10 times of intraspecies genetic distance (Hebert et al., 2003; Ward, 2009). In this study, the interspecific K2P distance estimated for most of the families was 15 to 30-fold higher than that of intraspecific distance. Using the geographically confined set of species suggests that the freshwater invertebrates of Bangladesh are composed of fairly distantly related components, rather than local radiations, perhaps indicating an origin in different regions. Careful comparisons with their respective sister groups based on the phylogenetic trees will illuminate these topics once the taxon sampling is more complete. Therefore, the present study provides a reference of macroinvertebrate barcodes with required criteria for effective identification and bioassessment in river and stream ecosystems. Furthermore, it has implications for the application of different metagenetic techniques such as environmental DNA barcoding (Baird and Hajibabaei, 2012) and metabarcoding and metagenomics.

In general, mitochondrial genomes or complete protein-coding genes are more powerful than partial or complete barcodes or even a full-length COI gene for the identification of species and also for a robust phylogenetic resolution. Therefore, a comprehensive mitogenome reference library is a pivotal component to achieve the full potential of metabarcoding for ecosystem assessment (Elbrecht et al., 2017). In this study, I also extracted mitogenomes from freshwater macroinvertebrates using MMG which is an undoubtedly robust approach for mitogenomes extraction from mixed samples. It relies on some sensitive issues like library preparation with quality DNA, quality sequences from NGS, filtering of putative mitochondrial sequences, assembly of putative reads, annotation of genes and finally the assignment of mitogenomes to accurate taxa with full-length sequences. In particular, the assembly of the total volume of reads is a computationally challenging step for complex samples. Due to time limitations and cost of the techniques, primarily 287 morphospecies belonging to different groups of macroinvertebrates were taken in this study to extract their

complete mitogenomes that produced 108 complete mitogenomes (consisting of all 13 protein-coding genes, 22 tRNA and 2 rRNA genes) and 89 partial contigs (consisted of single to multiple protein-coding and non-coding genes). Although the extraction rate of complete mitogenomes was less than 50%, the overall percentage was over 68 % including partial genomes. The total number of mitogenomes obtained from this pipeline was also affected by the generation of multiple contigs of the same species or same morphospecies. This result suggested for the inclusion of distantly related species in the same library for shotgun sequencing that may increase the productivity of this pipeline in terms of the higher number of complete and partial mitogenomes. Another challenging step was to assemble shorter contigs to full contigs using powerful assembly software. Although a number of different assemblers are available, a rigorously performing assembler on a variety of MMG datasets is still desirable. Therefore, the combination of different assemblers such as IDBA-UD, Celera Assembler and Newbler (Crampton-Platt et al., 2016) was used to assemble long mitogenome sequences that also showed variation in their outputs. The gene annotation with the MitoZ tool was found rapid and effective but generated some disparity for the placement of start and stop codons in many genes which required manual editions for almost all contigs. However, the efficacy of the gene annotation methods should be tracked using other annotation software such as MITOS (Bernt et al., 2013) or a reference sequence-based annotation pipeline (Zhou et al, 2013). The retrieval of sequence information from each MMG library is another crucial step for relating back to specific individuals in the pool of samples. Therefore, I used respective morphospecies barcodes as bait sequences to associate individual contigs to particular species in the pool.

So far, a total of 108 complete mitogenomes for different groups of macroinvertebrates were extracted in this study, which has expanded the genomic resources of freshwater macroinvertebrates of Bangladesh. These genomic resources can be used for phylogenetic and taxonomic studies as well as for understanding biogeographic affinities with macroinvertebrates elsewhere in the world. Biodiversity assessment and conservation of aquatic resources is often hampered by our limited knowledge of genetic diversity. The complete mitogenome exploration can address this challenge by offering muti-genes information and extending the scope to use single, multiple or all genes required in a specific study. For instance, the success of environmental metabarcoding or eDNA metabarcoding relies on the availability of a reference database for a particular gene, in this context,

mitogenomes provide a set of protein-coding and non-coding genes. Given the incomplete reference database for poorly explored regions such as the various biogeographic regions of Bangladesh, we will usually not find a perfect match to a reference sequence but instead a robust identification can be made against a well-annotated phylogenetic tree that places the short metabarcode sequence in the context of the phylogenetically more informative mitogenomes.

I reconstructed mitogenome-based phylogenetic trees with nucleotide sequences of protein-coding genes to acquire maximum resolution for the evolutionary relationship of aquatic insect species of the Order Diptera, Ephemeroptera, Trichoptera, Odonata, and Hemiptera. This result supported the potential of the mitogenome-based (all protein-coding genes) phylogenetic tree revealing the monophyly, paraphyly and polyphyly of families belonging to respective orders. The formation of branches and clades of the families of Diptera were corroborated by other studies (e.g., Cameron et al., 2007) where Syrphidae, Muscidae, Ephydridae and Empidae were grouped as different subclades and Tabanidae was their sister clade. In our analyses, some taxa exist as long branches, that were problematic for accurate estimation of the phylogenetic relationships. For instance, some lineages of Chironomidae were incorrectly placed with Simulidae in the phylogenetic tree. Due to long-branch attraction (LBA) artefacts, unrelated species can be grouped incorrectly which is a common phenomenon occurring in tree reconstructions (Hendy and Penny, 1989; Bergsten, 2005). In the Odonata tree, the families Macrmiidae and Cordulidae were found as sister clades and those are placed with Libellulidae. This finding was similarly revealed by another study concerning the molecular phylogenetic study of a few families of Anisoptera (Carle et al., 2015). As this study produced a limited number of mitogenome sequences, I included the barcode sequences of families in order tree that returned reliable results, for example, the barcode sequences of the Trichopteran families Odontoceridae and Philopotamidae placed in the right clades that were supported by Thomas et al. (2020). In addition, many barcode sequences of each insect order attained a deeper level of identification having the sound placement with available mitogenome sequences in the phylogenetic tree.

In this study, I used only nucleotide sequences for tree construction, though amino acid data are also used to generate a phylogeny and there is controversy about which is best (Simmons et al., 2002). The main argument for using amino acid sequence is that amino acids have

more possible character states than nucleotides (20 versus 4). However, nucleotide sequences can lead to better resolution of the tree due to an increased number of characters (Jill and Langdale, 2006). Therefore, inference of phylogeny using both nucleotide and amino acid sequences could be a more robust approach to obtain the best phylogenetic tree. Another limitation was that the mitogenome sequences and barcode sequences explored were preliminarily assigned to the family level only. Therefore, the tree did not depict the subfamily or genus level relationship for the species of all orders. One of the key challenges of the mitochondrial genome-based phylogenetic analysis is the paucity of complete mitogenomes of macroinvertebrates in existing reference databases. Even though, mitogenome sequences of some families (e.g., Philopotamidae, Odontoceridae and Hydroptilidae of Trichoptera) of macroinvertebrates are completely lacking in the database. Therefore, complete mitogenome generation should have priority for building a comprehensive database of mitogenomes to resolve the phylogeny of local taxa. Furthermore, sequences from the local individual (barcoding) or bulk samples (metabarcoding) can be compared against mitogenomes and may reveal more reliable biodiversity patterns with ecological and phylogenetic information.

2.6 References

Ahad, M. A., Raja, S. I., Khatun M. M. and Hossain, M. A., 2012. Aquatic Insect Biodiversity of Northern Region of Bangladesh. *International Journal of Bio-resource and Stress Management*, *3*(3), pp.387-393.

Ahmed, A.T.A. and Rahman, M.M., 2013. Heavy metal pollution and their bioaccumulation by fish and shellfish population in some rivers of Bangladesh. A Research Report. Department of Zoology, University of Dhaka, Dhaka.105 pp.

Ahmed, M.K., Baki, M.A., Islam, M.S., Kundu, G.K., Habibullah-Al-Mamun, M., Sarkar, S.K. and Hossain, M.M., 2015. Human health risk assessment of heavy metals in tropical fish and shellfish collected from the river Buriganga, Bangladesh. *Environmental Science and Pollution Research*, *22*(20), pp.15880-15890.

Ahsan M. A., Siddique, M. A. B., Munni, M. A., Akbor, M.A., Akter S., and Mia, M. Y., 2018. Analysis of Physicochemical Parameters, Anions and Major Heavy Metals of the Dhaleshwari River Water, Tangail, Bangladesh. *American Journal of Environmental Protection*, *7*(2), pp.29-39.

Ahmed, M.S., Chowdhury, M.M.K. and Nahar, L., 2019. Molecular characterization of small indigenous fish species (SIS) of Bangladesh through DNA barcodes. *Gene*, *684*, pp.53-57.

Ahmed, M.S., Datta, S.K., Saha, T. and Hossain, Z., 2021. Molecular characterization of marine and coastal fishes of Bangladesh through DNA barcodes. *Ecology and Evolution*, 11(9), pp.3696-3709.

Ali, S., AS.M. Haque, J.R. Oppenheimer and KM.S. Aziz. 1978. Studies on the bottom fauna of three fish ponds in Dhaka City, Bangladesh. *Bangladesh J. Zool. 6*(1), pp.43-55.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. *J Mol Biol*, 215(3), pp.403-10.

Andujar, C., Arribas P., Ruzicka F., Crampton-platt, A., Timmermans, M.J.T.N. and Vogler, A.P., 2015. Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology*, 24, pp.3603–3617.

Armitage, P.D., Moss, D., Wright, J.F. and Furse, M.T., 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research*, *17*(3), pp.333-347.

Aslam, A.F.M., Sultana, S., Rain, F.F., Das, S.R., Siddika, A. and Howlader, A.J., 2019. Molecular characterization and identification of three stored grain pests based on mitochondrial cytochrome c oxidase subunit I (COI) gene sequences. *Bangladesh Journal of Zoology*, *47*(1), pp.1-11.

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M. and Vogler, A.P., 2016. Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology and Evolution*, 7(9), pp.1071-1081.

Baird, D. J. and Hajibabaei, M., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, *21*, pp. 2039–2044.

Bergsten, J., 2005. A review of long-branch attraction. *Cladistics*, 21(2), pp.163-193.

Baird, D.J. and Sweeney, B.W., 2011. Applying DNA barcoding in benthology: the state of the science. *Journal of the North American Benthological Society*, *30*(1), pp.122-124.

Baird, D.J., Pascoe, T.J., Zhou, X. and Hajibabaei, M., 2011. Building freshwater macroinvertebrate DNA-barcode libraries from reference collection material: formalin preservation vs specimen age. *Journal of the North American Benthological Society*, *30*(1), pp.125-130.

Ball, S.L., Hebert, P.D., Burian, S.K. and Webb, J.M., 2005. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society*, *24*(3), pp.508-524.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D. and Pyshkin, A.V., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), pp.455-477.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling., 1999. *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish*, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.

Barber-James, H.M., Gattolliat, J.L., Sartori, M. and Hubbard, M.D., 2007. Global diversity of mayflies (Ephemeroptera, Insecta) in freshwater. *Hydrobiologia*, *595*, pp.339-350.

Bashar, K. Reza, M. S., Razzak, M. A., Rahman, K.M. Z., Goda, P. and Howlader A. J., 2014. Faunistic study of Odonata (Dragonfly & Damselfly) in some selected regions of Bangladesh. *Journal of Entomology and Zoology Studies*, 2 (4) pp. 1-6.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., et al., 2013. GenBank. *Nucleic Acids Research*, *41*, pp.36–42.

Benke, A.C., Van Arsdall Jr, T.C., Gillespie, D.M. and Parrish, F.K., 1984. Invertebrate productivity in a subtropical blackwater river: the importance of habitat and life history. *Ecological Monographs*, *54*(1), pp.25-63.

Bickel, D., Pape, T. and Meier, R. eds., 2009. *Diptera diversity: status, challenges and tools.* Leiden, Netherlands: Brill. pp. 459.

Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114-2120.

Cameron, S.L., Lambkin, C.L., Barker, S.C. and Whiting, M.F., 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Systematic Entomology*, 32(1), pp.40-59.

Cardoni, S., Tenchini, R., Ficulle, I., Piredda, R., Simeone, M.C. and Belfiore, C., 2015. DNA barcode assessment of Mediterranean mayflies (Ephemeroptera), benchmark data for a regional reference library for rapid biomonitoring of freshwaters. *Biochemical Systematics and Ecology*, *62*, pp.36-50.

Carle, F.L., Kjer, K.M. and May, M.L., 2015. A molecular phylogeny and classification of Anisoptera (Odonata). *Arthropod Systematics & Phylogeny*, 73(2), pp.281-301.

Carew, M.E., Pettigrove, V., Cox, R.L. and Hoffmann, A.A., 2007. DNA identification of urban Tanytarsini chironomids (Diptera: Chironomidae). *Journal of the North American Benthological Society*, *26*(4), pp.587-600.

Carew, M.E., Pettigrove, V.J., Metzeling, L. and Hoffmann, A.A., 2013. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in Zoology*, *10*(1), pp.1-15.

Carew, M.E. and Hoffmann, A.A., 2015. Delineating closely related species with DNA barcodes for routine biological monitoring. *Freshwater Biology*, *60*(8), pp.1545-1560.

Casas, P.A.S., Sing, K.W., Lee, P.S., Nuñeza, O.M., Villanueva, R.J.T. and Wilson, J.J., 2018. DNA barcodes for dragonflies and damselflies (Odonata) of Mindanao, Philippines. *Mitochondrial DNA Part A*, 29(2), pp.206-211.

Cavender-Bares, J.,Kozak, K.H., Fine, P.V.A. and Kembel, S.W., 2009. The merging of community ecology and phylogenetic biology. *Ecology Letter*, 12: 693–715.

Chen, J., Li, Q., Kong, L. and Yu, H., 2011. How DNA barcodes complement taxonomy and explore species diversity: the case study of a poorly understood marine fauna. *PloS One*, 6(6), p.e21326.

Chowdhuri, S. H. And Aktaruzzaman, M.,1981. Dragonfly (Odonata:Anisoptera) Larvae from Chittagong. *Bangladesh J. Zool*., 9(2), pp.131-144.

Chowdhury, S.H. and Mohiuddin, M., 2011. A checklist of the Odonata from the eastern region of Bangladesh with some taxonomic notes. *University Journal of Zoology*, Rajshahi University, 30, pp.61-66.

Collins, A., Ohandja, D.G., Hoare, D. and Voulvoulis, N., 2012. Implementing the Water Framework Directive: a transition from established monitoring networks in England and Wales. *Environmental Science & Policy*, *17*, pp.49-61.

Cordero, R. D., Sánchez-Ramírez, S. and Currie, D.C., 2017. DNA barcoding of aquatic insects reveals unforeseen diversity and recurrent population divergence patterns through broad-scale sampling in northern Canada. *Polar Biology*, *40*(8), pp.1687-1695.

Crampton-Platt, A., Yu, D.W., Zhou, X. and Vogler, A.P., 2016. Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience*, *5*(1), pp.s13742-016.

Crampton-Platt, A., Timmermans, M.J., Gimmel, M.L., Kutty, S.N., Cockerill, T.D., Vun Khen, C. and Vogler, A.P., 2015. Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. *Molecular Biology and Evolution, 32*(9), pp.2302-2316.

Daily, G. C. and Ellison, K., 2012. The new economy of nature: the quest to make conservation profitable. Island Press.

Dettai, A., Gallut, C., Brouillet, S., Pothier, J., Lecointre, G. and Debruyne, R., 2012. Conveniently pre-tagged and pre-packaged: extended molecular identification and metagenomics using complete metazoan mitochondrial genomes. *PloS one*, *7*(12), p.e51263.

De Moor, F.C. and Ivanov, V.D., 2007. Global diversity of caddisflies (Trichoptera: Insecta) in freshwater. *Hydrobiologia, 595,* pp. 393-407.

Dobson, M., Pawley, S., Fletcher M., and Powell, A., 2012. *Guide to Freshwater Invertebrates*. Freshwater Biological Association, UK. Scientific Publication No 68. 216 pp.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res*. 32(5), pp.1792-1797.

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, pp.2460–2461.

Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*, p.081257.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V., Mallet, J. and Jiggins, C.D., 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1627), pp.2881-2889.

Ekrem, T., Willassen, E. and Stur, E., 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution*, *43*(2), pp.530-542.

Ermakov, O.A., Simonov, E., Surin, V.L., Titov, S.V., Brandler, O.V., Ivanova, N.V. and Borisenko, A.V., 2015. Implications of hybridization, NUMTs, and overlooked diversity for DNA barcoding of Eurasian ground squirrels. *PLoS One*, *10*(1), p.e0117201.

Erasmus, D.J., Yurkowski, E.A. and Huber, D.P., 2018. DNA barcode-based survey of Trichoptera in the Crooked River reveals three new species records for British Columbia. *PeerJ*, *6*, p.e4221.

Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol*., 3 (5), pp.294–299.

Fujisawa, T. and Barraclough, T.G., 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology*, *62*(5), pp.707-724.

Fochetti, R. and De Figueroa, J.M.T., 2007. Global diversity of stoneflies (Plecoptera; Insecta) in freshwater. *Hydrobiologia*, *595*, pp. 365-377.

Ferrington, L.C., 2007. Global diversity of non-biting midges (Chironomidae; Insecta-Diptera) in freshwater. In *Freshwater animal diversity assessment* (pp. 447-455). Springer, Dordrecht.

Gattolliat, J.L., Vinçon, G., Wyler, S., Pawlowski, J. and Sartori, M., 2016. Toward a comprehensive COI DNA barcode library for Swiss Stoneflies (Insecta: Plecoptera) with special emphasis on the genus Leuctra. *Zoosymposia*,*11*, pp.135-155.

Ge, Y., Xia, C., Wang, J., Zhang, X., Ma, X. and Zhou, Q., 2021. The efficacy of DNA barcoding in the classification, genetic differentiation, and biodiversity assessment of benthic macroinvertebrates. *Ecology and Evolution*, *11*(10), pp.5669-5681.

Geraci, C.J., Al-Saffar, M.A. and Zhou, X., 2011. DNA barcoding facilitates description of unknown faunas: a case study on Trichoptera in the headwaters of the Tigris River, Iraq. *Journal of the North American Benthological Society*, *30*(1), pp.163-173.

Gillett, C.P.D.T., Crampton-Platt, A., Timmermans, M.J.T.N. Jordal, B. H., Emerson, B. C., and Vogler, A. P., 2014. Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: curculionoidea). *Molecular Biology and Evolution*, 31, pp.2223–2237.

Ghosh, A., Abedin, M.S., Howlader, A.J. and Hossain, M.M., 2019. Molecular identification and phylogenetic relationships of seven Satyrinae butterflies in Bangladesh using Cytochrome c oxidase subunit I gene. *Jahangirnagar University Journal of Biological Sciences*, 8(1), pp.67-74.

Gomez-Rodrıguez, C., Crampton-Platt, A., Timmermans, M.J. T. N., Baselga, A. and Vogler, A.P., 2015. Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex Assemblages. *Methods in Ecology and Evolution*, 6, pp.883–894.

Gwiazdowski, R.A., Foottit, R.G., Maw, H.E.L. and Hebert, P.D., 2015. The Hemiptera (Insecta) of Canada: constructing a reference library of DNA barcodes. *PLoS One*, *10*(4), p.e0125635.

Habib, K.A., Neogi, A.K., Rahman, M., Oh, J., Lee, Y.H. and Kim, C.G., 2021. DNA barcoding of brackish and marine water fishes and shellfishes of Sundarbans, the world's largest mangrove ecosystem. *Plos One*, 16(8), p.e0255110.

Hart, L.A., Bowker, M.B., Tarboton, W. and Downs, C.T., 2014. Species composition, distribution and habitat types of Odonata in the iSimangaliso Wetland Park, KwaZulu-Natal, South Africa and the associated conservation implications. *PLoS One*, *9*(3), p.e92588.

Hawlitschek, O., Morinière, J., Lehmann, G.U.C., Lehmann, A.W., Kropf, M., Dunz, A., Glaw, F., Detcharoen, M., Schmidt, S., Hausmann, A. and Szucsich, N.U., 2017. DNA barcoding of crickets, katydids and grasshoppers (Orthoptera) from Central Europe with focus on Austria, Germany and Switzerland. *Molecular Ecology Resources*, *17*(5), pp.1037-1053.

Havemann, N., Gossner, M.M., Hendrich, L., Moriniere, J., Niedringhaus, R., Schäfer, P. and Raupach, M.J., 2018. From water striders to water bugs: the molecular diversity of aquatic Heteroptera (Gerromorpha, Nepomorpha) of Germany based on DNA barcodes. *PeerJ*, 6, p.e4577.

Hebert, P.D., Cywinska, A., Ball, S.L. and DeWaard, J.R., 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), pp.313-321.

Hebert, P.D., Ratnasingham, S. and De Waard, J.R., 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*, pp. S96-S99.

Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H. and Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences*, *101*(41), pp.14812-14817.

Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P.D., Hausmann, A., Köhler, F. and Balke, M., 2015. A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Molecular Ecology Resources*, *15*(4), pp.795-818.

Hendy, M.D. and Penny, D., 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology*, *38*(4), pp.297-309.

Hering, D., Borja, A., Jones, J.I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert, M. and Leese, F., 2018. Implementation options for DNA-based

identification into ecological status assessment under the European Water Framework Directive. *Water Research*, *138*, pp.192-205.

Hering, D., Meier, C., Rawer-Jost, C., Feld, C.K., Biss, R., Zenker, A., Sundermann, A., Lohse, S. and Böhmer, J., 2004. Assessing streams in Germany with benthic invertebrates: selection of candidate metrics. *Limnologica*, 34(4), pp.398-415.

Hoareau, T.B. and Boissin, E., 2010. Design of phylum-specific hybrid primers for DNA barcoding: addressing the need for efficient COI amplification in the Echinodermata. *Molecular Ecology Resources*, *10*(6), pp.960-967.

Hossain, S., Aslam, A.F., Saha, B. and Howlader, A.J., 2015. Abundance of aquatic insects in relation to physico-chemical parameters of two highly polluted Rivers Sitalakkhya and the Buriganga. *Bangladesh J. Zool*., *43*(1), pp.63-72.

Hossain, T. and Rahman, M.H., 2018. A preliminary list of aquatic and semi-aquatic hemiptera (Arthropoda: Insecta) of Chittagong University Campus, Bangladesh. *Journal of Entomology and Zoology Studies, 6*(3), pp. 1010-1014.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.J. and Tappu, R., 2016. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Computational Biology*, *12*(6), p.e1004957.

Jäch, M.A. and Balke, M., 2008. Global diversity of water beetles (Coleoptera) in freshwater. *Hydrobiologia*, 595(1), pp.419-442.

Islam, M.M., Akhtar M.K., Masud M.S., 2006. Prediction of environmental flow to improve the water quality in the river Buriganga. (Proceedings of the 17th IASTED International Conference on Modelling and Simulation, Montreal, QC, Canada).

Islam, M.S., Meghla, N.T., Suravi, M.S. and Islam, M., 2012. Status of water quality in the Dhaleshwari River and its effect on aquatic organism. *Journal of Environmental Science and Water Resources, 1*(8), pp.192-201.

Janzen, D.H., Hallwachs, W., Blandin, P., Burns, J.M., Cadiou, J.M., Chacon, I., Dapkey, T., Deans, A.R., Epstein, M.E., Espinoza, B. and Franclemont, J.G., 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources*, *9*, pp.1-26.

Janzen, D.H., Hajibabaei, M., Burns, J.M., Hallwachs, W., Remigio, E. and Hebert, P.D., 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), pp.1835-1845.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A. and Larsen, T.H., 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters, 16*(10), pp.1245-1257.

Jill Harrison, C. and Langdale, J.A., 2006. A step-by-step guide to phylogeny reconstruction. *The Plant Journal*, *45*(4), pp.561-572.

Jung, S.W., Min, H.K., Kim, Y.H., Choi, H.A., Lee, S.Y., Bae, Y.J. and Paek, W.K., 2016. A DNA barcode library of the beetle reference collection (Insecta: Coleoptera) in the National Science Museum, Korea. *Journal of Asia-Pacific Biodiversity*, *9*(2), pp.234-244.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), pp.772-780.

Kaila, L. and Ståhls, G., 2006. DNA barcodes: Evaluating the potential of COI to diffentiate closely related species of Elachista (Lepidoptera: Gelechioidea: Elachistidae) from Australia. *Zootaxa, 1170*(1), pp.1-26.

Kalkman, V.J., Clausnitzer, V., Dijkstra, K.D.B., Orr, A.G., Paulson, D.R. and van Tol, J., 2007. Global diversity of dragonflies (Odonata) in freshwater. In Freshwater animal diversity assessment (pp. 351-363). Springer, Dordrecht.

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A. and Flouri, T., 2017. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, *33*(11), pp.1630-1638.

Karr, J.R. and Chu, E.W., 1999. Restoring Life in Running Waters: Better Biological Monitoring. Island Press, Washington, DC.

Karthika, P., Vadivalagan, C., Gunasekaran, C. and Anandhakumar, S., 2012. DNA Barcoding of selected dragonfly species (Libellulidae and Aeshnidae) for species authentication with phylogenetic assessment. *European Journal of Experimental Biology*, 2 (6), pp. 2158-2165.

Kenney, M.A., Sutton-Grier, A.E., Smith, R.F. and Gresens, S.E., 2009. Benthic macroinvertebrates as indicators of water quality: The intersection of science and policy. *Terrestrial Arthropod Reviews*, 2(2), p.99.

Kim, S., Song, K.H., Ree, H.I. and Kim, W., 2012. A DNA barcode library for Korean Chironomidae (Insecta: Diptera) and indexes for defining barcode gap. *Molecules and Cells*, *33*(1), pp.9-17.

Kullander, S.O., Rahman, M.M., Noren, M. and Mollah, A.R., 2015. Danio annulosus, a new species of chain Danio from the Shuvolong Falls in Bangladesh (Teleostei: Cyprinidae: Danioninae). *Zootaxa,* 3994(1), pp.53-68.

Kullander, S.O., Rahman, M.M., Norén, M. and Mollah, A.R., 2017. Devario in Bangladesh: Species diversity, sibling species, and introgression within danionin cyprinids (Teleostei: Cyprinidae: Danioninae). *PLoS One,* 12(11), p.e0186895.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), pp.1547-1549.

Leblanc, L., Hossain, M.A., Doorenweerd, C., Khan, S.A., Momen, M., San Jose, M. and Rubinoff, D., 2019. Six years of fruit fly surveys in Bangladesh: a new species, 33 new country records and discovery of the highly invasive Bactrocera carambolae (Diptera, Tephritidae*). ZooKeys*, 876, p.87.

Letunic, I. and Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), pp. W293-W296.

Lenat, D.R., 1993. A biotic index for the southeastern United States: derivation and list of tolerance values, with criteria for assigning water-quality ratings. *Journal of the North American Benthological Society*, *12*(3), pp.279-290.

Lenat, D.R. and Resh, V.H., 2001. Taxonomy and stream ecology—the benefits of genus-and species-level identifications. *Journal of the North American Benthological Society*, *20*(2), pp.287-298.

Li, L., Zheng, B. and Liu, L., 2010. Biomonitoring and bioindicators used for river ecosystems: definitions, approaches and trends. *Procedia Environmental Sciences*, *2*, pp.1510-1524.

Linard, B., Crampton-Platt, A., Gillett, C. P.D.T., Timmermans, M. J.T.N. and Vogler, A. P., 2015. Metagenome Skimming of Insect Specimen Pools: Potential for Comparative Genomics. *Genome Biol. Evol.*, 7(6), pp.1474–1489.

Lohman, D.J., Prawiradilaga, D.M. and Meier, R., 2009. Improved COI barcoding primers for Southeast Asian perching birds (Aves: Passeriformes). *Molecular Ecology Resources*, *9*(1), pp.37-40.

Lukhtanov, V.A. and Kuznetsova, V.G., 2009. Molecular and cytogenetic approaches to species diagnostics, systematics, and phylogenetics. Zhurnal Obshchei Biologii, 70(5), pp.415-437.

Mazumdar, S., Hebert, P.D.N., Bhuiya, B.A. and Miah, M.I., 2019. Parasitic Hymenoptera Recovered by DNA Barcoding of Malaise Trap Collection at the Chittagong University Campus, Bangladesh. *American Journal of BioScience*, *7*(6), pp.94-98.

Meier, R., Shiyang, K., Vaidya, G. and Ng, P.K., 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, *55*(5), pp.715-728.

Meghla, N.T., Islam, M.S., Ali, M.A. and Nargis, S., 2013. Assessment of physicochemical properties of water from the Turag River in Dhaka City, Bangladesh. *International Journal of Current Microbiology and Applied Sciences*, *2*(5), pp.110-122.

Meng, G., Li, Y., Yang, C. and Liu, S., 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, *47*(11), pp.e63-e63.

Miller, M.A., Pfeiffer, W., and Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees in *Proceedings of the Gateway Computing Environments Workshop* (GCE), 14 Nov. 2010, New Orleans, LA, pp 1-8.

Monaghan, M.T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D.J., Lees, D.C., Ranaivosolo, R., Eggleton, P., Barraclough, T.G. and Vogler, A.P., 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology*, *58*(3), pp.298-311.

Moulton, M.J., Song, H. and Whiting, M.F., 2010. Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources, 10*(4), pp.615-627.

Moriniere, J., Hendrich, L., Balke, M., Beermann, A.J., König, T., Hess, M., Koch, S., Müller, R., Leese, F., Hebert, P.D. and Hausmann, A., 2017. A DNA barcode library for Germany′ s mayflies, stoneflies and caddisflies (Ephemeroptera, Plecoptera and Trichoptera). *Molecular Ecology Resources*, *17*(6), pp.1293-1307.

Moniruzzaman, M., Elahib S.F., and Jahangir M.A.A., 2009. Study on temporal variation of physico-chemical parameters of Buriganga River. *Bangladesh J Sci. Indus. Res., 44,* pp.327–334.

Moriniere, J., Balke, M., Doczkal, D., Geiger, M.F., Hardulak, L.A., Haszprunar, G., Hausmann, A., Hendrich, L., Regalado, L., Rulik, B. and Schmidt, S., 2019. A DNA barcode library for 5,200 German flies and midges (Insecta: Diptera) and its implications for metabarcoding-based biomonitoring. *Molecular Ecology Resources*, *19*(4), pp.900-928.

Mustafa, T., Alam, S. M. I., Jamal, M. S., Johora, F.T., Likhon, M. A. K., Hoq, T. and Naser, M. N., 2013.Abundance of benthic fauna in winter and summer seasons at three water bodies of Dhaka, Bangladesh. *Bangladesh J. Zool.,* 41(1), pp.79-86.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A. and Anson, E.L., 2000. A whole-genome assembly of Drosophila. *Science*, *287*(5461), pp.2196-2204.

Nasiruddin, M., Azadi, M. A. and Reza, M. S., 2014. Abundance and diversity of aquatic insects in two water bodies of Chittagong university campus. *Bangladesh J. Zool.* 42(1), pp.19-33.

Nei, M. and Kumar, S., 2000. *Molecular evolution and phylogenetics*. Oxford University Press, USA.

Papadopoulou, A., Taberlet, P. and Zinger, L. 2015. Metagenome skimming for phylogenetic community ecology: a new era in biodiversity research. *Molecular Ecology*, 24, pp.3515–3517.

Peng, Y., Leung, H.C., Yiu, S.M. and Chin, F.Y., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, *28*(11), pp.1420-1428.

Pfenninger, M., Nowak, C., Kley, C., Steinke, D. and Streit, B., 2007. Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic Chironomus (Diptera) species. *Molecular Ecology*, *16*(9), pp.1957-1968.

Pilgrim, E.M., Jackson, S.A., Swenson, S., Turcsanyi, I., Friedman, E., Weigt, L. and Bagley, M.J., 2011. Incorporation of DNA barcoding into a large-scale biomonitoring program: opportunities and pitfalls. *Journal of the North American Benthological Society*, *30*(1), pp.217-231.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. and Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, *55*(4), pp.595-609.

Poulton, B.C., Monda, D.P., Woodward, D.F., Wildhaber, M.L. and Brumbaugh, W.G., 1995. Relations between benthic community structure and metals concentrations in aquatic macroinvertebrates: Clark Fork River, Montana. *Journal of Freshwater Ecology*, *10*(3): 277-293.

Puillandre, N., Lambert, A., Brouillet, S. and Achaz, G., 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, *21*(8), pp.1864-1877.

Polhemus, J.T. and Polhemus, D.A., 2007. Global diversity of true bugs (Heteroptera; Insecta) in freshwater. In Freshwater animal diversity assessment (pp. 379-391). Springer, Dordrecht.

Pape, T., Blagoderov, V., and Mostovski, M. B. (2011). Order Diptera Linnaeus, 1758. In Z.-Q. Zhang (Ed.), *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness*. Woodcroft, South Australia: Magnolia Press, pp. 222–229.

Rahman, M.M., Noren, M., Mollah, A.R. and Kullander, S.O., 2019. Building a DNA barcode library for the freshwater fishes of Bangladesh. *Scientific Reports*, 9(1), pp.1-10.

Rahman, M.M., Mollah, A.R., Norén, M. and Kullander, S.O., 2016. Garra mini, a new small species of rheophilic cyprinid fish (Teleostei: Cyprinidae) from southeastern hilly areas of Bangladesh. *Ichthyol. Explor. Freshw*, 27(2), pp.173-181.

Ratnasingham, S. and Hebert, P.D., 2007. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular Ecology Notes*, *7*(3), pp.355-364.

Ratnasingham, S. and Hebert, P.D., 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PloS one, 8(7), p.e66213.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, pp.53-65.

Raupach, M.J., Hannig, K., Moriniere, J. and Hendrich, L., 2016. A DNA barcode library for ground beetles (Insecta, Coleoptera, Carabidae) of Germany: The genus Bembidion Latreille, 1802 and allied taxa. *ZooKeys*, *592*, p.121.

Raupach, M.J., Hendrich, L., Küchler, S.M., Deister, F., Morinière, J. and Gossner, M.M., 2014. Building-up of a DNA barcode library for true bugs (Insecta: Hemiptera: Heteroptera) of Germany reveals taxonomic uncertainties and surprises. *PLoS One*, 9(9), p.e106940.

Resh, V.H., Norris, R.H. and Barbour, M.T., 1995. Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Australian Journal of Ecology*, 20(1), pp.108-121.

Sana, B. B. and Ali, M. S., 2011. A preliminary list of aquatic Coleoptera (Arthropoda: Insecta) in Natore and Rajshahi districts of Bangladesh. *J. Taxon. Biodiv. Res*., 5: 1-4.

Sharmin, S., Rahman, S.H., Naser, M.N. and Hoque, S., 2018. Macro benthic fauna in relation to limnological variables in a migratory bird visiting lake at Jahangirnagar University, Bangladesh. *Journal of Biodiversity Conservation and Bioresource Management*, *4*(2), pp.99-106.

Schmidt-Kloiber, A. and Nijboer, R.C., 2004. The effect of taxonomic resolution on the assessment of ecological water quality classes. *Hydrobiologia*, 516(1), pp.269-283.

Selvakumar, C., Sivaramakrishnan, K.G. and Janarthanan, S., 2016. DNA barcoding of mayflies (Insecta: Ephemeroptera) from South India. *Mitochondrial DNA Part B*, *1*(1), pp.651-655.

Shen, Y., Guan, L., Wang, D. and Gan, X., 2016. DNA barcoding and evaluation of genetic diversity in Cyprinidae fish in the midstream of the Yangtze River. *Ecology and Evolution*, *6*(9), pp.2702-2713.

Simmons, M.P., Ochoterena, H. and Freudenstein, J.V., 2002. Amino acid vs. nucleotide characters: challenging preconceived notions. *Molecular Phylogenetics and Evolution*, *24*(1), pp.78-90.

Sinclair, C.S. and Gresens, S.E., 2008. Discrimination of Cricotopus species (Diptera: Chironomidae) by DNA barcoding. *Bulletin of Entomological Research*, 98(6), pp.555-563.

Smolders, A.J.P., Lock, R.A.C., Van der Velde, G., Hoyos, R.M. and Roelofs, J.G.M., 2003. Effects of mining activities on heavy metal concentrations in water, sediment, and macroinvertebrates in different reaches of the Pilcomayo River, South America. *Archives of Environmental Contamination and Toxicology*, *44*(3), pp.0314-0323.

Srivathsan, A. and Meier, R., 2012. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNAbarcoding literature. *Cladistics* 28:190–194.

Stein, E.D., Martinez, M.C., Stiles, S., Miller, P.E. and Zakharov, E.V., 2014. Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States?. *PloS one*, *9*(4), p.e95525

Suh, K.I., Hwang, J.M., Bae, Y.J. and Kang, J.H., 2019. Comprehensive DNA barcodes for species identification and discovery of cryptic diversity in mayfly larvae from South Korea: Implications for freshwater ecosystem biomonitoring. *Entomological Research*, 49(1), pp.46-54.

Sweeney, B.W., Battle, J.M., Jackson, J.K. and Dapkey, T., 2011. Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality?. *Journal of the North American Benthological Society*, 30(1), pp.195-216

Tang, M., Tan, M., Meng, G., Yang, S., Su, X., Liu, S., Song, W., Li, Y., Wu, Q., Zhang, A. and Zhou, X., 2014. Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Research*, *42*(22), pp. e166.

Tang, M., Hardman, C.J., Ji, Y., Meng, G., Liu, S., Tan, M., Yang, S., Moss, E.D., Wang, J., Yang, C. and Bruce, C., 2015. High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution*, *6*(9), pp.1034-1043.

Thomas, J.A., Frandsen, P.B., Prendini, E., Zhou, X. and Holzenthal, R.W., 2020. A multigene phylogeny and timeline for Trichoptera (Insecta). *Systematic Entomology*, 45(3), pp.670-686.

van Velzen, R., Weitschek, E., Felici, G. and Bakker, F.T., 2012. DNA barcoding of recently diverged species: relative performance of matching methods. *PloS one*, 7(1), p.e30490.

Verneaux, V. and Verneaux, J., 2002. Assessing lake functioning using the macrobenthic community with special reference to Chironomidae (Diptera). A subalpine lake (Lake Annecy) as an example. *Archiv für Hydrobiologie*, pp.61-78.

Versteirt, V., Nagy, Z.T., Roelants, P., Denis, L., Breman, F.C., Damiens, D., Dekoninck, W., Backeljau, T., Coosemans, M. and Van Bortel, W., 2015. Identification of Belgian mosquito species (Diptera: Culicidae) by DNA barcoding. *Molecular Ecology Resources*, *15*(2), pp.449-457.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R. and Hebert, P.D., 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), pp.1847-1857.

Ward, R.D., 2009. DNA barcode divergence among species and genera of birds and fishes. *Molecular Ecology Resources*, *9*(4), pp.1077-1085.

Webb, J.M., Jacobus, L.M., Funk, D.H., Zhou, X., Kondratieff, B., Geraci, C.J., DeWalt, R.E., Baird, D.J., Richard, B., Phillips, I. and Hebert, P.D., 2012. A DNA barcode library for North American Ephemeroptera: progress and prospects. *PloS one*, *7*(5), p.e38063.

Weigand, H., Beermann, A.J., Čiampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M.F., Grabowski, M., Rimet, F., Rulik, B. and Strand, M., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of the Total Environment*, *678*, pp.499-524.

Whitworth, T.L., Dawson, R.D., Magalon, H. and Baudry, E., 2007. DNA barcoding cannot reliably identify species of the blowfly genus Protocalliphora (Diptera: Calliphoridae). Proceedings of the Royal Society B: *Biological Sciences*, 274(1619), pp.1731-1739.

Yu, H., Kong, L. and Li, Q., 2016. Evaluation of the efficacy of twelve mitochondrial protein-coding genes as barcodes for mollusk DNA barcoding. *Mitochondrial DNA Part A*, *27*(2), pp.1336-1339.

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C. and Ding, Z., 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*(4), pp.613-623.

Zamora-Munoz, C. and Alba-Tercedor, J., 1996. Bioassessment of organically polluted Spanish rivers, using a biotic index and multivariate methods. *Journal of the North American Benthological Society*, 15(3), pp.332-352.

Zamora-Munoz, C. and Alba-Tercedor, J., 1996. Bioassessment of organically polluted Spanish rivers, using a biotic index and multivariate methods. *Journal of the North American Benthological Society*, 15(3), pp.332-352.

Zhang, J., Kapli, P., Pavlidis, P. and Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29(22), pp.2869-2876.

Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30(5), pp.614-620.

Zhou, X., Adamowicz, S.J., Jacobus, L.M., DeWalt, R.E. and Hebert, P.D., 2009. Towards a comprehensive barcode library for arctic life-Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Frontiers in Zoology, 6*(1), pp.1-9.

Zhou, X., Jacobus, L.M., DeWalt, R.E., Adamowicz, S.J. and Hebert, P.D., 2010. Ephemeroptera, Plecoptera, and Trichoptera fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. *Journal of the North American Benthological Society*, *29*(3), pp.814-837.

Zhou, X., Robinson, J. L., Geraci, C. J., Parker, C. R., Flint, O. S. Jr, Etnier, D. A., and Hebert, P. D. N., 2011. Accelerated construction of a regional DNA-barcode reference

library: Caddisflies (Trichoptera) in the Great Smoky Mountains National Park. *Journal of the North American Benthological Society*, *30*(1), pp.131–162.

Zhou, X., Frandsen, P.B., Holzenthal, R.W., Beet, C.R., Bennett, K.R., Blahnik, R.J., Bonada, N., Cartwright, D., Chuluunbat, S., Cocks, G.V. and Collins, G.E., 2016. The Trichoptera barcode initiative: a strategy for generating a species-level Tree of Life. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702), p.20160025.

Zou, S., Li, Q. and Kong, L., 2012. Multigene barcoding and phylogeny of geographically widespread muricids (Gastropoda: Neogastropoda) along the coast of China. *Marine Biotechnology*, *14*(1), pp.21-34.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X.U., Zhou, L., Tang, M., Fu, R., Li, J. and Huang, Q., 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *Gigascience*, *2*(1), pp.2047-217X.

Chapter 3

Metabarcoding for high-throughput freshwater bioassessment: prospects and methodological challenges

Chapter 3: Metabarcoding for high-throughput freshwater bioassessment: prospects and methodological challenges

3.1 Abstract

Metabarcoding is a powerful technique for biological assessments of aquatic ecosystems. However, significant optimization and standardization of various methodological aspects are still required for its cost-effective, time-saving, and reliable application in large-scale biomonitoring programmes. In general, critical steps to be standardized include the preparation of bulk macroinvertebrate samples, optimizing the procedure for homogenization of samples for DNA extraction and the strategies for DNA pooling from these extracts. Metabarcoding also relies on the choice of robust universal primers for amplifying a wide range of bioindicator taxa, and reliable bioinformatic processing in particular OTU clustering for reliable diversity estimation. In this chapter, I compare three techniques of bulk sample preparation used in metabarcoding of macroinvertebrates to acquire maximum diversity outputs. The results implied that raw macroinvertebrates samples with minimum processing efforts can be used reliably in wide-ranging biodiversity assessment. As bulk complex samples need to be homogenised, I also evaluated the exigency of replicates of homogenate samples in metabarcoding pipeline which also supported the use of a properly homogenized single sample for reliable species recovery from each bulk sample. The primer used in this study was found as a prospective universal primer to amplify reasonably a wide range of macroinvertebrate taxa, in particular for mayflies, stoneflies, caddisflies, beetles, water bugs, dragon and damselflies and dipterans, although the amplification rate was low for annelids and molluscs. Investigation on DNA extractions from pooled samples and separately processed DNA extractions before sequencing together showed the surprising result that more OTUs were generated from the pooled DNA samples than the combined separately processed samples. Comparison of two clustering methods, Usearch and Swarm, indicated similar efficacy for OTU clustering, although Usearch was considered slightly superior by some criteria. These results have implications for how metabarcoding of macroinvertebrates should be used in biomonitoring of freshwater ecosystems.

3.2 Introduction

3.2.1 Metabarcoding - a metagenetic approach for biodiversity studies

Bioassessment of the freshwater ecosystem requires an accurate and comprehensive method for characterizing the bioindicator (e.g., macroinvertebrates) communities (Carrizo et al., 2017; Yu et al., 2012). The key challenge of existing morphology-based conventional methods of bioassessment is to ensure large-scale biodiversity data from complex community samples with maximum accuracy and consistency in taxonomic identification rapidly and cost-effectively. To address this challenge, DNA-based approaches have potential advantages for species (or OTU) level identification irrespective of body sizes and life stages of organisms (Creer et al., 2016; Hajibabaei et al., 2011). DNA barcoding has been a highly effective approach for building specimen-based reference libraries for different groups of animals (Ekrem, Willassen and Stur, 2007; Hebert et al., 2003). Although conventional DNA barcoding can be used to identify thousands or millions of individual specimens (Hajibabaei et al., 2007; Ivanova, deWaard, and Hebert, 2006), it is time-consuming, labour-intensive, and too expensive to apply in large-scale biomonitoring programmes.

Fortunately, recent technological advances in high throughput sequencing (HTS) and bioinformatic processing have surmounted the limitations of single-specimen DNA barcoding and extended the opportunity for mass amplification and sequencing to generate high-throughput biodiversity data for ecosystem monitoring (Bush et al., 2017). One of the emerging techniques is DNA metabarcoding, which holds great promises for rapid and low-cost biodiversity studies assessing the community composition of bulk or environmental samples (eDNA) (Andujar et al., 2018; Aylagas et al., 2016; Elbrecht et al., 2017; Emilson et al., 2017; Kuntke et al., 2020). This approach supports species identification for communities of individuals amplifying a standardized DNA fragment of taxonomically informative genes (COI or other genes) with universal primers (Cristescu et al., 2014; Hajibabaei et al., 2012; Yu et al., 2012). The resulting amplicons are then sequenced on an HTS platform (e.g., Illumina HiSeq or MiSeq) and compared to a previously generated DNA sequence reference database of well-characterized species or available public databases (e.g., GenBank/NCBI) for taxonomic assignment and subsequent analysis (Ji et al., 2013; Taberlet et al., 2012). Therefore, metabarcoding is now widely used for large-scale biodiversity assessment

addressing different aspects of ecology in freshwater and terrestrial ecosystems (Bohan et al., 2017; Creer et al., 2016).

3.2.2 Scope and present application extent of metabarcoding

Metabarcoding is being tested and validated for biodiversity studies including species richness, distribution and composition of community assemblages addressing taxonomical, ecological as well as evolutionary questions. The alpha and beta diversity metrics can be estimated reliably through metabarcoding despite the over or underestimation of biodiversity inherent to this technique and the loss of taxonomic information for incomplete reference datasets (Yu et al., 2012). Metabarcoding-based taxonomic inferences complement morphological approaches although results are still inconclusive for many taxonomic groups (Carew et al., 2013; Cowart et al., 2015; Gibson et al., 2015; Hajibabaei et al., 2012; Zimmermann et al., 2015). However, the potential of metabarcoding in describing more species than morphological approaches in a limited time is quite evident in different studies conducted in various ecosystems (Brandon-Mong et al., 2015; Bush et al., 2020; Creedy, Ng and Vogler, 2019; Elbrecht et al., 2017; Ji et al., 2013; Shokralla et al., 2015).

Metabarcoding has been used in aquatic ecosystems to characterize the marine metazoan community including meio and meso-benthic organisms (Fonseca et al., 2014; Leray and Knowlton, 2015), benthic macroinvertebrates (Aylagas et al., 2016; Aylagas et al., 2018), zooplankton (Chain et al., 2016), and deep-sea nematodes (Dell'Anno et al., 2015). In freshwater ecosystems, the breadth of application of metabarcoding is growing and several studies have performed a robust benchmarking of this technique for rapid and reliable assessment of invertebrate biodiversity (e.g., Andújar et al., 2018; Emilson et al., 2017; Elbrecht et al., 2017; Hajibabaei, 2012; Martins et al., 2019). Besides the aquatic environment, this approach has now been employed in other ecosystems, for example, to assess the diversity and composition of terrestrial vertebrates (Goldberg et al., 2011; Sato et al., 2017; Vences, et al., 2016;), terrestrial arthropods (Arribas et al., 2016; Beng et al., 2016; Braukmann et al., 2019; Ji et al., 2013; Oliverio et al., 2018; Marquina et al., 2018) and canopy arthropods (Creedy, Ng and Vogler, 2019). The applicability of diatom metabarcoding has also been evaluated in the bioassessment of rivers that supported it as a

valid approach for ecological quality assessment (Mortágua et al., 2019; Pérez-Burillo et al., 2020).

In addition to diversity estimations, the metabarcoding approach has been extended to address other environmental issues: detecting a pesticide spill in a river (Andujar, et al., 2018), ecogenomic responses of benthic communities to environmental stressors (Beermann et al., 2018; Xie et al., 2017), ballast water surveillance (Zaiko et al., 2015), monitoring biological invasions (Comtet et al., 2015), establishing host-parasitoid and predator-prey relationships (Galan et al., 2018; Sow et al., 2019) and phylogenetic placement of species/OTUs (Keck et al., 2018). eDNA metabarcoding collecting DNA from water has been attempted to evaluate freshwater macroinvertebrates (Fernández et al., 2018), which was found highly sensitive and required less sampling and identification efforts. However, the DNA recovery from eDNA samples (e.g., water, sediments) is trickier than from bulk samples and metabarcoding outputs especially the composition of invertebrate bioindicator strongly differs between two approaches (Macher et al., 2018). In a study of invertebrate metabarcoding, eDNA extracted from water samples was also evident as a poor proxy for DNA from bulk samples and the results supported the use of bulk benthic samples for metabarcoding based bioassessment (Hajibabaei et al., 2019). Contrary to invertebrate fauna, eDNA metabarcoding for detecting freshwater vertebrate fauna has been found effective, especially for fish and amphibians (e.g., Goldberg et al., 2011; Vences et al., 2016). In the context of ecosystems, metabarcoding studies with eDNA for meiobenthic and macrobenthic communities in marine and estuarine habitats are found more convenient and effective (Brannock et al., 2016; Lobo et al., 2017; Lanzén et al., 2016) than freshwater.

3.2.3 Challenges of metabarcoding for large-scale biodiversity assessment

Despite its potential for community species identification, metabarcoding faces some challenges to produce reliable biodiversity estimates, which should be addressed before application in regular biomonitoring programs. Generally, the success of metabarcoding relies on the rapid and reliable retrieval of a wide range of taxonomic groups from a given bulk or environmental sample. DNA metabarcoding is in essence a multifaceted approach based on many procedural steps as follows: a) collection of bulk (or environmental) samples b) sample preparation for DNA extraction c) primer selection and PCR amplification of a

taxonomically informative genomic region, d) high-throughput sequencing of the amplicons, e) bioinformatic processing of the sequences (quality filtering of amplicons and clustering into OTUs) and f) completeness of reference sets against which to identify the sequence data. Each of these steps can potentially introduce its own sources of artefacts and biases (Bik, et al., 2012; Creedy et al., 2021; Deagle et al., 2014; Kress et al., 2015; Zinger et al., 2019).

For example, after sample collection using a standardised protocol, sample preparation is a vital step for DNA extraction from bulk samples. The accurate assessment of the community composition of invertebrates relies on the inclusion of all life stages (eggs, pupa, larvae etc.). Hence, specimen sorting, and exclusion of unwanted matters may affect the maximum inclusion of organisms from a complex kick net which remains a key issue in large-scale freshwater biomonitoring. Meanwhile, manual cleaning and sorting of samples are laborious and time-consuming to remove unwanted debris (plant parts, pebbles, sediment particles) retaining target organisms. Variation in biomass of different species present in a bulk sample is another hurdle for equal amplification by the same primer, as large-bodied species are more likely to be recovered (Brandon-Mong et al., 2015; Elbrecht, Peinert and Leese, 2017). Therefore, the use of OTU read counts for measuring species abundance of a community is still a tricky issue in metabarcoding studies (Elbrecht and Leese, 2015; Piñol et al., 2015).

In most metabarcoding studies, invertebrate specimens are usually separated from debris and classified based on their size and taxonomy to avoid biomass biases and PCR inhibitors (Carew et al., 2013; Creedy, Ng and Vogler, 2019; Elbrecht, Peinert and Leese, 2017). Besides, taxonomic combination (Beentjes et al., 2019; Morinière et al., 2016), bulk DNA extractions from body parts including pooling effects (Braukmann et al., 2019) have been investigated to validate metabarcoding of arthropods mock communities, but these approaches may increase the chance of underestimation of some taxa (Haase et al., 2010). Alternatively, Arribas et al. (2016) proposed a Ludox (a colloidal solution) based flotation–Berlese–flotation (FBF) protocol for the processing of soil samples to obtain clean DNA from grassland arthropod mesofauna as part of the metabarcoding and metagenomic pipeline. This protocol provided contaminant (bacteria and inhibitors) free DNA of soil arthropods from a large volume of soil and exposed a diverse community of Acari and Collembola. The Ludox flotation technique has also been tested for marine and estuarine sediment samples for describing meiofaunal and microbial diversity through the metabarcoding approach which also outperforms their expected diversity (Lallias et al., 2015; Fonseca et al., 2017). For

freshwater invertebrate metabarcoding, the samples prepared by the flotation method (Andújar et al., 2018) and unsorted bulk macroinvertebrate samples (Nichols et al., 2019) have also been used to detect target macroinvertebrate taxa, but its utility still requires adequate evidence to use unsorted samples in the large-scale biomonitoring.

A recent study (Martins et al., 2019) tried to optimize non-destructive DNA recovery from 96% ethanol used to preserve macroinvertebrate samples sidestepping DNA extraction from cleaned, sorted, and homogenized bulk samples. This study with metabarcoding of a 313-bp COI fragment detected most taxa from previously built reference barcodes for the same organisms. Conversely, inconsistency in OTU generation between soil samples and preservative ethanol and homogenates was also revealed from Malaise trapped arthropod communities (Marquina et al., 2019). This result indicates preservative ethanol as a potential source of DNA for macroinvertebrate samples, but it still requires its reproducibility in real-world large complex samples. Therefore, it is imperative to find a method for sample preparation or a way of using raw samples and to test their effects on final species detection from large and complex community samples. It is still in flux how do the sediments in bulk samples affect the final species composition sequenced on the HTS platform with a higher sequencing depth? And to what extent does it affect the estimate of community diversity in a biomonitoring program?

Selecting a potential universal primer for the target gene marker is a primary issue for mass amplification of diverse taxa present in a complex community sample. DNA from bulk samples tends to be often differentially amplified (Elbrecht and Leese, 2015; Piñol, et al., 2019; Tedersoo et al., 2018) due to biases either from the PCR primers (Clarke, et al., 2014) or the DNA polymerase (Dabney and Meyer, 2012; Krehenwinkel et al., 2017; Nichols et al., 2018; Pan et al., 2014). Investigations of potential barcodes and primer choice (Andújar et al., 2018a; Elbrecht and Leese, 2017; Hajibabaei et al., 2019; Krehenwinkel et al., 2018) supported the COI barcodes to provide better biodiversity coverage from bulk samples than other markers and the efficacy of different primer sets varied in metabarcoding of arthropods and freshwater invertebrates. Integration of multiple primer sets, and barcodes have also been advocated to obtain a more complete biodiversity estimate (Hajibabaei et al., 2019) and for a more comprehensive and accurate understanding of ecological impacts on freshwater biodiversity (Ficetola et al., 2021). These studies suggest judiciously choosing of primers considering the target taxa prior to the application of metabarcoding in wide-ranging

biomonitoring programmes. As the alternatives to COI, the nuclear ribosomal genes 18S rRNA (Capra et al., 2016; Creer et al., 2010) or 28S rRNA (Hirai, et al., 2014), and the mitochondrial 12S rRNA (Machida, Kweskin, and Knowlton, 2012) or 16S rRNA (Elbrecht et al., 2016; Saitoh et al., 2016) are being explored in many studies but those are still limited by their reference database deficiencies.

The impact of replication of samples (technical and biological) on final results is an important aspect that is often poorly addressed in invertebrate metabarcoding. Replicates of homogenate bulk samples for DNA extraction and multiple PCRs may reduce the problems associated with the detection of missing taxa that are actually present (false negatives) and increase the chance to identify rare species (Ficetola et al., 2015). DNA extraction replicates could improve the estimates of eukaryote diversity and the ability to separate samples with different characteristics in metabarcoding of marine sediments (Lanzén et al., 2017). Furthermore, the effects of sample pooling (bulk and eDNA), pooled DNA from technical replicates (homogenate samples) and the pooling of amplicon products still remain understudied in metabarcoding studies with freshwater invertebrates. These approaches are assumed to ensure maximum species detection with minimum survey effort but may introduce their own source artefacts and biases for large-scale surveys of aquatic organisms. Wainer et al. (2020) compared the fungal communities in pooled and unpooled eDNA samples (soil samples) where fungal richness decreased in pooled samples, but the detection of rare and invasive plants increased, indicating that pooling might be effective to determine the composition of soil communities. Another study tested the effects of the pooling of eDNA samples (water samples) and found the pooling strategy was unsuitable to assess fish diversity, but this procedure could be useful to compare fish communities among sites (Sato et al., 2017). In the context of DNA pooling in metabarcoding of diverse arthropod mock communities, Braukmann et al. (2019) found that separately processed DNA (PCR of independent samples) produced more diversity than pooled DNA prior to PCR (PCR of pooled samples). However, this issue should be further investigated for invertebrate communities to test its efficacy in large-scale biomonitoring.

In addition, false positives are another crucial issue in DNA metabarcoding of bulk and environmental samples. False positives may arise at any step of the experimental workflow through the presence of reagent contaminants, PCR cross-contamination, replication errors by the DNA polymerase and sequencing errors (Taberlet et al., 2018; Willerslev et al., 2014).

HTS data may contain PCR chimeras, nuclear mitochondrial pseudogenes (Numts), and sequencing noise (Lenz and Becker, 2008; Tedersoo et al., 2018). Therefore, the quantification of OTUs mostly relies on clustering of reads after algorithmic denoising for removing chimeras, contaminants, and sequencing errors. These OTUs generated by clustering a large number of sequences into sets that ideally corresponds to the species in the original samples. Species are defined operationally as a cluster of similar sequences with a standard cut off value (generally 97%), which is also a challenging issue because the level of intraspecific variation and the divergence of species from each other are not uniform across taxa (Yu et al., 2012). Moreover, the obtained OTUs are not easily reconcilable across sites regarding species delimitation if the geographic variation is added (Bergsten et al., 2012; Cristescu, 2014).

The challenges for OTUs clustering and accompanying other issues raise the possibility of over or underestimation of species in mixed-species assemblages (Creedy, Ng and Vogler, 2019). Different clustering algorithms, the flexibility of quality filtering parameters, and sequence divergence threshold greatly affect the number of OTUs generated in both mock and natural samples (Flynn et al., 2015; Brannock and Halanych, 2015). Alternative strategies, including no clustering and clustering with varied divergence thresholds, were also tested and results showed that the number of OTUs estimated with 99% to 97% similarity thresholds varied greatly, but ≥97% divergence thresholds were reliable to reveal the composition of the complex community (Xiong and Zhan, 2018). These results suggest the need for focusing on different clustering methods with a range of parameters for quality filtering of sequences and to choose the best fit for answering the questions asked from the data. Therefore, considering the above-stated issues for validating metabarcoding, this approach should be highly standardized to address the challenges of exploring large-scale taxonomic and ecological information for biomonitoring of freshwater ecosystems.

3.2.4 Aims and research questions

In this study, I investigate the reproducibility of metabarcoding under different sample preparation techniques, strategies for using replicates of homogenate samples, and DNA pooling in the metabarcoding procedure to answer the following questions: (a) In the initial step of DNA extraction from standard aquatic sampling using kick samples, how do different methods of specimen extraction affect the species/OTUs detection under three processing techniques: i) using raw bulk samples mixed with sediment and debris; ii) bulk samples treated by Ludox colloidal solution for floating of specimens and removal of debris; iii) bulk samples treated in only water for floating specimens. b) How does the recovered OTU set relate to the technical replicates of homogenised samples? I did homogenisation of dried samples and tested the relevance of taking single or multiple homogenised samples for DNA extraction and sequencing and their effect on final outputs. c) How does DNA pooling from multiple replicates before PCR affect the final OTU generation? d) How do clustering methods influence the OTU delimitation? I tested two clustering algorithms among various algorithms available to cluster the sequence reads for the specific parameter settings that may have strong effects on the numbers of OTUs obtained. e) Is the universal primer applicable for amplifying COI barcodes and to what extent does its performance vary for a wide range of taxa in a complex community assessment? In this chapter, I describe the standard metabarcoding method with sampling, laboratory and bioinformatics protocols taking into account the aforementioned questions and challenges underlying the metabarcoding pipelines. In the next chapter, these methods are then applied to test if under these optimised conditions we can detect significant community differentiation across environmentally gradient sites.

3.3. Methods

3.3. 1 Study sites and sampling protocol

Samples for metabarcoding of freshwater macroinvertebrates were collected from selected upland and lowland rivers of Bangladesh (described in chapter 2). In total, 140 two-poled kick net samples were taken from 4 lowland and 16 upland rivers for characterizing macroinvertebrates communities (see chapter 4). Besides these, I collected additional samples for testing different sample preparation techniques and OTUs clustering methods of the metabarcoding pipeline.

For testing the effect of three sample preparation techniques, 18 samples were collected from 6 sites (3 samples per site) of an interconnected upland stream (Fig. 3.1) with a 500-µm mesh two-poled kick-net operating for three minutes for each sample. The entire contents of the samples from kick nets were placed in a container tray and processed by manually removing large debris and with consecutive sorting by different meshed (1 mm and 0.5 mm) sieve buckets. After preliminary processing, all samples were preserved with 95% ethanol in the field and transported on ice to the laboratory. To assess the performance of two different clustering methods (Swarm and Usearch) for OTU clustering (species estimation), 30 samples were also collected separately from 5 sites of 3 upland and 3 low land rivers. In this case, a single kick-net sampling was done at each of five sites of six rivers applying above stated sampling strategies in the field.

For investigating the relevance of taking multiple technical replicates for metabarcoding the bulk samples, 9 homogenate bulk samples were taken from 3 sites of 3 rivers (3 technical replicate samples per river). Furthermore, 10 DNA samples from 6 sites of upland rivers were tested to assess the DNA pooling effect (before PCR amplification) on the outputs of metabarcoding of macroinvertebrates. For the last two sets of experiments, samples were selected from the main sampling slot of macroinvertebrates metabarcoding (chapter 4).

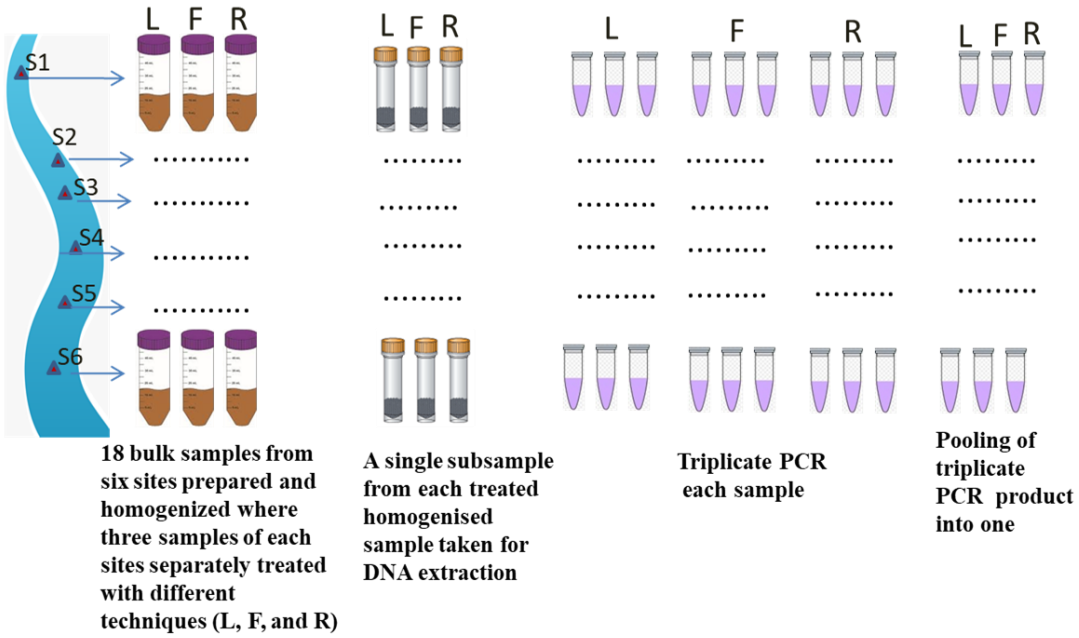

Figure 3. 1 Study design for metabarcoding bulk macroinvertebrate samples under three preparation techniques with Ludox flotation (L), Water flotation (F) and Raw sample (R).

3.3. 2 Sample processing of three techniques for bulk macroinvertebrate sample preparation

For DNA based biodiversity studies, bulk sample preparation techniques including flotation, decantation, and isopycnic separation (density gradient-based) have been applied for separating meso, meio and microbial eukaryotes from terrestrial soil and marine sediments using colloidal silica solution (Arribas et al., 2016, Briski et al., 2013; Burgess , 2001; Creer et al., 2015). These techniques were found effective for extracting the detached body parts and extracellular DNA along with active organisms and their dormant stages and for removing unwanted debris and bacteria. Nevertheless, the application of these techniques has hardly been described for the preparation of freshwater macroinvertebrate samples. As freshwater macroinvertebrate kick-net samples contain a lot of debris including soil and sand particles, plant parts, and leaf litter, I used Ludox colloidal solution (Colloidal silica polymer) as a flotation medium to clean the bulk samples along with two other techniques. In this study, we prepared bulk samples before DNA extraction with three methods: i) Ludox flotation; ii) flotation with water; iii) raw samples after manually removing debris for ten minutes. Before applying these three treatments of sample preparation, all large-sized (>15 mm) organisms (generally odonates, water bugs, molluscs) were isolated from all samples for

DNA extraction separately to minimize biomass biases created from large specimens. After excluding the large samples, firstly I mixed uniformly three bulk samples of each site to make one and then again equally divided into three samples for uniform distribution of organisms in three sample preparation methods. In this way, I tested three samples using three preparation techniques from each of the six sites.

3.3. 2a Ludox flotation protocol

A density separation-flotation protocol was developed using a colloidal silica solution namely Ludox HS 40 (Colloidal silica polymer, specific gravity of 1.31 g cm^{-3}) for removing both sediments and plant parts. This protocol works on the principle that the separation of macroinvertebrates occurs when the density of the solution is above the density of the macroinvertebrate specimens but below the density of sediments and debris. By trial and error, the desired specific density of the solution was optimized at 1.13 g cm^{-3} diluting with water to provide an operative medium for the separation of targeted macroinvertebrates from unwanted debris.

Each sample was transferred into a graduated flask and filled with deionised water. Covering the flask, vigorously shaking dislodged the organisms from sediments and plant parts. The flask was then allowed to settle for around 30 sec for the deposition of sediments and to suspend the organic matter. Water with suspended organisms was gently poured onto a 45 µm sieve retaining the sediment in the flask. I collected the organisms from the sieve with a spatula and transferred them into a pre-labelled tube filled with alcohol. After that, the remaining sample and sediments in the flask were exposed to density separation solution with Ludox (specific density 1.13g cm^{-3}) in a beaker. The beaker was placed on a magnetic stirrer with a Teflon-coated bar and stirred for about 2 minutes for dislodging and floating organisms from debris and sediments. The sample was kept for 1 minute for sediment particles and detritus to settle at the bottom and for floating and suspending macroinvertebrate samples in the column and near the surface. Then I gently poured Ludox solution with organisms into a 45 µm mesh sieve and rinse the organisms completely with distilled water to remove Ludox. These steps were repeated at least three times for the maximum inclusion of organisms. Finally, the remaining sediments/debris was also checked under a stereomicroscope for any organisms left. A diagrammatic outline for the Ludox flotation protocol is given in Fig. 3.2.

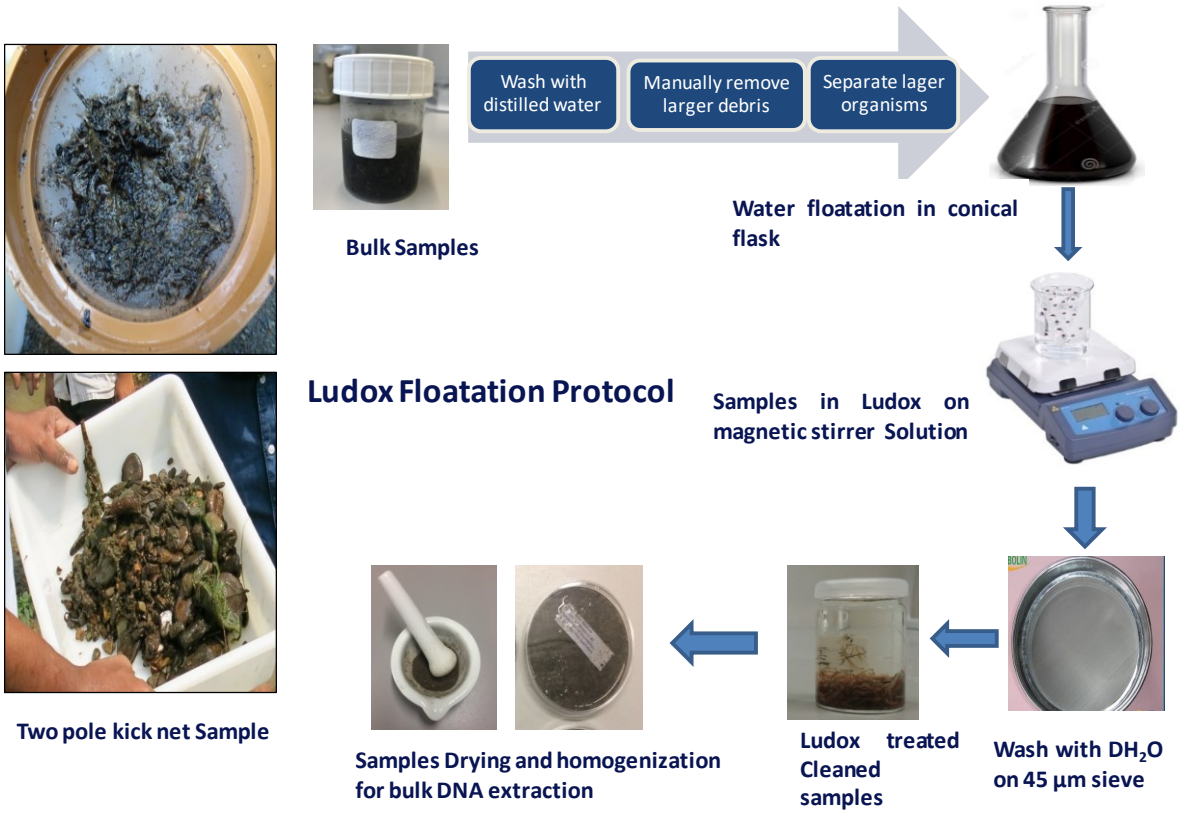

Figure 3. 2 Ludox flotation protocol for bulk sample preparation for metabarcoding of macroinvertebrates

3.3. 2b Water flotation protocol

For this method, the same steps of the Ludox floatation protocol were maintained using only water instead of Ludox. Avoiding the use of Ludox, I did this experiment on an assumption that dislodging of organisms, detached body parts, tissues remnants with extracellular DNA occurred during flotation and from this part, sufficient community DNA can be extracted for downstream analysis.

3.3. 2c Raw samples protocol

In this method, I just took raw samples after 10 minutes removal of unwanted debris and plant parts per sample. To sidestep the preparation task for saving cost, time and labour, I tested with raw samples to assess the sediment effects on final species estimation for complex community samples.

3.3. 3 Samples drying and homogenization

After treating with the above-stated techniques, the samples were loaded onto an individual Petri dish. Samples were dried in an incubator at 37^0C and the dried samples were homogenised through grinding with cleaned and sterilized mortar and pestle. For proper homogenization, samples were broken up several times with a clean sterilized spatula and then transferred into a new tube. Bulk organism samples consisting of large-sized animals were also homogenised using a tissue lyser. The schematic diagram of this protocol is as follows (Fig. 3.3).

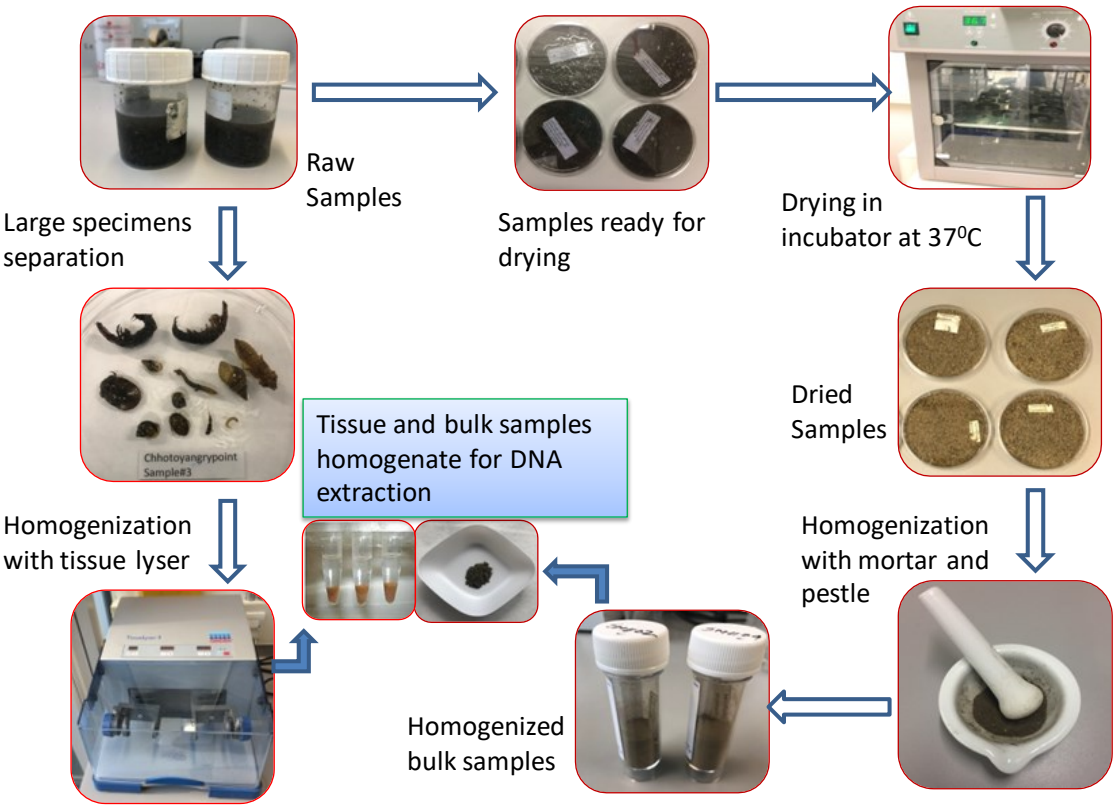

Figure 3. 3 Processing protocol for raw bulk samples of macroinvertebrates for use in the metabarcoding pipeline.

3.3. 4 Sample preparation for testing OTUs clustering methods

30 bulk samples were prepared by separating large size (>15 mm) organisms from each sample and then treating with Ludox flotation followed by drying and homogenization (Figure 3.3). The large-individuals separation from each sample additionally produced 30 bulk organisms' subsamples. DNA extraction, PCR amplification and sequencing were done separately for smaller individual and larger individual samples to minimize biomass biases created from large specimens. After sequencing, respective smaller and larger individual samples were combined for final OTU estimation from two OTU clustering pipelines (Usearch and Swarm) (Edgar, 2010; Mahé, et al., 2015).

3.3. 5 Sample preparation for testing replicates of homogenate bulk samples for DNA extraction

For investigating the effects of using multiple technical replicates of homogenate samples on species description (OTU generation), 9 Ludox treated homogenised samples were selected from 3 sites of 3 different rivers taking three replicates of the same amount from each site (Fig. 3.4). I did this experiment assuming that multiple replicates from the same samples will return more diversity output than a single sample though processing of multiple replicates is expensive in terms of DNA extraction and sequencing.

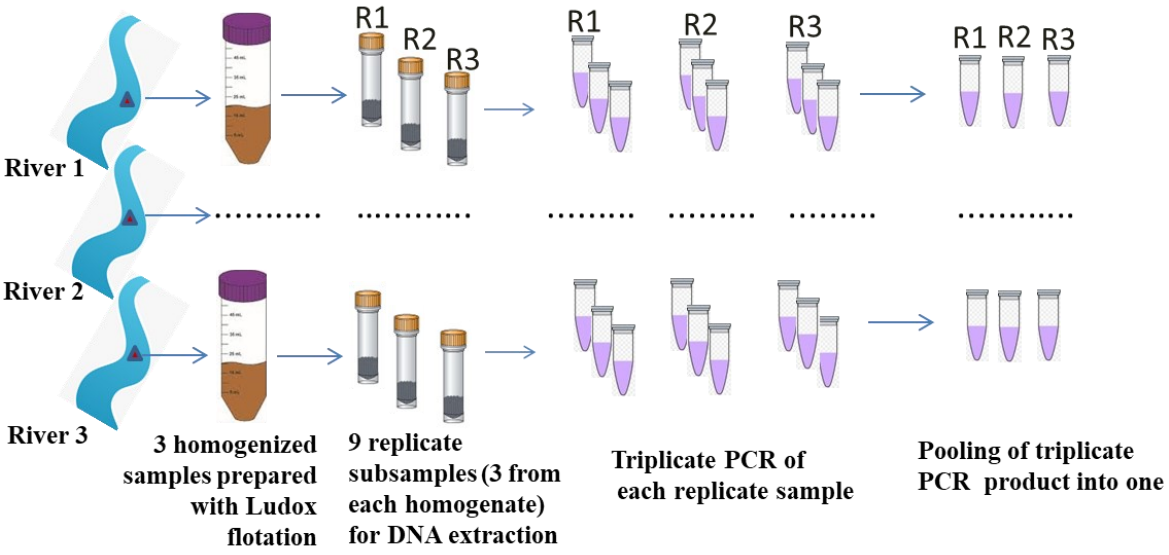

Figure 3. 4 Experimental design for testing the outputs of multiple technical replicates from homogenate bulk sample in metabarcoding pipeline.

3.3.6 Sample preparation for testing the effects of DNA pooling before PCR amplification

This test was carried out with pooled and non-pooled DNA from 20 replicate samples from 10 Ludox treated homogenate bulk samples taking two replicates from each of them. I carried out this test assuming that using pooled DNA in downstream steps is also effective for producing equivalent results. After extracting DNA from each replicate of a specific sample, this experiment was designed with two pathways: i) unpooled pathway where each replicate's DNA sample was kept separate, and amplification and sequencing were performed separately and ii) pooled pathway where DNA from two replicates of each sample was pooled together for PCR amplification and sequencing (Fig. 3.5). In the case of pooling, the same amount of DNA was pooled from their respective replicate DNA samples and after proper mixing (by vortexing and low rpm centrifugation), I took these samples for PCR amplification and subsequent downstream analysis. After bioinformatic processing, outputs from respective unpooled samples were combined to compare with outputs of the respective pooled sample.

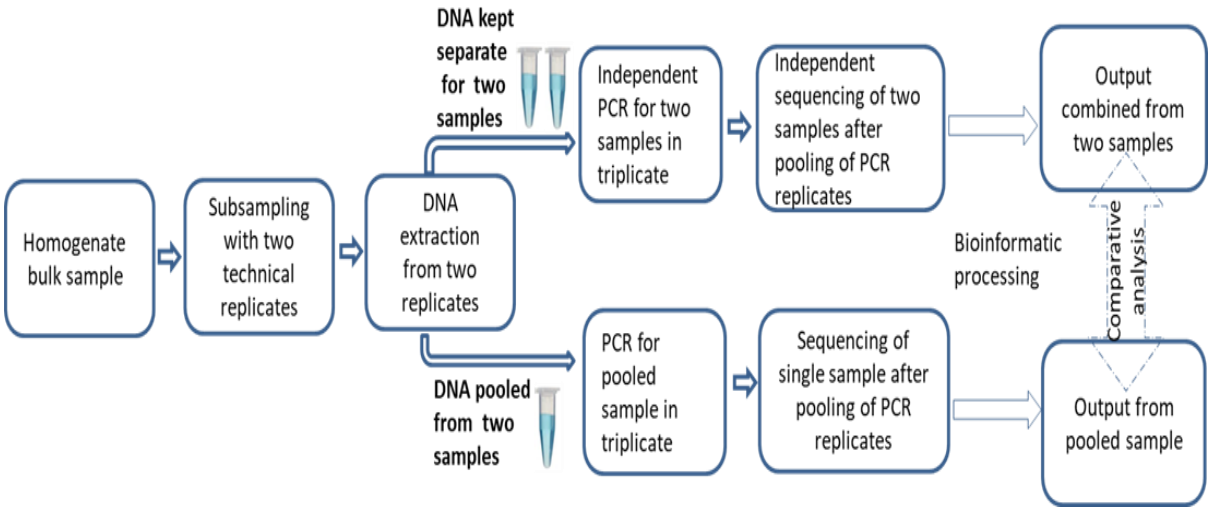

Figure 3. 5 Flow diagram showing study design of investigating the effects of unpooled and pooled DNA samples (before PCR) on the final outputs of metabarcoding of macroinvertebrates bulk samples.

3.3. 5 Bulk DNA extraction

DNA was extracted from homogenised powder samples using a DNeasy Power Soil Kit (Qiagen) eluting to a volume of 100 µl. Only 0.20 g of homogenised powder were taken from each bulk sample for DNA extraction. I modified the manufacturer's protocols increasing the amount of C1 solution (Cell lysis solution of kit) that was required to suspend the dried bulk sample. I used 180 µl of C1 solution (60 µl recommended in the main protocol) and the first centrifugation step was extended to 1 min (30 sec in manual) to extract the proper volume of supernatant required for subsequent steps in the extraction process. DNA from bulk organism samples/bulk tissue samples (prepared by separating large specimens from each raw sample) was extracted with a DNeasy 96 Blood and Tissue Kit following the manufacturer's protocols and eluted to a volume of 200 µl. DNA purity and concentration was determined using the Nanodrop ND-8000 (Thermo Scientific) system, prior to PCR amplification of extracted DNA samples.

3.3. 7 PCR amplification of Bulk DNA

Amplification of DNA was performed for 418 bp COI barcodes of multiple species in bulk community samples using degenerate primers (see chapter 2). For metabarcoding, each sample was amplified in three independent reactions with the same reaction volume and protocols (Figs. 3.1, 3.4). The PCR reactions contained 2 µL DNA template, 14.65 µL sterilized ultrapure water, 3.0 µL 10X TaKaRa buffer (Takara Bio Inc.), 0.15 µL MgCl2 (50 mM, Bioline), 0.40 µl of TaKaRa dNTPs (2.5 mM each), 0.70 µl of each primer (10 mM), 0.15 µl of Taq^{TM} Hot Start polymerase (5u/µl, Takara Bio Inc.), 0.25 µl of BSA (20 mg/ml, Thermo Scientific). The PCR conditions were started with preheated lid at 105°C and initial denaturation for 4 min at 95^0 C, followed by a total of 30 cycles of 95°C for 30 sec, 48°C for 30 sec, and 72°C for 1 min 45 sec, and a final extension at 72°C for 10 min. Successful PCR products were considered by a clear single band of expected size visualized on a 1.0 % agarose gel. Positive and negative controls were also included in PCR to check the expected amplified bands and contamination of PCR respectively.

3.3. 8 Sample Pooling, library preparation and multiplex amplicon sequencing

Taking the equal volume of PCR products from three replicates of each sample were pooled together in a single plate. The pooled PCR plates were then cleaned with Agencourt AMPure XP paramagnetic bead technology (Agencourt Bioscience Company, Massachusetts, USA) following the manufacturer's protocols with slight modification (e.g., increase of incubation period, adjustment of beads volume based on amplicon length). After quality control, library preparation with secondary PCR and indexing with Nextra XT tags, amplicons were sequenced with aim of 80k reads per sample on an Illumina MiSeq platform (2x300 bp paired end) at the sequencing facility of Earlham Institute, Norwich, UK.

3.3. 9 Bioinformatic processing of metabarcoding data

After a quality check with Fastq (http://www.Bioinformatics.babraham.ac.uk /projects/fastqc/), sequences were processed using the NAPtime pipeline (NGS Amplicon Pipeline) that was described in chapter 2 (Creedy, Ng and Vogler, 2019). This pipeline is potentially applicable for the analysis of both barcode and metabarcode sequences using some common scripts (e.g., NAPdemux, NAPtrim, NAPmerge and NAPconvert) except for two final scripts: NAPselect for barcoding and NAPcluster for metabarcoding. In Chapter 2, NAPdemux, NAPtrim, NAPmerge and NAPconvert were described. After taking the merged fasta file for each metabarcode sample, the NAPcluster script dereplicates and does size sorting of reads before denoising using USEARCH UNOISE (Edgar, 2016). NAPcluster carries out clustering using USEARCH cluster_ otus (Edgar, 2010) or Swarm (Mahé, et al., 2015), and mapping reads to OTUs using USEARCH usearch_global (Edgar, 2010). We also tested the effect of the Swarm clustering algorithm by incorporating in the NAPcluster scripts. NAPcluster also can assign OTUs a preliminary taxonomy based on parsing BLAST searches against the GenBank nt database, but I did this separately outside NAPcluster. I retained only contigs of 418 bp and unique sequences with >2 copies considering them to be sequencing errors rather than valid sequences. NAPcluster produces a combined fasta file with all OTUs and a table (OTU_Read table) of read numbers for each OTU in each metabarcode sample. A BLAST search of OTU representative sequences was conducted against a database created from NCBI (on 21-10-2020) for their taxonomic assignment. A bioinformatics workflow is shown below (Fig. 3.6).

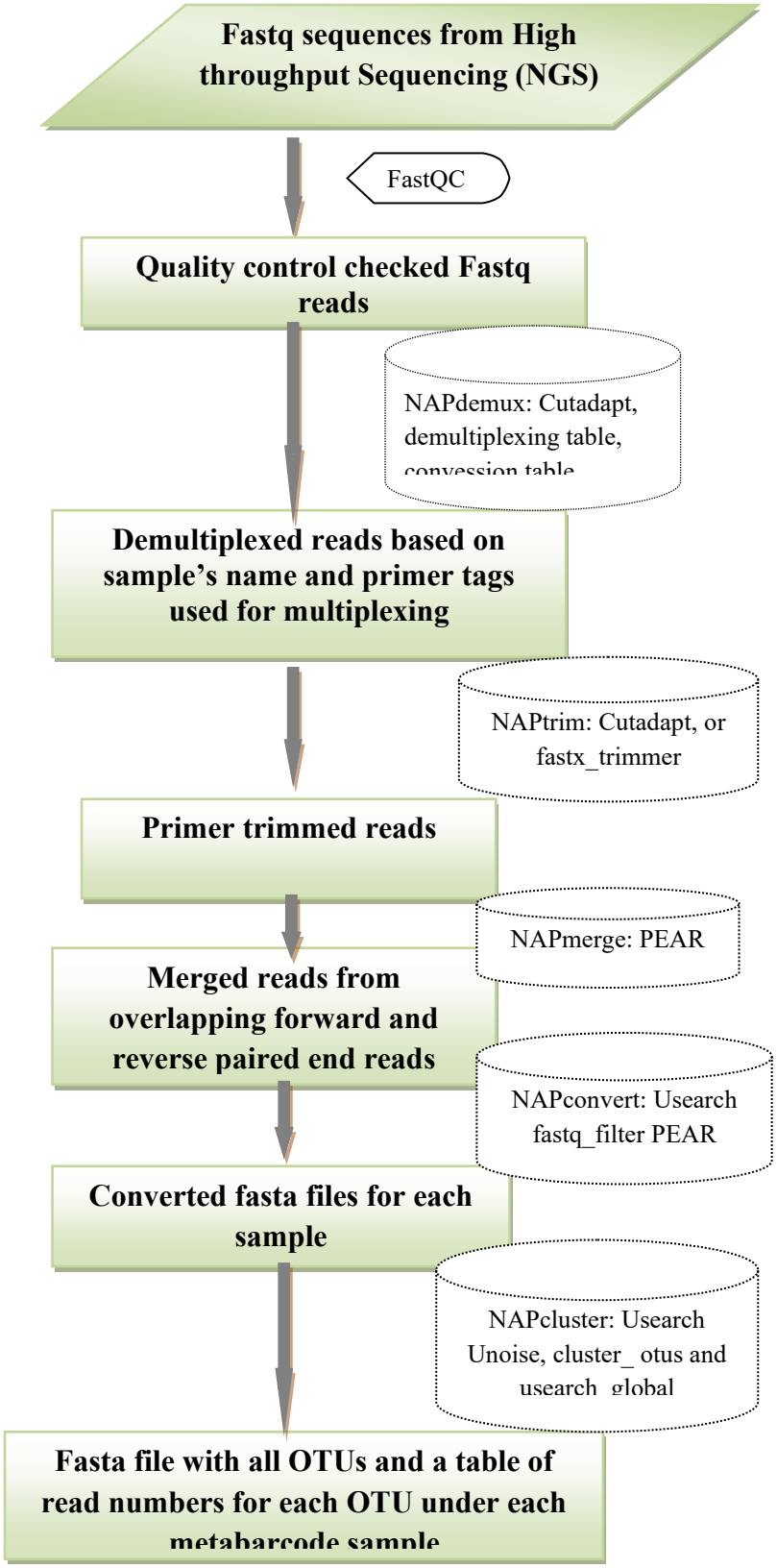

Figure 3. 6 Metabarcoding pipeline for the study of macroinvertebrate

3.3. 10 Data Analysis

Identities of each OTU representative were established with the lowest common ancestor (LCA) method in MEGAN Community Edition using the Lowest Common Ancestor methodology (Huson et al., 2016). As amplification of bulk samples of macroinvertebrates was performed with universal primers, the metabarcoding pipeline produced a remarkable amount of unwanted OTUs (e.g., bacteria, fungi, algae, vertebrates and from different invertebrate phyla). Firstly, OTUs of non-target groups (e.g bacteria, fungi, vertebrates) were removed from the dataset retaining only invertebrate OTUs. Furthermore, non-target invertebrates and even arthropods that were not assigned to target taxa were excluded from the dataset. OTUs assigned to Coleoptera, Diptera, Ephemeroptera, Plecoptera, Trichoptera, Hemiptera and Odonata were retained for final analysis. Data for each complex sample in the OTUs-reads output table and MEGAN-run OTU taxonomy table were used for further macroinvertebrate diversity studies in different experiments. Statistical analyses were conducted in R (R Core Team, 2018).

I evaluated the changes in diversity metrics of macroinvertebrate taxa within each sample preparation method across sampling sites fitting the generalized linear models (GLM) with Poisson error distribution and log-link function using the 'glm' function of the glm2 package. This included sample preparation methods/treatments (Ludox, Flotation and Raw) as predictor variables and the total number of OTUs and taxa-wise OTU richness as response variables in the separate model. Data dispersion test for the model was also done with the 'dispersiontest' function of the AER package along with the calculation of dispersion index from degrees of freedom and residual deviance of the model. I also assessed the effects of the three preparation methods on total reads counts across the sampling sites with negative binomial error distribution and log-link function of GLM. Further, GLMs with Gaussian error distribution with identity-link function were also fitted to assess the changes in abundances (percentage of reads as proxy) of different taxa. Model assumption and goodness of fit was checked with diagnostic graphs (Normal Q_Q plot) of residuals (function: autoplot, package: ggfortify) and the significance of the components was tested using F statistics. In addition, the Shannon diversity index (Shannon and Wiener, 1963) was also calculated for all samples of three treatments and a one-way ANOVA test was done with these index values to compare the outputs of treatments followed by a posthoc test (TukeyHSD) to detect the individual effect of each preparation technique. Taxonomic diversity and read based abundance of

OTUs from the three sample preparation techniques were visualised by Heat Trees using the metacoder package of R (Foster, Sharpton and Grunwald, 2017).

To investigate the relevance of taking multiple technical replicates from homogenate macroinvertebrate samples, OTU diversity and abundance of different taxa in three technical replicates were evaluated calculating the Shannon diversity index of replicate samples. A one-way ANOVA test with Shannon index values was performed to assess their significant differences in replicates followed by a posthoc test (TukeyHSD) to detect the individual contribution of each replicate. To compare the diversity properties of replicate samples of each river, species accumulation curves (SAC) were derived with the random method using the 'specaccum' function of the vegan package (Oksanen et al., 2017). SACs can also be used to indicate the adequacy of replicate numbers in representing the macroinvertebrate fauna in a particular homogenate sample. In addition, Venn diagrams were prepared using 'venn.diagram' function of the VennDiagram package to illustrate the number of unique and shared OTUs between the three replicate samples.

To explore the effect of pooled and unpooled DNA samples (before PCR amplification) on final diversity outputs from the metabarcoding pipeline, the significance of dissimilarity/turnover between complex pooled and separately processed samples were tested using GLM fitting the negative binomial distribution with the log-link function. The model assumption, residual analysis and data dispersion were also checked following the above-stated methods. Dissimilarity/association of samples was visualized with nonmetric multidimensional scaling (NMDS) ordination (Kruskal, 1964) using the vegan 'metaMDS' function. Further, a ranked dissimilarity based ANOSIM (The ANalysis Of SIMilarity) test was performed using the 'anosim' function of the vegan package. This test compares the mean rank within groups to the mean rank between groups.

I chose to test the performance of two commonly used clustering algorithms (Usearch and Swarm) for OTU generation. I evaluated OTU richness and diversity calculating the Shannon diversity index of samples and the Wilcoxon signed-rank test was performed to compare the outputs between two clustering methods as this data was not treated as normally distributed (Shapiro-Wilk test of normality, W = 0.95796, p-value = 0.03741). Heat trees were also made

to visualise the taxonomic diversity and read-based abundance of explored OTUs from both clustering methods using an R package ‘metacoder’ (Foster, Sharpton and Grunwald, 2017).

To test the universal primer’s efficacy for metabarcoding macroinvertebrates in this study, I compared the PCR success rate of COI barcodes of different macroinvertebrate taxa. Percentage data of amplification success from triplicate PCR was calculated and checked for normality using the Shapiro-Wilk test ($W = 0.8105$, p-value = 0.0001) and this data was not treated as normally distributed in comparisons. Therefore, I compared percentage data of amplification rate of different orders of Insecta, Crustacea, Mollusca and Annelida using a Kruskal Wallis test. Further, the Dunn test, a post-hoc analysis was performed to determine which variables (taxon) differ from each other. In addition, both forward and reverse primers sequences were mapped on multiple mitogenome sequences of said taxa to check the annealing positions of primers with their nucleotides matches/specificity on an online Primer Map platform (https://www. bioinformatics. org/sms2/ primer_map.html).

3.4 Results

3.4.1 Testing techniques of bulk samples processing for metabarcoding of macroinvertebrates

Three different techniques were tested in using bulk samples of macroinvertebrates to adopt the effective sample processing method for large-scale bioassessment with maximum outputs. These three techniques included: i) flotation with Ludox solution (referred to as Ludox in the text) ii) flotation in water (Float) and iii) usage of raw samples (Raw). After filtering and bioinformatic processing, a total of 8,31,636 reads were obtained from 18 metabarcodes libraries (18 samples) and each sample contained 46202 reads (COI 418 bp in length and with >5 copies) on average. OTU clustering with Usearch (cluster_otu and usearch_global) at a 3% threshold produced 577 OTUs across the entire dataset. After the taxonomic assignment of OTUs by MEGAN, OTUs of only target taxa (Coleoptera, Ephemeroptera, Trichoptera, Plecoptera, Diptera, Hemiptera and Odonata) were retained for further analysis. These taxa finally contained 220 (38.12 %) OTUs and 687784 (82.70 %) reads for all samples under three preparation techniques.

3.4.1a Generation of reads (sequences) and OTUs from three techniques

The number of reads from the metabarcoding pipeline ranged from 22904 to 51729 across the sites and sample preparation methods. Comparison of generated reads across the methods (Ludox, Float and Raw) showed that average read number varied within a range of 37966 to 39132 where Ludox produced the highest average number of reads followed by Float and Raw that was also supported by the median read numbers of three processing methods or treatments (Fig. 3.7). The percentage of reads assigned to all taxa was 34.15%, 32.74%, and 33.12% for the Ludox, Float and Raw samples respectively. Conversely, the lowest number of OTUs was produced from the Ludox method whereas Float and Raw methods returned a nearly equal number of OTUs (Fig. 3.8). The average OTUs numbers generated from Ludox, Float and Raw were estimated as 118 (SD=15), 127 (SD=15) and 128 (SD=12) respectively.

Overall, in terms of OTU generation, the Float and Raw samples of bulk macroinvertebrates were found to produce nearly similar outputs from three preparation techniques though the highest number of reads was estimated from Ludox.

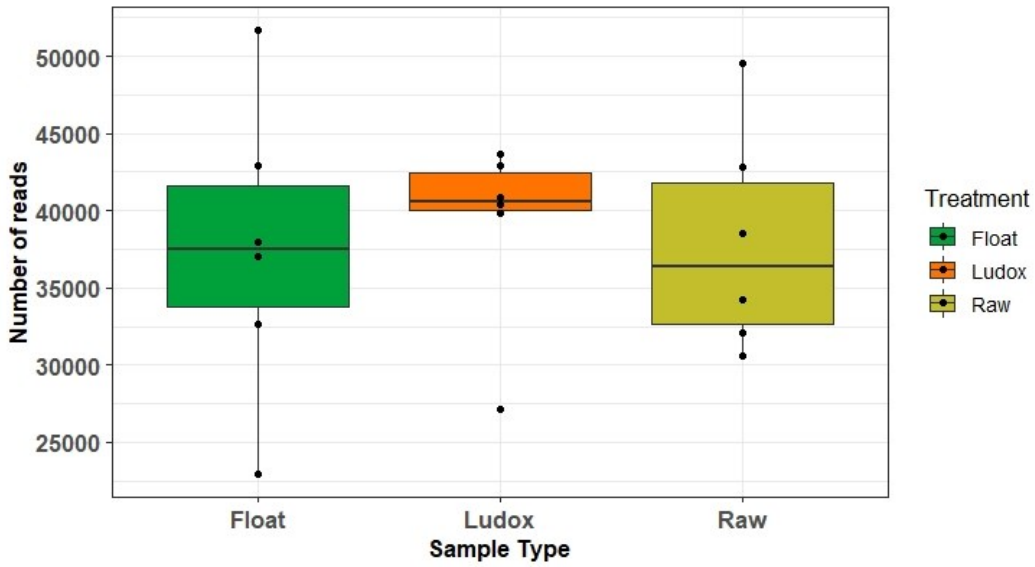

Figure 3. 7 Number of total reads under three sample processing techniques across the sites

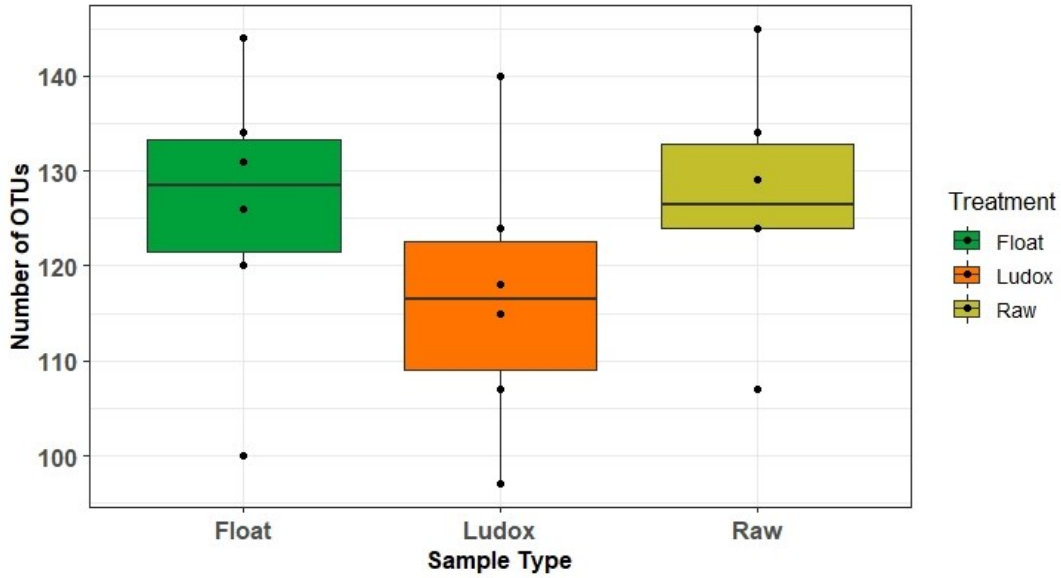

Figure 3. 8 OTU richness under three sample processing techniques across the sites

3.4.1b OTU richness in target groups of macroinvertebrate taxa from three techniques

Variation in OTU richness of target taxa including their abundances (reads as proxy) was also investigated for the three methods across the sites. Heat Trees were constructed with all target taxa estimating OTUs and read numbers that revealed a nearly similar pattern for three sample processing techniques (Fig. 3.9). The average number of Dipteran (53±4) OTUs per

sample was higher in all three treatments followed by Ephemeroptera (24±2), Trichoptera (22±1), Coleoptera (13±2), Odonata (6±1), Hemiptera (5±1), and Plecoptera (2±0). It is worth noting that even the lower OTUs containing taxa (e.g., Hemiptera and Plecoptera) also showed similar richness between the three treatments. Variations in taxa abundance (based on read percentage) between samples of three treatments were also lower where Ephemeroptera (87.42±0.38) was dominant over Trichoptera (6.08±0.53), Diptera (3.24±0.10), Hemiptera (2.16±0.70), Odonata (0.70±0.19), Coleoptera (0.34±0.22) and Plecoptera (0.06±0.03). The abundance of all taxa (except Hemiptera) was almost uniform for all samples.

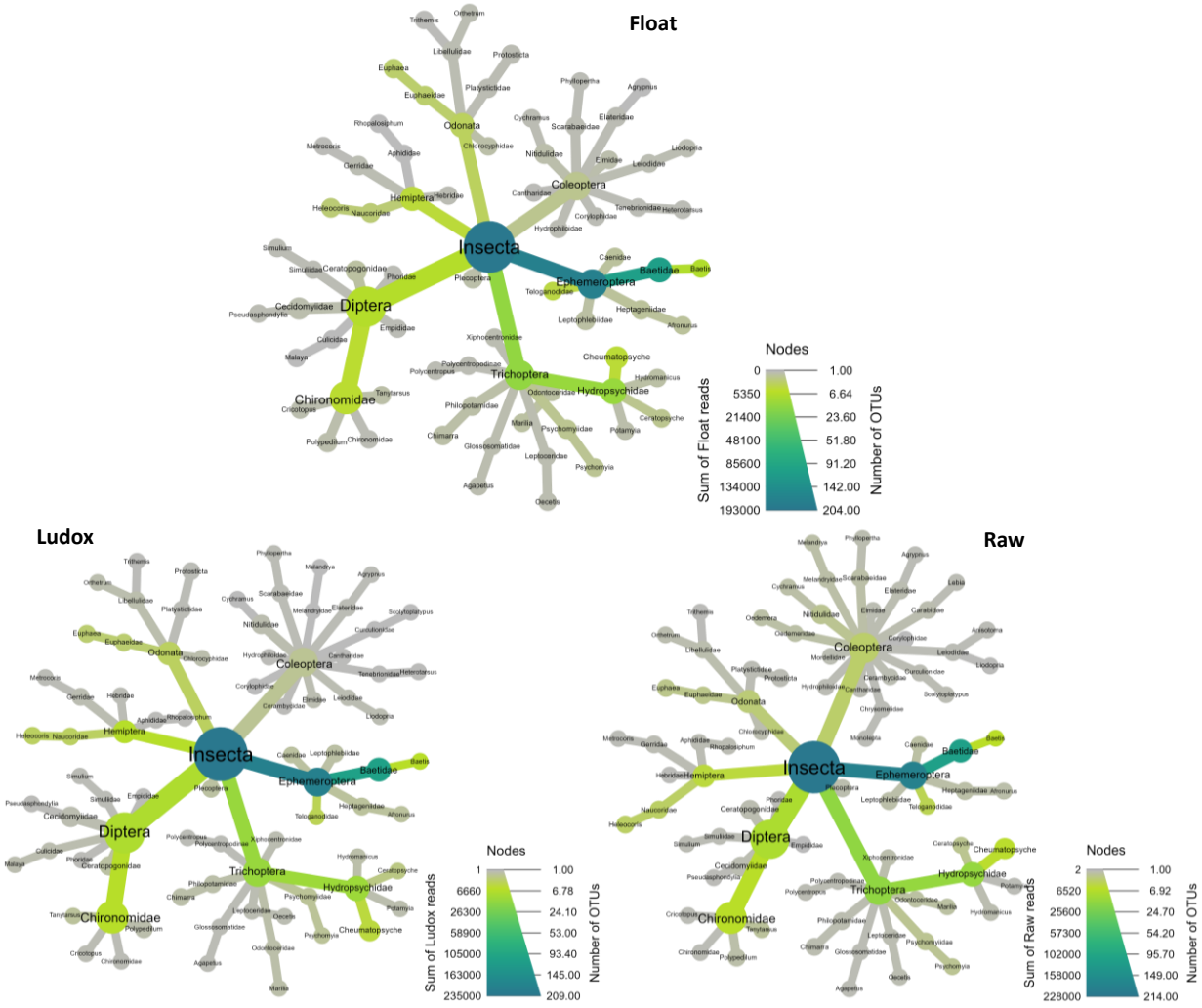

Figure 3. 9 Heat Trees showing taxa composition with OTU richness and read abundance in three sample processing techniques. Heat Tree is a type of taxonomic tree in which each node (the circles) is a taxon and the edges (lines) show hierarchical relationships between taxa. Sizes of circles denote the number of OTUs, and the colour codes indicate the number of reads assigned under each taxon.

3.4.1c Statistical analysis of sample preparation effects on diversity outputs

Statistical analysis of sample preparation effects was conducted by fitting generalized linear models using various diversity measures as response variables. All models identified no significant differences in diversity outputs between Ludox, Float and Raw samples (Table 3.1). A one-way ANOVA test with Shannon index values (as response variable) of Ludox, Float and Raw samples also revealed that the diversity of macroinvertebrates did not significantly differ between samples prepared by three different techniques (F=0.058, df=2,15, p= 0.944).

Table 3. 1 Generalised linear models with different distribution families (Poisson, Negative Binomial and Gaussian) fitted for respective diversity measures (response variables) with three levels of Treatment (Ludox, Float and Raw) as predictor variables.

Response variable	Predictor variable	Family	F Statistics	df	p-value
Total OTU count	Treatment (Levels-Ludox, Float and Raw)	Poisson	0.9398	2,15	0.4125
Ephemeroptera OTU counts	Treatment (Ludox, Float and Raw)	Poisson	0.0776	2,15	0.9257
Trichoptera OTU counts	Treatment (Ludox, Float and Raw)	Poisson	0.8435	2,15	0.4496
Coleoptera OTU counts	Treatment (Ludox, Float and Raw)	Poisson	0.8845	2,15	0.4334
Diptera OTU counts	Treatment (Ludox, Float and Raw)	Poisson	1.7832	2,15	0.2019
Plecoptera OTU counts	Treatment (Ludox, Float and Raw)	Poisson	0.01	2,15	0.99
Hemiptera OTU counts	Treatment (Ludox, Float and Raw)	Poisson	0.2857	2,15	0.7555
Odonata OTU counts	Treatment (Ludox, Float and Raw)	Poisson	1.0153	2,15	0.3859
Total reads count	Treatment (Ludox, Float and Raw)	Negative binomial	0.062	2,15	0.9402
Ephemeropteran abundance	Treatment (Ludox, Float and Raw)	Gaussian	0.0285	2,15	0.9719
Trichopteran abundance	Treatment (Ludox, Float and Raw)	Gaussian	0.2368	2,15	0.792
Coleopteran abundance	Treatment (Ludox, Float and Raw)	Gaussian	1.2928	2,15	0.3034
Dipteran abundance	Treatment (Ludox, Float and Raw)	Gaussian	3e-04	2,15	0.9997
Plecopteran abundance	Treatment (Ludox, Float and Raw)	Gaussian	0.9027	2,15	0.4264
Hemipteran abundance	Treatment (Ludox, Float and Raw)	Gaussian	0.5292	2,15	0.5997
Odonata abundance	Treatment (Levels-Ludox, Float and)	Gaussian	0.4926	2,15	0.6206

3.4.2 Investigating the outcome of multiple replicates of homogenate bulk samples

The relevance of using multiple technical replicates of homogenate bulk samples for DNA extraction, PCR amplification and sequencing in the metabarcoding pipeline were tested. In each stream, three subsamples (of equal amount) from one homogenate bulk sample made three replicates. The final outputs in terms of generation of OTUs and the composition of target taxa were compared among the replicates. After final filtering and bioinformatic processing, 9 metabarcode libraries resulted in about 520157 reads for 9 samples from three streams where each sample comprised 57k reads on average. OTU clustering was done with cluster_otu and usearch_global of Usearch at a 3% threshold produced 904 OTUs. After the taxonomic assignment of OTUs by MEGAN, non-targeted OTUs (taxa) were removed and the remaining OTUs assigned to target taxa groups were retained for final analysis. A set of 7 taxa groups finally contained 381102 (73.27 %) reads assigned to 260 (28.76 %) OTUs in 9 samples.

3.4. 2a Species/OTUs recovery trend of replicate samples

OTUs (species) accumulation curve across three replicates in each river indicated that each additional replicate contributed only a small number of additional OTUs (Fig. 3.10).

3.4. 2b OTU richness and taxa composition in three replicates of each river

In the context of OTUs number assigned to selected taxa in three replicates within each river, there was a slight variation in three replicates for the total number of OTUs which ranged from 87 to 105 in Betchora, 113 to 118 in Cheihkhiyang and 94 to 105 in Sangukhiang river. Two replicate samples of the Cheihkhiang river contained an equal number of OTUs and their taxa composition was also similar to each other. Taxa-specific OTUs number were also nearly equal between replicates of each stream (Fig. 3.11). The estimation of reads-based abundance also showed a similar pattern among the replicates across the rivers (Fig. 3.12).

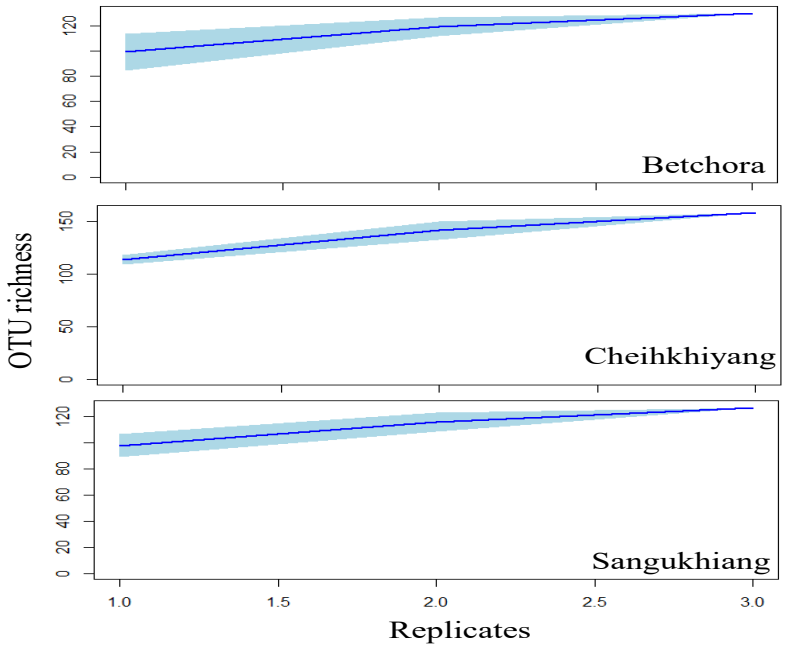

Figure 3. 10 Species Accumulation curve (made with specacum function of vegan package, method = random, not exact) for three replicates in each river (Upper- Betchhora, middle-Cheihkhiang and lower-Sangukhiang river)

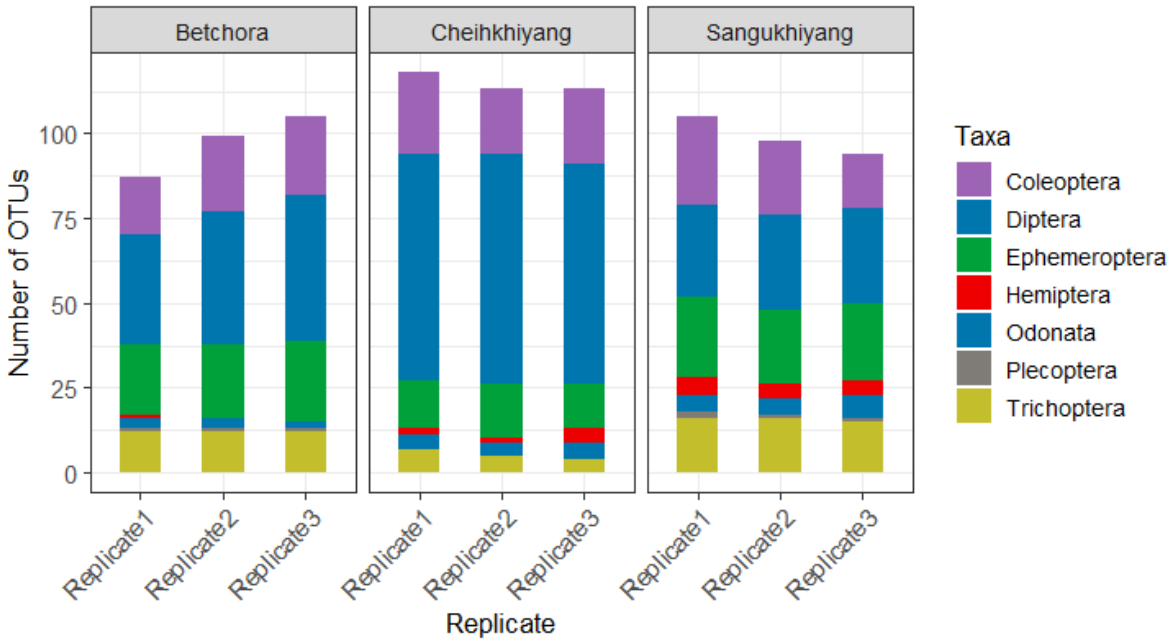

Figure 3. 11 OTUs richness in three replicates of each river under classified taxa group.

In the Betchhora river, three replicates produced 130 OTUs of which they shared 67 (52%) (Fig. 3.13). Thus, any single sample could lose 25 to 43 OTUs whereas the number of lost OTUs can be reduced to 8 to 17 using two replicates. For the Cheihkhiang river, three

replicate samples shared 81 (51%) of 159 OTUs, whereas a single sample underestimate was 41 to 46 OTUs. Conversely, using two replicates, an additional 23 to 28 OTUs were retained. The same trend was observed for the replicate samples of the Sangukhiang River as well as for three replicates across the rivers. Statistical analysis of a one-way ANOVA with Shannon diversity index of replicates showed that the mean index values did not differ significantly between replicates across the rivers ($F= 0.213$, $df=2,6$, $p=0.814$).

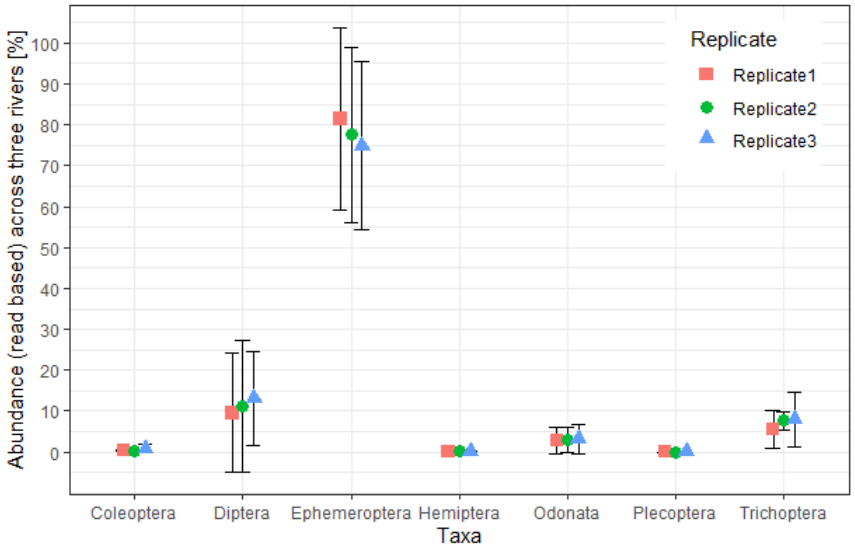

Figure 3. 12 Taxa wise abundance (read based) in three replicates across the rivers with only target taxa group (Coleoptera Ephemeroptera, Plecoptera, Trichoptera, Hemiptera, Odonata and Diptera).

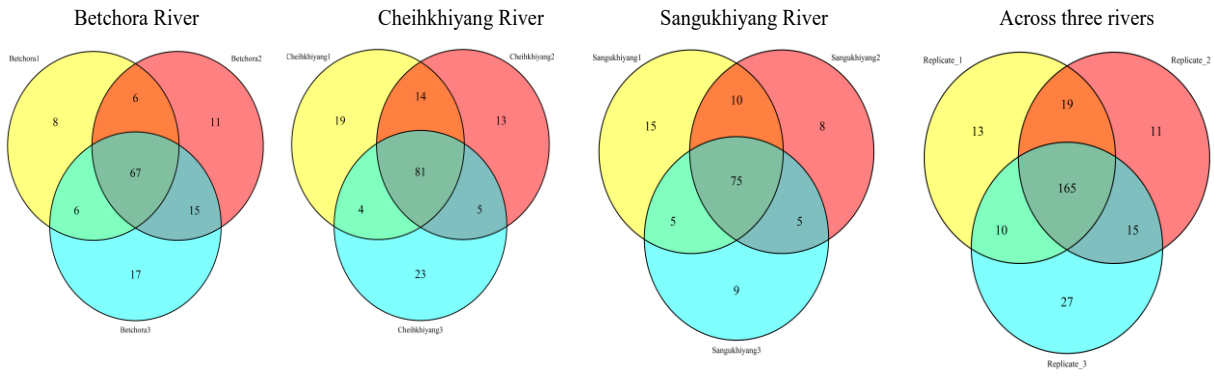

Figure 3. 13 Venn diagram showing the shared OTUs of targeted taxa among three replicates in each river (each colour represents a replicate).

3.4.3 Testing the effects of DNA pooling (before PCR amplification) and separately processed DNA

The outputs were compared from pooled DNA of two replicate samples and from the unpooled samples of the same replicates processed separately. The final outputs in terms of generation of total reads, number of OTUs, and the similarity and dissimilarity of OTUs were analysed between pooled DNA samples and unpooled DNA samples. In total, 30 metabarcode libraries (20 from separately processed samples and 10 from pooled DNA samples) were generated with 595228 reads containing 1046 OTUs. OTU clustering was done with cluster_otu and usearch_global of Usearch at a 3% threshold. After the taxonomic assignment of OTUs by MEGAN and subsequent exclusion of non-target taxa, the remaining 415 OTUs (40%) were classified under 7 taxa groups (Coleoptera, Ephemeroptera, Plecoptera, Trichoptera, Odonata, Hemiptera and Diptera) containing 436831 reads (73%) for final analyses.

Although the number of reads (range: 9812- 42200; average: 27416) in unpooled samples (mentioned here as SPDO-separately processed and combined DNA output) was expectedly higher than that (range: 5905-27265; average: 16266) of pooled samples (mentioned here as PDO- pooled DNA output), the rates of read generation were nearly equal in both treatment types. But surprisingly, the number of OTUs (range: 93-251; average: 182) were higher in pooled samples than that (range: 85-157; average 125) of SPDO samples (Fig. 3.14). From pair-wise comparison (SPDO sample with respective PDO sample) it was also obvious that in all pooled samples, OTUs were higher than separately processed samples. Of the ten pairs of samples, only in four pairs, the number of OTUs was close to each other but in other pairs, pooled samples contained remarkable higher numbers of OTU found in unpooled samples (Fig. 3.14). Statistical analysis of GLM (family=negative binomial with log-link function) with OTU richness also showed a significant difference ($F= 9.175$, $p= 0.007$) between two treatments (unpooled and pooled samples).

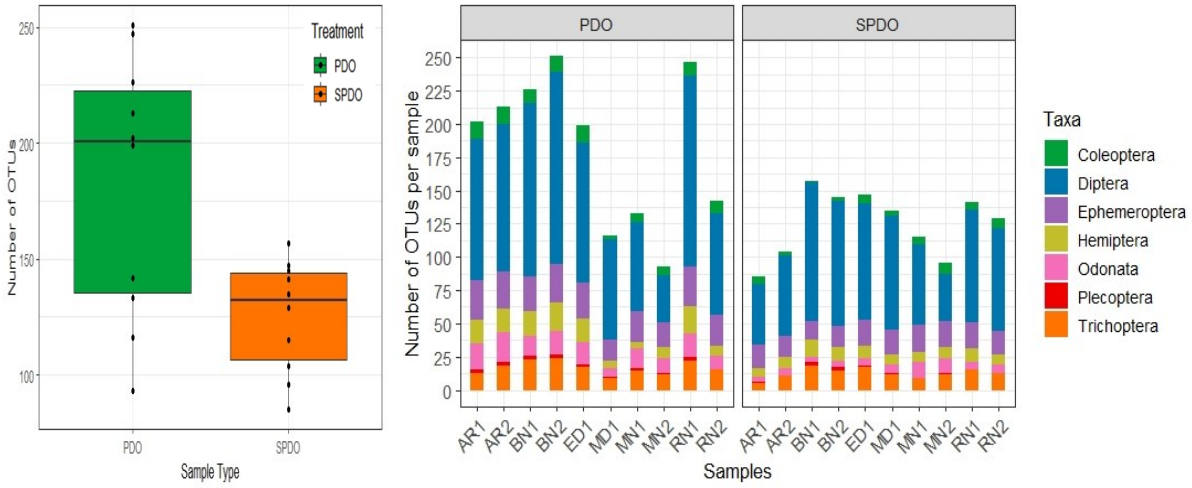

Figure 3. 14 Box plots for SPDO (separately processed and combined DNA output) and PDO (pooled DNA output) samples (Left). Right: Each pair of bars represent one SPDO sample (left bar) and their respective PDO sample (right bar).

Further, for measuring the dissimilarity between samples within pairs (between SPDO and PDO) and for putting all samples in a spatial configuration, Nonmetric Multidimensional Scaling (NMDS) was also performed. The NMDS plot (stress=0.04, distance=jaccard)) indicated that the OTU composition of pooled and unpooled samples of seven sites had a close association or less dissimilarity than those of the other three sites (AR1, RN1, BN2) of the rivers (Fig. 3.15). It also stated that the samples taken from the same rivers were placed together except for one river (MN2, Mongot River). A complementary statistical test of indirect gradient analysis (e.g., NMDS) called ANOSIM did not support significant rank-based dissimilarity between pooled and unpooled samples (R= -0.003333, p= 0.4558) (Fig. 3.16).

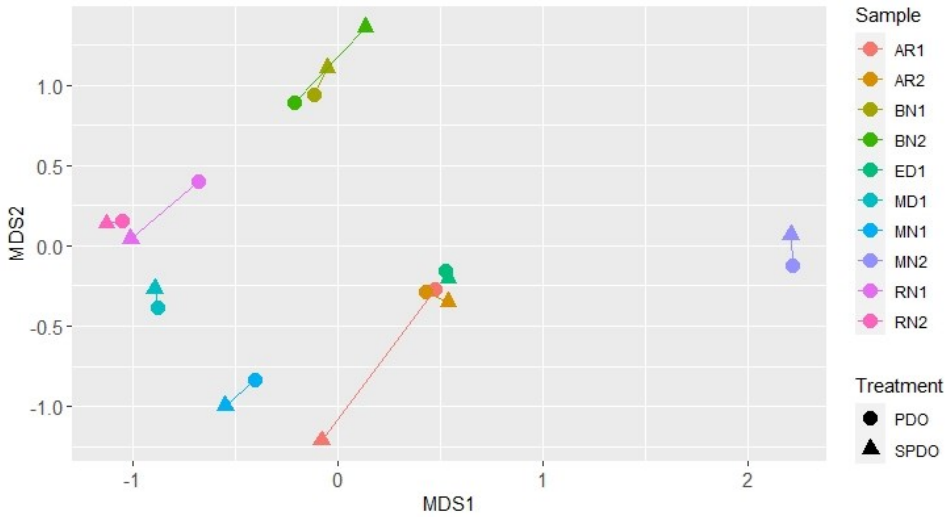

Figure 3. 15 NMDS plot (stress:0.04, distance: Jaccard) for pooled (PDO) and unpooled samples (SPDO) from ten sites of six rivers. Within each pair of samples one SPDO sample (marked with a triangle) and their respective PDO sample (marked with a circle). Each colour represents a sample site that contained a pair of samples.

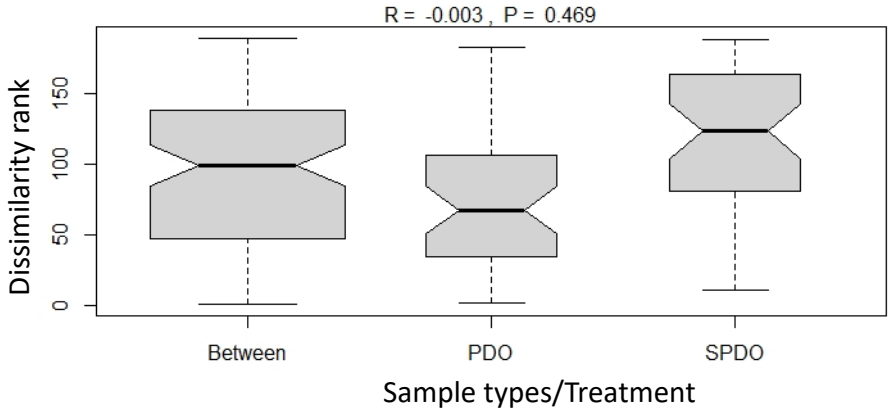

Figure 3. 16 ANOSIM plot (distance: Jaccard) for pooled (PDO) and unpooled samples (SPDO) from ten sites of six rivers.

3.4.4 Comparing outputs from two different clustering methods (Usearch and Swarm) in metabarcoding of macroinvertebrate samples

To test the effects of different clustering methods on OTUs or species estimation in the metabarcoding pipeline, the final outputs in terms of generation of OTUs from Usearch and Swarm clustering methods were compared for 30 composite samples from six rivers. OTU clustering was done with Usearch and Swarm clustering algorithm for entire sets of samples separately and then their outputs were assessed. In total, 30 metabarcode libraries were

generated with around 5.5 million reads for 30 composite samples from both clustering methods where Usearch contained a slightly higher number of reads. However, the number of total OTUs produced from Usearch and Swarm was estimated as 3211 and 3138 respectively. Taxonomic assignment of OTUs by MEGAN and subsequent exclusion of non-target OTUs, all samples retained around 3.95 (71%) and 3.39 million reads (62%) with 895 (27.87%) and 874 OTUs (27.85 %) in Usearch and Swarm clustering method respectively.

OTU clustering by Usearch and Swarm algorithm generated mostly similar results per sample across all rivers. The estimation of OTU richness slightly varied among rivers but the equal variation trend was for Usearch and Swarm. The average number of OTUs was estimated from Usearch and Swarm respectively as 212 and 209 in Cheihkhiang River; 192 and 188 in Sangukhiang; 173 and 170 in Betchhora; 131 and 130 in Buriganga; 140 and 135 in Dhaleshwari; 153 and 149 in Turag. The estimation of taxa composition in each river also showed variation in their OTU numbers among six rivers but with similar patterns produced from both clustering methods. Taxa-wise OTU estimation per sample across all rivers also revealed almost equal OTU counts for Usearch and Swarm algorithm. All samples were dominated by Diptera followed by Coleoptera Ephemeroptera, Hemiptera, Odonata, Trichoptera and Plecoptera in both clustering outputs with an almost equal number of OTUs (Fig. 3.17).

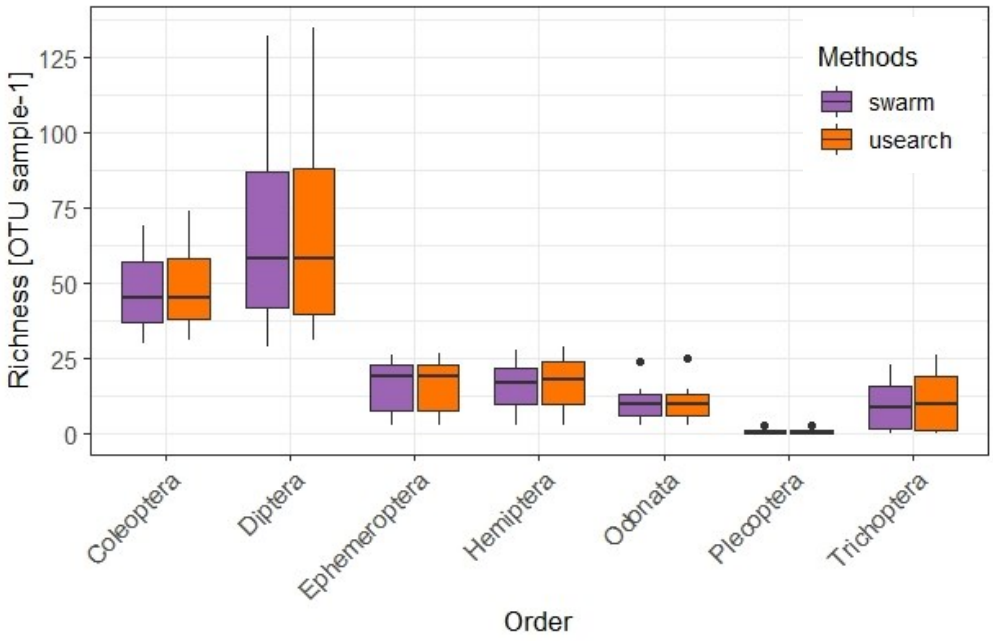

Figure 3. 17 OTU richness estimated from Usearch and Swarm (marked with different colours) clustering methods in 6 upland and lowland rivers. Each box plot included data from 5 sampling sites of each river.

Variation in OTU richness of target taxa including their abundances (reads as proxy) and the lower-level taxonomic assignment was also visualized with Heat trees that also revealed a nearly similar pattern in diversity outputs from two clustering algorithms (Fig. 3.18). A Mann-Whitney test indicated that there was no significant difference between Usearch and Swarm methods (W=380, p=0.5367) for exploring OTUs of target taxa groups. In addition, the Wilcoxon Rank Sum Test with Shannon diversity index did not show significant differences between outputs from Usearch and Swarm methods (W = 447, p= 0.9707) (Fig. 3.19). Overall, estimation of OTU richness in each river, OTU richness under selected taxa per sample and taxa composition and abundance (based on read number) in each river from two clustering methods provided congruent results except for a few incidences of the higher number of reads and OTUs generated by Usearch.

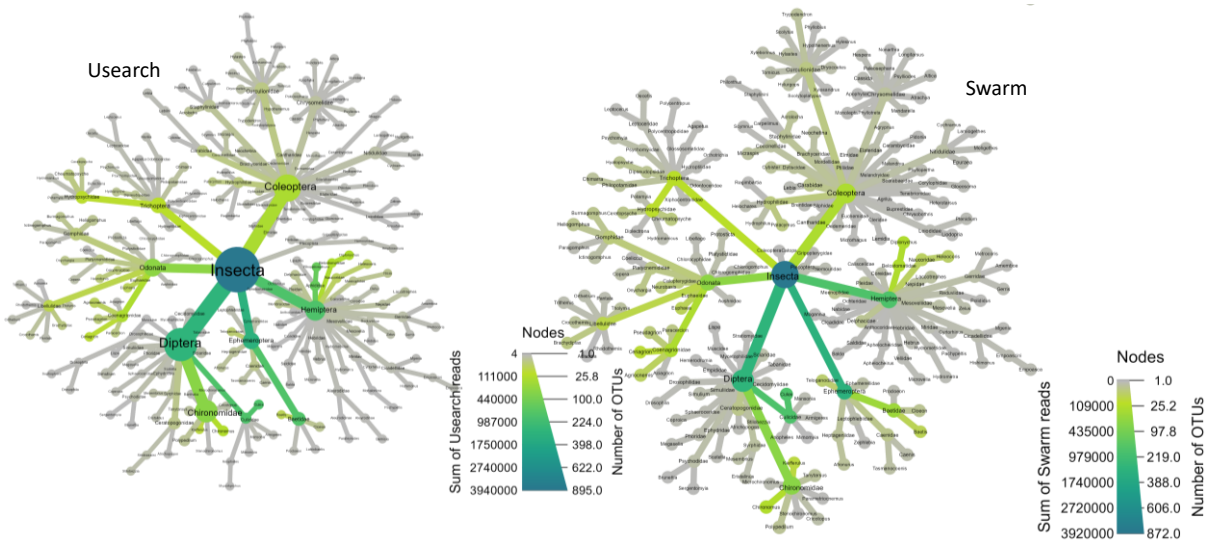

Figure 3. 18 Heat Trees showing taxa composition with OTU richness and read abundance of two clustering methods (Usearch and Swarm). In these taxonomic trees, each node (the circles) is a taxon and the edges (lines) show hierarchical relationships between taxa. Sizes of circles denote the number of OTUs and the colour codes indicate the number of reads assigned under each taxon.

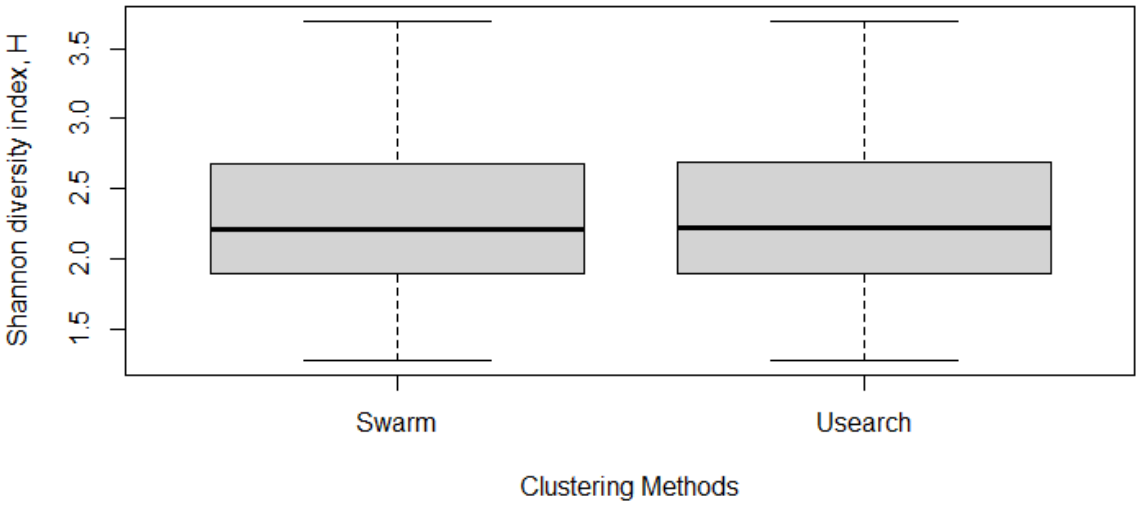

Figure 3. 19 Shannon diversity index values estimated from 30 macroinvertebrate samples using Usearch and Swarm clustering methods. Each box plot included data from 30 samples of 6 rivers.

3.4.5 Primer efficiency for metabarcoding of macroinvertebrates

As this study included a wide range of taxa for metabarcoding of macroinvertebrates (Chapter 4), the selected universal primer was tested in PCR amplification of COI barcode of individual specimens of target taxa for building barcode library (Fig. 3.21). In general, amplicon bands on agarose gel were quite indicative for the primer that fairly amplified the taxa of arthropods including Coleoptera, EPT, Hemiptera and Odonata and Diptera except for Mollusca and Annelida where the amplification rate was lower than that of Arthropoda taxa.

Based on the amplicons and barcodes number of morphospecies (Chapter 2), the amplification rate was over 90% for Coleoptera, Ephemeroptera, Trichoptera, Hemiptera, Odonata and Diptera. This rate slightly decreased for Plecoptera, and Crustacea compared to other arthropods but exceeded 84%. For Mollusca and Annelida, the amplification rate was accounted as 68.46% and 55.83 % respectively (Fig. 3.20). Statistical analysis of a one-way Kruskal-Wallis rank sum test showed that the amplification rates between major taxa groups (Insecta, Crustacea, Mollusca and Annelida) varied significantly ($KW = 17.979$, $p = 0.0004$). The post-hoc analysis (Dunn test) also confirmed this result but did not show significant variation in amplification rates for the insect Orders ($p > 0.05$). As expected, the barcoding

(barcode sequencing) rate of amplicons was less than amplification due to error of sequencing and bioinformatic processes, but it was more than 96% for Odonata, around 90% for Ephemeroptera, Trichoptera, Hemiptera and Diptera. As for the amplification rate, the barcoding rate (=successful sequence) also reduced to 60% and 68% for Annelida and Mollusca respectively (Fig. 3.23).

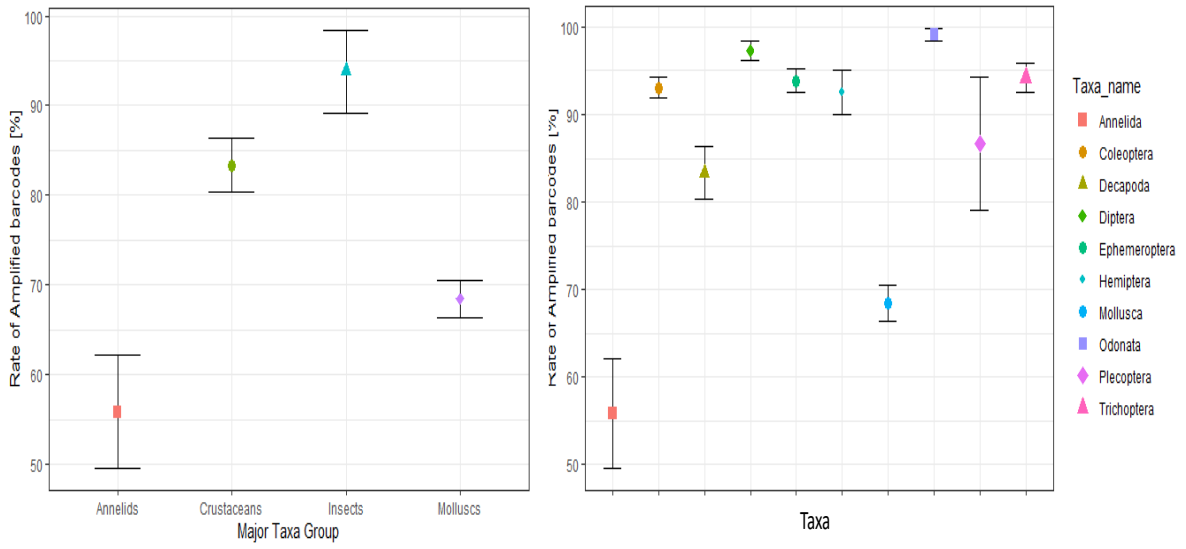

Figure 3. 20 Amplification success rate of major taxa group (left), and of different insect orders with annelids, crustaceans and molluscs (right). Different colours denote different taxa.

In metabarcoding, for PCR amplification of bulk complex samples of macroinvertebrates, the same primer was found quite effective across all samples (Figure 3.22) with a few exceptions. Although this effective amplification rate of bulk samples did not solely guarantee the final extraction of barcodes of all taxa present in the samples, however, the performance of the chosen primers for individual and bulk sample amplification indicated its potential for metabarcoding of macroinvertebrates. Moreover, morphospecies barcodes were also matched with metabarcodes to assess their inclusion rate (presence) in finally processed metabarcode samples. For this, we made a custom blast search with all obtained barcodes (taxa-wise) against metabarcodes samples. Blast search showed that most of the barcodes of morphospecies were present (at 99 to 100% similarity) in a metabarcode sample across all the taxa. The barcode inclusion rate of the taxa Coleoptera, Ephemeroptera, Trichoptera, Hemiptera, Odonata and Diptera in metabarcodes samples was higher than other groups and estimated at around 95%. Plecopteran and crustacean barcodes were identified in

metabarcodes sample with a rate of 85% and 82% respectively whereas the inclusion of molluscan and annelids barcodes was close to 80% (Figure 3.23).

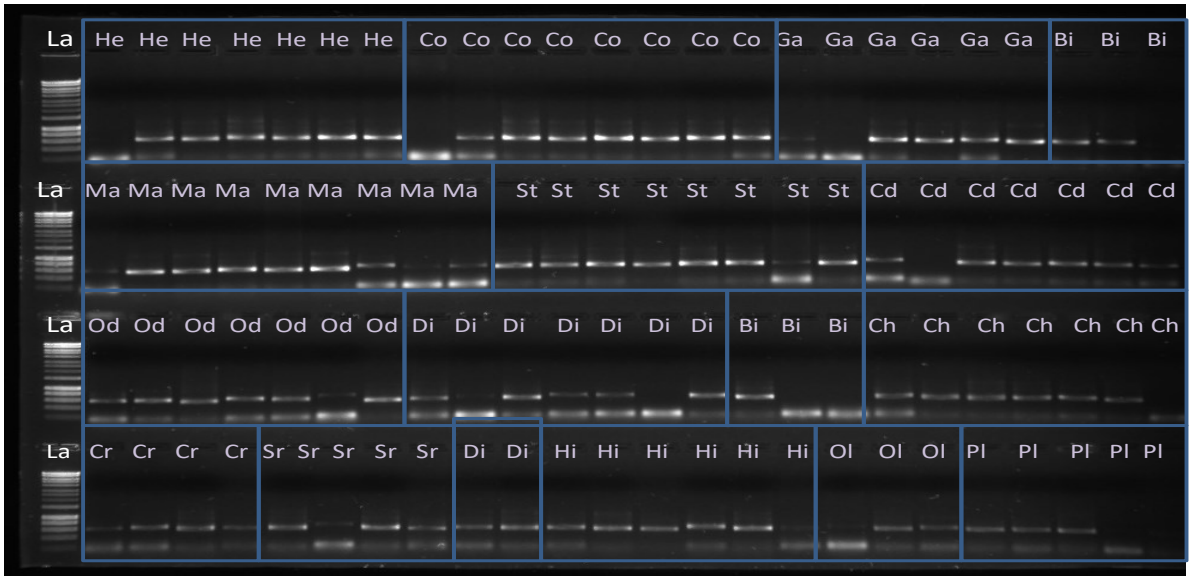

Figure 3. 21 COI amplicon bands of individual samples from different taxa of macroinvertebrates on the agarose gel. (He-Hemiptera; Co-Coleoptera; Ga-Gastropoda; BI-Bivalvia; Ma-Mayfly (Ephemeroptera); St-Stonefly (Plecoptera); Cd-Caddisfly (Trichoptera); Od (Odonata); DI (Diptera); Ch-Chironomidae; Cr-Crab (Decapoda); Sr-Shrimp (Decapoda); HI-Hirudinea; Ol-Oligochaeta; Pl-Polychaeta).

Figure 3. 22 COI amplicon bands of bulk samples sample of macroinvertebrates on an agarose gel.

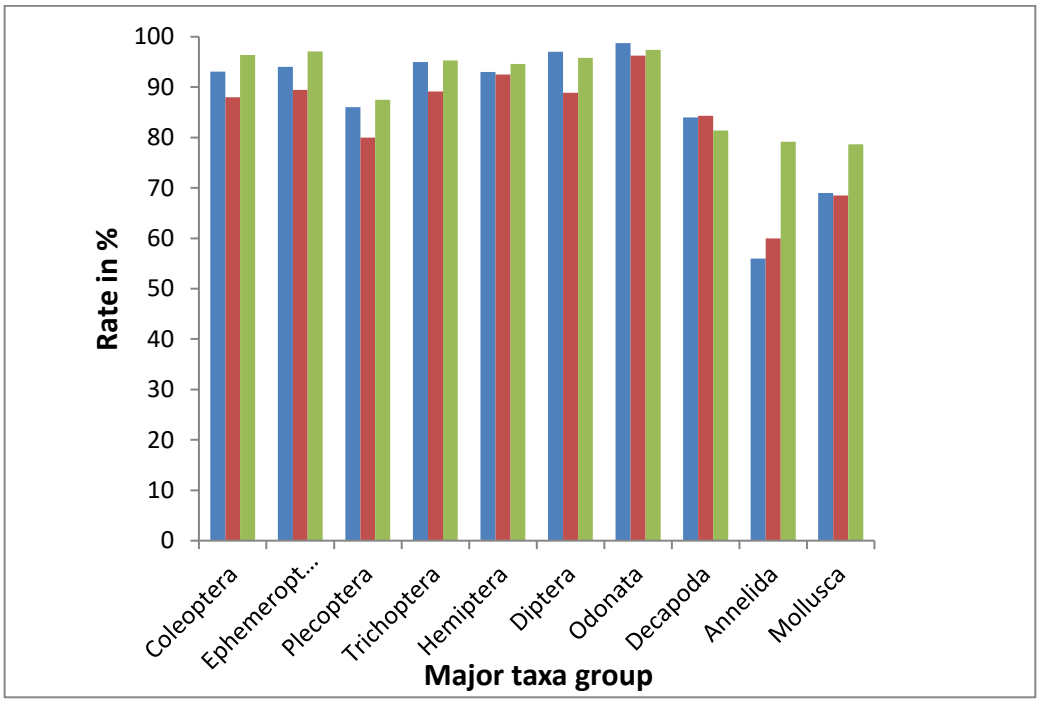

Figure 3. 23 Amplification, sequencing of amplicon and recovery of barcoded fragments from metabarcoding across the studied group of macroinvertebrates. Blue coloured bar shows the success rate of COI fragment amplified; the Purple coloured bar for COI barcodes obtained from sequencing and the green coloured bar shows the matching rate of barcodes with the barcodes from the metabarcoding pipeline.

3.5 Discussion

Metabarcoding of metazoan communities in freshwater, marine and terrestrial ecosystems has already gained wide acceptance to estimate biodiversity from complex bulk samples. Nonetheless, the application of this metagenetic technique for biomonitoring purposes requires the standardization of laboratory and bioinformatic procedures. In particular, preparation of bulk macroinvertebrate samples (often collected by kick-nets) for DNA extraction, optimizing the amount (number of technical replicates) of the homogenized sample, DNA pooling strategies and subsequent recovery of reliable diversity estimates from those complex samples are still challenging for large scale bioassessment programmes.

Preparation of bulk macroinvertebrate samples for DNA extraction

To address the sample preparation issue, I compared three sample preparation techniques (flotation with Ludox, flotation in water and using raw samples) to track their effectiveness for maximum biodiversity estimation taking issues of cost and time into account. The results showed that there was no significant variation among these techniques in terms of OTU generation and composition of target taxa. Therefore, metabarcoding with raw samples could be a prospective approach in large scale freshwaters biomonitoring to reduce the time and cost of sorting (size or taxonomy based) and chemical (e.g., using colloidal solution) processing of samples. The findings of the present study support a recent study where sediments and debris did not affect the detection of the target macroinvertebrate taxa in unsorted raw samples (Nichols et al., 2019) though that study was dealt with relatively less complex samples. Various measures or techniques are in practice to prepare complex samples including sample cleaning (removing debris manually or with chemical treatment), sorting with different size classes or taxonomic groups. However, sample processing techniques for metabarcoding of real-world complex freshwater macroinvertebrates samples are still less investigated especially in a time saving and cost-effective manner tested in the present study.

In most metabarcoding studies with soil and aquatic macroinvertebrates, samples were cleaned manually or chemically (Arribas et al., 2016) and then sorted based on different size classes of organisms (Elbrecht, Peinert and Leese, 2017; Creedy, Ng and Vogler, 2019). For instance, Arribas et al. (2016), used a flotation–Berlese–flotation (FBF) protocol for the

processing of arthropod mesofauna specimens from the soil. In addition, using Ludox colloidal solution (Lallias et al., 2015) and special flotation mechanical enrichment and homogenization techniques (Aylagus et al., 2016) have also been suggested for marine sediments sample for metabarcoding of macroinvertebrates and meiofauna. However, all of these techniques were compromised with additional time and costs associated with sample preparation for DNA extraction. In contrast, the usage of raw samples with minimum processing efforts was found to be a cost-effective, labour and time-saving technique in the present study. Although chemical treatment with Ludox colloidal solution was found effective to reduce unwanted elements, it was unlikely to separate every tiny and very delicate, immature organism of EPT and Diptera from a complex freshwater macroinvertebrate sample.

The kick net macroinvertebrate samples often contain the debris and other unwanted organic materials that are generally assumed to affect the extraction of quality DNA, and PCR amplification. In particular, the presence of non-target organic matter derived PCR inhibitors might influence the DNA metabarcoding success (Majaneva et al., 2018). Therefore, for DNA extraction from properly homogenised samples, I used the Qiagen Power Soil Kit which features a novel bead tube and optimized chemistry for more efficient lysis of soil organisms. This kit is also designed with Inhibitor Removal Technology (IRT) to eliminate the PCR inhibitors. Furthermore, using the optimised amount of BSA (Bovine serum albumin) in PCR reaction was also found effective for successful amplification. DNA metabarcoding of unsorted samples also rely on the choice of DNA extraction methods, especially with commercial kits which was revealed by a study with invertebrates (Majaneva et al., 2018) where the Qiagen PowerPlant Kit extractions resulted in the highest DNA yield and more repeatable estimation compared to other kits. Likewise, the Qiagen Power Soil Kit used in the current study also facilitated the DNA extraction from debris-mixed samples of macroinvertebrates.

Biomass biases from different-sized organisms are another concern in bulk community samples, especially larger organisms like odonates larvae, beetles, molluscs and water bugs can lead to overestimation while the low biomass organisms can be underestimated vice versa. Therefore, we have sorted larger organisms (larger individuals of >1.5 cm) for separate processing of the respective sample. Therefore, to minimize the variation of specimen

biomass in samples, only large organisms (>15 mm, e.g., adult beetles, dragonfly's nymph, snails, mussels, and water bugs) were separated and each large individual sample was prepared taking an equal amount of tissue from individuals (see Chapter 4) instead of samples with different size classes (Creedy, Ng and Vogler, 2019; Elbrecht, Peinert and Leese, 2017) and taxonomic group (Beentjes et al., 2019) of organisms. The optimization of sequencing depth of bulk amplified samples is also important for exploring rare species from unsorted samples. I aimed for deep sequencing (80000 reads per sample) of each bulk sample that seemed effective for sequencing rare species even with low biomass in raw samples. Although increased sequencing depth might facilitate sequencing frequencies of unwanted organisms, those sequences can be removed bioinformatically for the final estimation of biodiversity. For instance, stonefly species were rarely found during the sorting of morphospecies but they were recovered in metabarcode samples which was an indication of the potential of higher sequencing coverage of raw samples for reliable biodiversity estimation, but it also relies on proper primer selection. The requirement of deep sequencing of unsorted samples in metabarcoding pipeline was also implied by other studies (Alberdi et al., 2018; Elbrecht et al., 2021).

Therefore, the present study suggested the use of raw samples of macroinvertebrates (with minimum processing efforts, such as removing large debris, separating large organisms as individual samples) for biodiversity estimation at gradient sites of freshwater ecosystems. The result also implied that DNA metabarcoding with raw macroinvertebrate samples has the potential for bioassessment programmes by reducing the sample processing time and increasing the speed of macroinvertebrate species identification.

Optimization of the amount of homogenate samples (number of replicates)

The sample homogenisation and amount of homogenate of samples (number of replicates) for DNA extraction are also crucial in metabarcoding of macroinvertebrates when using unsorted raw samples. I did a small test taking multiple replicates (subsamples of the same amount) from the homogenised bulk sample to assess the community coverage of single, double and triple subsamples (replicates) in metabarcoding of macroinvertebrates. The results demonstrated the less impact of technical replication and had no statistically significant variation of community estimation between single, double and triple replicate samples.

Surprisingly, nearly stable (highest) species accumulation was found in the output of triple replicates. Using raw macroinvertebrate samples with proper homogenization were also found viable for tracking community changes across the gradient sites (Chapter 4).

Generally, an unsorted macroinvertebrate sample carries sediments, organic matter, plant debris and uneven specimens which need to be correctly processed so that the whole community can be represented from extracted DNA from that sample (Aylagas et al., 2016). In this case, it is a key challenge to ensure the DNA for representation of the whole community in every subsample which necessitates the homogenization of raw samples. It is evident that the isolation of organisms followed by homogenization allows a reliable characterization of the macroinvertebrate community through DNA metabarcoding (Aylagas et al., 2016). Therefore, mortar homogenization of macroinvertebrate samples was done after drying at ambient temperature which made homogenates for DNA extraction. Further, optimization of the number (amount) of homogenized samples is also essential for reducing the time and cost of sample processing. Although the best community characterization using metabarcoding would require the DNA extraction of the total homogenate sample, this cannot be achieved in a reasonable time because of its large volume even after proper homogenization. Moreover, commercial kits are generally designed for sample up to a certain volume (10-20 g) and this issue also concerns the higher cost of DNA extraction from an entire sample and its subsequent processing. In this case, an optimum number of technical replicates of DNA extractions on each homogenised sample and subsequent PCRs replication are important to maximise the outputs minimizing false (Ficetola et al., 2015; Willerslev et al., 2014). Aylagas et al. (2016) recommended performing two DNA extractions on two homogenized subsamples to return a reliable representation of the whole community. However, the effects of both technical and biological replication on final results are often overlooked in the metabarcoding of freshwater macroinvertebrates.

Studies of diet content metabarcoding revealed that technical replication affected the measures of diet descriptors (e.g., Alberdi et al., 2017; Pansu et al., 2015; Willerslev et al., 2014), though the impact of biological replication was much higher than that of technical replicates. Mata et al. (2019) also investigated the effects of technical and biological replication on metabarcoding gut contents and found that diet diversity increased steadily with the number of technical replicates of pellet samples. Investigation on replication effects

of genomic DNA extractions on marine sediment samples also demonstrated that DNA extraction replicates and sequencing depth could improve diversity measures of the benthic community (Lanzen et al., 2017). The results of the present study also apparently support all of these investigations, though the variation in diversity estimates of macroinvertebrates was not statistically significant in different replicates. Although the experiment was based on three homogenate samples from each of three streams that may be affected by some idiosyncrasies and limitations, this is unlikely to affect the generality of our conclusions to a significant extent. Therefore, this study may have implications for metabarcoding of raw macroinvertebrate samples in large-scale freshwater biomonitoring using a properly homogenised single subsample (replicate) that could be able to produce a maximum representation of community species.

The effects of pooling of extracted DNA before PCR

Pooling of bulk DNA or PCR product is also a concern for sample manipulation in macroinvertebrates metabarcoding because pooling of DNA or amplicon products from multiple replicates is required to reduce the cost of sequencing and also to save time for bioinformatic processing and subsequent analysis. But how does this approach affect the final estimation of diversity, in particular, the impact of Pre-PCR DNA pooling on final results is often poorly studied in macroinvertebrate metabarcoding studies. Here, I carried out an investigation assessing the final outputs of pooled DNA samples compared with the separately processed DNA output of those samples. The results demonstrated a striking result with higher OTU richness and taxa composition in pooled samples than in separately processed samples. This finding contrasted with a similar metabarcoding study where species recovery was efficient in separately processed leg samples of diverse arthropod mock communities (Braukmann et al., 2019). In some studies, pooling of samples before DNA extraction was done to reduce processing time and costs by combining multiple samples (Burgar et al., 2014; Clare et al., 2014; Jedlicka, Vo and Almeida, 2017), though this strategy has been opposed due to its substantial errors in the estimation of the diversity of fish and fungal communities (Sato et al., 2017; Wainer et al., 2020). For instance, in gut content metabarcoding, poor estimates of diet diversity and composition were evident in case of pooling of pellet samples before extraction (Mata et al., 2017).

This unexpected overestimation of diversity from pooled DNA samples suggests that the sequencing efficiency is increased with higher diversity in the sample. This phenomenon, if generally upheld, is difficult to understand based on the mechanics of the PCR that should broadly amplify from all templates in a similar way. However, composition of a metabarcoding sample is known to be important for the detection of all species present in a sample. An alternative is that the artefactual formation of chimeras in a more complex mixture would give rise to additional OTUs that are not filtered efficiently in the chimera removal steps. Another possibility is that my results might be further affected by triplicate PCRs that can trigger to produce False-positives in pooled samples because of PCR-induced artefacts and contamination (Ficetola et al., 2015). Therefore, the reproducibility of the present findings should be tested again as there is a level of stochasticity in high throughput metabarcoding experiments. However, the findings of the present study have implications for bulk sample manipulation in metabarcoding studies indicating the possibility of overestimated outputs.

The effect of clustering method on OTUs generation

To estimate species diversity in a complex sample, sequences are clustered into operational taxonomic units (OTUs), which are used as a proxy for species. Diversity estimates can vary greatly depending on the OTUs estimation methods (Bachy et al., 2013; Egge et al., 2013), and therefore, the selection of optimal procedures is valuable to ensure the reliability of OTU-based measurement for a particular study (Flynn et al., 2015). Several studies (Bannock and Halanych, 2015; Flynn et al., 2015; Xiong and Zhan, 2018) revealed that clustering outputs varied with different methods or algorithms and with the given threshold for quality filtering. These studies suggested the use of multiple clustering methods with carefully fixed parameters in particular for sequence divergence threshold for OTUs clustering. I used two different clustering algorithms (Usearch and Swarm) to optimize the number of exploited OTUs for biodiversity estimation from complex metabarcodes samples of freshwater macroinvertebrates. The result implied that both clustering algorithms were similarly effective in producing congruent results though Usearch clustering algorithm at 3% similarity was considered as an appropriate setting for OTU clustering.

Primer choice for freshwater macroinvertebrate metabarcoding

In this study, the mitochondrial cytochrome oxidase subunit I (COI) gene was selected for metabarcoding of freshwater macroinvertebrates as it is the most widely used marker in metabarcoding for the Metazoa (Andújar et al., 2018). In the context of amplification of COI, primer mismatches in the PCR step can prevent the amplification of certain taxa that result in the inaccurate detection of community composition (Deagle et al., 2014; Elbrecht and Leese, 2015). The selection of primer sets for the targeted ecosystem and taxonomic groups is still an unresolved concern albeit many COI primer sets are now available for metabarcoding of arthropods and the negative effects of primer bias are reduced through primer design incorporating primer degeneracy (Elbrecht and Leese, 2017). For example, Elbrecht and Leese (2017) developed four primer sets for metabarcoding of freshwater macroinvertebrates, but the efficiency rate still varied within four combinations. Recently, the usage of multiple primer sets for COI metabarcoding has also been suggested to recover a higher richness of macroinvertebrate taxa from freshwater ecosystems (Hajibabaei et al., 2019). In fact, there is no panacea primer set to recover all taxa from the community bulk sample. To maximise amplification of a diverse set of target sequences, I used the degenerate primers III_B_F (Shokralla et al., 2015) and Fol_degen_rev (Yu et al., 2012) which worked fairly well for barcoding and metabarcoding of macroinvertebrate samples in particular for all arthropod taxa except crustaceans. The amplification and sequencing rate of molluscs and annelid taxa was found lower than arthropods. I also tested this primer set amplifying all target taxa that showed a 95% to 100% amplification rate for arthropod taxa. In the context of primer choice, this study suggested repeated testing with target taxa before using it in metabarcoding of complex macroinvertebrate samples.

3.6 References

Alberdi, A., Aizpurua, O., Gilbert, M.T.P. and Bohmann, K., 2018. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), pp.134-147.

Andújar, C., Arribas, P., Gray, C., Bruce, C., Woodward, G., Yu, D.W. and Vogler, A.P., 2018. Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular Ecology*, *27*(1), 146-166.

Andújar, C., Arribas, P., Yu, D.W., Vogler, A.P. and Emerson, B.C., 2018a. Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, 27(20), pp. 3968-3975.

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M. and Vogler, A.P., 2016. Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology and Evolution*, 7(9), pp.1071-1081.

Aylagas, E., Mendibil, I., Borja, Á. and Rodríguez-Ezpeleta, N., 2016. Marine sediment sample pre-processing for macroinvertebrates metabarcoding: mechanical enrichment and homogenization. *Frontiers in Marine Science*, 3, p.203.

Aylagas, E., Borja, Á., Muxika, I. and Rodríguez-Ezpeleta, N., 2018. Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. *Ecological Indicators*, *95*, pp.194-202.

Bachy, C., Dolan, J.R., López-García, P., Deschamps, P. and Moreira, D., 2013. Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *The ISME journal*, 7(2), pp.244-255.

Beentjes, K.K., Speksnijder, A.G., Schilthuizen, M., Hoogeveen, M., Pastoor, R. and van der Hoorn, B.B., 2019. Increased performance of DNA metabarcoding of macroinvertebrates by taxonomic sorting. *PloS one*, *14*(12), p.e0226527.

Beermann, A.J., Zizka, V.M., Elbrecht, V., Baranov, V. and Leese, F., 2018. DNA metabarcoding reveals the complex and hidden responses of chironomids to multiple stressors. *Environmental Sciences Europe*, *30*(1), pp.1-15.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N. and Ribera, I., 2012. The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, *61*(5), pp.851-869.

Beng, K. C., Tomlinson K. W., Shen, X. H. Surget-Groba, Y., Hughes, A. C., Corlett, R. T. and Slik, J. W. F. 2016. The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Scientific Reports*, *6*, p.24965.

Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R. and Thomas, W.K., 2012. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, *27* (4), pp.233-243.

Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. and Woodward, G., 2017. Next-generation global biomonitoring: large-scale, automated reconstruction of ecological networks. Trends in Ecology & Evolution, 32(7), pp.477-487.

Brandon-Mong, G.J., Gan, H.M., Sing, K.W., Lee, P.S., Lim, P.E. and Wilson, J.J., 2015. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. Bulletin of entomological research, 105(6), pp.717-727.

Brannock, P.M. and Halanych, K.M., 2015. Meiofaunal community analysis by high-throughput sequencing: comparison of extraction, quality filtering, and clustering methods. *Marine Genomics*, 23, pp.67-75.

Brannock, P.M., Wang, L., Ortmann, A.C., Waits, D.S. and Halanych, K.M., 2016. Genetic assessment of meiobenthic community composition and spatial distribution in coastal sediments along northern Gulf of Mexico. *Marine Environmental Research*, *119*, pp.166-175.

Braukmann, T.W., Ivanova, N.V., Prosser, S.W., Elbrecht, V., Steinke, D., Ratnasingham, S., de Waard, J.R., Sones, J.E., Zakharov, E.V. and Hebert, P.D., 2019. Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources*, *19*(3), pp.711-727.

Burgar, J.M., Murray, D.C., Craig, M.D., Haile, J., Houston, J., Stokes, V. and Bunce, M., 2014. Who's for dinner? High-throughput sequencing reveals bat dietary differentiation in a biodiversity hotspot where prey taxonomy is largely undescribed. *Molecular Ecology*, *23*(15), pp.3605-3617.

Bush, A., Monk, W.A., Compson, Z.G., Peters, D.L., Porter, T.M., Shokralla, S., Wright, M.T., Hajibabaei, M. and Baird, D.J., 2020. DNA metabarcoding reveals metacommunity dynamics in a threatened boreal wetland wilderness. *Proceedings of the National Academy of Sciences*, 117(15), pp.8539-8545.

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C.A. and Dawson, T.P., 2017. Connecting Earth observation to high-throughput biodiversity data. *Nat Ecol Evol, 1*, p.0176.

Capra, E., Giannico, R., Montagna, M., Turri, F., Cremonesi, P., Strozzi, F., Leone, P., Gandini, G. and Pizzi, F., 2016. A new primer set for DNA metabarcoding of soil Metazoa. *European Journal of Soil Biology*, *77*, pp.53-59.

Carew, M.E., Pettigrove, V.J., Metzeling, L. and Hoffmann, A.A., 2013. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in Zoology*, 10(1): 45.

Carrizo, S.F., Jähnig, S.C., Bremerich, V., Freyhof, J., Harrison, I., He, F., Langhans, S.D., Tockner, K., Zarfl, C. and Darwall, W., 2017. Freshwater megafauna: Flagships for freshwater biodiversity under threat. *Bioscience*, 67(10), pp.919-927.

Chain, F. J. J., Brown, E.A., MacIsaac, H. J., and Cristescu, M.E., 2016. Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. *Diversity and Distribution*, *22*, pp.493–504.

Clare, E.L., Symondson, W.O., Broders, H., Fabianek, F., Fraser, E.E., MacKenzie, A., Boughen, A., Hamilton, R., Willis, C.K., Martinez-Nuñez, F. and Menzies, A.K., 2014. The diet of Myotis lucifugus across C anada: assessing foraging quality and diet variability. *Molecular Ecology*, *23*(15), pp.3618-3632.

Clarke, L.J., Soubrier, J., Weyrich, L.S. and Cooper, A., 2014. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular ecology resources*, *14*(6), pp.1160-1170.

Comtet, T., Sandionigi, A., Viard, F. and Casiraghi, M., 2015. DNA (meta) barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions*, *17*(3), pp.905-922.

Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J. and Arnaud-Haond, S., 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PloS one, 10(2), p.e0117562.

Creedy, T.J., Ng, W.S. and Vogler, A.P., 2019. Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. *Ecology and Evolution*, *9*(6), pp.3105-3116.

Creedy, T., Andujar, C., Meramveliotakis, E., Noguerales, V., Overcast, I., Papadopoulou, A., Morlon, H., Vogler, A., Emerson, B. and Arribas, P., 2021. Coming of age for COI metabarcoding of whole organism community DNA: towards bioinformatic harmonisation. *Molecular Ecology Resources, 22*(3), pp.847-861.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., Potter, C. and Bik, H.M., 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, 7(9), pp.1008-1018.

Creer, S., Fonseca, V.G., Porazinska, D.L., Giblin-Davis, R.M., Sung, W., Power, D.M., Packer, M., Carvalho, G.R., Blaxter, M.L., Lambshead, P.J.D. and Thomas, W.K., 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, *19*, pp.4-20.

Cristescu, M.E., 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution, 29*(10), pp.566-571.

Dabney, J. and Meyer, M., 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques*, 52(2), pp.87-94.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. and Taberlet, P., 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, 10(9), p.20140562.

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26: 2460–2461.

Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. BioRxiv, p.081257.

Egge, E., Bittner, L., Andersen, T., Audic, S., de Vargas, C. and Edvardsen, B., 2013. 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PloS one*, 8(9), p.e74371.

Ekrem, T., Willassen, E. and Stur, E., 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution*, *43*(2), pp.530-542.

Elbrecht, V. and Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PloS One*, *10*(7), p.e0130324.

Elbrecht, V., Peinert, B. and Leese, F., 2017. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution*, *7*(17), pp.6918-6926.

Elbrecht, V., Vamos, E.E., Meissner, K., Aroviita, J. and Leese, F., 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, 8(10), pp.1265-1275.

Elbrecht, V. and Leese, F., 2017. Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science*, *5*, p.11.

Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J.N., Coissac, E., Boyer, F. and Leese, F., 2016. Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ*, *4*, p.e1966.

Elbrecht, V., Bourlat, S.J., Hörren, T., Lindner, A., Mordente, A., Noll, N.W., Schäffler, L., Sorg, M. and Zizka, V.M., 2021. Pooling size sorted Malaise trap fractions to maximize taxon recovery with metabarcoding. *PeerJ,* 9, p.e12177.

Emilson, C.E., Thompson, D.G., Venier, L.A., Porter, T.M., Swystun, T., Chartrand, D., Capell, S. and Hajibabaei, M., 2017. DNA metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an environmental gradient. *Scientific Reports*, 7(1), pp.1-11.

Fernández, S., Rodríguez, S., Martínez, J.L., Borrell, Y.J., Ardura, A. and García-Vázquez, E., 2018. Evaluating freshwater macroinvertebrates from eDNA metabarcoding: A river Nalón case study. *PloS One*, *13*(8), p.e0201741.

Ficetola, G.F., Boyer, F., Valentini, A., Bonin, A., Meyer, A., Dejean, T., Gaboriaud, C., Usseglio-Polatera, P. and Taberlet, P., 2021. Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. *Molecular Ecology*, *30*(13), pp.3189-3202.

Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F. and Rayé, G., 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, *15*(3), pp.543-556.

Flynn, J.M., Brown, E.A., Chain, F.J., MacIsaac, H.J. and Cristescu, M.E., 2015. Toward accurate molecular identification of species in complex environmental samples: testing the

performance of sequence filtering and clustering methods. *Ecology and Evolution*, *5*(11), pp.2252-2266.

Fonseca, V.G., Carvalho, G.R., Nichols, B., Quince, C., Johnson, H.F., Neill, S.P., Lambshead, J.D., Thomas, W.K., Power, D.M. and Creer, S., 2014. Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. *Global Ecology and Biogeography*, *23*(11), pp.1293-1302.

Foster, Z.S., Sharpton, T.J. and Grünwald, N.J., 2017. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Computational Biology*, *13*(2), p.e1005404.

Fonseca, V.G., Sinniger, F., Gaspar, J.M., Quince, C., Creer, S., Power, D.M., Peck, L.S. and Clark, M.S., 2017. Revealing higher than expected meiofaunal diversity in Antarctic sediments: a metabarcoding approach. *Scientific Reports*, *7*(1), p.6094.

Galan, M., Pons, J.B., Tournayre, O., Pierre, E., Leuchtmann, M., Pontier, D. and Charbonnel, N., 2018. Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. *Molecular Ecology Resources*, *18*(3), pp.474-489.

Gibson, J.F., Shokralla, S., Curry, C., Baird, D.J., Monk, W.A., King, I. and Hajibabaei, M., 2015. Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. *PloS one*, *10*(10), p.e0138432.

Goldberg, C.S., Pilliod, D.S., Arkle, R.S. and Waits, L.P., 2011. Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. *PloS One*, *6*(7), p.e22746.

Haase, P., Pauls, S.U., Schindehütte, K. and Sundermann, A., 2010. First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: human error greatly lowers the precision of assessment results. *Journal of the North American Benthological Society*, *29*(4), pp.1279-1291.

Hajibabaei, M., Porter, T.M., Wright, M. and Rudar, J., 2019. COI metabarcoding primer choice affects the richness and recovery of indicator taxa in freshwater systems. *PloS one*, *14*(9), p.e0220953.

Hajibabaei, M., Porter, T.M., Robinson, C.V., Baird, D.J., Shokralla, S. and Wright, M.T., 2019. Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. *PloS one*, *14*(12), p.e0225409.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C. and Baird, D.J., 2011. Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PLoS One*, 6(4): e17497.

Hajibabaei, M., Spall, J. L., Shokralla, S. and Konynenburg, S., 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, 12:28.

Hajibabaei, M., Singer, G.A., Clare, E.L. and Hebert, P.D., 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC biology*, *5*(1), pp.1-7.

Hebert, P.D.N., Cywinska, A., BALL, S.L. and Dewaard, J.R. 2003. Biological identification through DNA barcodes. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 270:313–321.

Hirai, J., Kuriyama, M., Ichikawa, T., Hidaka, K. and Tsuda, A., 2015. A metagenetic approach for revealing community structure of marine planktonic copepods. *Molecular Ecology Resources*, *15*(1), pp.68-80.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.J. and Tappu, R., 2016. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Computational Biology*, *12*(6): e1004957.

Ivanova, N.V., Dewaard, J.R. and Hebert, P.D., 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, *6*(4), pp.998-1002.

Jedlicka, J.A., Vo, A.T.E. and Almeida, R.P., 2017. Molecular scatology and high-throughput sequencing reveal predominately herbivorous insects in the diets of adult and nestling Western Bluebirds (Sialia mexicana) in California vineyards. *The Auk: Ornithological Advances*, *134*(1), pp.116-127.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A. and Larsen, T.H., 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), pp.1245-1257.

Keck, F., Vasselon, V., Rimet, F., Bouchez, A. and Kahlert, M., 2018. Boosting DNA metabarcoding for biomonitoring with phylogenetic estimation of operational taxonomic units' ecological profiles. *Molecular Ecology Resources*, *18*(6), pp.1299-1309.

Krehenwinkel, H., Kennedy, S.R., Rueda, A., Lam, A. and Gillespie, R.G., 2018. Scaling up DNA barcoding–Primer sets for simple and cost efficient arthropod systematics by multiplex PCR and Illumina amplicon sequencing. *Methods in Ecology and Evolution*, *9*(11), pp.2181-2193.

Krehenwinkel, H., Wolf, M., Lim, J.Y., Rominger, A.J., Simison, W.B. and Gillespie, R.G., 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, *7*(1), pp.1-12.

Kress, W.J., García-Robledo, C., Uriarte, M. and Erickson, D.L., 2015. DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution*, *30*(1), pp.25-35.

Kruskal, J.B., 1964. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29(2), pp.115-129.

Kuntke, F., de Jonge, N., Hesselsøe, M. and Nielsen, J.L., 2020. Stream water quality assessment by metabarcoding of invertebrates. *Ecological Indicators*, *111*, p.105982.

Lallias, D., Hiddink, J.G., Fonseca, V.G., Gaspar, J.M., Sung, W., Neill, S.P., Barnes, N., Ferrero, T., Hall, N., Lambshead, P.J.D. and Packer, M., 2015. Environmental metabarcoding

reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *The ISME journal*, *9*(5), p.1208.

Lanzén, A., Lekang, K., Jonassen, I., Thompson, E.M. and Troedsson, C., 2016. High-throughput metabarcoding of eukaryotic diversity for environmental monitoring of offshore oil-drilling activities. *Molecular Ecology*, *25*(17), pp.4392-4406.

Lanzén, A., Lekang, K., Jonassen, I., Thompson, E.M. and Troedsson, C., 2017. DNA extraction replicates improve diversity and compositional dissimilarity in metabarcoding of eukaryotes in marine sediments. *PLoS One*, *12*(6), p.e0179443.

Leray, M., and Knowlton, N., 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proc. Natl. Acad. Sci.* USA.,11:2076–2081.

Lobo, J., Shokralla, S., Costa, M.H., Hajibabaei, M. and Costa, F.O., 2017. DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. *Scientific Reports*, *7*(1), p.15618.

Machida, R.J., Leray, M., Ho, S.L. and Knowlton, N., 2017. Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data*, *4*(1), pp.1-7.

Macher, J.N., Vivancos, A., Piggott, J.J., Centeno, F.C., Matthaei, C.D. and Leese, F., 2018. Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology Resources*, *18*(6), pp.1456-1468.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M., 2015. Swarm v2: Highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3, e1420.

Majaneva, M., Diserud, O.H., Eagle, S.H., Hajibabaei, M. and Ekrem, T., 2018. Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples. *Metabarcoding and Metagenomics*, *2*, p.e26664.

Marquina, D., Esparza-Salas, R., Roslin, T. and Rnquist, F., 2019. Establishing arthropod community composition using metabarcoding: Surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. *Molecular Ecology Resources*, *19*(6), pp.1516-1530.

Martins, F.M., Galhardo, M., Filipe, A.F., Teixeira, A., Pinheiro, P., Paupério, J., Alves, P.C. and Beja, P., 2019. Have the cake and eat it: Optimizing nondestructive DNA metabarcoding of macroinvertebrate samples for freshwater biomonitoring. *Molecular Ecology Resources*, *19*(4), pp.863-876.

Mata, V.A., Rebelo, H., Amorim, F., McCracken, G.F., Jarman, S. and Beja, P., 2019. How much is enough? Effects of technical and biological replication on metabarcoding dietary analysis. *Molecular Ecology*, *28*(2), pp.165-175.

Morinière, J., Cancian de Araujo, B., Lam, A.W., Hausmann, A., Balke, M., Schmidt, S., Hendrich, L., Doczkal, D., Fartmann, B., Arvidsson, S. and Haszprunar, G., 2016. Species

identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. *PloS one*, *11*(5), p.e0155497.

Mortágua, A., Vasselon, V., Oliveira, R., Elias, C., Chardon, C., Bouchez, A., Rimet, F., Feio, M.J. and Almeida, S.F., 2019. Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. *Ecological Indicators*, *106*, p.105470.

Nichols, R.V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M., Green, R.E. and Shapiro, B., 2018. Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*, *18*(5), pp.927-939.

Nichols, S.J., Kefford, B.J., Campbell, C.D., Bylemans, J., Chandler, E., Bray, J.P., Shackleton, M., Robinson, K.L., Carew, M.E. and Furlan, E.M., 2020. Towards routine DNA metabarcoding of macroinvertebrates using bulk samples for freshwater bioassessment: Effects of debris and storage conditions on the recovery of target taxa. *Freshwater Biology*, 65(4), pp.607-620.

Oliverio, A.M., Gan, H., Wickings, K. and Fierer, N., 2018. A DNA metabarcoding approach to characterize soil arthropod communities. *Soil Biology and Biochemistry*, 125: 37-43.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. and Oksanen, M.J., 2013. Package 'vegan'. Community ecology package, version, 2(9), pp.1-295.

Pan, W., Byrne-Steele, M., Wang, C., Lu, S., Clemmons, S., Zahorchak, R.J. and Han, J., 2014. DNA polymerase preference determines PCR priming efficiency. *BMC Biotechnology*, 14(1), pp.1-17.

Pansu, J., Giguet-Covex, C., Ficetola, G.F., Gielly, L., Boyer, F., Zinger, L., Arnaud, F., Poulenard, J., Taberlet, P. and Choler, P., 2015. Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. *Molecular Ecology*, *24*(7), pp.1485-1498.

Pérez-Burillo, J., Trobajo, R., Vasselon, V., Rimet, F., Bouchez, A. and Mann, D.G., 2020. Evaluation and sensitivity analysis of diatom DNA metabarcoding for WFD bioassessment of Mediterranean rivers. *Science of The Total Environmen*t, 727, p.138445.

Piñol, J., Mir, G., Gomez-Polo, P.R.I.S.C.I.L.A. and Agustí, N., 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, 15(4), pp.819-830.

Piñol, J., Senar, M.A. and Symondson, W.O., 2019. The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, *28*(2), pp.407-419.

R Core Team, 2018. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria.

Saitoh, S., Aoyama, H., Fujii, S., Sunagawa, H., Nagahama, H., Akutsu, M., Shinzato, N., Kaneko, N. and Nakamori, T., 2016. A quantitative protocol for DNA metabarcoding of springtails (Collembola). *Genome*, *59*(9), pp.705-723.

Sato, H., Sogo, Y., Doi, H. and Yamanaka, H., 2017. Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. *Scientific Reports*, *7*(1), pp.1-12.

Serrana, J.M., Miyake, Y., Gamboa, M. and Watanabe, K., 2019. Comparison of DNA metabarcoding and morphological identification for stream macroinvertebrate biodiversity assessment and monitoring. *Ecological Indicators*, *101*, pp.963-972.

Shannon, C.E. and Weaver, W., 1949. The Mathematical Theory of Communication. University of Illinois Press, Urbana.

Shokralla, S., Porter, T.M., Gibson, J.F., Dobosz, R., Janzen, D.H., Hallwachs, W., Golding, G.B. and Hajibabaei, M., 2015. Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports*, 5, p.9687.

Shokralla, S., Spall, J.L., Gibson, J.F., and Hajibabae, M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 21, pp.1794–1805.

Sow, A., Brévault, T., Benoit, L., Chapuis, M.P., Galan, M., Coeur d'acier, A., Delvare, G., Sembène, M. and Haran, J., 2019. Deciphering host-parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding. *Scientific Reports*, *9*(1), p.3646.

Taberlet, P., Bonin, A., Zinger, L. and Coissac, E., 2018. *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press.

Taberlet, P.,C. Eric, Pompanon F.,Brochmann C. and Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*. 21: 2045–2050.

Vences, M., Lyra, M.L., Perl, R.B., Bletz, M.C., Stanković, D., Lopes, C.M., Jarek, M., Bhuju, S., Geffers, R., Haddad, C.F. and Steinfartz, S., 2016. Freshwater vertebrate metabarcoding on Illumina platforms using double-indexed primers of the mitochondrial 16S rRNA gene. *Conservation Genetics Resources*, *8*(3), pp.323-327.

Wainer, R., Ridgway, H., Jones, E., Peltzer, D. and Dickie, I., 2020. Consequences of environmental DNA pooling. *Authorea Preprints*.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergård, M., Gussarova, G., Haile, J. and Craine, J., 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, *506* (7486), pp.47-51.

Xie, Y., Hong, S., Kim, S., Zhang, X., Yang, J., Giesy, J.P., Wang, T., Lu, Y., Yu, H. and Khim, J.S., 2017. Ecogenomic responses of benthic communities under multiple stressors along the marine and adjacent riverine areas of northern Bohai Sea, China. *Chemosphere*, *172*, pp.166-174.

Xiong, W. and Zhan, A., 2018. Testing clustering strategies for metabarcoding-based investigation of community–environment interactions. *Molecular Ecology Resources*, *18*(6), pp.1326-1338.

Yu, D., W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C. and Ding, Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*, pp.613–623.

Zaiko, A., Martinez, J.L., Schmidt-Petersen, J., Ribicic, D., Samuiloviene, A. and Garcia-Vazquez, E., 2015. Metabarcoding approach for the ballast water surveillance–An advantageous solution or an awkward challenge?. *Marine Pollution Bulletin*, *92*(1-2), pp.25-34.

Zaiko, A., Samuiloviene, A., Ardura, A., and Garcia-Vazquez, E. 2015. Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Mar. Pollut. Bull. 100*, pp.53–59.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle, B.E. and De Barba, M., 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*(8), pp.1857-1862.

Zimmermann, J., Glöckner, G., Jahn, R., Enke, N. and Gemeinholzer, B., 2015. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, *15*(3), pp.526-542.

Chapter 4

Metabarcoding of macroinvertebrates to assess diversity and environmental degradation in river ecosystems of Bangladesh

Based on this chapter a research article entitled "DNA-based assessment of environmental degradation in an unknown fauna: the freshwater macroinvertebrates of the Indo-Burmese hotspot" has been published in the Journal of Applied Ecology.

Authors:

Md. Mizanur Rahman[1, 2, 3], Alfred Burian[4, 5], Thomas J. Creedy[1, 2] & Alfried P. Vogler[1, 2]*

[1]Department of Life Sciences, Natural History Museum, London, SW7 5BD, UK
[2]Department of Life Sciences, Silwood Park Campus, Imperial College London, Ascot, SL5 7PY, UK
[3]Department of Zoology, University of Dhaka, Dhaka-1000, Bangladesh
[4]Marine Ecology Department, Lurio University, Nampula, Mozambique
[5]Department of Computational Landscape Ecology, UFZ–Helmholtz Centre for Environmental Research, Leipzig, Germany

Chapter 4: Metabarcoding of macroinvertebrates to assess diversity and environmental degradation in river ecosystems of Bangladesh

4.1 Abstract

Man-made stressors are causing various degrees of biodiversity loss and limit ecosystem functioning in many freshwater habitats in Bangladesh which makes it essential to assess their impacts as a first step of protecting and restoring aquatic ecosystems. Therefore, efficient methods are required for biomonitoring of poorly known tropical ecosystems, but biological assessments of environmental status are limited by insufficient information on taxonomy, composition, and ecology of local communities. Here I applied the metabarcoding technique to establish the macroinvertebrate diversity and impact of various types of anthropogenic disturbances on the freshwater macroinvertebrates in highland and lowland rivers of Bangladesh. Whole-community metabarcoding was used to investigate the distribution of hypothetical species-level clusters (Operational Taxonomic Units, OTUs) across sites of different impacts. From highland and lowland rivers respectively, I found 936 and 662 DNA clusters of insects, decapods and molluscs, dominated by Diptera, which revealed significant variation ($p<0.001$) in richness across sites. In highland streams, the type and strength of anthropogenic stressors varied greatly across streams but did not affect total OTU diversity. In contrast, EPT richness decreased by ~50% in response to habitat degradation. The environmental variables of lowland rivers, significantly varied ($p<0.001$) across the sampling sites reflecting their joint dependency on pollution pressure. Decreases in species richness and genetic diversity of lowland rivers were highly dependent on the high density of nitrate, phosphate and salinity strongly suggesting the joint importance of different environmental components. Partial-network analysis revealed 26 and 16 OTUs for highland and lowland rivers respectively that may serve as potential indicators for either good or poor ecological status. Overall, the results document high diversity, local endemicity and pronounced responses to disturbance in largely unexplored but threatened habitats of Bangladesh. The approach of the present study will have great value for applied conservation management as a step towards building a biomonitoring system in this region where currently little is known about the taxonomy, diversity and endemicity in both intact and disturbed ecosystems.

4.2 Introduction

4.2.1 Biomonitoring of freshwater ecosystems

Freshwater ecosystems provide a number of critical services, including the provision of clean drinking water, fish stocks, flood protection, carbon sequestration and mode of transportation besides supporting a vast habitat range for freshwater flora and fauna (Arthington et al., 2010; Hitzhusen et al., 2000). Improving our knowledge about the structure and function of freshwater ecosystems is therefore very essential for practical interest to mankind along with its biological implications. The availability of freshwater resources for the maintenance of life throughout the world depends on sustainable management and usage of river ecosystems (Loeb and Spacie, 1993). Unfortunately, freshwater environments have been subjected to unfavourable alterations and degradation caused by man-made and natural perturbations over the last few decades (Uherek and Pinto, 2014). Many freshwater habitats are threatened by anthropogenic disturbances such as pollution and eutrophication, over-harvesting, structural modifications, water abstraction, and invasion of exotic species (Dudgeon et al., 2006; Kuntke et al., 2020; Vörösmarty et al., 2010). Individually or combined, these stressors are causing various degrees of biodiversity loss and limit ecosystem functioning (Arthington et al., 2010; Chapin et al., 2000; Loreau et al., 2001), which makes it essential to assess their impacts as a first step of protecting and restoring freshwater ecosystems (Santos and Ferreira, 2020).

Ideally, a comprehensive monitoring programme should include physical, chemical and biological measurements because they provide the complete spectrum of structures, services and changes encountered in ecosystems. Biological monitoring with aquatic organisms (e.g., fish, macroinvertebrates, diatoms) has been proven to be necessary supplementary to other monitoring techniques (Rosenberg and Resh, 1993; Soininen and Könönen, 2004). Macroinvertebrates are effective sentinels of external disturbances and thus have been widely used as indicators of habitat integrity and degradation (Bonada et al., 2006; Menezes et al., 2010; Serrana et al., 2019). They have been already included in the regulatory framework for protecting aquatic resources, e.g. in the Water Framework Directive and the Marine Strategy Framework Directive in Europe (Hering et al., 2018; Leese et al., 2016) and similar legislation in North America, such as the US Clean Water Act (the United States, 1972).

4.2.2 **Characterizing macroinvertebrates for biomonitoring river ecosystems**

Ecological status assessment in aquatic ecosystems requires detailed knowledge about diversity (e.g., taxonomic, functional, and phylogenetic) and the sensitivity of bioindicator macroinvertebrates. Among many diversity measures, species richness, species turnover, species composition and functional feeding guilds are considered as fundamental components to produce diversity indices, biotic index, multimetric indices, or more complex multivariate predictive indices for biomonitoring (Bonada et al., 2006; Karr, 1999; Rosenberg and Resh, 1993; Li et al., 2010). Most of these approaches broadly rely on alpha and beta levels biodiversity information, tolerance/intolerance measures and trophic dynamics (DeShon, 1995; Kerans et al., 1992, Kerans and Karr, 1994; Barbour et al., 1995; Hilsenhoff, 1987; Merritt et al., 1996). Consequently, such assessments are mainly conducted in countries of the temperate zones where the fauna is sufficiently well characterised for routine species identification and understanding of ecological requirements. Exploiting these established responses to water quality, biomonitoring of riverine ecosystems in many developed countries uses well defined biotic indices such as the Biological Monitoring Working Party Score System (BMWP) and Average Score Per Taxon (ASPT), as well as statistical modelling of water quality using the River InVertebrate Prediction And Classification System (RIVPACS) (Hawkes, 1997; Wright et al., 2000).

However, this is not the case for most countries in tropical and subtropical parts of the world, including the Indo-Burmese region, where water quality evaluation is mainly based on physico-chemical data and bioindicators that are only rudimentarily developed (Chowdhury et al., 2016; Ofenböck et al., 2010; Eriksen et al., 2021b). The region is one of many examples of the intense threat to biodiversity in (sub)tropical areas that rank among both the potentially most diverse and also the least well-monitored freshwater ecosystems (Ahmed et al., 2013; Allen et al., 2012; Eriksen et al., 2021a). Thus, the challenge in many parts of the world is to develop methodologies for documenting the species and ecological diversity of local water bodies and, at the same time, to establish what are the most damaging practices affecting these ecosystems. Likewise, Bangladesh located in the western part of the Indo-Burmese hotspot of South Asia supports rich faunal diversity but experiences critical environmental degradation that escalated the pressing demand for biodiversity estimation of macroinvertebrates and ecological assessment of river systems to restore and protect the freshwater ecosystems.

4.2.3 Bioassessment of freshwater ecosystems in Bangladesh

Bangladesh is crisscrossed by many rivers, but only 405 of them are recognized by the Bangladesh Water Development Board (BWDB, 2012). Most of the rivers are either tributaries or distributaries of the Ganges, the Brahmaputra or the Meghna River. They are the vast reservoirs of biodiversity including fish, macroinvertebrates, plankton and aquatic plants which are maintaining the ecological processes of freshwater ecosystems. These rivers serve various ecosystem services including drinking water, cultivation and as the principal arteries of commercial transportation in the country. The National Water Management Plan (NWMP) divides the country's river systems into six regions that can be broadly categorised into plain or lowland rivers and highland rivers.

This high land area is representative of subtropical, mountainous regions of the Indo-Burma ecoregion generally covered by evergreen forest and crossed by numerous small streams. Expansion of traditional slash-and-burn agriculture, rapidly increasing population density and infrastructure development combine to impact these streams through erosion, pollution and extraction of building material, as also noted in adjacent Myanmar (Eriksen et al., 2021b; Bai, 2006). In particular, the second and third-order small streams crossing steep terrain in a mosaic landscape potentially hold high species diversity and great turnover of aquatic invertebrates. The lowland areas also support most of the river systems on the Ganges, Brahmaputra or the Meghna basins and are known to contain diverse aquatic fauna and flora including fish, macroinvertebrate, plankton and aquatic plants. With increasing population, industrialization, urbanization and discharges of agricultural run-off, lowland rivers and streams are getting more and more seriously polluted across the country.

However, biological assessment of riverine ecosystems is still in its infancy even though macroinvertebrates or other bioindicator diversity remain poorly studied except for the limited work of British naturalists in the early 20th century and a few recent taxon-specific studies. Faunal survey and biodiversity study in the Indo-Burmese region was so far commenced at the end 19th century and continued to the middle of 20th century by some British naturalists during their colonial government in India. They produced a multivolume book namely 'The Fauna of British India Including Ceylon and Burma' that included several volumes for invertebrates such as Coleoptera (Flower, 1912; Marshall 1916), Odonates (Fraser, 1934 & 1936) and molluscs (Preston, 1915). Currently available studies in

Bangladesh are limited to opportunistic surveys and sketchy diversity estimations of aquatic insects (e.g., Sana and Ali, 2011; Ahad et al., 2012; Mustafa et al., 2013; Nasiruddin et al., 2014), mosquitoes (Irish et al., 2016), odonates (Chowdhury and Akhteruzzaman, 1981; Bashar et al., 2014), and molluscs (Begum et al., 1989).

There is no established methodology for bioassessment of river ecosystems like the Water Framework Directive in Europe. Instead, major water projects are run by the government solely for its benefit to irrigation, flood control, drainage facilities, river navigation and hydroelectric power generation. Over the last two decades, Bangladesh has achieved some remarkable progress in Integrated Water Resources Management (IWRM) formulating policies and legislation (e.g., National Water Policy 1999, National Water Management Plan 2004, Bangladesh Water Act 2013 and Bangladesh Water Rules 2018). Each of these policies, plans and rules has emphasized on the ecological assessment and biomonitoring for conservation of aquatic biota. However, the country still lacks biological quality elements (e.g., macroinvertebrates) based classification system for evaluating and monitoring the health of rivers or streams.

Therefore, the first step in building any such system requires basic knowledge of the regional or country's species diversity, both for communities without anthropogenic impact (the reference state) and along gradients of environmental degradation. In addition, these assessments require an understanding of the degree of local endemicity due to climatic, seasonal and biogeographic variation across the study region, which may drive apparent species turnover among streams and sampling sites, even in unimpacted settings (Múrria et al., 2018; Seymour et al., 2021). Given this fact, to protect the freshwater ecosystems, Bangladesh should explore the monitoring data characterizing bioindicator macroinvertebrates and assess the impact of existing stressors on the degradation of river systems. However, it is a herculean task for traditional biomonitoring programs to accommodate these ecological attributes especially describing community composition through a rapid, reliable, and cost-effective process. In order to tackle the limitations of current identification methods, DNA-based metabarcoding has become a suitable approach for the simultaneous identification of individuals in large mixed communities. Therefore, the current study uses metabarcoding to establish the diversity and sensitivity to alterations in freshwater rivers and streams in highland and lowland areas of Bangladesh.

4.2.4 Morphology vs metabarcoding based assessment of macroinvertebrates in freshwater ecosystems

Morphology-based monitoring of freshwater invertebrates is increasingly complemented with metabarcoding, i.e., the mass amplicon sequencing of standardised short genomic regions and identification against reference databases, which can reveal the composition of entire communities (Emilson et al., 2017; Serrana et al., 2019). The ability of DNA metabarcoding for mass identification of organisms has been demonstrated in the assessment of aquatic macroinvertebrate communities (Carew et al., 2013; Hajibabaei et al., 2012). Studies also have highlighted the potentiality of this technique to characterize the alpha, beta, and gamma diversity of complex macroinvertebrate community and their ecological assessment metrics (Elbrecht et al., 2017; Gibson et al., 2015). Stream water quality assessment with biological quality elements (BQEs) has been performed by metabarcoding that also complemented the results of the conventional assessment producing the distinct clusters of taxonomic units of macroinvertebrates (Kuntke et al., 2020). Moreover, metabarcoding data are used to cross-reference species detectable among different sites, to assess responses to environmental impacts on the level of entire communities (Andújar et al., 2018; Beng et al., 2016; Carew et al., 2018; Emilson et al., 2017; Xie et al., 2017). The species-level resolution of metabarcoding may also refine the taxonomic level at which these analyses of turnover and environmental sensitivity can be conducted (Bush et al., 2020). Beyond the recognition of certain easily recognisable species, morphological monitoring even in biologically well-known regions is frequently based on higher taxa, e.g., at the level of families in insects ("Chironomidae"), which blurs the many differences in response to environmental parameters that exist even among closely related species (Beermann et al., 2018). Instead, metabarcoding allows searching for broad patterns emerging in communities distinguished by many hundreds of taxa. This fine-scale taxonomic resolution is highly valuable even if an exact Linnaean species identification is not possible due to the incompleteness of reference databases, as expected in poorly studied tropical communities.

In this study, the metabarcoding technique was applied to test its performance in routine assessment of macroinvertebrate fauna along with the environmentally gradient sites of lowland and highland river systems. This approach is likely to facilitate OTU/species-level identification and detect the factors responsible for the deterioration of aquatic ecosystems in Bangladesh.

4.2.5 General aims and research questions

Here I assessed macroinvertebrate diversity and the impact of anthropogenic activities in 16 highland streams of the Bandarban district and 4 lowland rivers around the capital city, Dhaka. The highland second-order streams have remained in a fairly pristine state but are increasingly affected by anthropogenic pressure. All of them are tributaries of the Sangu River and traverse similar terrain and altitudinal ranges, while the degree of disturbance varies among streams, making them useful replicates for regional sampling. The 4 lowland rivers are interconnected formed a river network and are heavily impacted by industrial and urban effluents and agricultural run-off directly discharged into the river. I assume that DNA-based operational taxonomic units (OTUs) from the metabarcoding pipeline can be potentially placed in taxonomic and phylogenetic frameworks of macroinvertebrate communities to assess the alpha and beta diversity across the human-induced pressure or pollution gradient sites. Metabarcoding was used to evaluate the degree to which total species diversity of macroinvertebrates and the diversity of the disturbance-sensitive and pollution tolerant macroinvertebrate taxa were impacted by human-induced stressors. Secondly, we screened for potential indicator species associated either with poor or good ecological status by correlation with a set of environmental variables evaluated for each sampling site. Together, these steps allowed us to provide the baseline data needed for evaluating habitat alterations and to develop tools for future bioassessment in the poorly known river systems of Bangladesh.

4.3 Methods

4.3.1 Study site and sample collection

Highland streams

Macroinvertebrate samples (*n*=80) were collected from relatively pristine upland streams located in Bandarban, south-eastern Bangladesh (Fig. 4.1). Upland streams locally called Chhora flow through tropical evergreen or semi-evergreen hill forests and drain into the Sangu River. Their riparian vegetation is rich in forbs and shrubs and the streams show high densities of invertebrates and fish (Ahmed et al., 2013). Sixteen physically similar second-order streams were selected at approximately equal distances from each other, as much as logistically possible in this area of limited access to sampling locations.

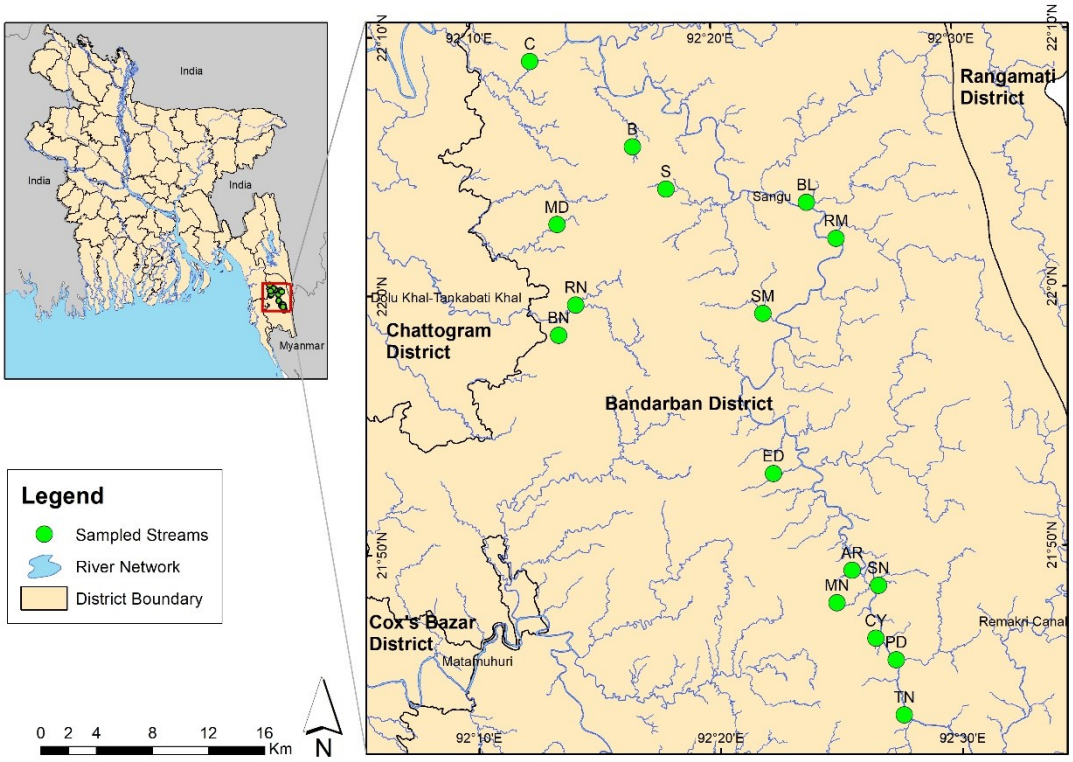

Figure 4. 1 Map of the sixteen highland streams in Bandarban district of south-eastern Bangladesh. Green circles highlight the 16 sampled streams (S-Sangukhiang Chhora, B-Betchhora, C-Cheihkhiang Chhora, BN-Bangchhora, RN-Ranginmukh Chhora, MD-Maddyamkhal Chhora, ED-Eddmara Jhirri, AR-Army camp Chhora, MN-MongotJhirri, SN-Sandak Jhirri, PD-Paddayo Jhirri, TN-Tindupoint, CY-Chhotoyangry point, SM-Semakhal Chhora, BL-Belden Chhora, RM-Rumakhal Chhora).

All streams were located in the Sangu River basins and sampling sites mostly included the riffles and shallow pools (about 30-100 cm deep). Substrate condition and composition varied between streams containing different types and amount of large stone, boulders, gravels pebbles, fine sediments, leaf litters and detritus depending on anthropogenic pressure (Rahman et al., 2016). The climate of the region is tropical monsoon with distinct wet (May to October) and dry (December to April) seasons. Annual temperature varies from 10^0 C to 35^0C with a mean minimum of 24^0C in (December-January) and a mean maximum of 34^0 C (April -June). The average annual rainfall is 2540 to 3810 mm in this region (Bai, Z.G., 2006). Current pressures on the streams network in this area include deforestation, conversion of forested areas to agricultural fields (jhum cultivation), roads construction and irresponsible tourist visits at some locations (Bai, 2006). Furthermore, sewage and other wastes enter the streams from the adjacent household area. These lead to potential hydro-morphological changes of the streams and increase pollution loads with excess sedimentation, nutrients and xenochemicals that influence the aquatic environment negatively (Erikson et al., 2021). In addition, fishing by illegal means such as electrofishing and applying pesticides is also assumed as crucial stressors to the aquatic fauna of the streams. GPS coordinates, elevation, and substrate type of the streams are given in the appendix (Table 4.1).

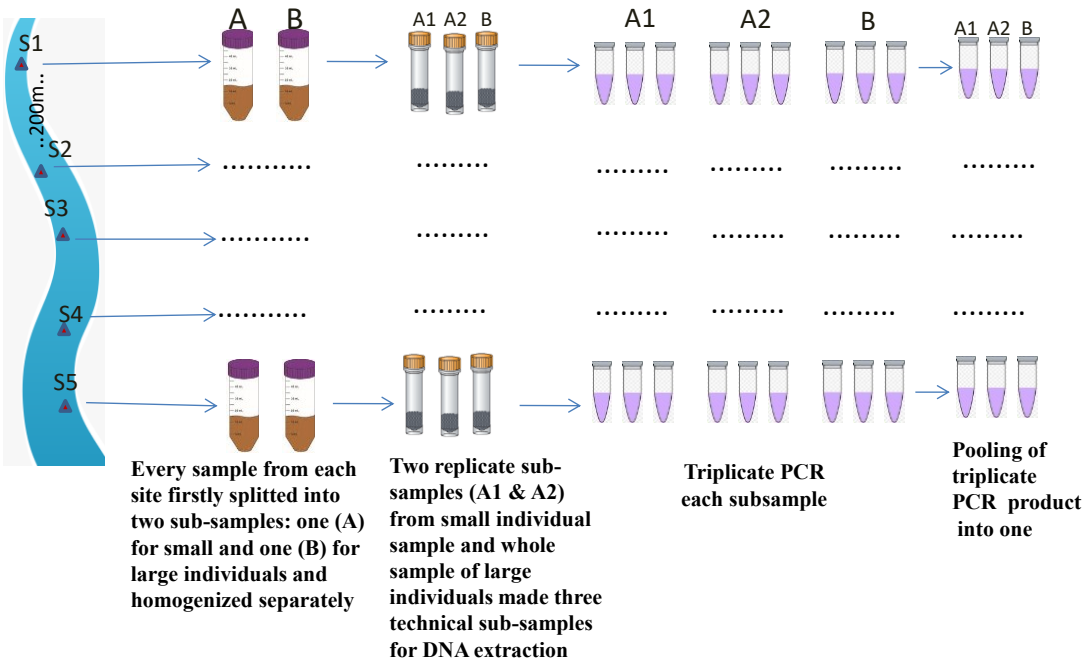

Figure 4. 2 Study design for metabarcoding bulk macroinvertebrate samples from each upland stream. Five kick-net samples were collected from five sites (S1-S5) of a stream.

Sampling took place in the mid-winter season when the water was relatively fast-flowing and cold at temperatures of 15-18 °C. Per river, 5 sites were sampled at distances of 200 m over a 1000 m river section (Fig. 4.2). At each of the 80 sites, two-pole kick-net samples (3 minutes collection time per sample) were obtained, and samples were sieved (0.5 mm) in the field and preserved in absolute ethanol after removing large debris.

Anthropogenic impacts at each site were quantified based on the assessment of components of environmental intactness: (i) hydromorphological intactness, (ii) substrate intactness, (iii) absence of pollution, (iv) absence of fishing pressure and (v) catchment intactness (Table 4.1). These components of environmental intactness were derived from 14 binomial variables (Table 4.5, 4.6,4.7 in the appendix) that were evaluated based on direct observations by the authors and interviews with 3-4 informed locals at each river (Table 4.1). Reliance on community knowledge has the advantage to attain time-integrated data that can be more robust than single-time measurements of fluctuating physical variables such as oxygen and nutrient concentrations. Interviewees were frequently local farmers and partly illiterate who may struggle with strictly quantitative concepts (e.g., grading the amount of change). Hence, we used a large number of binomial variables to increase the robustness of our results. Further, overall environmental status was calculated as the grand average of all five components of environmental intactness.

Table 4. 1 Overview of binomial variables assessed during field sampling to evaluate different dimensions of human influence.

No	Binomial variables	References	Higher-order criteria	Assessment method
1	Substrate composition appropriate for stream order and slope	Hughes, 1995; Barbour et al., 1996	Substrate intactness	Visual inspection
2	No Sand, gravel or stone excavation	Nijboer et al., 2004	Substrate intactness	Visual inspection, Interview
3	Natural stream structure	Hughes, 1995; Barbour et al., 1996	Hydromorphological intactness	Visual inspection
4	No Significant water extraction	This study	Hydromorphological intactness	Interview, Visual inspection
5	No Damming or diversion of water flow	Nijboer et al., 2004	Hydromorphological intactness	Visual inspection
6	No Dumping of household wastes	This study	Absence of pollution	Interview, Visual inspection
7	Minimum washing and bathing activities	Nijboer et al., 2004	Absence of pollution	Interview
8	Minimum Run-off from adjacent cropland	This study	Absence of pollution	Interview, Visual inspection
9	Natural water colour and odour	Nijboer et al., 2004	Absence of pollution	Visual inspection
10	Minimum Tourist pressure	This study	Absence of pollution	Interview
11	Absence of fishing pressure	This study	Absence of fishing pressure	Interview
12	Apropriate riparian vegetation	Hughes, 1995; Barbour et al., 1996	Catchment intactness	Visual inspection
13	Representative Diversity of terrestrial wildlife	Barbour et al., 1996	Catchment intactness	Interview, Visual inspection
14	No significant intervention of exotic plant or animal species	Sánchez-Montoya et al., 2009	Catchment intactness	Visual inspection

Lowland rivers

Samples (*n*=60) were collected from four non-wadable and interconnected lowland rivers (including highly polluted and least polluted) of which the Buriganga and the Turag partially surround the capital city Dhaka and the other two, the Dhaleshwari and Kaliganga Rivers are about 10-20 km away to the west from the city (Fig. 4.3). These rivers are commercially important for agricultural, sanitary, and industrial purposes (Alam et al., 2002) and also serve as a major transportation route and flood control and drainage outlet. They also play an important role in the livelihood of local people by providing fish and fisheries resources. Rapid industrialization and unplanned urbanization have encroached on most of the banks of the River Buriganga. The bank of the River Turag is also occupied by factories, but some parts are used for agricultural production. Therefore, the discharge of industrial effluents sewerage toxic wastes and agricultural runoff have resulted in increased water pollution (Moniruzzaman et al., 2009; Islam et al., 2006; Ahmed et al., 2015). The land-use patterns of the Dhaleshwari and Kaliganga River mainly include agricultural activities and the water is generally used for irrigation. Although some industrial plants are active along the banks, they are facing relatively less anthropogenic pressure than the Buriganga and Turag. These rivers also support aquatic vegetation on the riverbank. However, agricultural runoff with insecticides, pesticides and fertilizers is frequently released into the river (Islam et al., 2012; Ahsan et al., 2018). Details of selected lowland rivers are also described in chapter 2.

The Rivers Buriganga and Turag are situated in the Ganges basin (known as the Padma basin in Bangladesh) whereas the Rivers Dhaleshwari and Kaliganga are in the Brahmaputra basin (Jamuna basin). The climatic condition of the basin area is tropical monsoon type which influences the hydrodynamic features of these rivers. The rainfall mainly occurs (80–90%) during the monsoon (June to September), estimated annually at over 2,000 mm. In the rainy season, the temperature ranges from 25 to 31 °C, evaporation ranges from 80 to 130 mm and the average humidity fluctuates from 80 to 90 %. Variation in rainfall intensity, temperature, relative humidity, evaporation and wind velocity affects the water quality of the river. Generally, the river experiences low tidal and semi-tidal influences in downstream reaches during wet (monsoon) and dry seasons respectively (Bangladesh Meteorological Department).

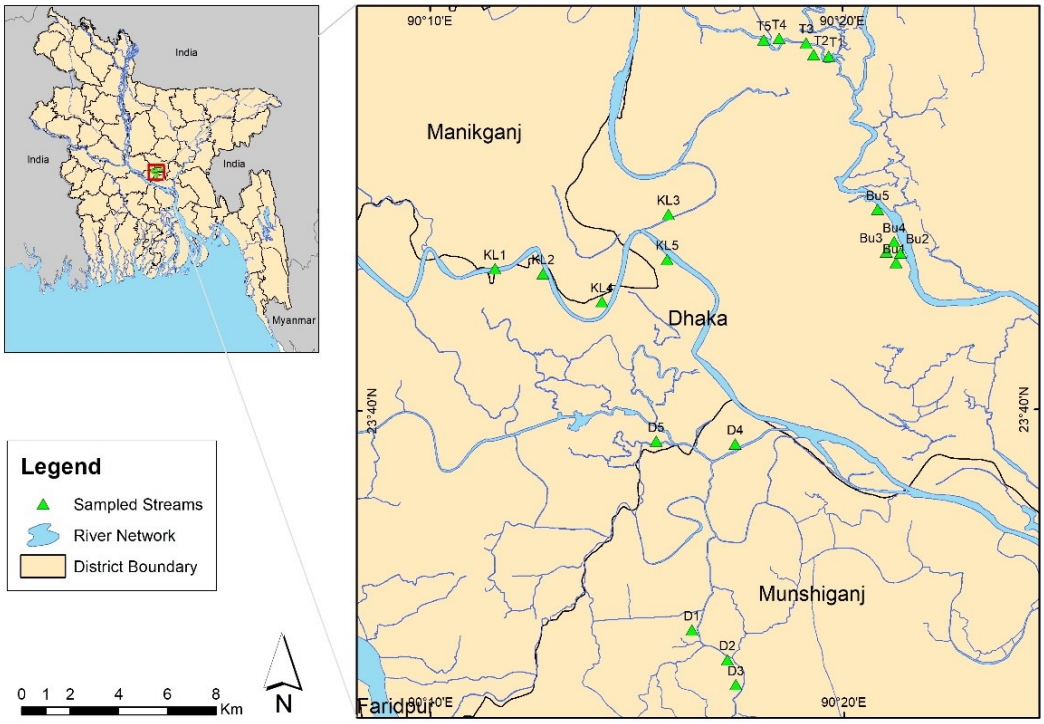

Figure 4. 3 Map of the 20 study sites in 4 lowland rivers around Dhaka city of central Bangladesh. Green triangles highlight the 5 study sites in each river (Bu- Buriganga, D-Dhaleshwari, KL-Kaliganga, T-Turag). GPS coordinates of each site of four rivers are given in the appendix Table 4.2.

Sampling took place in the mid-winter season when the water was slow-flowing and cold at temperatures of 16-20 °C. Five sites in each of the four rivers were sampled at 2000-3000 m river section distances. Each site extended around 400-600 m from where three replicate samples were taken that produced 60 samples from 4 lowland rivers (Fig. 4.3). A two-pole kick-net was used for sample collection following the same protocols applied for upland rivers. To support the evaluation of the environmental degradation in selected sites of the rivers, hydro-chemical parameters like pH, dissolved oxygen (DO), salinity, conductivity, and total dissolved solids (TDS) were measured onsite immediately after sample collection using a HACH Water Quality Multimeter (Model: HQ40d, USA). Nitrate and phosphate were also measured by using a Spectrophotometer (Model: DR/1900, HACH, USA) following the manufacturer's protocols.

4.3.2 Samples processing and DNA methods

All large-sized (body length >15mm) individuals were separated to avoid overrepresentation in mixed community sequencing (Creedy et al., 2019; Elbrecht et al., 2021). For these individuals, a tissue sample (~20 mg) was obtained, and pooled tissues were subsequently dried at room temperature and homogenized using a TissueLyser (Qiagen, Hilden, Germany). Smaller individuals (<15 mm) were still mixed with debris and small sediment particles. Hence, I homogenized these samples after drying them at 37°C with sterilised pestles and mortars. To account for the heterogeneity of the sample, DNA extraction was carried out on 0.20 g homogenised material in duplicate. Thus, each sample was split into three technical sub-samples (two sub-samples for small and one sub-sample for large individuals) for DNA extraction, PCR amplification and sequencing (Fig. 4.2).

DNA was extracted from small and large individuals using the DNeasy Power Soil Kit and the DNeasy 96 Blood and Tissue Kit, respectively (Qiagen, Hilden, Germany). For small individual samples, the manufacturer's extraction protocol was slightly adjusted by increasing the used volume of the C1 solution from 60 to 180 µL and the first centrifugation step from 0.5 to 1 min. These adjustments were made for managing highly dried samples and extracting the proper volume of supernatant required for subsequent extraction steps.

DNA purity and concentration was determined using the NanodropND-8000 (Thermo Scientific) system for PCR amplification of extracted DNA samples. Metabarcoding of samples followed a standard protocol, targeting a 418 bp region of the Cytochrome Oxidase subunit I (COI) gene with invertebrate-specific primers (fwd: CCNGAYATRGCNTTYCCNCG and rev: TANACYTCNGGRTGNCCRAARAAYCA) (Arribas et al., 2016). The standard Illumina tails (TCGTCGGCAGCGTCAGATGTGTATAA GAGACAG and GTCTCGTGGGCTCGGAGA TGTGTATAAGAGACAG) were attached for forward and reverse respectively and 6-bp different tags to build Illumina ready PCR amplicon to differentiate the reads belonging to each sample.

Each PCR reactions contained 2 µL DNA template, 14.65 µL sterilized ultrapure water, 3.0 µL 10X TaKaRa buffer (Takara Bio Inc.), 0.15 µL MgCl2 (50 mM, Bioline), 0.40 µl of TaKaRa dNTPs (2.5 mM each), 0.70 µl of each primer (10 mM), 0.15 µl of TaqTM Hot Start

polymerase (5u/µl, Takara Bio Inc.), 0.25 µl of Bovine Serum Albumin (20mg/ml, Thermo Scientific). After an initial 4 min denaturation at 95˚C, the PCR ran for 30 cycles of 95°C for the 30s, 48°C for 30 sec, and 72°C for 1 min 45 sec, and a final extension at 72°C for 10 min. Amplification success was evaluated using gel electrophoresis. PCR on each extraction was done in triplicate and equal aliquots of PCR triplicates were pooled and cleaned with AMPure XP paramagnetic beads (Agencourt Bioscience, Massachusetts, USA) following the manufacturer's protocols with a slight adjustment of the first incubation period to 7 min and second incubation to 4 min, and the volume of magnetic beads to 6µl/10 µl of PCR product for 418 bp amplicons and 18µl/10 µl for 100 bp amplicons. Samples were indexed with Nextera XT tags during a secondary PCR for library preparation, and amplicons were sequenced on an Illumina MiSeq (2x300 bp paired-end) aiming for 65000 and 30000 reads per sample of small and large individuals, respectively.

4.3.3 Bioinformatic processing

Bioinformatic processing followed an established pipeline, a set of Perl scripts that wrap software for quality filtering, merging and clustering of sequence reads (Creedy et al., 2019). After primer trimming using cutadapt (Hannon Lab, 2012), read merging was performed with PEAR (Zhang et al., 2014; PEAR-q value of 26), and conversion of multiple fastq to fasta files and error rate filter with the maxee parameter (–eemax value of 1) were implemented USEARCH110 fastq_filter. All resulting reads were concatenated and dereplicated to form a set of unique sequences. Sequences that differed from the modal length of 418 bps and occurring in ≤2 copies were removed. The remaining sequences were denoised (Edgar, 2016) using default settings and clustered into OTUs using USEARCH v11.0 (Edgar, 2010).OTU clustering was performed using USEARCH v11.0 (Edgar, 2010) and the most representative sequences were identified against the NCBI *nr* database for taxonomic assignment with the lowest common ancestor (LCA) method in MEGAN (Huson et al., 2016). OTUs only assigned to Insecta (Coleoptera, Diptera, Ephemeroptera, Hemiptera, Odonata, Plecoptera, and Trichoptera), Decapoda and Mollusca were retained for final analysis. Phylogenetic trees required for calculating phylogenetic diversity were constructed under maximum likelihood (ML) with RAxML using default settings and the GTRMIX-model (Stamatakis, 2006) on the CIPRES server (Miller et al., 2010). For abundance-based analyses, I combined technical sub-samples for each field sample, by first combining the two technical replicates of small-

bodied individuals and removing four and five samples (from highland and lowland respectively) with a sequencing depth of <8500 total reads. A joint rarefaction-extrapolation approach was implemented with the remaining samples using the iNEXT package in R to calculate richness for each of the two sample types (Hsieh et al., 2016). For all community assessments, I used the relative proportion of unique OTUs to the total number of reads of non-rarefied samples, which I will refer to as relative abundance. For samples of large-bodied individuals, I prepared equal amounts of tissue from each sampled individual. As the maximum number of individuals per sample was low (max. 15-20), the sequencing depth was large enough to capture the full species diversity and rarefaction was not required. However, several samples (n = 9 from upland and 2 from lowland) did not successfully amplify (< 500 reads per sample) and were excluded from the dataset.

4.3.4 Statistical analysis

Biodiversity indices were calculated for each local site and correlated with environmental intactness (highland) and hydro-chemical measures (lowland). To calculate the overall OTU richness at a site, I transformed the small and large-specimen samples into binomial presence-absence datasets and merged the two sample types. For density-dependent measures (evenness), I used relative read abundance as a measure of the population size of the species present, which was performed separately on samples with small and large-bodied individuals. Phylogenetic diversity (PD) was calculated as unweighted total branch length (Faith, 1992) based on ML trees. Further downstream analyses for highland and lowland rivers were carried out with slightly different statistical methods as follows:

Highland streams

In case of highland streams, the impact of overall environmental intactness on biodiversity measures of OTU richness, evenness and PD was assessed in regression analyses. In addition, I tested which aspects of environmental intactness had the highest impact on species diversity, using a full model building approach established with all possible combinations of the five components of environmental intactness as predictors and the Bayesian Information Criterion (BIC) to determine the most parsimonious models. Regression residuals were

controlled for autocorrelation, homogeneity of variance and remaining patterns. Moreover, I tested differences between the beta diversity of small and large-specimen samples using paired t-tests after calculating pairwise Jaccard similarity metrics for both sample types. The same procedure was repeated using Bray-Curtis as density-dependent similarity index using non-rarefied proportion data.

Further, I screened our community data for OTUs sensitive to anthropogenic influence using an indicator species analysis (ISA) (Cáceres and Legendre, 2009) modified for the use of continuous variables. I first assessed the relationship between each of the five environmental intactness measures and the relative abundance of OTUs. Standard ISA identifies taxa that show different abundance in two or more sample categories. However, I aimed to identify taxa that responded to continuous variables and hence I used Spearman rank correlations to screen for indicator taxa. For each river I averaged the environmental variables and relative abundance of individual OTUs, to avoid potential confounding issues emerging from a nested data structure and to increase the robustness of our analyses. Potential indicators for different components of environmental intactness were identified as the OTUs with the highest correlation coefficient ($r>0.31$) that showed a significant relationship with the respective component of environmental intactness. Due to the high number of calculated correlations and related issues with multiple testing, the reliability of p-values is undermined. Therefore, I used this approach to identify the most promising candidate species for further investigations, which still need to confirm the causality between species response and environmental degradation (also see Discussion). Further, correlations with an environmental variable sometimes might emerge because of an indirect influence of interactions with a second species in the community rather than resulting from an independent response to the environment. Consequently, I visualised among-OTU correlations and OTU-environment intactness correlations together in a network analysis to visualise these potential interactions (Seymour et al., 2020). For each component of environmental intactness, subnetworks were created based on significant Spearman correlations (displayed as edges) with potential indicator species (nodes) and displayed using the package *qgraph* (Epskamp et al., 2012) in R (R Development Core Team, 2018).

Lowland rivers

In case of lowland streams, the impact of seven hydro-chemical parameters on biodiversity measures of OTU richness and PD was assessed in regression analyses. The potential confounding effects of river identity were also taken into account by using a mixed effect regression model including river as a random effect. I investigated important predictors of the highest impact on species diversity and the correlation among hydro-chemical parameters. Because of the strong correlations among predictors, I did PCA (Principal Component Analysis) on the different covarying predictors to determine the effective factors (predictors) responsible for shaping the richness and phylogenetic diversity. However, DO, conductivity and TDS were removed from the dataset because they caused an aggregation effect (i.e., results were driven by river identity rather than by variation in those parameters). Nitrate, phosphate, and salinity were finally selected for running the PCA. The model selection process was conducted by building Linear Mixed-Effects Models (using the lmer function of the package, lme4) with pH and all possible combinations of the three axes of PCA as predictors and the Akaike Information Criterion (AIC) to determine the most parsimonious models.

I also analysed the overall change in OTU richness and approximated the underlying hydro-chemical parameters using a constrained ordination technique, redundancy analysis (RDA). A permutation test with forward selection was performed to examine if the considered explanatory variables were significant ($P \leq 0.05$ after 999 random permutations) in governing the OTU richness. Furthermore, I ran variation partitioning (Borcard, 1992) to quantify the proportion of the variation in community composition explained by variation in each of the three explanatory variables (nitrate, phosphate, and salinity). All ordination analyses were done using the R-language (R Development Core Team 2018) functions in the vegan package. Moreover, a ranked dissimilarity based ANOSIM (The ANalysis Of SIMilarity) test was performed to test the differences between the beta diversity of small and large-specimen samples using the 'anosim' function of the vegan package. The indicator species analysis was done following the same method above stated for highland streams.

4.4 Results

4.4.1 Highland Streams

A total of 4,658,829 reads passed the quality control and bioinformatic pipeline and were subsequently assigned to 3439 OTUs. The total number of reads per sample (80 independent samples, each carried out in triplicates) ranged from 15505 to 115126, with an average of 58235 reads. The total number of OTUs ranged from 163 to 1000 with an average of 540 per sample. 2613 of the 3439 OTUs could only be assigned to the level of "Eukaryota" using the NCBI *nr* database. For the current study, we retained altogether 936 OTUs of the target groups Insecta, Mollusca and Decapoda, which represented 67.12% of total read abundance and 27.21% of OTU richness.

4.4.1.1 Diversity measurement of macroinvertebrates

Metabarcoding revealed a total of 165 ± 42 OTUs per site (Fig. 4.4-A) across the 16 sampled streams. The small-specimen fractions (body length <15 mm) produced 153 ± 42 OTUs per site (mean of 29125 reads per sample) while the larger individuals contributed 37 ± 28 OTUs per site (12502 reads per sample). On average 9% of OTUs were found in both fractions. In most of the 80 sampling sites, evenness ranged between 0.4 and 0.6. Notably, the small-bodied fraction showed significantly higher evenness (T-test; T-value = 8.67, $p<0.001$) while no significant correlation was found between the evenness of small and large individuals across samples ($p = 0.79$). Analysis of Faith's PD index showed that on average only 14 ± 3% of the total phylogenetic diversity was present in each sample (Fig. 4.4-A).

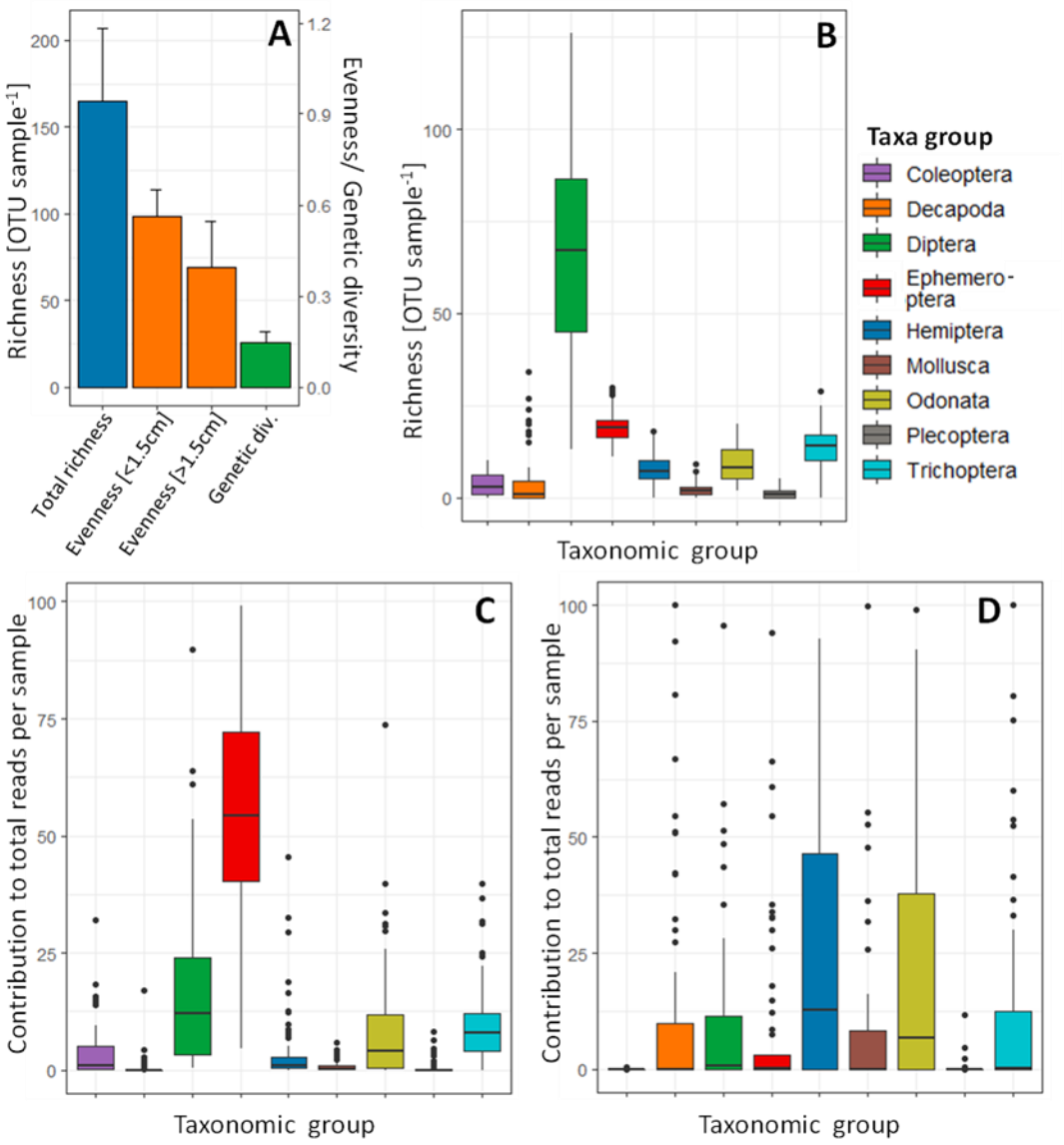

Figure 4. 4 Species diversity and relative abundance of analysed invertebrate taxa across all samples. (**A**) Average OTU richness, evenness (calculated separately for both sample groups) and genetic diversity of macroinvertebrates recorded in our study. Error bars denote standard deviations. (**B**) Boxplots of OTU richness per order (Bivalvia and Gastropoda have been merged into Mollusca) across all 80 sites. (**C**) Relative abundance (sequence reads) of different orders in the small- and (**D**) large-bodied fraction. C and D are based on different data collection protocols and hence the results are displayed separately.

Diptera were the most OTU-rich group, with an average of 66 ± 27 OTUs per sample and a total of 506 OTUs across all sites (Fig. 4.4-B). The diversity of Ephemeroptera, Plecoptera and Trichoptera (EPT) was lower, with 66, 9 and 71 OTUs, respectively, across all sites, although Ephemeroptera still accounted for 36 ± 9 OTUs per site. High diversity also was found for Decapoda, which each comprised ≥20 taxa in seven sites. The predator dominated taxa Odonata and Hemiptera also reached considerable average OTU numbers (Fig. 4.4-B). In contrast to richness, relative abundance (calculated as the proportion of the total number of reads attributed to a taxon) for the small-bodied fraction was dominated by Ephemeroptera (Fig.4.4-C), which together with Plecoptera and Trichoptera had high numbers of reads (65%) per site while it was only 16% for Diptera. Both the OTU richness and proportion of read numbers significantly varied among the insect orders (Kruskal-Wallis Test, $p<0.001$). The large-bodied specimens were dominated by Hemiptera and Odonata, although there was a large degree of variation across samples evident from numerous outliers in boxplots (Fig. 4.4-D). This variability was also reflected in the significantly higher beta-diversity in the large-bodied than small-bodied fraction (Mann-Whitney-Wilcoxon Test, $p<0.001$ for comparisons based on Jaccard and Bray-Curtis indices; Fig. 4.5).

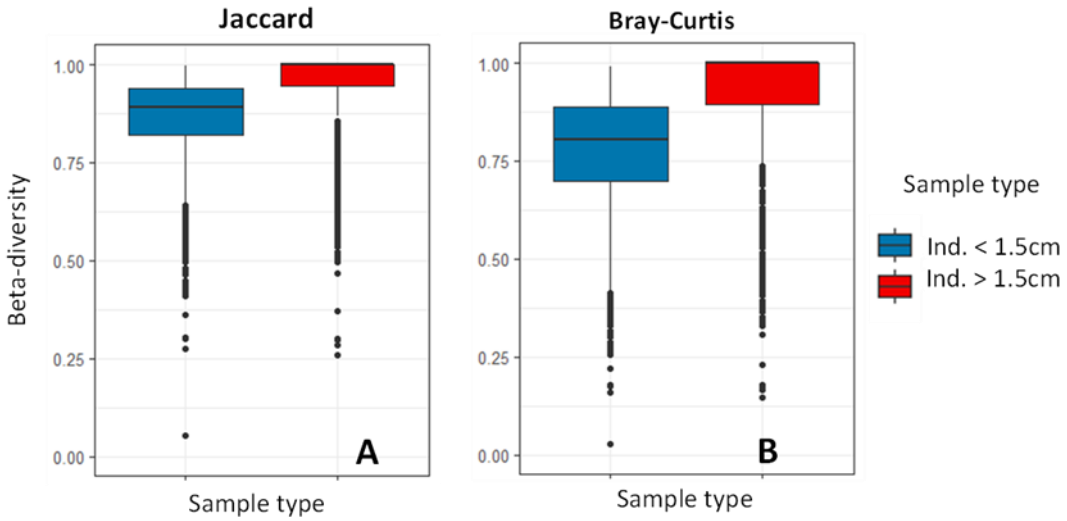

Figure 4. 5 Beta-diversity of different sample types. Both Jaccard based on presence-absence data (A) and Bray-Curtis on relative read counts (B) indices highlight that samples of small individuals showed a substantially higher median community similarity (lower dissimilarity, i.e. beta-diversity) than samples of large individuals. Displayed are boxplots of all possible pairwise within-group comparisons.

4.4.1.2 Taxonomic affiliations of OTUs

OTU identification against Genbank entries using Blast showed sequence similarity levels mostly outside of the widely applied 3% threshold of within-species diversity, indicating the lack of close relatives in the database. For most entries, the divergence to the nearest entry was within the range of 10-20% which is generally too distant for reliable identifications. However, matches within the 3% interval were obtained in all target groups, except Plecoptera, as follows: Ephemeroptera (2 of 66 OTUs), Trichoptera (17 of 71 OTUs), Coleoptera (1 of 30 OTUs), Hemiptera (11 of 97 OTUs), Odonata (24 of 75 OTUs), Diptera (53 of 506 OTUs), Decapoda (1 of 59 OTUs), and Mollusca (7 of 23 OTUs), resulting in an overall proportion of hits at the 3% level (presumed species-level) of 12.8%. The proportion of OTUs matched within the 3% level reflected known differences among major lineages in dispersal propensity and geographic ranges, which are generally highest in Odonata, followed by Trichoptera. These groups also showed the highest proportion of exact sequence matches or sequences with divergences of only 1 to 3 nucleotides. The geographic provenance of closely similar Genbank entries was dominated by China and Thailand, and to a smaller degree by India, Bangladesh, Malaysia, Pakistan, and Vietnam, presumably representing species that are widespread in the South Asian region. High-similarity matches from elsewhere were linked to cosmopolitan and invasive species. For example, the top Genbank hit of *Corbicula fluminea*, the Asian Clam, was a sequence obtained from Argentina, where this species is invasive. Two other molluscans with perfect Genbank matches, *Ferrissia fragilis* and *Mieniplotia scabrescens*, are originally from North America and the Indo-Pacific region, respectively, but were widely introduced into Europe and Eastern Asia. The odonatan *Pantala flavescens*, the Globe Skimmer, was previously sequenced from Liberia, but the species is the most widespread dragonfly in the world known for huge migrations (Troast et al., 2016).

4.4.1.3 Environmental variables

The environmental intactness of sampled sites, assessed based on interviews with local community members and visual observations of five disturbance categories, demonstrated a wide range of anthropogenic influences across the investigated streams (Fig. 4.6). While overall environmental intactness was very high in five streams, others displayed a relatively poor status. Among the five criteria, catchment intactness as a summary parameter for

disturbances beyond the immediate river ecosystem, was the most widely reduced measure across the river systems. In general, individual components of environmental intactness were positively correlated (Fig. 4.7) reflecting their joint dependency on human population pressure.

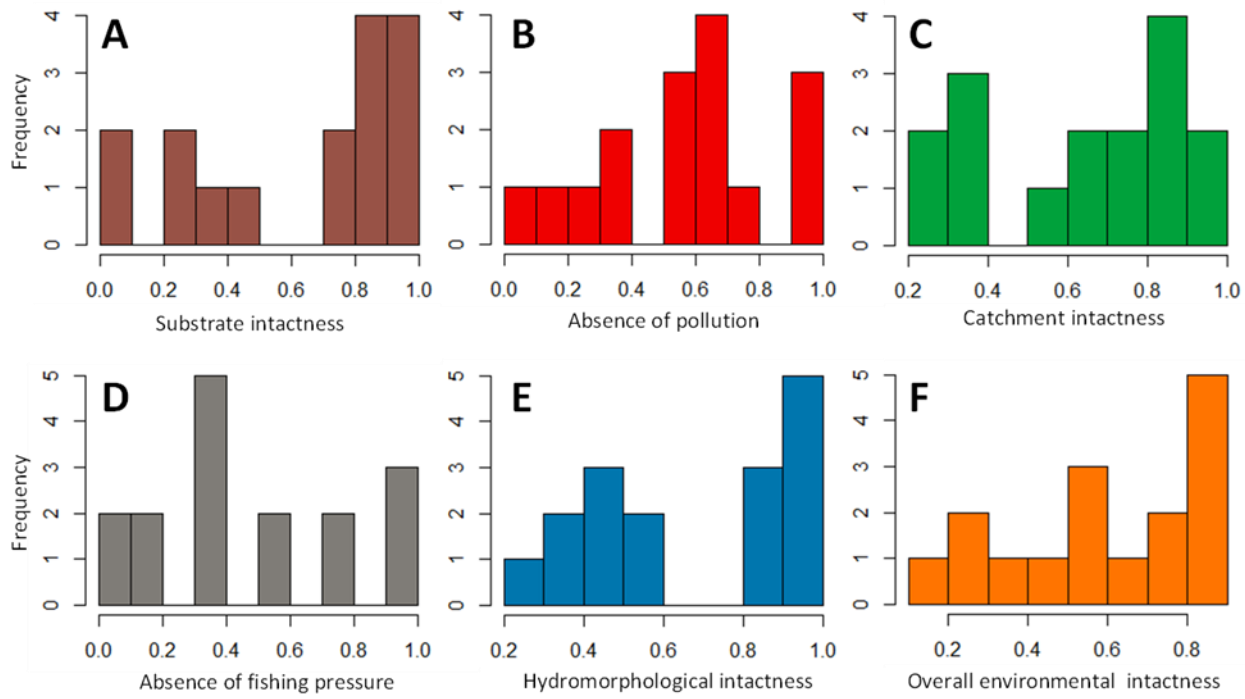

Figure 4. 6 Frequency of disturbance severity in 16 mountain streams, scoring each stream at five sampling sites for five environmental criteria, resulting in an average score between 0 (low intactness) and 1 (high intactness) for each river (**A-E**), which were used to calculate overall environmental intactness (**F**).

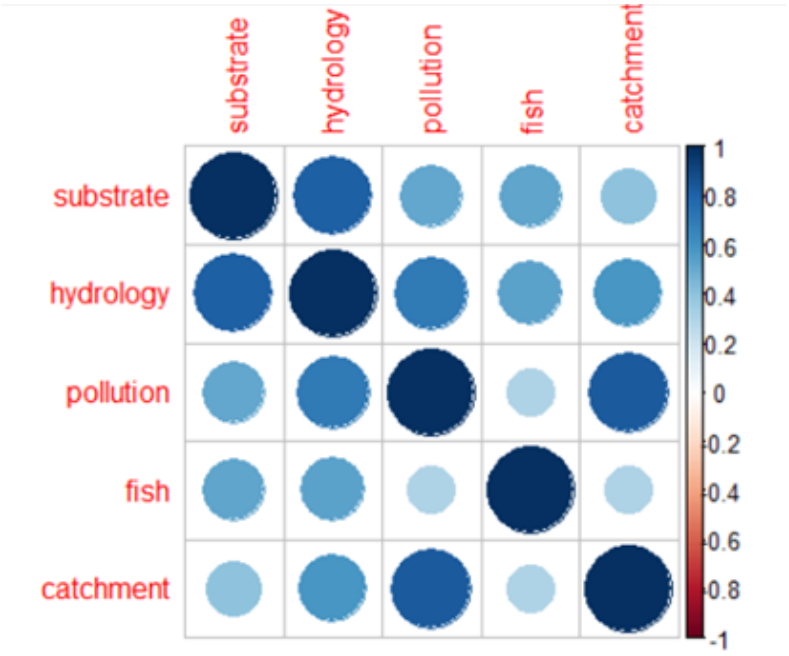

Figure 4. 7 Correlation matrix showing correlations between all measured components environmental intactness. Blue colour for positive correlation and red for negative correlation where the size and shade of the ball indicate the strength of the correlation.

4.4.1.4 Relationship of environmental variables with diversity measures of macroinvertebrates

Regression analysis

There was no significant relationship between overall environmental intactness with any of the alpha diversity measures, including OTU richness, evenness or phylogenetic diversity (p>0.20). In contrast, EPT species richness was highly sensitive to a number of anthropogenic disturbances. Decreases in EPT richness were well explained by models using multiple components of environmental intactness as predictors attained lower BIC scores and better model fits with models containing only a single component (Table 4.2 in the appendix). Hence, decreases in EPT richness was dependent on a number of habitat alterations that included hydromorphological and substrate alterations, pollution and high fishing pressure, strongly suggesting the joint importance of different environmental components. This was corroborated by the fact that using overall environmental intactness, i.e., the compound value from all five environmental parameters as explanatory variable, resulted in the model with the clearly lowest BIC (y = 22.8x +20.4, r2 = 0.49, p< 0.001, BIC= 107.0; Table 4.2 in the appendix). The importance of this relationship was reflected in the steep slope of the

regression highlighting that EPT richness was twice as high in streams with a high overall environmental intactness (Fig. 4.8-A). Similar to OTU richness, OTU evenness (y = 0.11x +0.40, r2 = 0.09, p = 0.006) and phylogenetic diversity (y = 0.03x +0.04, r2 = 0.45, p = 0.003) of EPT also showed best fit with overall environmental intactness. In contrast, EPT relative abundance was largely independent of environmental intactness (p = 0.82; Fig. 4.8-B). Therefore, although many EPT OTUs were absent in ecosystems exposed to high anthropogenic impacts, others apparently benefited from environmental changes.

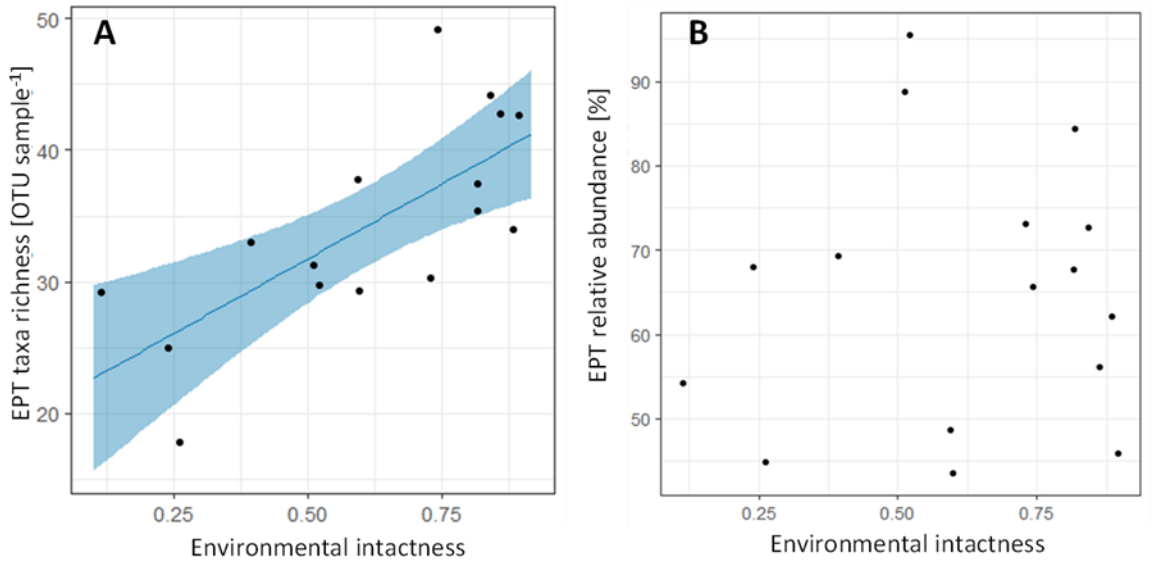

Figure 4. 8 The impact of overall environmental intactness on the richness (**A**) and abundance (**B**) of key indicator taxa (the Ephemeroptera, Plecoptera, Trichoptera fraction; EPT). Overall environmental intactness was calculated from five criteria (hydromorphological changes, substrate intactness, pollution, fishing pressure, and catchment intactness) evaluating human impact on mountain streams. The linear regression (y = 23x +20, r^2 = 0.52, p< 0.001) and the confidence interval (shading) are shown in (A).

Indicator Species Analysis

Finally, we investigated potential indicator taxa of anthropogenic activities using a partial-network approach. Our correlation-based species filtering resulted in the identification of 26 potential indicator taxa (Table 4.2, Fig. 4.9). Blast searches were conducted to determine their lowest identifiable taxon affiliation, which assigned these potential indicator OTUs to various taxonomic groups, with the highest number in Diptera (n = 13) followed by Ephemeroptera (n=4) (Table 4.2).

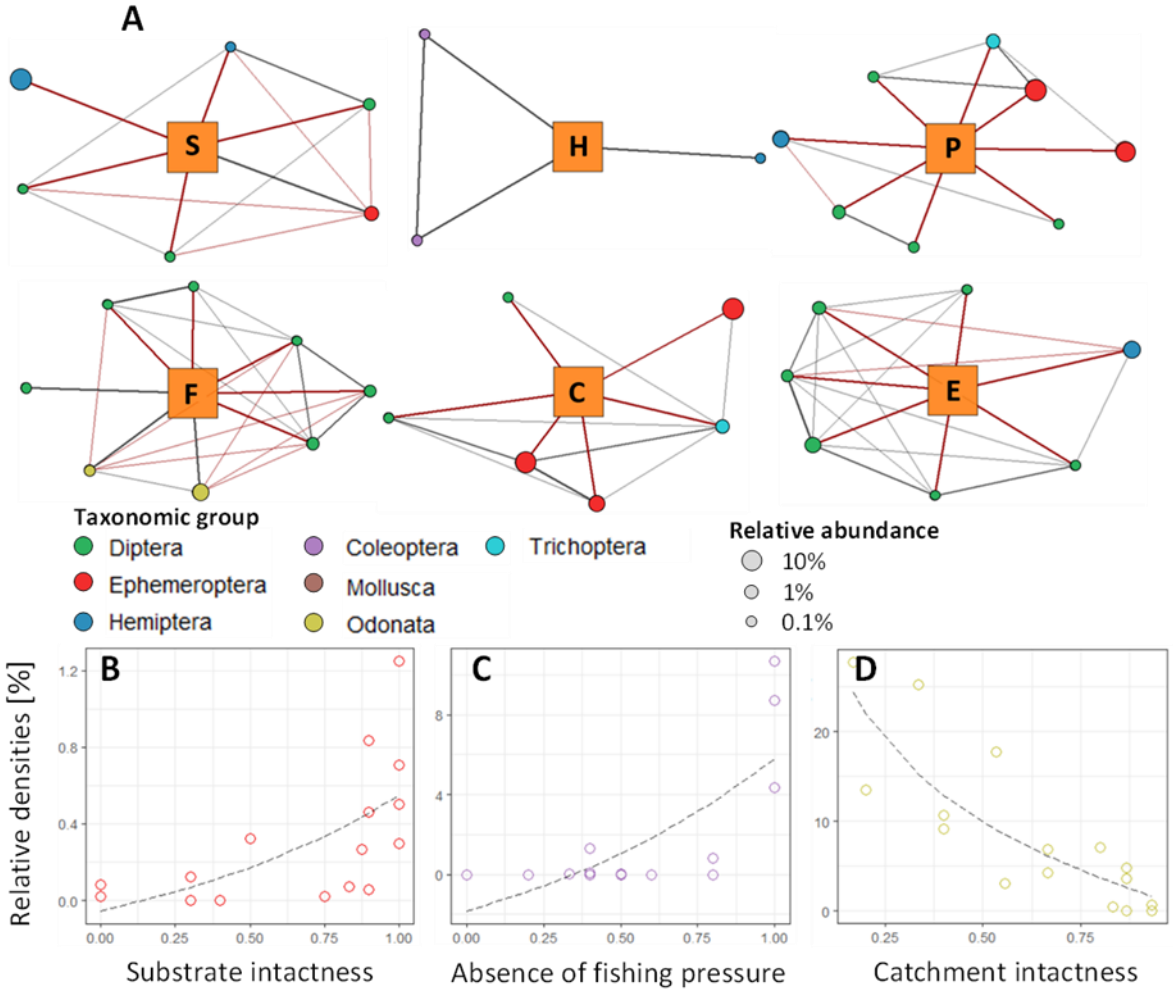

Figure 4. 9 (**A**) Partial networks linking the abundance of potential indicator species to environmental conditions and other potential indicator species. Orange squares represent different mechanistic drivers (S: substrate intactness, H: hydrological intactness, P: minimum pollution, F: absence of fishing pressure, C: catchment intactness, E: overall environmental intactness). The edges of the networks represent potential indicator species. The colour of the circles depicts taxonomic affiliations, whereas the size indicates their average contribution to total reads per sample. Red and grey lines indicate negative and positive relationships, respectively. Note the connections between species indicating interactions of co-distributed potential indicators. (**B**-**D**) Examples of the relationship between potential indicator OTUs and environmental conditions (OTU 174, 53 and 2 in Table 4B). Examples are: (**B**) a species of Ephemeroptera as a potential indicator of natural substrate intactness, (**C**) an Odonata affected negatively by fishing pressure, indicating the cascading effects of the removal of larger fish species, and (**D**) an Ephemeroptera species as a generalised disturbance indicator of catchment intactness.

Read abundance in each of these OTUs was associated with one or multiple criteria of environmental intactness, and potential indicator taxa were predominantly correlated with poor environmental conditions (Table 4.2). A few indicator OTUs of Coleoptera and Odonata interacted positively with hydromorphological intactness and minimum fishing pressure respectively and a highly sensitive indicator OTU of Ephemeroptera also was positively associated with substrate intactness. However, across all criteria, negative interactions had significantly higher relative frequency (T-test, T-value = 5.38, $p<0.001$) revealing a greater number of potential disturbance indicators than indicators of good ecological conditions (Fig. 4.9). Several potential indicator species also showed significant correlations with each other (e.g., sub-network for pollution) suggesting either a common ecological response or an effect of direct species interactions.

Table 4. 2. Potential indicator species identified from metabarcoding community data, their closest match with the NCBI and BOLD databases and their relation to facets of environmental intactness. Numbers in brackets indicate correlation coefficients. The BOLD assignments are presented in a second line under each taxon (where available). The environmental criteria according to their environmental intactness (i.e. "pollution" and "fishing" have a score of 0 in highly affected and of 1 in pristine environments).

OTU ID	Order	Family	NCBI (Genbank)/ BOLD result	Match [%]	Env. criterion
460	Coleoptera	Elmidae	*Stenelmis* sp.	85.20	hydrology (0.63)
			Stenelmis crenata	87.75	
463	Coleoptera	Elmidae	*Stenelmis fuscata*	86.33	hydrology (0.64)
			Stenelmis crenata	87.57	
106	Diptera	Ceratopogonidae	*Jenkinshelea* sp.	95.69	overall env. intactness (-0.65)
283	Diptera	Chironomidae	*Parachironomus* sp.	89.95	pollution (-0.73), fishing (-0.63),
			Parachironomus sp.	91.07	overall env. intactness (-0.74)
299	Diptera	Ceratopogonidae	*Bezzia fuliginata*	91.38	fishing (-0.68),
			Bezzia nigrita	91.05	overall env. intactness (-0.63)
406	Diptera	Chironomidae	*Conchapelopia togapallida*	88.27	substrate (-0.72), catchment (-0.65),
			Chironomidae	91.05	overall env. intactness (-0.67)
582	Diptera	Chironomidae	*Polypedilum decematoguttatum*	95.93	pollution (-0.65)
			Polypedilum sp.	97.58	
894	Diptera	Chironomidae	*Tanytarsus* sp.	89.49	overall env.
			Chironomid sp.	98.56	intactness (-0.69)

Table 4. 2. (Cont.)

OTU ID	Order	Family	NCBI (Genbank)/ BOLD result	Match [%]	Env. criterion
1047	Diptera	Chironomidae	*Chironomus* sp.	98.80	pollution (-0.74)
			Chironomus sp.	99.71	
1357	Diptera	Chironomidae	Benthalia dissidens	89.73	substrate (-0.65)
			Chironomus sp	91.58	
1383	Diptera	Chironomidae	*Polypedilum okiharaki*	96.17	fishing (-0.74)
			Polypedilum okiharaki	96.14	
1516	Diptera	Chironomidae	*Rheotanytarsus rivulophilus*	88.75	pollution (-0.68), fishing (-0.63)
			Chironomid sp	96.14	
1742	Diptera	Ceratopogonidae	*Bezzia fuliginata*	88.54	fishing (-0.68)
			Ceratopogonid sp.	93.39	
1788	Diptera	Ceratopogonidae	Ceratopogonidae sp.	84.65	fishing (0.64)
			Ceratopogonidae sp.	87.1	
3025	Diptera	Chironomidae	Cladotanytarsus sp.	96.89	substrate (-0.74), catchment (-0.71), overall env. intactness (-0.73)
			Chironomidae	100	
2	Ephemeroptera	Baetidae	*Baetis maculosus*	92.58	pollution (-0.74), catchment (-0.83)
			Baetis sp.	99.00	
17	Ephemeroptera	Siphlaenigmatidae	*Siphlaenigma janae*	87.11	pollution (-0.70), catchment (-0.64)
			Siphlaenigma janae	86.71	
21	Ephemeroptera	Baetidae	*Baetis maculosus*	92.34	catchment (-0.67)
174	Ephemeroptera	Ephemerellidae	*Torleya* sp.	91.60	substrate (0.62)
			Torleya sp.	91.55	
22	Hemiptera	Nepidae	*Laccotrephes maculatus*	98.19	pollution (-0.70), overall env. intactness (-0.70)
			Laccotrephes ruber	98.92	
29	Hemiptera	Naucoridae	*Heleocoris rotundatus*	87.53	substrate (-0.66)
			Heleocoris sp.	88.35	
2233	Hemiptera	Hebridae	*Hebrus axillaris*	87.38	substrate (-0.68)
3261	Hemiptera	Cicadellidae	*Typhlocybini* sp.	89.13	hydrology (0.64)
53	Odonata	Gomphidae	*Stylurus intricatus*	86.37	fishing (0.76)
			Nepogomphus walli	97.83	
724	Odonata	Libellulidae	*Orthetrum glaucum*	86.87	fishing (0.68)
			Orthetrum glaucum	87.1	
83	Trichoptera	Hydroptilidae	*Hydroptila thuna*	100	pollution (-0.64), catchment (-0.73)
			Hydroptila thuna	100	

4.4.2 Lowland rivers

A total of 3954918 reads passed the quality control and bioinformatics pipeline and were subsequently assigned to 2552 OTUs. The total number of reads per sample ranged from 5859 to 136727, with an average of 65915 reads. The total number of OTUs ranged from 82 to 900 with an average of 481 per sample. 2094 of the 2552 OTUs could only be assigned to the level of "Eukaryota" using the NCBI *nr* database. We retained altogether 662 OTUs of the target groups Insecta, Decapoda, Oligochaeta, Hirudinea and Mollusca, which represented 70.12% of total read abundance and 25.94% of OTU richness.

4.4.2.1 Diversity measurement of macroinvertebrates

Metabarcoding revealed a total of 152 ± 45 OTUs per site across 4 lowland rivers (Fig. 4.10-A). The small-specimen fractions (body length <15 mm) produced 126 ± 36 OTUs per site (mean of 47278 reads per sample) while the larger individuals contributed 29 ± 16 OTUs per site (9298 reads per sample). Across the 60 sampling sites, the evenness in smaller and large individual samples ranged between 0.46 and 0.41. Analysis of Faith's PD index showed that on average only 23 ± 6% of the total phylogenetic diversity was present in each sample (Fig. 4.10- A).

Oligochaeta were the most OTU-rich group, with an average of 32 ± 13 OTUs per sample though a total of 205 Diptera OTUs were found across all sites with an average of 26 ± 9 (Fig. 4.10-B). The combined diversity of Ephemeroptera, Plecoptera and Trichoptera (EPT) was lower, with an average of 6±8 OTUs across all sites. A total of 13 trichopteran OTUs were found only in 18 sites of 3 rivers whereas only two plecopteran OTUs were recorded in six sites of one river. The OTU richness of predator dominated taxa Coleoptera, Odonata and Hemiptera were calculated with an average of 5±3, 8±7 and 7±3 respectively per sample across the rivers. High diversity also was found for Decapoda and Mollusca, which each comprised 50 and 59 OTUs with 9±8, 9±5 per sample respectively. The second annelid taxa Hirudinea contained only 12 OTUs with a very low average (<1) (Fig. 4.10-B).

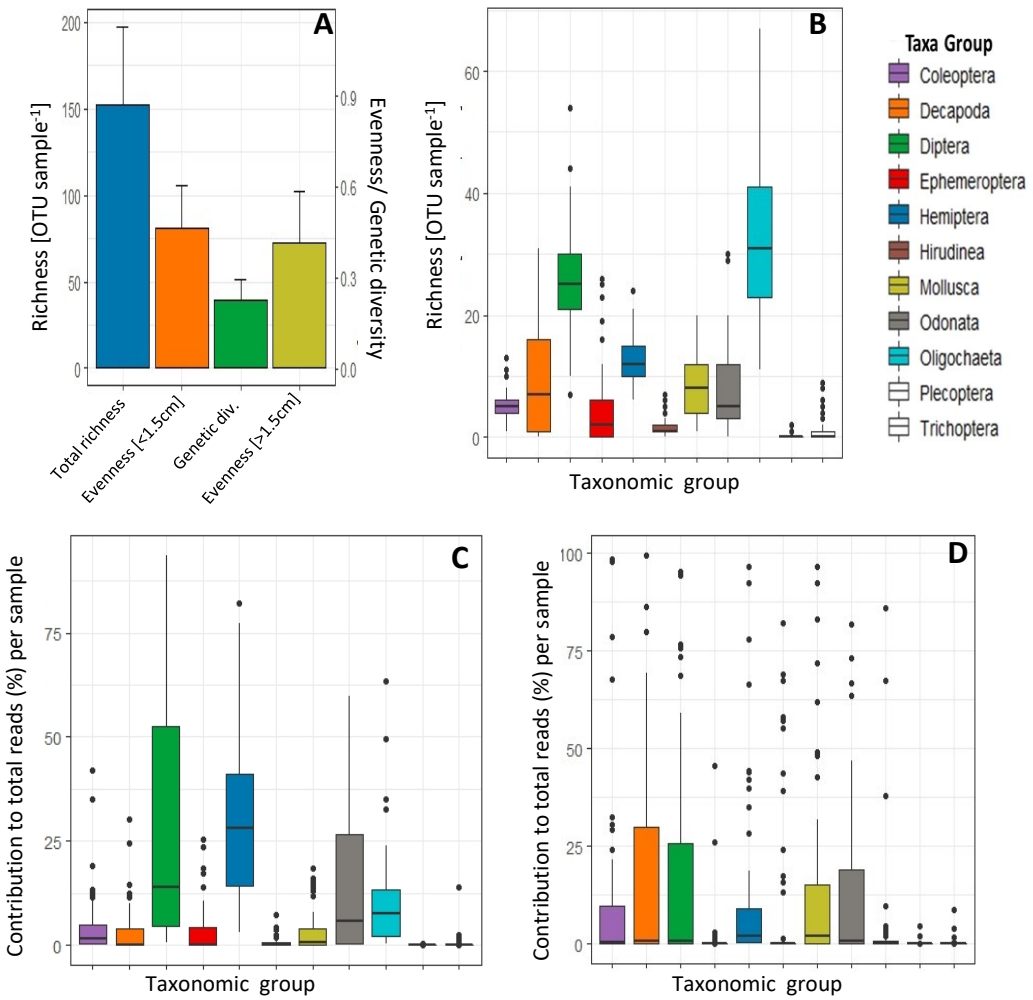

Figure 4. 10 Species diversity and relative abundance of analysed invertebrate taxa across all samples. (**A**) Average OTU richness, evenness (calculated separately for both sample groups) and genetic diversity of macroinvertebrates recorded in our study. Error bars denote standard deviations. (**B**) Boxplots of OTU richness per order (Bivalvia and Gastropoda have been merged into Mollusca) across all 60 sites. (**C**) Relative abundance (sequence reads) of different orders in the small- and (**D**) large-bodied fraction. C and D are based on different data collection protocols and hence the results are displayed separately.

In contrast to richness, relative abundance (calculated as the proportion of the total number of reads attributed to a taxon) for the small-bodied fraction was dominated by Hemiptera (Fig. 4.10-C), which had a high number of reads (30%) per site while it was only 28% and 10% for Diptera and Oligochaeta respectively. Notably, the relative abundance was very low for highly sensitive taxa Ephemeroptera (3%), Trichoptera (<1%) and Plecoptera (<1%). Both the OTU richness and proportion of read numbers significantly varied among the insect orders (Kruskal-Wallis Test, p<0.001). The large-bodied specimens were dominated by

Hemiptera and Mollusca, although there was a large degree of variation across samples evident from numerous outliers in boxplots (Fig. 4.10-D).

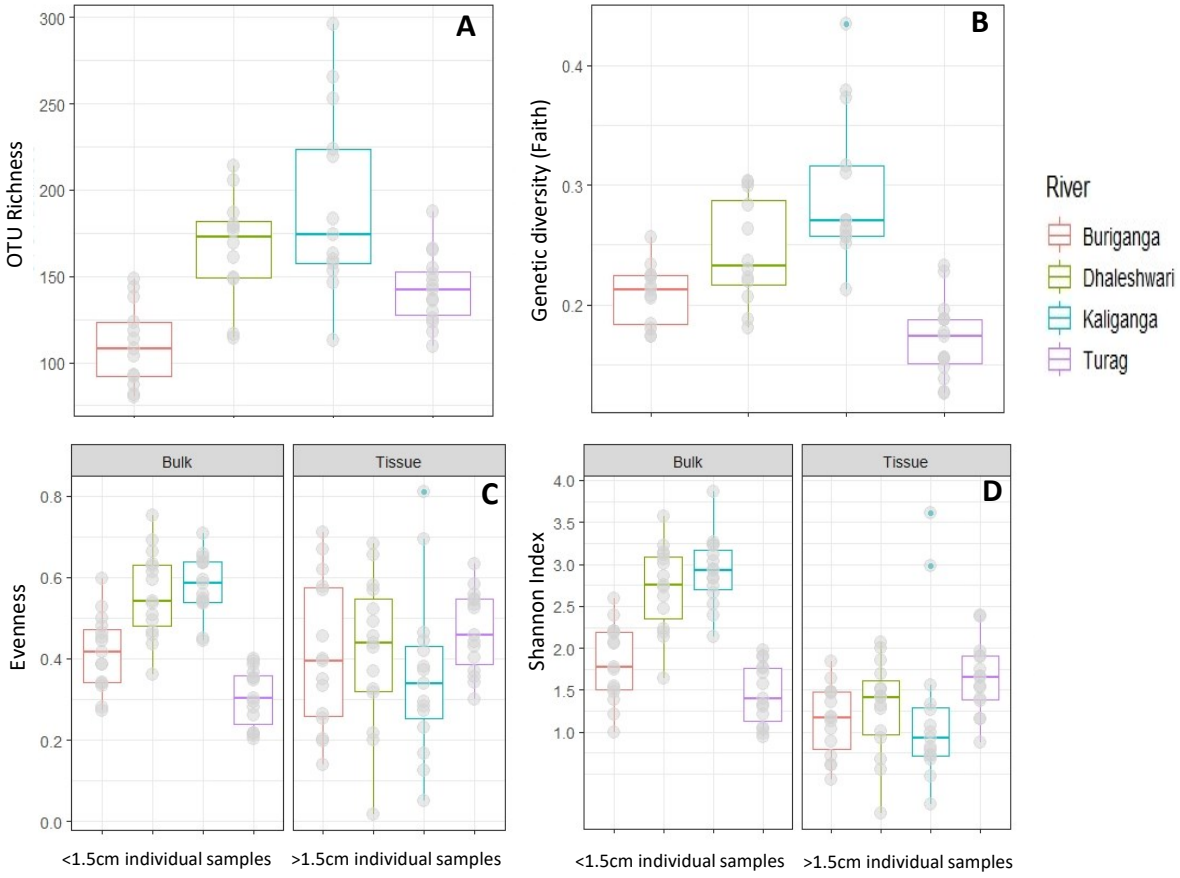

Figure 4. 11 Box plots showing the values of alpha diversity measures in four rivers (denoted by 4 colours). A-OTU Richness, B- alpha phylogenetic diversity index, C-evenness for small-bodied (Bulk) and large-bodied (Tissue), and D- The Shannon Diversity Index for small-bodied (Bulk) and large-bodied (Tissue).

4.4.2.2 Alpha and beta diversity measures of macroinvertebrates

The number of OTUs varied from 80 to 296 in four lowland rivers though there was apparently pairwise (river pair) similarity in species richness with the average number of OTUs between Dhaleshwari-Kaliganga and Buriganga-Turag. It was notable that the river Kaliganga supported almost double the number of OTUs than Buriganga (Fig. 4.11-A). One-way ANOVA ($p<0.001$) showed a significant variation in OTU richness among the rivers. A Post-hoc test (TukeyHSD) of ANOVA also showed pairwise significant differences between rivers ($p<0.001$) except for the pairs (Kaliganga-Dhaleshwari, $p= 0.25$; Turag-Dhaleshwari,

p= 0.28). The alpha phylogenetic diversity also significantly varied (Kruskal-Wallis Test, $p<0.001$) among the rivers with a range of 0.2 to 0.32 where the highest diversity was observed in Kaliganga River followed by Dhaleshwari. The Dunn Test also confirmed variations ($p<0.001$) in Faith genetic diversity between rivers (Fig. 4.11-B).

In case of evenness, there was no significant difference between small individual and large individual samples ($t = 1.6205$, $p = 0.1105$). Like OTU richness, evenness in smaller individual samples was also higher in the rivers Dhaleshwari and Kaliganga than in the other two rivers. However, in case of a larger individual fraction, the evenness was nearly similar in four rivers within a range of 0.33 to 0.45. The evenness measured with small individual samples significantly varied among (Kruskal-Wallis Test, $p<0.001$) and between (DunnTest, $p<0.001$) rivers. In contrast, larger-bodied samples did not show significant variation (Kruskal-Wallis Test, p= 0.2465) in evenness among rivers (Fig. 4.11-C). The Shannon diversity index of small individual fractions followed a similar pattern with genetic diversity where the highest value was estimated for the Kaliganga river followed by Dhaleshwari. Surprisingly, a reverse pattern of Shannon diversity in large individual fractions was observed for Kalinganga and Turag. The Shannon diversity significantly varied between small and large individual samples across the rivers (Wilcoxon rank-sum Test, $p<0.001$) (Fig. 4.11-D).

This significant variation was also reflected by higher beta-diversity in large-bodied and small-bodied fractions that was supported by rank-based dissimilarity analysis, ANOSIM (Analysis of similarity, distance =Bray-Curtis, p=0.001, R=0.78 and 0.44 for small individual and larger individual samples) (Fig. 4.12).

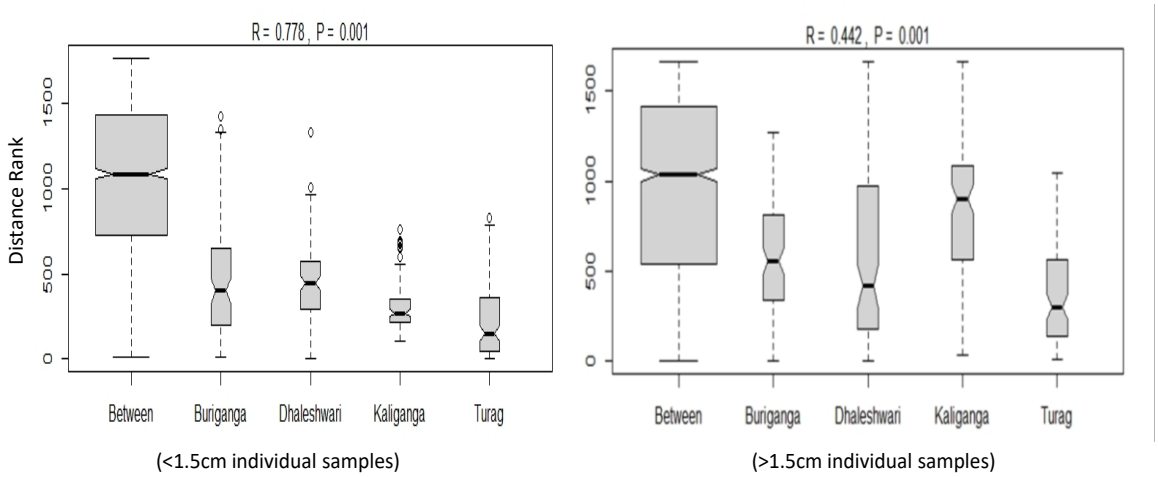

Figure 4. 12 ANOSIM plot (distance: Bray-curtis) for Beta-diversity in smaller individual and larger individual fractions of four rivers.

4.4.2.3 Taxonomic affiliations of OTUs

Like upland rivers, OTU identification against Genbank entries showed sequence similarity levels mostly outside of the widely applied 3% threshold of within-species diversity, indicating the lack of close relatives in the database. Matches within the 3% interval were almost double (24.62%) of upland's groups, as follows: Ephemeroptera (4 of 29 OTUs), Trichoptera (2 of 13 OTUs), Coleoptera (7 of 42 OTUs), Hemiptera (25 of 70 OTUs), Odonata (31 of 54 OTUs), Diptera (67 of 205 OTUs), Decapoda (7 of 50 OTUs), Oligochaeta (16 of 126 OTUs), Hirudinea (3 of 12 OTUs) and Mollusca (17 of 59 OTUs). The proportion of OTUs matched within the 3% level was generally highest in Diptera, followed by Odonata and Hemiptera. These groups also showed the highest proportion of exact sequence matches. The geographic provenance of closely similar Genbank entries was dominated by India and China and to a smaller degree by Thailand, Malaysia, and Pakistan, presumably representing species that are widespread in the South Asian region. Surprisingly, only a few OTUs of Decapoda (e.g., *Macrobrachium kistnense*) and Odonata (e.g., *Tholymis tillarga, Rhodothemis rufa*) perfectly matched with Genbank entries sequenced from Bangladesh. High similarity matches from elsewhere were linked to cosmopolitan and invasive species. For example, a red worm with a complete match, *Limnodrilus hoffmeisteri* is one of the most widespread and abundant oligochaetes in the world. One mollusc with perfect Genbank match, *Ferrissia fragilis* is originally from North America but was widely introduced into Europe and Eastern Asia. The top Genbank hit of *Rhopalosiphum*

nymphaeae, an aphid, was sequenced from South Korea, but this species is commonly found in Europe.

4.4.2.4 Environmental variables

The environmental condition of sampled sites, assessed based on onsite measurement of seven water quality parameters, demonstrated clear pairwise (river pair: Dhaleshwari-Kaliganga and Buriganga-Turag) differences of pollution impacts across four investigated rivers (Fig. 4.13). However, all parameters significantly varied (One way ANOVA, $p<0.001$) among the four rivers. The overall water quality was relatively better in Dhaleshwari-Kaliganga than Buriganga-Turag. The individual components of seven parameters (except DO) were positively correlated, reflecting their joint dependency on pollution pressure. As expected, OTU richness was positively correlated only with DO and had a negative interaction with other parameters (Fig. 4.14).

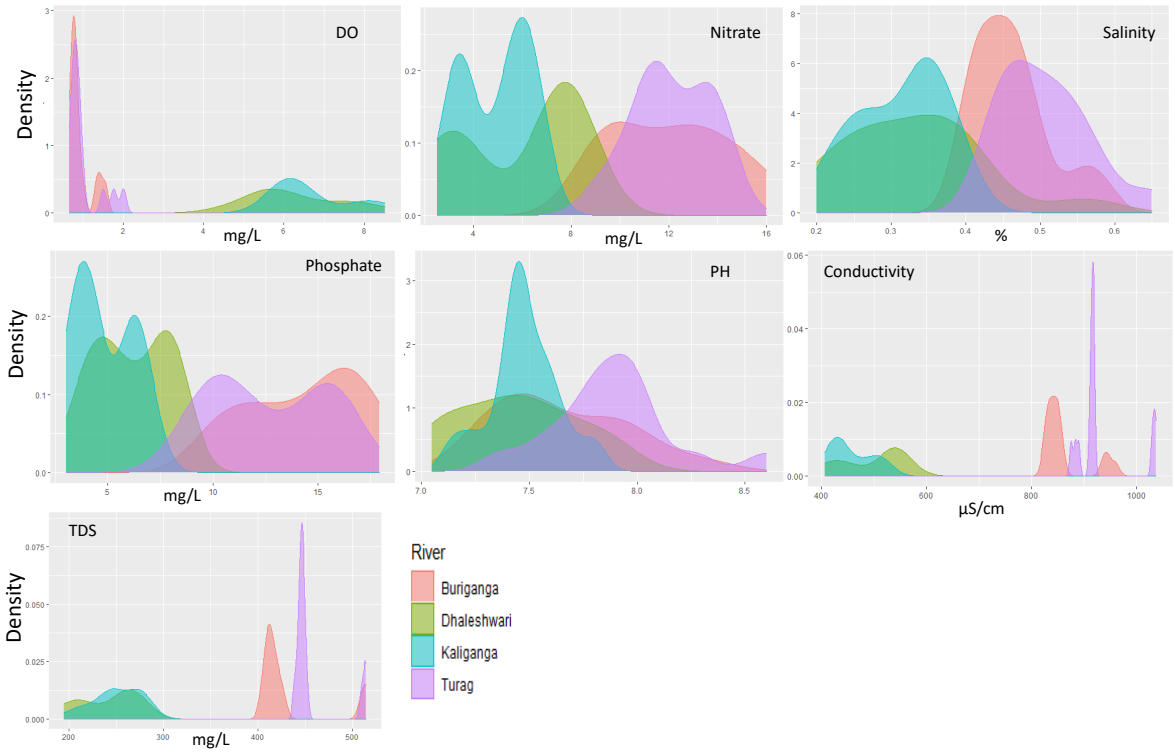

Figure 4. 13 Density plots showing the distribution of the proportion of seven water quality parameters (DO, nitrate, salinity, phosphate, P^H, conductivity and TDS) in each of the four rivers.

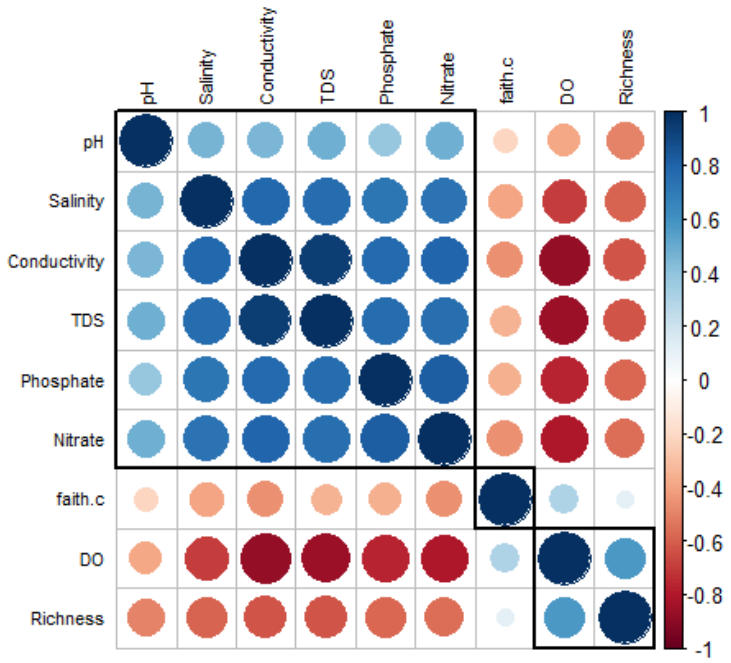

Figure 4. 14 Correlation matrix showing correlations between all measured environmental parameters and total richness across the sampled sites. Blue colour for positive correlation and red for negative correlation where the size and shade of the ball indicate the strength of the correlation.

4.4.2.5 Relationship of environmental variables with diversity measures of macroinvertebrates

Regression and correlation analysis

There was a significant relationship of environmental variables (DO, P^H, phosphate, nitrate, salinity, conductivity and TDS) with two alpha diversity measures (OTU richness and phylogenetic diversity) indicating species richness and genetic diversity was highly sensitive to a number of pollution disturbances (Fig 4.1 in the appendix). Correlation analysis also revealed a strong negative correlation between diversity measures and environmental variables except for DO. While there was also a positive relationship among covarying factors, DO negatively correlated with other factors (Fig. 4.14). Further, to investigate the important predictors of the highest impact on species and genetic diversity, the Principal Component Analysis with different covarying predictors supported the exclusion of DO, conductivity and TDS as they were causing additive effects. The first PCA axis represented almost equal loadings for nitrate, phosphate and salinity that could be triggered by strong correlations among predictors. The linear mixed-effect regression models with multiple

components of environmental variables (such as PCA axis 1 of nitrate, phosphate & salinity and P^H as predictors) and OTU richness and phylogenetic diversity (as response variables) produced the best model that attained lower AIC scores. Hence, decreases in species richness and genetic diversity were dependent on a number of environmental factors that included the high density of nitrate, phosphate and salinity strongly suggesting the joint importance of different environmental components. This was corroborated by the models, (y = 13x +108, r2 = 0.40, p< 0.001, AIC= 508.46) for species richness and (y = 1.2x +20, r^2 = 0.71, p= 0.005, AIC=240) for genetic diversity (Fig. 4.15-A, B). Surprisingly, there was no clear relationship between OTU richness and alpha phylogenetic diversity (Fig. 4.15-C).

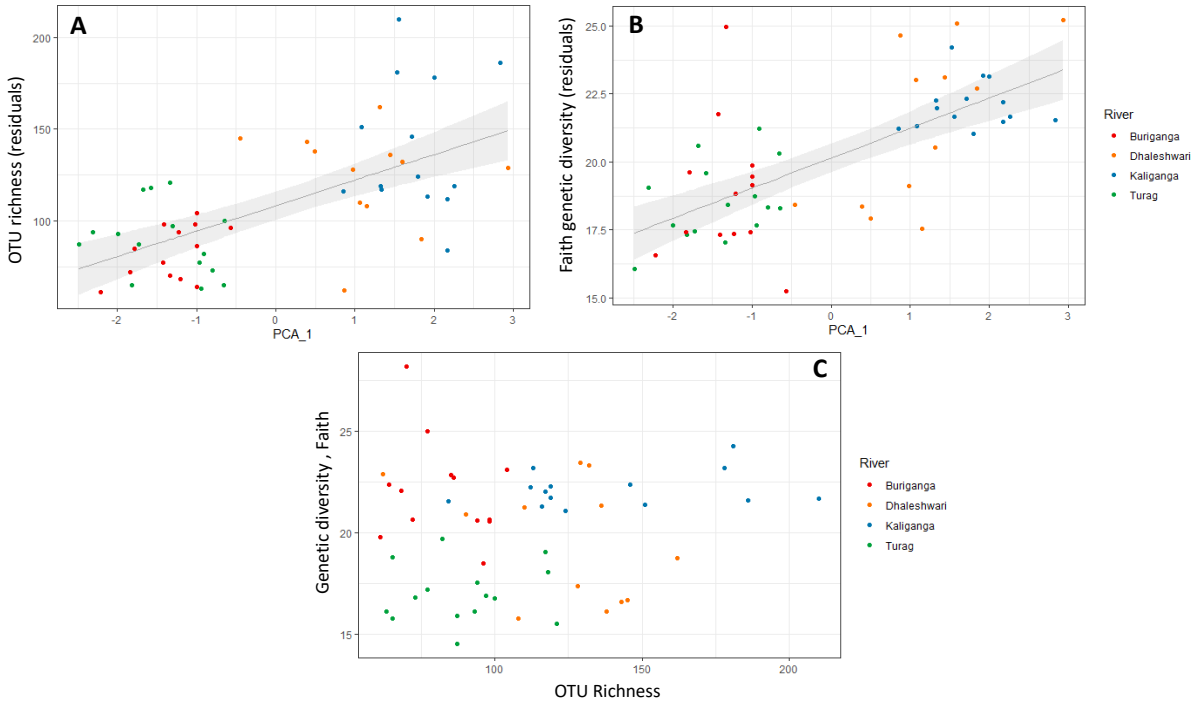

Figure 4. 15 The impact of hadrochemical parameters on the richness (**A**) and alpha phylogenetic diversity (**B**) of all target taxa. PCA was done with three parameters (Phosphate, Nitrate and Salinity). The mixed-effect linear regression A: (y = 13x +108, r^2 = 0.40, p< 0.001) and B: [(y = 1.2x +20, r^2 = 0.24 and 0.71 (marginal and conditional respectively), p= 0.005). The confidence intervals are shown by shading in A & B. Relationship between OTU richness and alpha phylogenetic diversity (C).

Ordination and variation partitioning analysis

RDA was performed to explore the relationship between environmental factors (predictor variables) and macroinvertebrate OTUs in sixty sites of four lowland rivers (Fig. 4.16-A). Results showed that macroinvertebrate composition was apparently affected by environmental factors. A total of 17.09% of the cumulative variance in OTUs was explained by the three RDA axes of which 81.22 % variance was explained by axis 1. Multiple environmental variables (e.g., Nitrate, phosphate and salinity) played a significant role in shaping the OTUs richness. A permutation test with a forward selection also supported the significance of explanatory variables (P= 0.001 after 999 random permutations) in governing the OTU richness. In RDA ordination, the length of the arrow was proportionally related to the importance of variables. Arrows of all variables were located at the same site in the RDA ordination indicating their positive relationships among them and the influences on species composition in the same direction. Finally, variation partitioning was done to quantify the proportion of variation in OTU composition (for both sample types) by each of the three explanatory variables (Fig. 4.16-B). The results of smaller individual samples showed that the conditional effect of nitrate (5.0%) and salinity (5.0%) was the same and higher than that of phosphate (1%). The higher shared variance (12 %) indicated that the species composition of macroinvertebrates was affected by the combined impacts of three hydro-chemical variables. The marginal (simple) effects of three environmental variables were also estimated as 17 %, 10% and 14% for nitrate, phosphate and salinity respectively. The simple (marginal) effects of the three predictors were significant (permutation test ANOVA, $p < 0.001$). The conditional (partial) effects of nitrate and salinity were also significant at p= 0.005 and p= 0.008 respectively while it was insignificant for phosphate (p=0.064). Notably, the effect of three predictors (nitrate, phosphate and salinity) was too low to estimate for larger individual samples.

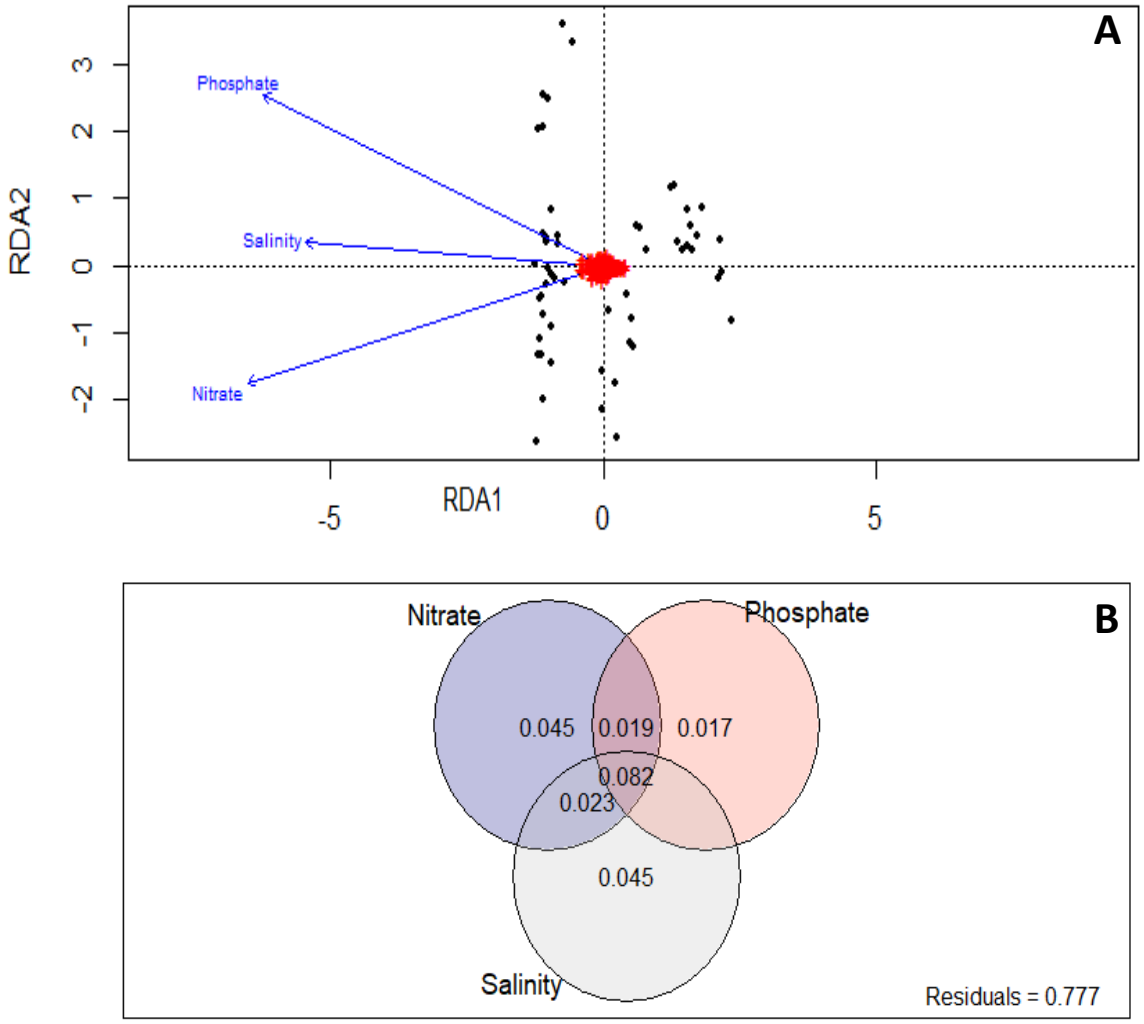

Figure 4. 16 Ordination diagram of Redundancy analysis (RDA) exhibits macroinvertebrate taxa (red shapes) and environmental variables (arrows). Black shapes denote the sampled sites. B. Venn diagram displaying the results of a variation partitioning analysis. Three environmental factors, nitrate, phosphate and salinity were used here as explanatory variables. The bounding rectangle represents the total variation in the response variable (here OTU richness) while each circle represents the portion of variation accounted for by an explanatory matrix or a combination of explanatory matrices.

Indicator species analysis

Finally, I investigated potential indicator taxa of anthropogenic activities and the correlation-based species filtering resulted in the identification of 16 potential indicator taxa (Table 4.3). Blast searches were conducted to determine their lowest identifiable taxon affiliation, which assigned these potential indicator OTUs to various taxonomic groups, with the highest number in Diptera (n = 5) followed by Odonata (n=4) (Table 4.3). Read abundance in each of these OTUs was associated with one or multiple parameters of water quality, and potential indicator taxa were predominantly correlated with poor environmental conditions (e.g., presence of low dissolved oxygen and higher p^H, nitrate, phosphate, salinity, TDS and conductivity). For instance, all dipteran (Culicidae and Chironomidae) OTUs negatively responded to dissolved oxygen and positively responded to P^H, nitrate, phosphate, and other water quality criteria. The same pattern of response was also observed for an alien invasive hemipteran (*Rhopalosiphum nymphaeae*) while for other hemipteran OTUs interacted positively with DO and negatively with other parameters. All indicator OTUs of Odonata, Decapoda and Coleoptera interacted positively with higher dissolved oxygen and negatively with higher contents of all other parameters. A single indicator OTU of Ephemeroptera also was positively associated with dissolved oxygen. However, across all criteria, negative and positive interactions had nearly equal relative frequency revealing a similar number of potential disturbance indicators and indicators of good ecological conditions.

Table 4. 3 Potential indicator species identified from metabarcoding community data, their closest match with the NCBI and their relation to environmental factors. Numbers in brackets indicate correlation coefficients.

OTU ID	Major taxon	Family	NCBI Result	Genbank Match (%)	Environmental criterion and Corr. Coeff.
1	Hemiptera	Aphididae	*Rhopalosiphum nymphaeae*	100	Ph (0.62), DO (-0.68), Salinity (0.55), conductivity (0.65), TDS (0.62), Phosphate (0.64), Nitrate (0.69)
2	Diptera	Culicidae	*Culex quinquefasciatus*	100	Ph (0.49), DO (-0.42), Salinity (0.55), conductivity (0.54), TDS (0.53)
3	Diptera	Culicidae	*Culex tritaeniorhynchus*	100	Ph (0.45), DO (-0.61), Salinity (0.61), conductivity (0.59), TDS (0.56), Phosphate (0.81), Nitrate (0.65)
7	Decapoda	Atyidae	*Caridina babaulti*	92.66	Ph (-0.34), conductivity (-0.29), TDS (-0.35)
9	Oligochaeta	Naididae	*Slavina* sp.	91.83	Ph (-0.38), conductivity (-0.27)
11	Hirudinea	Erpobdellidae	*Dina lineata*	84.45	Ph (0.56), DO (-0.61), Salinity (0.47), conductivity (0.59), TDS (0.60), Phosphate (0.61), Nitrate (0.48)
13	Diptera	Chironomidae	*Chironomus javanus*	100	Ph (0.59), DO (-0.58), Salinity (0.50,) Phosphate (0.45), conductivity (0.55), TDS (0.51), Nitrate (0.70)
14	Odonata	Libellulidae	*Brachythemis contaminata*	100	Ph (-0.35)
18	Decapoda	Atyidae	*Caridina babaulti*	90.96	Ph (-0.46), DO (0.32), Salinity (-0.37), conductivity (-0.32), TDS (-0.34)
19	Diptera	Ephydridae	*Brachydeutera* sp.	99.49	Ph (0.49), DO (-0.53), Salinity (0.50), conductivity (0.54), TDS (0.51), Phosphate (0.43), Nitrate (0.58)
20	Odonata	Pseudocopera ciliata	Platycnemididae	97.85	Ph (-0.33)

Table 4. 3(Cont.)

OTU ID	Major taxon	Family	NCBI Result	Genbank Match (%)	Environmental criterion and Corr. Coeff.
23	Ephemeropt-era	Baetidae	*Cloeon virens*	88.76	Ph (-0.34), DO (0.61), Salinity (-0.65), conductivity (-0.65), TDS (-0.57), Phosphate (-0.59), Nitrate (-0.60)
12	Hemiptera	Gelastocoridae	*Nerthra adspersa*	84.17	DO (0.33), Salinity (-0.34), conductivity (-0.35), TDS (-0.32), Nitrate (-0.37)
16	Odonata	Libellulidae	*Crocothemis servilia*	100	DO (0.44), Salinity (-0.48), TDS (-0.39), conductivity (-0.45), Phosphate (-0.52), Nitrate (-0.48)
22	Odonata	Pseudagrion microcephalum	Coenagrionidae	98.36	DO (0.42), Salinity (-0.47), conductivity (-0.48), TDS (-0.38), Phosphate (-0.37), Nitrate (-0.38),
24	Diptera	Diptera sp.	Diptera	91.87	DO (-0.31), Conductivity (0.34), TDS (0.42), Phosphate (0.43)

4.5 Discussion

The highlands and lowlands of Bangladesh currently face rapid human population growth and increases in anthropogenic stressors (Aukema et al., 2017). Mountain streams and lowland rivers are functionally important landscape elements in this region, and I, therefore, assessed various types of ecosystem degradation using the sensitivity of macroinvertebrate communities to disturbance. The overall diversity of mountain streams was high, and a total of 936 OTUs (as proxy of species-level entities; see below) of macroinvertebrates were found in the three target groups (Insecta, Decapoda and Mollusca) in 16 second-order streams. From four lowland rivers, more than 650 OTUs were explored under the four major taxa groups viz. Insecta, Decapoda, Annelida and Mollusca. Insect OTUs belonged to Coleoptera, Diptera, Ephemeroptera, Hemiptera, Odonata, Plecoptera and Trichoptera. Most of these OTUs were not represented in public databases, potentially indicating high levels of endemicity or a general lack of DNA taxonomic data in many remote regions of the world (McGee et al., 2019).

In highland streams, total OTU richness was not greatly altered by anthropogenic stressors, the presumably sensitive EPT lost approximately half of their diversity along the studied disturbance gradient, similar to findings from a global survey of temperate and tropical regions (Eriksen et al., 2021a; Jähnig et al., 2021). In addition, I found great species turnover among sites and in response to various disturbance regimes. In case of lowland rivers, there was an obvious negative response of species to the alteration of hadrochemical elements in the water. The decrease in OTU richness and phylogenetic diversity was highly associated with the increase of some pollution borne chemical parameters in particular phosphate, nitrate and salinity. This high community sensitivity highlights the urgent need to closely monitor anthropogenic impacts in the region and to implement a regular species monitoring program to assess overall environmental intactness.

As field visits to the highland region were constrained by logistic issues, the assessment of environmental variables was limited to 14 physical variables, which were then averaged to obtain one overall environmental intactness value. This approach is a simplification, and it can be argued that the unweighted averaging is somewhat arbitrary. However, the approach is valuable and largely robust. First, different environmental characteristics such as an intact catchment and the absence of pollution, etc. were correlated as they all are the result of

human activity, which varies across rivers (Fig. 4.7). These consistently positive correlations imply that I do not obscure anthropogenic impacts by averaging different environmental characteristics but instead obtain a more complete picture of the overall effects of humans on the ecosystem. Second, the simplification of all ecosystem characteristics into an overall environmental intactness score has the advantage of providing an easily understandable index. For lowland rivers, seven environmental variables (hydrochemical parameters) and correlation analysis also revealed a strong negative correlation between diversity measures and water quality variables except for DO (Fig. 4.14). Among covarying factors, there was a positive relationship, whereas DO negatively correlated with other factors. The linear mixed-effect regression and ordination analysis strongly suggested the joint importance of different water quality components for shaping the invertebrate composition of lowland rivers. In particular, the changes in nitrate and salinity contents equally affected the invertebrate community of lowland rivers (Fig. 4.16-A). These findings can be very valuable in communications with decision-makers in the region and convince them of necessary conservation actions.

Given these poorly known ecosystems, our method for characterising the species diversity differed from studies in better known temperate faunas that usually start with a reference set of validated DNA barcodes against which the metabarcoding sample is identified. Here, taxonomic processing is reversed in that these entities were used for *de novo* generation of a reference set that needs to be validated in future. I retained specimen vouchers for the detailed taxonomic and phylogenetic characterisation of species. However, even the preliminary recognition of indicator species will require a solid definition of biological entities that can be linked to the ecological status of water bodies. OTU clustering at the 97% similarity threshold is widely used for approximating the species-level, based on an underlying phenetic species concept that defines species as genetic clusters separate from other such sets (Mallet, 1995). With improving read quality in sequencing technology, Amplified Sequence Variants (ASVs; Callahan et al., 2016) that correspond to the presumed true genotype variation may be employed, for a greater resolution possibly at the population level or for employing advanced methods of species delimitation from genotype information (e.g., Pons et al. 2006). However, while grouping of individuals into species-level entities is required given that we assess species traits for the response to environmental parameters, an ASV-based approach for species delimitation remains largely untested, and the improvement over standard OTU clustering may be comparatively small. Finally, we used the number of

reads representing each cluster as a measure of local species abundance, given a broad, albeit sometimes weak correlation of read abundance and biomass (Elbrecht and Leese, 2015). Improving these values through taxon-specific calibration and spike-ins as internal standards correcting for among-library variation will be needed to refine the indicator species modelling (see Ji et al., 2020; Creedy et al., 2019).

A crucial step in our assessment was to attain representative sampling of the local invertebrate communities. To reduce sampling error, field sampling used a triplicate design at each site to minimise the effects of small-scale heterogeneity. Further, each highland stream was sampled at five sites along a 1 km transect to reduce the effect of sampling stochasticity and microheterogeneity of species distributions. In each of the four lowland rivers, five sites were sampled at 2000-3000 m river section distances and three replicate samples were taken within 400-600 m sampling reach of each site. The repeatability of DNA extraction and PCR amplification was also tested using multiple technical replicates from three bulk homogenate samples (Betchora, Sangukhiang, Cheihkhiang), which each recovered the same respective communities for the overwhelming majority of OTUs, showing that the DNA extraction and sequencing protocol produce a largely complete record of the species present in each sample (Chapter 3). Finally, we processed large and small-bodied specimens separately, to limit stochasticity and to increase species detection across size classes.

Clear differences were evident between the two body size classes. Besides profound differences in composition, I also found a greater turnover of large-bodied species across the streams (Figs. 4.4, 4.5, 4.10, 4.12). Body size impacts several key parameters that influence species distribution, population size (carrying capacity), and the scale of active and passive dispersal. Hence, high turnover in large-bodied species may be a result of comparatively low abundance and thus sampling probability. However, higher beta diversity also may indicate differences in dispersal rates constrained by cross-valley active movements that potentially represent greater barriers to large-bodied species. The separate processing of different body size classes requires substantial extra efforts, but this may be justified as the apparent differences in abundance and turnover hold important implications for ecosystem management and conservation of species of different body sizes.

Further, in highland streams, the total species diversity did not greatly change as a function of environmental intactness. This is not overly surprising as intermediate levels of

anthropogenic stress do not necessarily lead to loss of overall diversity (Primack et al., 2018). For example, mild pollution and associated eutrophication may even lead to an increase in species richness highlighted by hump-shaped productivity-biodiversity relationships found in many systems (Grace et al., 2016). However, total species diversity presents the sum of numerous individual species responses, and here this hides the substantial loss of species in the EPTs (Fig. 4.8). But even in these groups, responses were not universal, as illustrated by a widespread OTU of Ephemeroptera that responds positively to poor overall environmental intactness, as well as to each of the individual anthropogenic stressors (Table 4.2 – OTU ID 2, Fig. 4.9B). Thus, the results indicate a mixture of positive and negative shifts in species responses and a profound reorganisation of macroinvertebrate assemblages in the wake of anthropogenic change.

The partial network analysis revealed only a small number of potential indicators of 'good' environmental conditions, i.e., very few OTUs were consistently present across undisturbed habitats (Fig. 4.9). In contrast, the number of OTUs designated as indicators of disturbed conditions was greater, as a set of species seemed to either directly or indirectly (through reduced competition) profit from anthropogenic disturbances. This dominance of disturbance indicators highlights the increase of widespread species under environmental degradation that are recovered consistently across sites and ultimately indicate the homogenisation of species assemblages (Carvalheiro et al., 2013), even when total species richness is not changing decisively. Vice versa, the scarcity of indicators for environmental intactness reflects the high heterogeneity of species assemblages at undisturbed sites and underlines the high overall level of biodiversity found in the highlands of Eastern Bangladesh.

Using conventional methods, the designation of indicator species requires a large number of observations for establishing correlation with the habitat status of solidly identified species. Metabarcoding greatly simplifies the problem of taxonomic identifications in assessing complex but poorly studied communities, and thus makes it easier to find the candidate species among the many species to be assessed. Nine of twenty-six potential indicators were among the ~100 top-most abundant metazoan OTUs (by read numbers) across the dataset. However, because of multiple statistical testing, which is part of our screening approach, there remains uncertainty about the candidate taxa and their environmental associations. Further field experiments are necessary to confirm the current

findings and allow their ultimate use as valid indicators. The analysis of sub-networks assists in the process, because it helps to visualize potential indicator species within their community context. If indicator species show strong positive co-occurrence, this may either indicate multiple species' independent responses to the environmental driver or constitute an indirect response mediated via synergistic relationships with a co-occurring species. The latter would perhaps decrease the indicator value of this species, but also provide a route towards the study of complex interaction webs in the river ecosystem and their sensitivity to environmental degradation.

Many countries of the Global South face the problem of biodiversity loss associated with rapid industrialisation and urbanisation, in what are frequently the most species-rich biomes on Earth (Allen et al., 2012; Eriksen et al., 2021a). Consequently, calls are being made for concerted efforts and deployment of resources to advance the study of taxonomic diversity and overcome the uncertainty in trends of change in freshwater communities globally (Van Klink et al., 2021; Jähnig et al., 2021; Maasri et al., 2021). The use of indicator species for ecological status assessment widely established in North America and Europe is not easily transferable to the diverse and less well-known ecosystems of tropical and subtropical regions (Morse et al., 2007). However, recent attempts to compile relevant studies at the global scale are underway, showing for example the high sensitivity of EPT in any ecosystem around the world (Eriksen et al., 2021a). These studies are mostly conducted at genus and family level only, but metabarcoding is an obvious tool to address the 'taxonomic impediment' for aquatic bioassessment (Hering et al., 2018; Eriksen et al., 2021a). We show here that with this approach even a single short-term field study can provide a basic reference set for biodiversity monitoring, including the preliminary designation of indicators of disturbance and pristine conditions. The key question now is about the reliability and specificity of these indicators, which needs to be studied across further sites and complemented by experimental approaches, and the degree of endemism of these potential indicators, which determines if they can be used universally at wider geographical scales across the Indo-Malayan hotspot.

4.6 References

Ahad, M. A., Raja, S. I, Khatun, M.M, and Hossain, M.A., 2012. Aquatic Insect Biodiversity of Northern Region of Bangladesh. *International Journal of Bio-resource and Stress Management*, *3*(3), pp. 387-393.

Ahmed, A. T. A., Rahman, M., and Mandal, S., 2013. Biodiversity of Hillstream fishes in Bangladesh. *Zootaxa*, *3700*(2), pp.283–292.

Allen, D.J., Smith, K.G., and Darwall, W. R. T., 2012. The status and distribution of freshwater biodiversity in Indo-Burma. Cambridge, UK and Gland, Switzerland: IUCN.

Andújar, C., Arribas, P., Gray, C., Bruce, C., Woodward, G., Yu, D. W., and Vogler, A. P., 2018. Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular Ecology*, *27*(1), pp.146–166.

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M., and Vogler, A. P., 2016. Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology and Evolution*, *7*(9), pp. 1071–1081.

Arthington, A. H., Naiman, R. J., McClain, M. E., and Nilsson, C., 2010. Preserving the biodiversity and ecological services of rivers: New challenges and research opportunities. *Freshwater Biology*, *55*(1), pp.1–16.

Aukema, J. E., Pricope, N. G., Husak, G. J., and Lopez-Carr, D., 2017. Biodiversity areas under threat: Overlap of climate change and population pressures on the world's biodiversity priorities. *PLoS ONE*, *12*(1).

Bashar, K., Reza, M.S., Razzak, M.A., Rahman, K.M.Z., Goda. P., and Howlader, A.J., 2014. Faunistic study of Odonata in some selected regions of Bangladesh. *Journal of Entomology and Zoology Studies*, 2 (4), pp.1-6.

Begum, M., Ismail, M., and Ali, S., 1989. A note on Freshwater mussels of Bangladesh. *Bangladesh J. Zool. 17*(2), pp. 203-204.

Bai, Z. G., 2006. Assessing land degradation in the Chittagong Hill Tracts, Bangladesh, using NASA GIMMS. Report 2006/06, ISRC-World Soil information. Wageningen. The Netherland.

Beermann, A. J., Zizka, V. M. A., Elbrecht, V., Baranov, V., and Leese, F., 2018. DNA metabarcoding reveals the complex and hidden responses of chironomids to multiple stressors. *Environmental Sciences Europe*, *30*(1).

Beng, K. C., Tomlinson, K. W., Shen, X. H., Surget-Groba, Y., Hughes, A. C., Corlett, R. T., and Slik, J. W. F. (2016). The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Scientific Reports*, *6*.

Bonada, N., Prat, N., Resh, V. H., and Statzner, B., 2006. Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. *Annual Review of Entomology*, *51*, pp.495–523.

Borcard, D., Legendre, P. and Drapeau, P., 1992. Partialling out the spatial component of ecological variation. *Ecology*, *73*(3), pp.1045-1055.

Bush, A., Monk, W.A., Compson, Z.G., Peters, D.L., Porter, T.M., Shokralla, S., Wright, M.T., Hajibabaei, M. and Baird, D.J., 2020. DNA metabarcoding reveals metacommunity dynamics in a threatened boreal wetland wilderness. *Proceedings of the National Academy of Sciences*, *117*(15), pp.8539-8545.

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), pp.581–583.

Carew, M.E., Pettigrove, V.J., Metzeling, L. and Hoffmann, A.A., 2013. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in Zoology*, *10*(1), pp.1-15.

Carew, M. E., Kellar, C. R., Pettigrove, V. J., and Hoffmann, A. A., 2018. Can high-throughput sequencing detect macroinvertebrate diversity for routine monitoring of an urban river? *Ecological Indicators*, *85*, pp.440–450.

Carvalheiro, L. G., Kunin, W. E., Keil, P., Aguirre-Gutiérrez, J., Ellis, W. N., Fox, R., Groom, Q., Hennekens, S., van Landuyt, W., Maes, D., van de Meutter, F., Michez, D., Rasmont, P., Ode, B., Potts, S. G., Reemer, M., Roberts, S. P. M., Schaminée, J., Wallisdevries, M. F., and Biesmeijer, J. C., 2013. Species richness declines and biotic homogenisation have slowed down for NW-European pollinators and plants. *Ecology Letters*, *16*(7), pp.870–878.

Cáceres, M. D., and Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. *Ecology*, *90*(12), pp.3566-3574.

Chapin, F.S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., Hooper, D. U., Lavorel, S., Sala, O. E., Hobbie, S. E., Mack, M. C., and Díaz, S., 2000. Consequences of changing biodiversity. *Nature, 405* (6783), pp.234-242.

Chowdhury, G. W., Gallardo, B., and Aldridge, D. C., 2016. Development and testing of a biotic index to assess the ecological quality of lakes in Bangladesh. *Hydrobiologia*, *765*(1), pp.55–69.

Chowdhuri, S.H., Aktaruzzaman, M., 1981. Dragonfly (odonata:Anisoptera) Larvae from Chittagong. *Bangladesh J. Zool., 9*(2), pp.131-144.

Creedy, T. J., Ng, W. S., and Vogler, A. P., 2019. Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. *Ecology and Evolution*, *9*(6), pp.3105–3116.

Davies, P.E., 1994. Monitoring river health initiative. River bioassessment manual. National River Process and Management Program. Freshwater system, Tasmania.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A. H., Soto, D., Stiassny, M. L. J., and Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews of the Cambridge Philosophical Society*, *81* (2), pp.163–182.

Edgar, R. C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), pp.2460–2461.

Edgar, R. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. BioRxiv, 081257.

Elbrecht, V., and Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PloS one*, *10*(7), e0130324.

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., and Leese, F., 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, *8*(10), pp.1265–1275.

Elbrecht, V., Bourlat, S. J., Hörren, T., Lindner, A., Mordente, A., Noll, N. W., ... and Zizka, V. M., 2021. Pooling size sorted Malaise trap fractions to maximize taxon recovery with metabarcoding. *PeerJ*, 9, e12177.

Emilson, C. E., Thompson, D. G., Venier, L. A., Porter, T. M., Swystun, T., Chartrand, D., Capell, S., and Hajibabaei, M., 2017. DNA metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an environmental gradient. *Scientific Reports*, *7*(1).

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., and Borsboom, D., 2012. qgraph: Network Visualizations of Relationships in Psychometric Data. *JSS Journal of Statistical Software, 48*.

Eriksen, T. E., Brittain, J. E., Søli, G., Jacobsen, D., Goethals, P., and Friberg, N., 2021a. A global perspective on the application of riverine macroinvertebrates as biological indicators in Africa, South-Central America, Mexico and Southern Asia. *Ecological Indicators*, *126*, 107609.

Eriksen, T. E., Friberg, N., Brittain, J. E., Søli, G., Ballot, A., Årstein-Eriksen, E., ... and Braaten, H. F. V. (2021b). Ecological condition, biodiversity and major environmental challenges in a tropical river network in the Bago District in South-central Myanmar: First insights to the unknown. *Limnologica*, *86*, 125835.

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation, 61*.

Fowler, W., 1912. The Fauna of British India, Including Ceylon and Burma. Vol. Coleoptera. General introduction and Cicindelidae to Paussidae. Taylor and Francis, London, p.529.

Fraser, F.C., 1934. The Fauna of British India, Including Ceylon and Burma. Vol. Odonata. 2: Agriidae, Gomphidae. Taylor and Francis, London, p.398.

Fraser, F.C. (1936). The Fauna of British India, Including Ceylon and Burma. Vol. Odonata. 3 Cordulegasteridae, Aeshnidae, Libellulidae. Taylor and Francis, London, p. 461.

Gibson, J.F., Shokralla, S., Curry, C., Baird, D.J., Monk, W.A., King, I. and Hajibabaei, M., 2015. Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. *PloS one*, *10*(10), p.e0138432.

Grace, J. B., Anderson, T. M., Seabloom, E. W., Borer, E. T., Adler, P. B., Harpole, W. S., Hautier, Y., Hillebrand, H., Lind, E. M., Pärtel, M., Bakker, J. D., Buckley, Y. M., Crawley, M. J., Damschen, E. I., Davies, K. F., Fay, P. A., Firn, J., Gruner, D. S., Hector, A., … Smith, M. D. 2016. Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature*, *529*(7586), pp.390–393.

Hajibabaei, M., Spall, J.L., Shokralla, S. and van Konynenburg, S., 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, 12(1), pp.1-10.

Hawkes, H.A., 1997. Origin and development of the biological monitoring working party score system. *Water Research*, *32* (3).

Hering, D., Borja, A., Jones, J. I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A., and Kelly, M., 2018. Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Research,*138, pp.192–205.

Hsieh, T. C., Ma, K. H., and Chao, A., 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, *7*(12), 1451–1456.

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H. J., and Tappu, R., 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Computational Biology*, *12*(6).

Hughes, R.M. (1995). Defining acceptable biological status by comparing with reference conditions, in: Davies, W. S. & T. P. Simon (eds,), Biological Assessment and Criteria. Tools for Water Resource Planning and Decision Making. Lewis Publishers, Boca Raton, FL: 31–47.

Irish, S.R., Al-Amin, H.M., Alam, M.S., Harbach, R.E. (2016). A review of the mosquito species (Diptera: Culicidae) of Bangladesh. *Parasites & Vectors*, 9:559.

Jähnig, S. C., Baranov, V., Altermatt, F., Cranston, P., Friedrichs-Manthey, M., Geist, J., ... and Domisch, S., 2021. Revisiting global trends in freshwater insect biodiversity. *Wiley Interdisciplinary Reviews: Water*, *8*(2), p.e1506.

Ji, Y., Huotari, T., Roslin, T., Schmidt, N. M., Wang, J., Yu, D. W., and Ovaskainen, O., 2020. SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes. *Molecular Ecology Resources*, *20*(1), pp.256-267.

Kuntke, F., de Jonge, N., Hesselsøe, M., and Lund Nielsen, J., 2020. Stream water quality assessment by metabarcoding of invertebrates. *Ecological Indicators*, *111*.

Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., Mergen, P., Pawlowski, J., Piggott, J., Rimet, F., Steinke, D., Taberlet, P., Weigand, A., Abarenkov, K., Beja, P., Bervoets, L., Björnsdóttir, S., Boets, P., Boggero, A., … Zimmermann, J., 2016. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Research Ideas and Outcomes*, *2*, e11321.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J. P., Hector, A., Hooper, D. U., Huston, M. A., Raffaelli, D., Schmid, B., Tilman, D., and Wardle, D. A., 2001. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science 294* (5543), 804-808.

Maasri, A., Jähnig, S., Adamescu, M., Adrian, R., Baigun, C., Baird, D., ... and Worischka, S., 2021. A Global Agenda for Advancing Freshwater Biodiversity Research. *Authorea* Preprints.

Mahé, F., Rognes, T., Quince, C., de Vargas, C. and Dunthorn, M., 2014. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ*, *2*, p.e593.

Mallet, J., 1995. A species definition for the modern synthesis. *Trends in Ecology & Evolution*, *10*(7), pp. 294-299.

Marshall, G.A.K., 1916. The Fauna of British India, Including Ceylon and Burma. Vol. Rhynchophora, Curculionidae. Taylor and Francis, London. Pp. 367.

Mustafa, T., Alam, S.M.I., Jamal, M.S., Johora, F.T., Likhon, M.A.K., Hoq, T., Naser, M.N., 2013. Abundance of benthic fauna in winter and summer seasons at three water bodies of Dhaka, Bangladesh. *Bangladesh J. Zool. 41*(1), pp.79-86.

McGee, K. M., Robinson, C. V, and Hajibabaei, M., 2019. Gaps in DNA-Based Biomonitoring Across the Globe. *Frontiers in Ecology and Evolution*, 7, p.337.

Menezes, S., Baird, D. J., and Soares, A. M. V. M., 2010. Beyond taxonomy: A review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. *Journal of Applied Ecology*, *47*(4), pp.711–719.

Miller, M., A., Pfeiffer, W., Schwartz, T. (2010). "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" *In* Proceedings of the Gateway Computing Environments Workshop New Orleans.

Múrria, C., Bonada, N., Vellend, M., Zamora-Muñoz, C., Alba-Tercedor, J., Sainz-Cantero, C.E., Garrido, J., Acosta, R., El Alami, M., Barquín, J. and Derka, T., 2017. Local environment rather than past climate determines community composition of mountain stream macroinvertebrates across Europe. *Molecular Ecology*, *26*(21), pp.6085-6099.

Morse, J. C., Bae, Y. J., Munkhjargal, G., Sangpradub, N., Tanida, K., Vshivkova, T. S., ... and Yule, C. M., 2007. Freshwater biomonitoring with macroinvertebrates in East Asia. *Frontiers in Ecology and the Environment*, *5*(1), pp.33-42.

Nasiruddin, M., Azadi, M.A., and Reza, M.S., 2014. Abundance and diversity of aquatic insects in two water bodies of Chittagong university campus. *Bangladesh J. Zool.*, *42* (1), pp.19-33.

Nijboer, R.C., Johnson, R.K., Verdonschot, P.F.M., Sommerhäuser, M. and Buffagni. A. 2004. Establishing reference conditions for European streams. *Hydrobiologia*, *516*.

Ofenböck, T., Moog, O., Sharma, S., and Korte, T., 2010. Development of the HKHbios: A new biotic score to assess the river quality in the Hindu Kush-Himalaya. *Hydrobiologia*, *651*(1), pp.39–58.

Primack, R. B., Miller-Rushing, A. J., Corlett, R. T., Devictor, V., Johns, D. M., Loyola, R., Maas, B., Pakeman, R. J., and Pejchar, L., 2018. Biodiversity gains? The debate on changes in local- vs global-scale species richness. *Biological Conservation, 219*.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., ... & Vogler, A. P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55(4), pp. 595-609.

Preston, H.B., 1915. The Fauna of British India, Including Ceylon and Burma. Vol. Mollusca: Freshwater Gastropoda and Pelecypoda. Taylor and Francis, London, pp.244.

R Development Core Team. 2018. R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sana, B.B., Ali, M.S., 2011. A preliminary list of aquatic Coleoptera (Arthropoda: Insecta) in Natore and Rajshahi districts of Bangladesh. *J. Taxon. Biodiv. Res*., 5, pp.1-4.

Sánchez-Montoya, M.M., Vidal-Abarca, M.R., Puntí, T., Poquet, J.M., Prat, N., Rieradevall, M., Alba-Tercedor, J., Zamora-Muñoz, C., Toro, M., Robles, S. and Álvarez, M., 2009. Defining criteria to select reference sites in Mediterranean streams. *Hydrobiologia, 619*(1), pp.39-54.

Santos, J. M., and Ferreira, M. T. (2020). Use of aquatic biota to detect ecological changes in freshwater: Current status and future directions. *Water*, 12(6).

Serrana, J. M., Miyake, Y., Gamboa, M., and Watanabe, K., 2019. Comparison of DNA metabarcoding and morphological identification for stream macroinvertebrate biodiversity assessment and monitoring. *Ecological Indicators*, *101*, 963–972.

Seymour, M., Edwards, F. K., Cosby, B. J., Kelly, M. G., de Bruyn, M., Carvalho, G. R., and Creer, S. (2020). Executing multi-taxa eDNA ecological assessment via traditional metrics and interactive networks. *Science of The Total Environment*, *729*, 138801.

Seymour, M., Edwards, F. K., Cosby, B. J., Bista, I., Scarlett, P. M., Brailsford, F. L., Glanville, H. C., de Bruyn, M., Carvalho, G. R., and Creer, S., 2021. Environmental DNA provides higher resolution assessment of riverine biodiversity and ecosystem function via Spatio-temporal nestedness and turnover partitioning. *Communications Biology*, *4*(1), 512.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics, 22*:2688–2690.

Troast, D., Suhling, F., Jinguji, H., Sahlén, G., Ware, J., 2016. A global population genetic study of Pantala flavescens. *PloS one, 11*(3), p.e0148949.

United States, 1972. Federal Water Pollution Act Amendments of 1972. Pub. L. 90-500,October 18.

Van Klink, R., Bowler, D. E., Gongalsky, K. B., Swengel, A. B., Gentile, A., and Chase, J. M., 2020. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*, *368*(6489), pp.417-420.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., and Davies, P. M., 2010. Global threats to human water security and river biodiversity. *Nature*, *467*(7315), pp.555–561.

Wright, J. F., Sutcliffe, D. W., and Furse, M. T., 2000. *Assessing the biological quality of fresh waters. RIVPACS and other techniques.* Freshwater Biological Association, Ambleside, England.

Xie, Y., Hong, S., Kim, S., Zhang, X., Yang, J., Giesy, J. P., Wang, T., Lu, Y., Yu, H., and Khim, J. S., 2017. Ecogenomic responses of benthic communities under multiple stressors along the marine and adjacent riverine areas of northern Bohai Sea, China. *Chemosphere*, *172*, pp.166–174.

Zhang J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics, 30* (5), pp.614–620.

Chapter 5

General Discussion

Chapter 5: General Discussion

Biomonitoring has been a central issue in the assessment of environmental or man-made impacts on natural systems to protect and restore both terrestrial and aquatic ecosystems. Biomonitoring underpins much of freshwater resource management and has received significant research efforts in the developed world. It is mandatory in many European and North American countries, but conventional methods rely on the expertise of a declining number of taxonomists and many taxa lack the taxonomic resolution for identifications at the species level. Molecular methods are now widely expected to replace or at least complement these existing approaches. A key aspect of the existing biomonitoring system is the background knowledge on the susceptibility of many taxa (at species or higher levels), which permits the use of presence/absence data to classify each water body. This is complicated by the fact that different countries apply classification schemes that have to be calibrated against each other. In Bangladesh, no such status assessment of water bodies exists to date, but it will be required urgently, given the dependence on the numerous rivers in this country. To some extent, the absence of an existing system of biomonitoring simplifies the introduction of molecular-based methods for this purpose, as there is no need to calibrate this methodology against morphology-based systems. However, the introduction of metabarcoding techniques can only be successful if it is easily standardised and widely applicable, at a low cost. In these contexts, this study is a contribution to establishing biomonitoring of freshwater ecosystems in the understudied regions of the world using high throughput sequencing methods.

In this thesis, I expended a great effort on building a DNA barcode reference library for Bangladeshi macroinvertebrates and generating the complete mitogenomes of a good number of species of aquatic insect orders. Further, issues related to bulk sample preparation in metabarcoding pipeline have been addressed for their standardization of metabarcoding of freshwater macroinvertebrates which can be easily implemented in the field and requires only limited resources, time and expertise in molecular biology and bioinformatics. Finally, I tried to estimate the species diversity and turnover of macroinvertebrates in highland and lowland river systems and established their responses to anthropogenic stressors and environmental degradation using the metabarcoding technique. This approach will have great implications for the conservation of freshwater ecosystems as a step towards building a biomonitoring system in this region.

5.1 DNA barcodes of freshwater macroinvertebrates

Being in a region where essentially no taxonomic and ecological knowledge about these freshwater organisms exists, a key step is the generation of basic taxonomic and distributional information. Morphology-based identification of invertebrate organisms to lower taxonomic ranks (e.g., family, genus or species) is a great challenge for this step in understudied aquatic ecosystems. The results can be variable in the same waterbody depending on the taxonomists' expertise, experience and opinion, which can potentially lead to contrasting bioassessments (Carstensen and Lindegarth, 2016; Clarke, 2013). In this context, DNA barcoding offers a less biased approach than morphology-based identification of aquatic macroinvertebrates (Leese et al., 2018) using short, standardized gene markers. However, the effective utility of those barcodes largely relies on a well-curated reference database of target organism groups with metadata. Currently, the largest DNA barcode reference library, the Barcode of Life Data System (BOLD (Ratnasingham and Hebert, 2007) along with GenBank (Benson et al., 2013), is playing a pivotal role in biodiversity assessment and monitoring providing barcode sequences of multiple gene markers of organisms with their autecological and biogeographic information. In essence, a local or regional barcode database of particular bioindicator organisms can be more effective in the bioassessment or biomonitoring of particular freshwater ecosystems.

The exploration of DNA barcodes for Bangladeshi fauna remains in a rudimentary stage which is quite evident in the publicly accessible databases. As of December 2022, ~ 44 COI barcode sequences of coleopterans (only terrestrial beetles), ~450 sequences of dipterans (mostly fruit flies), ~176 sequences of molluscs (mostly from several marine species) and only 14 sequences of odonates have been submitted in the GenBank database. Strikingly, there was no sequence for other arthropods (mayflies, caddisflies, and stoneflies) and annelids (oligochaetes and polychaetes). In contrast, a barcode GAP analysis for European macroinvertebrates showed that there were comparatively few species of some insect orders (e.g., Hemiptera, Odonata and Trichoptera) missing sequences in the BOLD and GenBank databases (Weigand et al., 2019). Basically, the numbers of barcode sequences of macroinvertebrate groups from Bangladesh are remarkably low except for a comprehensive DNA barcode library for 243 freshwater fish species (Rahman et al., 2019) and a partial DNA barcode database for marine fishes (Ahmed et al., 2021; Habib et al., 2021). Overall, this estimation clearly indicates a large gap in the reference database for the country’s aquatic

macroinvertebrates though this is an important prerequisite for biodiversity and ecological assessment using genetic, metagenetic and metagenomic techniques (Weigand et al., 2019).

I started with the generation of ~ 812 COI barcodes that represents more than 300 species (delimited by three methods) of macroinvertebrates in the highland and lowland rivers of Bangladesh which had not been studied before. In essence, this number of species hints at the highly diverse macroinvertebrates of Bangladesh though the present study included 16 upland streams in a hilly district of the south-east region and 4 lowland rivers in the central part of the country. Furthermore, species accumulation curves made for major macroinvertebrate groups also showed that the diversity of these groups is yet to be discovered from the studied area (Fig. 2.9 in the appendix). However, in the comparison of explored barcodes to sequences (from other regions of the world) in the existing molecular and taxonomic databases (BOLD, GenBank and GBIF), many species encountered here revealed their local endemism, mostly were new records in the country or possibly even undescribed new species.

For species delimitation, three approaches (e.g., distance-based Usearch clustering at a 3 % threshold, phylogenetic tree-based bPTP and RESL on the BOLD platform) produced incongruent results for some taxa. This was expected as there is no stand-alone comprehensive method for species delimitation using genomic data. It is noteworthy that presumed morphospecies of all selected specimens was taken for the exploration of COI barcodes in this study and the family-level identification of each specimen was confirmed primarily based on their morphology. Notably, during the taxonomical assignment of explored barcode sequences against the BOLD, GenBank and GBIF databases, most of the barcodes could not be assigned to their species name, therefore, I retained their voucher specimens with locality data. These could be useful for further taxonomic assignment to lower levels (species or genus) by Linnean classification in future.

Nowadays, understanding the genetic diversity of organisms has attracted great attention in biodiversity and ecological studies. Genetic distances among and between species may be crucial for the fitness of a population and provide a way for populations to adapt to environmental changes (Xu et al., 2021). The inter and intraspecific genetic distance analyses with barcodes of major groups of macroinvertebrates of this study provided baseline information on the genetic variation of the country's aquatic insects and molluscan species.

This result will also have great value in population studies and also could contribute to prioritizing conservation measures of potential bioindicator macroinvertebrates.

5.2 Mitogenome-based phylogeny of freshwater macroinvertebrates

Both mitochondrial and nuclear genomes or genes are useful for understanding the morphological, physiological, behavioural, ecological and evolutionary processes affecting any organism. Generally, nuclear genes might be assumed as a powerful source of phylogenetic information for understanding more ancient levels of divergences as multiple substitutions at variable sites of nuclear genes can result in the elimination of phylogenetic signals (Caravas, 2012). Conversely, the rapid and higher degree of sequence variation in mitochondrial genes makes it possible to resolve lower taxonomic levels for organisms (Hwang and Kim,1999; Chan et al.,2021). Although multiple types of genetic markers are suitable for molecular systematics and identification purposes, the varying properties of the genetic markers complicate the choice for their respective applications (Chan et al.,2021).

Mitogenomes have contributed to resolving the taxonomy and understanding the adaptation and evolutionary mechanisms in vertebrates (Parhi et al., 2019; Li et al., 2013) and invertebrates (Jacobsen et al., 2012; Stokkan et al., 2018). For the phylogenetic and evolutionary analysis of insects, mitogenomes have been powerful markers (e.g., Condamine et al., 2018; Crampton-Platt et al., 2015; Yan et al., 2019) due to their small genome size (consisting of a set of 37 genes of which 13 protein-coding, two ribosomal RNAs (rRNAs), and 22 transfer RNAs genes), maternal inheritance, low sequence recombination and fast evolutionary rates (Curole and Kocher, 1999; Lin et al., 2022). Within the mitochondrial genome, generally used markers are the protein-coding genes and the 12S and 16S ribosomal RNA (rRNA) genes which have different evolutionary rates and functions that make some of the genes quite conserved and the others more variable. These marker genes lack introns (common in single-copy nuclear genes), contains small intergenic regions (or are absent), and rare heteroplasmy (coexistence of different mtDNA within a cell or individual) (Bruvo-Mađarić, 2009). In addition, mitochondria lack proofreading power (i.e., error-repair) and mtDNA does not code for proteins directly involved with its own replication, transcription or translation. These attributes also lead to a larger number of length mutations and transitions than single-copy nuclear DNA. Mitochondrial genomes or genes are generally easier to amplify (by widely available mitochondrial primers) or to sequence (with the advent of high-

throughput sequencing technology) than nuclear genes. All these features make mtDNA an ideal candidate for phylogenetic investigations on different taxonomic levels (Alberts et al, 2002).

If considering the disadvantageous features of mitochondrial DNA for phylogenetic studies, the higher rate of substitution of mitochondrial genes can be problematic to resolve divergences of more than 5–10 million years. In addition, mitochondrial genes have attributes that tend to lead to high levels of homoplasy when analyzed by standard phylogenetic methods, such as an extreme A/T bias in third positions (Mooers and Holmes, 2000; Lin and Danforth, 2004).

Conversely, the nuclear genome or genes, particularly the nuclear rRNA genes, is more conserved than mtDNA which makes them a potentially helpful source of genetic markers for resolving higher taxonomic levels for organisms (Hwang and Kim,1999; Chan et al.,2021). Within nDNA, the internal transcribed spacer (ITS) regions possess a higher degree of sequence variation than the nuclear rRNA genes because of a faster nucleotide substitution rate (Hwang and Kim, 1999). However, from a practical point of view, mitochondrial genomes the have gained much attraction in the evolutionary studies of animals.

In this study, I have made great efforts to build mitochondrial genomes (~108) for a large selection of the local morphospecies from each of the major macroinvertebrate taxa. The comparative study of these mitogenomes with regions elsewhere in the world, need to be performed in the context of evolutionary analyses. Phylogenetic trees generated with these mitogenomes supported the sound placement of local fauna in the global framework of major lineages, which is of great interest for the analysis of the local study, but equally these taxa will also contribute to an understanding of the phylogeny of these arthropod orders which currently lack representatives of this region. The phylogenetic trees constructed with concatenated sequences of mitochondrial protein-coding genes (local and global) using the maximum likelihood approach have revealed the intra-order evolutionary placement of different families with their monophyletic, paraphyletic and polyphyletic status. For instance, this study has confirmed the monophyly of two suborders (Anisoptera: Dragonfly and Zygoptera: Damselflies) where the families Euphaeidae and Calopterigidae formed two sister clades with the family Cenagrionidae of damselflies. The suborder (Anisoptera) of

dragonflies also recovered the monophyly of the families Gomphidae, Libellulidae and Macromiidae. This phylogenetic placement within the order Odonata was also corroborated by other studies (Bybee et al., 2016; Carle et al., 2016). Most of the true fly families were monophyletic which was supported by Cranston, et al. (2012) while the non-biting midge (Chironomidae) and the biting midge (Ceratopogonidae) were found as paraphyletic groups. A phylogenetic study of caddisflies (Trichoptera) combining mitochondrial COI and nuclear 18S rRNA, 28S rRNA revealed the monophyly of Hydropsychidae, Polycentropodidae, Psychomyiidae, Stenpsychidae Hydroptilidae Philopotamidae Glossosomatidae and Leptoceridae (Thomas et al., 2020) which was also evident in the present study. In the case of Ephemeroptera, non-monophyly of Baetidae, Ephemeridae, and Heptageniidae was recovered by a combined multi-gene (18S rDNA, 28S rDNA, 16S rDNA, 12S rDNA) based phylogeny (Ogden, and Whiting, 2005). Conversely, those families were found as monophyletic in this study where phylogenetic construction was made using 13 mitochondrial protein-coding genes.

Despite the power of the mitochondrial genome to infer phylogeny, the limitation of mitogenome data is a major hurdle for multi-genes or mitogenome based phylogenetic studies which makes it essential to build a mitogenome reference database including the maximum number of taxa worldwide. Already I noticed complete mitogenomes for several families (e.g., Odontoceridae, Philopotamidae, Glossosomatidae) of insect orders entirely absent from the Genbank database, and thus the mitogenome sequences generated in this study will contribute to the increasingly complete public records. Second, the mitogenomes trees were also effective to place the barcodes and OTUs confidently into a phylogenetic framework, which was not possible with the short barcode sequences alone. With this tree in hand, it can then be established what kind of lineages are specific to highland and lowland streams, and to what extent biogeographic differences shape the respective freshwater fauna. In addition, this information can be exploited for better understanding of species sensitivity. For example, indicator species revealed by the correlation with physical habitat parameters can be placed in the phylogenetic tree to predict the sensitivity of other species not explicitly tested for their responses to habitat alteration or pollution, as a way of expanding the ecological status assessment to all freshwater species even for those with limited available observations.

Therefore, the mitogenome data explored in this study will provide important information to both ecologists and evolutionary biologists by providing the required biological perspectives of different insects' orders in a phylogenetic context. Future efforts should be made for

exploring more mitogenomic data for other species of insects to resolve the phylogeny that will further provide greater insight into their evolutionary biology.

5.3 Standardization of some methodological aspects in metabarcoding pipeline

Metabarcoding of invertebrate samples has some technical challenges starting with bulk sample preparation for DNA extractions (Creedy et al., 2019; Zinger et al., 2019; Pawlowski et al., 2018). A major consideration is the way the samples are gathered in the field. Standard methods of kick sampling produce large volumes of material, including sediments and plant material from surrounding trees, which have to be removed from the actual specimens. Floating of invertebrates is an established procedure, but usually involves the use of a colloidal solution such as Ludox to adjust the specific gravity for improved separation of invertebrate specimens from other material. I spent some effort to optimise the conditions of the Ludox extractions, which is useful information for any type of specimen extraction from ecological samples obtained in aquatic habitats. Surprisingly, I found that even these optimal conditions do not perform better than raw samples or floating in water only (sample extraction method of Chapter 3). This is important in the context of sampling that needs to be simple and low-cost to be adopted in countries where biomonitoring of invertebrates is not currently prescribed. While simple, this method retains a great advantage over DNA extractions directly from the sediment, as has been proposed (Nichols, et al., 2019), because a much greater sample volume can be processed from which the specimens are extracted. In addition, I split the very large (>15 mm) specimens for separate DNA extraction, which reduces the problem of different biomass and thus different detection limits of small-bodied and rare species. After these large specimens are removed, metabarcoding is expected to be fairly robust to differences in biomass (Creedy et al., 2019).

In a metabarcoding pipeline, the homogenization of bulk macroinvertebrate samples is another important step for exploring maximum diversity estimates using a reasonable amount of homogenate from a sample (Aylagas et al. 2016; Majaneva et al., 2018). As a standard kick-net sample contains a lot of debris of different kinds, the sample volume remains high even after a proper homogenization, which prohibits the DNA extraction from the whole sample. Therefore, the optimization of the amount/number of replicate samples is also crucial

in metabarcoding of macroinvertebrates. Here I did an experiment taking single, duplicate and triplicate samples that showed convincingly that there were no statistically significant differences in OTUs richness and composition among subsamples from a single homgenate. This result might rely on proper homogenization, replication of PCRs and deep sequencing amplified sample which I maintained in this study. Therefore, this finding has implications for large-scale biomonitoring programmes providing cost-effective, time-saving standardized procedures in the metabarcoding pipeline.

Primer and barcode choice has been a recurring theme in the metabarcoding literature of recent years, without the emergence of a clear favourite (Creedy et al., 2022; Andújar et al., 2018). The barcode region of the COI gene has been widely used for species identification and biodiversity assessment. The primers for COI used here had a very high success rate across different major taxa and only performed slightly less well on Mollusca, which are notoriously problematic for mitochondrial work. As universal primers are not efficient for the amplification of COI barcode of all eukaryotes (Geller et al., 2013), taxon-specific primers were also developed, including universal primers for invertebrate organisms (Folmer et al., 1994; Leray et al., 2013). Even though amplification efficiencies of those universal primers vary within the group of invertebrates (e.g., variable results between classes of arthropods, annelids and molluscs) where multiple primers are suggested for amplifying macroinvertebrates fauna (Pfrender et al., 2010). Alternatively, some studies advocated the use of nuclear 18S rRNA gene (for nematodes), a combination of mitochondrial COI and Cyt b genes or a combination of mitochondrial COI and nuclear 18S rRNA genes (for invertebrate communities) (Cowart et al., 2015). However the advantage of the COI primer also includes the fragment length which is at the limit of what can be achieved with the Illumina technology, but at the same time provides greater amounts of sequence information than virtually all alternatives (Elbrecht et al., 2018). The methodology used here is now increasingly mature, showing breadth in the taxa being targeted across three major phyla that make up the majority of freshwater organisms. In addition, the use of a protein- coding region has great advantages for detecting read errors and pseudogenes in metabarcoding, e.g. using the metaMATE software for establishing read number thresholds based on sequencing errors that can be detected predictably in protein-coding regions (Andujar et al., 2021). Therefore, for the identification of invertebrate communities, COI has been widely applied in biomonitoring freshwater ecosystems (Elbrecht and Leese, 2017).

The bioinformatics pipeline used well-established procedures of clustering of reads to recognise OTUs, which are taken as species equivalent. More elaborate methods can be applied by removing the clustering step, i.e. working at the ASV level (Edgar, 2018), which has already been applied in various studies for a true haplotype level analysis of entire mixed communities (Elbrecht et al., 2018; Arribas et al., 2021). This can be useful to study turnover at the genetic level in future with the same data and provides an exciting prospect in particular for the study of the largely unexplored landscapes of the Bangladesh hill regions, which are characterised by great landscape heterogeneity presumably promoting genetic differentiation. Using ASV data will resolve questions about connectivity within and between river basin at various spatial scales, and ultimately will provide information for conservation management that requires detailed knowledge about the extent of species ranges and within-species phylogeographic structure. The prospect of having this information available for entire species assemblages, rather than individual species of concern in conventional conservation genetics, will allow conservation and biological status assessment at the scale of entire ecosystems, possibly driven by the indicator species established from this initial analysis presented here. Yet, the straightforward clustering approach provided a good approximation of the species present. Detailed analyses of the correspondence to morphologically defined species are still outstanding, but I already provided a reference barcode library for ~300 morphospecies to which the metabarcode OTUs can be assigned. At this stage, only a few examples can be given to demonstrate the likely success of this assignment. For example, I was able to distinguish a total of 36 species of EPTs from highland streams, while the metabarcoding revealed a total of 66 OTUs. Given that the greater sensitivity of the DNA methods that also generate data from early developmental stages and the potential lumping of morphospecies, this roughly doubled number of OTUs was expected and constitutes a meaningful measure of the total species present in these samples. The use of haplotype data was already trialled here, by applying methods for molecular species delimitation, which correspond closely to the clustering.

The DNA methods, especially metabarcoding of bulk sample DNA, used here were highly reliable and applicable universally. However, this method has some limitations including the invasive sampling of specimens and the destruction of whole samples which also have been a concern in biodiversity and ecological assessment in aquatic systems. To overcome these limitations, DNA can be extracted from various environmental samples (e.g., water and

sediments) and also from preservatives where specimens are preserved for a long period of time for various purposes. This environmental DNA or eDNA (nuclear or mitochondrial) is released from an organism into the environment with secreted faeces, mucous, gametes, shed skin and carcasses. Recently eDNA metabarcoding has gained special attention from ecologists, researchers and policymakers in biodiversity and ecological assessment programmes in marine and freshwater ecosystems. To date, eDNA-based assessment has been reported for a range of aquatic species, including fish (e.g., Shu et al., 2020; Valentini et al., 2016; Balasingham et al., 2016; Olds et al., 2016; Turner et al., 2015; Janosik et al., 2014; Minamoto et al., 2012), amphibians (Valentini et al., 2016; Fukumoto et al., 2015; Ficetola et al., 2008), mammals (Thomsen et al., 2012), and invertebrates (Deiner et al., 2015; Mächler et al., 2014; Lim et al., 2016; Fernandez et al., 2018). These studies applied eDNA metabarcoding basically for the detection of targeted or invasive species of vertebrates and invertebrates and some dealt with overall biodiversity assessment in a specific water body showing equal or higher performances than the conventional approach for species detection. Compared to vertebrate organisms (e.g., fish and amphibians), the eDNA application in the assessment of invertebrates is relatively low in freshwater ecosystems which might be linked to less availability of DNA from smaller invertebrates and the higher degradation rate of DNA. For instance, Fernandez et al. (2018) tested the reliability of eDNA metabarcoding to record river macroinvertebrates and found it more sensitive than the conventional suggesting it for an alternative assessment of freshwater quality. The eDNA approach was recommended for the detection of specific indicator macroinvertebrates (Mächler et al., 2014) and also for overall metazoan diversity in freshwater systems (Lim et al., 2016).

However, the effectiveness of eDNA-metabarcoding has not been extensively unveiled for the assessment of freshwater macroinvertebrates. Several studies have looked at DNA metabarcoding and bulk sample (tissue) metabarcoding to assess the efficacy of each method for characterizing aquatic macroinvertebrate communities (Macher et al., 2018; Hajibabaei et al., 2019; Gleason et al., 2021). It is evident from these studies that eDNA metabarcoding is a poor replacement for bulk-sample metabarcoding of benthic macroinvertebrates, as eDNA is more prone to the water down of macroinvertebrates due to much co-amplification of nontarget taxa (e.g., fungi, algae, and bacteria) (Leese et al., 2021).

In addition to using bulk sample DNA and eDNA (from water and sediments) for metabarcoding, some attempts (Hajibabaei et al., 2012; Martins et al., 2019) were made to extract DNA from preservative ethanol (sample preservation medium) showing variable

results with a single case of better performance for preservative ethanol than water samples in detecting invertebrate diversity to a local scale (Wang et al., 2021) but it is not fully explored yet for the large-scale assessment of freshwater macroinvertebrates. Considering all these aspects, this study supports the use of DNA from bulk benthic samples avoiding an extensive sorting and separation of all individuals from sediments, which could be a very good source for DNA extraction in the metabarcoding-based biomonitoring programme.

5.4 Macroinvertebrate diversity and their responses to environmental degradation in highland and lowland rivers

This study was the first step in the country towards the use of high throughput metabarcoding technique to assess the macroinvertebrate diversity and their responses to man-made pressure in freshwater ecosystems. I selected representative regions for the two major biotas in Bangladesh, the lowland and highland rivers. These systems couldn't be more different in terms of climate and geographical features, while physical attributes and human impact are also different to some extent. The species equivalent OTU database (936 and 662 OTUs from highland and lowland rivers respectively) generated here now constitutes the first set of (meta) barcodes of highly diverse freshwater macroinvertebrate fauna for both regions of which one (highland streams) belongs to Indo-Burmese biodiversity hotspot. This OTU-based species database of insects, molluscs and annelids can be augmented with incoming samples in an iterative fashion that ultimately result in an increasingly complete and ecologically informed database that can be linked to status assessments at some point. Already now I can see clear differences that point to the great utility of these data. For example, EPTs were essentially absent from the more polluted lowland rivers but were present in all of the highland streams. Yet there were recognisable differences even within the highlands, e.g. the absence of stoneflies (Plecoptera) from all but four rivers, which also happen to be those with the least amount of disturbance. Other notable differences were the high abundance of Hemiptera in the lowlands, whereas Ephemeroptera were highly abundant in the highland rivers. In terms of species richness, Diptera and Hemiptera were the highest species-rich taxa in highland and lowland rivers respectively. Furthermore, as the species distribution and abundance depend on the ecological quality of the streams, the distribution of reads obtained from different taxa varied greatly between sampling locations.

These systems differ in numerous ways but exemplify the two main types of freshwater habitats in Bangladesh, which are of great importance and deserve specific considerations that explain the diversity and threats, while the methodology of DNA-based monitoring is universal. Lowland rivers are slow-moving, very large water bodies, that in many ways have properties of lentic (standing-water) systems, while highland streams are typical lotic (running- water) systems. These types have been recognised as a fundamental distinction between freshwater ecosystems, differing in the population dynamic. Whereas lotic populations are long-lived persisting in a stable habitat (even if rivers change their course the populations can easily track this change), lentic populations are short-lived due to the ephemeral nature of most lakes and ponds, requiring dispersal of populations for long term persistence (Ribera, 2008). As a consequence, lotic populations are genetically more isolated, species have smaller geographic ranges, and speciation and extinction rates are higher than in lentic systems. The highland communities reflect these features perfectly, as we see great turnover even among the 16 closely adjacent rivers used for this analysis (overlain by the environmental impact). This was partly evident from the lack of close matches in Genbank data, which only included several widespread, invasive species. In the lowlands, the proportion of 'known' BINs (matched presumed species within 3% threshold) was almost double of upland group. As neither of these regions have been sampled before, this difference probably reflects a true difference in the range sizes of lowland species. The lotic-lentic framework therefore can be a useful model for the two major freshwater ecoregions of Bangladesh, even if strictly speaking the lowland systems are not lentic either. In addition, we need to recognise the different volume of the water bodies, as large rivers and lakes usually differ in the composition of major groups, especially certain groups of insects that are mostly lost with an increase in the size of water bodies, while the amount of vegetation also differs, e.g., in allowing herbivorous aphids, including the top-abundant in the lowland rivers.

Finally, the differences between the two habitat types differ in the way I surveyed the various anthropogenic impacts. Lowland rivers were assessed with standard chemical methods, showing the correlation of 'good' quality mainly with high DO versus high nitrogen, phosphate and salinity etc. in poor-quality systems. In contrast, the highland systems could be assessed based on externally visible physical parameters, without the need for chemical analysis. In addition, these analyses resulted in different indicator species, none of which were shared between the two regions. In both regions, the number of disturbance indicators

was greater than indicators of healthy ecological conditions which suggests that a number of species are getting benefits from human-induced disturbances. One of the limitations of this study was the higher-level taxonomic assignment of OTUs as most of the metabarcodes could not be assigned to species rank of Linnean classification due to the lack of a complete reference database of barcode sequences though an attempt was made to build a reference database of ~300 species from insects, annelids and molluscs (Chapter 2). The complete mitogenome-based reference database would be more useful for using different marker genes (besides COI) for assessing invertebrate communities (e.g., Aylagas et al., 2016; Elbrecht and Leese, 2017). However, incomplete reference sets of organisms are one of the major challenges of metabarcoding for its successful application in the biomonitoring of invertebrate communities. Therefore, further efforts are to be made by extensive sampling with taxonomic resolution in near future.

As a general principle it was possible to recognise high turnover in the mountain regions, in particular in the undisturbed areas, and the gradual increase in widespread species in disturbed habitats, as evident from the greater proportion of species associated with 'poor' quality conditions with broad distributions, which ultimately are also homogenised across wider regions, as seen by the larger proportion of species with known representation in public databases as a proxy for cross-region distributions. More detailed analyses are required to confirm these hypotheses as the reference sets become more complete, but these preliminary findings show the power of local studies in the context of the growing barcode databases. Likewise, the different indicator sets can be compared across sites and biogeographic regions, which ultimately will reveal the similarities and differences. For example, while EPTs are widely recognised as indicators of good quality, they are an inhomogeneous group in terms of their response to different quality parameters, and this becomes even more of a problem when an ever-greater diversity of species from different biogeographic regions are considered. An approach that considers only particular subgroups within these three orders, guided by the phylogenetic tree, is likely to be more powerful in detecting smaller changes in water quality, before "EPT" as a whole show a decline. This is even more true for other groups that respond much less clearly, as seen for example in the complete community of the highland streams whose total species richness was not affected by apparent environmental degradation (unlike the EPT). Only the greater phylogenetic resolution of sub-lineages will be able to differentiate the mixed responses of the various components of complex communities studied

with the metabarcoding approach. As more regions are studied, in particular outside of the well-studied Palaearctic region, and as the phylogenetic trees become more complete and better supported with the inclusion of more taxa and more genes, our use of the metabarcoding approach will become ever more powerful globally. Further exploration of the diversity in natural and disturbed systems in the Indo-Malayan and other biogeographic region will be needed to establish the broader biotic differences and how they link to the responses of species and higher taxa.

The present study shows the potential of metabarcoding for assessing stream invertebrate diversity and environmental degradation based on obtained species-equivalent DNA sequences without prior sample sorting procedures. The approach standardized here has formed a foundation for applying a molecular method in Bangladesh that could be applied to evaluate an unknown invertebrate sample based on its placement within the obtained OTU clusters in this study. The result highlights the further acquisition of sequencing data from macroinvertebrate samples from additional streams across the country to begin the development of robust models for the assessment of stream health.

In conclusion, this thesis is a first step towards a DNA-based monitoring system of the rivers of Bangladesh, which is urgently required and can be expected to be taken up by the national authorities in charge of aquatic resources. At the same time, the molecular approach holds exciting prospects for the study of biodiversity in this poorly known region. In particular, the hill region of Bangladesh is topographically extremely complex and benefits from a (sub) tropical climate that combined contributes to great species richness. Knowledge about the local diversity, the spatial scale of species turnover and environmental degradation in river ecosystems in addition to the placement of local species in a global phylogenetic framework, will be important to the much-needed understanding of Bangladeshi biodiversity and ultimately contribute to its conservation.

5.5 References

Ahmed, M.S., Datta, S.K., Saha, T. and Hossain, Z., 2021. Molecular characterization of marine and coastal fishes of Bangladesh through DNA barcodes. *Ecology and Evolution*, *11*(9), pp.3696-3709.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P., 2002. *Molecular Biology of the Cell*. Fourth Edition. Garland Science.

Andújar, C., Arribas, P., Yu, D.W., Vogler, A.P. and Emerson, B.C., 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, *27*(20), pp. 3968-3975.

Andujar, C., Creedy, T.J., Arribas, P., López, H., Salces-Castellano, A., Pérez-Delgado, A.J., Vogler, A.P. and Emerson, B.C., 2021. Validated removal of nuclear pseudogenes and sequencing artefacts from mitochondrial metabarcode data. *Molecular Ecology Resources*, *21*(6), pp.1772-1787.

Aylagas, E., Mendibil, I., Borja, Á. and Rodríguez-Ezpeleta, N., 2016. Marine sediment sample pre-processing for macroinvertebrates metabarcoding: mechanical enrichment and homogenization. *Frontiers in Marine Science*, *3*, p.203.

Balasingham, K.D., Walter, R.P., Mandrak, N.E. and Heath, D.D., 2018. Environmental DNA detection of rare and invasive fish species in two Great Lakes tributaries. *Molecular Ecology*, *27*(1), pp.112-127.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., et al., Bruvo-Mađarić, B., 2009. Molecular phylogenetic methods in entomology-new insights on the evolutionary relationships of hexapods and arthropods. *Entomologia Croatica*, *13*(2), pp.69-84.

Bybee, S., Córdoba-Aguilar, A., Duryea, M.C., Futahashi, R., Hansson, B., Lorenzo-Carballa, M.O., Schilder, R., Stoks, R., Suvorov, A., Svensson, E.I. and Swaegers, J., 2016. Odonata (dragonflies and damselflies) as a bridge between ecology and evolutionary genomics. *Frontiers in Zoology*, *13*(1), pp.1-20.

Caravas, J., 2012. Phylogenetic utility of mitochondrial and nuclear genes: a case study in the diptera (true flies). Wayne State University Dissertations. Paper 427.

Carle, F.L., Kjer, K.M. and May, M.L., 2015. A molecular phylogeny and classification of Anisoptera (Odonata). *Arthropod Systematics & Phylogeny*, *73*(2), pp.281-301.

Carstensen, J. and Lindegarth, M., 2016. Confidence in ecological indicators: a framework for quantifying uncertainty components from monitoring data. *Ecological Indicators*, *67*, pp.306-317.

Chan, A.H.E., Chaisiri, K., Saralamba, S., Morand, S. and Thaenkham, U., 2021. Assessing the suitability of mitochondrial and nuclear DNA genetic markers for molecular systematics and species identification of helminths. *Parasites & Vectors*, 14(1), pp.1-13.

Clarke, R.T., 2013. Estimating confidence of European WFD ecological status class and WISER Bioassessment Uncertainty Guidance Software (WISERBUGS). *Hydrobiologia*, *704* (1), pp.39-56.

Condamine, F.L., Nabholz, B., CLAMENS, A.L., Dupuis, J.R. and Sperling, F.A., 2018. Mitochondrial phylogenomics, the origin of swallowtail butterflies, and the impact of the number of clocks in B ayesian molecular dating. *Systematic Entomology*, *43*(3), pp.460-480.

Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J. and Arnaud-Haond, S., 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PloS one, *10*(2), p.e0117562.

Crampton-Platt, A., Timmermans, M.J., Gimmel, M.L., Kutty, S.N., Cockerill, T.D., Vun Khen, C. and Vogler, A.P., 2015. Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. Molecular Biology and Evolution, 32(9), pp.2302-2316.

Cranston, P.S., Hardy, N.B. and Morse, G.E., 2012. A dated molecular phylogeny for the Chironomidae (Diptera). *Systematic Entomology*, *37*(1), pp.172-188.

Creedy, T.J., Andujar, C., Meramveliotakis, E., Noguerales, V., Overcast, I., Papadopoulou, A., Morlon, H., Vogler, A.P., Emerson, B.C. and Arribas, P., 2022. Coming of age for COI metabarcoding of whole organism community DNA: towards bioinformatic harmonisation. *Molecular Ecology Resources*, 22(3), pp.847-861.

Creedy, T.J., Ng, W.S. and Vogler, A.P., 2019. Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. *Ecology and Evolution*, *9*(6), pp.3105-3116.

Curole, J.P. and Kocher, T.D., 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends in Ecology & Evolution, 14(10), pp.394-398.

Deiner, K., Walser, J.C., Mächler, E. and Altermatt, F., 2015. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, *183*, pp.53-63.

Edgar, R.C., 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics*, *34*(14), pp.2371-2375.

Elbrecht, V., Vamos, E.E., Steinke, D. and Leese, F., 2018. Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ,* 6, p.e4644.

Elbrecht, V. and Leese, F., 2017. Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science*, 5, pp 1–17.

Fernández, S., Rodríguez, S., Martínez, J.L., Borrell, Y.J., Ardura, A. and García-Vázquez, E., 2018. Evaluating freshwater macroinvertebrates from eDNA metabarcoding: A river Nalón case study. *PLoS One*, *13*(8), p.e0201741.

Ficetola, G.F., Miaud, C., Pompanon, F. and Taberlet, P., 2008. Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), pp.423-425.

Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates, *Mol. Mar. Biol. Biotechnol*. 3, 294–299.

Fukumoto, S., Ushimaru, A. and Minamoto, T., 2015. A basin-scale application of environmental DNA assessment for rare endemic species and closely related exotic species in rivers: A case study of giant salamanders in Japan. *Journal of Applied Ecology*, *52*(2), pp.358-365.

Geller, J., Meyer, C., Parker, M. and Hawk, H., 2013. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, *13*(5), pp.851-861.

Gleason, J.E., Elbrecht, V., Braukmann, T.W., Hanner, R.H. and Cottenie, K., 2021. Assessment of stream macroinvertebrate communities with eDNA is not congruent with tissue-based metabarcoding. *Molecular Ecology*, *30*(13), pp.3239-3251.

Habib, K.A., Neogi, A.K., Rahman, M., Oh, J., Lee, Y.H. and Kim, C.G., 2021. DNA barcoding of brackish and marine water fishes and shellfishes of Sundarbans, the world's largest mangrove ecosystem. *Plos One*, *16*(8), p.e0255110.

Hajibabaei, M., Porter, T.M., Robinson, C.V., Baird, D.J., Shokralla, S. and Wright, M.T., 2019. Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. *PloS one*, *14*(12), p.e0225409.

Hajibabaei, M., Spall, J.L., Shokralla, S. and van Konynenburg, S., 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, *12*(1), pp.1-10.

Hwang, U.W. and Kim, W., 1999. General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. *The Korean Journal of Parasitology*, *37*(4), p.215.

Jacobsen, M.W., Hansen, M.M., Orlando, L., Bekkevold, D., Bernatchez, L., Willerslev, E. and Gilbert, M.T.P., 2012. Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (Coregonus spp.). *Molecular Ecology*, *21*(11), pp.2727-2742.

Janosik, A.M. and Johnston, C.E., 2015. Environmental DNA as an effective tool for detection of imperiled fishes. *Environmental Biology of Fishes*, *98*(8), pp.1889-1893.

Leese, F., Bouchez, A., Abarenkov, K., Altermatt, F., Borja, Á., Bruce, K., et al., 2018. Why we need sustainable networks bridging countries, disciplines, cultures and generations for aquatic biomonitoring 2.0: a perspective derived from the DNAqua-Net COST Action. In: Bohan, D.A., Dumbrell, A.J., Woodward, G., Jackson, M. (Eds.), Next Generation Biomonitoring. Part 1. vol. 58. Academic Press, pp. 63–99.

Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P. and Zizka, V.M., 2021. Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification. *Environmental DNA*, *3*(1), pp.261-276.

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T. and Machida, R.J., 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1), pp.1-14.

Li, Y., Ren, Z., Shedlock, A.M., Wu, J., Sang, L., Tersing, T., Hasegawa, M., Yonezawa, T. and Zhong, Y., 2013. High altitude adaptation of the schizothoracine fishes (Cyprinidae) revealed by the mitochondrial genome analyses. *Gene, 517*(2), pp.169-178.

Lim, N.K., Tay, Y.C., Srivathsan, A., Tan, J.W., Kwik, J.T., Baloğlu, B., Meier, R. and Yeo, D.C., 2016. Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. *Royal Society Open Science*, *3*(11), p.160635.

Lin, C.P. and Danforth, B.N., 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Molecular Phylogenetics and Evolution*, *30*(3), pp.686-702.

Lin, X.L., Zhao, Y.M., Yan, L.P., Liu, W.B., Bu, W.J., Wang, X.H. and Zheng, C.G., 2022. Mitogenomes provide new insights into the evolutionary history of Prodiamesinae (Diptera: Chironomidae). *Zoologica Scripta*, *51*(1), pp.119-132.

Macher, J.N., Vivancos, A., Piggott, J.J., Centeno, F.C., Matthaei, C.D. and Leese, F., 2018. Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology Resources*, *18*(6), pp.1456-1468.

Mächler, E., Deiner, K., Steinmann, P. and Altermatt, F., 2014. Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshwater Science*, *33*(4), pp.1174-1183.

Majaneva, M., Diserud, O.H., Eagle, S.H., Hajibabaei, M. and Ekrem, T., 2018. Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples. *Metabarcoding and Metagenomics*, *2*, p.e26664.

Martins, F.M., Galhardo, M., Filipe, A.F., Teixeira, A., Pinheiro, P., Paupério, J., Alves, P.C. and Beja, P., 2019. Have the cake and eat it: Optimizing nondestructive DNA metabarcoding of macroinvertebrate samples for freshwater biomonitoring. *Molecular Ecology Resources*, *19*(4), pp.863-876.

Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M.N. and Kawabata, Z.I., 2012. Surveillance of fish species composition using environmental DNA. *Limnology*, *13*(2), pp.193-197.

Mooers, A.Ø. and Holmes, E.C., 2000. The evolution of base composition and phylogenetic inference. *Trends in Ecology & Evolution*, *15*(9), pp.365-369.

Nichols, R.V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M., Green, R.E. and Shapiro, B., 2018. Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*, *18*(5), pp.927-939.

Olds, B.P., Jerde, C.L., Renshaw, M.A., Li, Y., Evans, N.T., Turner, C.R., Deiner, K., Mahon, A.R., Brueseke, M.A., Shirey, P.D. and Pfrender, M.E., 2016. Estimating species richness using environmental DNA. *Ecology and Evolution*, *6*(12), pp.4214-4226.

Ogden, T.H. and Whiting, M.F., 2005. Phylogeny of Ephemeroptera (mayflies) based on molecular evidence. *Molecular Phylogenetics and Evolution*, *37*(3), pp.625-643.

Parhi, J., Tripathy, P.S., Priyadarshi, H., Mandal, S.C. and Pandey, P.K., 2019. Diagnosis of mitogenome for robust phylogeny: A case of Cypriniformes fish group. *Gene, 713*, p.143967.

Pfrender, M.E., Hawkins, C.P., Bagley, M., Courtney, G.W., Creutzburg, B.R., Epler, J.H., Fend, S., Ferrington Jr, L.C., Hartzell, P.L., Jackson, S. and Larsen, D.P., 2010. Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. *The Quarterly Review of Biology*, *85*(3), pp.319-340.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I. and Feio, M.J., 2018. The future of biotic indices in the ecogenomic era: Integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of the Total Environment*, *637*, pp.1295-1310.

Rahman, M.M., Noren, M., Mollah, A.R. and Kullander, S.O., 2019. Building a DNA barcode library for the freshwater fishes of Bangladesh. *Scientific Reports*, 9(1), pp.1-10.

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system (www. barcodinglife.org). *Mol. Ecol. Notes, 7*, pp 355–364.

Ribera, I., 2008. Habitat constraints and the generation of diversity in freshwater macroinvertebrates. Aquatic insects: challenges to populations, pp. 289-311.

Shu, L., Ludwig, A. and Peng, Z., 2021. Environmental DNA metabarcoding primers for freshwater fish detection and quantification: In silico and in tanks. *Ecology and Evolution*, *11*(12), pp.8281-8294.

Stokkan, M., Jurado-Rivera, J.A., Oromí, P., Juan, C., Jaume, D. and Pons, J., 2018. Species delimitation and mitogenome phylogenetics in the subterranean genus Pseudoniphargus (Crustacea: Amphipoda). *Molecular Phylogenetics and Evolution, 127*, pp.988-999.

Thomas, J.A., Frandsen, P.B., Prendini, E., Zhou, X. and Holzenthal, R.W., 2020. A multigene phylogeny and timeline for Trichoptera (Insecta). *Systematic Entomology*, *45*(3), pp.670-686.

Thomsen, P.F., Kielgast, J.O.S., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L. and Willerslev, E., 2012. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, *21*(11), pp.2565-2573.

Turner, C.R., Uy, K.L. and Everhart, R.C., 2015. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, *183*, pp.93-102.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F. and Gaboriaud, C., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*(4), pp.929-942.

Wang, Y., Chen, K., Gao, J., Wang, M., Dong, J., Xie, Y., Giesy, J.P., Jin, X. and Wang, B., 2021. Environmental DNA of preservative ethanol performed better than water samples in detecting macroinvertebrate diversity using metabarcoding. *Diversity and Distributions*, *27*(10), pp.1989-2002.

Weigand, H., Beermann, A.J., Čiampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M.F., Grabowski, M., Rimet, F., Rulik, B. and Strand, M., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of the Total Environment*, *678*, pp.499-524.

Xu, D., Han, Z., Shan, T., Yano, C. F., eds. 2021. *Exploring the Genetic Diversity of Marine Organisms Based on the Analysis of Chromosome and Genomic DNA Markers*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-693-7.

Yan, L., Pape, T., Elgar, M.A., Gao, Y. and Zhang, D., 2019. Evolutionary history of stomach bot flies in the light of mitogenomics. *Systematic Entomology*, *44*(4), pp.797-809.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle, B.E. and De Barba, M., 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*(8), pp.1857-1862

6.0 Appendix

Table 2.1 Sampling sites in lowland River with GPS coordinates

Rivers	Sites no	Site Code	Latitude	Longitude
Dhaleshwari	1	D1	23°35.101′ N	090º16.317′ E
	2	D2	23°34.434′ N	090º17.175′ E
	3	D3	23°33.876′ N	090º17.380′ E
	4	D4	23°39.245′ N	090º17.378′ E
	5	D5	23°39.308′ N	090º15.466′ E
Buriganga	1	Bu1	23°43.279′ N	090º21.283′ E
	2	Bu2	23°43.497′ N	090º21.387′ E
	3	Bu3	23°43.522′ N	090º21.047′ E
	4	Bu4	23°43.748′ N	090º21.239′ E
	5	Bu5	23°44.469′ N	090º20.837′ E
Turag	1	T1	23°47.887′ N	090º19.659′ E
	2	T2	23°47.923′ N	090º19.297′ E
	3	T3	23°48.184′ N	090º19.112′ E
	4	T4	23°48.287′ N	090º18.463′ E
	5	T5	23°48.252′ N	090º18.091′ E
Kaliganga	1	KL1	23°43.164′ N	090º11.565′E
	2	KL2	23°43.053′ N	090º12.724′E
	3	KL3	23°44.371′ N	090º15.772′E
	4	KL4	23°42.425′ N	090º14.154′E
	5	KL5	23°43.364′ N	090º15.738′E

Table 2.2 Name of upland rivers with GPS coordinates

No.	Name of Stream/Rivers	River code	Longitudes	Longitudes
1	Sangukhiang Chhora	S	22°03.986′N	092º18.061′E
2	Betchhora	B	22°05.640′ N	092º16.694′ E
3	Cheihkhiang Chhora	C	22°09.000′ N	092º12.484′ E
4	Bangchhora	BN	21°58.392′ N	092º13.523′ E
5	Ranginmukh Chhora	RN	21°59.555′ N	092º14.242′ E
6	Maddyamkhal Chhora	MD	22°02.672′ N	092º13.519′ E
7	Eddmara Jhirri	ED	21°52.923′ N	092º22.340′ E
8	Armycamp Chhora	AR	21°49.133′ N	092º25.535′ E
9	Mongot Jhirri	MN	21°47.889′ N	092º24.882′ E
10	Sandak Jhirri	SN	21°48.538′ N	092º26.620′ E
11	Paddayo Jhirri	PD	21°45.649′ N	092º27.305′ E
12	Tindupoint	TN	21°43.512′ N	092º27.598′ E
13	Chhotoyangrypoint	CY	21°46.496′ N	092º26.480′ E
14	SemakhalChhora	SM	21°59.116′ N	092º22.007′ E
15	Belden Chhora	BL	22°03.393′ N	092º23.883′ E
16	RumakhalChhora	RM	22°01.976′ N	092º25.090′ E

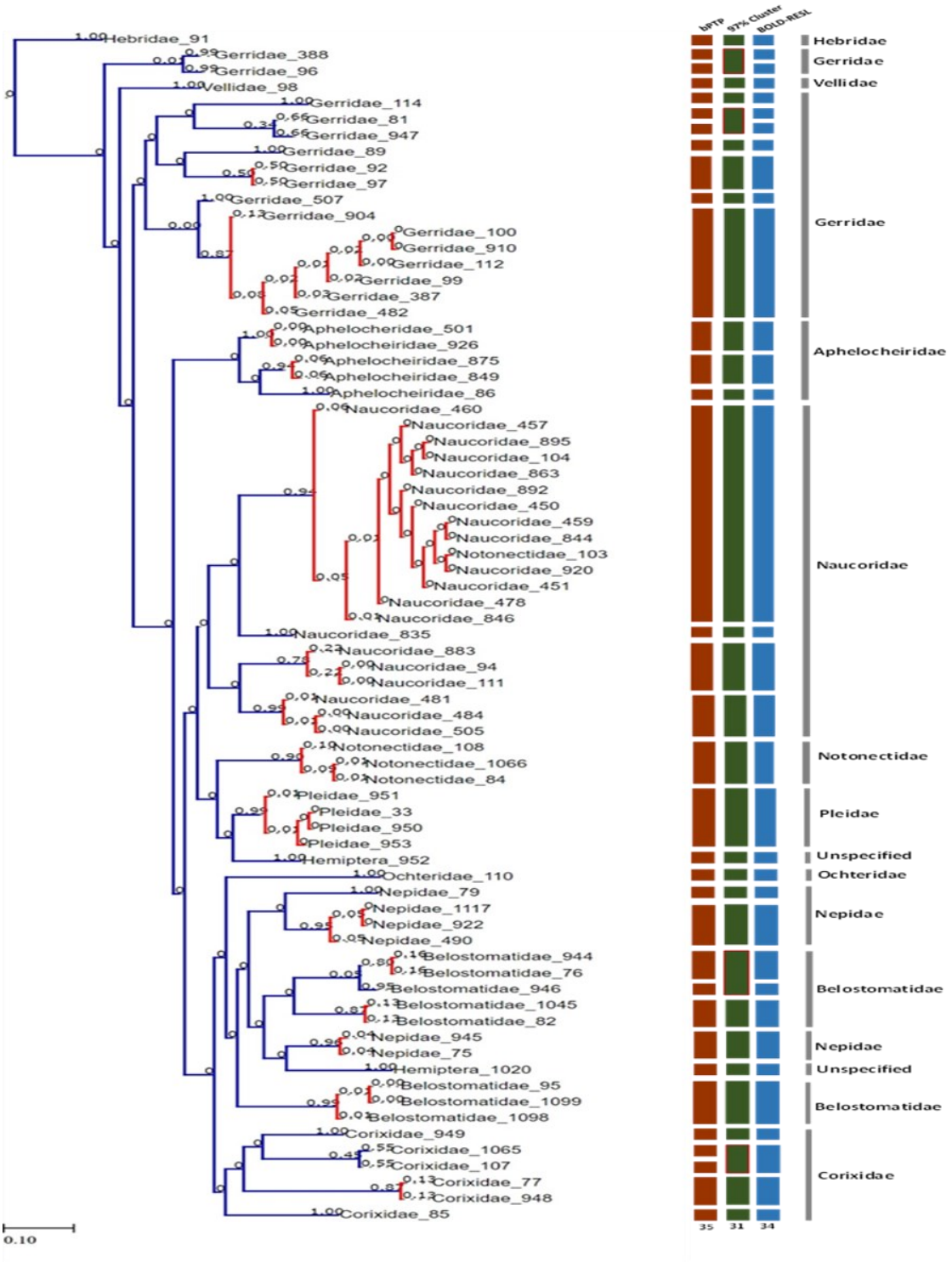

Figure 2.1 Molecular species-delimitation analysis of the Hemiptera spp. by three methods: a Bayesian implementation of the Poisson tree processes (bPTP), Sequence clustering by 97% threshold clustering and Refined Single Linkage (RESL) analysis. Delimitation results are visualized as bars on a Bayesian maximum clade credibility tree of the cytochrome c oxidase subunit 1 gene.

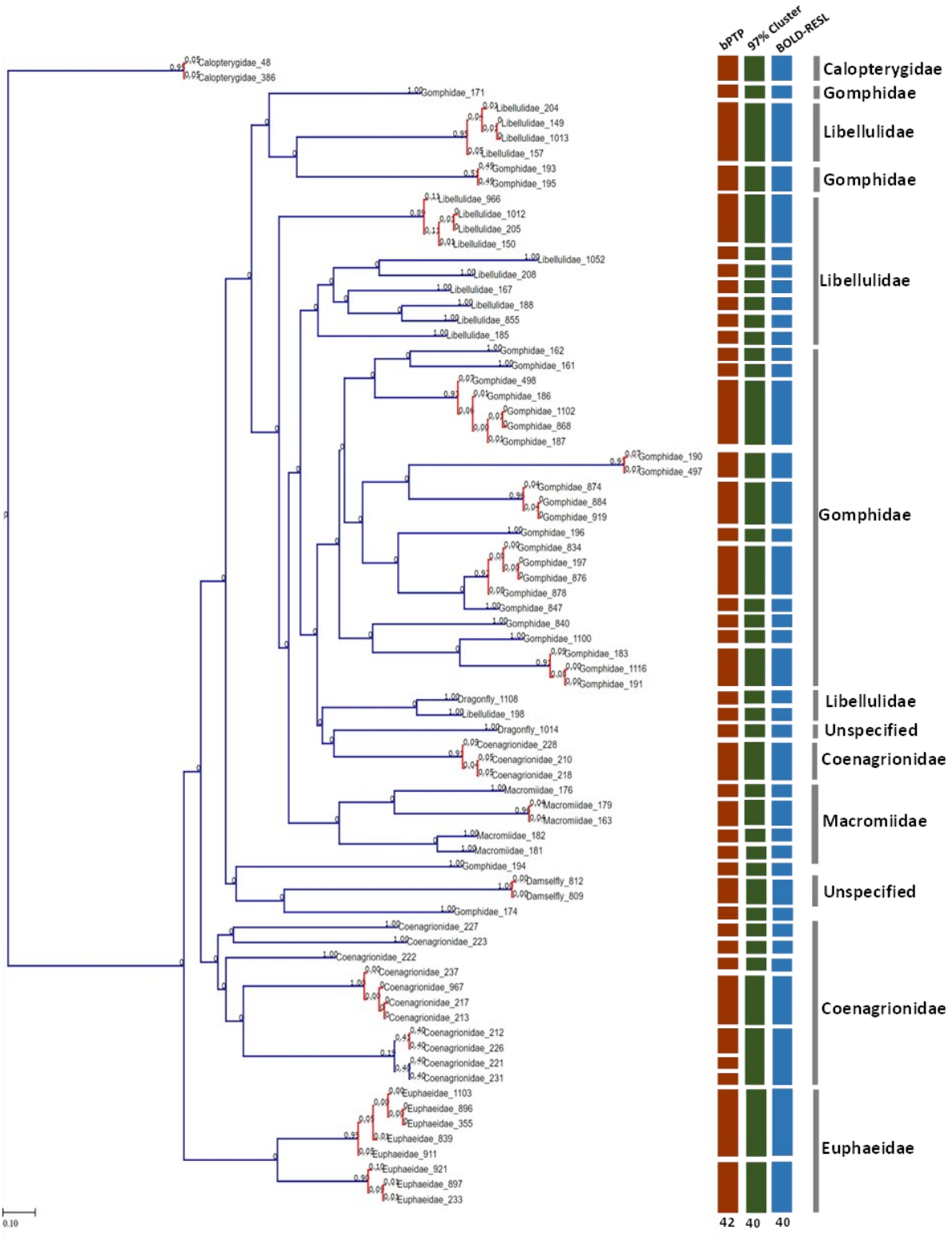

Figure 2.2 Molecular species-delimitation analysis of the Odonata spp. by three methods: a Bayesian implementation of the Poisson tree processes (bPTP), Sequence clustering by 97% threshold clustering and Refined Single Linkage (RESL) analysis. Delimitation results are visualized as bars on a Bayesian maximum clade credibility tree of the cytochrome c oxidase subunit 1 gene.

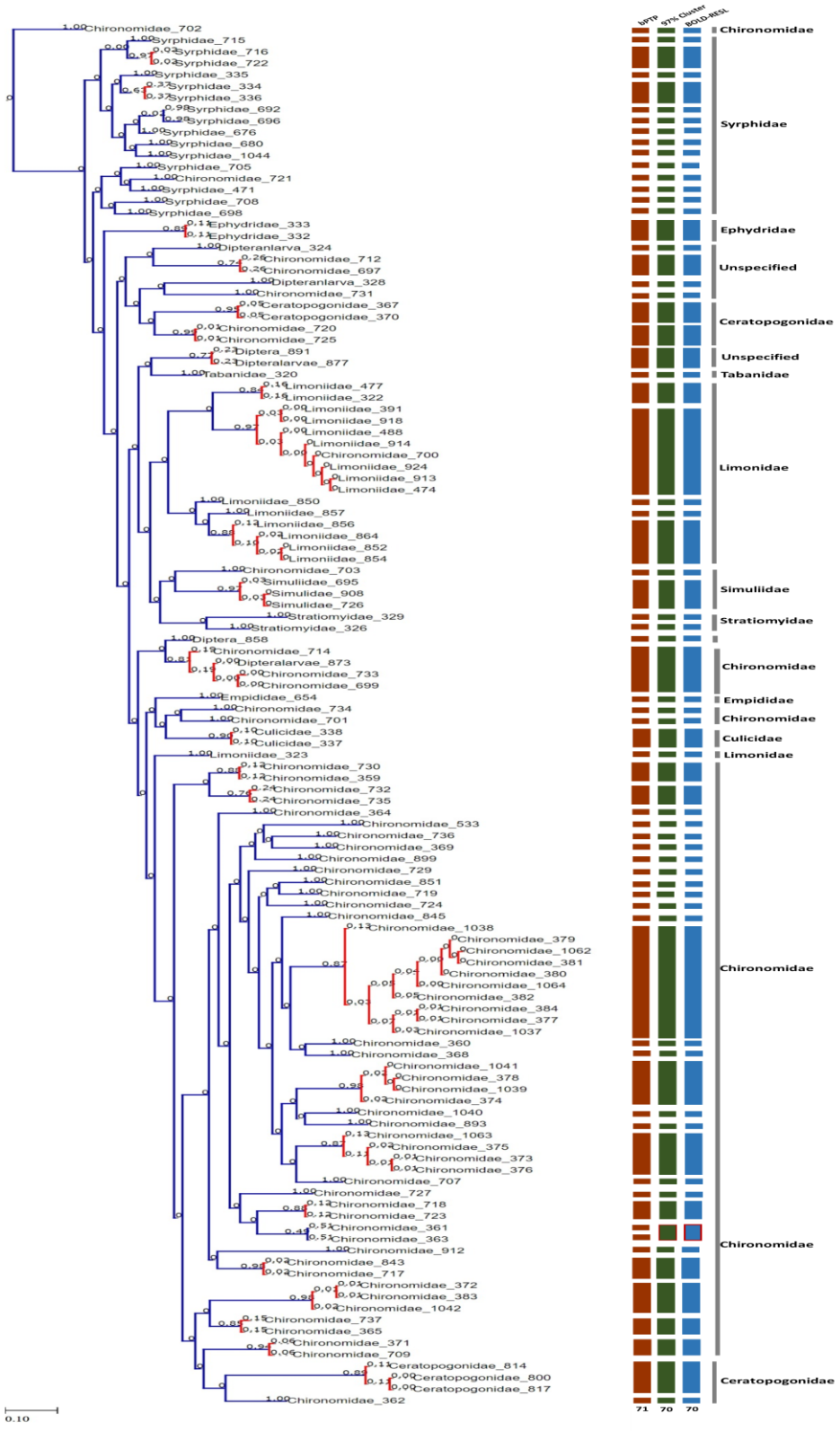

Figure 2.3 Molecular species-delimitation analysis of the Diptera spp. by three methods: a Bayesian implementation of the Poisson tree processes (bPTP), Sequence clustering by 97% threshold clustering and Refined Single Linkage (RESL) analysis. Delimitation results are visualized as bars on a Bayesian maximum clade credibility tree of the cytochrome c oxidase subunit 1 gene.

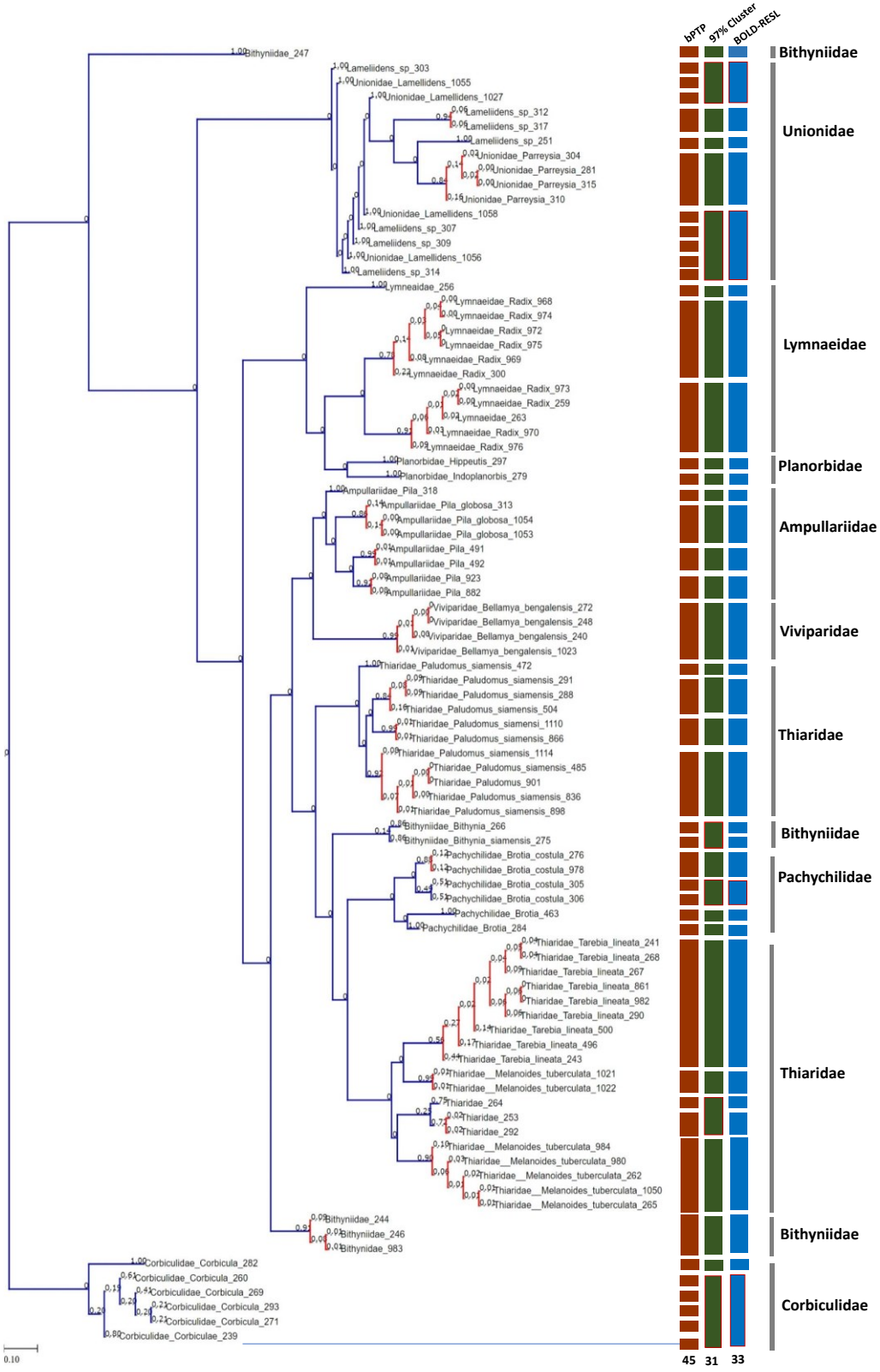

Figure 2.4 Molecular species-delimitation analysis of the Mollusca spp. by three methods: a Bayesian implementation of the Poisson tree processes (bPTP), Sequence clustering by 97% threshold clustering and Refined Single Linkage (RESL) analysis. Delimitation results are visualized as bars on a Bayesian maximum clade credibility tree of the cytochrome c oxidase subunit 1 gene.

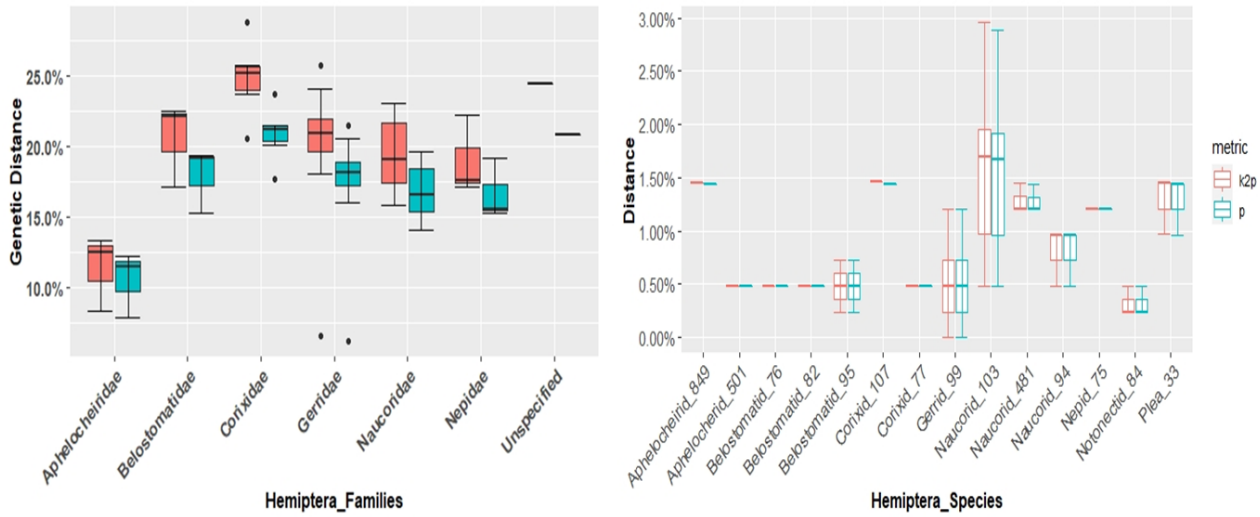

Figure 2.5 Interspecific (within each family) (left) and intraspecific (right) genetic distance of Hemiptera based on K2P and p-distance models.

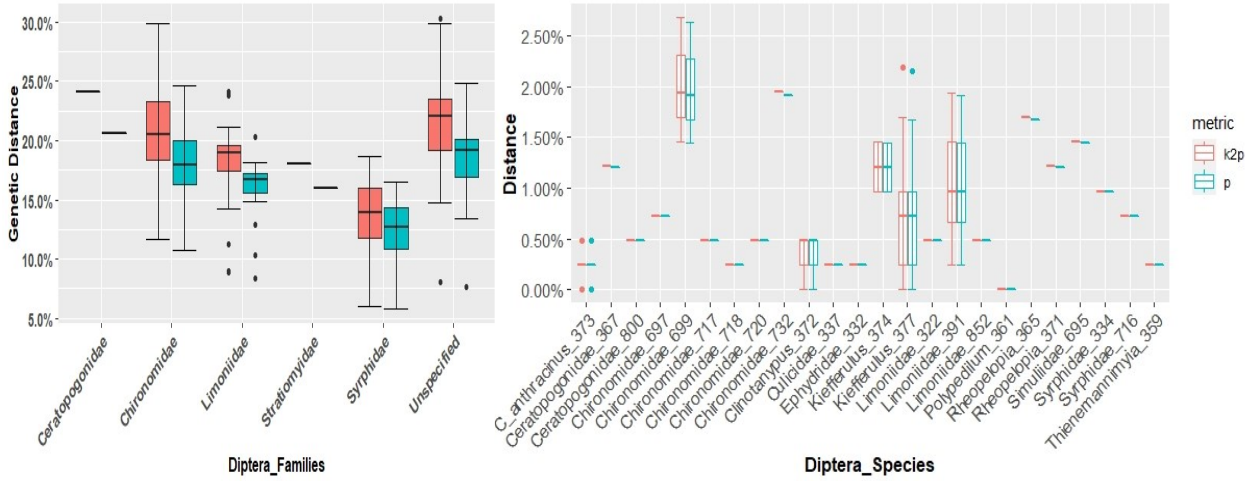

Figure 2.6 Interspecific (within each family) (left) and intraspecific (right) genetic distance of Diptera based on K2P and p-distance models.

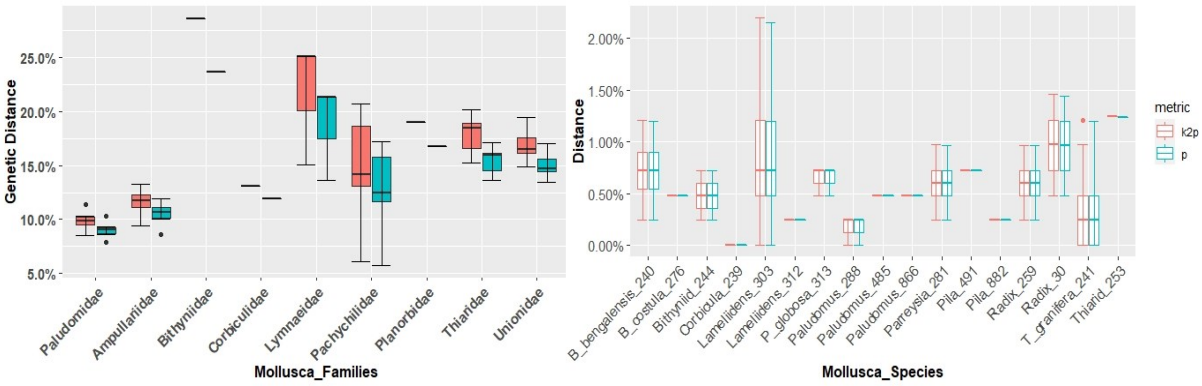

Figure 2.7 Interspecific (within each family) (left) and intraspecific (right) genetic distance of Mollusca based on K2P and p-distance models.

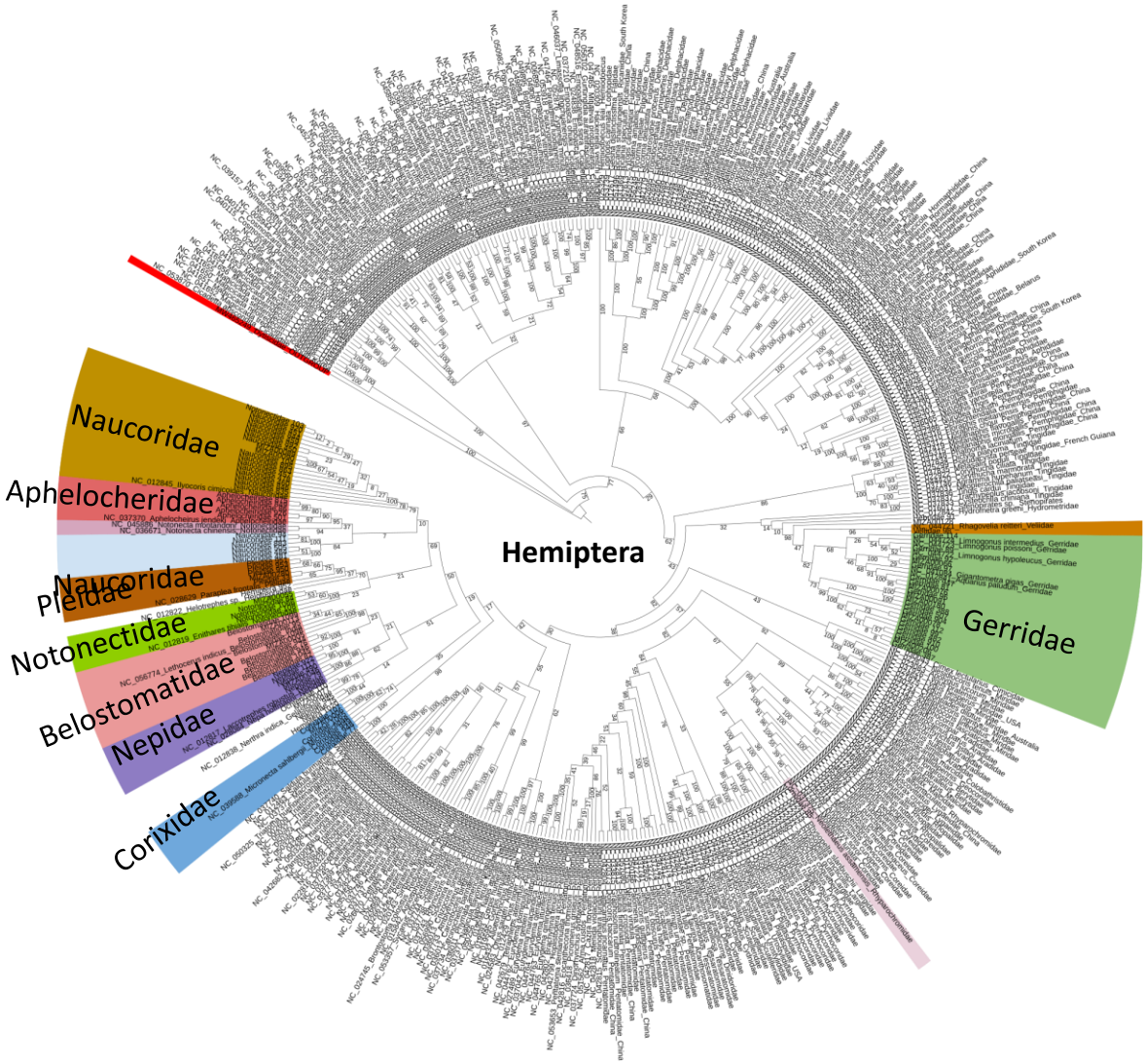

Figure 2.8 Maximum likelihood tree for hemipteran species constructed with nucleotide sequences of protein-coding genes and COI barcodes of mitogenomes

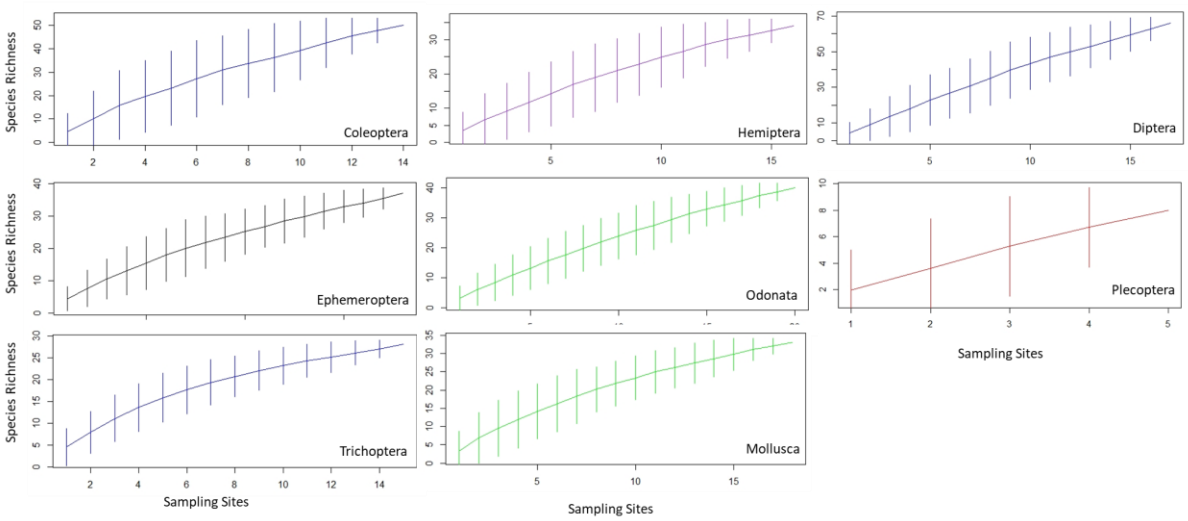

Figure 2.9 Major taxa-wise species accumulation curve showing the species richness in sampling sites

Table 2.3 Species identification of Odonata against Genbank and GBIF entries (within 3% level) using BLAST and sequence ID tool.

Calopterygidae_386	100	Neurobasis chinensis	100	Neurobasis chinensis
Coenagrionidae_218	100	Ceriagrion coromandelianum	99.761	Ceriagrion coromandelianum
Libellulidae_167	100	Trithemis pallidinervis	99.761	Trithemis pallidinervis
Libellulidae_185	100	Trithemis festiva	99.761	Trithemis festiva
Libellulidae_188	100	Orthetrum pruinosum	99.761	Orthetrum
Coenagrionidae_222	99.761	Aciagrion olympicum	99.761	Pseudagrion decorum
Coenagrionidae_227	99.761	Onychargia atrocyana	99.522	Onychargia atrocyana
Lebellulidae_1052	99.761	Brachythemis contaminata	99.043	Brachythemis contaminata
Lebellulidae_855	99.761	Orthetrum glaucum	99.761	Orthetrum glaucum
Libellulidae_208	99.754	Zyxomma petiolatum	99.522	Zyxomma petiolatum
Lebellulidae_1012	99.522	Crocothemis servilia	99.522	Crocothemis servilia
Lebellulidae_157	99.522	Urothemis signata	99.522	Libellulidae
Dragonfly_1014	99.519	Brachydiplax chalybea	99.519	Brachydiplax chalybea
Coenagrionidae_217	99.282	Pseudagrion rubriceps	99.043	Pseudagrion rubriceps
Libellulidae_198	99.282	Zygonyx iris malayanus	99.282	Zygonyx iris
Gomphidae_1100	99.043	Paragomphus capricornis	99.043	Paragomphus capricornis
Dragonfly_171	98.638	Aethriamanta brevipennis	98.638	Aethriamanta brevipennis
Coenagrionidae_212	98.325	Pseudagrion microcephalum	97.608	Pseudagrion microcephalum
Coenagrionidae_223	98.086	Pseudocopera ciliata	98.086	Pseudocopera ciliata
Euphaeidae_897	97.917	Euphaea ochracea	97.917	Euphaea ochracea
Gomphidae_1116	---	---	99.282	Paragomphus lineatus
Macromiidae_182	---	---	98.325	Epophthalmia frontalis

Table 2.4 Species identification of Hemiptera against Genbank and GBIF entries (within 3% level) using BLAST and sequence ID tool

Major taxa group	Species_ID	NCBI		GBIF	
		Identity %	Nearest taxa matched	Identity %	Nearest taxa matched
Hemiptera	Belostomatidae_1098	99.761	Lethocerus patruelis	99.522	Lethocerus
	Gerridae_89	99.522	Gerris sp.	99.522	Gerris
	Nepidae_79	97.837	Laccotrephes griseus	97.837	Laccotrephes griseus
	Nepidae_1117	98.45	Laccotrephes maculatus	98.45	Laccotrephes maculatus
	Belostomatidae_1045	97.567	Diplonychus rusticus	97.567	Diplonychus rusticus
	Pleidae_33	97.122	Paraplea frontalis	97.122	Paraplea frontalis
	Gerridae_97	98.81	Amemboa kumari		
	Gerridae_387	---	---	97.129	Gerridae
	Corixidae_107	---	---	99.761	Corixidae
Mollusca	Corbicula_282	100	Corbicula sp.	99.761	Corbicula
	Lameliidens_312	100	Pilsbryoconcha	100	Pilsbryoconcha
	Brotia costula_276	100	Brotia costula	100	Brotia costula
	Indoplanorbis_279	99.761	Indoplanorbis exustus	99.761	Indoplanorbis exustus
	Radix_30	99.522	Cerasina oxiana	99.282	Radix
	Melanoides tuberculata_262	99.522	Melanoides tuberculata	99.522	Melanoides tuberculata
	Bellamya bengalensis_272	99.522	Bellamya bengalensis	99.522	Filopaludina bengalensis
	Melanoides tuberculata_1021	99.043	Melanoides tuberculata	99.036	Melanoides tuberculata
	Lameliidens_307	99.282	Lamellidens marginalis	98.804	Lamellidens
	Lameliidens_251	99.281	Unionidae	98.561	Indonaia
	Pila globosa_1054	98.03	Pila globosa	98.03	Pila globosa
	Thiaridae_263	98.325	Radix auricularia	97.368	Radix
	Parreysia_281	98.086	Parreysia sp.	97.368	Parreysia corrugata
	Lymneaidae_256	97.368	Succinea vitrea	---	----
	Tarebia_lineata_241	100	Tarebia granifera	---	-----

Table 2.5 Species identification of Diptera against Genbank and GBIF entries (within 3% level) using BLAST and sequence ID tool

Species_ID	NCBI		GBIF	
	Identity %	Nearest taxa matched	Identity %	Nearest taxa matched
Chironomidae_373	100	Chironomus circumdatus	100	Chironomus circumdatus
Chironomidae_382	100	Chironomus javanus	99.761	Chironomus javanus
Chironomidae_721	100	Melanostoma mellinum	99.522	Melanostoma mellinum
Culicidae_338	100	Culex pipiens	100	Culex
Dipteranlarva_328	100	Psychodidae	99.761	Clogmia albipunctata
Stratiomyidae_329	100	Hermetia illucens	97.122	Hermetia illucens
Syrphidae_1044	100	Eristalis pertinax	99.761	Eoseristalis pertinax
Syrphidae_335	100	Eristalinus sp.	100	Eristalinus
Chironomidae_729	99.761	Tanytarsus pollexus	99.761	Tanytarsus pollexus
Syrphidae_336	99.761	Helophilus hybridus	99.761	Helophilus hybridus
Syrphidae_708	99.761	Sphaerophoria scripta	99.761	Syrphidae
Syrphidae_471	99.761	Cheilosia variabilis	99.761	Cheilosia variabilis
Syrphidae_680	99.761	Eristalis arbustorum	99.522	Kiefferulus calligaster
Syrphidae_698	99.761	Syrphus ribesii	99.761	Syrphus ribesii
Tabanidae_320	99.761	Tabanus megalops	99.522	Tabanus striatus
Syrphidae_715	99.522	Helophilus pendulus	99.522	Helophilus pendulus
Syrphidae_705	99.522	Rhingia campestris	97.847	Rhingia laevigata
Chironomid_1040	99.282	Kiefferulus tainanus	97.608	Kiefferulus
Ephydridae_332	99.235	Brachydeutera sp.	99.043	Ephydridae
Simuliidae_695	99.277	Simulium aureohirtum	99.043	Simulium aureohirtum
Chironomidae_912	98.561	Chironomidae sp.	----	----
Syrphidae_692	97.368	Eristalis tenax	---	----
Chironomidae_378	99.761	Kiefferulus calligaster	---	---
Syrphidae_722	99.761	Helophilus hybridus	---	---
Chironomidae_360	99.522	Chironomus sp.	---	---
Chironomidae_703			98.854	Limoniidae
Chironomidae_360			98.804	Chironomus
Chironomidae_712			97.525	Sciaridae
Chironomidae_701			97.368	
Syrphidae_692			97.368	Eristalis tenax
Limoniidae_323			99.282	Tipulidae
Chironomidae_359			99.76	Chironomidae
Syrphidae_680			99.522	Syrphidae
Stratiomyidae_326			99.282	Stratiomyidae

Table 4.1 Name of upland rivers with GPS coordinates, elevation and substrate condition

No.	Name of streams	GPS coordinates		Elevation (m)	Substrate type
		Latitudes	Longitudes		
1	Sangukhiang Chhora	22°03.986′N	092º18.061′E	79	Boulders, gravels, pebbles, fine sediments, leaf litters
2	Betchhora	22°05.640′ N	092º16.694′ E	86	Boulders, gravels, pebbles, fine sediments, detritus
3	Cheihkhiang Chhora	22°09.000′ N	092º12.484′ E	60	Fine sediments, leaf litters, detritus
4	Bangchhora	21°58.392′ N	092º13.523′ E	67	Boulders, fine sediments, leaf litters, detritus
5	Ranginmukh Chhora	21°59.555′ N	092º14.242′ E	55	Pebbles, fine sediments, detritus
6	Maddyamkhal Chhora	22°02.672′ N	092º13.519′ E	52	Fine sediments, leaf litters, detritus
7	Eddmara Jhirri	21°52.923′ N	092º22.340′ E	66	Boulders, gravels, pebbles, fine sediments, leaf litters
8	Armycamp Chhora	21°49.133′ N	092º25.535′ E	61	Gravels, fine sediments, leaf litters, detritus
9	Mongot Jhirri	21°47.889′ N	092º24.882′ E	165	Boulders, gravels, fine sediments, leaf litters
10	Sandak Jhirri	21°48.538′ N	092º26.620′ E	70	Gravels, fine sediments, leaf litters, detritus
11	Paddayo Jhirri	21°45.649′ N	092º27.305′ E	77	Gravels, pebbles, fine sediments, detritus
12	Tindupoint	21°43.512′ N	092º27.598′ E	53	Boulders, gravels, pebbles, fine sediments, leaf litters
13	Chhotoyangrypoint	21°46.496′ N	092º26.480′ E	45	Boulders, gravels, pebbles, fine sediments, leaf litters
14	Semakhal Chhora	21°59.116′ N	092º22.007′ E	144	Boulders, pebbles, fine sediments, leaf litters
15	Belden Chhora	22°03.393′ N	092º23.883′ E	43	Gravels, fine sediments, leaf litters, detritus
16	RumakhalChhora	22°01.976′ N	092º25.090′ E	63	Gravels, fine sediments, leaf litters, detritus

Table 4.2. Regression model outputs of ETP richness with different environmental criteria (explanatory variables)

Environmental criteria	Intercept	Slope	R^2	p-value	BIC
Hydromorphology intactness + absence of pollution	22.172	9.414 10.428	0.250	3.634e-05	490.22
Absence of pollution +substrate intactness	24.203	10.581 6.807	0.247	4.3e-05	490.57
Hydromorphology + absence of pollution + substrate intactness	21.882	6.422 9.404 4.422	0.264	5.296e-05	492.15
Absence of fishing pressure + absence of pollution + substrate intactness	23.676	3.691 9.182 5.934	0.268	4.58e-05	491.83
Overall environmental intactness	20.407	22.659	0.4903	0.001	104.71

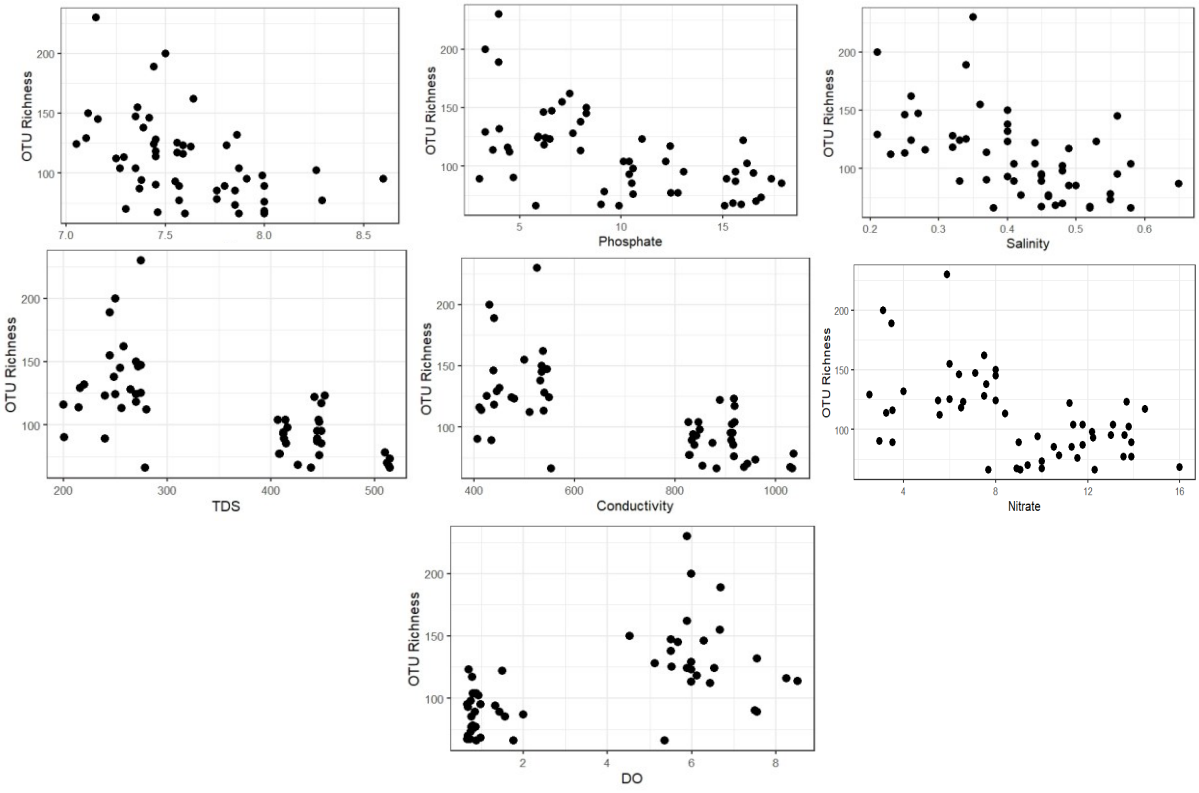

Figure 4.1A The linear regression graph showing the impact of different environmental variables on the OTU richness of macroinvertebrates

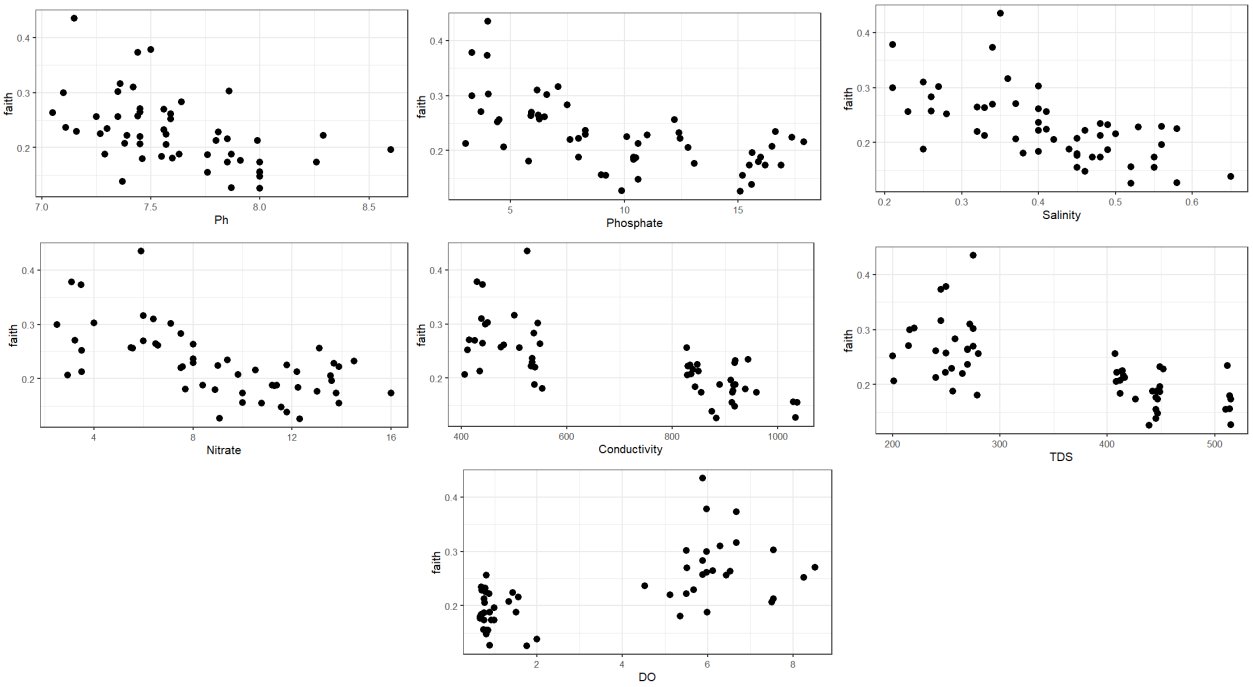

Figure: 4.1B The linear regression graph showing the impact of different environmental variables on genetic diversity (faith index) of macroinvertebrates

Table 4.3: Hydro-chemical parameters measured from lowland rivers

River	Site	PH	DO (mg/L)	Salinity (%)	Conductivity µmhos/cm	TDS (mg/L)	Phosphate (mg/L)	Nitrate (mg/L)
Turag	1	8	1.76	0.52	883	439	15.1	12.3
Turag	1	7.63	1.5	0.44	890	442	16	11.2
Turag	1	7.37	2	0.65	875	445	15.61	11.78
Turag	2	7.56	0.78	0.49	919	449	12.42	14.5
Turag	2	7.81	0.7	0.53	918	452	11.02	13.69
Turag	2	7.91	0.66	0.45	915	445	13.08	13
Turag	3	8.6	0.99	0.56	911	449	15.62	13.6
Turag	3	8.26	0.93	0.48	914	447	16.2	13.78
Turag	3	8	0.85	0.45	913	445	15.2	13.9
Turag	4	7.87	0.89	0.44	919	446	10.4	11.38
Turag	4	7.76	0.76	0.49	916	449	10.52	11.3
Turag	4	8	0.81	0.46	918	447	10.6	11.55
Turag	5	7.87	0.88	0.58	1034	515	9.9	9.07
Turag	5	8	0.74	0.52	1030	514	9	10
Turag	5	7.76	0.8	0.55	1037	510	9.18	10.76
Dhaleshwari	1	7.39	5.5	0.4	532	249	8	7.58
Dhaleshwari	1	7.16	5.67	0.56	534	255	8.3	8
Dhaleshwari	1	7.64	5.88	0.26	537	258	7.48	7.5
Dhaleshwari	2	7.29	5.99	0.25	538	256	8	8.4
Dhaleshwari	2	7.11	4.52	0.4	534	270	8.3	8
Dhaleshwari	2	7.45	5.12	0.32	540	265	7.62	7.5
Dhaleshwari	3	7.05	6.53	0.33	549	270	5.9	8
Dhaleshwari	3	7.35	5.5	0.27	545	275	6.6	7.1
Dhaleshwari	3	7.6	5.36	0.38	553	279	5.8	7.68
Dhaleshwari	4	7.1	5.98	0.21	445	216	3.3	2.5
Dhaleshwari	4	7.54	6.68	0.34	452	218	4.98	3.47
Dhaleshwari	4	7.86	7.55	0.4	450	220	4.02	4
Dhaleshwari	5	7.76	7.91	0.2	408	195	4.3	2.5
Dhaleshwari	5	7.45	7.5	0.37	406	201	4.7	2.94
Dhaleshwari	5	7.89	8.4	0.28	411	198	4.54	3.4

Table 4.4 Hydro-chemical parameters measured from lowland rivers

River	Site	PH	DO (mg/L)	Salinity (%)	Conductivity µmhos/cm	TDS (mg/L)	Phosphate (mg/L)	Nitrate (mg/L)
Buriganga	1	7.46	0.66	0.45	938	514	15.9	8.9
Buriganga	1	7.3	0.69	0.48	944	512	16.64	9.38
Buriganga	1	7.85	0.75	0.55	960	515	16.9	10
Buriganga	2	7.35	0.81	0.41	827	407	12.2	13.1
Buriganga	2	7.57	0.77	0.42	828	408	12.8	13.56
Buriganga	2	8.29	0.87	0.46	830	409	12.46	13.9
Buriganga	3	7.55	0.69	0.4	843	412	10.4	12.23
Buriganga	3	7.27	0.8	0.58	847	414	10.1	11.78
Buriganga	3	7.99	0.75	0.48	850	416	10.6	12.2
Buriganga	4	7.57	1.43	0.41	834	413	17.38	9
Buriganga	4	7.38	1.33	0.45	836	412	16.5	9.82
Buriganga	4	7.85	1.56	0.5	839	415	17.9	10.53
Buriganga	5	7.85	0.79	0.42	851	422	14.92	14.6
Buriganga	5	7.52	0.86	0.44	855	424	14.4	15.4
Buriganga	5	8	0.99	0.47	855	426	15.5	16
Kaliganga	1	7.59	5.98	0.4	480	240	6.5	6.58
Kaliganga	1	7.36	6.67	0.36	500	245	7.1	6
Kaliganga	1	7.44	5.88	0.26	475	250	6.28	5.5
Kaliganga	2	7.42	6.29	0.25	438	272	6.18	6.4
Kaliganga	2	7.56	5.52	0.34	425	275	5.92	6
Kaliganga	2	7.45	6.12	0.32	440	270	6.22	6.5
Kaliganga	3	7.25	6.43	0.23	510	280	4.5	5.56
Kaliganga	3	7.15	5.88	0.35	525	275	3.99	5.89
Kaliganga	3	7.68	6.36	0.38	515	280	4.1	4.98
Kaliganga	4	7.5	5.98	0.21	430	250	3.3	3.1
Kaliganga	4	7.44	6.68	0.34	440	245	3.98	3.47
Kaliganga	4	7.8	7.55	0.33	435	240	3.02	3.5
Kaliganga	5	7.46	7.99	0.28	422	220	3.3	3.5
Kaliganga	5	7.45	8.52	0.37	415	215	3.7	3.24
Kaliganga	5	7.59	8.25	0.28	411	200	4.42	3.5

Table 4.5 Components of environmental intactness derived from 14 binomial variables: A-Substrate composition appropriate; B-No sand, gravel or stone excavation, C-Natural channel structure; D-No damming or diversion; E-No significant water extraction; F-No dumping of household wastes; G-Minimal washing and bathing; H-Minimum Run-off from cropland; I-Natural colour and odour; J-Minimal Tourist pressure; K-Minimal Fishing Pressure; L-Presence of adjacent natural vegetation; M-Representative diversity of wild animals; N-No significant intervention by exotic plant or animal.

River	Site	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Betchhora	1	1	0	0	0	1	0	0	1	0	1	0	1	1	1
Betchhora	2	0	0	1	0	1	0	1	1	1	1	1	1	1	1
Betchhora	3	0	0	1	1	1	0	1	0	1	1	0	1	1	1
Betchhora	4	0	0	0	0	1	1	1	0	1	1	0	1	1	1
Betchhora	5	1	1	0	0	1	1	1	1	1	1	1	1	1	0
Sangukhiang Chhora	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1
Sangukhiang Chhora	2	1	1	1	1	0	1	1	1	1	1	0	1	1	1
Sangukhiang Chhora	3	1	1	1	1	1	1	1	0	1	1	0	1	1	1
Sangukhiang Chhora	4	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Sangukhiang Chhora	5	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Cheihkhiang Chhora	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0
Cheihkhiang Chhora	2	0	0	0	0	1	0	0	0	0	0	0	0	1	0
Cheihkhiang Chhora	3	0	1	0	1	1	0	0	0	0	0	0	0	0	0
Cheihkhiang Chhora	4	1	1	0	1	1	0	0	0	0	0	0	0	1	0
Cheihkhiang Chhora	5	0	0	0	0	1	0	0	1	0	0	0	1	1	1
Bangchhora	1	1	1	1	1	0	1	1	0	1	1	0	1	1	1
Bangchhora	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Bangchhora	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Bangchhora	4	0	1	1	1	1	1	1	0	1	1	1	1	1	1
Bangchhora	5	1	1	1	0	1	1	1	1	1	1	1	1	1	0
Ranginmukh Chhora	1	1	1	1	0	0	1	0	0	1	1	0	0	0	0
Ranginmukh Chhora	2	0	0	0	0	0	1	1	1	0	1	0	0	0	0
Ranginmukh Chhota	3	0	0	0	0	1	1	1	1	1	0	0	0	0	1
Ranginmukh Chhora	4	1	0	1	1	0	0	1	1	1	0	0	1	0	1
Ranginmukh Chhora	5	1	1	1	1	1	0	0	0	1	0	0	1	1	1
Eddmara Jhirri	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1
Eddmara Jhirri	2	1	1	1	1	1	1	1	1	1	1	0	1	1	0
Eddmara Jhirri	3	1	1	1	1	1	1	1	0	1	1	0	1	1	1
Eddmara Jhirri	4	1	0	1	1	1	1	1	1	1	1	1	1	1	1
Eddmara Jhirri	5	1	0	1	1	1	1	1	1	1	1	1	1	1	0

Table 4.6 Components of environmental intactness derived from 14 binomial variables: A-Substrate composition appropriate; B-No sand, gravel or stone excavation; C-Natural channel structure; D-No damming or diversion; E-No significant water extraction; F-No dumping of household wastes; G-Minimal washing and bathing; H-Minimum Run-off from cropland; I-Natural colour and odour; J-Minimal Tourist pressure; K-Minimal Fishing Pressure; L-Presence of adjacent natural vegetation; M-Representative diversity of wild animals; N-No significant intervention by exotic plant or animal.

River	Site	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Rumakhal Chhora	1	1	1	0	0	1	1	0	0	1	0	1	1	0	0
Rumakhal Chhora	2	0	1	1	0	1	0	0	0	0	0	0	0	0	0
Rumakhal Chhora	3	1	1	1	0	1	1	0	0	1	0	0	0	1	1
Rumakhal Chhora	4	1	1	1	0	0	1	1	0	1	0	0	0	0	1
Rumakhal Chhora	5	1	1	0	1	0	1	1	0	1	0	1	0	0	1
Armycamp Chhora	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Armycamp Chhora	2	0	0	0	0	1	0	0	1	0	0	1	0	0	0
Armycamp Chhora	3	0	0	0	0	1	1	0	0	0	0	1	0	0	0
Armycamp Chhora	4	0	0	0	0	1	0	0	0	1	0	1	0	0	1
Armycamp Chhora	5	0	0	0	0	1	0	0	0	1	0	0	1	0	0
Sandak Jhirri	1	0	1	1	1	1	0	0	0	0	1	1	1	1	0
Sandak Jhirri	2	0	1	1	0	1	0	0	0	1	1	0	1	1	1
Sandak Jhirri	3	1	1	1	0	0	0	0	0	1	1	0	1	1	1
Sandak Jhirri	4	1	1	1	1	0	1	1	0	1	1	0	1	0	1
Sandak Jhirri	5	1	1	1	0	0	0	1	1	1	1	0	0	0	1
Tindupoint	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1
Tindupoint	2	1	1	1	1	1	1	1	1	1	0	1	1	1	0
Tindupoint	3	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Tindupoint	4	1	1	1	1	1	1	1	0	1	0	1	0	1	1
Tindupoint	5	1	1	1	1	1	0	1	1	1	0	1	0	1	1
Chhotoyangrypoint	1	1	1	1	1	1	0	1	0	1	0	1	1	1	0
Chhotoyangrypoint	2	1	1	1	1	1	1	1	1	1	0	1	1	0	1
Chhotoyangrypoint	3	1	1	1	1	1	1	1	0	1	0	1	1	0	1
Chhotoyangrypoint	4	1	1	1	1	1	1	1	1	1	0	1	0	1	1
Chhotoyangrypoint	5	1	1	1	1	1	0	1	1	1	0	1	0	1	1
Belden Chhora	1	1	1	1	1	1	0	0	0	1	1	1	0	0	0
Belden Chhora	2	1	1	1	1	1	0	0	0	0	1	0	1	0	0
Belden Chhora	3	1	1	1	1	1	1	0	0	1	0	0	1	1	1
Belden Chhora	4	1	1	1	1	1	0	1	0	0	0	1	1	0	1
Belden Chhora	5	1	1	1	1	0	0	1	1	1	0	1	0	1	1

Table 4.7 Components of environmental intactness derived from 14 binomial variables: A-Substrate composition appropriate; B-No sand, gravel or stone excavation; C-Natural channel structure; D-No damming or diversion; E-No significant water extraction; F-No dumping of household wastes; G-Minimal washing and bathing; H-Minimum Run-off from cropland; I-Natural colour and odour; J-Minimal Tourist pressure; K-Minimal Fishing Pressure; L-Presence of adjacent natural vegetation; M-Representative diversity of wild animals; N-No significant intervention by exotic plant or animal.

River	Site	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Mongot Jhirri	1	0	0	0	0	0	1	1	0	0	1	1	0	1	1
Mongot Jhirri	2	0	0	1	0	1	1	1	1	1	1	1	1	1	1
Mongot Jhirri	3	0	0	1	0	1	0	0	0	1	1	1	1	1	1
Mongot Jhirri	4	0	1	1	0	0	0	0	0	1	1	1	0	0	1
Mongot Jhirri	5	1	1	0	1	0	0	0	0	1	0	1	1	0	0
Maddyamkhal Chhora	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maddyamkhal Chhora	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Maddyamkhal Chhora	3	0	0	1	0	1	0	0	0	0	1	0	0	0	0
Maddyamkhal Chhora	4	0	0	0	0	1	0	0	0	1	1	0	0	0	1
Maddyamkhal Chhora	5	0	0	0	0	1	0	0	0	0	1	1	0	1	1
Semakhal Chhora	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1
Semakhal Chhora	2	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Semakhal Chhora	3	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Semakhal Chhora	4	1	1	1	1	1	1	1	0	1	0	1	1	1	1
Semakhal Chhora	5	1	0	1	1	1	1	0	0	1	0	0	0	1	0
Padday Jhirri	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0
Padday Jhirri	2	1	1	1	1	1	0	1	0	1	1	0	1	1	1
Padday Jhirri	3	1	1	1	0	0	1	0	0	1	1	0	0	1	1
Padday Jhirri	4	0	1	1	1	1	1	0	0	1	1	1	1	1	1
Padday Jhirri	5	1	1	1	1	1	0	1	1	1	1	1	1	1	1