Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biology

Nina Østby

DNA Metabarcoding Diet Analysis of Water Vole (*Arvicola amphibius*) in Northern Norway

Master's thesis in Natural Science with Teacher Education
Supervisor: Bernt-Erik Sæther
Co-supervisors: Henrik Pärn, Michael D. Martin
June 2019

NTNU
Norwegian University of
Science and Technology

Nina Østby

DNA Metabarcoding Diet Analysis of Water Vole (*Arvicola amphibius*) in Northern Norway

Master's thesis in Natural Science with Teacher Education
Supervisor: Bernt-Erik Sæther
Co-supervisors: Henrik Pärn, Michael D. Martin
June 2019

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biology

NTNU
Kunnskap for en bedre verden

Abstract

Understanding what factors affect population dynamics is important in ecology, as it provides important knowledge for conservation purposes in a time where global climate change and habitat destruction and degradation threaten the earth's biodiversity. The extreme population size fluctuations in rodent populations provide good opportunities for studying the multiple factors influencing population dynamics. Food abundance and quality may be a factor contributing to the extreme population cycles seen in many rodents. For investigating the impact of food abundance and quality on a population's state, it is necessary to know the diet of the focal species. In this study, the water vole (*Arvicola amphibius*) diet was explored using DNA metabarcoding on stool samples collected from four islands off the Helgeland coast in northern Norway. The diet was dominated by Poaceae (57% on average of identified plant families per sample), but with large individual variation. Rosaceae was the second most abundant plant family in the diet (14%) with especially high proportions on the smaller islands sampled.

Key words

Metabarcoding, environmental DNA, diet analysis, water vole, *Arvicola amphibius*, population dynamics

Sammendrag

Studier og kunnskap om hvilke faktorer som påvirker populasjonsdynamikk er svært viktig i økologi. Det gir viktig informasjon som er nyttig for blant annet bevaring av biodiversitet i en tid der den trues av globale klimaendringer samt tap og ødeleggelser av habitat. De typisk ekstreme populasjonssvingningene i gnager-populasjoner gir et godt utgangspunkt for å studere de mange faktorene som påvirker populasjonsdynamikken. Mattilgang – og kvalitet kan være en faktor som bidrar til å påvirke disse svingningene typisk for mange gnager-populasjoner. For å undersøke om dette er tilfellet, er det nødvendig å vite hva den aktuelle arten som studeres spiser. I dette studiet har jeg kartlagt våndens (*Arvicola amphibius*) diett ved bruk av DNA metabarcoding på avføring samlet fra fire øyer på Helgelandskysten i Nord-Norge. Dietten var dominert av Poaceae (57% gjennomsnittlig per vånd), men med stor individuell variasjon. Rosaceae var den andre mest spiste plantefamilien (14%) med spesielt høy proporsjon i prøvene hentet fra de mindre øyene i studiet.

Acknowledgements

This master's thesis is a part of the water vole project at the Centre of Biodiversity Dynamics (CBD) at the Norwegian University of Science and Technology (NTNU) in Trondheim. The thesis was developed in cooperation with Vitenskapsmuseet where all laboratory work and part of the data analysis was conducted. It is the last part of my teacher education, and I want to thank my supervisors Michael D. Martin, Henrik Pärn and Bernt-Erik Sæther for all help. I especially want to thank Mike for all help with the lab work and initial data analysis, and Henrik for the help with analysis and plotting in R. Also a thanks to Jacob Rasmussen for assistance in the lab.

I have learnt very much about being a scientist in the past year since I were in the field collecting samples. The days of field work at Helgeland was truly a fantastic experience. Being a part of the process all the way from sampling and handling of water voles to a finished master thesis has made me appreciate my work with the thesis a lot more. Thanks to Sindre Sommerli for the training in field work methods and for all bad biology puns heard at Helgeland. Thanks to the landowners in the Skålvær archipelago, and to Bjørn and Finn for transport to the study islands. Also thanks to Pål Thorvaldsen for providing information about the vegetation on similar islands nearby the study area. Lastly, I would like to thank my fellow students at LUR and Vitenskapsmuseet for the much-needed lunch breaks and laughs along the way.

Table of contents

1 INTRODUCTION 1

1.1 UNDERSTANDING POPULATION DYNAMICS 1
1.2 POPULATION CYCLES IN RODENTS 1
1.3 HERBIVORE DIET SELECTION 2
1.3.1 Diet variation between seasons 3
1.3.2 Diet variation between sexes 4
1.4 ENVIRONMENTAL DNA AND METABARCODING 4
1.5 DIET ANALYSIS USING METABARCODING 6
1.5.1 Previous studies on herbivore diets using metabarcoding 6
1.5.2 The water vole diet 7
1.5.3 The Helgeland water vole diet 8

2 MATERIAL AND METHODS 9

2.1 STUDY SPECIES 9
2.2 STUDY SITE 9
2.3 SAMPLING METHODS 11
2.3.1 The diet analysis sample collection 12
2.4 DNA EXTRACTION, AMPLIFICATION AND SEQUENCING 13
2.5 DATA ANALYSIS 14

3 RESULTS 17

3.1 TAXONOMIC RESOLUTION OF THE SEQUENCE DATA 17
3.2 OVERALL WATER VOLE DIET 17
3.3 DIET VARIATION BETWEEN SITES 21

4 DISCUSSION 23

4.1 THE WATER VOLE DIET 23
4.1.1 Taxonomic resolution 23
4.1.2 Overall diet 23
4.1.3 Variation in diet between sites 26
4.1.4 Individual variation 27
4.2 METABARCODING ANALYSIS 28
4.2.1 Metabarcoding limitations and improvements 28
4.2.2 Limitations of the trnL P6 loop marker 29
4.3 FURTHER STUDIES 31

5 CONCLUSION 32

6 REFERENCES 33

7 APPENDIX 37

1 Introduction

1.1 Understanding population dynamics

Understanding population dynamics and community ecology is a central part of ecology and conservation. The science of community ecology includes understanding mechanisms that determine the abundances, numbers and types of species in the same geographical location (Kotler and Brown, 2007). The ongoing sixth mass extinction caused by global climate change, habitat degradation and fragmentation has massive implications on ecosystems. Thus, knowledge about community ecology has become very important for conservation biologists to be able to evaluate conservation efforts to preserve our valuable species diversity (Flather and Sieg, 2007). Species diversity is important to conserve because of each species' value, and for maintaining the overall ecosystem functionality (Flather and Sieg, 2007).

1.2 Population cycles in rodents

Rodent populations are known to fluctuate much in size (Krebs et al., 1973), and to have cycles in their population sizes of 3-4 years (Boonstra et al., 1998; Krebs et al., 1973), sometimes up to 9 years (Brzeziński et al., 2018). The same pattern is seen in rodents in many different ecological communities across the world (Krebs et al., 1973). Huge efforts have been done to understand why these cyclic fluctuations occur and whether they are caused by one or several factors (Krebs et al., 1973; Lidicker, 1988).

Lidicker (1988) is convinced that both extrinsic (such as weather conditions, vegetation productivity and predation) and intrinsic factors (such as presaturation dispersal, fecundity factors and social interactions) contribute to the extreme fluctuations in microtine populations. This multi-factorial perspective on microtine population fluctuations has been well established also in later years. The relative contributions of each factor, however, is not so well known (Cerquiera et al., 2006). Much of the debate of population regulation in e.g. small mammals relates to bottom-up (e.g. food availability) versus top-down (e.g. predators) regulation of population growth (Holt and Kimbrell, 2007). The typical top-down, bottom-up snowshoe hare and bobcat-example on predator-prey interactions is well known in ecology and established as an important factor in population regulation in both prey and predator (Holt and Kimbrell, 2007).

Food availability is a density-dependent factor that affects the dynamics of populations. Inter- and intraspecific competition for food and other resources grows with population density.

When population sizes are growing towards the carrying capacity of the environment the population exists in, competition for resources increases, the mortality rates increase, and birth rates decrease (Holt and Kimbrell, 2007). Food may thus be a limiting factor for population growth. Foraging behavior is a central part of understanding population dynamics and community ecology, and interest in foraging theory grew rapidly among ecologists in the 1960's (Ydenberg et al., 2007). Foraging theory includes models of animal foraging behavior and how foraging shapes the ecosystem of the animals foraging. Animals may change their foraging decisions in different situations in order to meet their energy- and nutritional demands (Newman, 2007). They are affected by, and also affect, the community they are part of through foraging (Kotler and Brown, 2007).

In a study on the population structure and the reproductive pattern of water voles, Cerquiera et al. (2006) claims that the survival of juveniles plays an important role in creating these fluctuations. Predation is believed to be the main cause of death among juvenile water voles (Cerquiera et al., 2006), but also the mother's living conditions (e.g. resource availability such as food) can affect juvenile survival through the mother's capacity to take care of her offspring. The most critical phase is at or near weaning, and when populations decline, evidence from several studies on microtines suggest that lactation and maternal condition is associated with juvenile survival (Boonstra, 1994). Low resource availability and/or quality (of e.g. food) for the mother can thus indirectly affect juvenile survival. High mortality of juveniles in the peak phases of vole populations create an ageing population and leads to a decline in population size the following years, as the number of actively reproducing females declines (Cerquiera et al., 2006). However, since juvenile mortality rates must be extremely high for the population to decline, Cerquiera et al. (2006) concludes that it plays an important role in determining population cycles, but cannot be the only factor doing so.

1.3 Herbivore diet selection

Animals are shown to prefer food that corrects a nutritional deficit or imbalance in their demand for nutrients (Newman, 2007). Plants are limited in their nutritional contents and are harder to extract nutrients from compared to animal-derived food. They are also believed to be more varied in the ratios of nutrients between them. Because of the limited and varied nutrient content of plants, herbivores would benefit from having a varied diet composed of different plant species that covers their nutritional demands (Newman, 2007). How animals distinguish between plants with different nutrient compositions is still somewhat unclear, but they are believed to be able to use visual cues, odor and taste (Newman, 2007).

As well as providing herbivores with nutrients, plants produce secondary metabolites that are toxic for the herbivores in bigger amounts. These compounds may influence the diet selection of herbivores (Newman, 2007). Herbivores cannot totally avoid toxic chemicals in their diet, as plants contain a variety of different defensive chemicals. However, the ability of vertebrate herbivores to detect chemicals in plants enable them to regulate the amount of harmful chemicals consumed, below a threshold where the body is able to detoxify and excrete them (Heiska et al., 2007). They stay below the threshold by having a varied diet consisting of different plants with different toxins (Newman, 2007). Some studies have found that plant quality (nutrients, toxins) plays a key role in determining the fluctuations in population sizes of voles (Heiska et al., 2007), while others have found no such connection (Boonstra et al., 1998). Cerquiera et al. (2006) argues that extrinsic factors such as food availability, as well as predation and parasitism, might play a role in the diversity of the *duration* of the cycles, rather than *creating* the cycles.

The degree to which food abundance and quality affect the population cycles of microtines is thus somewhat unclear but is suggested as one out of many explanatory factors for the cycles.

1.3.1 Diet variation between seasons

The high energetic requirements of subterranean species in many cases forces them to have a wide diet, eating what is available to them. The cost of burrowing gives underground foragers large search expenditures, and with variable food resources encountered while burrowing, they benefit from being opportunistic generalists (Comparatore et al., 1995). However, studies on rodent diets have detected variations in food preferences between seasons and between sexes (Puig et al., 1999). Several species of Poaceae (grass) are rich in proteins and are easily digested, and therefore make a good diet choice for subterranean rodents (Puig et al., 1999). Water voles seem to select plant species for their nutritional content (Neyland, 2011). In a study conducted on a Welsh water vole population, Neyland (2011) found some seasonal variation in the voles' preference for certain plant species. In their study on the burrowing rodent *Ctenomys mendocinus* (among others), Puig et al. (1999) found lower individual variation in their diets during summer when food was abundant and nutritious. According to the optimal foraging theory, specialization to fewer species tend to happen when the food items are abundant or has a high nutritional value (Pyke et al., 1977). During winter, mature grasses often have a lower nutritional quality. *C. mendocinus* was found to compensate for this by increasing their foraging on other plants like shrubs and succulents during the winter season (Puig et al., 1999).

1.3.2 Diet variation between sexes

Differences in dietary composition between rodent males and females are mainly due to differences in nutritional requirements in growth and reproduction stages (Puig et al., 1999). Females show higher nutritional requirements during pregnancy and lactation period, preferring grasses which are abundant and nutritious. Pregnant water voles have also been observed eating other non-plant taxa like frogs (Neyland, 2011). Males tend to have a more varied winter diet, before the reproductive season. In this period, males increase their burrowing activity to find mates and encounter a higher number of different food items (Puig et al., 1999).

1.4 Environmental DNA and metabarcoding

To investigate the contribution of food abundance and quality to population dynamics, we first need to know the diet of the focal species. When the diet is known, connections between population state and plant abundance, availability, quality and growth seasons can be investigated. If enough quantitative data is available, one can use this to detect shifts in diet between conspecific populations depending on the presence or absence of competitors, predators and impact of human activities (Hawlitschek et al., 2018). Diet analysis can be performed using different techniques, and one technique that is becoming more and more common is the eDNA metabarcoding approach.

Environmental DNA (eDNA) is becoming an important part of ecology and environmental management. eDNA is genomic DNA from different species in a single sample retrieved from the environment (Bohmann et al., 2014). An eDNA sample can be collected from soil, sediment, water, feces, etc. As an eDNA sample consist of different organisms' DNA and can be highly degraded and in low concentrations, it may pose more analytical challenges in comparison to a tissue sample from a single organism (Taberlet et al., 2018). However, the development of techniques for analyzing eDNA samples provides great potential for studying environments and ecosystems. It is a good tool for monitoring the health of entire ecosystems, as it can reveal the presence of rare species, invasive species and parasites that are otherwise not easy to detect (Bohmann et al., 2014; Gomes et al., 2017; Taberlet et al., 2018). The impact of pollution and other anthropogenic pressures on ecosystem functionality can also be assessed. These kinds of environmental studies provide important information that can be used for e.g. conservation purposes (Bohmann et al., 2014; Taberlet et al., 2018). Environmental sampling also enables non-invasive detection, which is especially handy when

working with endangered or otherwise vulnerable species (Gomes et al., 2017; Hawlitschek et al., 2018; Taberlet et al., 2018).

One of the main characteristics of eDNA is that it is composed of a mixture of DNA from different types of organisms, all extracted from the same sample. With DNA metabarcoding, it is possible to identify all species within a clade that is contained in bulk samples without isolating each of them before sequencing, given that the corresponding barcode sequences for all these species are included in the reference database used (Taberlet et al., 2018; Hawlitschek et al., 2018). Among other things, diet studies have become easier with the eDNA technology as droppings from animals can be collected in a non-invasive manner and analyzed for all food items in one single metabarcoding session (Taberlet et al., 2018).

A metabarcode is a short fragment of taxonomically informative DNA with a conserved region on each end that makes up an anchor for PCR (Polymerase Chain Reaction) primers. The ideal metabarcode is short and highly variable between different species (Taberlet et al., 2018). Plant DNA is often targeted using primers for the P6 loop of the chloroplast trnL intron (10-143 bp (base pairs)) (Taberlet et al., 2006). The resolution for this primer is lower than that of the whole trnL (UAA) intron (254-767 bp), but using this shorter intron for metabarcoding of eDNA samples comes with some great advantages: the amplification system of this short P6 loop is very robust and is able to amplify highly degraded DNA. This intron is therefore often used in analysis of ancient DNA and diet analysis using stool samples (Taberlet et al., 2006). Furthermore, the resolution of the P6 loop is higher than alternative systems (Taberlet et al., 2006).

In addition to being able to identify different taxa in a single sample, metabarcoding also allows genetic analysis of pooled samples from different sources by indexing each samples' DNA molecules with a unique combination of index primers before multiplexing (pooling) the samples (Debrovolny et al., 2019). When tagged with unique index primer combinations, the samples can be demultiplexed after sequencing, so that sequences of DNA detected in the sequencer can be allocated to its original sample (Taberlet et al., 2018). This parallel sequencing method is easily implemented using Next Generation Sequencing (NGS), which has revolutionized biological science since 2005, because it enables sequencing of up to millions of different DNA fragments from different samples simultaneously at a relatively low cost (Shokralla et al., 2012; Taberlet et al., 2018; Hawlitschek et al., 2018).

Briefly described, metabarcoding by NGS includes extraction of DNA from samples, PCR amplification of a target region (e.g. for vascular plants: chloroplast P6 loop of the trnL intron), followed by data generation on a sequencing instrument (Taberelet et al., 2018). Computer programs are then used for organizing the sequences into clusters of identical/similar sequences and comparing them to reference sequences in taxonomic databases, with which the sequences are assigned to specific taxa (Taberlet et al., 2018). One such program is OBITools (Boyer et al., 2016).

The metabarcoding technology is highly dependent on availability of extensive taxonomic reference databases, such as GenBank, EMBL (European Molecular Biology Laboratory) and BOLD (Barcode of Life Data System). Each sequence obtained from the samples is compared to the sequences in the reference libraries to determine what species it originates from (Soininen et al., 2015; Taberlet et al., 2018). Each sequence is assigned to its closest match in the reference library. If the reference library is incomplete and no match is found, a sequence may be assigned to a closely related, yet incorrect taxon (Taberlet et al., 2018).

1.5 Diet analysis using metabarcoding

Metabarcoding has become a very popular technique for diet analysis in the recent years. Diet compositions can be determined by metabarcoding samples from gut content or feces (Taberlet et al., 2018). It is less costly and less time consuming than manually identifying food fragments in feces using macro- or microhistology, or directly observing foraging behavior (Taberlet et al., 2018). Metabarcoding also yields a finer taxonomic resolution and potentially identifies more taxa, compared to these traditional methods (Soininen et al., 2015). Not needing to directly handle animals in the sampling process has also made metabarcoding more popular (Hawlitschek et al., 2018). Feces is relatively easy to obtain, generally non-invasive, and therefore offers a good source of DNA for diet analysis. However, the main issue with feces is that the DNA in it is often highly degraded as a result of passing through an animal's digestive tract (Taberlet et al., 2018; Hawlitschek et al., 2018).

1.5.1 Previous studies on herbivore diets using metabarcoding

There are several studies where metabarcoding has been used for herbivore diet analysis. Soininen et al. (2009, 2015) used metabarcoding to map the diets of both voles (*Microtus oeconomus* and *Myodes rufocanus*) and lemmings (*Lemmus lemmus*). Lopes et al. (2015) compared the diets of two subterranean rodent species (*Ctenomys minutus* and *C. flamaroini*). Iwanowicz et al. (2016) mapped the diet of the Pacific pocket mouse (*Perognathus*

longimembris pacificus). Willerslew et al. (2014) used metabarcoding on extinct megafauna gut content. In all these studies, Poaceae seems to be the most eaten plant family.

In most of these studies, the universal trnL g-h primer pair has been used. Soininen et al. (2009) also used the trnL c-h primer pair which is universal for all plant taxa, in order to identify bryophytes in the lemming diet. As the *Ctenomys* rodents are omnivorous, Lopes et al. (2015) used a variety of different primers. For mollusks, arthropods and vertebrates: the primer pair 16SMAV-F and 16SMAV-R, which amplifies a fragment of the 16S rDNA mitochondrial gene. Blocking primers were also used to avoid amplification of the rodents themselves (Lopes et al., 2015). In addition to the trnL g-h primer pair, Lopes et al. (2015) also included primers specific for the Asteraceae family (ITS1-F and ITS1Ast-R) and the Poaceae family (ITS1-F and ITS1Poa-R). Iwanowicz et al. (2016) did not use any trnL-primer, but instead used ITS5a and ITS4, primers targeting ribosomal DNA. In addition to the trnL g-h primer pair, Willerslev et al. (2014) used the primer pairs for Asteraceae and Poaceae (same as Lopes et al. (2015)), and also Cyperecae (ITS1-F and ITS1Cyp-R).

1.5.2 The water vole diet

Many previous studies on water voles have been conducted with the more aquatic vole type in England. Only in Britain, 227 different plant species have been identified in the water voles' diet. This long list of plant species is compiled from several different habitats over large parts of Britain (Strachan et al., 2011). It will be unrealistic to expect one individual or geographically restricted population to eat such a wide variety of plants, but it gives an indication of the versatile diet possibilities of the vole. Strachan et al. (2011) mention lush aerial stems and waterside grasses as the most important parts of the British vole diet, and in a few instances, insects, mollusks, crabs and fish have been eaten. Ashby and Vincent (1976) used microscopical analysis of faeces to determine the diet of a population of *A. amphibius* in Britain. They found grasses (Poaceae) to comprise 80% of the winter diet, 65% of the spring and approximately 90% of the summer and autumn diet (Ashby and Vincent, 1976). In a population of Welsh water voles, Neyland (2011) detected 23 plant species (and three non-plant species – two frogs and one snail in three single events) in the water vole diet. A species of Juncaceae, *Jucus effusus*, was found to be the most important plant in the focal water vole. The second most eaten species was a Poaceae species, *Typha latifolia*, followed by *Carex riparia* (Cyperaceae) and *Epilobium hirsutum* (Onagraceae).

In Belarus, the water vole diet has been found to consist of 92 different species (Wilson et al., 2017). Generally, for the Eurasian water vole, many different species of grass (Poaceae) comprise their diet. Furthermore, families like Typhaceae, Juncaceae, Polygonaceae, Boraginaceae, Cyperaceae, Urticaceae, Lamiaceae and Rosaceae are mentioned (Wilson et al., 2017).

1.5.3 The Helgeland water vole diet

The overall aim of the water vole project is to understand causes and consequences of the extreme population fluctuations of small mammals, which also has been observed in the Helgeland water vole metapopulation: In 2014, the local media reported large water vole population sizes at the island Austbø, which is located near the 13 study islands of the water vole project (Figure 2, Appendix 1). Sampling of the study islands started in 2015 and has shown a net decrease in the population sizes, with a drastic population crash on all islands except Geiterøya N (n: number of individuals = 20) in spring 2018. Furthermore, the subpopulation on that island also declined drastically towards fall, when only two individuals were captured on that island (one female, one probable male).

Food quality may possibly affect demographic rates in wild Arvicoline rodent populations (Moorhouse et al., 2008). Therefore, basic knowledge on the water vole diet could provide a better understanding of the vole-vegetation interaction and its potential role in population regulation. The aquatic water vole in Britain is well studied, but not so much is known about the more fossorial type of vole in the rest of Europe, and especially not in Scandinavia. In this study, I investigated the diet of the water vole (*Arvicola amphibius*) on four islands in the Helgeland archipelago in northern Norway using metabarcoding. As far as I know, this study is the first to map the water vole diet in Scandinavia, and to use eDNA metabarcoding to do it.

Based on the versatile diet of British and Belarusian voles, I expect to find several different families and species of plants in the Helgeland water vole diet – though a bit restricted by the diversity of plants in the study area.. I especially expect species of Poaceae (grass) to be important in their diet.

2 Material and methods

2.1 Study species

The European water vole, *Arvicola amphibius* (formerly *A. terrestris*), is a herbivorous rodent in the subfamily Arvicolinae, along with other voles, lemmings and muskrats. Their fur is chestnut brown to black, they have a rounded body with a blunt nose and small, rounded ears and a short tail covered with fur (Figure 1). They weigh 140-350 grams, with females slightly smaller than males (Strachan et al., 2011).

Figure 1: *Water vole from Gulbrandsøyan S. Photo: Nina Østby*

The water vole has a wide geographic distribution over Britain, Europe and Russia (Strachan et al., 2011). They are burrowing animals, and each vole creates a network of burrows with several food storage chambers, nest chambers and entrances. When out of their burrows, they prefer running in dense vegetation that gives them protection from predators (Strachan et al., 2011). *A. amphibius* is a polytypic species with many subspecies having slightly different morphological and behavioral traits (Wilson et al., 2017). The British water vole belong to a more aquatic morphotype, living near fresh water. Much of the rest of the European water vole is the more fossorial type (living largely underground) and are mainly found in higher altitudes (Strachan et al., 2011). The Norwegian water voles, including those in this study are of the fossorial type.

In the International Union for Conservation of Nature (IUCN) Red List *A. amphibius* is classified as “least concern” (Batsaikhan et al., 2016), and it is not protected in Norway.

2.2 Study site

The study site of the water vole project is located in the Skålvær archipelago at Helgeland, northern Norway (65.879°N, 12.215°E), just south of the Arctic circle (Figure 2). There are 13 small islands included in the project, all uninhabited by humans. These islands are a part of a bigger water vole metapopulation. In this study, the fecal samples originate from four of these 13 islands: Geiterøya N (N = north, 9 692 m^2), Gulbrandsøyan S (S = south, 10 980 m^2), Gullrekka 1 (4 070 m^2) and Gullrekka 2 (2 053 m^2) (Figure 2).

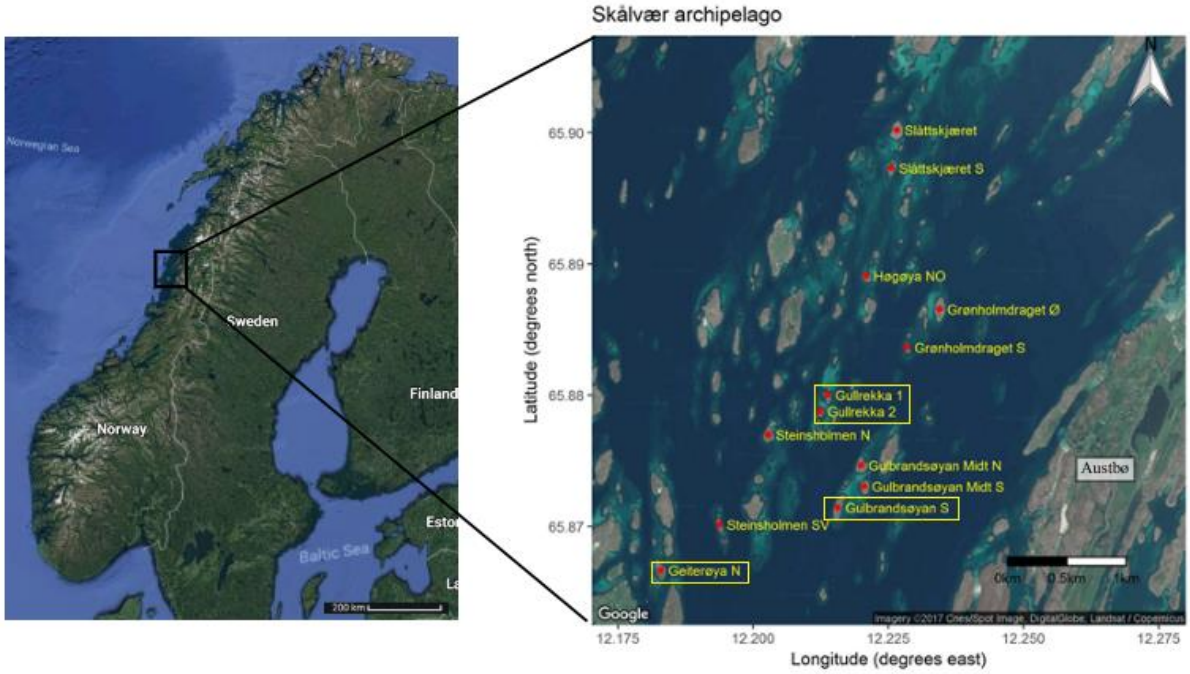


Figure 2: *The Skålvær archipelago on Helgeland with all 13 islands of the water vole project marked with a red dot and the four study islands in this study marked with yellow rectangles.*

All four study islands have very few, small or no trees. Along the sea and a few meters up, the islands have a rocky ground with no vegetation, uninhabitable for water voles. Other than that, the four study islands represent two quite different types of islands in terms of size and vegetation. On the two bigger islands, Geiterøya N and Gulbrandsøyan S, the vegetation is mainly coastal heath (from damp heath to dry, herb rich) and semi-natural grasslands (see Appendix 1). Shrubs of common juniper (*Juniperus communis*) are also quite common on these islands. The terrain is mostly even, with some ponds and trenches.

The other type of islands, Gullrekka 1 and Gullrekka 2, are smaller, so-called "bird manured islands" (Fremstad, 1997) where nesting birds (e.g. European Herring Gull, *Larus argentatus* and common gull, *Larus canus*) have fertilized the soil. Also, occasional run-up of seaweed during high tide brings up nutrients to the soil. The high levels of phosphorus and nitrogen affect the vegetation, which is dominated by herbs and grasses and few or one species may dominate (Fremstad, 1997). In a vegetation survey on similar small islands approximately seven km south west of the study area, Carlsen et al. (2011) found that plant species like *Rhodiola rosea, Filipendula ulmaria, Tripleurospermum maritimum, Angelica archangelica* and *Valeriana sambucifolia* have outcompeted species of Ericaceae on such islands.

Due to difference in the vegetation , Geiterøya N and Gulblandsøyan S are referred to as "heath/grassland islands", and Gullrekka 1 and 2 as "bird manured islands" in this study.

On all islands, the short-eared owl (*Asio flammeus*) and the common kestrel (*Falco tinnunculus*) are the most common predators. Juvenile water voles may also risk predation from black-backed gull (*Larus marinus*) and grey heron (*Ardea cinerea*).

2.3 Sampling methods

The sampling of the water voles was performed in spring, summer and autumn. For the capturing, Sherman XLF15 folding traps were used. The number of traps used per islands depend on island size and ranged from 80 traps for the smallest islands to 170 for the largest. Exact trap sites were determined by burrow openings, latrines and other signs of vole presence. The entrance of the traps were put as near such signs as possible without blocking their lanes or burrow entrances. Each trap was covered by moss and grass to stabilize them in case of wind, and to prevent overheating from the sun (Appendix 2). The traps were filled with chunks of carrot and potato as food and some dry grass serving as bedding material for the voles while captured. Furthermore, some small chunks of carrot were scattered around the entrance. The traps were checked every 1,5-2 hours.

When captured, sex, age and reproductive state and body mass were determined. Sex was determined by investigating reproductive organs (Stoddart, 1971). Since the sex of young, sexually immature individuals was hard to determine, there were four classes of sex: male (m), female (f), probably male (pm) and probably female (pf). For simplicity, all pm and pf are treated as m and f in this study, respectively. Age (juvenile/adult) was determined based on body mass (Stoddart, 1971), where voles weighing <110g were considered to be juveniles. Sexually active females have a vaginal opening and everted nipples, while in sexually active males, the testes can be felt in the scrotum and the penis tip is easily protruded. In some cases where these cues were not very obvious, the categories “probably active”, “probably not active” and “unknown” were used.

Unique PIT-tags (TROVAN unique ID-100B; 11.5 x 2.12 mm.) were inserted in the back using an IM-200 syringe implanter. For practical and ethical reasons, voles weighing less than 45 grams were not tagged. A biopsy sample was taken from the ear. Four vole dropping pellets were collected from the trap the respective vole was captured in, and immediately wrapped in a filter paper. The filter paper was then put in a Corning 15 mL centrifuge tube with approximately 10 mL silica gel to dry the samples to slow down further degradation. After handling, the voles were released in the same area they were captured. The traps were

cleaned, refilled with bait and placed out again in the same position. When recaptured, the voles were only registered using a Dorset LID575 tag reader and released. The fecal samples were frozen the same day and later brought to the lab at Vitenskapsmuseet in Trondheim for analysis.

Retention time of the digestive tract of similar species is 3.3-4.8 hours for the field vole *Mictrotus Agrestis* (Hagen et al., 2018) and > 5 hours for 100 g herbivorous hindgut fermenters (Sakaguchi, 2003). This suggests that water voles have a retention time that is longer than the maximum time between capture and collection of the fecal sample. Thus, I expect contamination of diet from bait and bedding material inside the trap to be negligible. However, it is not unlikely that some water voles may have eaten bait outside the trap a long enough time before capture so that carrot may be detected in the diet analysis.

2.3.1 The diet analysis sample collection

The 48 fecal samples analyzed in this study were sampled in mid-April, mid-July and the beginning of September in 2018. The main goal when putting together the sample collection was to pick a set of samples that together represent the overall water vole diet at metapopulation level. Since we may expect sexual, spatial and temporal variation in the diet, the sample collection is designed to capture as much of this variation as possible. They are therefore collected at different seasons, different sites (with different vegetation types) and evenly between sexes. However, in the study year the population had crashed, which prevented a perfectly balanced sampling. All 13 islands are sampled during spring and autumn, and only four islands (Gulbrandsøyan Midt S, Gulbrandsøyan Midt N, Gulbrandsøyan S and Geiterøya N) are sampled in the summer season. On three of these islands, zero voles were captured in the summer session (Gulbrandsøyan S was not visited due to nesting Arctic tern, *Sterna paradisaea*). Thus, Geiterøya N was the only island with samples from all three seasons, though with only two samples from September. Samples from this season were supplemented from Gulbrandsøyan S which should be comparable to Geiterøya N in terms of vegetation. In addition to the big islands, a few samples from the small islands (Gullrekka 1 and 2) were included as these nutrient-rich islands represent a bit different vegetation composition. See Appendix 3 for a complete list of the samples with capture site, month and sex.

2.4 DNA extraction, amplification and sequencing

The DNA extraction on the 48 samples was performed using a Qiagen DNeasy PowerSoil DNA extraction kit according to the manufacturers' protocol. No blanks were included in the DNA extraction. In all laboratory steps following DNA extraction, positive control samples were included to ensure the protocol worked. The positive controls were DNA extracts from another experiment; each contained DNA from a single vascular plant of known taxonomic identity. A quantitative PCR (qPCR) analysis was run to find the optimal number of cycles for the first amplification PCR. For each extract, the short and variable P6 loop region of the trnL (UAA) intron (Sjögren et al., 2017) was amplified using a modified version of the primer set which is universal for vascular plants. The complementary primers trnL-g-Fus (5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GGG CAA TCC TGA GCC AA-3') and trnL-h-Fus (5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GCC ATT GAG TCT CTG CAC CTA TC-3') were used as forward and reverse primer, respectively.

3 uL extracted eDNA was used as template in the amplification PCR in total volumes of 25 µL, containing PCR Buffer II (final conc. 1X), $MgCl_2$ (final conc. 2.5 mM), dNTP mix (final conc. 0.2 mM), primer trnL-g-Fus (final conc. 0.6 µM), primer trnL-h-Fus (final conc. 0.6 µM), BSA (final conc. 0.5 mg/mL), AmpliTaq Gold polymerase (final conc. 0.04 U/µL) and molecular biology water. The cycling protocol was the same as for the qPCR, only with 25 cycles. After the amplification cycles, it finished with 72°C for 5 min, before storing at 4°C. Three negative controls (H_2O, no DNA) and four positive controls (plant DNA extracts) were used during the PCR and processed in parallel throughout the remainder of the laboratory work.

The quality of PCR amplifications was controlled using gel electrophoresis. The amplified regions were purified with SPRI beads (Rohland and Reich, 2012), and then tagged with unique double index combinations with an Illumina Nextera XT DNA Library Preparation kit. The indexing PCR was carried out in 25-µl volumes containing: 2.5 µL template DNA, 2.5 µl Nextera index primer 1 (N7xx), 2.5 µl Nextera index primer 2 (S5xx), 12.5 µl 2X KAPA HiFi HotStart Ready Mix and mol. biol. water. The index PCR protocol was: 95°C for 3 min; 17 cycles of 95°C for 30 secs, 55°C for 30 secs and 72°C for 30 secs; before final extension with 72°C for 5 mins. The indexed libraries were then purified into 30 µl Qiagen EB buffer using a Qiagen QIAQuick PCR Purification Kit. The concentration of each sample was measured with a Qubit 2.0 fluorometer, and equal masses of each sample were combined into a

sequencing pool with 50 ng DNA from each sample and 28.8 ng of each of the negative and positive controls. Finally, 25% PhiX was added to the pool to increase complexity. Using a MiniSeq MO 300 cycles kit, the pooled sample was sequenced on an Illumina MiniSeq instrument.

2.5 Data analysis

The next-generation sequence data from the sequencer were analyzed and filtered using OBITools version 1.2.12 software package (Boyer et al., 2016; http://metabarcoding.org/obitools/doc/index.html). The OBITools data handling and processing was done following Sjögren et al. (2017). Briefly, for each sample the raw sequence read pairs were aligned/merged, and unaligned read pairs were removed. The sequence data for the 55 samples were then combined and reduced to unique sequences (*obiuniq* tool) that were counted using the *obiannotate* and *obistat* tools. Grouping identical sequences makes the taxonomy identification more efficient by avoiding identification of the same sequence several times. Furthermore, groups of very similar sequences (e.g. one different base or slight difference in sequence length due to sequencing/PCR errors or intraspecific variability) were clustered (denoising) as described in Taberlet et al. (2018), p. 70-71. The data was filtered again, keeping only sequences less than 151 base pairs (bp) and 10 or more occurrences. The length of the trnL intron P6 loop amplified with primers *g* and *h* is 10-143 bp (Taberlet et al., 2007). The choice of count threshold is based on other metabarcoding diet and ancient DNA studies such as Sjögren et al. (2017) and Alsos et al. (2016) who also set the count threshold on 10 counts or more per sequence.

The trnL P6 loop reference sequence database was constructed with OBITools using sequences from 835 of the most common northern boreal vascular plant species (Willerslev et al., 2014) as well as 815 arctic plant vascular species (Sønstebø et al., 2010). In addition, the baitspecies used as trap bait, potato (*Solanum tuberosum*) and carrot (*Daucus carota*), were added to the database as a control measure. Sequences generated in this study were then assigned to taxa using the *ecotag* tool in OBITools. 17 different sequences accounted for 90% of the unidentified sequence counts (no match to the reference database) and were subjected to a BLAST search using the software Geneious prime version 11.1.5. Seven out of these 17 sequences (55% by sequence count) were identified as GenBank sequences for near-perfect matches (100% identity over 100% of the query length). The seven sequences were retrieved and added to the reference database. The vole diet study sequences were then again assigned to taxa using the *ecotag* tool. After the second run, all sequences still unidentified were

merged to make one single sequence count. The NCBI taxonomy database was retrieved on 12.02.2019.

When identified, each sequence or cluster of sequences were allocated to a taxonomic rank (e.g. family, genus, species). A sequence only matching a single species were given the rank "species". A sequence or sequence cluster matching two or more species from the same genus (e.g. *Carex aquatilis* and *Carex atrata*) was given the rank "genus" and so on. The merged unidentified sequences were marked "no rank". In addition to these, two other unique sequences were identified as Campanulids, but were assigned "no rank". The rank was manually changed to "superfamily clade" (Campalunids is a clade within the clade Asterids (Byng et al., 2018). The two sequences are one or two of the species *Gnaphalium norvegicum, Pilosella lactucella, Sambucus racemose, Leontodon hispidus* and/or *Menyanthes trifoliata*.

Further data processing was performed in R using RStudio version 1.1.456 (R Core Team, 2018). The positive control samples were therefore removed from the data before further analysis. The sequence counts in negative controls were used to roughly correct the counts of each plant sequence in each vole sample by subtracting by a correction factor. The corrected count *F* for each sequence in each vole sample *n* was calculated according to the following equation:

$$F = A_n - \frac{\sum_{t=1}^{3} C_t}{3} * \frac{K_N}{K_n} * \frac{\sum_{i=1}^{x} A_{n,i}}{\sum_{t=1}^{3}(\sum_{i=1}^{x} C_{t,i})/3}, \quad (1)$$

where A_n is the observed sequence count for vole sample *n*, C_t is the observed sequence count for negative control sample *t*, $A_{n,i}$ is the observed sequence count for unique sequence cluster *i* in vole sample *n*, $C_{t,i}$ is the observed sequence count for unique sequence cluster *i* in negative control sample *t*, K_N is the mean post-PCR concentration of the three negative controls, K_n is the post-PCR concentration of vole sample *n*, and *x* is the total number of unique sequence clusters defined by OBITools after denoising. Because it is a count, a minimum value of $F = 0$ was enforced.

After correcting the counts for all samples, all sequence counts comprising the bottom 5% of each sample were removed. The justification for this procedure is that it simplifies visualization of results and that these sequences account for such a small fraction of the total diet that they are uninteresting in the context of diet analysis. Many of them also had a count of 0 after the correction. These removed families are listed in Appendix 4.

Plots were created using the ggplot2 version 3.1.0 package in R (Wickham, 2016). Since only 54% of the sequences are determined to species level, most of the results are presented on family level to give a good overview of the water voles' diet. Since the unidentified sequences consist of many different, non-similar sequences, they were excluded from most plots as they are not comparable to the count of the clustered identical or nearly identical sequences. Overall diet composition is presented with boxplot on family level, species level and on family level grouped by site. Individual proportions of different plant families in each sample was calculated and visualized with a heatmap. Mean diet proportions per plant family and species were calculated overall (Soininen, 2015) and within groups defined by site, by calculating the mean of the proportions in all samples. In addition to the proportion of each cluster/sequence in the samples, the number of samples each family/species was found in is also presented (Appendix 6). A family or species was considered as present in a sample if it had a sequence count > 0 in the sample. The last method puts less weight on the relative abundance of each sequence but gives valuable information about the diet on population level. Nevertheless, since herbivores often eat a variety of different plants, the two methods often yield similar results (Taberlet et al., 2018). In all plots and tables, families and species are grouped by relatedness (see Appendix 5).

Each species identified through the genetic analysis was checked for likeliness of occurrence at the study site using Global Biodiversity Information Facility – GBIF (GBIF.org, 2019) and the Norwegian flora by Lid and Lid (2013). The species were also checked against a species list from a vegetation survey report containing observation data on 101 vascular plants (not all of them found) from different islands in the same study area (Thorvaldsen et al., 2019).

3 Results

3.1 Taxonomic resolution of the sequence data

The total number of sequences obtained from OBITools after concatenating and aligning the sequences was 5 473 774, grouped into 171 clusters of identical or similar sequences. After accounting for negative controls and filtering out the 5% lowest sequence count for each sample, we were left with 39 clusters/ with a total sequence count of 4 968 401. 88% of the sequences were determined to family level, 55% to genus level and 54% to species level. 12% of the sequences were not identified to any rank (Figure 3). Those sequences are a cluster of all sequences, similar or not, that had no match to the databases.

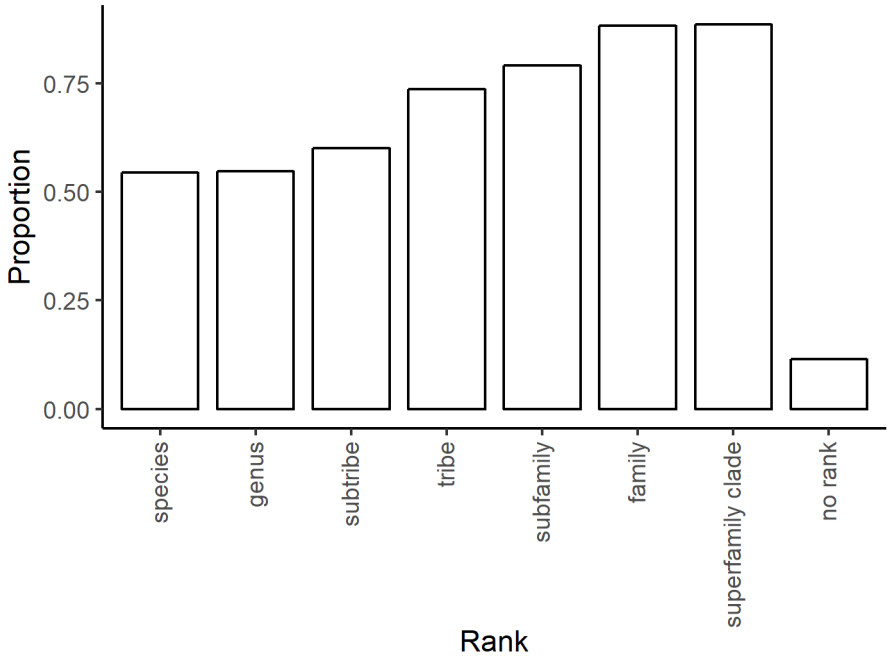


Figure 3: *Taxonomic rank of the 95% most abundant sequences. The proportion increases with higher taxonomic levels as sequences determined to a lower level also are determined at higher levels.*

3.2 Overall water vole diet

After filtering and removing the lowest counts in each sample, 17 plant families remained (Figure 4). When calculating the proportions of the different families in the total diet by sequence count, 90% of the total diet consisted of only four plant families (Poaceae, Rosaceae, Fabaceae, Cyperaceae) together with the unidentified sequences. The by far most abundant plant family, Poaceae, was found in all 48 samples (Figure 6 and Appendix 6) and accounted for 57% of the total diet (mean proportion). Rosaceae was second, detected in 31

samples, with a mean proportion of 14%, followed by Fabaceae and Cyperaceae comprising 5% and 4% of the diet, respectively. Apiaceae and Juncaceae comprised 3% and 1%, and each of the remaining plant families accounted for < 1% (Figure 4).

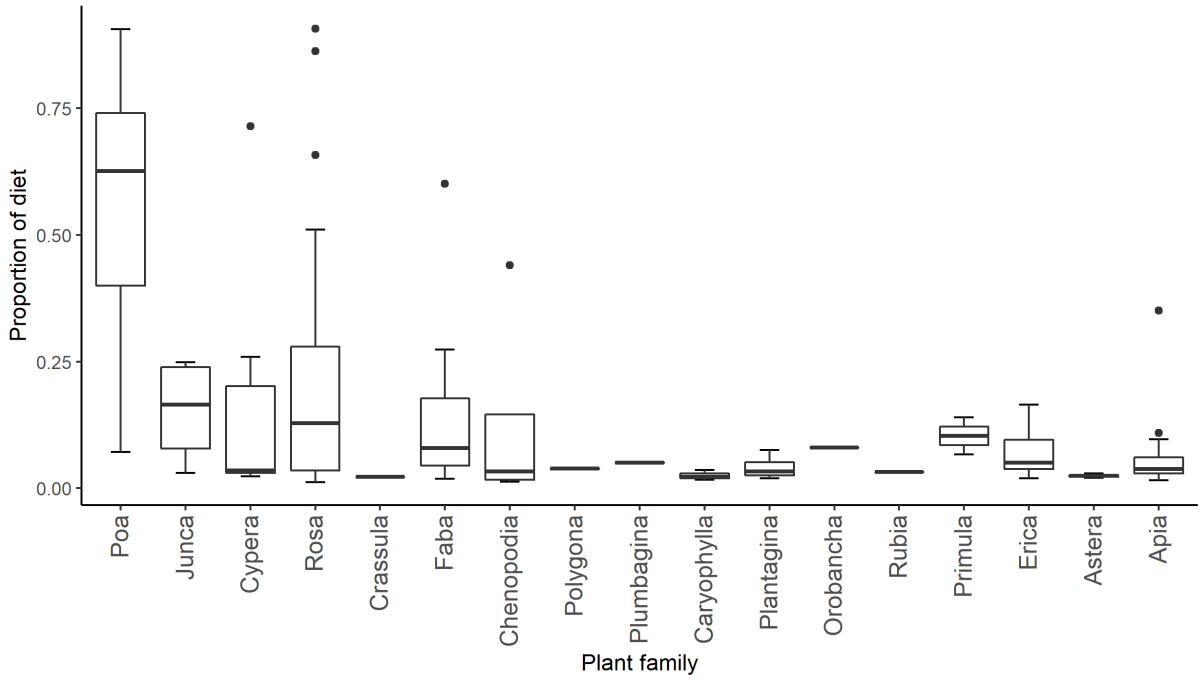


Figure 4: *Composition of plant families in water vole diet (n=48) based on sequence count per sample. Unidentified sequences were found in 39 samples and accounted for 12% of the diet (not shown on graph). The suffix "ceae" has been removed from all family names. Families are ordered by relatedness (Appendix 5). The black horizontal line is the median of the value of all samples, and the box contains the middle 50% of the samples: 25% above (upper quartile), and 25% (lower quartile) below the median. The whiskers represent values outside the middle 50%, and the points are outliers.*

36 different species were found in the total diet, and 28 of these were identified to a specific species (Figure 5). On average, each sample contained 6.2 species, with an average of 3.5 species of Poaceae per sample. The two most abundant species, Unknown species 1 and *A. pubescens*, both belong to the Poaceae family. They comprise 27% and 16% of the total diet, respectively. On average, 75% of all Rosaceae in the whole diet is *F. ulmaria*, which is the third most abundant diet species, comprising 11% of the total diet. All the remaining species each comprise < 5% of the total diet.

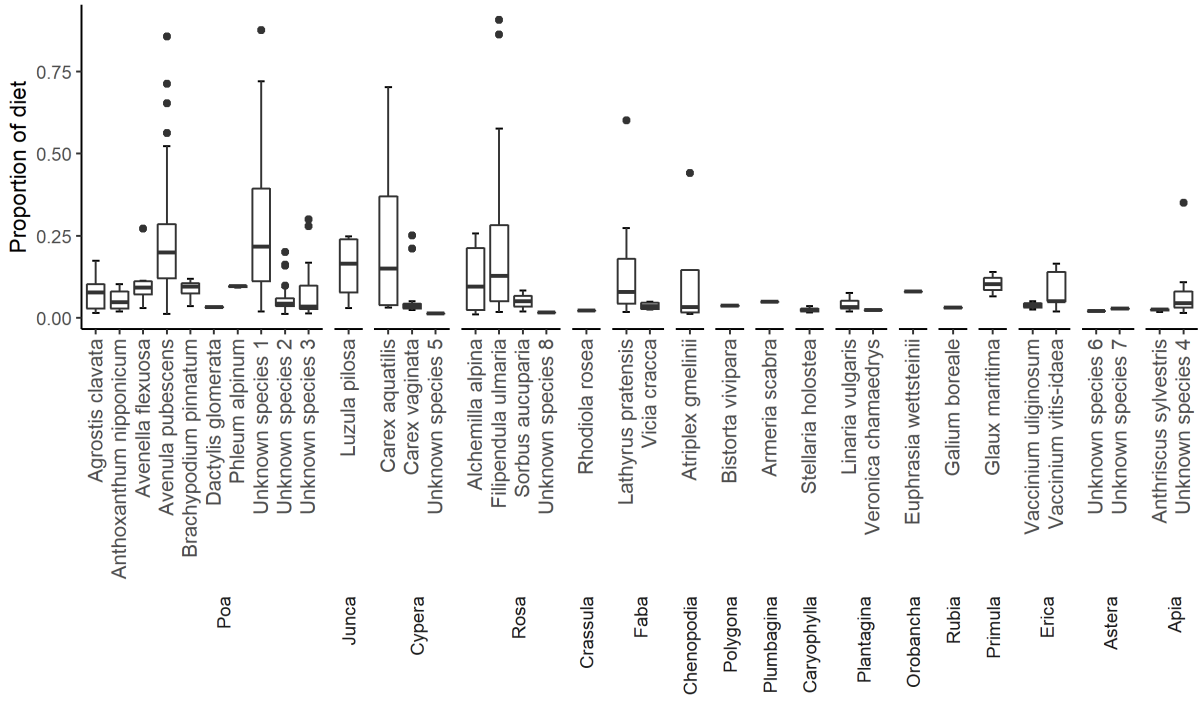


Figure 5: *Proportion of plant species detected in water vole fecal samples (n = 48). Unknown sequences are single sequences with more than one match to the reference database (Table 1). The species are grouped by family and the families are ordered by relatedness (Appendix 5). The suffix "ceae" has been removed from all family names. The black horizontal line is the median of the value of all samples, and the box contains the middle 50% of the samples: 25% above (upper quartile), and 25% (lower quartile) below the median. The whiskers represent values outside the middle 50%, and the points are outliers..*

Eight unique sequences matched two or more taxa at the species level and thus were assigned to higher taxonomic levels. A list of the species matches for these sequences and their common higher taxonomic levels was provided by OBITools (Table 1). Note that these "unknown sequences" are not the same as the "unidentified sequences". They are unknown because they had several matches in the database, and thus it is unknown which one of them they are.

Unknown species 4 was found in 17 of the 48 samples (Appendix 6) and could be either of the two species *Pimpenella saxifraga* or *Daucus carota*. The sequence accounted for 2% of the total diet by sequence count. The common carrot used in and around the traps is *D. carota subsp. sativus*. Low sequence counts of potato (*Solanum tuberosum*) was found in some of the samples after sequence filtering in OBITools, but these were removed with the lowest 5% counts of each sample.

Table 1: *Taxonomic information on the Unknown species in Figure 5. The species are unknown at species level because they had multiple multiple best matches to the reference database. The matches for Unknown species 5 are all* Carex *species, thus it is identified at genus level. All the other unknown species belong to different genus but same family and are thus identified at family level. The latter are the sequences represented in the 34% rise in taxonomic resolution between genus and family level in Figure 3. The most likely of the suggested species to occur at the study site (checked with GBIF.org (2019), Thorvaldsen et al. (2019) and Lid and Lid (2005)) are marked with bold text.*

Unknown species	**Family**	**Genus**	**Species suggestions**
Unknown species 1	Poaceae	*Alopecurus* *Phalaris*	***Alopecurus geniculatus*** ***Phalaris arundinacea***
Unknown species 2	Poaceae	*Festuca* *Helictochloa*	*Festuca viviparoidea* *Helictochloa hookeri*
Unknown species 3	Poaceae	*Agrostis* *Calamagrostis*	***Agrostis clavata*** *Calamagrostis deschampsioides*
Unknown species 4	Apiaceae	*Daucus* *Pimpenella*	***Daucus carota*** ***Pimpenella saxifraga***
Unknown species 5	Cyperaceae	*Carex*	***C. atrata, C. krausei, C. aquatilis, C. pallescens,*** *C. stylosa, C. microchaeta, C. alba, C. pediformis, C. supina*
Unknown species 6	Asteraceae	*Anaphalis* *Crepsis* *Grindelia*	***Anaphalis margaritacea*** ***Crepis paludosa*** *Crepis nana* *Crepis chrysantha* *Grindelia squarrosa*
Unknown species 7	Asteraceae	*Anaphalis* *Crepsis*	***Anaphalis margaritacea*** ***Crepis paludosa*** *Crepis nana* *Crepis chrysantha*
Unknown species 8	Rosaceae	*Comarum* *Rosa* *Rubus*	***Comarum palustre*** ***Rosa mollis*** ***Rubus arcticus***

There is a clear pattern with Poaceae being the most abundant plant family (mean proportion: 57%, range: 7-91%), followed by Rosaceae (Mean proportion: 14%, range 0-91%) (Figure 4 and 6). Many of the samples with a low proportion of Poaceae have a high proportion of Rosaceae, and this is especially observed in some samples from Gullrekka 2 (Figure 6). The

unidentified sequences are also among the most abundant in some samples. Note that they are not directly comparable to the identified families as they represent a lot more sequences than each family does.

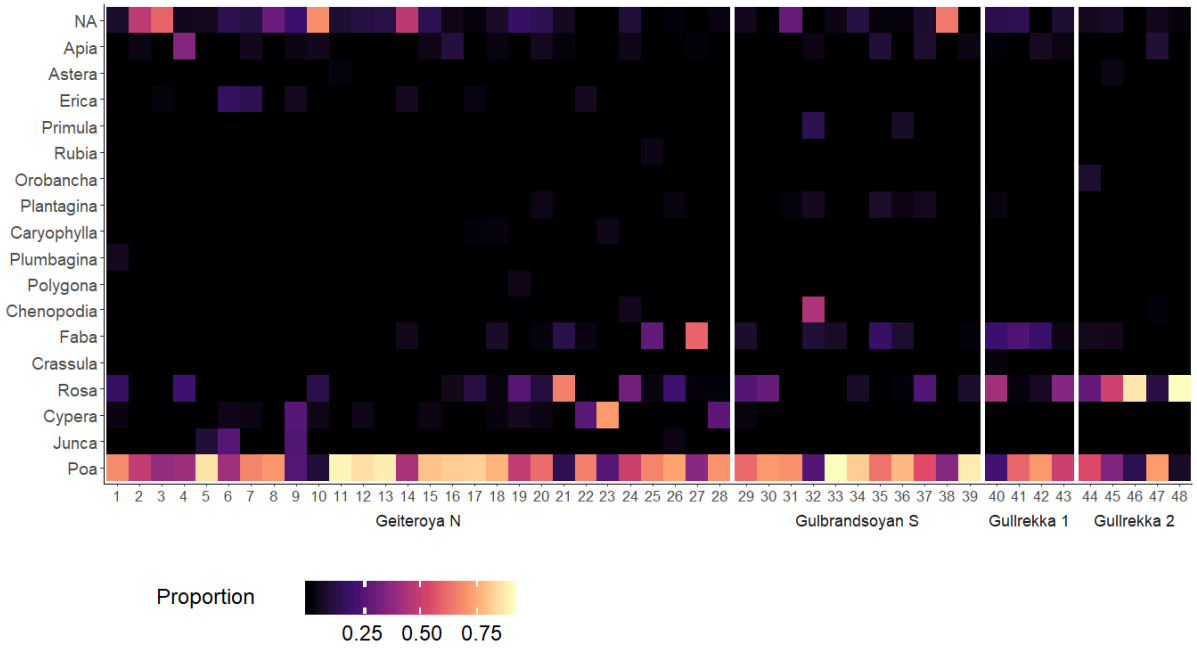


Figure 6: *Proportion of plant families detected in 48 water vole fecal samples grouped by site. The samples are also ordered by the month they were collected: April (1-16), July (17-26), September (27-48). Proportions are calculated within each sample from the sequence count of each family. Unidentified sequences are included to make the proportions sum up to 1 in each sample. The suffix "ceae" has been removed from all family names.*

3.3 Diet variation between sites

Geiterøya N has the most sampled individuals (n = 28) and also the most plant families represented. A few of the families only or nearly only occur on Geiterøya N. For instance, all four samples with Juncaceae detected in the diet originate from Geiterøya N (Figure 7). This family was only represented by one species, *Luzula pilosa* (Figure 5, Appendix 6). Another example is Cyperecae which occurs in 14 samples in the overall diet (Appendix 6), and 13 of these are from Geiterøya N with an average diet proportion of 6%. The last sample is from Gulbrandsøyan S and accounts for only 0.2% of the diet of the 11 voles on that island.

Rosaceae is found in the diet of the voles on all four islands, however a lot more on Gullrekka 2 than the other islands. Gullrekka 2 even has a larger proportion of Rosaceae (54%) than Poaceae (36%). The samples from Gullrekka 1 also has a higher proportion of Rosaceae (21%) compared to Geiterøya N (14%) and Gulbrandsøyan S (16%). In all 48 samples, there

are only four species of Rosaceae; *F. ulmaria, Alchemilla alpina, Sorbus aucuparia* and Unknown species 7. *F. ulmaria* comprise 86% of the Rosaceae in the diet of the five voles on Gullrekka 2. The remaining 14% is *A. alpina*, found in only one of the samples. *F. ulmaria* accounts for all the Rosaceae found in the samples on Gullrekka 1. When comparing the island types ("heath/grassland" vs. "bird manured"), the mean proportion of *F. ulmaria* on the "heath/grassland islands" is much lower (15%) than on the "bird manured islands" (36%).

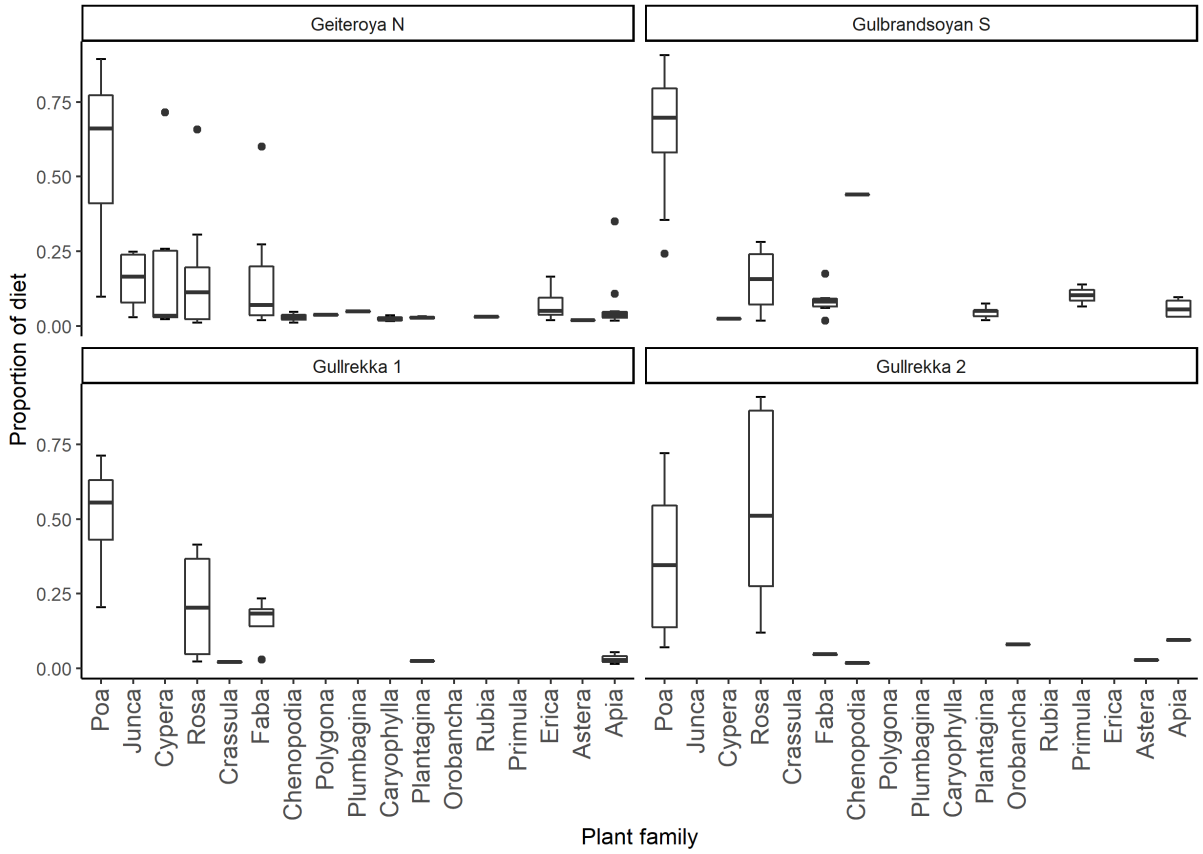


Figure 7: *Compostion of water vole diet on each of the four islands Geiterøya N (n=28), Gulbrandsøyan S (n=11), Gullrekka 1 (n=4) and Gullrekka 2 (n=5). The islands represent two different island types – "heath/grassland islands": Geiterøya N and Gulbrandsøyan S and "bird manured islands": Gullrekka 1 and 2. The suffix "ceae" has been removed from all family names. The black horizontal line is the median of the value of all samples, and the box contains the middle 50% of the samples: 25% above (upper quartile), and 25% (lower quartile) below the median. The whiskers represent values outside the middle 50%, and the points are outliers.*

4 Discussion

4.1 The water vole diet

4.1.1 Taxonomic resolution

Compared to other studies, the taxonomic resolution obtained in this study (Figure 3) is quite good at genus (55%) and species level (54%). From the trnL marker alone (without the family-specific ITS-markers they used), Lopes et al. (2015) only identified 20% and 23% of the sequences to genus and species level, respectively. Soininen et al. (2015) also had much lower resolution on genus and species level – 33% and 8% of the sequences, respectively, despite using both trnL g-h and c-h. On the other hand, they had a much higher resolution on family level; ≈ 100%. The reason for their poor resolution on lower levels is due to a plant family abundant in the lemming diet, Salicaceae, which generally has a low resolution for the g-h primer (Sønstebø et al., 2010). With this family excluded from the data, the genus and species resolution in Soininen et al. (2015) increased to 72% and 17%, however still much poorer on species level than in this study. A possible explanation for the good resolution on species level in this study is that the few species that exist on the study islands happen to be well represented in the reference databases used in the analysis.

4.1.2 Overall diet

It is clear from the results of this diet analysis that species of grass (Poaceae) are very important in the water vole diet at the four study islands at Helgeland (Figure 4 and 6). This was indeed expected based on the diet of the British and Belarusian water voles (Ashby and Vincent, 1976; Neyland, 2011; Strachan et al., 2011; Wilson et al., 2017). The proportion of Poaceae in the spring diet (60%) of the Helgeland water voles is approximately similar to that found in the diet of the British water voles (65%) studied by Ashby and Vincent (1976). The proportions are lower than that of this British vole population in summer (56% vs. 90%) and autumn (55% vs. 90%), but still high in both diets. Se Appendix 7 for a comparison of diet between seasons in the Helgeland water vole diet. It is common practice of water voles to collect food items and bring them to “feeding stations” where they are eaten (Neyland, 2011). The Welsh water vole diet was based on surveying what species of plants were found in such feeding stations. A species of Poaceae was the second most observed species in these feeding stations, and so Poaceae seems to be important food to water voles in all studies mentioned.

The proportion of Poaceae in the Helgeland water vole diet is also similar to the proportions found in the metabarcoding diet studies on other rodents. In the diet analysis of the two

subterranean rodents *C. minutus* and *C. flamaroini*, Lopes et al. (2015) detected at least one species of the Poaceae family in 66 of the 67 specimens of *C. minutus*, and in 97 of the 98 of the specimens of *C. flamaroini.* This is very similar to the findings in this study, where 48 out of the 48 vole feces samples contained Poaceae (Appendix 6). The overall proportion of the Poaceae family in this study is also similar to that found by Lopes et al. (2015), where Poaceae comprised 52% (*C. minutus*) and 66% (*C. flamaroini*) of the two rodent diets. Comparatore et al. (1995) also found two other rodents of the genus *Ctenomys* to prefer grass over forbs. Similar results are also reported by Khanam et al. (2016), who conducted a diet analysis study on three rodent species (*Suncus murinus, Rattus rattus* and *Mus musculus castaneus)* using DNA barcoding. Here, Poaceae was the most common plant family in all three diets, with a proportion of 64% (*S. murinus*), 58% (*M. musculus castaneus*) and 29% (*R. rattus*) (Khanam et al., 2016). Note that these proportions are shared with other taxa of birds and invertebrates, so the proportions would be somewhat higher if only analyzed for plants as in this study. Also Soininen et al. (2013) found grass to be the dominating food item for lemmings (*L. lemmus*), comprising 49% of the total diet. Judging from the similarity to other voles and rodents, the high proportion of Poaceae detected in the Helgeland water vole diet is very believable.

Rosaceae is the second most common plant species in the water vole diet in this study (mean proportion 14%). *F. ulmaria* represent most (75%) of all Rosaceae in the diet and is also the third most eaten species overall (mean prop. 11%). Rosaceae is not mentioned in either of the British/Welsh vole diets (Ashby and Vincent (1976); Neyland (2011)), even though plants of the family (e.g. *F. ulmaria*) are common across Britain (GBIF.org, 2019). *F. ulmaria* is mentioned as a part of the European water vole diet (Wilson et al., 2017). Furthermore, the plant species is common on the study islands (Appendix 1) and is thus not unlikely to be a part of the Helgeland water vole diet as well. However, in a case study on the impact of water voles on the agricultural landscape in the same area (e.g. on Austbø, see Figure 2 and Appendix 1), Thorvaldsen et al. (2019) suggest that *F. ulmaria* is *not* eaten or otherwise destroyed by voles, and that there may even be a positive correlation between the abundance of *F. ulmaria* and population sizes of water voles: The plant can grow freely without being eaten or otherwise disturbed by voles, and in return it provides good shelter from aerial predators (Thorvaldsen et al., 2019). Note that Thorvaldsen et al. (2019) did not conduct a diet analysis on the voles, and this assumption is based on a general impression of the correspondence between voles and this plant. They also took a brief check for signs of voles

chewing on the roots of *F. ulmaria* in an area with high density of voles, where no signs were found. When asking farmers at Austbø about these suggestions, they share the impression that *F. ulmaria* is left alone by water voles, and some even believe there is *more* of the plant in periods with high vole density, which match well with the assumptions of Thorvaldsen et al. (2019). The explanation for this may be found in the difference between island types like that of Austbø (here considered as more similar to the "heath/grassland islands" Geiterøya N and Gulbrandsøyan S) and the "bird manured islands" Gullrekka 1 and 2. See section 4.1.3 Variation in diet between sites.

Rosaceae does not seem to be common in the diets of the rodents investigated in Lopes et al. (2015), Khanam et al. (2016) and Soininen et al. (2013). Rosaceae only accounts for 2% of the *L. lemmus* diet (Soininen et al., 2013) and is only represented by one species; *F. ulmaria*. Rosaceae is not mentioned in neither of the studies by Lopes et al. (2015), nor Khanam et al. (2016). This does not mean, however, that these rodents avoid eating species of the Rosaceae family, as it could be that these kinds of species are not available to them. Lopes et al. (2015) conducted their study in Brazil, and Khanam et al. (2016) in Pakistan. Both of these study sites have a climate very different from Helgeland, and consequently, the plant community the rodents live in will be different. The lemming fecal samples in Soininen et al. (2013) were collected in Varanger in northern Norway, so it is the most comparable study to this in terms of climate and plant community similarities. For better comparing diets between different species and different sites, a selectivity coefficient should be calculated for all plant species by dividing the relative proportion of the plant species in the diet by the relative proportion of the plant species at the site, as is done in Soininen et al. (2015).

Some Juncaceae was detected in the Helgeland vole diet, however only in four samples and only represented by one species (*L. pilosa*). A species of the same family was the most important diet component of the Welsh vole, and was found to have a high nitrogen to water ratio (yielding more nitrogen per gram eaten than other plants in the voles' diet) (Neyland, 2011). Whether or not the *L. Pilosa* on Helgeland provide the voles with much nitrogen is not known, but if it does (and thus should be a preferred species), the low contribution of Juncaceae could be due to low abundance on the study islands. As seen in Figure 7, Juncaceae was only detected in the diet on Geiterøya N, which presumably has a higher plant diversity. If the species is a preferred food item and abundance is not limited, much more of the species would be detected in the diet. Comparing plant nutrient composition and vole preferences could be a topic for future studies.

A. sylvestris was detected only in three samples (Appendix 6) and comprised less than 0.2% of the total diet. Thorvaldsen et al. (2019) found traces of voles digging up the thick roots of *A. sylvestris, Urtica dioica,* and *Rumex longifolius*. These are species with high root mass that the voles collect during summer and autumn to store in their burrows for use in the winter (Thorvaldsen et al., 2019). The farmers on Austbø claim to see a negative correlation between the abundance of water voles and *A. sylvestris.* If assuming that this plant is just as abundant on the study islands as on Austbø, it could be that voles mainly collect and don't eat *A. sylvestris* to store it for winter use. In that case, analyzing samples from winter season would be interesting.

4.1.3 Variation in diet between sites

Since Geiterøya N is the biggest of the four islands and has the most samples, it is not surprising that this site also has the most variation in plant families occurring in the water vole diet. A bigger island with more possibilities for different micro-environments may give more room for variation in niches and more biodiversity. The species Cyperaceae, for instance, was only found in one vole outside Geiterøya N. This could both be due to the three other islands comprising less than half of the sample collection (20/48), or simply because Cyperaceae is scarce on these islands, both of which gives less probability for any of the voles to have encountered Cyperaceae.

F. ulmaria was detected in 21 samples (Appendix 6), so it is indeed not correct that water voles don't eat *F. ulmaria* (Figure 5) as was assumed by Thorvaldsen et al. (2019) and the farmers on Austbø. Nevertheless, it may be correct that this species is not eaten at Austbø and other islands with high plant diversity – at least not enough to make a noticeable difference in the abundance of the plant. The explanation for this could be found in the difference in the study islands. The proportion of *F.ulmaria* eaten on the two bigger "heath/grassland islands" in this study, Geiterøya N and Gulbrandsøyan S are lower (mean 15%) than that of the "bird manured islands" Gullrekka 1 and 2 (mean 36%) (Figure 7). *F. ulmaria* is one of the species that thrive on small "bird manured islands" rich in phosphorus and nitrogen (Carlsen et al., 2011). Since they also tend to dominate the areas they grow in (Thorvaldsen et al., 2019), the voles may eat this species simply because there is more of it available, and less available of other species. On Austbø, where there is a lot of other food items available, there may be no need for the vole to eat *F. ulmaria*. Thus, the diet of the water vole may just reflect the composition of the food species available to them (they are generalists). It should also be

noted that 10% is a small portion of the diet and may not have a noticeable impact on the abundance of *F. ulmaria*.

Nevertheless, it must be kept in mind that the sample sizes on Gullrekka 1 and 2 ($n_{tot} = 9$) could be poor representatives of the general diet of voles on small islands. There are only two samples (both from Gullrekka 2) that contain a very large proportion of *F. ulmaria* (86% and 93%). These voles may live in a patch dominated by *F. ulmaria*, which makes it the most convenient food source. The unusual high proportions could also be a one-time-event where the whole sample only represent one single meal. It could also result from PCR bias, in which case PCR replicates would be useful.

Furthermore, the ratio between sexes on the two islands is very uneven and could have been a potential confounding factor. However, all individuals on Gullrekka 1 are females and four out of five on Gullrekka 2 are males, so the ratio between the sexes within the island type ("bird manured island") is even. Since the two islands still show the same trend of high proportions of Rosaceae compared to the "heath/grassland islands", the high proportion of Rosaceae consumed on these two islands is thus likely to be caused by the similarity of the sites and not the composition of the samples in terms of sex. See Appendix 8 for diet comparison between sexes. It is also very important to note that these two islands are in immediate vicinity of each other (< 40m) and thus may have very similar environments. It cannot be said with certainty that two distant, small "bird manured islands" would show the same results. Furthermore, all the samples from these two small islands are from September, which may also be a confounding factor if there are seasonal differences in the preference for *F. ulmaria*.

4.1.4 Individual variation

The water voles in this study contained an average of 7.0 plant species each, and 3.5 of them were grass species. With a total of 36 different species across all 48 samples, the water vole diet on Helgeland is also quite versatile as for the British and Belarusian voles (Neyland, 2011; Strachan et al., 2011; Wilson et al., 2017). Still, only two grass species (Unknown species 1 and *A. pubescens*) comprise almost half of the total water vole diet. Poaceae must be both important to the voles and abundant in their environment, although the variation of Poaceae in the diet between individuals is quite large, as seen in Figure 6. Unknown species 1 and *A. pubescens* make a very high proportion in some voles (Figure 5) and are probably nutritious and contain little toxins. Whether or not the diet selection and population cycles of

the voles is affected by plant toxins is not clear from the results here, and must be investigated further by determining the plant toxins/nutrient concentrations.

4.2 Metabarcoding analysis

4.2.1 Metabarcoding limitations and improvements

The metabarcoding analysis of the Helgeland water vole diet has worked well, but some limitations of this method must be considered. In herbivorous diets, it is challenging to quantify the relative abundance of each plant in the sample through the raw counts of the obtained sequences (Shokralla et al., 2012; Taberlet et al., 2018). Many different factors may bias the relative abundance of plants identified in the metabarcoding analysis: different digestibility of the plant species, differences in the metabarcode's copy number between species and PCR bias, where shorter sequences may be more efficiently amplified (Soininen et al., 2013). The PCR protocol also influences the quantitative results. A higher annealing temperature will decrease the amplification of species with primer mismatches (Taberlet et al., 2018). For these reasons, the relative abundance of sequences may thus be biased. Yet, metabarcoding is as good as other diet quantification analysis methods and "could perhaps be considered as better" (Taberlet et al., 2018, p. 137). Indeed, in comparison studies between metabarcoding and the traditional microhistology approach for diet analysis, Soininen et al. (2013) found that the two methods yield similar relative abundance of the plants in the feces. Soininen et al. (2009) also concludes that metabarcoding has a far better taxonomic resolution than microhistology, not to mention that microhistology is very time-consuming and delicate work. There is no reason to believe the proportions of the Helgeland water vole diet is biased, at least not more than other herbivore diet studies where metabarcoding was used. However, some measures could be done to further improve the validity of the results:

A few samples had a very large proportion of some species or families compared to other samples (Figure 4 and 5). One of these is *A. gmelinii* (Chenopodiaceae) which only occurred in four samples (Appendix 6) and had a low proportion in three of them (mean 3%). The fourth sample had 44% of the *A. gmelinii* (Figure 5). The species was not found in any of the negative controls, and so is unlikely to be contamination. PCR replicates would be useful to test if the outliers' proportions are correct, rather than just PCR/sequencing bias (Taberlet et al., 2018). However, it is not unlikely that some samples have large proportions of few or one species, as the sample only represent a limited time of foraging, perhaps just one "meal". This

meal could have been intense foraging on a single plant species if it were abundant at the voles' feeding location.

According to GBIF.org (2019), Lid and Lid (2013) and Thorvaldsen et al. (2019), all species identified in the diet except *Atriplex gmelinii* was likely to occur in Norway, but other species of the *Atriplex* genus are reported by Thorvaldsen et al. (2019) to exist nearby. Even though the species is unlikely, there is not enough evidence to exclude any species from the diet. A complete list of plant species occurrence at the exact sample site at the time of sampling would be helpful to exclude unlikely species from the diet with certainty. For now, it can only be discussed to what degree the results should be trusted.

Since negative controls were included in the analysis *after* extraction of the DNA from the feces samples, potential cross contamination during the extraction would not have been detected. The sequence count in each of the fecal samples after correction could potentially have been different if the negative controls had higher sequence count because of cross contamination.

4.2.2 Limitations of the trnL P6 loop marker

The trnL primer is a universal primer conserved only in vascular plants, thus bryophytes would not be detected in this diet analysis even though they may be a part of the diet. Mosses play an important part in arctic ecosystems and are important food items for many herbivores (Soininen et al., 2015). In their analysis of the two lemming species on Bylot island, Soininen et al. (2015) found 10 different families of mosses in their diets. The lemmings also showed a relatively high selectivity for some moss species. It is not unlikely that the water voles at Helgeland also eat mosses. For further studies on water vole diet analysis, markers and databases for identifying bryophytes (trnL c-h) should be included, as done by Soininen et al. (2015). Lichens would also not be detected, and primers targeting fungi would have to be used.

As British voles have been observed eating insects and mollusks (Strachan et al., 2011), we cannot exclude the possibility that the Helgeland voles also eat insects or other animals. For detecting these taxa, markers targeting insects and animals must be used, as is done by Lopes et al. (2015). On the other hand, there is no reason to believe insects and mollusks is an important part of the water vole diet. Not accounting for this has probably not had a great impact on the result of this study.

Each Unknown sequence in Table 1 is determined to two or more species that belong to the same family. It can be challenging to distinguish between closely related plant species due to a lack of taxonomic resolution of the trnL metabarcode marker in the reference libraries available (Taberlet et al., 2018). Taberlet et al. (2018) mentions Asteraceae, Cyperaceae, Poaceae and Rosaceae as plant families for which the trnL marker has especially poor resolution. These are indeed the families that the Unknown species in Table 1 belong to, in addition to Apiaceae. This may thus explain why these sequences had more than one match to the reference database.

Unknown species 4 (Table 1) may have been the carrot used as trap bait. Despite the precautions taken with respect to retention time, there are some ways in which carrot could have ended up in the fecal samples collected. Some chunks of carrot were scattered by the entrance of the traps (Appendix 2), so there may be individuals which have eaten carrot outside the trap and been captured and sampled later that day with enough time for the carrot to have passed through the digestive tract. Carrot, potato and bedding grass may also be in their faeces if voles have entered a trap where the mechanism did not work properly. This could happen for example if a small chunk of carrot had ended up under the trigger plate and prevented release when the vole stepped on it, or if the trap had moved slightly after it was set up. Another explanation is that retention time in some cases may be more rapid than suggested by Sakaguchi (2003) and Hagen et al. (2018). The consequences of the contamination from bait species is that the ratios of the other species in the voles' actual diet are skewed, though quite little. Keep in mind that Unknown species 4 may just as likely be *P. saxifraga* (Table 1) – in which case no contamination from carrot has occurred. The grass used as bedding material for the voles inside the traps are not identified to any specific species, and it is not possible to know if it may be a contamination.

Despite being identified as to two different species, Unknown species 6 and 7 (Table 1) may be the same species. Unknown species 6 had a best match on five different species in the reference database, of which four were also the best match of Unknown species 7. Apart from one nucleotide position, the sequences are identical (see below). It is not unlikely that the sequences originate from the same species that differ in one base due to intraspecific variation (Taberlet et al. 2018).

Unknown species 6: atcacgttttccgaaaacaaacaaaggttcagaaagcgaaaatcaaaaag
Unknown species 7: atcacgttttccgaaaacacacaaaggttcagaaagcgaaaatcaaaaag

4.3 Further studies

This study is the first to map the water vole diet in Scandinavia using metabarcoding, and as such it opens for many further studies. In the context of understanding population cycle mechanisms, it provides the basic knowledge about the water vole diet on Helgeland which can be used to investigate the correlation between food species abundance and water vole population sizes. Thorvaldsen et al. (2019) has suggested that regularly removing weeds (providing food and protection) will contribute in reducing the vole population sizes, which may indicate that plant availability influences vole population density.

Comparing observed diet with relative abundance of their diet species may provide knowledge about the voles' foraging behaviour, e.g. if they are specialists for certain food types or generalists. With this information one may gain more knowledge about plant-vole interactions and consequences of shifts in the food species abundance, and whether this is one of the factors that influences population dynamics. In addition, if the voles have a high preference for a relatively non-abundant plant species (specialist), links between the plants' nutrient and toxin content and observed diet can be investigated (diet selection). The diet may also be affected by the growth phase of different plants (e.g. flowering season). In all cases mentioned, a complete list of plant species composition with relative abundance on the exact sample sites at the time of sampling would be useful.

A possible next step for diet analysis would be to compare groups (e.g. site, sex) using multivariate statistics. Ordination plots may provide a new, exciting dimension to these kinds of analysis. In this case, the diet analysis should probably be based on bigger sample collections altogether and within seasons and sites to avoid confounding factors between them. Primers targeting specific plant families and other taxa than vascular plants (e.g. bryophytes) should be included, as well as PCR replicates.

5 Conclusion

Regarding the good taxonomic resolution and the results, this study has proved that metabarcoding works well for diet analysis of the water voles on Helgeland. The study gives a good overview of the water vole diet, at least at plant family level. However, there is potential for improvements for an even better result. Primers detecting other taxa than non-vascular plants (e.g. bryophytes) should be included in the analysis. In addition, it could be useful to add primers for amplifying specific plant families, especially those known to have a low resolution for the trnL g-h primer.

The Helgeland water vole diet consist mainly of Poaceae, although with high individual variation. The most abundant grass species was one of the species *Alopecurus geniculatus* or *Phalaris arundinacea*, and the second most abundant was *A. pubescens.* Individuals with a low proportion of Poaceae often had a high proportion of Rosaceae, which is the second most abundant plant family in the diet. The Rosaceae family was mainly represented by *F. ulmaria*, which was especially abundant in the diet of the water voles on the small study islands. Some samples had a relatively large proportion of unidentified sequences that had no match to the reference database used.

In future studies one may use the knowledge about the water vole diet to investigate vole-plant interaction and whether plant species abundance and quality has any impact on the population size fluctuations seen it the water vole population on Helgeland.

6 References

Alsos IG, Sjögren P, Edwards ME, Landvik JY, Gielly L, Forwick M, et al. (2016) Sedimentary ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. Holocene 26(4): 627–642. doi: 10.1177/0959683615612563

Ashby KR, Vincent MA (1976) Individual and population energy budgets of the water vole. Acta Theriol 21(34): 499-512.

Batsaikhan N, Henttonen H, Meinig H, Shenbrot G, Bukhnikashvili A (2016) *Arvicola amphibius* (errata version published in 2017). The IUCN Red List of Threatened Species 2016. Retrieved from: https://www.iucnredlist.org/species/2149/115060819

Boonstra R (1994) Population cycles in microtines: The senescence hypothesis. Evol Ecol 8(2):196-219. Doi: 10/1007/BF01238250

Boonstra R, Krebs CJ, Stenseth N (1998) Population cycles in small mammals: the problem of explaining the low phase. Ecol 79(5): 1479-1488. doi: 10.2307/176770

Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E (2016) OBITOOLS: a unix-inspired software package for DNA metabarcoding. Mol Ecol Resour 16: 176–182. doi: 10.1111/1755-0998.12428

Brzezinski M, Chibowska P, Zalewski A, Borowik T, and Komar E (2018) Water vole Arvicola amphibius population under the impact of the American mink Neovison vison: Are small midfield ponds safe refuges against this invasive predator? Mamm Biol 93: 182-188. doi: 10.1016/j.mambio.2018.06.002

Byng JW, Smets E, van Vugt R, Bidault E, Davidson C, et al. (2018). The phylogeny of angiosperms poster. The Global Flora 1: 4-35. Downloadable from: http://www.plantgateway.com/poster/

Carlsen TH, Aune S, Bär, A. (2011) Vegetasjonskartlegging I Vegaøyan verdensarvområde 2011. Bioforsk 6: 1-25.

Cerqueira D, De Sousa B, Gabrion C, Giraudoux P, Quéré JP, Delattre P (2006) Cyclic changes in the population structure and reproductive pattern of the water vole, Arvicola terrestris Linnaeus, 1758. Mamm Biol 71: 193–202. doi: 10/1016/j.mambio.2006.01.004

Comparatore VM, Cid MS, Busch C (1995) Dietary preferences of two sympatric subterranean rodent populations in Argentina. Revista Chilena de Historia Natural, 68: 197-206.

Debrovolny S, Blaschitz M, Weinmaier T, Pechatschek J, Cichna-Markl M, Indra A, Hufnagl P, Hochegger R (2019) Development of a DNA metabarcoding method for the

identification of fifteen mammalian and six poultry species in food. Food Chem 272: 354-361. doi: 10.1016/j.foodchem.2018.08.032

Flather CH, Sieg CH (2007) Species rarity: Definition, Causes, and Classification. In: Raphael, M. G., and Molina, R. (eds) Conservation of Rare and Little-known Species, 40-65.

Fremstad E (1997) Vegetasjonstyper i Norge. NINA Temahefte. 12: 1-279. ISSN: 0804-421X

GBIF.org. (2019) GBIF Home Page. Available from: https://www.gbif.org

Geneious Prime 2019.1.1. https://geneious.com

Gomes GB, Hutson KS, Domingos JA, Chung C, Hayward S, et al. (2017) Use of environmental DNA (eDNA) and water quality data to predict protozoan parasites outbreaks in fish farms. Aquacult 479: 467-473. doi: 10.1016/j.aquaculture.2017.06.021

Hagen KB, Müller DWH, Ortmann S, Kreuzer M, Clauss M (2018) Digesta kinetics in two arvicoline rodents, the field vole (*Microtus agrestis*) and the steppe lemming (*Lagurus lagurus*). Mamm Biol 89: 71-78. doi: 10.1016/j.mambio.2018.01.003

Hawlitschek O, Angel F, Balmori-de La Puente A, Castresana J (2018) A pipeline for metabarcoding and diet analysis from fecal samples developed for a small semi-aquatic mammal. PLoS One 13(8): 1-19. doi: 10.1371/journal.pone.0201763

Heiska S, Tikkanen O, Rousi M (2007) Bark salicylates and condensed tannins reduce vole browsing amongst cultivated dark-leaved willows (*Salix myrsinifolia*). Chemoecol 17(4): 245-253. doi: 10.1007/s00049-007-0385-9

Holt RD, Kimbrell T (2007) Foraging and population dynamics. In: Stephens, D. W., Brown, J. S. and Ydenberg, R. C. (eds). Foraging: Behavior and Ecology: 365-395.

Iwanowicz DD, Vandergast AG, Cornman RS, Adams CR, Kohn JR, et al. (2016) Metabarcoding of Fecal Samples to Determine Herbivore Diets: A Case Study of the Endangered Pacific Pocket Mouse. PLoS One 11(11): 1-23. doi: 10.1371/journal.pone.0165366

Khanam S, Howitt R, Mushtaq M, Russell JC (2016) Diet analysis of small mammal pests: A comparison of molecular and microhistological methods. Integrative Zool 11(2): 98-110. doi: 10.1111/1749-4877.12172

Kotler BP, Brown JS (2007) Community ecology. In: Stephens DW, Brown JS, Ydenberg RC (eds). Foraging: Behavior and Ecology: 397-434.

Krebs CJ, Gaines MS, Keller BL, Myers JH, Tamarin RH (1973) Population cycles in small rodents. Science 179(4068): 35-41.

Lid J, Lid DT (2005) Norsk flora. Det norske samlaget. ISBN: 978-82-521-6029-1

Lopes CM, De Barba M, Boyer F, Mercier C, da Silva Fihlo PJS, et al. (2015) DNA metabarcoding diet analysis for species with parapatric vs sympatric distribution: a case study on subterranean rodents. Heredity 114(5): 525-536. doi: 10.1038/hdy.2014.109

Moorhouse TP, Gelling M, Macdonald DW (2008) Effects of Forage Availability on Growth and Maturation Rates in Water Voles. 77(6): 1288-1295. doi: 10.1111/j.1365-2656.2008.01431.x

Newman J (2007) Herbivory. In: Stephens DW, Brown JS, Ydenberg RC (eds). Foraging: Behavior and Ecology: 175-218.

Neyland PJ (2011) Habitat, home range, diet and demography of the water vole (Arvicola amphibious): Patch-use in a complex wetland landscape.. thesis, Swansea University.

Puig S, Rosi MI, Cona MI, Roig VG, Monge SA (1999) Diet of Piedmont populations of *Ctenomys mendocinus* (Rodentia, Ctenomyidae): seasonal patterns and variations according sex and relative age. Acta Theriol 44(1): 15–27.

Pyke G, Pulliam H, Charnov E (1977) Optimal Foraging: A Selective Review of Theory and Tests. Q Rev Biol 52(2): 137-154.

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://R-project.org/.

Rohland N, Reich D (2012) Cost-Effective, High-Throughput DNA Sequencing Libraries for Multiplexed Target Capture. Genome Res 22(5): 939–46. doi: 10.1101/gr.128124.111

Sakaguchi E (2003) Digestive strategies of small hindgut fermenters. Anim Sci J 74(5): 327-337. doi: 10.1046/j.1344-3941.2003.00124.x

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. Mol Ecol 21(8): 1794-1805. doi: 10.1111/j.1365-294X.2012.05538.x

Sjögren P, Edwards ME, Gielly L, Langdon CT, Croudace IW, et al. (2017) Lake sedimentary DNA accurately records 20th Century introductions of exotic conifers in Scotland. New Phytol 213(2): 929-941. doi: 10.1111/nph.14199

Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, et al. (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Front Zool 6(16): 1-9. doi: 10.1186/1742-9994-6-16

Soininen EM, Zinger L, Gielly L, Bellemain E, Bråthen KA, et al. (2013) Shedding new light on the diet of norwegian lemmings: DNA metabarcoding of stomach content. Polar Biol 36(7): 1069-1076. doi: 10.1007/s00300-013-1328-2

Soininen E, Gauthier G, Bilodeau F, Berteaux D, Gielly L (2015) Highly overlapping winter diet in two sympatric lemming species revealed by DNA metabarcoding. PLoS ONE 10(1). doi: 10.1371/journal.pone.0115335

Stoddart DM (1971) Breeding and Survival in a Population of Water Voles. J Anim Ecol 40: 487-494. doi: 10.2307/3257

Strachan R, Moorhouse T, Gelling M (2011) Water vole conservation handbook, Wildlife Cons Resour Unit, University of Oxford.

Sønstebø JH, Gielly L, Brysting A, Elven R, Edwards M, et al. (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. Mol Ecol Resour 10: 1009–1018. doi: 10.1111/j.1755-0998.2010.02855.x

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, et al. (2006) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res, 35(3): 1-8. doi: 10.1093/nar/gkl938

Taberlet P, Bonin A, Zinger L, Coissac E (2018) Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press, Oxford. doi: 10.2093/oso/9780198767220.001.0001

Thorvaldsen P, Dyrhaug M, Dorn W (2019) Vånd (Arvicola amphibius) i jordbrukslandskapet. Resultat fra casestudie på faktorer som påvirker skadeomfang av vånd på Helgelandskysten og forslag til skadedempende tiltak. NIBIO-rapport. 5(20): 1-38.

Wickham H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Willerslev E, Davison J, Moora M, Zobel M, Coissac E, et al. (2014) Fifty thousand years of Arctic vegetation and megafaunal diet. Nature 506: 47–51. doi: 10.1038/nature12921 PMID: 24499916

Wilson DE, Lacher Jr. TE, Mittermeier RA (2017) Handbook of the Mammals of the World (HMW). Lynx Edicions. 7: Rodents II. ISBN: 9788416728046

Ydenberg RC, Brown JS, Stephens DW (2007) Foraging: An overview. In: Stephens DW, Brown JS, Ydenberg RC (eds). Foraging: Behavior and Ecology: 1-28.

7 Appendix

Appendix 1: *Gulbrandsøyan S with Gulbrandsøyan Midt S (left) and Austbø in the background across the sound. The yellow marks are trap-flags marking trap sites. High abundance of Poaceae and some F. ulmaria (bottom left corner). Photo: Nina Østby.*

Appendix 2: *Loaded trap on Gulbrandsøyan S with moss for protection against wind and sun and carrot scattered by the entrance. In the background are Gulbrandsøyan Midt S (middle) and Gulbrandsøyan Midt N (left). Photo: Nina Østby.*

Appendix 3: *Site, time of capture and sex of the vole each sample belong to. The sample ID correspond to the sample numbers in Figure 6 and the PIT-tag to the specific vole each sample belonged to.*

Site	Date of capture	Month	PIT-tag	Sex	Sample ID
Geiterøya N	2018-04-16	april	000799CA65	m	5
Geiterøya N	2018-04-16	april	000799A6B5	f	3
Geiterøya N	2018-04-16	april	00079AA479	m	12
Geiterøya N	2018-04-16	april	0007997828	m	14
Geiterøya N	2018-04-16	april	000799BBCE	f	10
Geiterøya N	2018-04-16	april	000799C2B6	m	15
Geiterøya N	2018-04-16	april	000799A06A	f	7
Geiterøya N	2018-04-16	april	000799C0DF	f	16
Geiterøya N	2018-04-17	april	000799AF12	f	4
Geiterøya N	2018-04-17	april	000799713A	f	8
Geiterøya N	2018-04-17	april	000799B154	f	6
Geiterøya N	2018-04-17	april	0007999FAE	f	9
Geiterøya N	2018-04-17	april	00079A071B	m	11
Geiterøya N	2018-04-17	april	000799D8B8	m	13
Geiterøya N	2018-04-17	april	000799896F	m	2
Geiterøya N	2018-04-17	april	000799FFDF	m	1
Geiterøya N	2018-07-19	july	000799F8D6	m	22
Geiterøya N	2018-07-19	july	00079A2F7D	m	19
Geiterøya N	2018-07-19	july	000799C2B6	m	18
Geiterøya N	2018-07-19	july	00079A3590	pm	20
Geiterøya N	2018-07-19	july	000799C53D	m	21
Geiterøya N	2018-07-19	july	000799F7A2	f	17
Geiterøya N	2018-07-19	july	000799ED79	m	26
Geiterøya N	2018-07-19	july	00079A071B	m	24
Geiterøya N	2018-07-19	july	00079988B0	pm	23
Geiterøya N	2018-07-19	july	00079A363C	pm	25
Geiterøya N	2018-09-04	september	000799CD1E	f	28
Geiterøya N	2018-09-04	september	00079A25E7	pm	27
Gulbrandsøyan S	2018-09-05	september	00079992B7	f	31

Gulbrandsøyan S	2018-09-05	september	000799BA5B	f	34
Gulbrandsøyan S	2018-09-05	september	000799A328	f	33
Gulbrandsøyan S	2018-09-05	september	000799B43D	m	29
Gulbrandsøyan S	2018-09-05	september	000799B6F0	f	36
Gulbrandsøyan S	2018-09-05	september	000799C403	m	35
Gulbrandsøyan S	2018-09-05	september	000799856A	m	38
Gulbrandsøyan S	2018-09-05	september	00079971BA	m	37
Gulbrandsøyan S	2018-09-05	september	000799BB09	m	30
Gulbrandsøyan S	2018-09-05	september	000799AE90	m	39
Gulbrandsøyan S	2018-09-06	september	00079AA50C	f	32
Gullrekka 2	2018-09-08	september	000799B158	f	47
Gullrekka 2	2018-09-08	september	0007997E2D	m	44
Gullrekka 2	2018-09-08	september	000799C361	m	46
Gullrekka 2	2018-09-08	september	0007998B73	m	48
Gullrekka 2	2018-09-08	september	0007999CA9	m	45
Gullrekka 1	2018-09-09	september	000799BC66	f	40
Gullrekka 1	2018-09-09	september	00079997B3	f	43
Gullrekka 1	2018-09-10	september	0007997F74	f	41
Gullrekka 1	2018-09-10	september	000799C4A1	f	42

Appendix 4: *Families removed in the filtering of the 5% lowest sequence count in each sample.*

Removed families		
Solanaceae	Cornaceae	Hydrocharitaceae
Brassicaceae	Dryopteridaceae	Caprifoliaceae
Ranunculaceae	Cystopteridaceae	Ulmaceae
Onagraceae	Menyanthaceae	Violaceae
Araceae	Boraginaceae	Asparagaceae
Campanulaceae	Thelypteridaceae	

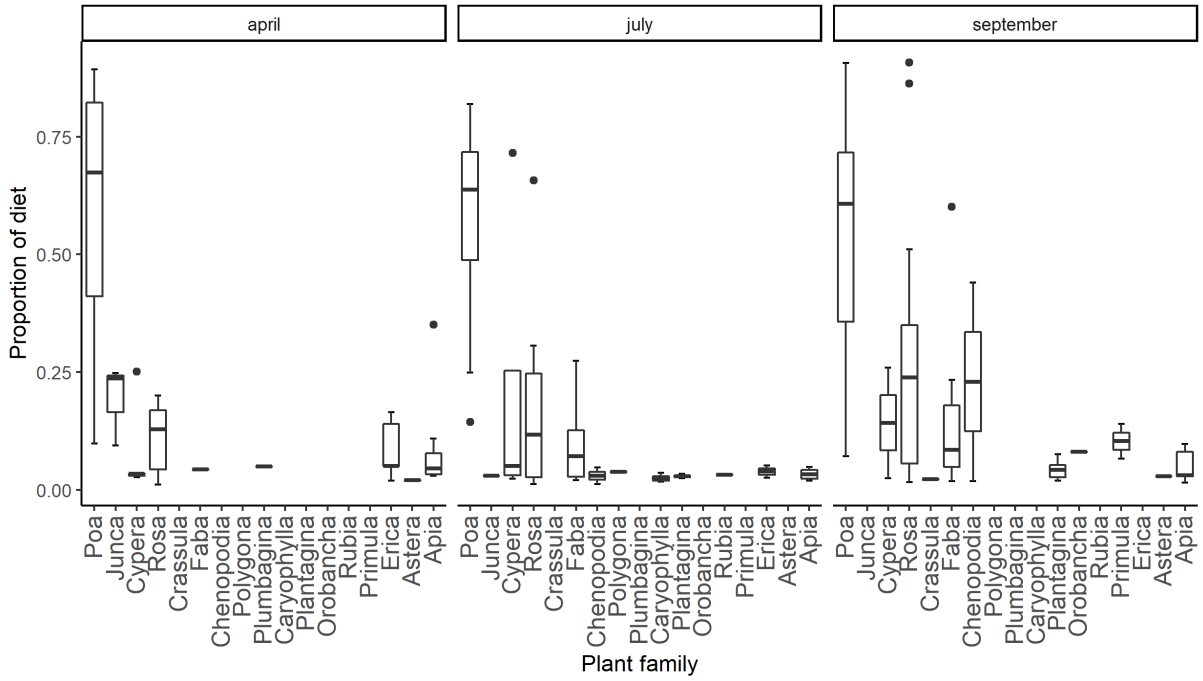
Appendix 5: *Phylogenetic information/relatedness of the most abundant plant families detected in the water vole diet, according to the Angiosperm phylogeny group (APG IV) classification system (Byng et al., 2018).*

Family	**Order**	**Class**
Poaceae	Poales	Angiospermae
Juncaceae	Poales	Angiospermae
Cyperaceae	Poales	Angiospermae
Rosaceae	Rosales	Angiospermae
Crassulaceae	Saxifragales	Angiospermae
Fabaceae	Fabales	Angiospermae
Chenopodiaceae	Caryophyllales	Angiospermae
Polygonaceae	Caryophyllales	Angiospermae
Plumbaginaceae	Caryophyllales	Angiospermae
Caryophyllaceae	Caryophyllales	Angiospermae
Plantaginaceae	Lamiales	Angiospermae
Orobanchaceae	Lamiales	Angiospermae
Rubiaceae	Gentianales	Angiospermae
Primulaceae	Ericales	Angiospermae
Ericaceae	Ericales	Angiospermae
Asteraceae	Asterales	Angiospermae
Apiaceae	Apiales	Angiospermae

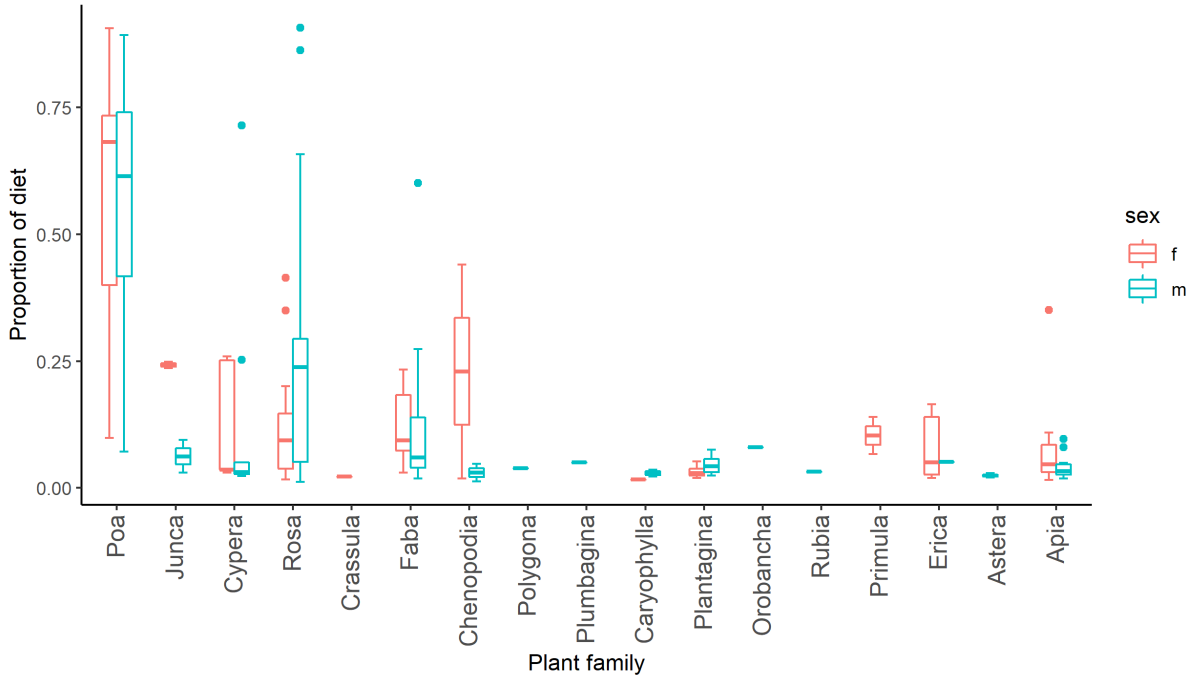
Appendix 6: *Number and proportion of samples each plant family and species were detected in. A family/species was considered present in a sample if it had a count > 0. The "Unknown species" are the unknown species that had two or more matches to the reference database (Table 1).*

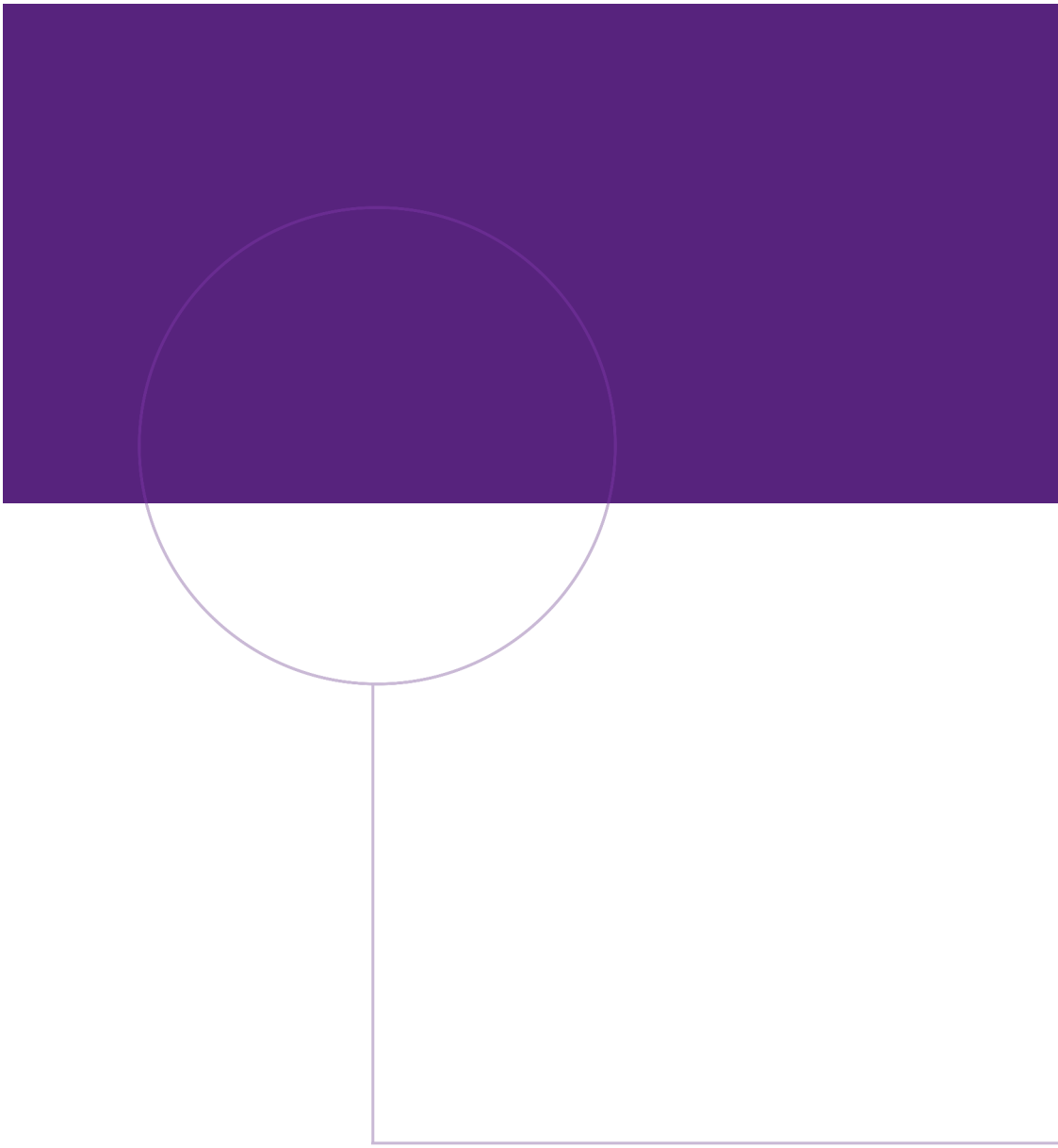
Family name	Samples with family: n (%)	Species name	Samples with species: n (%)
Poaceae	48 (100%)	Unknown species 1	46 (96%)
		Avenula pubescens	30 (63%)
		Unknown species 2	29 (60%)
		Agrostis clavata	20 (42%)
		Unknown species 3	18 (38%)
		Anthoxanthum nipponicum	12 (25%)
		Avenella flexuosa	6 (13%)
		Brachypodium pinnatum	4 (8%)
		Phleum alpinum	2 (4%)
		Dactylis glomerata	1 (2%)
Juncaceae	4 (8%)	Luzula pilosa	4 (8%)
Cyperaceae	14 (29%)	Carex vaginata	11 (23%)
		Carex aquatilis	4 (8%)
		Unknown species 5	1 (2%)
Rosaceae	31 (65%)	Filipendula ulmaria	21 (44%)
		Alchemilla alpina	14 (29%)
		Sorbus aucuparia	2 (4%)
		Unknown species 8	1 (2%)
Crassulaceae	1 (2%)	Rhodiola rosea	1 (2%)
Fabaceae	19 (40%)	Lathyrus pratensis	17 (35%)
		Vicia cracca	4 (8%)
Chenopodiaceae	4 (8%)	Atriplex gmelinii	4 (8%)
Polygonaceae	1 (2%)	Bistorta vivipara	1 (2%)
Plumbaginaceae	1 (2%)	Armeria scabra	1 (2%)
Caryophyllaceae	3 (6%)	Stellaria holostea	3 (6%)
Plantaginaceae	8 (17%)	Linaria vulgaris	7 (15%)
		Veronica chamaedrys	1 (2%)
Orobanchaceae	1 (2%)	Euphrasia wettsteinii	1 (2%)
Rubiaceae	1 (2%)	Galium boreale	1 (2%)
Primulaceae	2 (4%)	Glaux maritima	2 (4%)
Ericaceae	7 (16%)	Vaccinium vitis-idaea	5 (10%)
		Vaccinium uliginosum	2 (4%)
Asteraceae	2 (4%)	Unknown species 6	1 (2%)
		Unknown species 7	1 (2%)
Apiaceae	20 (42%)	Unknown species 4	17 (35%)
		Anthriscus sylvestris	3 (6%)
NA	39 (81%)		

Appendix 7: *Proportion of plant families in water vole diet in April (n=16), July (n=10) and September (n=22). The black horizontal line is the median of the value of all samples, and the box contains the middle 50% of the samples: 25% above (upper quartile), and 25% (lower quartile) below the median. The whiskers represent values outside the middle 50%, and the points are outliers.*



Appendix 8: *Proportion of plant families in diet in male (n=28) and female (n=20) water voles. Samples were collected on four different islands along the Helgeland coast in three different seasons (April, July, September). The black horizontal line is the median of the value of all samples, and the box contains the middle 50% of the samples: 25% above (upper quartile), and 25% (lower quartile) below the median. The whiskers represent values outside the middle 50%, and the points are outliers.*





NTNU
Norwegian University of
Science and Technology