The Efficacy of DNA sequencing on samples of terrestrial invertebrates 2018/2019

Published March 2022

Natural England Research Report NECR388

www.gov.uk/natural-england

Natural England Research Report NECR388

The Efficacy of DNA sequencing on samples of terrestrial invertebrates 2018/2019

REES, H.C., MADDISON, B.M., OWEN, J.P., BAKER, C.A., BISHOP, K., GOUGH, K.C., WEBB, J.R

Published March 2022

ISBN: 978-1-78354-894-1

Project details

This report was first published 31st January 2022. The section on costs has been revised in this version for clarity.

This report should be cited as: Rees, H.C., Maddison, B.M., Owen, J.P., Baker, C.A., Bishop, K., Gough, K.C., Webb, J.R. 2022. The Efficacy of DNA sequencing on samples of terrestrial invertebrates 2018-19. Natural England Commissioned Reports, Number 388

Natural England Project manager

Natural England Project Manager: Debbie Leatherland and Jon Webb

Contractor

Contractor details: RSK ADAS

Author

Author details: Dr Helen C. Rees, helen.rees@adas.co.uk

Keywords

Keywords: invertebrates, metabarcoding, beetles, spiders, flies, vane trap, pitfall trap, hand search, abundance

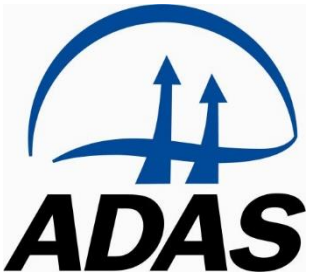
Acknowledgements

We would like to thank the University of Nottingham's DeepSeq: Next Generation Sequencing Facility for their assistance in troubleshooting the indexing PCR issues and for their Illumina sequencing service. Thanks also to the Natural England Field unit, Area Team staff and national staff for collecting and IDing the samples. Thank you to the rangers from Forestry England and Nottinghamshire County Council for their monthly collection of samples from the traps at Sherwood Forest.

Further information

This report can be downloaded from the Natural England Access to Evidence Catalogue: **http://publications.naturalengland.org.uk/** . For information on Natural England publications contact the Natural England Enquiry Service on 0300 060 3900 or e-mail **enquiries@naturalengland.org.uk**.

Contents



The Efficacy of DNA sequencing on samples of terrestrial invertebrates

2018/2019

Natural England Commissioned Report NECR388

Foreword

Natural England (NE) aims to make monitoring programmes more efficient and to investigate this they wish to compare the efficacy of DNA sequencing and how it compares to hand identification of invertebrates. This project will deliver important baseline data on the applications of DNA technologies, specifically mass DNA sequencing (metabarcoding) of terrestrial invertebrates to survey and monitor biodiversity.

This report:

- Investigates the accuracy of DNA species identification for invertebrates.
- Investigates the reliability of the BOLD database for undertaking invertebrate identification.
- Investigates whether abundance can be inferred as a part of DNA sequencing.
- Presents the methods and results of metabarcoding as compared to traditional techniques (including a comparison of turnaround time and cost).
- Makes recommendations for future work in line with the pros and cons of the two methods.

This report is focused on the DNA element of the work, with a separate report by the Natural England Field Unit discussing the invertebrates found through the field work.

1. Introduction

Natural England is the Government's advisor for the natural environment. It provides practical advice on how to safeguard England's natural wealth for the benefit of everyone. ADAS is an environmental consultancy which exists to provide ideas, specialist knowledge and solutions to secure our food and enhance the environment.

Natural England has recently begun to explore the application of DNA based technologies to biodiversity monitoring programmes with the hope of its uptake leading to efficiencies in current monitoring programmes. There are well over 30,000 different species of invertebrates in the UK (Key et al. 2000) and it can take many years to become an expert on species identification. Natural England are interested in the identification of all beetles and bycatch from veteran trees in Sherwood Forest (Vane trapping) and soft rock cliff seepages at Highcliffe, Dorset (hand collecting and pitfall trapping) to the species level where possible using a curated reference database such as the BOLD database.

1.1 Metabarcoding

DNA metabarcoding is a method used to rapidly assess biodiversity and combines two techniques: DNA based identification and high-throughput sequencing (Margulies et al. 2005), allowing for the DNA sequencing of bulk samples without a prior step of specimen sorting. Using 'universal' PCR primers (primers that work across a wide range of taxa in the target sample) to amplify specific target sequences (usually mitochondrial DNA sequences) the mass-amplification of the target of interest from multiple species can be achieved. Metabarcoding has proven an effective technique for community biodiversity assessment across a range of taxa and environments (Deiner et al. 2016; Drummond et al. 2015; Hajibabaei et al. 2011; Murray et al. 2012; Valentini et al. 2016) and is able to generate comprehensive data sets many times quicker than traditional hand identification methods. This so-called 'metabarcoding' approach is therefore a powerful means to study and understand the diversity and distribution of fauna and flora.

The amplified DNA fragments are mass sequenced using next generation DNA sequencing methods which returns large numbers of high quality sequence reads. Each organism present in the sample will contribute many copies of its mitochondrial DNA so low numbers of individual species should be detected. Sequence data is usually reduced down to a single representative of each species mitochondrial DNA sequence - an operational taxonomic unit (OTU). The individual OTUs can then be compared against existing DNA databases to identify the organisms that they represent.

The success of metabarcoding is dependent upon the primer set chosen for use and its target loci. Ideally primers should target a hypervariable region (for high resolution taxonomic discrimination) and thus will determine the efficiency and accuracy of species detection and identification. Additionally, primers should target short DNA fragments (around 400 bp or less) which allows for the recovery of potentially degraded target DNA which may have been subjected to long term storage or that has been taken from hostile

sample matrices. DNA is liable to degradation by factors such as nucleases, UV light, microbial action and the temperature and humidity of storage conditions will affect DNA quality after sample collection. Some environments will be more detrimental to DNA quality than others and are therefore described as ‘hostile’ environments.

Universal primers are available for a wide range of gene fragments across a range of taxa for example nuclear 18S and 28S ribosomal RNA markers (Machida et al. 2012a) the mitochondrial 12S rRNA gene (Machida et al. 2012b), and the mitochondrial Cytochrome c Oxidase I gene (COI) which has been adopted as the standard ‘taxon barcode’ for many taxa (Hebert et al. 2003). The standard COI target primer set was developed to amplify a 658 bp region (Folmer et al, 1994), however, as this fragment was too long for metabarcoding an attempt was made to design a primer set which would amplify a shorter ‘mini-barcode’ (Meusnier et al. 2008). This primer set was of limited use due to its poor efficiency over a large range of taxa therefore a primer set which was based on a modified version of the ‘Folmer’ reverse primer and a newly designed forward primer was created and was shown to have a higher amplification success rate than the original ‘Folmer’ primers (Leray et al. 2013). This primer set has since been used in several peer reviewed studies and was also the primer set of choice within Natural England report NECR252 (Tang et al. 2018).

1.2 The BOLD Database

The BOLD database is a publicly available database of DNA sequences which has been generated by the ‘Barcode of Life’ initiative which aims to build a reference library of standardized DNA sequences able to identify hundreds of thousands of species. The sequences populating the BOLD database are those agreed internationally as being regions of the genome which allow good discrimination between species (with little variation between individuals of the same species). The mitochondrial cytochrome oxidase subunit I gene (COI) is commonly used, with other more suitable regions being used for plants and fungi. The database is fully curated and has been created such that each sequence can be linked back to a preserved specimen which has been previously identified by taxonomy experts to the species level.

1.3 Aims and Objectives

The aim of this study was to investigate the efficacy of DNA sequencing and how it compares to hand identification. This study aims to take samples previously identified by hand and subjects them to metabarcoding to investigate both the accuracy of DNA based species identification compared to traditional hand identification and the reliability of the BOLD database for undertaking invertebrate identification. The aim of identification by metabarcoding will be to identify the individuals present down to the species level and in circumstances where this is not possible to identify down to the genus level. The results of this will be compared to taxonomic identification and species lists generated for all samples. Additionally, the resulting data set will be used to investigate whether species abundance can be inferred as a part of DNA sequencing. This report outlines the

methodology employed in this study along with the results obtained and discussion on the use of metabarcoding in terrestrial invertebrate analysis. A time and cost analysis is also included along with future recommendations for additional studies. All data will be made available for further study and training material used for a training day for Natural England staff on the DNA approaches used can be made available upon request.

2. Methods

2.1 Sample Collection

Natural England survey protocols were used to collect samples from veteran trees in Sherwood Forest (Vane trapping) and soft rock cliff seepages (hand collecting and pit-falling) at Highcliffe, Dorset. Samples were identified by hand and individuals of each species were counted and recorded on an Excel spreadsheet. The time incurred for their identification and the storage methods were noted. All samples were collected from Natural England (Peterborough office) on 11th December 2018 and transported to the ADAS laboratories for processing. Samples were supplied in a variety of tubes and vials, with multiple tubes per sample with the exception of the pitfall trap samples and two of the spider samples which were provided as a single sample (Figure 2.1).

Vane trap samples from Sherwood Forest were collected in 50% propylene glycol, sorted and then transferred to 50% ethanol for a few weeks. These were then transferred to specialists for identification and then storage in 95% ethanol.

Pitfall samples from Highcliffe were collected in a 50% propylene glycol over a two week period, sorted and then transferred to 50% ethanol for 8 weeks. These were then transferred to specialists for identification and storage in 95% ethanol.

Hand collected samples were stored in 50% ethanol for 8 weeks. These were then transferred to specialists for identification and then storage in 95% ethanol.

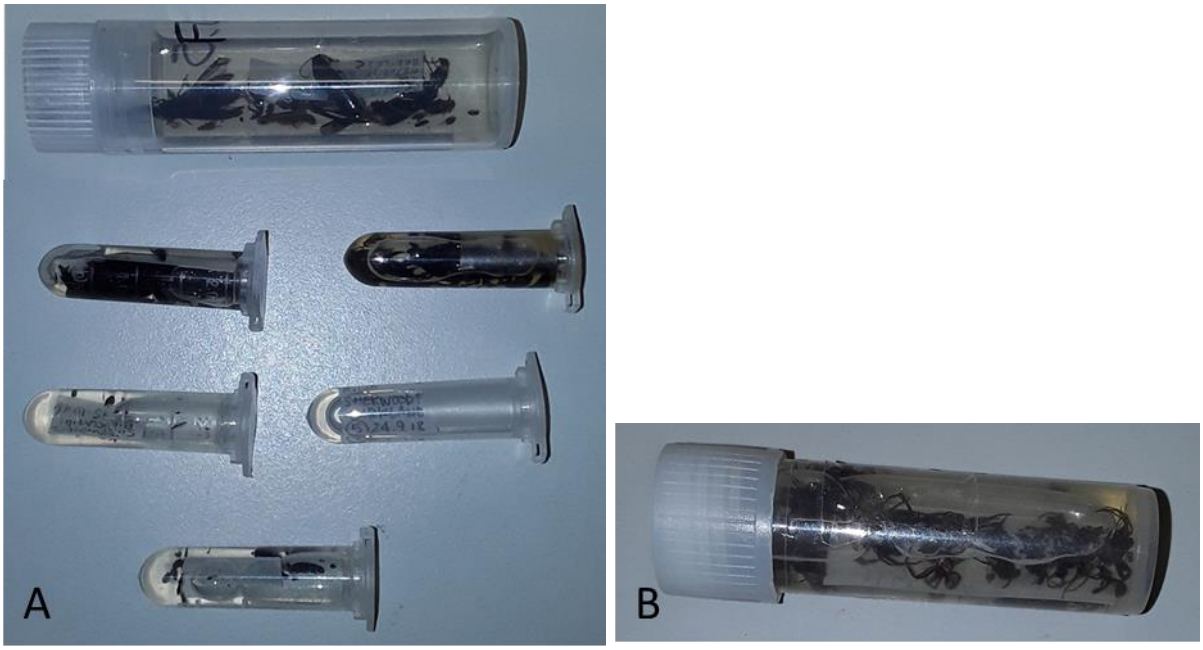


Figure 2.1 Sample images. Images of two of the samples provided by Natural England: A) Sherwood FE Vane Trap T5; B) Hand search Highcliffe Spiders 1. (photo: Helen Rees / ADAS)

2.2 Laboratory Standards and Specifications

Establishing and maintaining quality standards is essential for the efficient and effective operation of a diagnostics laboratory. This is important for ensuring the quality and traceability of results and reagents. All laboratory activities associated with DNA analysis are subject to errors if quality control is inadequate. Our DNA analysis follows a unidirectional workflow with separate laboratories and staff to act as a physical separation of the different aspects of the analysis work greatly reducing the potential for contamination of samples or the PCR amplicons. 'Blank' PCRs (sterile water rather than DNA) are used to monitor for reagent/procedural contamination, and in addition positive control samples are used to increase confidence in the results and identify any cross-contamination issues, should they occur.

2.3 Specimen Size Sorting

All samples were size sorted by eye using a cut-off of approximately <7mm to denote 'small' specimens and ≥7mm to denote 'big' specimens into fresh sterile petri dishes (Figure 2.2). It has been shown that worldwide there are two peaks in invertebrate size at approximately 4mm and 10mm therefore the 7mm cut off being in between these two values should allow us to capture those species that generally fit into these categories (Webb, J., personal communication). As agreed with Natural England, three particularly large beetles were removed from the samples, the heads cut off and placed back into the sample that they originally came from. It was felt that the significantly larger size of these three beetles within their respective samples could skew the resulting sequence data by effectively 'swamping' the total sample with their DNA. The remaining bodies of the beetles were stored in individual 7 mL bijoux tubes. Size sorted sub-samples were then placed into petri-dishes and allowed to dry in a fume hood prior to DNA extraction.

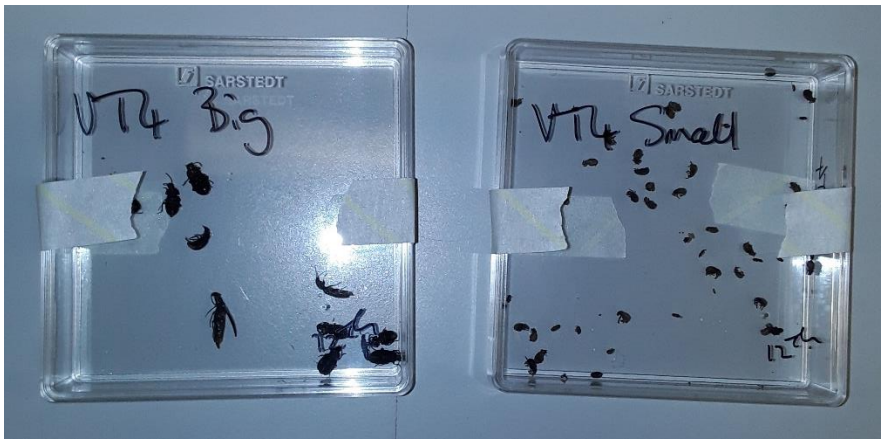


Figure 2.2 Size separated samples. An example of a size separated sample within petri dishes (Sherwood NCC Vane Trap 4) (Photo: Helen Rees / ADAS)

2.4 DNA Extraction

Each sub-sample (small or big) was individually transferred to a clean, sterile mortar and ground into a powder using a pestle and liquid nitrogen. A few of the small sub-samples with low numbers of specimens were ground within a sterile Eppendorf tube using an Eppendorf pestle and liquid nitrogen. The powdered sub-sample was then divided into two 50 mL tubes (one for long term storage at -20 ºC and the other for further processing) and the weight recorded for each sub-sample. After use mortar and pestles were immediately immersed in 10 % bleach for a minimum of 10 minutes and then cleaned in between samples with 10 % Distel (Tristel™), rinsed with dH2O and then autoclaved at 121 ºC for 15-20 minutes.

DNA extraction was performed using the DNeasy blood and tissue kit (Qiagen) following the manufacturer's instructions (Appendix 1) with some initial optimisation of sub-sample quantities (based on the recorded weights) and finally resuspended in 200 µL of elution buffer. Three extraction blanks to test for cross-contamination were also included. All extractions were quantified using a Qubit 3.0 Fluorometer (Invitrogen) following the manufacturer's instructions then stored at -20 ºC prior to PCR set up.

2.5 PCR Amplification

PCR amplification was performed in a separate laboratory to DNA extractions with dedicated equipment and PPE; PCRs were set up in a clean 'PCR room' within a UV sterilisable PCR cabinet. To ensure the unidirectional workflow DNA extracts are collected from the DNA extraction laboratory and transferred to the PCR set-up laboratory. Laboratory personnel do not return to the DNA extraction laboratory during that same day thus maintaining the unidirectional workflow.

The primer combinations used for the first round PCR amplification were mICOIintF/jgHCO2198 or HexCOIF4/HexCOIR4 both of which additionally included overhang adapter sequences (Table 2.1, Figure 2.3A) at the 5' end of the primers for compatibility with Illumina index and sequencing adapters (Illumina 2011). These primers amplify a fragment of the Cytochrome c Oxidase subunit I gene (COI) and have been shown to perform well in invertebrate metabarcoding studies (Leray et al. (2013); Geller et al. (2013)). PCRs included two negative controls (ddH2O in place of DNA); the three DNA extraction blanks; two positive control samples (*Esox lucius* DNA (pike) and *Allolobophora chlorotica* DNA (earthworm)) and all 67 invertebrate sub-samples (Appendix 1).

Table 2.1 Primers used in PCR rounds one and two this study

Primer Name	Oligonucleotides (5'-3')	%GC	Tm	Reference
mICOIintF	TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGGGWACWGGWTGAACWGTWTAYCC YCC	50.8	>75	Leray *et al.* (2013)
jgHCO2198	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGTAIACYTCIGGRTGICCRAARAAYCA	47.5	>75	Geller *et al.* (2013)
HexCOIF4	TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGHCCHGAYATRGCHTTYCC	51.9	>75	Marquina *et al.* (2018)
HexCOIR4	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGTATDGTRATDGCHCCNGC	51.9	>75	Marquina *et al.* (2018)
Index 1	CAAGCAGAAGACGGCATACGAGAT**XXXXX XXX**GTCTCGTGGGCTCGG	-	-	Illumina (2011)
Index 2	AATGATACGGCGACCACCGAGATCTACAC **XXXXXXXX**TCGTCGGCAGCGTC	-	-	Illumina (2011)

Sequences marked in blue are Illumina overhang adapter sequences, Index 1 and 2 sequences are in purple and are marked with X's as this sequence is variable for each different sample, those in red are the P5 and P7 sequences, and those in black are locus-specific sequences. Index 1 (i7) and Index 2 (i5) are examples of the type of primers used with the Index sequence itself being altered for different samples.

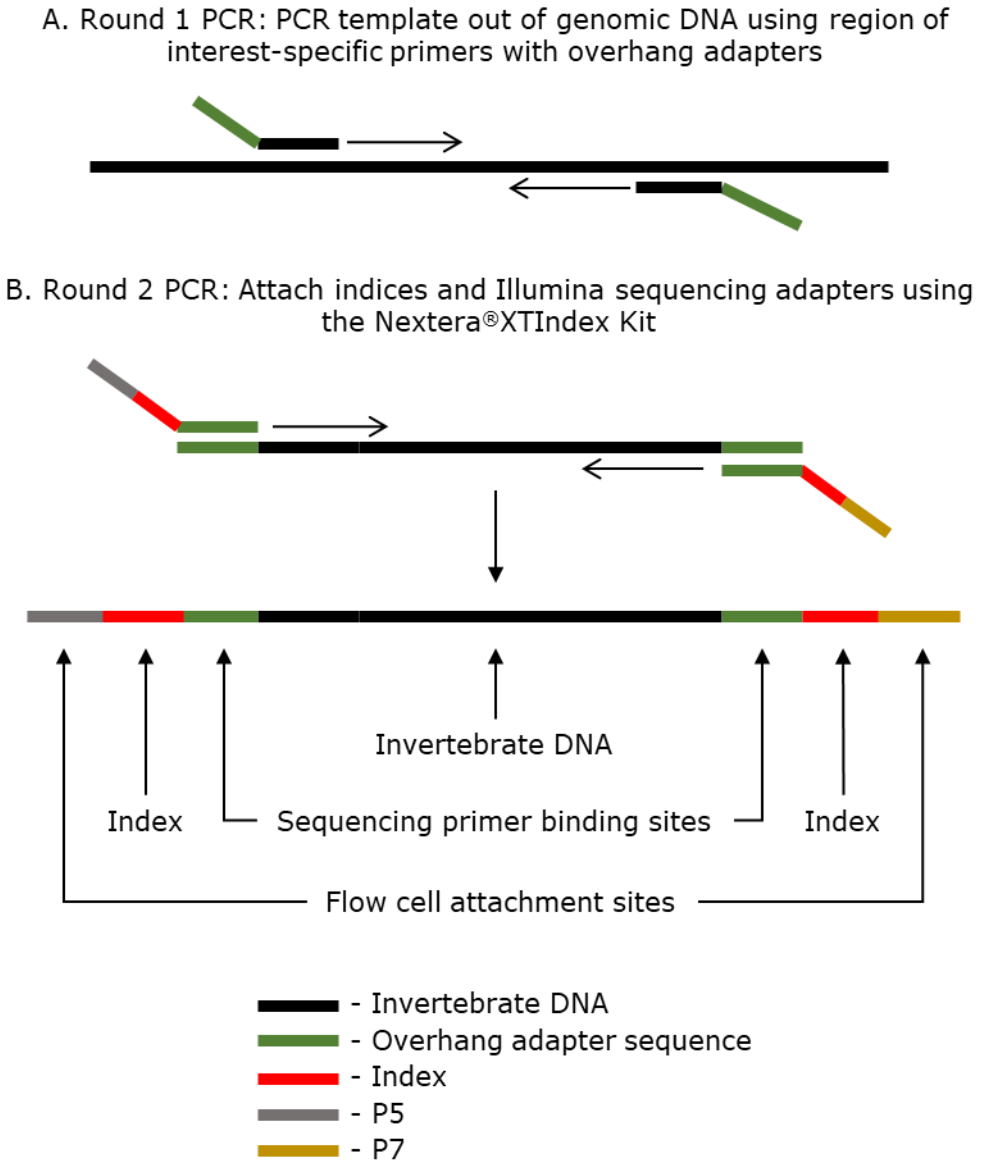
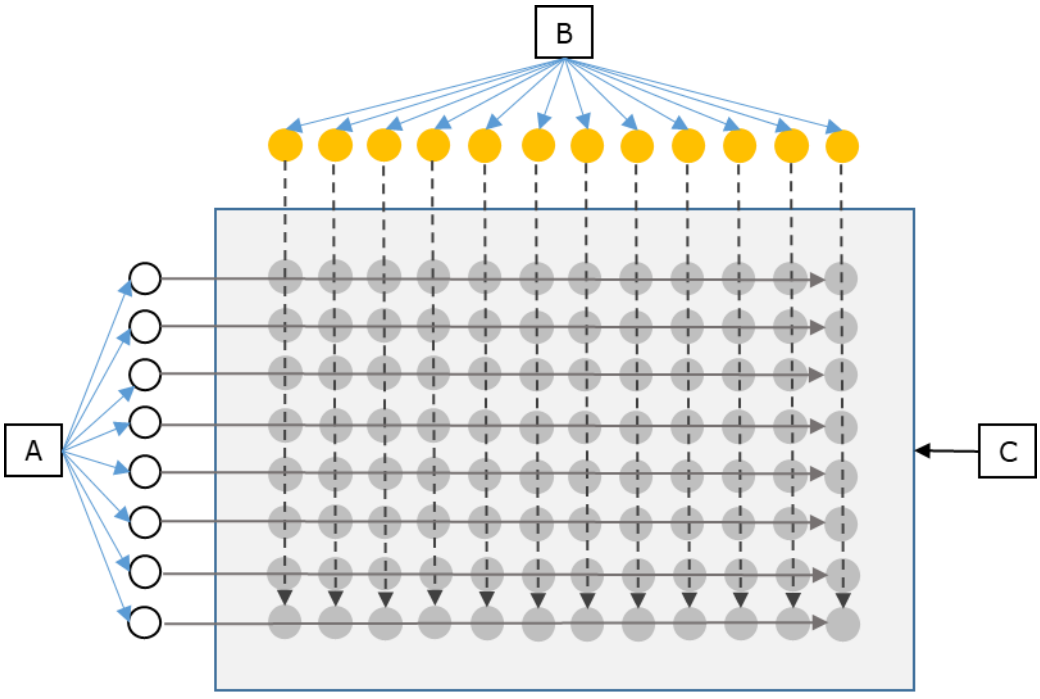


Figure 2.3 PCR Amplicon Workflow. A) User-defined forward and reverse primers complementary to the region of interest and including overhang adapters used to amplify region of interest from genomic DNA (see Table 2.1). B) Subsequent limited-cycle amplification step used to add indices and Illumina sequencing adapters.

2.6 Library Preparation

The first round PCR products were purified using NucleoSpin® Gel and PCR Clean-up purification columns (Machery-Nagel) to remove any free primers and primer dimer species according to the manufacturers' instructions (Appendix 1). Two other purification systems were also tested these being AMPure XP-beads (Beckman Coulter) and ProNex® beads (Promega) according to the manufacturer's instructions (Appendix 1). The second round of PCR or 'Index' PCR using the Nextera XT index kit v2 Set A (Illumina) added the molecular identification (MID) tags (unique 8-nucleotide sequences) and Illumina MiSeq sequencing adapter to the first round PCR products using a 'dual indexing principle' (Figure 2.3B, Figure 2.4). This strategy used two 8-nucleotide indices, Index 1

(i7) adjacent to the P7 sequence, and Index 2 (i5) adjacent to the P5 sequence (Table 2.1, Figure 2.3). In this process a unique Index 1 and Index 2 are added to each first round PCR product on a 96-well plate (Figure 2.4, Appendix 1). After PCR products were purified with AMPure XP beads according to the manufacturer's instructions (Appendix 1). A random selection of PCR products were then size verified using a D1000 ScreenTape Assay on the 4200 TapeStation system (Agilent). The second round PCR products were then quantified using a Qubit Flourometer and the Qubit dsDNA HS Assay kit (ThermoFisher), normalized to 2 nM, and pooled in equimolar amounts to create one library for Illumina sequencing. The library pool was quantified using the KAPA Library Quantification Kit for Illumina Platforms (Roche) and the Applied Biosystems 7500 fast Real-Time PCR system. The library pool was also quantified using the Qubit dsDNA HS Assay kit and the Aglient 4200 TapeStation and Agilent High Sensitivity D1000 ScreenTape Assay (Agilent). A Qubit quantification measurement of 2.01 nM was used to adjust the concentration of library pool for sequencing. The amplicon library pool was diluted to 10 pM, spiked with 10 % PhiX Control v3 library (Illumina) and run on the Illumina MiSeq using a MiSeq Reagent Kit v2 500 cycle kit (Illumina), to generate 250-bp paired-end reads. PhiX DNA is derived from the small, well characterized bacteriophage PhiX genome. It is a concentrated Illumina library (10 nM in 10 µl) that has an average size of 500 bp and consists of balanced base composition at ~45% GC and ~55% AT and serves as an in-run QC for the Illumina sequencing.



A. 8 different Index 2 primers (white caps)
B. 12 different Index 1 primers (orange caps)
C. 96-well plate

Figure 2.4 Dual Indexing Principle for Illumina sequencing. Index 2 primers are added across the plate (arrows) and the Index 1 primers are added down the plate (dashed arrows) resulting in 96 separate combinations of primers.

2.7 Bioinformatic Processing

Data processing was performed on an Intel i7 PC running Ubuntu Linux 18.04.1 LTS. The program FLASH 1.2.11 (Fast Length Adjustment of SHort reads, Magnoc and Salzberg 2011) was used to convert paired end reads (R1 and R2 in the MiSeq platform) to a single merged read, using a minimum overlap length of 10 nucleotides (the default) and a maximum of 150 nucleotides to calculate the alignment. Reads were trimmed reading from the 5' end using trimmomatic 0.38 (Bolger, Lohse and Usadel 2014) to truncate the sequence if the average phred score of a 5nt sliding window dropped below 30. Those reads that matched the template specific primers at the 5' and 3' ends (maximum error rate of 0.1% within target specific primer site i.e. 2 bp variants allowed) and had a target region of >120bp were then pulled out of the data using Cutadapt 1.18 (Martin 2011). Degeneracy within the primer sequences was accounted for when identifying primer sequences within the dataset. Data was next converted from fastq to fasta format using seqtk-1.3 (r106) (github) (Seqtk, 2012).

Before taxonomic assignment standard Linux tools were used to identify 100% identical reads and condense them down to a single read to minimise time-consuming repetitive BLAST searches, however a record of the frequency of replicate sequences was maintained. Any reads with less than 3 replicates were excluded from the BLAST search.

A custom arthropod BLAST database was created from the BOLD database using the search term 'arthropoda' and 'COI' before downloading the records in FASTA format. Sequences marked as "SUPPRESSED" within the data were discarded, as were sequences where full taxonomic assignment was not available i.e. where only a genus or higher classification was present, or genus was present with "sp.". From a total of 4.21M sequences downloaded from BOLD, 1.83M sequences were included in the final database.

BLAST searching was performed using the "megablast" program which is optimised to identify alignments in highly similar sequences, and returned the top hit for each query sequence in a custom tabulated format. An e-value of 1e-15 was set; higher values such a 1 or 10 return a larger list of more low-scoring hits, and actual e-values returned were in the order of 1e-150 for a full length alignment.

A custom perl script filtered the BLAST output, identifying hits sharing an accession number and passing a set of criteria covering the percentage similarity between the query sequence and the database sequence (typically 99%), and having a query alignment length difference less than 6 bp. Note that ≥ 99% similarity indicates an approximately three-base difference between query and reference sequences because the maximum sequence length subjected to taxonomic assignment are around 300 bp. Read counts for each sequence passing the similarity and query alignment length filters were pooled based on accession number to generate a final frequency count for each accession. Taxonomic assignments were then compared to data provided by Natural England.

3. Results

3.1 DNA Extraction

DNA was extracted from all 67 sub-samples: 32 extracted from 'big' invertebrates (>7mm) and 35 extracted from 'small' invertebrates (<7mm). DNA quantification showed that only low concentrations of DNA had has been extracted (Appendix 1), however, these amounts were sufficient for PCR amplification.

3.2 PCR Amplification

Initial tests (data not presented) using both the mICOIintF/jgHCO2198 and HexCOIF4/HexCOIR4 primer combinations demonstrated that the mICOIintF/jgHCO2198 primer pair had a greater success rate at amplifying the DNA extracted from the different sub-samples than the HexCOIF4/HexCOIR4 primer pair. The mICOIintF/jgHCO2198 primer pair was therefore used for this study.

All 67 DNA extracts were amplified successfully, although DNA extract 3HPB (Highfield pitfall 3 'big') showed a lower level of amplification as compared to the other DNA extracts (Figure 3.1). Extraction blanks and PCR negative controls were all negative for amplification, positive control DNAs were successfully amplified.

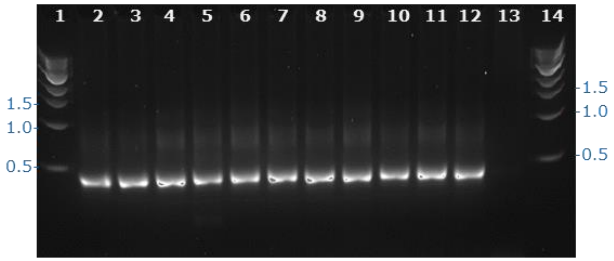
PCR products (of the expected size) were confirmed via agarose gel and Tapestation analysis of the PCR products (Figures 3.1 and 3.2, example shown in Appendix 2).

3.3 Library Preparation

Initial results suggested that the second round PCR or 'indexing' PCR was very inefficient. All primer sequences were cross checked against both the suppliers note to ensure that the sequence had been synthesised correctly and they were also checked by the University of Nottingham's Deep Seq team to ensure that primer design was correct. The primers passed all these checks. The suggestion that the degenerate primers might not be amplifying a COI sequence, was investigated by sequencing the PCR products generated from DNA extracted from *Esox Lucius* (Pike) and *Allolobophora chlorotica* (earthworm) DNA that was archived within the laboratory. Both PCR products were fully Sanger sequenced and confirmed the target amplicon as COI. In addition it demonstrated that the amplicons contained the correct PCR adaptors for the second round PCR. It was also discussed whether the Taq polymerase used was not suitable for generating the first round products or if the purification systems that had been used was somehow interfering with the indexing PCR.

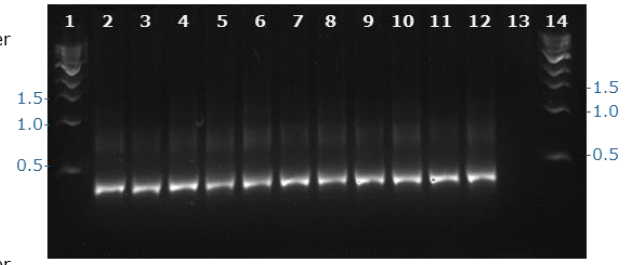
An extensive troubleshooting of the indexing PCR was carried out with different Taq polymerases and purification methods. In summary, three different high fidelity Taq

polymerases were tested namely: Phusion (Thermo Fisher Scientific); HiFi (PCR Biosystems); Q5 (New England Biolabs) none of which were successful in generating sufficient second round PCR product. These demonstrated that very little second round product (at ~ 541 bp) was produced and first round product (at ~437 bp) was still being detected when the reaction products were analysed using the Tapestation system (Figure 3.2). Three different purification methods were tested to purify the products of the round 1 PCR (AMPure XP Beads, ProNex® beads, and spin column purification). These extraction methods had no effect on getting the high fidelity DNA polymersases to work. Comparisons of a standard AmpliTaq Gold (Life Technologies), which had been used for the round 1 PCR, versus high fidelity Taq (Q5) and a real time PCR (SYBR green enzyme mix, Bio-Rad) was used to demonstrate whether the second round PCR could actually work (Figure 3.3). These results demonstrated that purified first round PCR products could not be amplified by the Q5 Taq polymerase (high fidelity), whereas the rtPCR master mix gave weak amplification. The Amplitaq gold polymerase used in round 1 gave good reaction products of a size consistent with the addition of the PCR adaptor sequences. Given these results, the indexing PCR was performed on a small number of first round PCR products using AmpliTaq Gold before confirming successful amplification and bands of the correct size using the Tapestation system (Figure 3.4). Once confirmation was achieved all first round PCR products were subjected to indexing PCR using AmpliTaq Gold before cleaning the PCR amplicons with AMPure XP beads and preparing them for Illumina sequencing. Illumina sequencing was carried out using a commercial service provided by the University of Nottingham's DeepSeq service resulting in a raw reads file which was uploaded to their server for access and downloading.



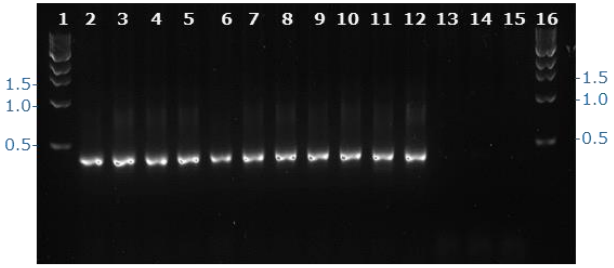
Gel 1: Top

1. Markers: 1 Kb Ladder
2. T1B
3. T2B
4. T3B
5. T4B
6. T5B
7. T6B
8. T7B
9. T8B
10. T9B
11. T10B
12. T11B
13. Neg 1
14. Markers: 1 Kb Ladder



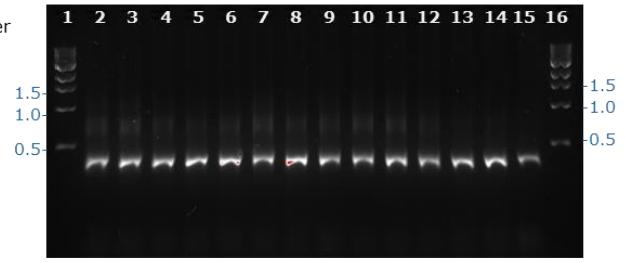
Gel 1: Bottom

1. Markers: 1 Kb Ladder
2. T1S
3. T2S
4. T3S
5. T4S
6. T5S
7. T6S
8. T7S
9. T8S
10. T9S
11. T10S
12. T11S
13. Neg 2
14. Markers: 1 Kb Ladder



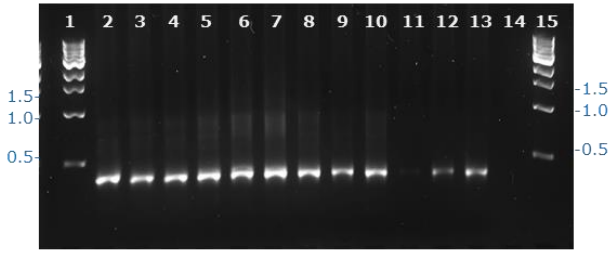
Gel 2: Top

1. Markers: 1 Kb Ladder
2. VT1B
3. VT2B
4. VT3B
5. VT4B
6. VT5B
7. VT6B
8. VT7B
9. VT8B
10. VT9B
11. VT10B
12. VT11B
13. EB1
14. EB2
15. EB3
16. Markers: 1 Kb Ladder



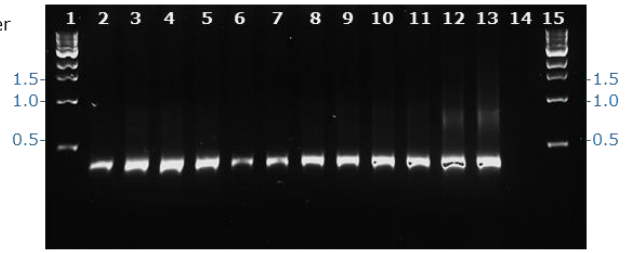
Gel 2: Bottom

1. Markers: 1 Kb Ladder
2. VT1S
3. VT2S
4. VT3S
5. VT4S
6. VT5S
7. VT6S
8. VT7S
9. VT8S
10. VT9S
11. VT10S
12. VT11S
13. Worm 1
14. Worm 2
15. Pike 1
16. Markers: 1 Kb Ladder



Gel 3: Top

1. Markers: 1 Kb Ladder
2. 1HHB
3. 3HHB
4. 4HHB
5. 1HHS
6. 2HHS
7. 3HHS
8. 4HHS
9. 1HPB
10. 2HPB
11. 3HPB
12. 4HPB
13. 4HCB
14. No Sample
15. Markers: 1 Kb Ladder



Gel 3: Bottom

1. Markers: 1 Kb Ladder
2. 1HPS
3. 2HPS
4. 3HPS
5. 4HPS
6. 1HSB
7. 2HSB
8. 1HSS
9. 2HSS
10. 3HSS
11. 4HSS
12. HDB
13. HDS
14. No Sample
15. Markers: 1 Kb Ladder

Figure 3.1 First round PCR results

Successful amplification of all 67 DNA extracts: 5 µL PCR product loaded per well; 3 µL 1 Kb Ladder loaded. Extraction blanks (EB1-3) and PCR negative controls (Neg 1-2) were negative for amplification. Positive control DNAs also successfully amplified.

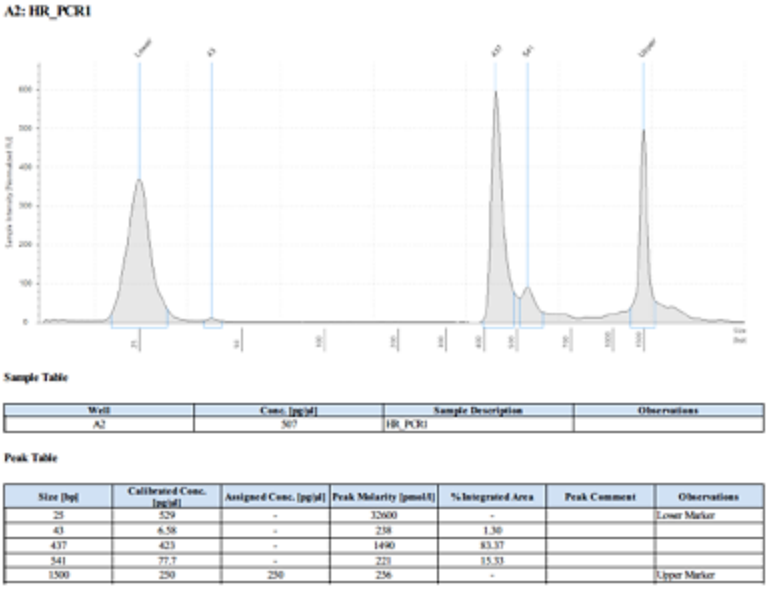
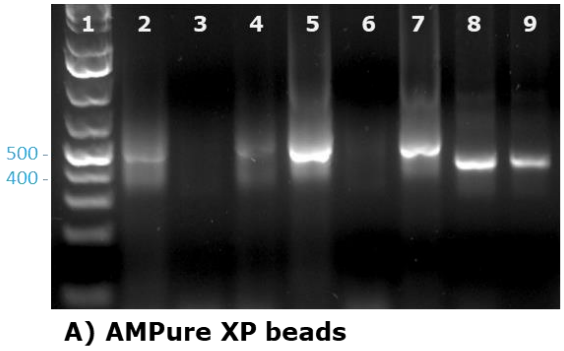
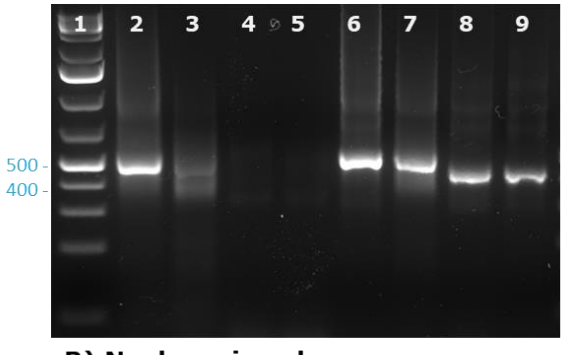


Figure 3.2 Example Tapestation readout. Indexing PCR result for first round PCR product 1HHS using high fidelity Taq. A large peak at 437 bp representing the first round PCR product can be seen along with a very small peak at 541 bp which represents a very small amount of indexing PCR product which is not sufficient for Illumina sequencing.



A) AMPure XP beads
1: 1Kb DNA ladder
2: T2S – AMPure Clean – Sybr
3: T2S – AMPure Clean – Q5 taq
4: T2S – AMPure Clean – AmpliTaq Gold
5: T2S – Non-cleaned – Sybr
6: T2S – Non-cleaned – Q5 taq
7: T2S – Non-cleaned – AmpliTaq Gold
8: T2S – Rd 1 PCR amplicon – Non-cleaned
9: T2S – Rd 1 PCR amplicon – AMPure cleaned



B) Nucleospin column
1: 1Kb ladder
2: VT6S - Sybr green
3: VT6S - Sybr green – Column clean up
4: VT6S - Q5 Taq
5: VT6S - Q5 Taq – Column clean up
6: VT6S - AmpliTaq Gold
7: VT6S AmpliTaq Gold – Column clean up
8: VT6S - Rd 1 PCR product
9: VT6S - Rd 1 PCR product – Column clean up

Figure 3.3 Agarose gel of indexing PCR products. Indexing PCR products (lanes 2 to 7) when amplified with different Taq polymerases (SYBR green, Q5, AmpliTaq Gold) when either purified with A) AMPure XP beads (AMPure Clean) or B) Nucelospin® column purification (column clean up) versus unpurified (non-cleaned) and round one PCR amplicons (lanes 8 and 9) to illustrate the size shift that should occur with the indexing PCR.

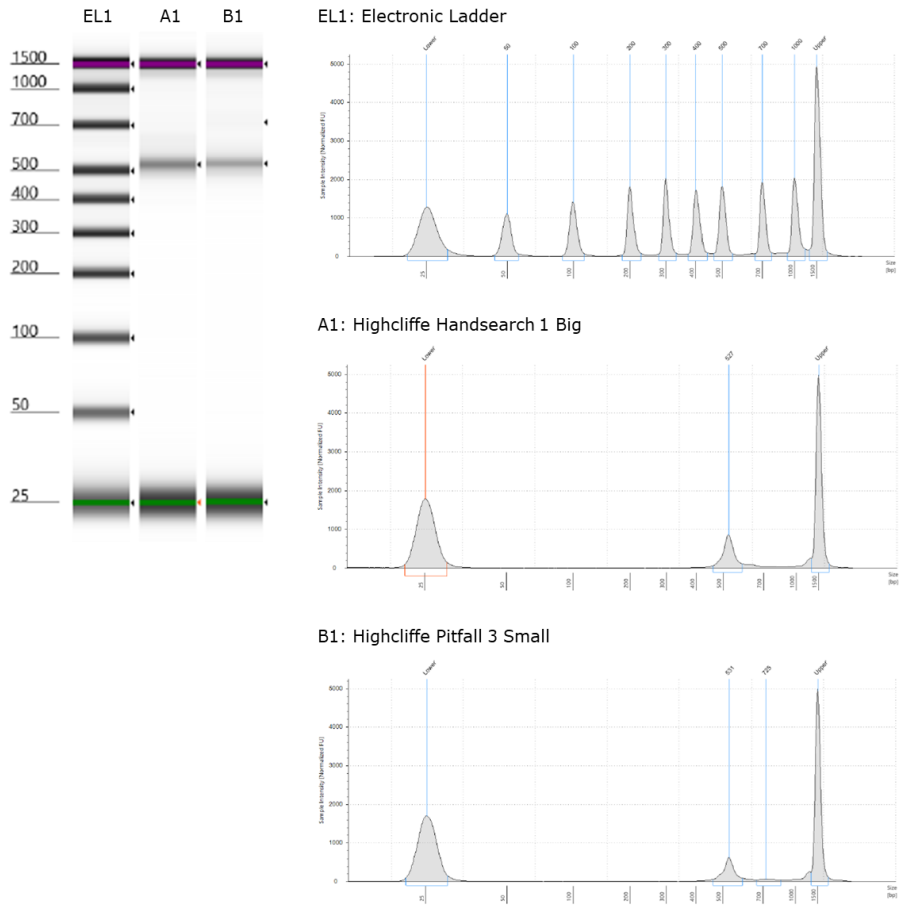


Figure 3.1 Tapestation result showing successful indexing PCR. Successful amplification of two first round PCR products subjected to indexing PCR using AmpliTaq Gold illustrating correct band size (~530 bp).

3.4 Bioinformatics and Data Analysis

A total of 27.2M raw reads (13.6M read pairs) assigned to sub-sample barcodes were returned from sequencing. A further 635K raw reads were attributable to control samples and finally 4.28M raw reads could not be assigned to a barcode. The mean number of raw read pairs per sub-sample barcode was 203K, ranging from 1047 reads for sub-sample Highcliffe Pitfall 3 Big to 342K for sub-sample Highcliffe Diptera Small. After bioinformatics processing to convert paired end reads to a single merged read, trimming these, and identifying those that contained the target specific primer site around 20% of the raw read pairs went onto taxonomic assignment.

As an example, for sub-sample T3 Big Sherwood FE, the reduction in read number relative to the raw read count is shown for each processing step:

- Raw read pairs: 169,675 (100%)
- Flash merged: 168,074 (99.06%)
- Trimmomatic filtered: 111,136 (65.49%)
- Cutadapt 5': 69,881 (41.18%)

- Cutadapt 3': 38,728 (22.82%)
- seqtk (fastq to fasta): 38,728 (22.82%)
- ≥ 3 replicate reads: 30,301 (17.85%) representing 1,488 unique sequences

Overall, a total of 1,930,532 sequences were assigned a taxonomic identification which represented 279 species. This corresponded to 79 species for Sherwood FE Birklands, 61 species for Sherwood NCC, 103 species for Highcliffe beetles, 49 species for Highcliffe spiders, and 17 species for Highcliffe flies. There are a number of species not found on the BOLD database which if added would likely increase the number of species found in each of the sampling areas.

After taxonomic assignment the data was compared to information as supplied by Natural England and numerous tables were produced to illustrate the community composition of each individual sample as found by taxonomic identification, metabarcoding of 'small' sub-samples, and metabarcoding of 'big' sub-samples (Appendices 3-7). The following alterations to the data were made prior to further analysis: 1. All species which do not appear on the BOLD database were removed as they would never be found by metabarcoding; 2. All species that were effectively removed from the initial sample by being retained by Natural England were removed from the overall species list; and 3. All species which were not in the order of interest which for example may be potential prey species were removed from the overall species list. This revised data was then used to create bar charts of the percentage of total individuals per species (taxonomic identification) and the percentage of total read counts per species (metabarcoding, small and big data pooled). Resulting data was plotted side-by-side for each individual sample for Sherwood FE Birklands, Sherwood NCC, Highcliffe beetles, Highcliffe spiders, and Highcliffe flies samples respectively (Figure 3.5 to 3.9). Results illustrate that within sampling locations (Highcliffe, Sherwood FE, and Sherwood NCC) and within methods there were significant differences in species composition at each sampling site. Additionally, different species profiles were obtained for taxonomic identification when compared to metabarcoding for each of the samples.

To investigate the relatedness of taxonomic identification versus sample metabarcoding, Venn diagrams were plotted (Figure 3.10). These results indicate that a proportion of species are found by both methods ranging from 39% in Highcliffe beetles to 63% for spiders, and that each method identifies species within the same order that are not found by the other which could illustrate potential taxonomic misidentification or errors within the BOLD database (Table 3.1).

Table 3.1 Summary of findings

	All species	**Beetles**	**Spiders**	**Flies**
Total number of species found (both methods)	448	359	64	30

Taxonomy	384	318	41	24
Metabarcoding	228	202	49	17
Taxonomy (BOLD/retained/non order removed)	337	279	40	18
Metabarcoding (non order removed)	246	193	36	17
Revised total number of species found (both methods – BOLD/retained/non-order removed)	379	309	46	24
Species found by both methods	186	146	29	11
Taxonomy not metabarcoding	152	134	11	7
Metabarcoding not taxonomy	41	29	6	6
Species not found on BOLD database	30 (7.8%)	22 (6.9%)	3 (7.7%)	6 (25%)
Species retained by Natural England	28 (7.4%)	28 (8.8%)	0	0
Species removed as non order	45 (11.9%)	30 (9.4%)	15 (36.6%)	0

The total number of species row, is the grand total of species found by both methods added together. The taxonomy row indicates the total number of species found before removal of any species. The taxonomy (BOLD/retained/non order removed) row indicates the number of species left after the removal of those species which do not appear on the BOLD database; those which were retained by Natural England; and those which were not in the order of interest. The metabarcoding row indicates the number of species found by metabarcoding. The revised total number of species found is the total number of species found by both methods added together minus those not on the BOLD database, retained by Natural England or non order species. The species found by both methods row indicates the number of species found by both taxonomic identification and metabarcoding. The taxonomy not metabarcoding row indicates the number of species found by taxonomy and not metabarcoding (after removal of species from data). The metabarcoding not taxonomy row indicates the number of species not found by taxonomic identification (after removal of species). The species not found on BOLD database indicates the number and percentage of all species found by taxonomic identification not found on the BOLD database. The species retained by Natural England row indicates the

number of species retained by Natural England at the taxonomic identification stage. The species removed as non order row indicates the number of species which are not within the order of interest) beetles, spiders, flies).

To determine whether there was any correlation between numbers of individuals and the number of sequence reads that are assigned to each species, plots were generated for the both total beetles and total spiders numbers (at the individual species level) plotted against sequence read number. The individual plots for beetles, spiders, and flies is summarised in figure 3.11 A-C respectively. The mean sequence number plotted against individual numbers (beetles) is plotted in figure 3.11D. These figures illustrate very poor relationship between number of individuals and the number of sequences found (across all species identified).

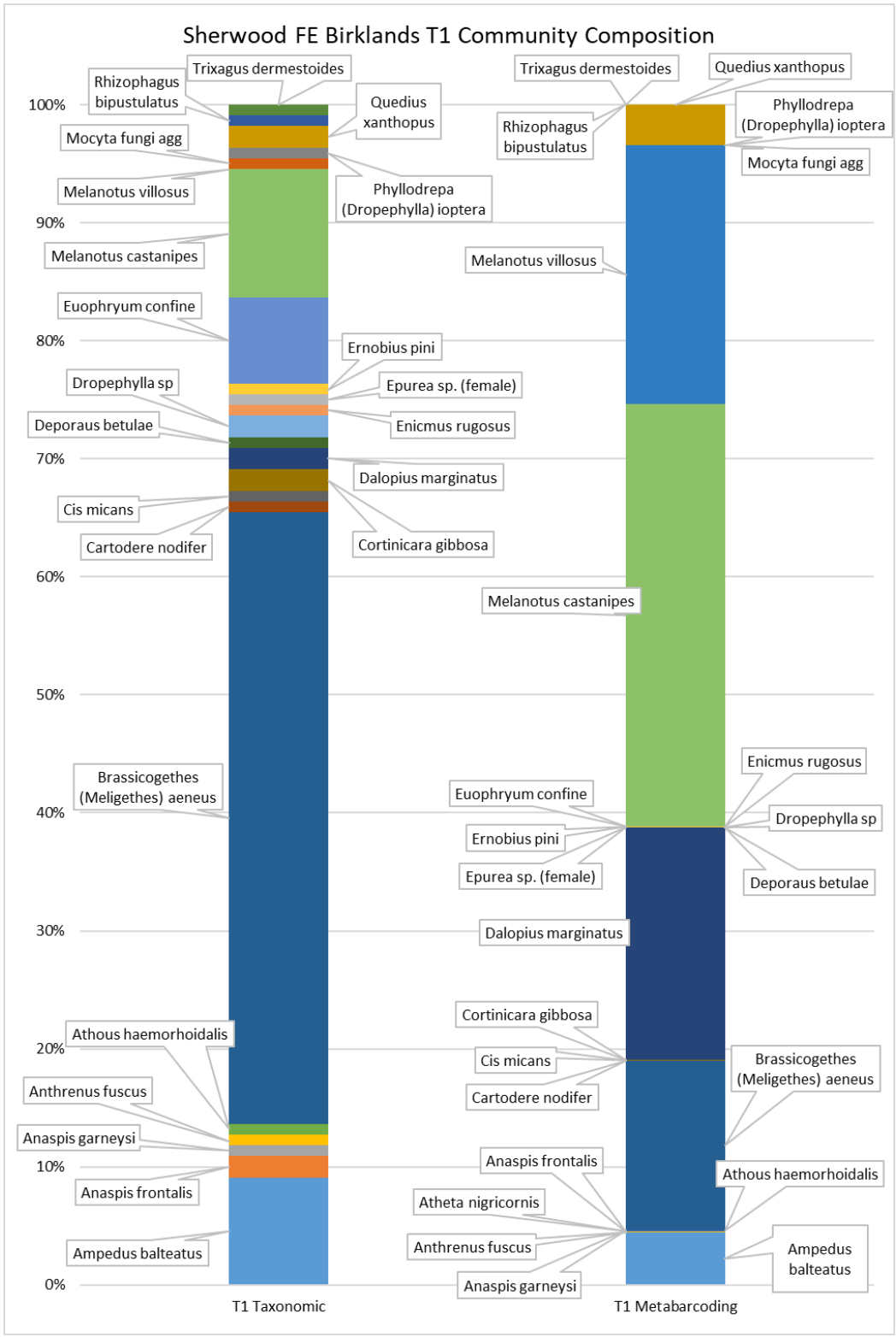


Figure 3.2 Community composition of Sherwood FE Birklands sample T1. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

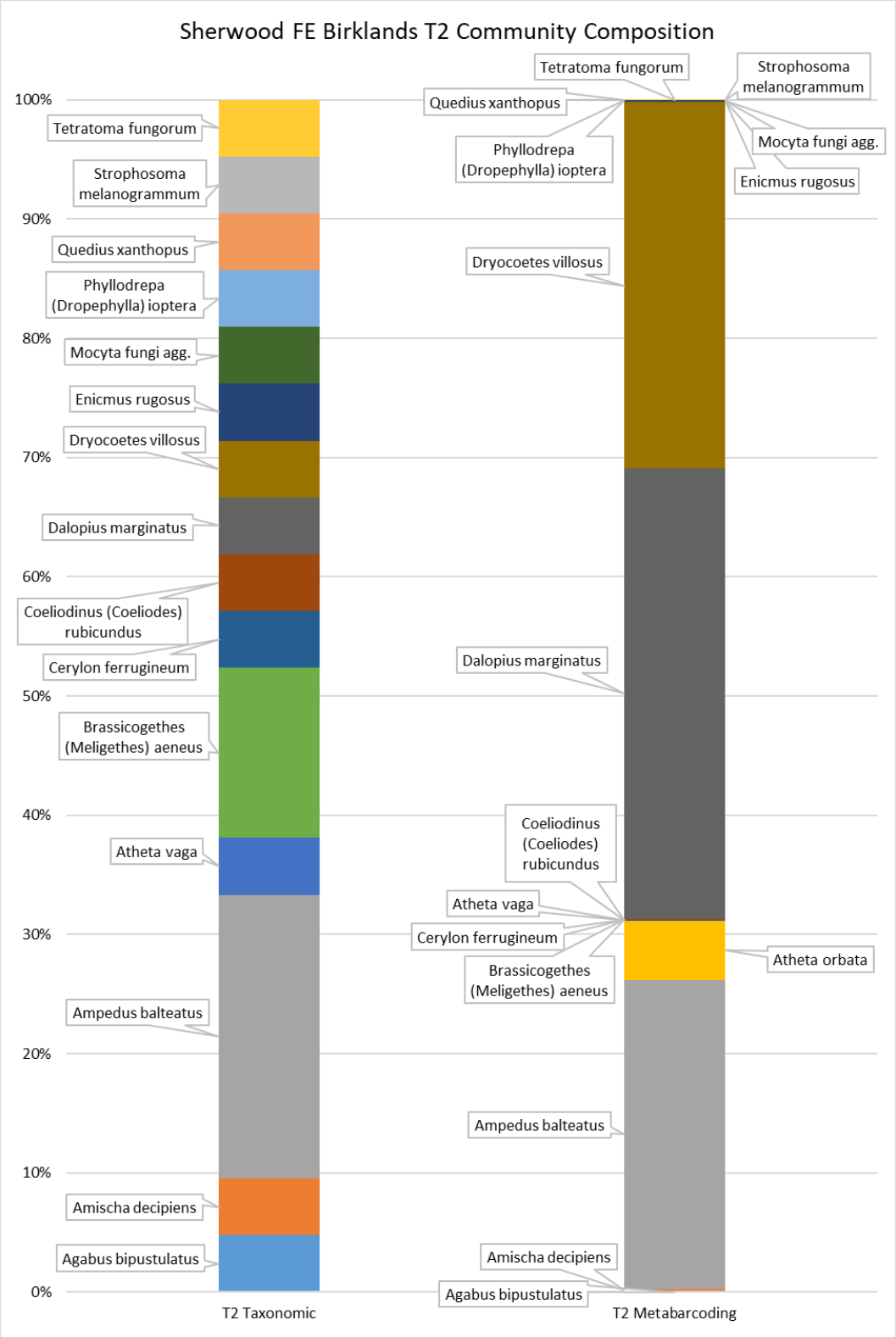


Figure 3.3 Community composition of Sherwood FE Birklands sample T2. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

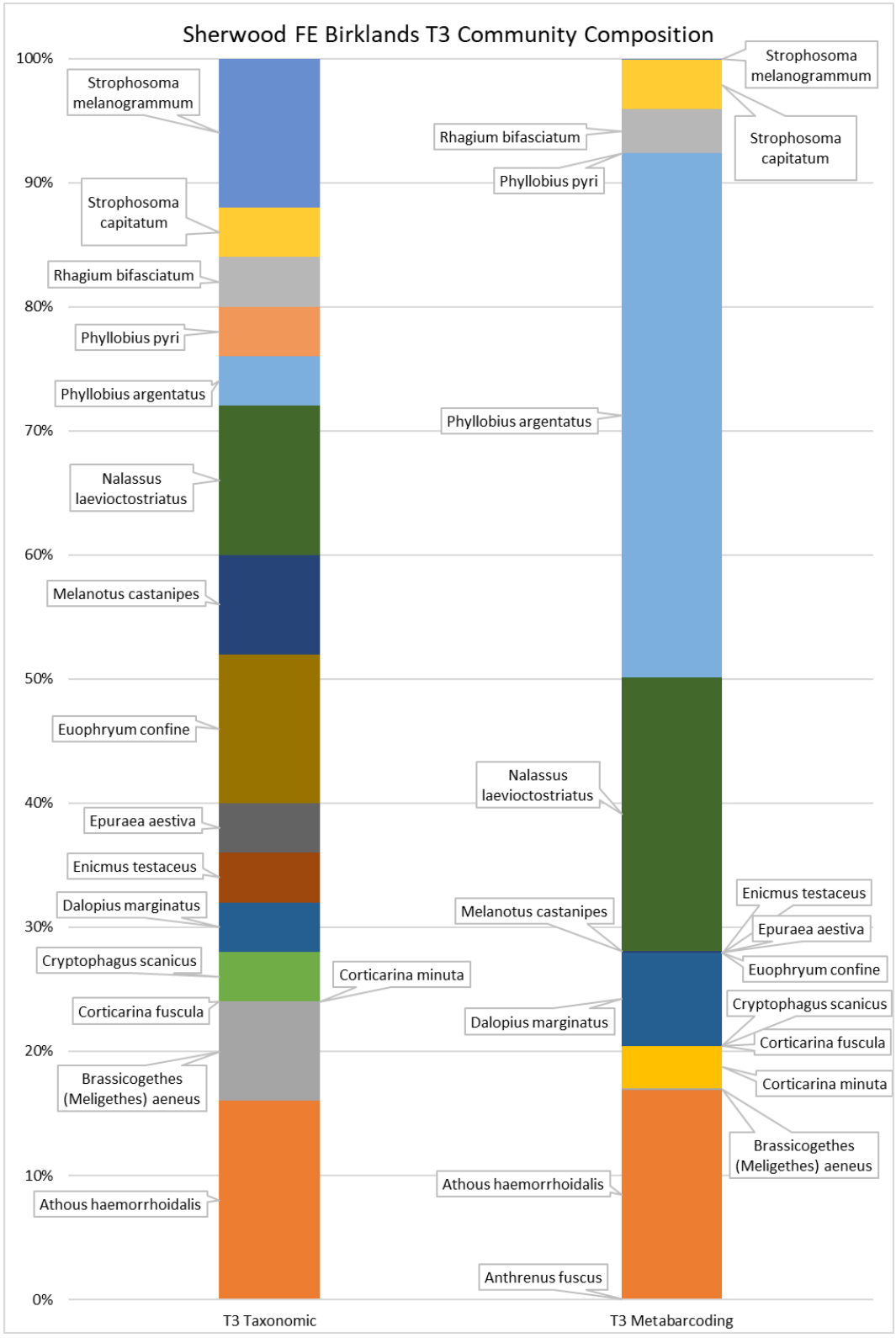


Figure 3.4 Community composition of Sherwood FE Birklands sample T3. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation**.**

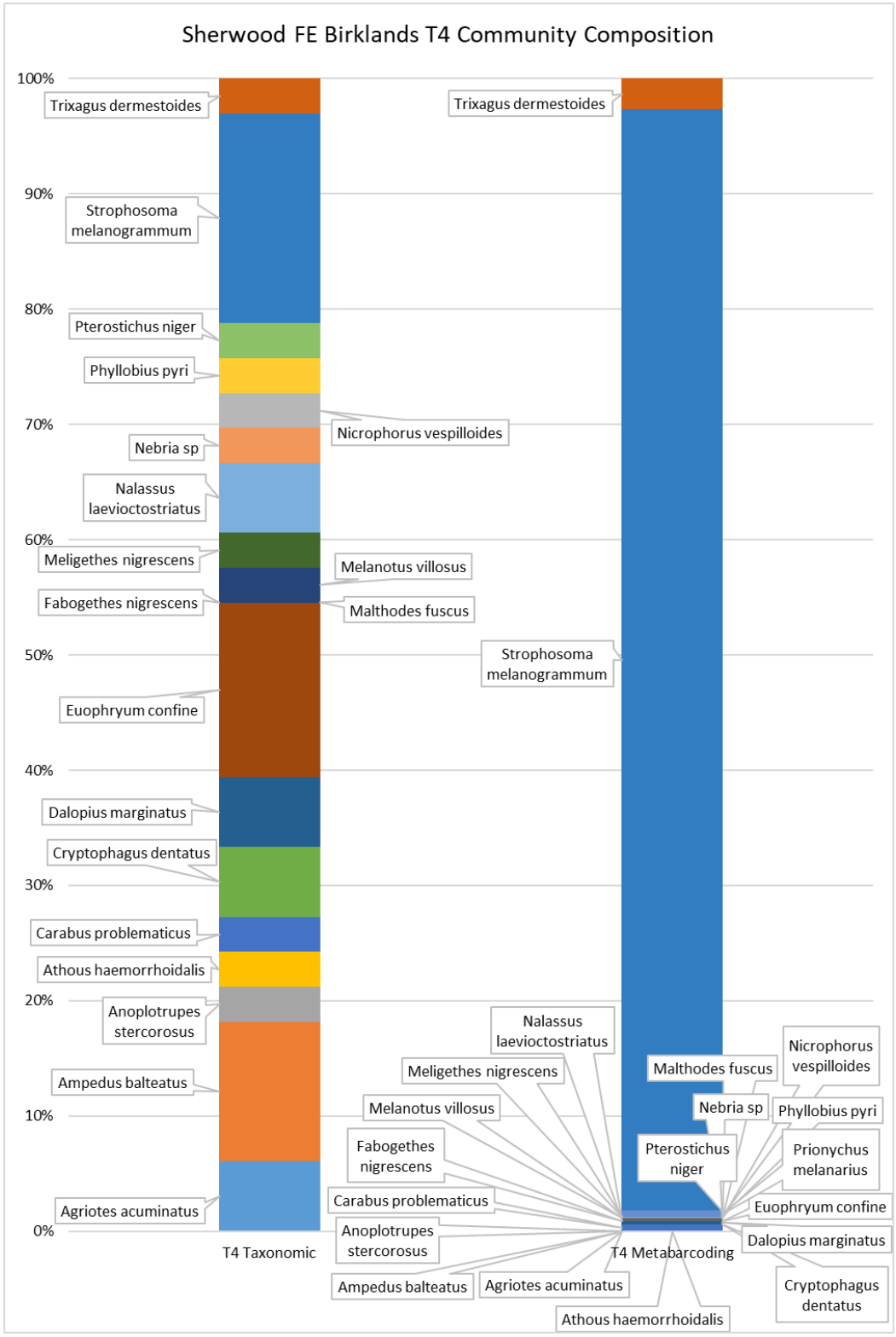


Figure 3.8 Community composition of Sherwood FE Birklands sample T4. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

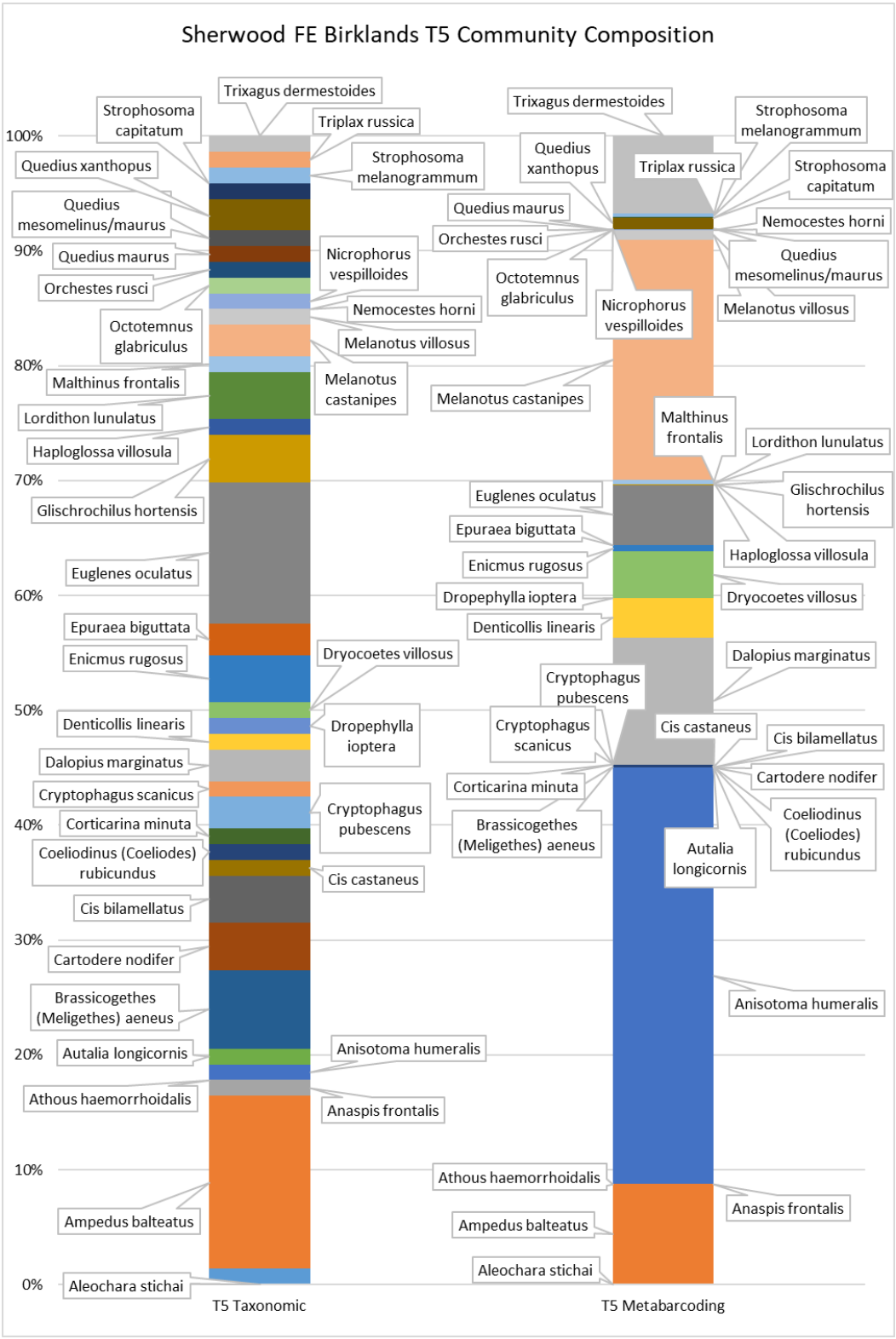


Figure 3.5 Community composition of Sherwood FE Birklands sample T5. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

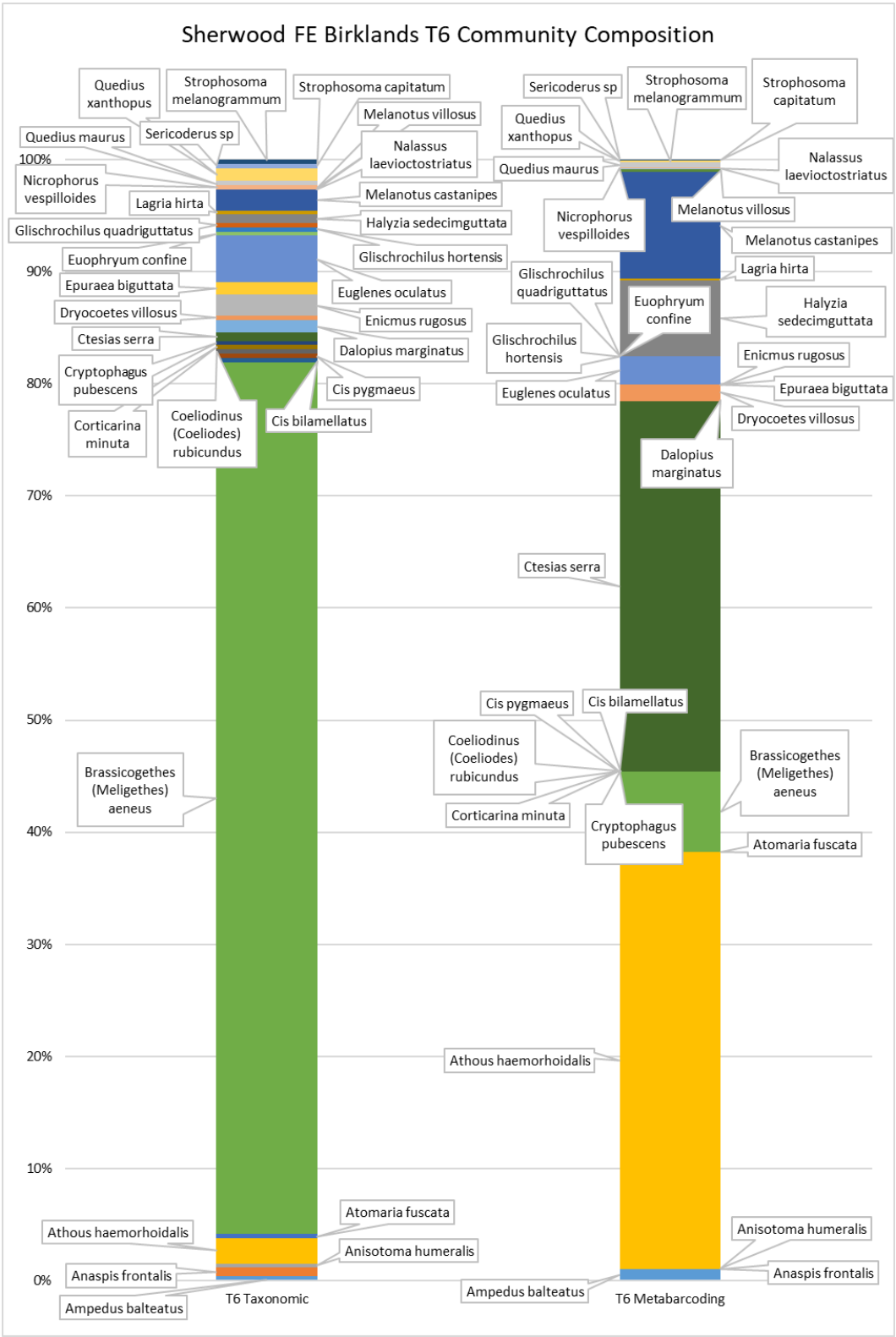


Figure 3.6 Community composition of Sherwood FE Birklands sample T6. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

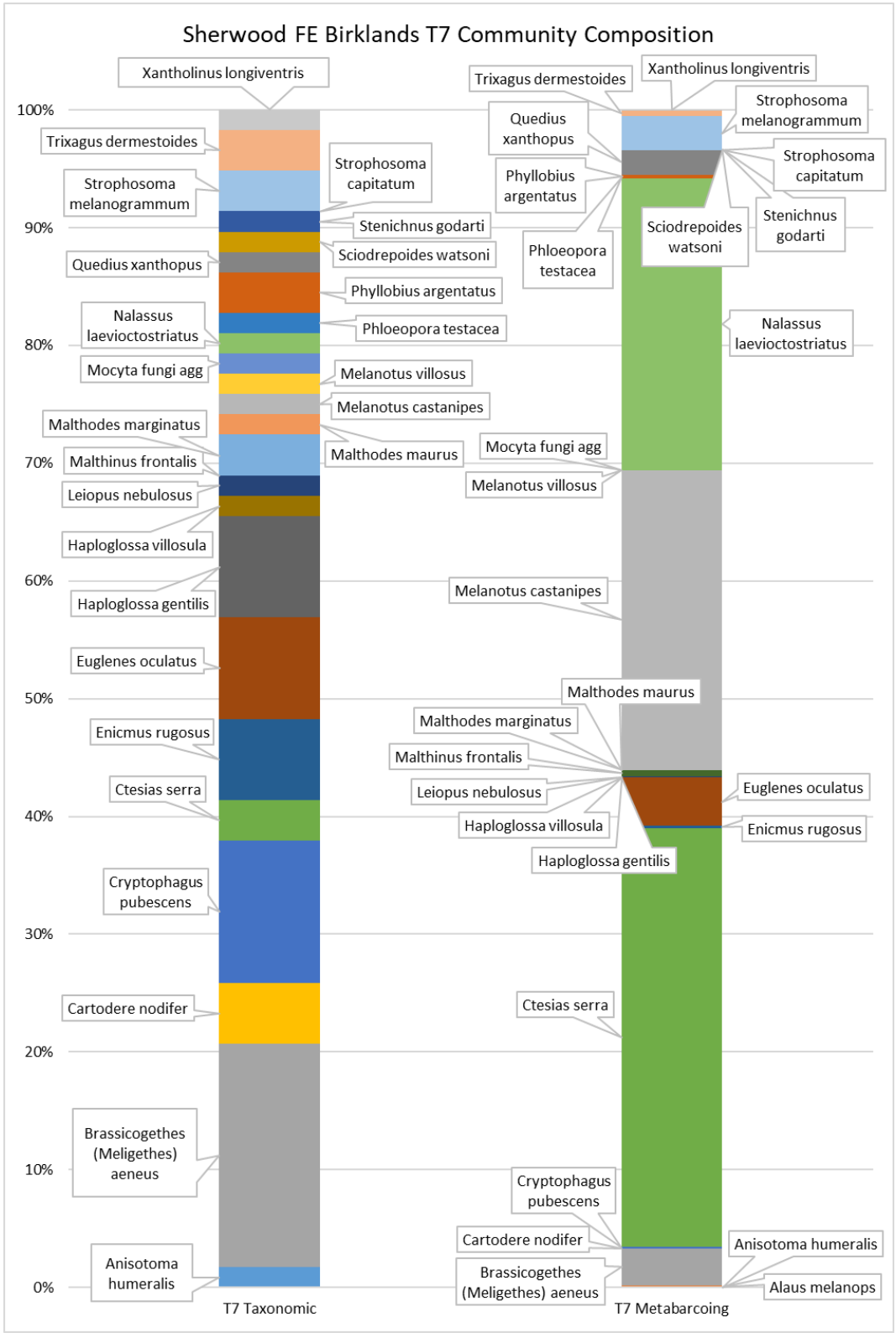


Figure 3.7 Community composition of Sherwood FE Birklands sample T7. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

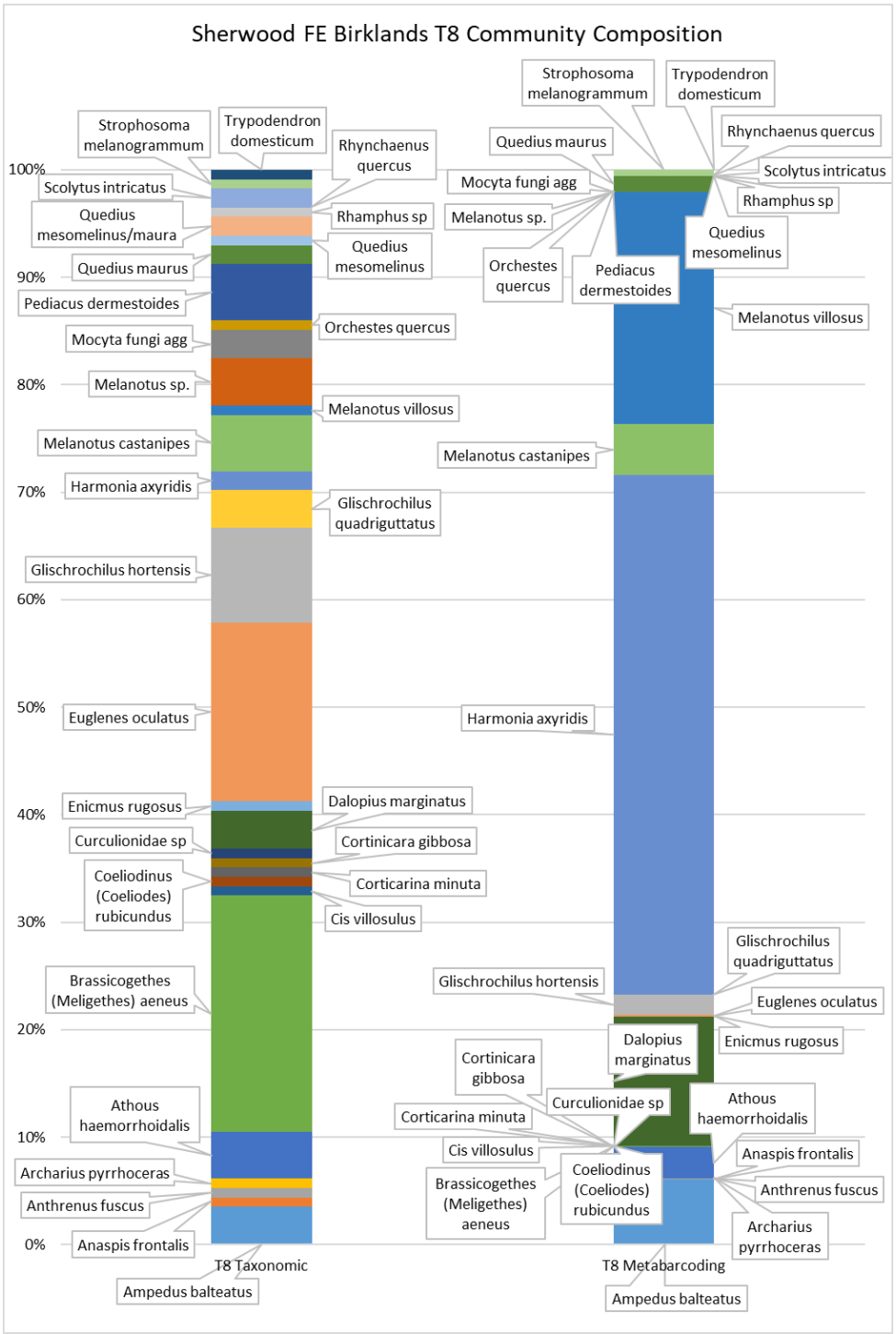


Figure 3.8 Community composition of Sherwood FE Birklands sample T8. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). ‘Taxonomic’ refers to those species that were hand identified and ‘metabarcoding’ refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

Figure 3.9 Community composition of Sherwood FE Birklands sample T9. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

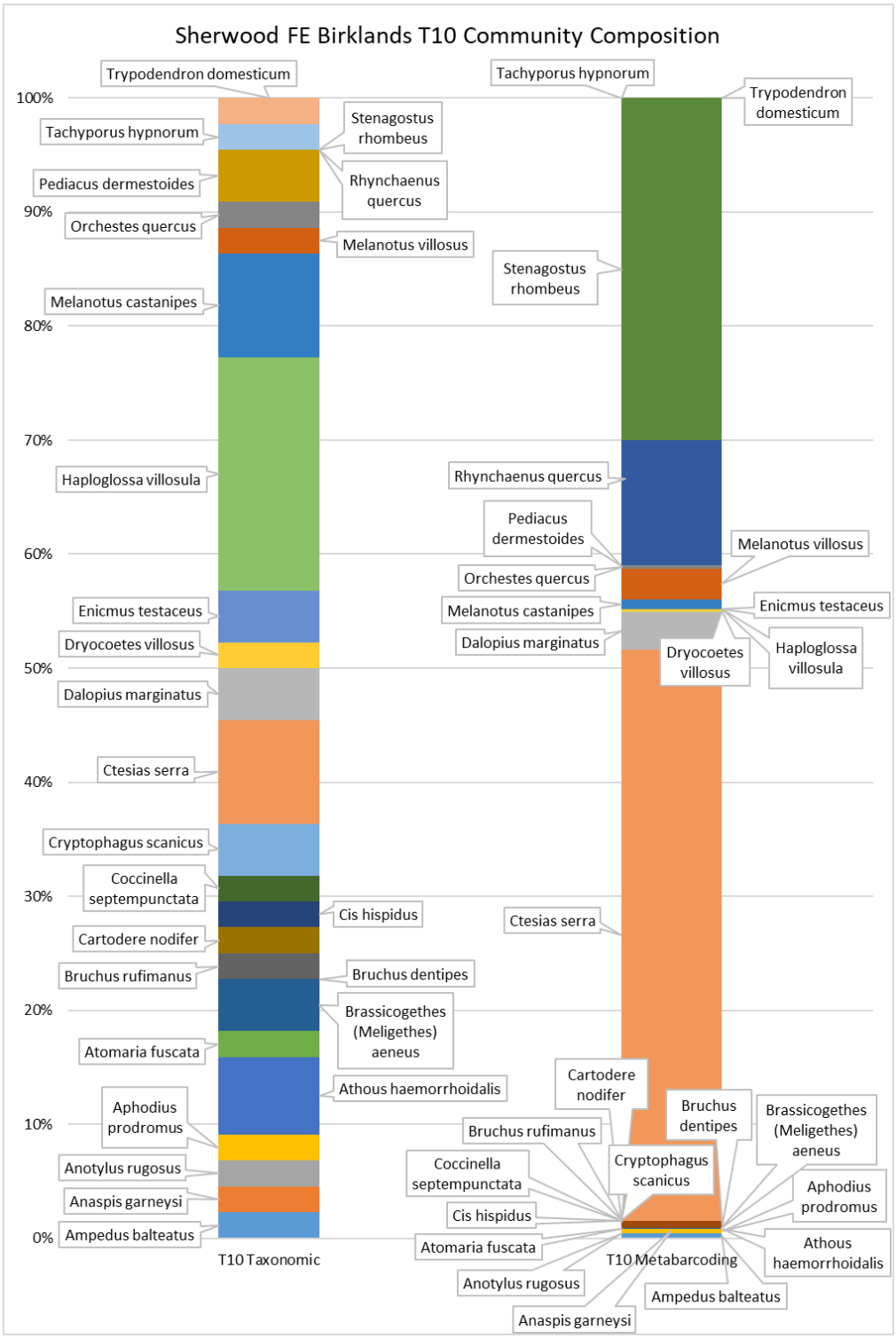


Figure 3.10 Community composition of Sherwood FE Birklands sample T10. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

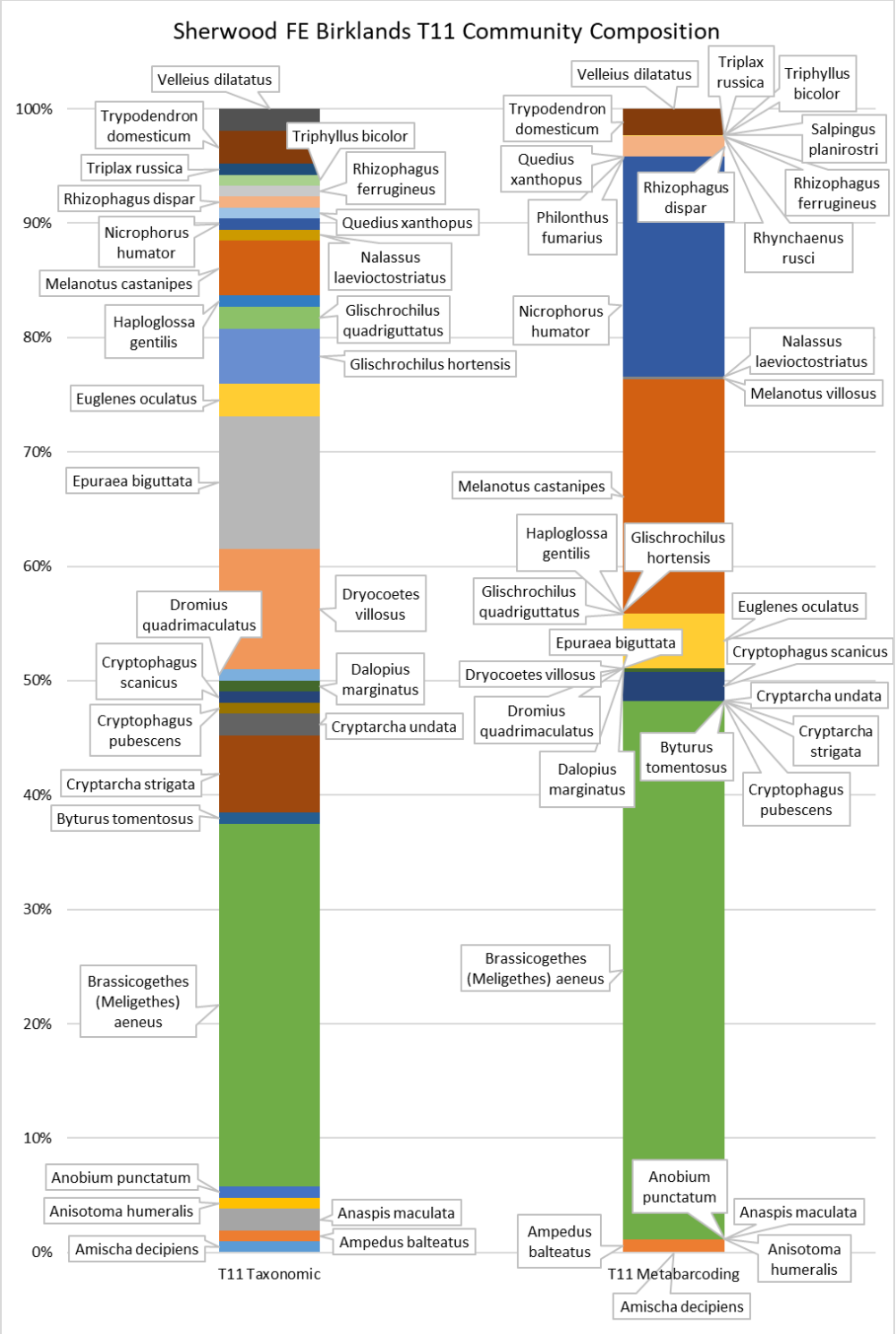


Figure 3.11 Community composition of Sherwood FE Birklands sample T11. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

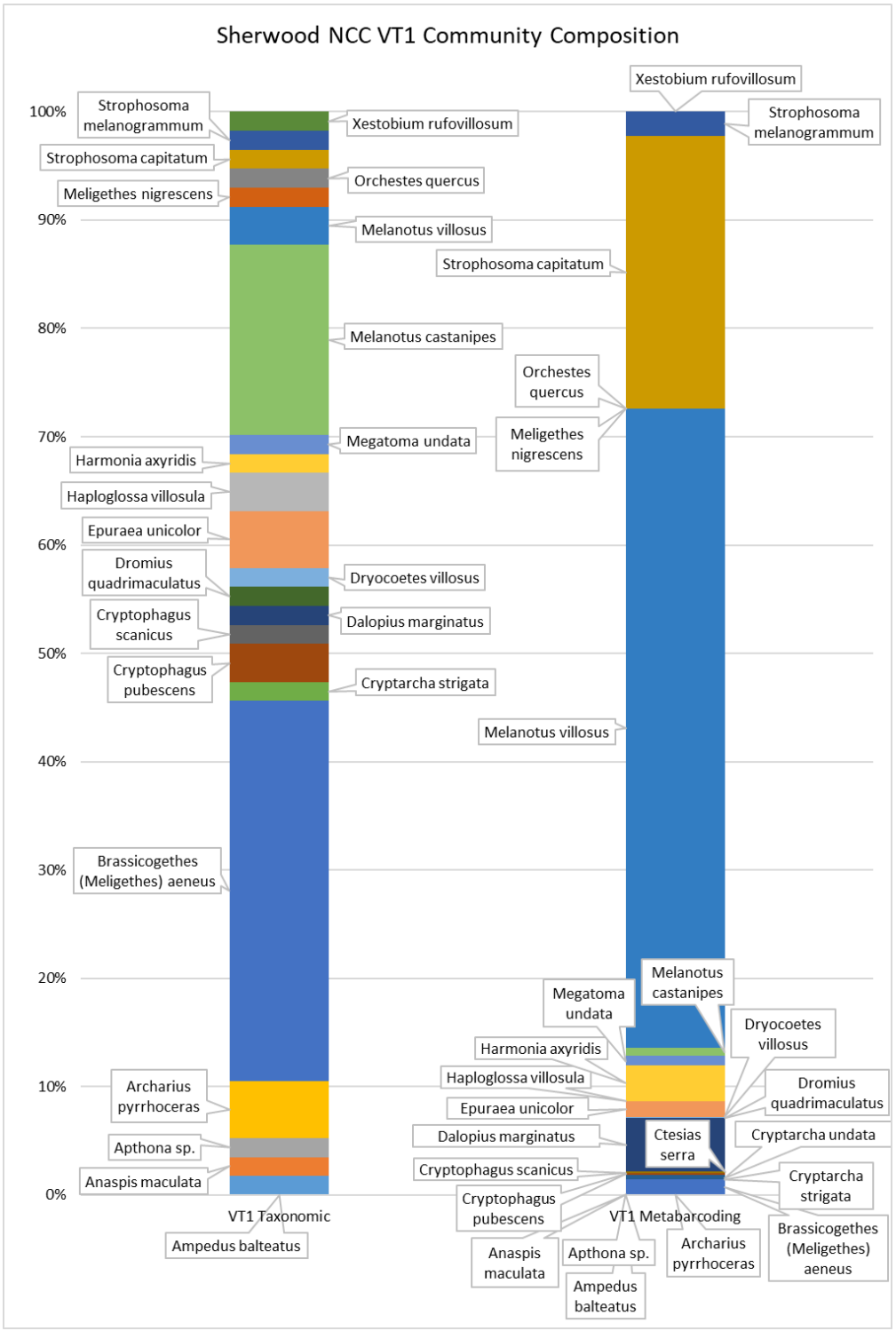


Figure 3.12 Community composition of Sherwood NCC sample VT1. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

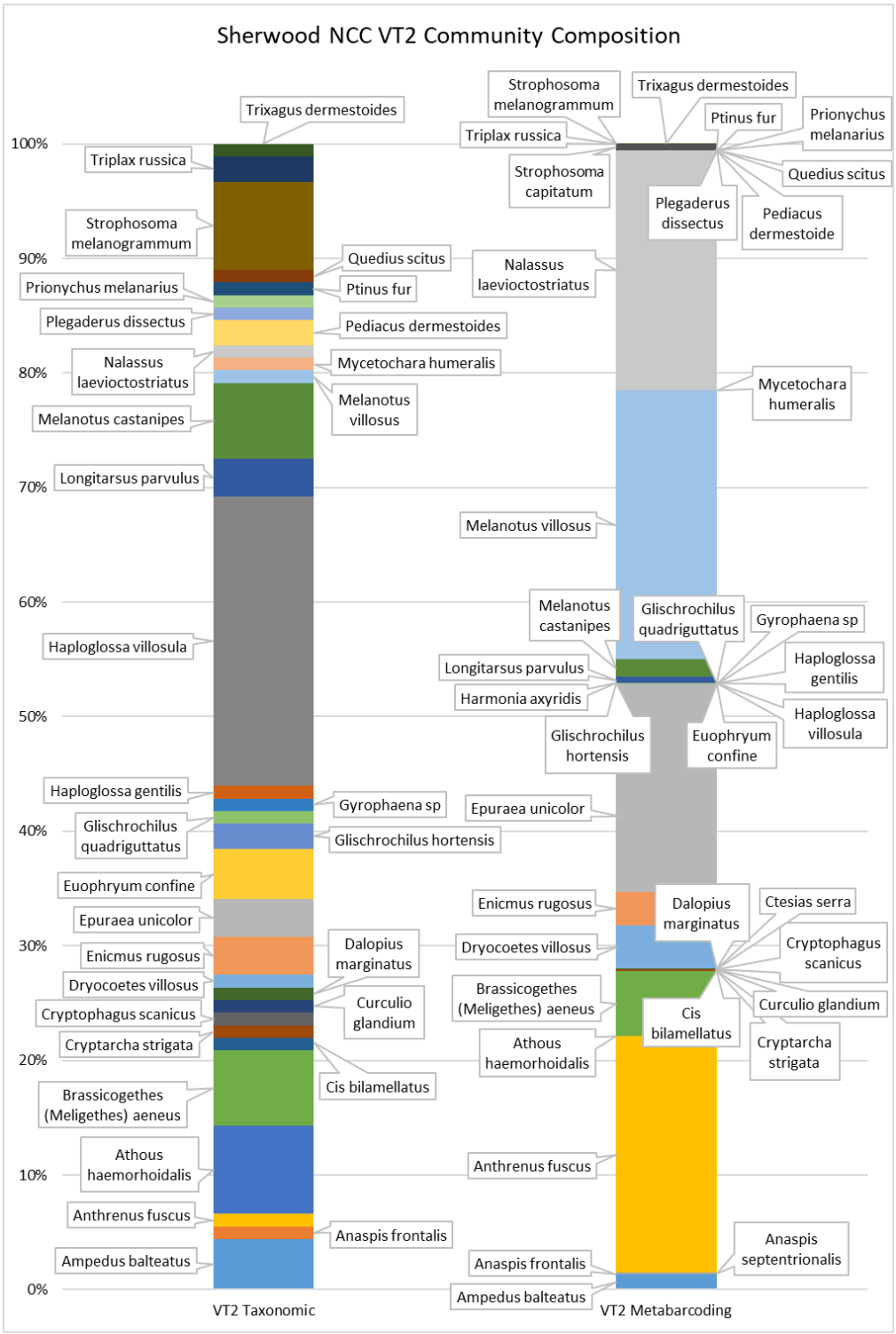


Figure 3.13 Community composition of Sherwood NCC sample VT2. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

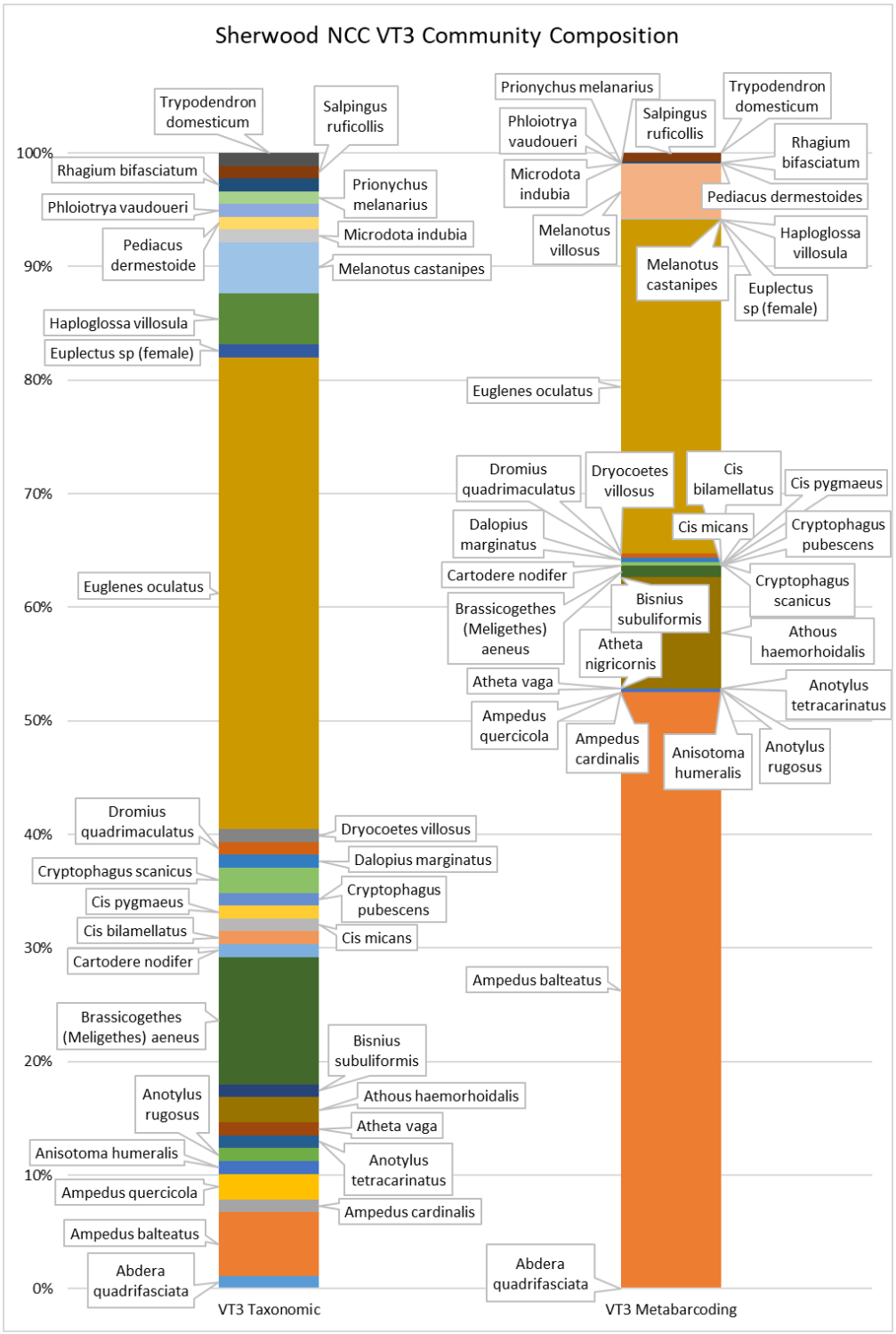


Figure 3.14 Community composition of Sherwood NCC sample VT3. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

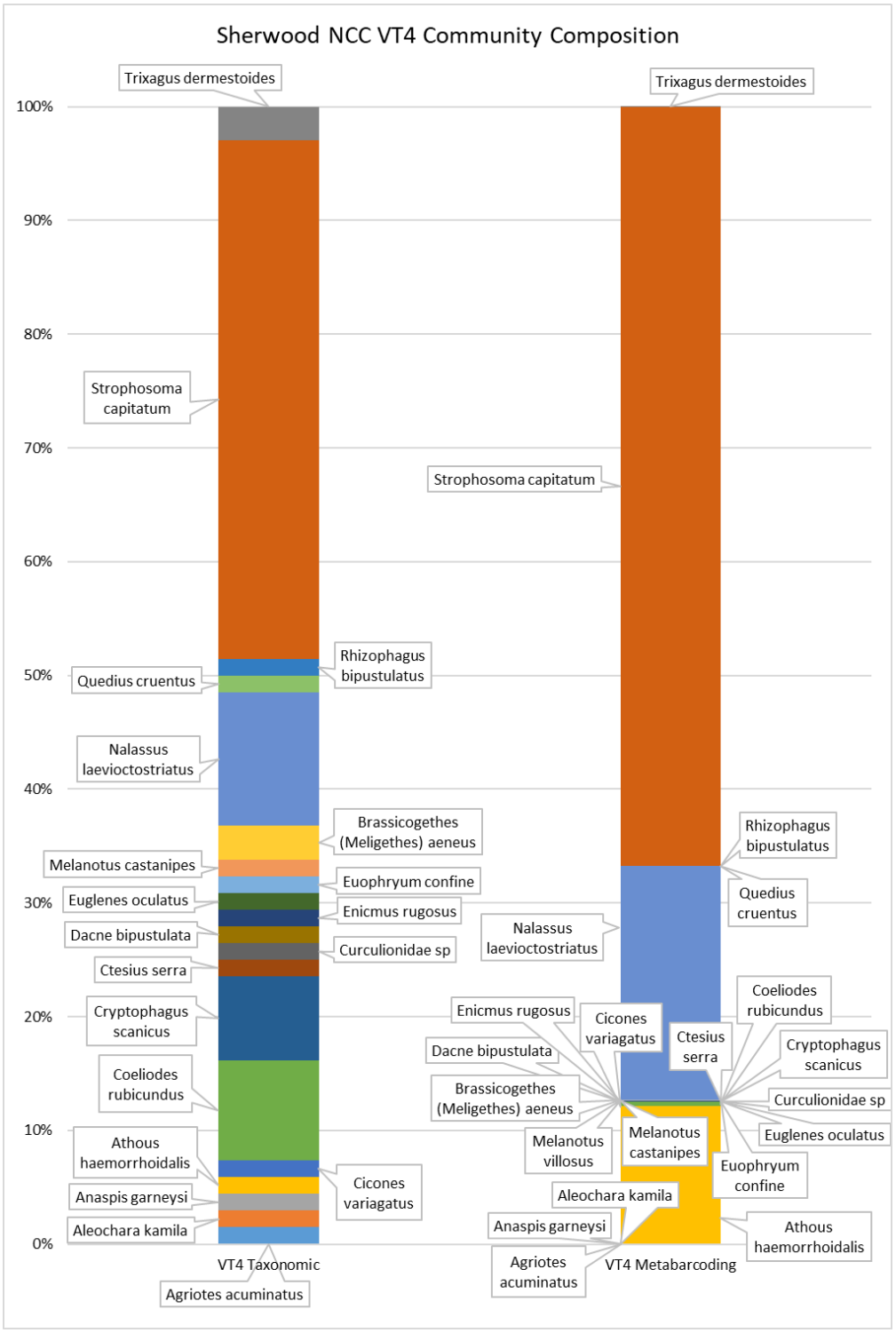


Figure 3.15 Community composition of Sherwood NCC sample VT4. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

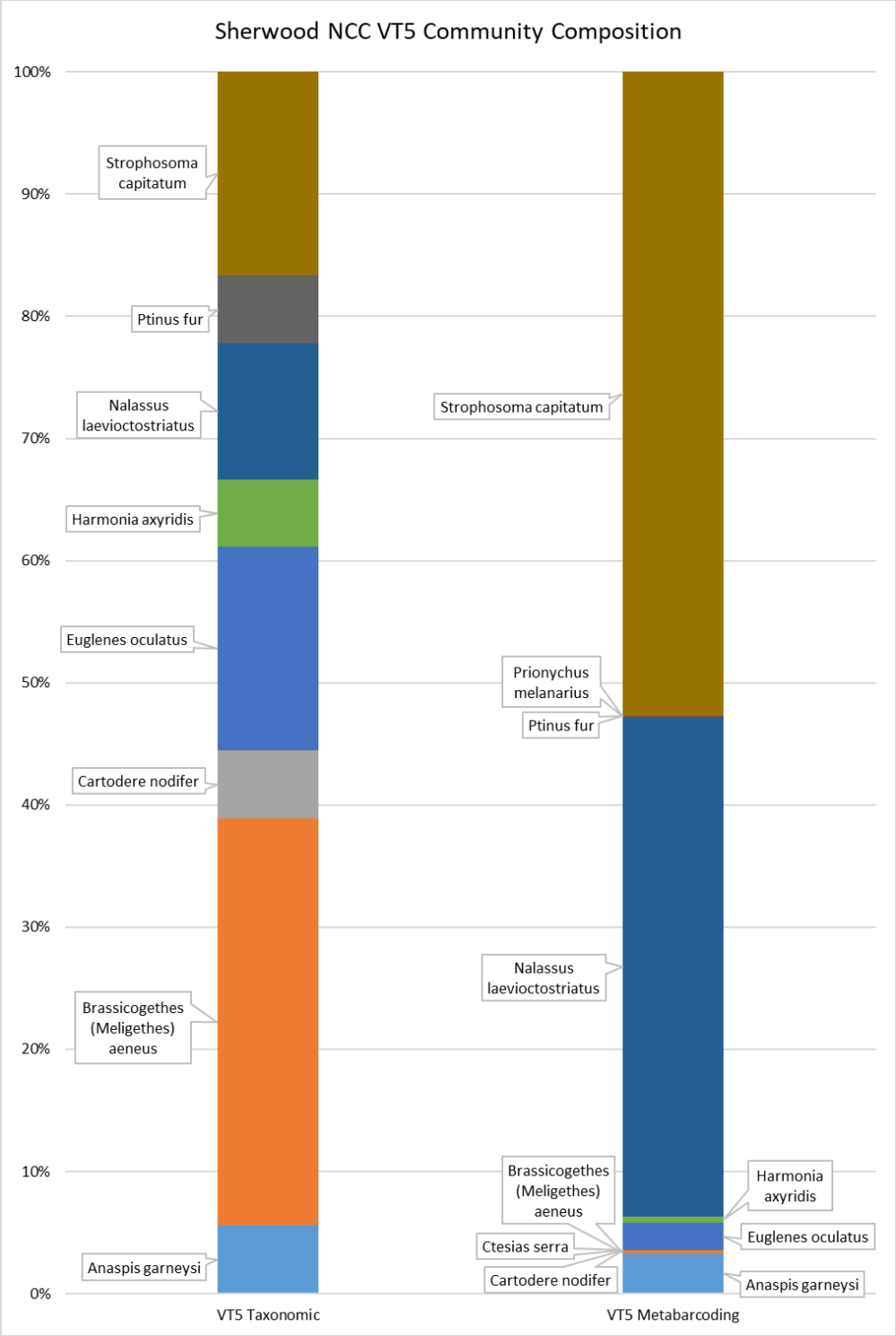


Figure 3.16 Community composition of Sherwood NCC sample VT5. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

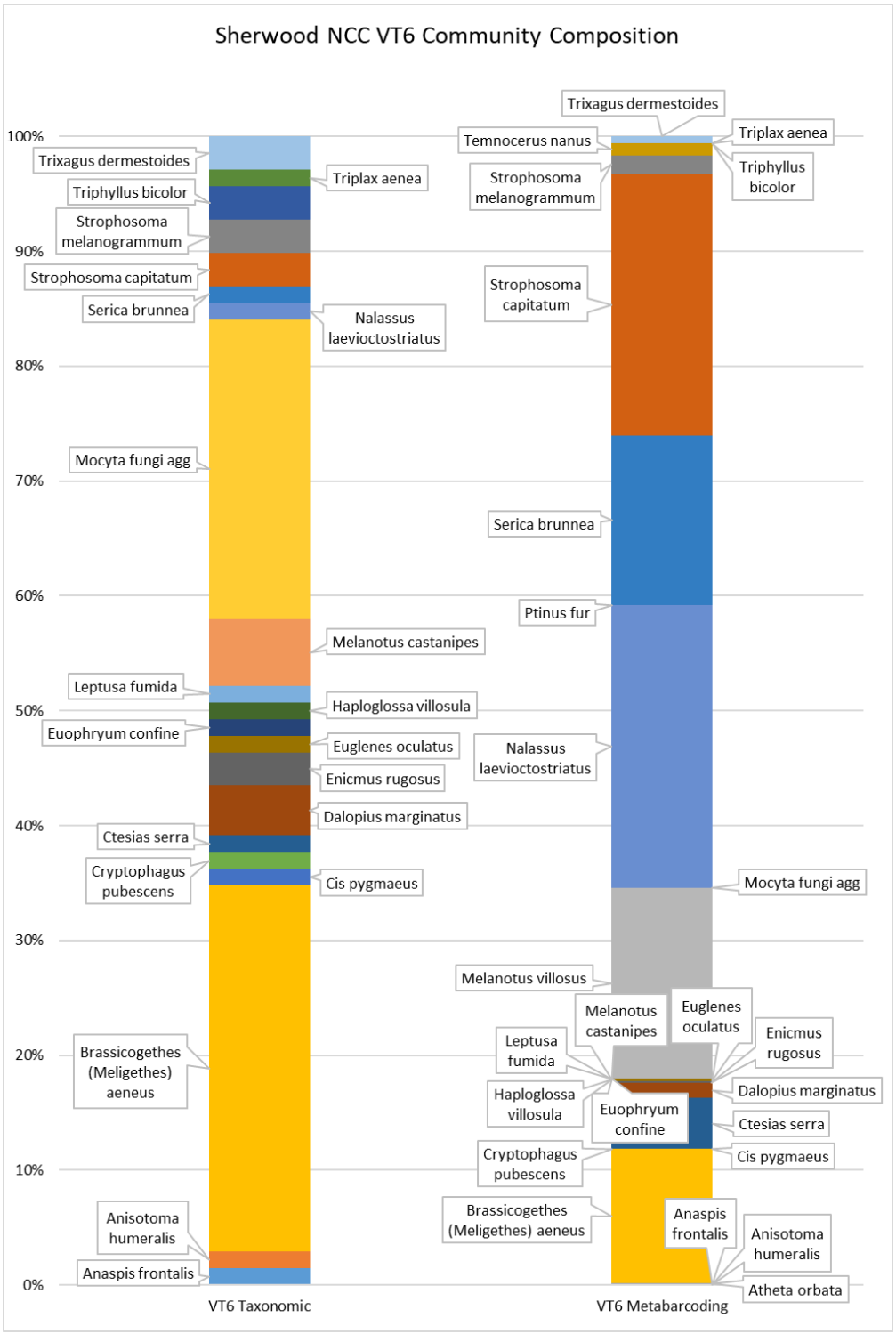


Figure 3.17 Community composition of Sherwood NCC sample VT6. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

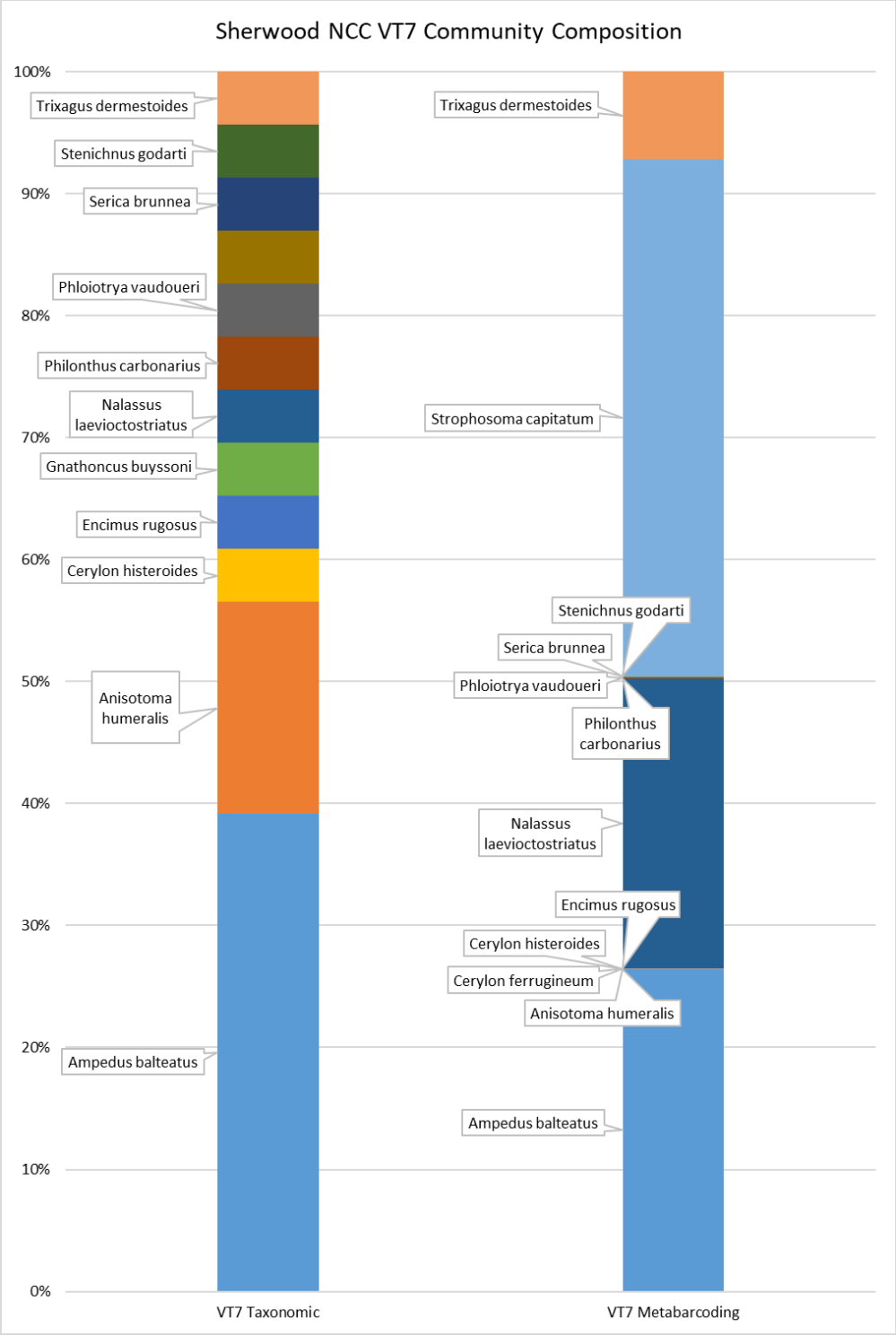


Figure 3.18 Community composition of Sherwood NCC sample VT7. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

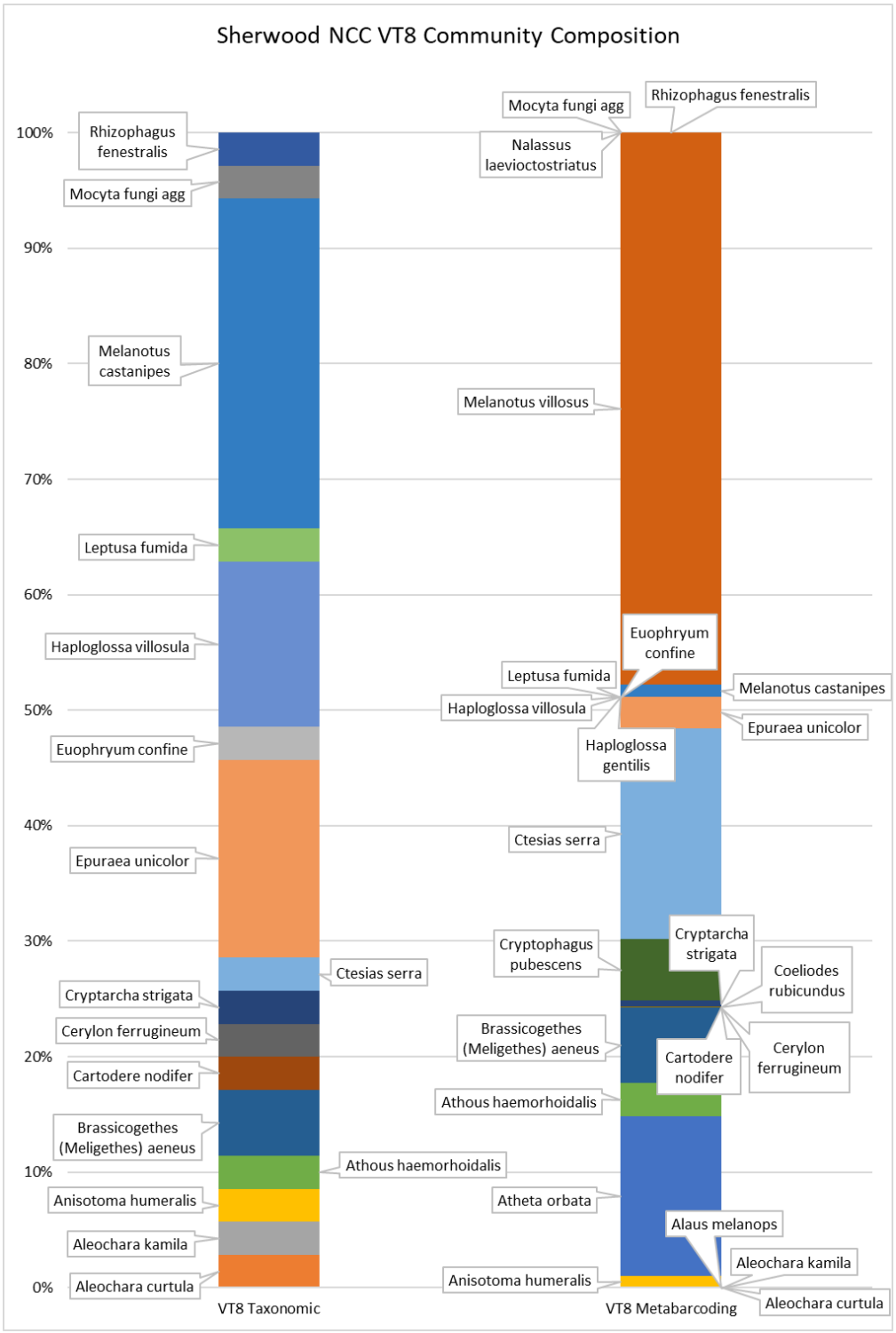


Figure 3.19 Community composition of Sherwood NCC sample VT8. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

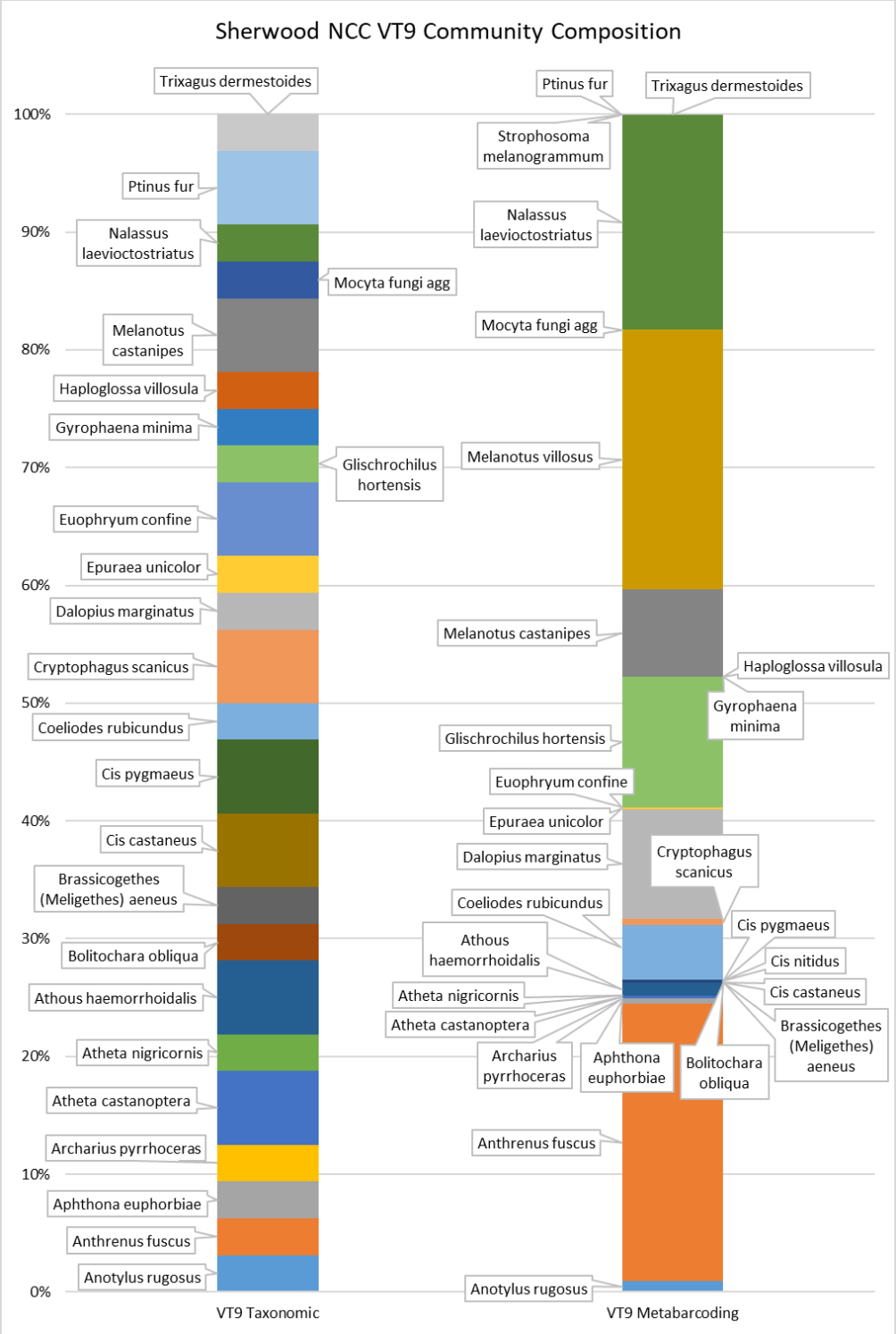


Figure 3.20 Community composition of Sherwood NCC sample VT9. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

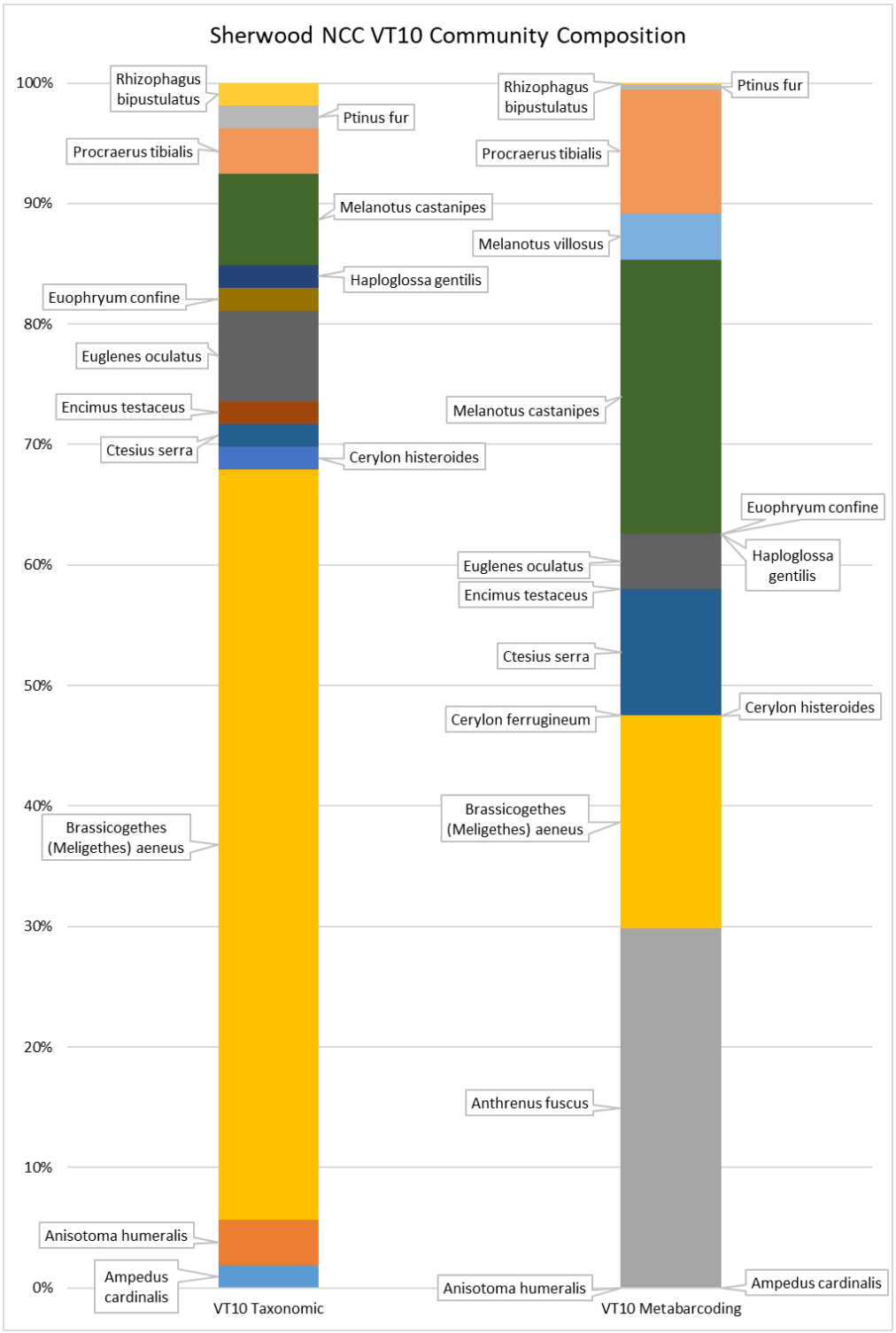


Figure 3.21 Community composition of Sherwood NCC sample VT10. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

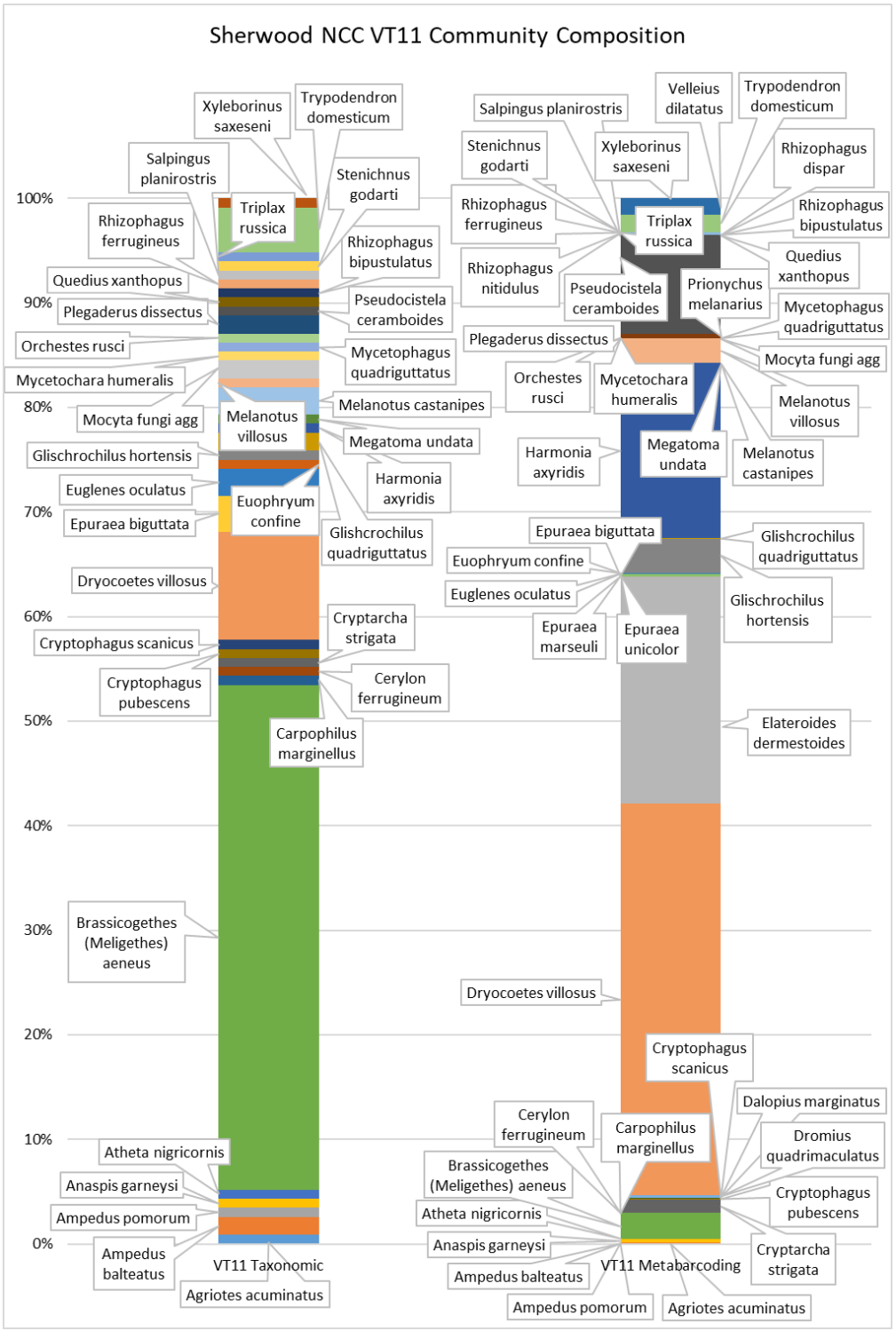


Figure 3.22 Community composition of Sherwood NCC sample VT11. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). ‘Taxonomic’ refers to those species that were hand identified and ‘metabarcoding’ refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation. All species that were effectively removed by being retained by Natural England were removed from the overall species list prior to bar chart creation.

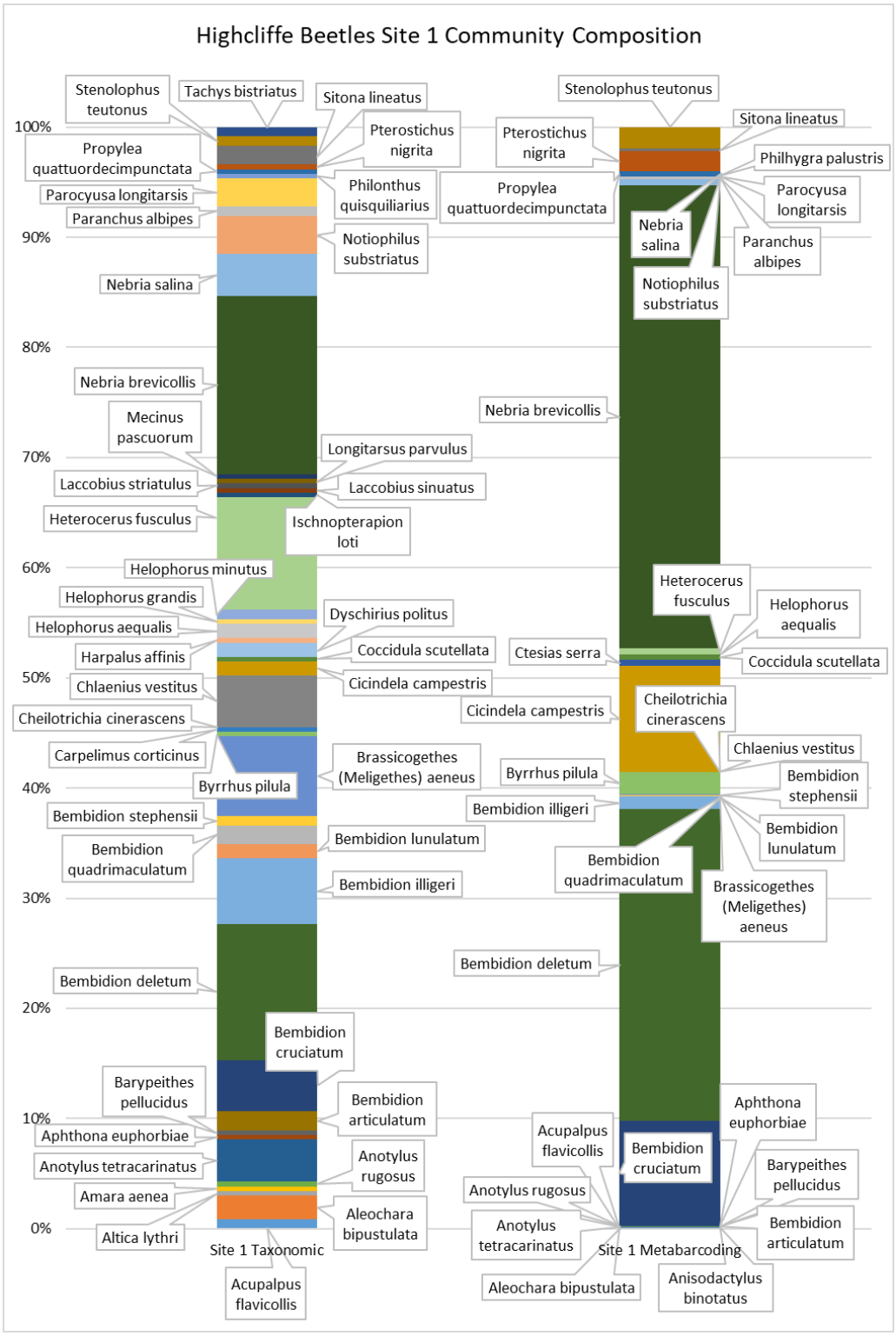


Figure 3.23 Community composition of Highcliffe beetles site 1. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation.

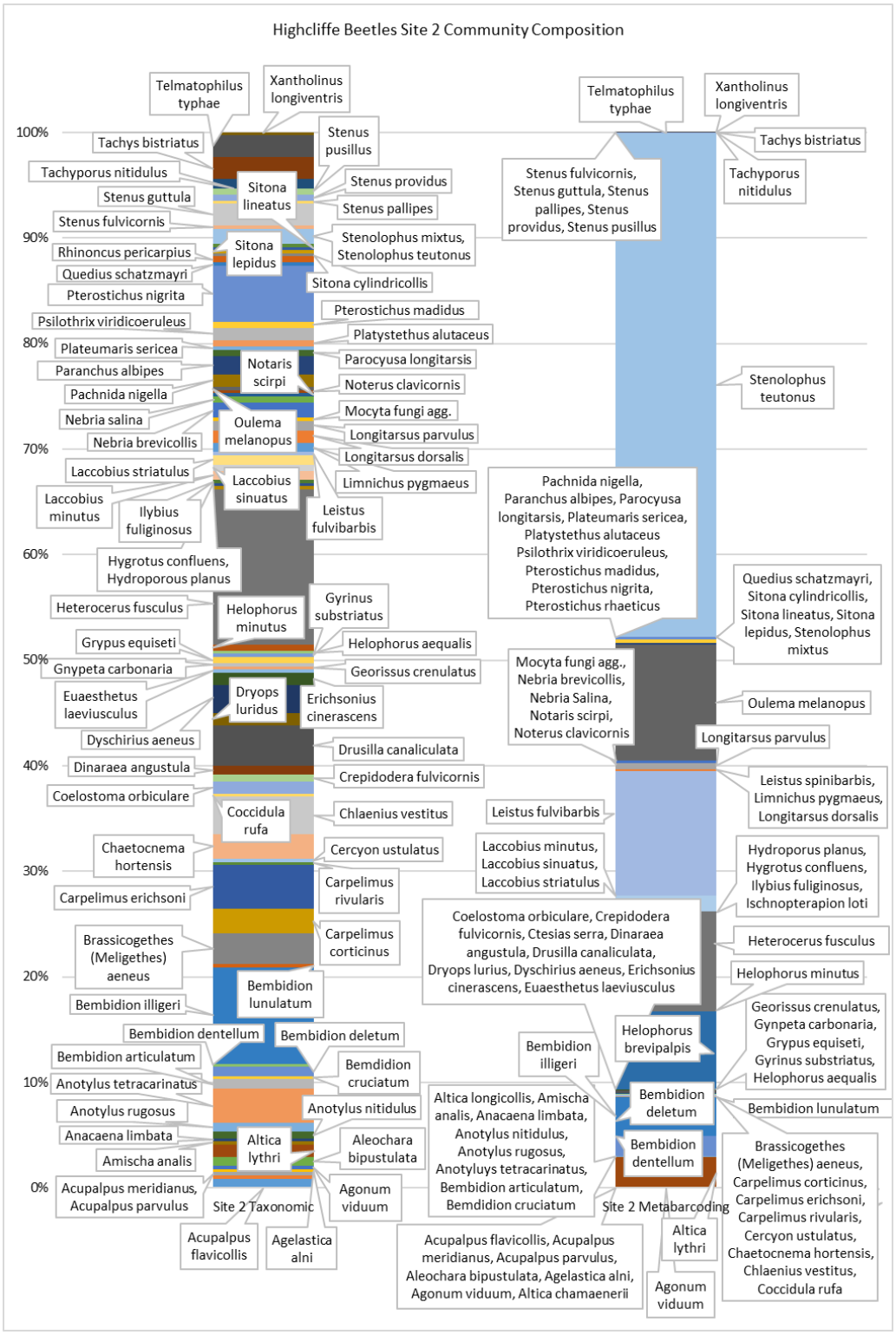


Figure 3.24 Community composition of Highcliffe beetles site 2. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation.

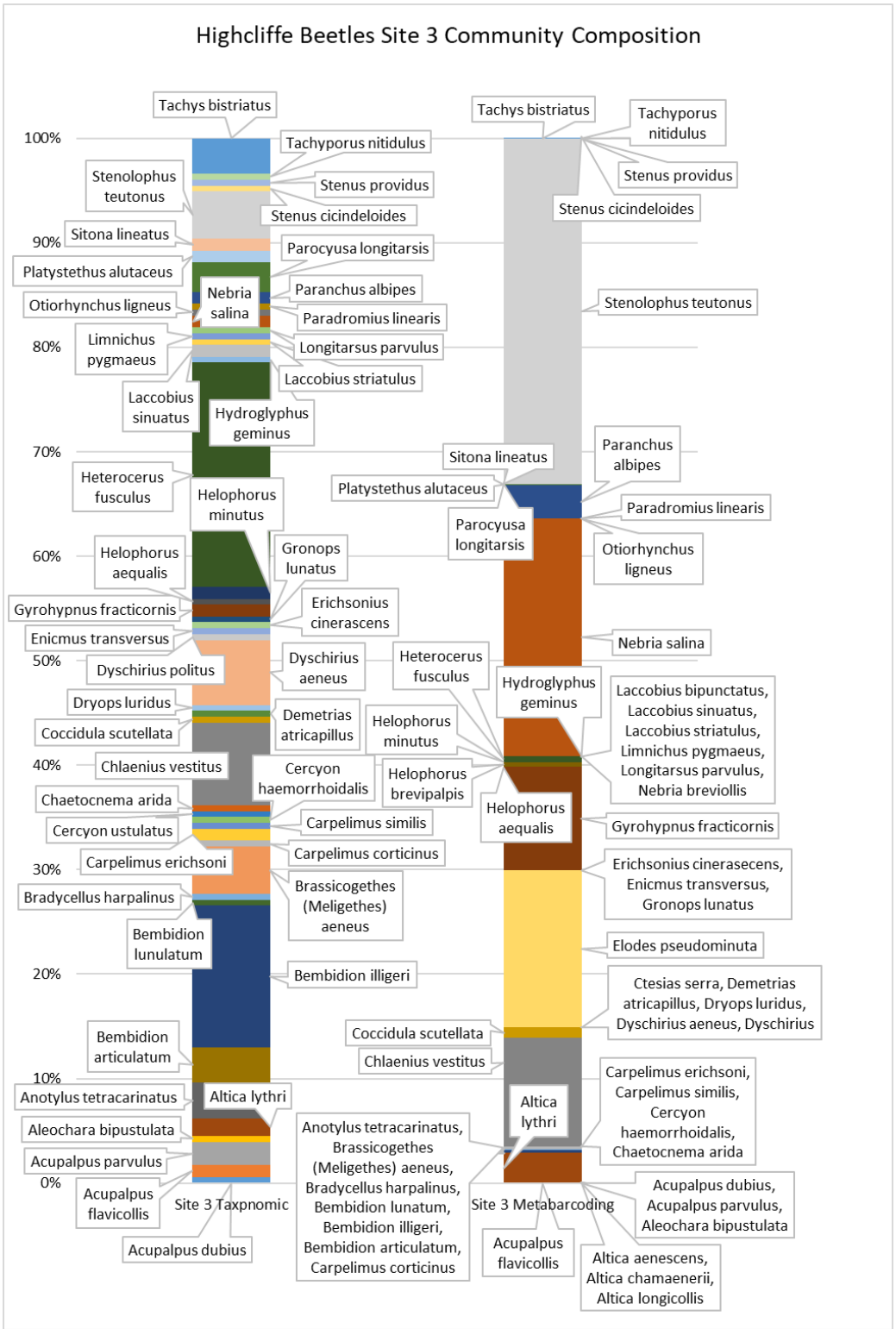


Figure 3.25 Community composition of Highcliffe beetles site 3. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation.

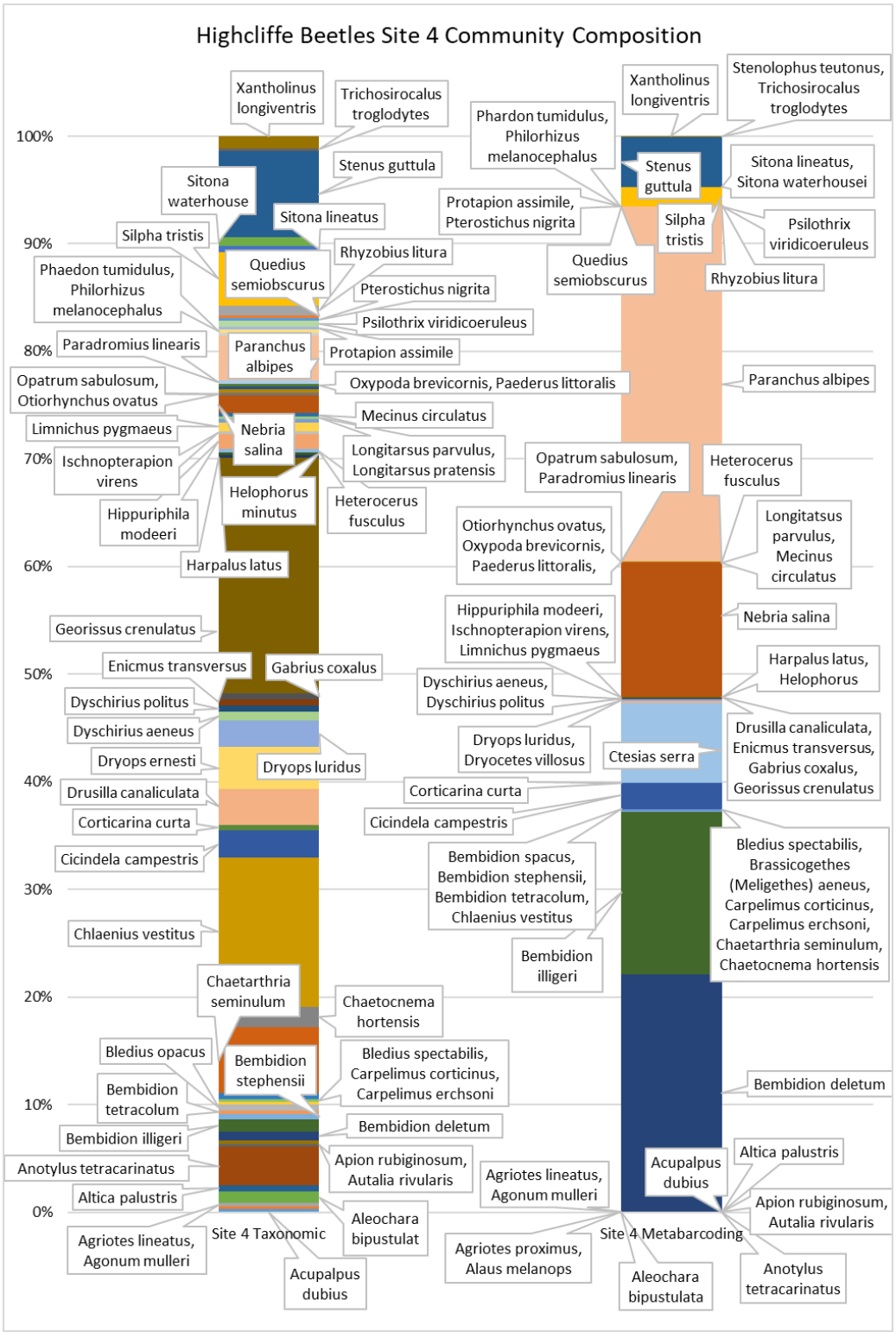


Figure 3.26 Community composition of Highcliffe beetles site 4. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation.

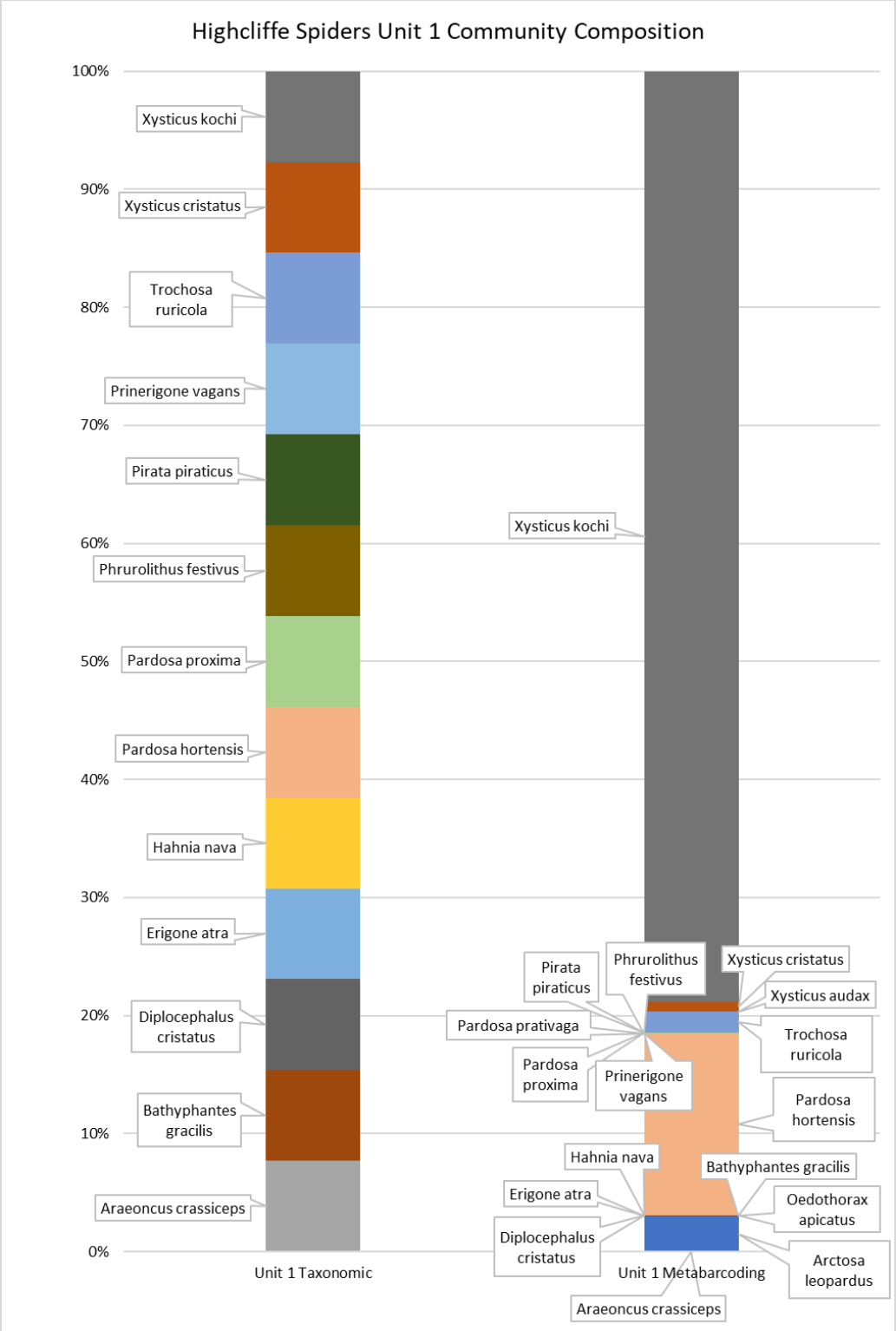


Figure 3.27 Highcliffe Spiders community composition Unit 1. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation.

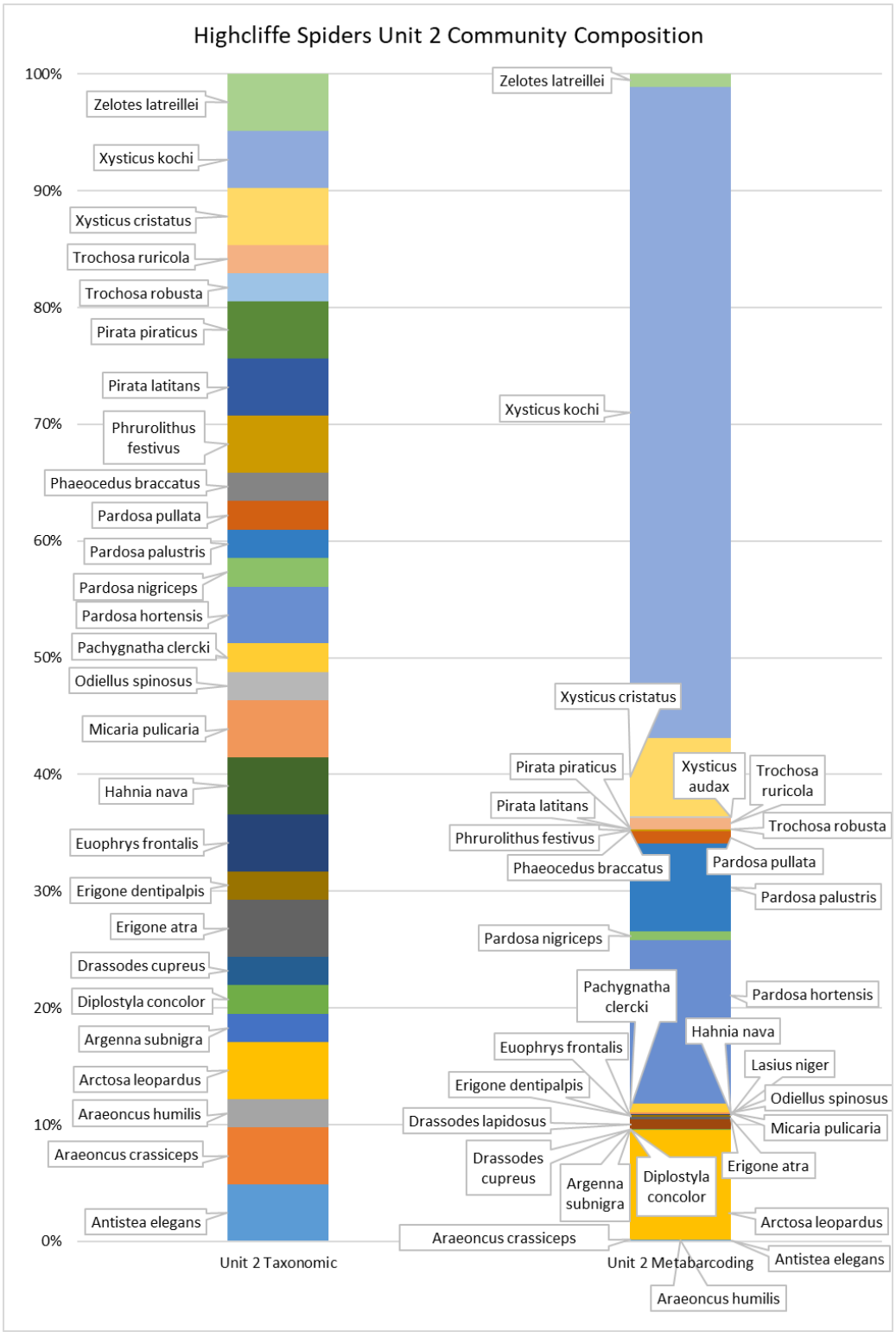


Figure 3.28 Highcliffe Spiders community composition Unit 2. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). ‘Taxonomic’ refers to those species that were hand identified and ‘metabarcoding’ refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation.

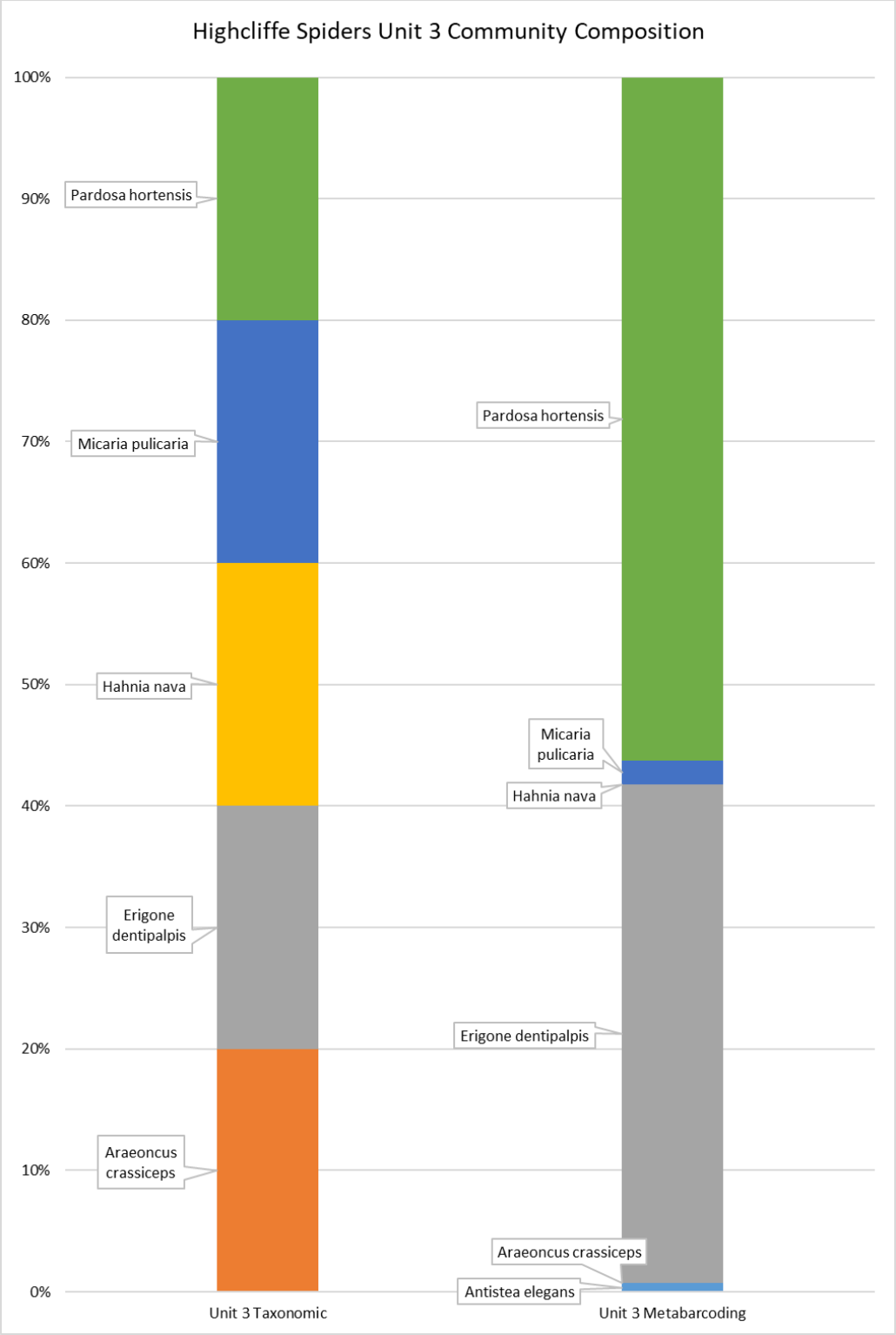


Figure 3.29 Highcliffe Spiders community composition Unit 3. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation.

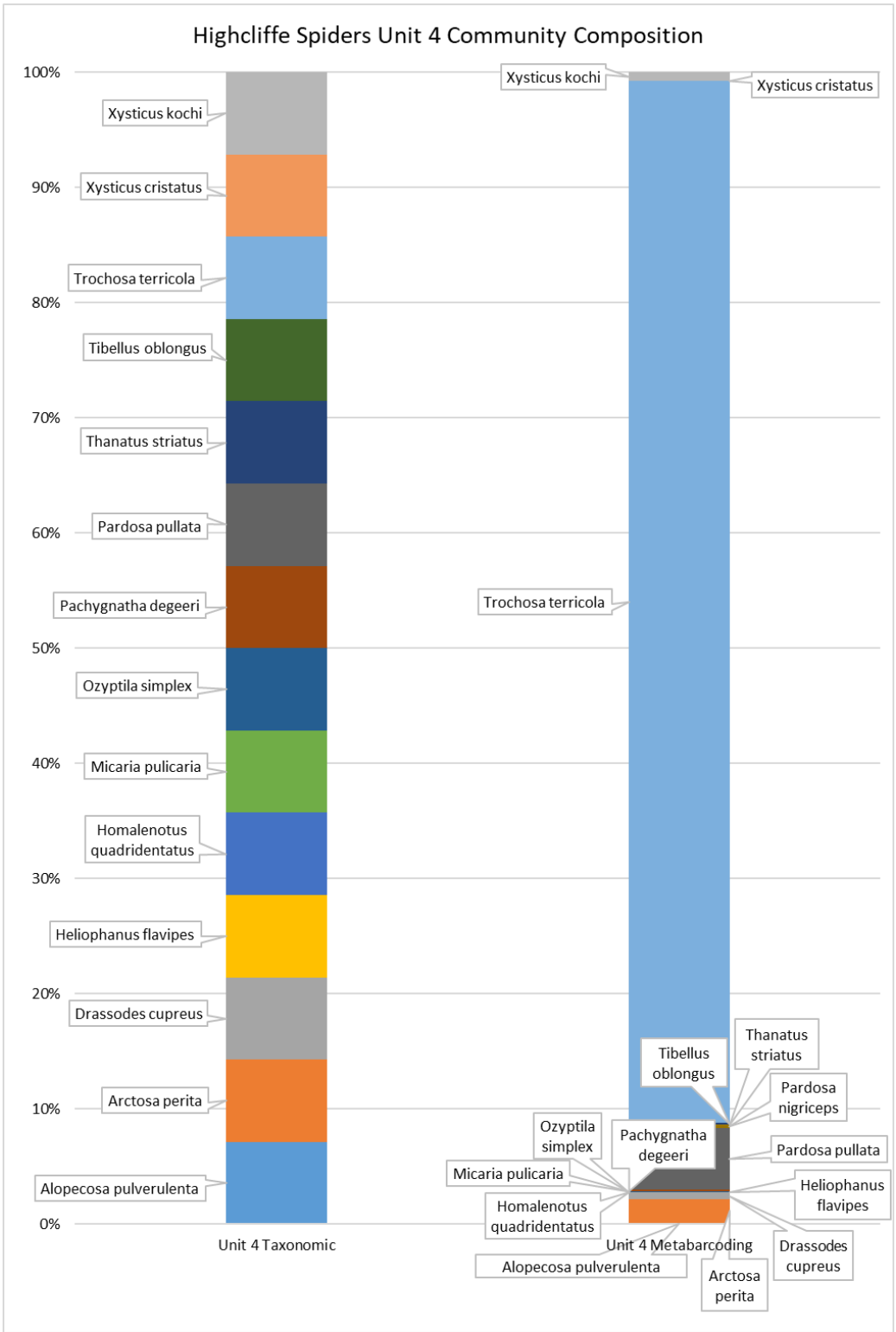


Figure 3.30 Highcliffe Spiders community composition Unit 4. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation.

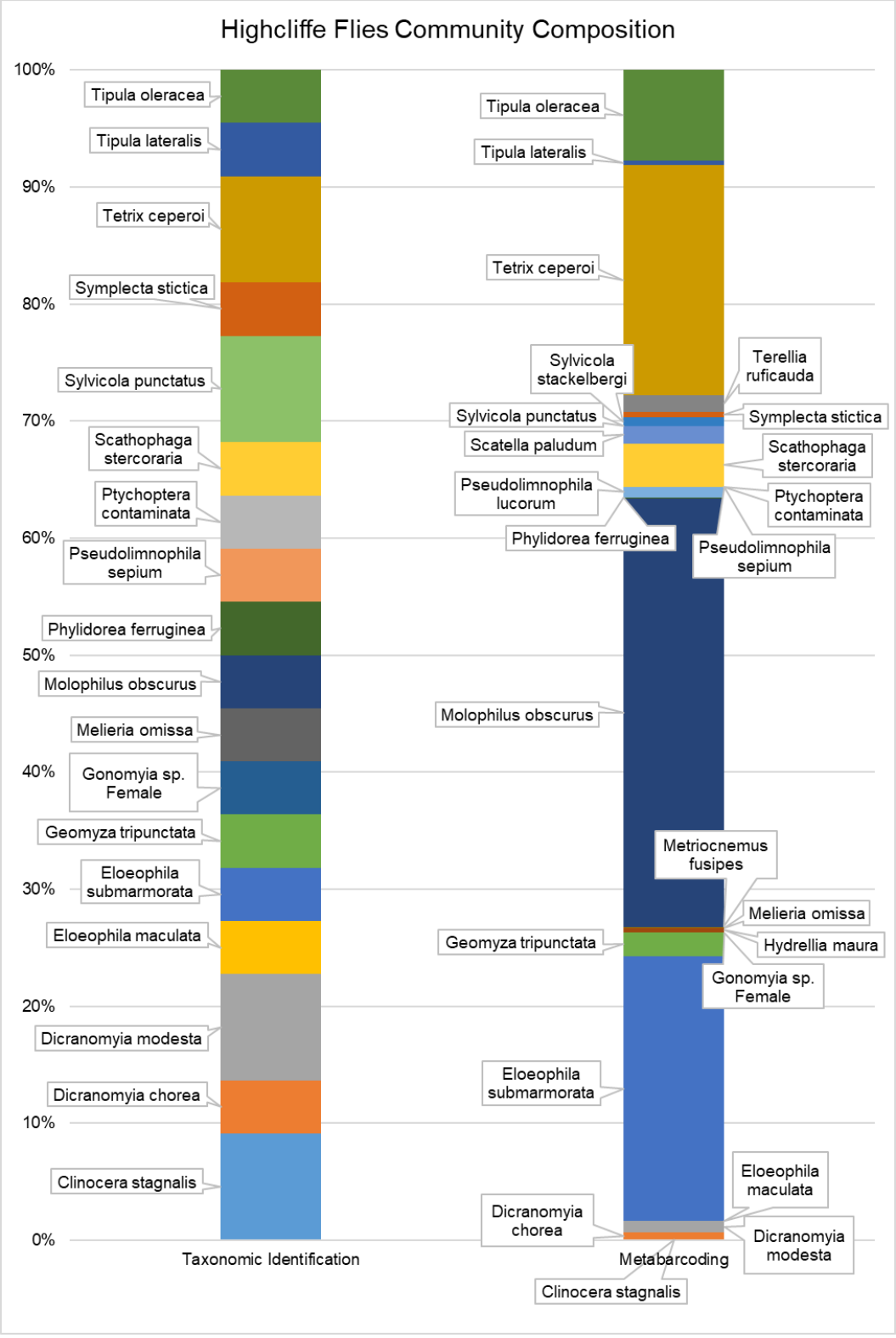


Figure 3.31 Diptera community composition. Bar sizes are based on the percentage abundance of number of individuals (taxonomic) or the percentage of number of sequence reads matching those species (metabarcoding). 'Taxonomic' refers to those species that were hand identified and 'metabarcoding' refers to those species that were identified by metabarcoding. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to bar chart creation.

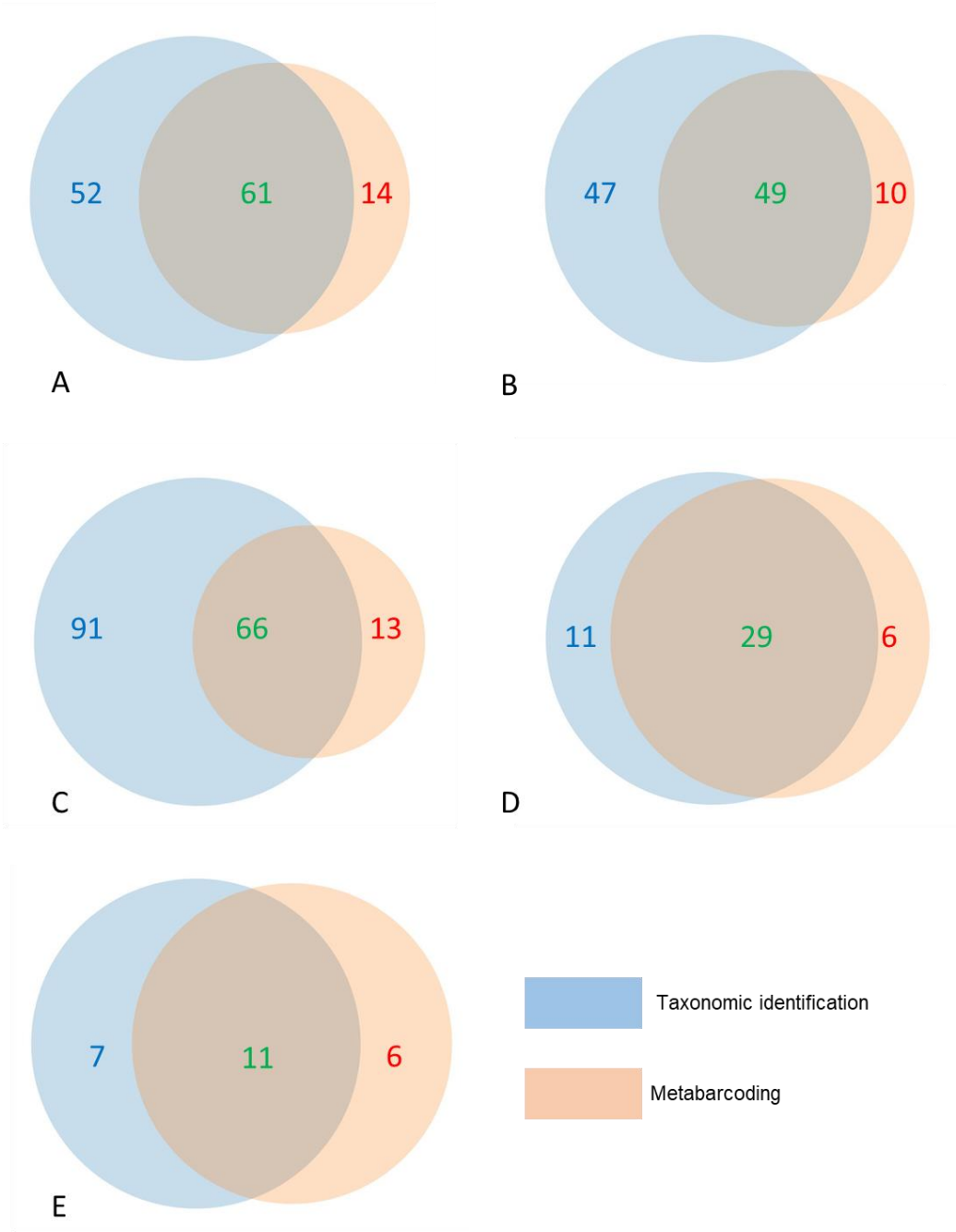


Figure 3.32 Relationship between the number of species identified by taxonomic assignment and the number identified by metabarcoding.

Venn diagrams represent the number of shared species for A. Sherwood FE samples, B. Sherwood NCC samples, C. Highcliffe beetles (pitfall traps and hand searches combined), D. Highcliffe Spiders, and E. Highcliffe Diptera. For all samples, those species that do not appear in the BOLD database and those which are not in the order of interest were removed from the overall species list prior to Venn diagram creation. For both Sherwood samples all species that were effectively removed by being retained by Natural England

were removed from the overall species list prior to Venn diagram creation. Coloured numbers represent the number of species found by i) one method: blue = taxonomic identification, and red = metabarcoding of samples; ii) two methods: green = taxonomic and metabarcoding.

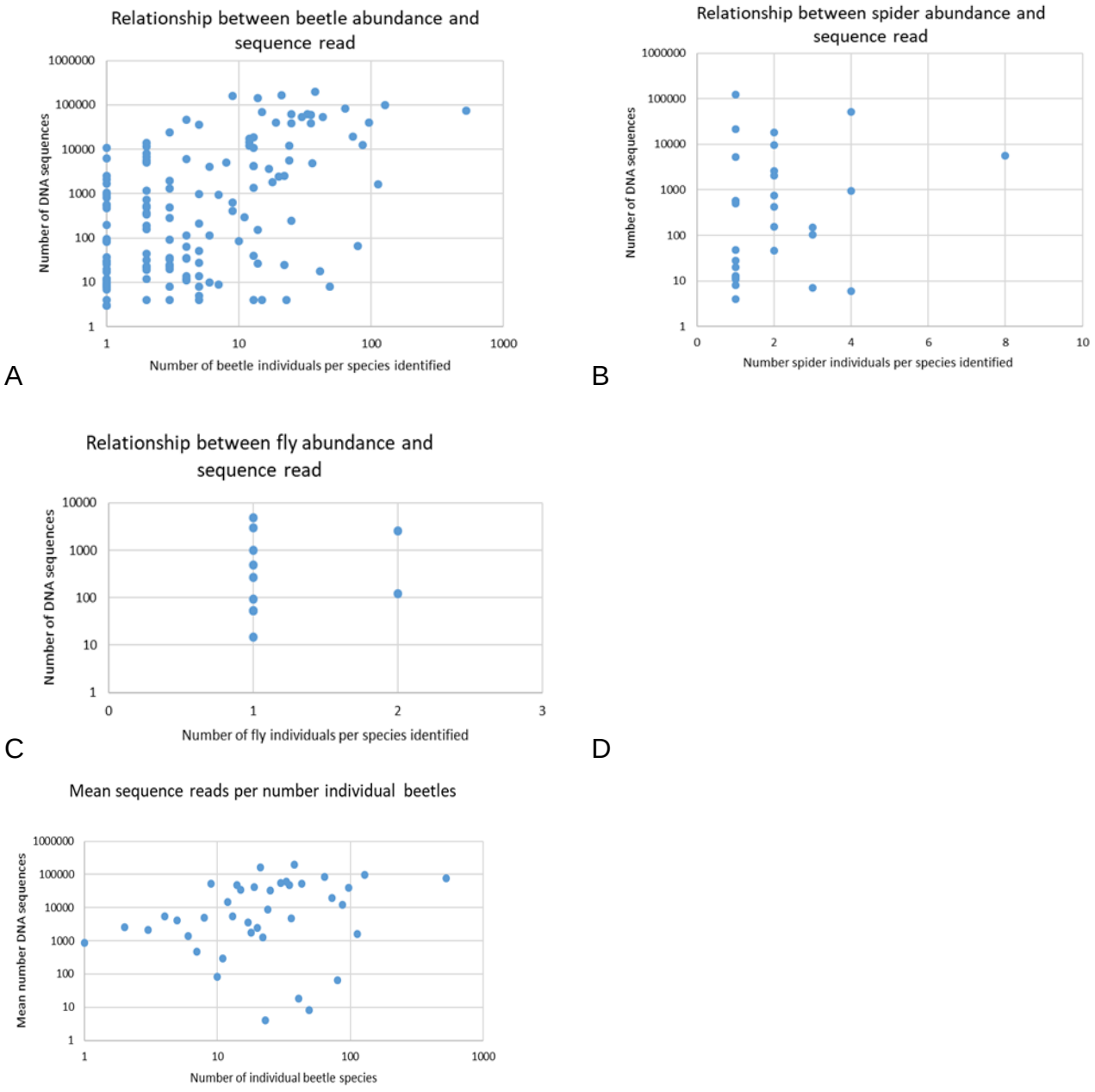


Figure 3.33 Relationship between species abundance and sequence read counts.

The total number of reads for each species identified by DNA sequencing plotted against the total number of times that species was identified in the samples: A. beetles, B. spiders, C. flies. Each point on the plot represents an individual species, and how many times that was seen both in the specimens and total number of sequence reads. D. The number of times a beetle species is identified plotted against the mean number of sequence reads.

4. Discussion

Overview. This work was undertaken to determine the applicability of metabarcoding methodology to monitor the presence (and possible abundance) of invertebrates that had been collected during ecological monitoring of two UK sites. The project was undertaken, not to necessarily develop new methodology, but more to apply the currently available methods and the associated DNA sequence reference databases that could be used to uncover any gaps in this approach and highlight areas of research and development effort that may make this approach more applicable and viable for use by Natural England (and others). Wherever possible this project followed previous examples of similar invertebrate metabarcoding work that had been published in peer reviewed articles (for example Deiner *et al.* 2016; Drummond *et al.* 2015; Hajibabaei *et al.* 2011; Murray *et al.* 2012; Valentini *et al.* 2016) with additional information being found within Natural England commissioned report NECR252 (Tang *et al.* 2018).

Size sorting. The first step of the analysis was to sort the samples by size. Several samples contained a number of species that were much bigger than the rest of those collected and there was concern that sequencing these within the context of the single sample would cause problems during the analysis of the sequence data - it would be very likely that the number of sequence reads would mostly represent those species with the largest biomass. Samples were therefore size sorted to greater/smaller than 7mm, and analysed as two separate sub-samples. Additionally, three very large beetles were removed from two of the samples and their head only returned to the sample to attempt to mimic the size of the next largest beetles present. In sample Sherwood FE T11, there was a single very large beetle (*Nicrophorus humator*) which was also found via metabarcoding of the 'big' sub-sample. This species accounted for ~46% of all sequence reads for the 'big' sub-sample with a further ~50% being assigned to *Melanotus sp.* of which five were known to be within the 'big' sub-sample. Therefore, despite the fact that only the head was added to the sub-sample this species still potentially 'species masks' other species present in the sub-sample (Brandon-Mong et al. 2015; Evans et al. 2016; Kelly et al. 2014). Only three other beetles were present within this 'big' sub-sample all of which were of a smaller size than their *Nicrophorus humator* and *Melanotus castanipes* counterparts. Of the remaining sequence reads three other species were within the hundreds of reads range which could account for these three individuals. If species-masking occurred, it is likely to have been at the first round PCR stage due to the relative amounts of DNA extracted from each species. In sample Sherwood FE T4 there were an additional two large individuals (which we believe to be *Anoplotrupes stercorosus* and *Carabus problematicus*) which were again removed and the head only returned to the sub-sample. Unfortunately for this sample, only a small number of reads were obtained so it is difficult to draw many conclusions, however, 43% of the sequence reads were assigned to *Carabus problematicus*, again suggesting that this species may have masked the others present. Interestingly, no reads were assigned to *Anoplotrupes stercorosus* but 47% were assigned to *Prionychus melanarius*.

DNA extraction and analysis. At the project start it was unknown whether there would be enough intact DNA that could be extracted from these sub-samples given that the time from collection to archiving and the fact that the solutions used for storage of the samples was probably less than optimal for the long term archiving of specimen DNA. We are unable to determine how storage has affected these samples as to do this we would require fresh material that was exactly the same as that which had been stored (and multiple replicates of samples). Such material would need to be extracted and then a total yield of DNA recorded and the amount of COI target for each sample quantified by qPCR. For this study all sub-samples were extracted after sub-sample drying and grinding to powder in liquid nitrogen. The Qiagen DNeasy blood and tissue kit that was used for DNA extraction has been used in metabarcoding studies before (Andruszkiewicz *et al.* 2017; Blackman *et al.* 2017; Sato *et al.* 2017) and gave us DNA suitable for amplification by PCR albeit in low quantities. However, in all cases we were able to extract quantifiable amounts of DNA that was suitable for PCR. As a general rule we would recommend that samples are preserved in the best state as possible for the preservation of DNA and suggest that samples collected spend the minimal time in the propylene glycol solution, before being stored in 95% ethanol and not 50% ethanol.

Instead of designing and trialling new PCR primers, which was beyond the scope of this project, primers that had previously been described (and are in widespread use) were used to generate the COI PCR amplicons from each sub-sample (Leray *et al.* 2013). Two different amplicons were initially tested with DNA extracted from a number of sub-samples, the primer pair mICOIintF/ jgHCO2198, amplified a greater number of sub-samples and with higher amounts of product than the alternate HexCOIF4/ HexCOIR4 primers that were also trialled, therefore these were taken forward as the primers used for the project. In carrying out this first round PCR the aim is to capture as much of the sequence diversity as possible that is contained within the samples. In order to do this the primers are degenerate that is they contain variations at some of the nucleotide positions within primer sequence (Table 2.1). Degenerate primers can be more difficult to use, because there will inevitably be some nucleotides that are mismatched upon primer binding to the target sequences. A modification of the jgHCO2198 forward primer as described by Geller *et al.* (2013) was to use the nucleotide ‘Inosine’ at three positions within the primer. Inosine is useful in that it can base pair with any natural base, resulting in a more stable primer/target duplex, and hence a more efficient PCR. These Inosine containing degenerate primers, were probably the cause of considerable delay in this project as described in detail later.

PCR was carried out used an ‘environmental mastermix’ containing the polymerase Amplitaq Gold (a standard Taq polymerase as opposed to a high fidelity Taq polymerase). This matermix/enzyme was chosen as it has been prepared to have good tolerance to PCR inhibitors such as may be co-extracted from invertebrate samples, which ordinarily may have inhibited the PCR. Previous experience with this enzyme in metabarcoding experiments suggested that the enzyme fidelity was good enough to retrieve good quality sequence data. Higher fidelity enzymes can also be used but are more difficult to use where target DNA concentration may be low and where degenerate primers are employed.

There is therefore a trade-off in getting the PCR to work effectively with the choice between a high fidelity polymerase which should be highly accurate (but may not be as sensitive as a lower fidelity enzyme) and a lower fidelity enzyme (such as amplitaq gold) which may be better at generating the DNA product to start with. All first round PCR products were made with Amplitaq gold before we attempted the second round PCR to add the sequencing adaptors. Using higher fidelity polymerases within the second round PCR (using the first round product as template), we had little success in adding the adapters by PCR. Sanger sequencing of amplicons made from two known species confirmed the primers were targeting the correct COI sequence, and alternate methodology for the purification of the first round PCR product did not improve these second round amplifications. Considerable effort was taken to resolve this issue of not being able to generate a second round PCR product. We concluded that it was likely that DNA polymerisation by high fidelity proof-reading enzymes that were trialled, fail after encountering the Inosine residues within the primer sequence of the first round PCR product. This is a plausible reason for the repeated failures of second round PCRs using the higher fidelity enzymes. As such the second round PCR (the addition of the index and flow cell attachment sequence) was therefore carried out with the same AmpliTaq Gold mastermix as first round PCRs.

BOLD database. The Barcode of Life Data system (BOLD) (http://www.barcodinglife.org/) is a publicly available cloud based storage and analysis platform designed to aid the acquisition, storage, analysis and publication of DNA barcode records with aim of eventually recording a barcode library for all eukaryotic life. In order to create our own searchable database that could be used to compare (Blast) our COI sequences against, a custom COI sequence database was assembled by downloading all sequences from BOLD as of January 2019 recorded as 'Arthropoda'. Of all sequences that were downloaded, (4,554,420 sequences) this represented a total of 178,408 individual species. Initial BLAST searches assigned a large number of sequence reads to for example Melanotus sp. In that reads were only being assigned to the genus level. To improve the sequence assignment it was decided to discard all sequences with a 'sp' from our downloaded BOLD database and additionally those saying 'SUPPRESSED' (it is unknown what this denotes but sequences containing this sequenced the sequence assignment). Once the BLAST search was run again with this modified database we were able to obtain assignments to the species level in most cases.

During the comparison of the sequence assignments with data provided by Natural England it was noted that 30 species identified by entomologists as being present in the supplied samples were not represented within the custom database, and either have not been sequenced/entered or have been entered but not identified to species level. The 30 species not included within the BOLD database include: *Anaspis fasciata, Argenna subnigra, Bledius atricapillus, Cassida hemisphaerica, Cathormiocerus socius, Cis villosulus, Curimopsis setigera, Cyphon pubescens, Dicranomyia goritiensis, Dicranophragma nemorale, Erioptera fusculenta, Homalenotus quadridentatus, Hylocereus dermestoides, Ilisia maculata*, *Kissister minimus, Laccobius atratus, Leiodes lunicollis, Lobrathium multipunctum, Meligethes carinulatus, Meligethes lugubris, Neliocarus faber, Oomorphus*

concolor, Othiorhyncus singularis, Parydra littoralis, Pirata latitans, Stenichnus poweri, Suillia imberbis, Tetartopeus angustatus, Tetralaucopora longitarsis, and Thinobius brevipennis. It is also worth noting that *Meligethes aeneus* was found to have been renamed *Brassicogethes aeneus*; *Coeliodes rubicundus* is known as *Coeliodinus rubicundus*; *Xyleborous saxeseni* is known as *Xyleborinus saxeseni*; and *Romualdius angustisetulus* is known as *Trachyphloeus angustisetulus* meaning that on first inspection these species were thought not to be present on the BOLD database. To add the missing species to either the BOLD database or our own custom curated database would require the further taxonomic identification of these individuals (multiple individuals), and the sequencing of their mitochondrial COI genes using Sanger sequencing.

Sample Analysis. In order to allow a fairer comparison of taxonomic identification and metabarcoding, the species not present on the BOLD database were removed from the data set so that we only compared the species that could be identified by metabarcoding. Taxonomic identification identified 336 individual species and metabarcoding 228 species so the addition of these 30 'missing' species should improve the rate of metabarcoding identification. The sequencing datasets also record additional species that were not identified by the entomologists. Overall 91 species were identified by sequencing that were not identified in the samples by taxonomy (41 after species not appearing on BOLD, non order species, and those retained by Natural England were removed). It is possible that these could represent species that were from gut contents i.e. prey species, contamination from the traps from previous usage if not properly cleaned, misidentification by taxonomists, or may indicate potential errors within the BOLD database. However, this still leaves species identified by the BLAST of our database (downloaded from the BOLD database) which are unlikely to be present such as *Bolla atahuallpai* a species of butterfly found in Peru within the spider samples. The presence of this unlikely species could be as a result of the actual species that the DNA sequence corresponds to not being present on the BOLD database or an error of some sort on the BOLD database.

Focusing on beetles, overall DNA metabarcoding missed 134 out of 279 (48%) of total beetles that had been collected and identified to species level (and were also present within BOLD), 23 additional species were not present on the BOLD database. However, metabarcoding detected an additional 29 species of beetle not identified by taxonomists, additional species were also detected which may represent prey species. For species that were not detected by the DNA metabarcoding we analysed the data from the Sherwood FE (T1-T11) samples to see how often species that were identified multiple times were not detected by sequencing. 65 species were identified on multiple occasions (they were collected in more than one trap). Of this number 25 (38%) were missed by the DNA sequencing. Species that are missed on two or more occasions maybe suggests that these are species that are simply being missed by the initial metabarcoding PCR step. There are several reasons why this could be the case: 1) DNA from certain species may be misrepresented in the pool of invertebrates, either by coming from invertebrates that are much smaller in size than others within the sample pool, or being present in much smaller numbers than the dominant species both scenarios contributing to differences in starting

biomass? 2) DNA may have been inefficiently extracted from different species, there may have been differential degradation of the DNA within some species depending on the time that they were trapped to the time they were collected, again this could contribute to a low DNA target number at the start? 3) Perhaps the biggest source of bias may be in the primers that are used in the initial PCR which may have missed some of these species, the primers used may simply not work efficiently for some of these invertebrate species. DNA extracted from community samples may be subject to potential amplification bias where different species' DNA is in competition to bind to the universal primers which can prevent the capture of all species present in a given sample as more common template DNAs are likely to be amplified (Kelly et al. 2014). This in turn can mean that for very large individuals, high abundance species can prevent the detection of low abundance species resulting in 'species masking' (Brandon-Mong et al. 2015; Evans et al. 2016; Kelly et al. 2014). Metabarcoding may therefore be less capable of identifying the DNA of less abundant species within a community than a species-specific qPCR for example. 4) Individuals could have been mis-identified at the taxonomic identification step.

To try and further investigate this, the COI sequences for some six species that were missed on multiple occasions were taken from the BOLD database and used to align with the Leray PCR primers used. This analysis demonstrated that the forward primer used should anneal and work with these during the PCR step. The BOLD database sequences however, all appear to be truncated just before the reverse primer binding site. This reverse primer is the same primer (Folmer primer) used in the construction of the 658bp reference sequences within the BOLD database, and as such is likely omitted before uploading to the database. As such it is likely that the primers used will amplify all target DNAs present (in the database), but it may be that amplification efficiencies vary considerably, especially within the context of additional species DNA. Further study of this could entail the PCR of some of these undetected species as single species target (to demonstrate that the PCR primers work). Further refinement of methodology could look at optimising primer sets that are more applicable to these sample types, perhaps targeting beetles, spiders and flies separately.

Comparing sequence read to number of individuals present in a sample was not consistent. A number of species were seen in multiple samples; for example in Sherwood FE (T1-T11) samples, 19 species were identified in over five different samples. Sequence reads from the same species within these samples vary considerably and read number will be influenced by total amount of biomass per sample. Grouping all species to individual numbers and plotting these against numbers of sequence (Figure 11) demonstrates poor correlation between read count and species number.

Time taken. In total we analysed 67 sub-samples (plus positive/negative controls and an extraction blank), taking into account that some samples were size fractioned into 'large' and 'small'. Ignoring the time taken for the extensive trouble shooting and the development of the tools for the analysis of the sequence data, the time taken for the sample analysis can be broken down as follows: The DNA extraction for this number of sub-samples

including the liquid Nitrogen grinding followed by the DNA extraction took six days of person time. The quantification of all DNA extracts and the setting up of the first round PCR took two person days. At this point the first round PCR products were passed to the Deepseq team. Ignoring time for troubleshooting the second round PCR, the first round clean up and quantification, the second round PCR, clean up and quantification followed by the extensive QC followed by the sequencing run took one week. The downloading of sequence data, the processing of the sequences and the blast searching against our custom arthropod sequence database took approximately three days person time. The data analysis including generation of community composition tables and bar charts and Venn diagrams showing the relatedness of the techniques used took approximately three days to complete. We would suggest that this whole process could be accomplished within four weeks which is comparable to the time taken for taxonomic identification of 160 hours or roughly 20 days (Webb, J., personal communication). In terms of costs, at the time of publication, metabarcoding worked out at approximately £155/hour, totalling £21,032 respectively. We have used both Natural England commercial rates and 'invertebrate consultants' as comparators. Using Natural England commercial rates, taxonomic identification worked out at approximately £182/hour, totalling approximately £29,120. Identification charges from invertebrate consultants will vary between both the consultants used and taxonomic groups to be identified but is roughly in the range of £25 to £65 per hour. This would cost between £4,000 to £10,400 for the identification outlined above[1].

Data Availability. The details of the taxonomic surveys and write up of the taxonomic identification is available on request from Natural England. The custom pipeline and sequencing data used for metabarcoding analysis has been deposited in the GitHub repository (https://github.com). Training material on the DNA methods used in the study is available upon request to ADAS.

[1] This section on estimated costs has been updated in this version to make it clearer and include the cost of invertebrate consultants

5. Recommendations

- It is recommended that in order to improve the custom sequence database that the species which are not currently found on the BOLD database are caught by either hand searching or trapping before being taxonomically identified and subjected to Sanger sequencing of their COI gene using the 'Folmer' barcoding primers and subsequent submission of sequence data to the BOLD database. This would ideally need to be carried out on multiple individuals of each species.
- It is recommended that the issue of 'species-masking' is investigated.
  A. Although the target specific primers used in this study successfully amplified the DNA from a large number of species it was still the case that many species were not detected via metabarcoding despite the fact that the primers should work on these species. It could for example be investigated whether the primer design could be improved specifically for UK species or for specific genus/families which are underrepresented or not seen in the metabarcoding data. These re-designed primers could be used in conjunction with the existing primers and any improvements documented.
  B. Likewise it could be investigated how species size and/or abundance effects metabarcoding outputs via the creation of 'mock' samples containing known biomass and or numbers of different species (see for example Braukmann *et al.* 2019).
  C. It could be investigated whether sample storage/preservation has an effect on samples. If not adequately stored samples run the risk of being subject to degradation of the sample condition which when you consider the already small amounts of DNA available for extraction could have a significant knock-on effect. We suggest that samples collected spend the minimal time in the propylene glycol solution, before being stored in 95% ethanol and not 50% ethanol
  D. Finally, where data deviates from taxonomic identification re-confirmation of species identification by taxonomic experts could be carried out if the samples were well documented by photograph prior to destructive sampling.

List of tables

List of figures

References

ANDRUSZKIEWICZ, E.A., STARKS, H.A., CHAVEZ, F.P., SASSOUBRE, L.M., BLOCK, B.A., BOEHM, A.B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. PLoS ONE, 12 (4): e0176343.

BLACKMAN, R.C., CONSTABLE, D., HAHN, C., SHEARD, A.M., DURKOTA, J., HANFLING, B., LAWSON-HANDLEY, L. (2017). Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples – first record of Gammarus fossarum in the UK. Aquatic Invasions, 12, 177-189.

BOLGER, A. M., LOHSE, M., & USADEL, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30:15, 2114-20.

BRAUKMANN, T.W.A., IVANOVA, N,V., PROSSER, S,W,J., ELBRECHT, V., STEINKE, D., RATNASINGHAM, S., DEWAARD, J.R., SONES, J.E., ZAKHAROV, E.V., HEBERT, P.D.N. (2019). Metabarcoding a Diverse Arthropod Mock Community. Molecular Ecology Resources, epublished ahead of print.

DEINER, K., FRONHOFER, E.A., M€ACHLER, E., WALSER, J.-C., & ALTERMATT, F. (2016). Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nature Communications, 7, 12544.

DRUMMOND, A. J., NEWCOMB, R. D., BUCKLEY, T. R., XIE, D., DOPHEIDE, A., POTTER, B. C., GROSSER, S. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. GigaScience, 4, 1.

FOLMER, O., BLAKC, M., HOEH, W., LUTZ, R., VRIJENHOEK, R. (1994). DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol, 3, 294-299.

HAJIBABAEI, M., SHOKRALLA, S., ZHOU, X., SINGER, G. A., & BAIRD, D. J. (2011). Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE, 6, e17497.

HEBERT, P.D.N., CYWINSKA, A, BALL, S.L., DEWAARD, J.R. (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences. 270, 313-321.

KEY, R.S., DRAKE, M., SHEPPARD, D.A. (2000). Conservation of invertebrates in England: a review and framework. Natural England Commissioned Reports, Number 35.

LERAY, M., YANG, J.Y., MEYER, C.P., MILLS, S.C., AGUDELO, N., RANWEZ, V., BOEHM, J.T., MACHIDA, R.J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in Zoologhy, 10, 34.

MACHIDA, R.J., KNOWLTON, N. (2012) PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences. PLoS ONE 7(9): e46180.

MACHIDA, R.J., KWESKIN, M., KNOWLTON, N. (2012b) PCR Primers for Metazoan Mitochondrial 12S Ribosomal DNA Sequences. PLoS ONE 7(4): e35887.

MAGOC, T. & SALZBERG, S. (2011) Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:21 , 2957-63.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., CHEN, Z. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437, 376–380.

MARTIN, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, 17(1), 10-12.

MEUSNIER, I., SINGER, G.A.C., LANDRY, J-F., HICKEY, D.A., HEBERT, P,D.N., HAJIBABAEI, M. (2008). A universal DNA mini-barcode for biodiversity analysis. BMC Genomics, 9, 214.

MURRAY, D. C., PEARSON, S. G., FULLAGAR, R., CHASE, B. M., HOUSTON, J., ATCHISON, J., MACPHAIL, M. (2012). High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. Quaternary Science Reviews, 58, 135–145.

SATO, H., SOGO, Y., DOI, H., YAMANAKA, H. (2017). Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Nature Scientific Reports, 7, 14860.

Seqtk (2012). Seqtk (Version 1.3) [Software]. Available at: https://github.com/lh3/seqtk (accessed February 2019)

TANG, C.Q., CRAMPTON-PLATT, A., TOWNEND, S., BRUCE, K., BISTA, I. & CREER, S. 2018. Development of DNA applications in Natural England 2016/2017. Natural England Commissioned Reports, Number 252.

VALENTINI, A., TABERLET, P., MIAUD, C., CIVADE, R., HERDER, J., THOMSEN, P.F., BOYER, F. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology, 25, 929–942.

WEBB, J. R., DREWITT, A., MOTT, N. (2018). Guidelines for soft rock cliff surveys. Natural England Survey Guidelines.

WEBB, J. R., HACKMAN, J., BOARDMAN, P. (2018). Saproxylic beetle surveys using vane traps and other supplementary methods for site quality assessments. Natural England Invertebrate Survey Guidelines.

APPENDIX 1. Detailed materials and methods

A. Sampling Methodology – vane traps for site quality assessments (Webb *et al.* 2018)

Decaying wood in veteran trees supports an important and diverse invertebrate saproxylic fauna. This includes beetles, flies, and smaller numbers of groups such as bees, wasps, moths and bugs.

In recent years trapping techniques have developed significantly and flight interception or vane traps in particular have proved a successful means to survey saproxylic beetles both in the UK and across Europe. Beetles are considered to be a good indicator of habitat quality and fidelity to veteran trees.

Vane traps can be deployed over a long period of time and are a method for catching insect samples that avoids destruction of the habitat. Vane trapping is also a readily repeatable method that allows for comparative analysis between sites and samples. This guidance aims to explain how a vane trap works, how it should be installed and maintained and how to deal with the resultant collections. Vane traps are considered to be easy to manage by non-specialists. Once trapping is complete all collections are identified by an entomologist and analysed using Pantheon.

Description of the vane traps:

Vane traps are durable, most being made of plastic, and can be re-used from year to year. They are light and relatively easy to transport and assemble. There are a number of designs but the following one has been used by Natural England staff on numerous occasions. It consists of intersecting panels of Perspex, around 45cm high, connected through a funnel to a screw-on collecting bottle. A Perspex roof slots on top of the panels which has two holes at each corner. A cable tie goes through each hole to attach string, which is then tied to the other 3 strings to form a loop. The collection bottles are part-filled, usually to about 5cm (about a third of the way up) with a mixture of 50% preserving fluid (propylene glycol) and 50% water and a drop of surfactant (washing up liquid).

A layer of chicken wire is laid between the bottom of the panels and the top of the funnel to stop unwanted objects falling into the bottle and act as a barrier to larger animals. Rope is used to attach the loop of string to the tree. The traps provide a vertical barrier to insect flight that is thought to be invisible to them. On collision with the panel beetles and other insects will often drop down and fall into the collection bottle.

Figure S1.1 Example of a vane trap installed within a hollow in an oak tree. The Perspex panels sit above the funnel which feeds into the collecting bottle. (photo Jon Webb / Natural England)

Choosing the location for the traps:

The approach is to hang the traps in trees which have been identified as having suitable rot and holes and at the same time ensuring the traps are safely secured. Information from previous tree surveys may help source suitable trees along with advice from an entomologist. In practice, these tend to be trees with exposed cavities containing red rot, white rot and those with cracks or fissures where sap runs out. Very hollow old trees that lack wood mould or fungal rot are possibly not as good as trees with a large amount of moist available decaying wood. Tree species will play a role, but more important is the type of rot within a tree, which can be broadly split into red rot or white rot (image?). Other issues to take into account when siting traps include potential interference from livestock or people where there is public access. This might be overcome by siting the traps a little higher or in less visible places. See Annex 1 for examples of vane traps in a variety of locations.

To date, Natural England staff have only sited traps next to cavities on veteran trees; it would be worth testing trap sitings away from cavities in the future.

As standard protocol, Natural England currently deploy 10 traps within a site on suitable trees. We have often deployed a few more than 10 (between 11 and 13 in case of unforeseen emergencies where a location might have to be discarded). These ten traps can cover the whole site, or a sub-sample of such a site. Tree species against which traps are set are chosen as a representation of the site as a whole (e.g. At Brocton Coppice in Cannock Chase, all 10 traps were sited on oaks; at Burghley Park in Lincolnshire, 5 were

placed on oak and 5 on sweet chestnut). On larger sites, where resources are not such an issue, it would be possible to deploy 10 traps on one tree species and ten on another, thus creating two samples. Multiples of 10 can also be used to investigate different parts of a site.

Should be erected by early-mid April and taken down in late October. The collecting bottle on each trap is numbered with a permanent marker pen e.g. 1 to 10. The traps do not need to be spaced at any particular density.

In most cases a ladder will be required to install, service and remove the traps. There is guidance on the Natural England Intranet on how to assess and control the risks associated with working at height which includes an on-line training course. It is a two person job.

It is helpful to record if any of the chosen trees have number tags and to take a GPS grid reference and at least one photo of each tree. Then put them into document that can be taken out on future visits to help locate them again.

Servicing the traps:

The traps must be regularly serviced, ideally every 3-4 weeks. Generally about 8 visits will be required to the site over the trapping period. In periods of high rainfall it is advisable to service the bottles as and when they fill up as efficiency becomes reduced. Trap servicing is at least two person job for health and safety reasons.

During the check, each bottle is removed and replaced with a fresh collecting bottle of the same number which has been filled with preserving fluid. The bottle screws on and off the trap. So prior to each site check you should ensure you fill up and take the correct number of bottles and that each is clearly numbered. When you take the lid off the fresh bottle you can reuse it to cover the bottle coming off the trap.

Try to retain all of the contents in the collecting bottle that comes off the trap even if it contains more liquid from an ingress of rain than when it was put out. Collect everything – some species can be very small. The contents from each collecting bottle should be initially sorted (be advised that some species are less than 2mm long). Some collecting bottles can be dark and difficult to look through if wood falls in so a white tray can be useful to separate things out. Even if you are intending just to look at the beetles, keep the by catch for others. If possible separate into taxonomic groupings, at least to order in small batches under the microscope.

Once sorted, put the samples put in separate tubes with 50% ethanol or stronger if to be left longer than a couple of years (95% used in this study). Label them with the site name, tree number and date of trap emptying. It is generally best to write details in pencil and place the label inside the container to avoid the risk of labels falling off over time. The invertebrates will be preserved quite successfully within the bottles of alcohol until they can be passed over to the entomologist for identification and interpretation. Keep the collected bottles in a cool and dry place out of direct sunlight.

Each time you change the bottles check to ensure that each trap is still in place and correctly assembled. If the trap has fallen down, the rope is broken or the piece of net above the collecting bottle has dislodged then make any necessary repairs. Take some spare rope, tape and scissors and a permanent marker pen with you on each visit and your document of tree photos, grid references and any tree tag numbers. Keep a record of the dates that you have changed the bottles.

B. Sampling Methodology – Hand Searching for Soft Rock Cliff Surveys (Webb *et al* 2018)

Hand searching can be a very effective method for sampling riparian invertebrates, particularly in terms of recording the smaller, cryptic species and those which are subterranean for most of the time. The method below is based on Derek Lott's protocol (Drake *et al*. 2007).
Each sample station consists of a Soft Rock Cliff seepage and its associated riparian habitat, such as eroding banks, the edges of water, vegetated sand, stretches of emergent vegetation, etc.
Each sample consists of the combined catches of three separate 10-minute sub-searches within a 30 minute period at each sample station. The aim of these separate searches is to target specific habitat types. This searching includes the time involved in transferring specimens to collecting tubes, preparing equipment etc. So the actual search time spent searching tends to be in the range of 8-5 minutes per sub-search.
At each sample station one or more of the following techniques are used to find animals, depending on the habitats present:

1. Soft sediments are trampled or patted, and surface-active insects pooted up directly from the ground.
2. Next to water margins, exposed sediment is splashed with water. This works best on steeper banks where a plastic kitchen sieve can be used to catch insects washed into the water or beetles can simply be pooted as they run back up the slope.
3. The basal parts of plants are examined or pulled apart; tussocks can be dissected over a sheet or tray using a small hand-saw and sieve and insects then pooted.
4. Litter and dense mats of fallen vegetation are sieved over a plastic sheet or tray, using a sieve with a mesh of 4 to 8 mm.
5. Emergent vegetation is submerged and the insects that float to the surface are scooped up with a plastic kitchen sieve.
6. Large stones can be lifted and species pooted from below and large woody debris can be broken apart before pooting.

C. Sampling Methodology – Pitfall Trapping

Based on Sadler & Bell (2000), ten small plastic cups, c. 10cm diameter, are dug into the sediment so that the rim is flush with the surface. These are filled one third full of a 50:50 mixture of commercial anti-freeze and water, with a small amount of detergent added to break the surface tension. Antifreeze both assists in sample preservation and reduces evaporation. For species that are collected for analysis via DNA meta-barcoding the current advice is to use propylene glycol rather than antifreeze.

At each site, pitfalls are placed sufficiently high up along a wetland edge to lessen the risk of flooding and/or hidden away to avoid detection. Pitfalls were left on each site for at least two weeks, and not more than four weeks, before collection and storage in 50% ethanol.

D. DNA Extraction

1. Add 360µl of buffer ALT from the DNeasy Blood and Tissue kit to the sub-sample.
2. An extra 1.5 mL tube must be set up to act as an extraction blank for every set of extractions performed. Therefore, add 360 µL of buffer ATL into a 1.5 mL microfuge tube and perform the DNA extraction as per steps below. Label this tube as extraction blank (EB).
3. Add 20 µL of proteinase K and 200 µL buffer AL. Mix thoroughly by vortexing. Heat at 56ºC for 10 min.
4. Add 200 µL of 100% ethanol. Mix thoroughly by vortexing.
5. Pipet the mixture into a DNeasy Mini spin column placed in a 2 mL collection tube.
6. Centrifuge at ≥ 6000 xg (8000 rpm) for 1 min. Discard the flow-through and collection tube.
7. Place the spin column in a new 2 mL collection tube. Add 500 µL Buffer AW1.
8. Centrifuge for 1 min at ≥6000 xg. Discard the flow-through and collection tube.
9. Place the spin column in a new 2 mL collection tube, add 500 µL Buffer AW2.
10. Centrifuge for 3 min at 20,000 xg (14,000 rpm). Discard the flow-through and collection tube.
11. Transfer the spin column to a new pre-labelled 1.5 mL microcentrifuge tube.
12. Elute the DNA by adding 200 µL Buffer AE to the centre of the spin column membrane. Incubate for 1 min at room temperature (15–25°C).
13. Centrifuge for 1 min at ≥6000 xg.

E. DNA Quantification

DNA extracts were quantified using the Qubit® dsDNA BR assay kit and Qubit 3.0 fluorimeter as follows:

1. The Qubit® working solution was prepared by diluting the Qubit® dsDNA BR reagent 1:200 in Qubit® dsDNA BR buffer.
2. Make up two standards by adding 190 µL Qubit® working solution into each of two tubes before adding 10 µL of each Qubit® standard to the appropriate tube. Mix by vortexing.
3. For each extract make up a tube with a final volume of 200 µL containing 1-20 µL extract and 180-199 µL Qubit® working solution.
4. Allow all tubes to incubate for two minutes before reading the standards and extracts on the Qubit® 3.0 fluorimeter.

F. DNA Purification

AMPure XP PCR Purification

1. Add 1.8 µL AMPure XP per 1.0 µL of PCR product and mix thoroughly by pipette mixing.
2. Incubate at room temperature for five minutes to allow DNA fragments to bind to the paramagnetic beads.
3. Separate the beads from the solution using a magnetic plate by waiting for the solution to clear before aspirating and discarding the solution leaving ~5 µL behind so as not to disturb the separated magnetic beads.

4. Wash beads twice with 200 µL of 70% Ethanol to remove contaminants, aspirating and discarding the solution for each wash (tubes remain on the magnetic plate throughout).
5. Remove the tubes from the magnetic plate and add 40 µL of elution buffer to elute purified DNA fragments from beads. Mix ten times by pipette mixing and incubate for two minutes.
6. Place the tubes back onto the magnetic plate and leave for one minute to separate the beads from the solution.
7. Transfer the eluate to a fresh tube.

ProNex® Size-Selective Purification System

1. Equilibrate the ProNex® bottle to ambient temperature for up to one hour prior to beginning purification then resuspend by vigorous vortexing.
2. Mix the ProNex® solution into the PCR products at a ratio of 3:1 v/v (Pronex® to PCR product) by pipetting ten times.
3. Incubate at room temperature for ten minutes and then place onto a magnetic stand for two minutes.
4. Carefully remove and discard the supernatant.
5. Wash beads twice with 200 µL of wash buffer to remove contaminants, incubating for 30-60 seconds before aspirating and discarding the solution after each wash (tubes remain on the magnetic plate throughout).
6. Air dry for five minutes (for high sensitivity downstream application drying times of up to one hour can be used).
7. After removing the tube from the magnetic stand add 50 µL of elution buffer and resuspend by pipetting.
8. Incubate for five minutes to elute the DNA then return the tube to the magnetic stand for one minute.
9. Transfer the eluate to a fresh tube.

Nucleospin® Gel and PCR Cleanup

1. If using small volumes (< 30 µL) adjust the volume of the reaction mixture to 50-100 µL with ultrapure water.
2. Mix one volume of PCR product with two volumes of Buffer NTI.
3. Place a NucleoSpin® Gel and PCR clean-up column into a collection tube and load onto the spin column.
4. Wash the silica membrane by adding 700 µL Buffer NT3 to the column and centrifuge for 30 seconds and 11,000 xg.
5. Discard the flow-through and place the column back into the collection tube before repeating this wash step.
6. Dry the silica membrane for one minute at 11,000 xg to remove Buffer NT3 completely.
7. Elute the DNA by placing the column into a fresh 1.5 mL microcentrifuge tube and add 15-30 µL Buffer NE and incubate at room temperature for one minute before centrifuging for one minute at 11,000 xg.

G. Polymerase Chain Reaction (PCR)

PCRs were set up in a total volume of 100 µL consisting of:

a. 2 µL of extracted template DNA,
b. 3 µL of each primer (0.4 µmol/L),
c. 50 µL of TaqMan® Environmental Master Mix 2.0 (containing AmpliTaq GOLD DNA polymerase),
d. 42 µL ddH2O.

A touchdown PCR (Don *et al.* (1991)) was used to amplify the invertebrate DNA extracted from sub-samples and included: an initial incubation for 5 minutes at 95ºC; then 17 cycles (denaturation at 95ºC for 30 seconds, annealing temperature for 30 seconds, and extension at 72ºC for 60 seconds) where the annealing temperature is reduced by 1ºC each cycle from 62ºC down to 47ºC; followed by 30 cycles at an annealing temperature of 46ºC and a final extension step at 72ºC for 30 seconds before holding at 4ºC until collection of PCR products for analysis.

H. Sequence Library Preparation

Illumina sequencing requires that sequences are able to physically attach to the high throughput sequencer. In order to achieve this, adapter sequences are added to the target amplicons (Table 2.1, Figure 2.3) thus allowing them to attach to the complementary adapters on the sequencer.

Sequencing libraries were prepared according to Illumina's '16S rRNA Sequencing Protocol'. Briefly, this requires:

1. Purification of the target specific PCR amplicons (including overhang adapters) was performed with Nucleospin® Gel and PCR cleanup columns (as above).
2. Second round PCRs (indexing PCRs using the Nextera XT index Kit v2 Set A kit) were set up in a total volume of 50 µL consisting of:
    a. 5 µL of first round PCR amplicon,
    b. 5 µL of each primer,
    c. 25 µL of Taqman Environmental Mastermix 2.0 (containing AmpliTaq GOLD DNA polymerase,
    d. 10 µL ddH_2O.
3. The indexing PCR included:
    a. an initial incubation for 3 minutes at 95°C,
    b. 12 cycles of 95ºC for 30 seconds,
    c. 55ºC for 30 seconds,
    d. 72ºC for 30 seconds,
    e. a final extension step at 72ºC for five minutes,
    f. hold at 4ºC until collection of PCR products.
4. The second round PCR products were then quantified using a Qubit 3.0 Fluorometer (see above).
5. Indexed PCR products were normalized by diluting to 2 nM using 10 mM Tris pH 8.5 before pooling (in equimolar amounts) of 5 µL aliquots of each to create a single pooled library for one Illumina MiSeq run.

6. The pooled library was then denatured with NaOH and diluted with hybridization buffer.
7. A PhiX library was also prepared in the same fashion.
8. The amplicon library pool was diluted to 10 pM, spiked with 10 % PhiX.
9. The combined library was then heat denatured at 96ºC for 2 minutes, inverted to mix and placed in an ice-water bath for 5 minutes. This heat denaturation step was performed immediately before loading the combined library into the MiSeq reagent cartridge to ensure efficient template loading on the MiSeq flow cell.
10. The library was run on the Illumina MiSeq using a MiSeq Reagent Kit v2 500 cycle kit, to generate 250-bp paired-end reads.

Table S1 Table showing further sample information.

Sample Name	**Shortened Sample Name**	**DNA Conc^{n} ng/µL**	**1st round PCR result**	**Notes**
Highcliffe Hand search 1 Big	1HHB	3.50	Bright Band	
Highcliffe Hand search 1 Small	1HHS	3.50	Bright Band	
Highcliffe Hand search 2 Big	2HHB	0	0	There were no specimens ≥7mm
Highcliffe Hand search 2 Small	2HHS	4.30	Bright Band	
Highcliffe Hand search 3 Big	3HHB	3.80	Bright Band	
Highcliffe Hand search 3 Small	3HHS	7.00	Bright Band	
Highcliffe Hand search 4 Big	4HHB	4.10	Bright Band	
Highcliffe Hand search 4 Small	4HHS	12.70	Bright Band	
Highcliffe Pitfall 1 Big	1HPB	3.30	Bright Band	
Highcliffe Pitfall 1 Small	1HPS	3.14	Bright Band	
Highcliffe Pitfall 2 Big	2HPB	2.70	Bright Band	
Highcliffe Pitfall 2 Small	2HPS	5.50	Bright Band	
Highcliffe Pitfall 3 Big	3HPB	3.80	Faint Band	
Highcliffe Pitfall 3 Small	3HPS	10.30	Bright Band	
Highcliffe Pitfall 4 Big	4HPB	1.22	Bright Band	
Highcliffe Pitfall 4 Small	4HPS	too low to measure	Bright Band	
Highcliffe Spiders 1 Big	1HSB	1.50	Bright Band	

Sample Name	Shortened Sample Name	DNA Conc^{n} ng/µL	1st round PCR result	Notes
Highcliffe Spiders 1 Small	1HSS	1.80	Bright Band	
Highcliffe Spiders 2 Big	2HSB	3.20	Bright Band	
Highcliffe Spiders 2 Small	2HSS	5.00	Bright Band	
Highcliffe Spiders 3 Big	3HSB	0	0	There were no specimens ≥7mm
Highcliffe Spiders 3 Small	3HSS	4.30	Bright Band	
Highcliffe Spiders 4 Big	4HSB	0	0	There were no specimens ≥7mm
Highcliffe Spiders 4 Small	4HSS	1.40	Bright Band	
Highcliffe Diptera Big	HDB	19.00	Bright Band	
Highcliffe Diptera Small	HDS	34.7	Bright Band	
VT1 Big Sherwood NCC	VT1B	4.60	Bright Band	
VT1 Small Sherwood NCC	VT1S	9.06	Bright Band	
VT2 Big Sherwood NCC	VT2B	4.64	Bright Band	
VT2 Small Sherwood NCC	VT2S	8.70	Bright Band	
VT3 Big Sherwood NCC	VT3B	4.90	Bright Band	
VT3 Small Sherwood NCC	VT3S	2.60	Bright Band	
VT4 Big Sherwood NCC	VT4B	2.30	Bright Band	
VT4 Small Sherwood NCC	VT4S	5.60	Bright Band	
VT5 Big Sherwood NCC	VT5B	1.55	Bright Band	

Sample Name	Shortened Sample Name	DNA Concⁿ ng/µL	1st round PCR result	Notes
VT5 Small Sherwood NCC	VT5S	12.20	Bright Band	
VT6 Big Sherwood NCC	VT6B	2.60	Bright Band	
VT6 Small Sherwood NCC	VT6S	8.03	Bright Band	
VT7 Big Sherwood NCC	VT7B	4.41	Bright Band	
VT7 Small Sherwood NCC	VT7S	14.30	Bright Band	
VT8 Big Sherwood NCC	VT8B	4.35	Bright Band	
VT8 Small Sherwood NCC	VT8S	5.80	Bright Band	
VT9 Big Sherwood NCC	VT9B	3.53	Bright Band	
VT9 Small Sherwood NCC	VT9S	6.60	Bright Band	
VT10 Big Sherwood NCC	VT10B	4.60	Bright Band	
VT10 Small Sherwood NCC	VT10S	7.60	Bright Band	
VT11 Big Sherwood NCC	VT11B	5.90	Bright Band	
VT11 Small Sherwood NCC	VT11S	1.70	Bright Band	
T1 Big Sherwood FE	T1B	4.00	Bright Band	
T1 Small Sherwood FE	T1S	8.60	Bright Band	
T2 Big Sherwood FE	T2B	3.80	Bright Band	
T2 Small Sherwood FE	T2S	1.41	Bright Band	
T3 Big Sherwood FE	T3B	4.37	Bright Band	
T3 Small Sherwood FE	T3S	7.54	Bright Band	

Sample Name	Shortened Sample Name	DNA Conc^{n} ng/µL	1st round PCR result	Notes
T4 Big Sherwood FE	T4B	4.50	Bright Band	
T4 Small Sherwood FE	T4S	15.20	Bright Band	
T5 Big Sherwood FE	T5B	18.40	Bright Band	
T5 Small Sherwood FE	T5S	14.30	Bright Band	
T6 Big Sherwood FE	T6B	16.10	Bright Band	
T6 Small Sherwood FE	T6S	3.14	Bright Band	
T7 Big Sherwood FE	T7B	6.80	Bright Band	
T7 Small Sherwood FE	T7S	13.50	Bright Band	
T8 Big Sherwood FE	T8B	21.10	Bright Band	
T8 Small Sherwood FE	T8S	3.80	Bright Band	
T9 Big Sherwood FE	T9B	4.70	Bright Band	
T9 Small Sherwood FE	T9S	4.42	Bright Band	
T10 Big Sherwood FE	T10B	6.40	Bright Band	
T10 Small Sherwood FE	T10S	8.04	Bright Band	
T11 Big Sherwood FE	T11B	3.30	Bright Band	
T11 Small Sherwood FE	T11S	8.80	Bright Band	

APPENDIX 2 *Meligethes aeneus* COI sequence (AJ536173.1) and primer binding sites for fragment size prediction

ATTTAAAATTTTTCGAATAAATGGCTATTTTCAACTAACCATAAAGATATCGGAACTTTATATTTTATTTTTG
GAGCTTGATCTGGAATAGTAGGTACTTCTTTAAGTATATTAATTCGGACAGAATTAGGTAACCCGGGATCA
CTAATTGGAAATGACCAAATCTATAATGTTATTGTAACAGCCCATGCATTTGTTATAATTTTTTTTATAGTTA
TACCATTTATAATTGGAGGATTTGGAAATTGGCTAGTGCCTCTAATACTAGGGGCCCCTGATATAGCTTTC
CCTCGAATAAATAATATAAGATTTTGACTACTACCTCCTTCCTTGTCCTTACTTTTAATAAGAAGAATTGTA
GAAAGAGGAGCTGGTACTGGATGAACAGTGTACCCACCTTTATCCTCAAATATTGCTCATGGGGGGGCAT
CTGTTGATTTAGCTATTTTTAGCCTTCATTTAGCTGGTATCTCATCTATCTTAGGGGCAGTAAATTTCATTA
CAACTGTAATTAATATACGTCCAAAAGGAATAACATTTGATCGAATACCTTTATTTGTATGAGCAGTAATAA
TTACAGCTATTCTCCTTCTACTATCACTACCAGTATTAGCAGGAGCTATTACAATACTATTAACAGACCGAA
ATCTAAACACAACATTTTTTGACCCTTCTGGGGGAGGTGACCCAATTTTATACCAACATTTATTTTGATTTT
TTGGACATCCAGAAGTATACATTTTAATCCTACCAGGATTTGGAATAATCTCTCATATTATTAGACAAGAAA
GTAGAAAAAAGGAAGCATTCGGAACCCTGGGTATAATTTATGCTATAATAGCAATTGGGCTATTAGGATTT
GTAGTATGAGCTCATCATATATTCACTGTAGGAATAGATGTTGACACACGAGCATATTTTACCTCTGCAAC
TATAATTATTGCAGTACCCACAGGTATTAAAATTTTTAGTTGATTAGCAACTTTACATGGAACACAAATTAA
CTATAGACCTGTAACTTTATGGGCGTTAGGGTTTGTATTTTTATTTACAGTAGGAGGATTAACAGGAGTAA
TTTTAGCAAACTCTTCAATTGATATTGTTTTACATGATACATACTATGTAGTAGCACATTTCCATTATGTATT
ATCAATAGGGGCAGTATTTGCTATCATAGCCGGGCTAGTTCAATGATTCCCATTAATTACAGGATTAACTT
TAAACAATAAATTTTTAAAAATTCAATTCTTTACTATATTTATTGGAGTTAACCTAACATTCTTTCCTCAACAT
TTCTTAGGATTAAGCGGAATACCACGACGATATTCTGATTACCCAGATGCTTATACTCTATGAAATATAACT
TCATCAATTGGATCTTTAATTTCCTTAGTAAGAGTATTATTCTTAATTTTTACAATTTGAGAGGCTTTCTCAG
TTAAACGATTAAATCTTTCATCATTAAATTTAAATACATCTATTGAATGAATACAATCGTACCCACCTGCAG
AACATAGCTATAATGAGCTACCTATCCTAACAAATTTCTAA

mICOIintF

5’-GGWACWGGWTGAACWGTWTAYCCYCC-3’

jgHCO2198 (reverse compliment)

5’-TGYTTRTTRGGICARCCICGAYGTITA-3’

The predicted size of the first round PCR product will be as follows:

- 364 bp for the specific *Meligethes aeneus* COI fragment being amplified
- 67 bp for the Illumina overhang adapters
- Total amplicon size: 431 bp

Table S2 DNA base degeneracy table showing single letter abbreviations for base combinations.

IUPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U for RNA)	Thymine (or Uracil)
I	Inosine
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base

APPENDIX 3. Sherwood FE Birklands sample community composition

Table S3.1 Sample T1 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Ampedus balteatus	10	2,511		2,511
Anaspis frontalis	2			
Anaspis garneysi	1		5	5
Anthrenus fuscus	1		3	3
Atheta nigricornis	0		41	41
Athous haemorrhoidalis	1	36		36
Brassicogethes (Meligethes) aeneus	57		8,044	8,044
Cartodere nodifer	1		6	6
Cis micans	1			
Cortinicara gibbosa	2		33	33
Dalopius marginatus	2	40	10,975	11,015
Deporaus betulae	1		30	30
Dropephylla sp	2			
Enicmus rugosus	1			
Epurea sp. (female)	1			
Ernobius pini	1		12	12
Euophryum confine	8			
Melanotus castanipes	12	22,074		20,074
Melanotus villosus	0	12,251		12,251
Mocyta fungi agg	1			
Phyllodrepa (Dropephylla) ioptera	1		13	13
Quedius xanthopus	2	1,920		1,920
Rhizophagus bipustulatus	1			
Trixagus dermestoides	1			

Table S3.2 Sample T2 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sub-sample	Read Count – 'Small' sub-sub-sample	Total Read Count
Agabus bipustulatus	1			
Amischa decipiens	1		149	149
Ampedus balteatus	5	13,932		13,932
Atheta orbata	0		2,680	2,680
Atheta vaga	1			
Brassicogethes (Meligethes) aeneus	3			
Cerylon ferrugineum	1		26	26
Coeliodinus (Coeliodes) rubicundus	1		51	51
Dalopius marginatus	1	20,304		20,304
Dryocoetes villosus	1		16,516	16,516
Enicmus rugosus	1		78	78
Mocyta fungi agg.	1			
Phyllodrepa (Dropephylla) ioptera	1		22	22
Quedius xanthopus	1			
Strophosoma melanogrammum	1			
Tetratoma fungorum	1		4	4

Table S3.3 Sample T3 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Anthrenus fuscus	0		5	5
Athous haemorrhoidalis	4	9,482		9,482
Brassicogethes (Meligethes) aeneus	2		57	57
Corticarina minuta	0		1,920	1,920
Corticarina fuscula	0		9	9
Cryptophagus scanicus	1		3	3
Dalopius marginatus	1	4,189		4,189
Enicmus testaceus	1		14	14
Epuraea aestiva	1		7	7
Euophryum confine	3			
Melanotus castanipes	2	80		80
Nalassus laevioctostriatus	3	12,351	9	12,360
Phyllobius argentatus	1	7	23,683	23,690
Phyllobius pyri	1			
Rhagium bifasciatum	1	1,990		1,990
Strophosoma capitatum	1		2,249	2,249
Strophosoma melanogrammum	3	56		56

[±] *species nor genus found on BOLD database*

Table S3.4 Sample T4 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Agriotes acuminatus	2			
Ampedus balteatus	4			
Anoplotrupes stercorosus	1			
Athous haemorrhoidalis	1			
Carabus problematicus	1	194		194
Cryptophagus dentatus	2			
Dalopius marginatus	2		96	96
Euophryum confine	5			
Fabogethes nigrescens	0		39	39
Malthodes fuscus	0		29	29
Melanotus villosus	1	33		33
Meligethes nigrescens	1			
Nalassus laevioctostriatus	2			
Nebria sp	1			
Nicrophorus vespilloides	1	11		11
***Othiorhyncus singularis*[±]**	3			
Phyllobius pyri	1			
Prionychus melanarius	0	211		211
Pterostichus niger	1			
Strophosoma melanogrammum	6		32,265	32,265
Trixagus dermestoides	1		892	892

[±] *species nor genus found on BOLD database*

Table S3.5 Sample T5 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Aleochara stichai	1		3	3
Ampedus balteatus	11	4,235		4,235
Anaspis frontalis	1			
Athous haemorrhoidalis	0	3		3
Anisotoma humeralis	1		17,561	17,561
Autalia longicornis	1			
Brassicogethes (Meligethes) aeneus	5		15	15
Cartodere nodifer	3			
Cis bilamellatus	3			
Cis castaneus	1			
Coeliodinus (Coeliodes) rubicundus	1		127	127
Corticarina minuta	1			
Corynoptera trepida[#]	0		98	98
Cryptophagus pubescens	2			
Cryptophagus scanicus	1			
Dalopius marginatus	2	5,346		5,346
Denticollis linearis	1	1,669		1,669
Dropephylla ioptera	1			
Dryocoetes villosus	1		1,961	1,961
Enicmus rugosus	3		274	274
Epuraea biguttata	2			
Euglenes oculatus	9		2,534	2,534
Glischrochilus hortensis	3	36		36
Haploglossa villosula	1			
Lordithon lunulatus	3		4	4
Malthinus frontalis	1		173	173
Melanotus castanipes	2	6,311	3,824	10,135
Melanotus villosus	1	412		412
Nemocestes horni[#]	0	7		7
Nicrophorus vespilloides	1	10		10
Octotemnus glabriculus	1			
Orchestes rusci	1			
Otiorhynchus singularis[*]	1			
Quedius maurus	1	12	3	15
Quedius mesomelinus/maurus	1			
Quedius xanthopus	2	493		493
Sericoderus sp	1			
Strophosoma capitatum	0		8	8

Strophosoma melanogrammum	1		145	145
Triplax russica	1			
Trixagus dermestoides	1		3,282	3,282

[#]*Not a beetle – potential prey species?*

[*]*species not found on BOLD database but genus present*

Table S3.6 Sample T6 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Ampedus balteatus	1	826		826
Anaspis frontalis	2			
Anisotoma humeralis	1			
Athous haemorrhoidalis	6	30,872		30,872
Atomaria fuscata	1			
Brassicogethes (Meligethes) aeneus	207	7	5,932	5,939
Cis bilamellatus	1			
Cis pygmaeus	1			
Coeliodinus (Coeliodes) rubicundus	1			
Corticarina minuta	1			
Cryptophagus pubescens	1			
Ctesias serra	2		27,377	27,377
Dalopius marginatus	3	13		13
Dryocoetes villosus	1		1,205	1,205
Enicmus rugosus	5			
Epuraea biguttata	3			
Euglenes oculatus	11		2,061	2,061,
Euophryum confine	1			
Glischrochilus hortensis	1			
Glischrochilus quadriguttatus	1		6	6
Halyzia sedecimguttata	2		5,653	5,653
Lagria hirta	1	96		96
Melanotus castanipes	5	7,885		7,885
Melanotus villosus	0	248		248
Nalassus laevioctostriatus	0	3		3
Nicrophorus vespilloides	1	71		71
Quedius maurus	1	397	3	397
Quedius xanthopus	3	167		167
Sericoderus sp	1			
Strophosoma capitatum	0		8	8
Strophosoma melanogrammum	1		19	19

Those species highlighted in red were removed from the samples and retained by Natural England.

Table S3.7 Sample T7 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Anisotoma humeralis	1			
Alaus melanops[≠]	0		120	120
Brassicogethes (Meligethes) aeneus	11		2,508	2,508
Cartodere nodifer	3			
Cryptophagus pubescens	7		82	82
Ctesias serra	2		28,029	28,029
Enicmus rugosus	4		188	188
Euglenes oculatus	5		3,226	3,226
Haploglossa gentilis	5			
Haploglossa villosula	1			
Leiopus nebulosus	1		86	86
Malthinus frontalis	0		387	387
Malthodes marginatus	2		19	19
Malthodes maurus?	1			
Melanotus castanipes	1	19,984	122	20,106
Melanotus villosus	1	3		3
Mocyta fungi agg	1			
Nalassus laevioctostriatus	1	19,553	4	19,557
Phloeopora testacea	1			
Phyllobius argentatus	2		217	217
Porcellio scaber[#]	0		4	4
Quedius xanthopus	1	1,647		1,647
Sciodrepoides watsoni	1			
Stenichnus godarti	1			
Strophosoma capitatum	0	4		4
Strophosoma melanogrammum	2		2,260	2,260
Trixagus dermestoides	2		395	395
unknown beetle larvae[±]	1			
Xantholinus longiventris	1		41	41

Those species highlighted in red were removed from the samples and retained by Natural England.

[≠]*Alaus melanops is not found on the UK species list – potential metabarcoding misidentification*

[#]*Not a beetle*

[±] *species nor genus found on BOLD database*

Table S3.8 Sample T8 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Ampedus balteatus	4	2,353		2,353
Anaspis frontalis	1			
Anthrenus fuscus	1		28	28
Archarius pyrrhoceras	1			
Athous haemorrhoidalis	5	1,132		1,132
Brassicogethes (Meligethes) aeneus	25		22	22
Cis villosulus*	1			
Coeliodinus (Coeliodes) rubicundus	1			
Corticarina minuta	1			
Cortinicara gibbosa	1			
Curculionidae sp	1			
Dalopius marginatus	4	3,970	687	4,657
Enicmus rugosus	1			
Euglenes oculatus	19		68	68
Glischrochilus hortensis	10		709	709
Glischrochilus quadriguttatus	4			
Harmonia axyridis	2		18,661	18,661
Melanotus castanipes	6	1,822		1,822
Melanotus villosus	1	8,357		8,357
Melanotus sp.	5			
Mocyta fungi agg	3			
Orchestes quercus	1			
Pediacus dermestoides	6			
Quedius maurus	2	517	34	551
Quedius mesomelinus	1	3	8	11
Quedius mesomelinus/maura	2			
Rhamphus sp	1			
Rhynchaenus quercus	0		3	3
Scolytus intricatus	2			
Strophosoma melanogrammum	1		223	223
Trypodendron domesticum	1			

**species not found on BOLD database but genus present*

Table S3.9 Sample T9 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Ampedus balteatus	6	4,954		4,954
Ampedus quercicola	1	35		35
Anotylus tetracarinatus	1			
Brassicogethes (Meligethes) aeneus	4		419	419
Cartodere nodifer	1		17,456	17,456
Corynoptera trepida[#]	0	38		38
Cryptophagus scanicus	1		3	3
Ctesias serra	2		14,760	14,760
Entomobrya nivalis[#]	0		8	8
Euglenes oculatus	1		6,044	6,044
Glischrochilus hortensis	1		15	15
Haploglossa villosula	2			
Melanotus castanipes	5	20,134	2,959	23,093
Melanotus villosus	0	986		986
Pediacus dermestoides	1			
Phloeopora testacea	4		10	10
Quedius xanthopus	1			
Rhizophagus bipustulatus	1			
Trypodendron domesticum	1		203	203

[#]*Not a beetle – potential prey species?*

Table S3.10 Sample T10 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Ampedus balteatus	1	472		472
Anaspis garneysi	1			
Anotylus rugosus	1		10	10
Aphodius prodromus	1		357	357
Athous haemorrhoidalis	3	66		66
Atomaria fuscata	1			
Brassicogethes (Meligethes) aeneus	2		3	3
Bruchus dentipes[¥]	0		669	669
Bruchus rufimanus	1			
Cartodere nodifer	1			
Cis hispidus	1		6	6
Coccinella septempunctata	1	9		9
Cryptophagus scanicus	2		3	3
Ctesias serra	4		52,225	52,225
Dalopius marginatus	2	3,485		3,485
Dryocoetes villosus	1		241	241
Enicmus testaceus	2			
Haploglossa villosula	9			
Melanotus castanipes	4	859		859
Melanotus villosus	1	2,882		2,882
Orchestes quercus	1		283	283
Pediacus dermestoides	2			
Rhynchaenus quercus	0		11,397	11,397
Stenagostus rhombeus[§]	0	31,319		31,319
Tachyporus hypnorum	1		9	9
Trypodendron domesticum	1		5	5

[*] *species not found on BOLD database but genus present*

[¥] *Bruchus dentipes not found on UK species list – potential metabarcoding misidentification of Bruchus rufimanus*

[§] *Potential taxonomic misidentification*

Table S3.11 Sample T11 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Amischa decipiens	1		18	18
Ampedus balteatus	1	622		622
Anaspis maculata	2		23	23
Anisotoma humeralis	1			
Anobium punctatum	1			
Brassicogethes (Meligethes) aeneus	33	14	26,655	26,669
Byturus tomentosus	1		3	3
Corynoptera trepida[#]	0		5	5
Cryptarcha strigata	7			
Cryptophagus pubescens	1			
Cryptophagus scanicus	1		1,436	1,436
Dalopius marginatus	1	178		178
Dromius quadrimaculatus	1			
Dryocoetes villosus	11			
Epuraea biguttata	12			
Euglenes oculatus	3		2,701	2,701
Glischrochilus hortensis	5			
Glischrochilus quadriguttatus	2			
Haploglossa gentilis	1			
Melanotus castanipes	5	11,630		11,630
Melanotus villosus	0	122		122
Nalassus laevioctostriatus	1			
Nicrophorus humator	1	10,886		10,889
Philonthus fumarius	0		4	4
Quedius xanthopus	1			
Rhizophagus dispar	1		1,027	1,027
Rhizophagus ferrugineus	1			
Rhynchaenus rusci	0		45	45
Salpingus planirostris	0		11	11
Triphyllus bicolor	1			
Triplax russica	1			
Trypodendron domesticum	3		1,301	1,301
Velleius dilatatus	2			

[#]*Not a beetle - potential prey species?*

Table S3.12 Total Sherwood FE community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Agabus bipustulatus	1			
Agriotes acuminatus	2			
Alaus melanops≠	0		120	120
Aleochara stichai	1		3	3
Amischa decipiens	2		367	367
Ampedus balteatus	43	29,909		29,905
Ampedus quercicola	2	35		
Anaspis frontalis	6			
Anaspis garneysi	2		5	5
Anaspis maculata	2		23	23
Anisotoma humeralis	4		17,561	17,561
Anobium punctatum	1			
Anoplotrupes stercorosus	1			
Anotylus rugosus	1		10	10
Anotylus tetracarinatus	1			
Anthrenus fuscus	2		36	36
Aphodius prodromus	1		357	357
Archarius pyrrhoceras	1			
Atheta nigricornis	0		41	41
Atheta orbata	0		2,680	2,680
Atheta vaga	2			
Athous haemorrhoidalis	20	41,591		41,591
Atomaria fuscata	2			
Autalia longicornis	1			
Brassicogethes (Meligethes) aeneus	349	21	43,655	43,676
Bruchus dentipes¥	0		669	669
Bruchus rufimanus	1			
Byturus tomentosus	1		3	3
Carabus problematicus	1	194		194
Cartodere nodifer	9		17,462	17,462
Cerylon ferrugineum	1		26	26
Cis bilamellatus	4			
Cis castaneus	1			
Cis hispidus	1		6	6
Cis micans	2			

Cis pygmaeus	1			
Cis villosulus*	1			
Coccinella septempunctata	1	9		9
Coeliodinus (Coeliodes) rubicundus	5		178	178
Corticarina fuscula	0		9	9
Corticarina minuta	3		1,920	1,920
Cortinicara gibbosa	3		33	33
Corynoptera trepida[#]	0	38	103	141
Cryptarcha strigata	7			
Cryptarcha undata	1			
Cryptophagus dentatus	2			
Cryptophagus pubescens	12		85	85
Cryptophagus scanicus	6		1,445	1,445
Ctesias serra	10		122,391	122,391
Curculionidae sp	1			
Dalopius marginatus	18	37,525	11,758	49,283
Denticollis linearis	1	1,669		1,669
Deporaus betulae	1		30	30
Dromius quadrimaculatus	1			
Dropephylla ioptera	3			
Dropephylla sp	2			
Dryocoetes villosus	15		19,923	19,923
Enicmus rugosus	15		540	540
Enicmus testaceus	3		14	14
Entomobrya nivalis[#]	0		8	8
Epuraea aestiva	2		7	7
Epuraea biguttata	19			
Ernobius pini	1		12	12
Euglenes oculatus	48		16,634	16,634
Euophryum confine	17			
Fabogethes nigrescens	0		39	39
Glischrochilus hortensis	20	36	730	766
Glischrochilus quadriguttatus	7		6	6
Halyzia sedecimguttata	2		5,653	5,653
Haploglossa gentilis	6			
Haploglossa villosula	13			
Harmonia axyridis	2		18,661	18,661
Lagria hirta	1	96		96
Leiopus linnei	0		86	86

Leiopus nebulosus	1			
Lordithon lunulatus	3		4	4
Malthinus frontalis	1		560	560
Malthodes fuscus	0		29	29
Malthodes marginatus	2		19	19
Malthodes maurus?	1			
Melanotus castanipes	42	90,779	6,905	95,684
Melanotus sp.	5			
Melanotus villosus	5	25,294		25,294
Meligethes nigrescens	1			
Mocyta fungi agg	7			
Nalassus laevioctostriatus	7	31,907	13	31,920
Nebria sp	1			
Nemocestes horni[#]	0	7		7
Nicrophorus humator	1	10,886		10,889
Nicrophorus vespilloides	3	92		92
Octotemnus glabriculus	1			
Orchestes quercus	2		283	283
Orchestes rusci	1			
Othiorhyncus singularis[±]	4			
Pediacus dermestoides	9			
Philonthus fumarius	0		4	4
Phloeopora testacea	6		10	10
Phyllobius argentatus	3	7	23,900	23,900
Phyllobius pyri	4			
Phyllodrepa (Dropephylla) ioptera	3		35	35
Porcellio scaber[#]	0		4	4
Prionychus melanarius	0	211		211
Pterostichus niger	1			
Quedius maurus	5	926	40	966
Quedius mesomelinus	1	3	8	11
Quedius mesomelinus/maurus	3			
Quedius xanthopus	12	4,227		4,227
Rhagium bifasciatum	1	1,990		1,990
Rhamphus sp.	1			
Rhizophagus bipustulatus	2			
Rhizophagus dispar	1		1,027	1,027
Rhizophagus ferrugineus	1			
Rhynchaenus quercus	0		11,400	11,400

Rhynchaenus rusci	0		45	45
Salpingus planirostris	0		11	11
Sciodrepoides watsoni	1			
Scolytus intricatus	2			
Sericoderus sp	2			
Stenagostus rhombeus	0	31,319		31,319
Stenichnus godarti	1			
Strophosoma capitatum	1	4	2,265	2,269
Strophosoma melanogrammum	15	56	34,912	34,968
Tachyporus hypnorum	1		9	9
Tetratoma fungorum	1		4	4
Triphyllus bicolor?	1			
Triplax russica	2			
Trixagus dermestoides	5		4,569	4,569
Trypodendron domesticum	6		1,509	1,509
***unknown beetle larvae*[±]**	1			
Velleius dilatatus	3			
Xantholinus longiventris	1		41	41

[*]*species not found on BOLD database but genus present*

[≠]*Alaus melanops not found on the UK species list – potential metabarcoding misidentification*

[#]*not a beetle – potential prey species?*

[±] *species nor genus found on BOLD database*

[¥]*Bruchus dentipes not found on UK species list – potential metabarcoding misidentification of Bruchus rufimanus*

APPENDIX 4. Sherwood NCC sample community composition

Table S4.1 Sample VT1 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Ampedus balteatus	1		55	55
Anaspis maculata	1			
Apthona sp.	1			
Brassicogethes (Meligethes) aeneus	20		1,157	1,157
Cryptarcha undata	0		354	354
Cryptophagus pubescens	2		138	138
Cryptophagus scanicus	1			
Ctesias serra	0		108	108
Dalopius marginatus	1		4,089	4,089
Dromius quadrimaculatus	1		19	19
Dryocoetes villosus	1		84	84
Epuraea unicolor	3		1,216	1,216
Haploglossa villosula	2			
Harmonia axyridis	1		2,766	2,766
Megatoma undata	1		717	717
Melanotus castanipes	10	600		600
Melanotus villosus	2	49,265		49,265
Meligethes nigrescens	1			
Orchestes quercus	1			
Strophosoma capitatum	1		20,960	20,960
Strophosoma melanogrammum	1		1,879	1,879
Xestobium rufovillosum	1			
Xysticus kochi[#]	0	4		4

[#]*not a beetle – potential prey species?*

Table S4.2 Sample VT2 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Ampedus balteatus	4	824		824
Anaspis frontalis	1			
Anaspis septentrionalis	0		84	84
Anthrenus fuscus	1		12,764	12,764
Brassicogethes (Meligethes) aeneus	6		3,512	3,512
Cis bilamellatus	1			
Cryptarcha strigata	1		99	99
Ctesias serra	0		5	5
Curculio glandium	1			
Dalopius marginatus	1	9		9
Dryocoetes villosus	1		2,345	2,345
Enicmus rugosus	3		1,789	1,789
Epuraea unicolor	3		11,254	11,254
Euophryum confine	4			
Glischrochilus hortensis	2			
Glischrochilus quadriguttatus	1		28	28
Gyrophaena sp	1			
Haploglossa gentilis	1			
Haploglossa villosula	23			
Harmonia axyridis	0		3	3
Longitarsus parvulus	3		318	318
Melanotus castanipes	6	957		957
Melanotus villosus	1	14,504		14,504
Mycetochara humeralis	1			
Nalassus laevioctostriatus	1	12,952	5	12,957
Pediacus dermestoides	2			
Plegaderus dissectus	1			
Prionychus melanarius	0 (1 retained)		12	12
Ptinus fur	1			
Strophosoma capitatum	0		330	330
Strophosoma melanogrammum	7			
Triplax russica	2			
Trixagus dermestoides	1		6	6

Those species highlighted in red were removed from the samples and retained by Natural England.

[±]*species nor genus found on BOLD database*

Table S4.3 Sample VT3 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Ampedus balteatus	5	4,208	28,062	32,270
Ampedus cardinalis	1			
Anisotoma humeralis	1		209	209
Anotylus rugosus	1			
Anotylus tetracarinatus	1			
Atheta vaga	1			
Atheta nigricornis	0		3	3
Athous haemorrhoidalis	2	6,016		6,016
Brassicogethes (Meligethes) aeneus	10		644	644
Cartodere nodifer	1			
Cis bilamellatus	1			
Cis micans	1			
Cis pygmaeus?	1			
Cryptophagus pubescens	1			
Cryptophagus scanicus	2		200	200
Dalopius marginatus	1	192		192
Dromius quadrimaculatus	1		247	247
Dryocoetes villosus	1		6	6
Euglenes oculatus	37		18,051	18,051
Haploglossa villosula	4			
Hyperlasion wasmanni[#]	0		110	110
Melanotus castanipes	4	51		51
Melanotus villosus	0	3,020		3,020
Pediacus dermestoides	1			
Phloiotrya vaudoueri	1			
Rhagium bifasciatum	1	77		77
Salpingus ruficollis	1		480	480
Trypodendron domesticum	1			

[#]*not a beetle – potentioal prey species?*

Table S4.4 Sample VT4 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Agriotes acuminatus	1			
Anaspis garneysi	1			
Athous haemorrhoidalis	1	11,463		11,463
Brassicogethes (Meligethes) aeneus	2			
Coeliodinus (Coeliodes) rubicundus	6		371	371
Cryptophagus scanicus	5			
Ctesius serra	1		47	47
Curculionidae sp.	1			
Dacne bipustulata	1			
Enicmus rugosus[±]	1		15	15
Euglenes oculatus	1			
Euophryum confine	1			
Melanotus villosus	1	123		123
Nalassus laevioctostriatus	8	19,428		19,428
Quedius cruentus	1			
Rhizophagus bipustulatus	1			
Strophosoma capitatum	31		63,198	63,198
Trixagus dermestoides	2		25	25

Table S4.5 Sample VT5 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Anaspis garneysi	1		3,699	3,699
Brassicogethes (Meligethes) aeneus	6		299	299
Cartodere nodifer	1			
Corynoptera trepida[#]	0	4		4
Ctesias serra	0	22		22
Euglenes oculatus	3		2,540	2,540
Harmonia axyridis	1	526		526
Nalassus laevioctostriatus	2	46,037	3	46,040
Prionychus melanarius	0	69		69
Ptinus fur	1			
Strophosoma capitatum	3		59,220	59,220

[#]*not a beetle – potential prey species?*

Table S4.6 Sample VT6 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Anaspis frontalis	1			
Anisotoma humeralis	1		60	60
Atheta orbata	0		74	74
Brassicogethes (Meligethes) aeneus	22		10,962	10,962
Cis pygmaeus	1			
Cryptophagus pubescens	1			
Ctesias serra	1		4,178	4,178
Dalopius marginatus	3	1,059	167	1,226
Enicmus rugosus	2		134	134
Euglenes oculatus	1		197	197
Euophryum confine	1			
Haploglossa villosula	1			
Leptusa fumida	1			
Melanotus castanipes	4			
Melanotus villosus	0	15,620		15,620
Mocyta fungi agg	18			
Nalassus laevioctostriatus	1	23,058	6	23,064
Ptinus fur	0		4	4
Serica brunnea	1	13,852		13,852
Strophosoma capitatum	2	5	21,323	21,328
Strophosoma melanogrammum	2		1,557	1,557
Temnocerus nanus	0		945	945
Triphyllus bicolor	2		4	4
Triplax aenea	1			
Trixagus dermestoides	2		585	585

Table S4.7 Sample VT7 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Ampedus balteatus	9	19,836		19,836
Anisotoma humeralis	4		4	4
Cerylon ferrugineum	0		5	5
Cerylon histeroides	1			
Enicmus rugosus	1			
Nalassus laevioctostriatus	1	17,879		17,879
Philonthus carbonarius	1	17		17
Phloiotrya vaudoueri	1			
Rhagonycha fulva	1	85		85
Strophosoma capitatum	0		31,811	31,811
Trixagus dermestoides	1		5,392	5,392

Table S4.8 Sample VT8 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Alaus melanops[≠]	0		4	4
Aleochara curtula	1			
Anisotoma humeralis	1		653	653
Atheta (Acrotona) orbata	0		8,655	8,655
Athous haemorrhoidalis	1	1,827		1,827
Brassicogethes (Meligethes) aeneus	2		4,087	4,087
Cartodere nodifer	1			
Cerylon ferrugineum	1			
Coeliodinus (Coeliodes) rubicundus	0		59	59
Corynoptera trepida[#]	0		8	8
Cryptarcha strigata	1		338	338
Cryptophagus pubescens	0		3,326	3,326
Ctesias serra	1		11,470	11,470
Cyphon pubescens[±]	1			
Epuraea unicolor	6		1,708	1,708
Euophryum confine	1			
Haploglossa gentilis	0			
Haploglossa villosula	5		8	8
Melanotus castanipes	10	661		661
Melanotus villosus	0	29,540	482	30,022
Mocyta fungi agg	1			
Nalassus laevioctostriatus	0	3		3
Rhizophagus fenestralis	1			

[≠]*Alaus melanops not found on the UK species list – potential metabarcoding misidentification*

[#]*not a beetle – potential prey species?*

[±]*species nor genus found on BOLD database*

Table S4.9 Sample VT9 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Anotylus rugosus	1		688	688
Anthrenus fuscus	1		18,290	18,290
Aphthona euphorbiae	1		340	340
Archarius pyrrhoceras	1		5	5
Atheta castanoptera	2		190	190
Atheta nigricornis	1			
Athous haemorrhoidalis	2	843		843
Bolitochara obliqua	1			
Brassicogethes (Meligethes) aeneus	1		9	9
Cis castaneus	2			
Cis nitidus[§]	0		189	189
Cis pygmaeus	2			
Coeliodinus (Coeliodes) rubicundus	1		3,605	3,605
Corticarina gibbosa	1			
Cryptophagus scanicus	2		392	392
Dalopius marginatus	1		7,191	7,191
Epuraea unicolor	1		157	157
Euophryum confine	2			
Glischrochilus hortensis	1		8,550	8,550
Gyrophaena minima	1			
Haploglossa villosula	1			
Melanotus castanipes	2	5,791		5,791
Melanotus villosus	0	17,049		17,049
Mocyta fungi agg	1			
Nalassus laevioctostriatus	1	14,159		14,159
Ptinus fur	2		6	6
Strophosoma melanogrammum	0		3	3
Trixagus dermestoides	1		20	20

[§]*potential taxonomic misidentification*

Table S4.10 Sample VT10 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Anisotoma humeralis	2		5	5
Anthrenus fuscus	0		14,742	14,742
Brassicogethes (Meligethes) aeneus	33		8,752	8,752
Cerylon histeroides	1			
Cerylon ferrugineum	0		4	4
Ctesius serra	1		5,194	5,194
Enicmus testaceus	1			
Euglenes oculatus	4		2,264	2,264
Euophryum confine	1			
Melanotus castanipes	4	11,248		11,248
Melanotus villosus	0	1,930		1,930
Mocyta fungi agg.	1			
Procraerus tibialis	2	5,087		5,087
Ptinus fur	1		199	199
Rhizophagus bipustulatus	1		52	52

Table S4.11 Sample VT11 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Agriotes acuminatus	1			
Ampedus balteatus	2	91		91
Ampedus pomorum	1			
Anaspis fasciata*	1			
Anaspis garneysi	1		320	320
Atheta nigricornis	1			
Brassicogethes (Meligethes) aeneus	56		2,080	2,080
Carpophilus marginellus	1		37	37
Cerylon ferrugineum	1			
Corticarina gibbosa	1			
Cryptarcha strigata	1		1046	10465
Cryptophagus pubescens	1		84	84
Cryptophagus scanicus	1		5	5
Dalopius marginatus	0	36		36
Dromius quadrimaculatus	0		217	217
Dryocoetes villosus	12		31,643	31,643
Elateroides dermestoides	0	18,275	61	18,336
Epuraea biguttata[§]	4			
Epuraea marseuli	0		4	
Epuraea unicolor	0		185	
Euglenes oculatus	3		100	100
Euophryum confine	1			
Glischrochilus hortensis	1		2,773	2,773
Glishcrochilus quadriguttatus	2		50	50
Harmonia axyridis	1	13,994	148	14,142
Hylocereus dermestoides[±]	6			
Megatoma undata	1			
Melanotus castanipes	3	27		27
Melanotus villosus	1	1,295	669	1,964
Mocyta fungi agg	2			
Mycetophagus quadriguttatus	1			
Orchestes rusci	1			
Plegaderus dissectus	2			
Prionychus melanarius	0	330	10	340
Pseudocistela ceramboides	1	8,013		8,013
Rhizophagus bipustulatus	1			
Rhizophagus dispar	0		13	13
Rhizophagus nitidulus	0		206	206
Salpingus planirostris	1			
Stenichnus godarti	1			
Triplax russica	1			
Trypodendron domesticum	5		1,419	1,419

Velleius dilatatus	0	1,320		1,320
Xyleborinus saxeseni	1		12	12

[*]*species not found on BOLD database but genus present*

[±]*species nor genus found on BOLD database*

Table S4.12 Total Sherwood NCC community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Agriotes acuminatus	2			
Alaus melanops[≠]	0		4	4
Aleochara curtula	1			
Ampedus balteatus	21	24,959	28,117	53,076
Ampedus pomorum	1			
Ampedus quercicola	1			
Anaspis fasciata[*]	1			
Anaspis frontalis	2			
Anaspis garneysi	4		4,019	4,019
Anaspis maculata	1			
Anaspis septentrionalis	0		84	84
Anisotoma humeralis	9		931	931
Anotylus rugosus	2		688	688
Anotylus tetracarinatus	1			
Anthrenus fuscus	2		45,796	45,796
Aphthona euphorbiae	1		340	340
Apthona sp.	1			
Archarius pyrrhoceras	4		5	5
Atheta castanoptera	2		190	190
Atheta nigricornis	2		3	3
Atheta (Acrotona) orbata	0		8,729	8,729
Atheta vaga	1			
Athous haemorrhoidalis	6	20,149		20,149
beetle bits	?			
Bolitochara obliqua	1			
Brassicogethes (Meligethes) aeneus	158		31,502	31,502
Carpophilus marginellus	1		37	37
Cartodere nodifer	3			
Cerylon ferrugineum	2		9	9
Cerylon histeroides	2			
Cis bilamellatus	2			
Cis castaneus	2			
Cis micans	1			
Cis nitidus[§]	0		189	189
Cis pygmaeus	4			

Coeliodinus (Coeliodes) rubicundus	7		4,035	4,035
Corticarina gibbosa	2			
Corynoptera trepida[#]	0	4	8	12
Cryptarcha strigata	3		1,837	1,837
Cryptophagus pubescens	5		3,548	3,548
Cryptophagus scanicus	12		597	597
Ctesias serra	4	22	21,002	21,024
Curculio glandium	1			
Curculionidae sp	1			
Cyphon pubescens[±]	1			
Dacne bipustulata	1			
Dalopius marginatus	7	1,296	11,447	12,746
Dromius quadrimaculatus	2		483	483
Dryocoetes villosus	15		34,078	34,078
Elateroides dermestoides	0	18,275	61	18,336
Enicmus rugosus	7		1,938	1,938
Enicmus testaceus	1			
Epuraea biguttata[§]	4			
Epuraea marseuli	0		4	4
Epuraea unicolor	13		14,520	14,520
Euglenes oculatus	49		23,152	23,152
Euophryum confine	11			
Glischrochilus hortensis	4		11,323	11,323
Glischrochilus quadriguttatus	3		78	78
Gyrophaena minima	1			
Gyrophaena sp	1			
Haploglossa gentilis	2			
Haploglossa villosula	36		8	8
Harmonia axyridis	3	14,520	2,917	17,437
Hylocereus dermestoides[±]	6			
Hyperlasion wasmanni[#]	0		110	110
Leptusa fumida	2			
Longitarsus parvulus	3		318	318
Megatoma undata	2		717	717
Melanotus castanipes	44	600		19,335
Melanotus villosus	4	132,336	669	133,005

Meligethes nigrescens	1			
Mocyta fungi agg	23			
Mycetochara humeralis	2			
Mycetophagus quadriguttatus	1			
Nalassus laevioctostriatus	14	133,516	14	133,530
Orchestes quercus	1			
Orchestes rusci	1			
Pediacus dermestoides	3			
Philonthus carbonarius	1	17		17
Phloiotrya vaudoueri	2			
Plegaderus dissectus	3			
Prionychus melanarius	0 (both retained)	399	22	421
Procraerus tibialis	2	5,087		5,087
Pseudocistela ceramboides	1	8,013		8,013
Ptinus fur	5		209	209
Quedius cruentus	1			
Rhagium bifasciatum	1	77		77
Rhagonycha fulva	1	85		85
Rhizophagus bipustulatus	3		52	52
Rhizophagus dispar	0		13	13
Rhizophagus fenestralis	1			
Rhizophagus nitidulus	0		206	206
Salpingus planirostris	1			
Salpingus ruficollis	1		480	480
Serica brunnea	2	13,852		13,852
Stenichnus godarti	1			
Strophosoma capitatum	37		196,842	196,842
Strophosoma melanogrammum	10		3,439	3,439
Temnocerus nanus	0		945	945
Triphyllus bicolor	2		4	4
Triplax aenea	1			
Triplax russica	3			
Trixagus dermestoides	7		6,028	6,028
Trypodendron domesticum	6		1,419	1,419
Velleius dilatatus	0	1,320		1,320

Xestobium rufovillosum	1			
Xyleborinus saxeseni	1		12	12
Xysticus kochi[#]	0	4		4

[≠]Alaus melanops not found on the UK species list – potential metabarcoding misidentification

[§]potential taxonomic misidentification

[#]not a beetle – potential prey species?

[*]species not found on BOLD database but genus present

[±]species nor genus found on BOLD database

APPENDIX 5. Highcliffe Beetles community composition

Table S5.1 Highcliffe beetles site 1 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Acupalpus flavicollis	2			
Aleochara bipustulata	5		28	28
Altica lythri	1			
Amara aenea	1			
Anisodactylus binotatus	0	192		192
Anotylus rugosus	1		23	23
Anotylus tetracarinatus	9		4	4
Aphthona euphorbiae	1			
Barypeithes pellucidus	1		26	26
Bembidion articulatum	4		4	4
Bembidion cruciatum	11		12,020	12,020
Bembidion deletum	29		25,063; 10,454	35,517
Bembidion illigeri	14		1,094; 329	1,423
Bembidion lunulatum	3		4	4
Bembidion quadrimaculatum	4		10; 24	34
Bembidion stephensii	2		6,116	122
Brassicogethes (Meligethes) aeneus	17		123; 11	134
Byrrhus pilula	1	2,461	9	2,470
Carpelimus corticinus	1			
Carpelimus incongruous	1			
Cheilotrichia cinerascens[#]	0		6	6
Chlaenius vestitus	11	23	5	28
Cicindela campestris	3	12,102		12,102
Ctesias serra	0		664	664
Coccidula scutellata	1			
Corynoptera trepida[#]	0		9	9
Dyschirius politus	3			
Harpalus affinis	1			
Helophorus aequalis	3		4	4

Helophorus grandis	1			
Helophorus minutus	2			
Heterocerus fusculus	24		591; 67	658
Ischnopterapion loti	1			
Laccobius sinuatus	1			
Laccobius striatulus	1			
Lobrathium multipunctum*	1			
Longitarsus parvulus	1			
Mecinus pascuorum	1			
Nebria brevicollis	38	42,898; 9,842		52,740
Nebria salina	9	455; 189	20	664
Notiophilus substriatus	8			
Paranchus albipes	2	304	51	355
Parocyusa (Tetralaucopora) longitarsis	6		12	12
Philonthus quisquiliarius	1			
Philhygra palustris	0		5	5
Prinerigone vagans[#]	0	7		7
Propylea quatuordecimpunctata	1		849	549
Pterostichus nigrita	1	2,272		2,272
Scatella paludum[#]	0	3		3
Sitona lineatus	4		83; 281	364
Stenolophus teutonus	2		2,427	2,427
Tachys bistriatus	2			

Hand search counts are shown in black and pitfall trap counts in blue text.

[*]*species not found on BOLD database but genus present*

[#]*not a beetle – potential prey species?*

Table S5.2 Highcliffe beetles site 2 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Acupalpus flavicollis	3			
Acupalpus meridianus	1			
Acupalpus parvulus	1			
Agelastica alni	1			
Agonum viduum	1			
Aleochara bipustulata	3		6	6
Altica chamaenerii[≠]	0		3	3
Altica lythri	4		1,328; 296	1,624
Altica longicollis	0		12	12
Amischa analis	1			
Anacaena limbata	1			
Anotylus nitidulus	2			
Anotylus rugosus	3		233	23
Anotylus tetracarinatus	11			
Bembidion articulatum	3			
Bemdidion cruciatum	1		4	4
Bembidion deletum	3		1,084	1,084
Bembidion dentellum	1			
Bembidion illigeri	31	4	26; 2,035	2,065
Bembidion lunulatum	1			
Brassicogethes (Meligethes) aeneus	10	4	13; 8	25
Carpelimus corticinus	8			
Carpelimus erichsoni	14			
Carpelimus rivularis	1			
Cercyon ustulatus	1			
Chaetocnema hortensis	8			
Chlaenius vestitus	12	29		29
Coccidula rufa	1		7	7
Coelostoma orbiculare	4		64	64
Crepidodera fulvicornis	2		22	22
Ctesias serra	0	5	79	84
Curimopsis setigera[*]	1			
Dinaraea angustula	3			
Drusilla canaliculata	13	20	9; 107	136
Dryops luridus	4			
Dyschirius aeneus	9		4	4
Entomobrya lanuginose[#]	0		67	67

Erichsonius cinerascens	4		16	16
Euaesthetus laeviusculus	1			
Georissus crenulatus	1			
Gnypeta carbonaria	1			
Grypus equiseti	2		23	23
Gyrinus substriatus	1			
Helophorus aequalis	1			
Helophorus brevipalpis[§]	0		4,135	4,135
Helophorus minutus	2			
Heterocerus fusculus	50		132; 5,163	5,295
Hydroporus planus	1			
Hygrotus confluens	1			
Ilybius fuliginosus	1			
Ischnopterapion loti	0		826	826
Isotomurus plumosus[#]	0		40	40
Laccobius minutus	3			
Laccobius sinuatus	2			
Laccobius striatulus	3			
Lasius niger[#]	0		36	36
Leiodes lunicollis[*]	2			
Leistus fulvibarbis	2	6,629		6,629
Leistus spinibarbus	1			
Limnichus pygmaeus	3			
Longitarsus dorsalis	4		112	112
Longitarsus parvulus	3		313	313
Mocyta fungi agg.	1			
Nebria brevicollis	5	117		117
Nebria salina	2	4		4
Notaris scirpi	1			
Noterus clavicornis	1			
Oulema melanopus	1		6,176	6,176
Pachnida nigella	4			
Paranchus albipes	6	8	73	81
Parocyusa (Tetralaucopora) longitarsis	2		5	5
Phalangium opilio[#]	0		15	15
Plateumaris sericea	1			
Platystethus alutaceus	2			
Psilothrix viridicoeruleus	4			
Pterostichus madidus	2	159		159
Pterostichus nigrita	18	95	39	134
Pterostichus rhaeticus	0	6		6
Quedius schatzmayri	1			
Rhinoncus pericarpius	2			
Sitona cylindricollis	1			
Sitona lepidus	1			

Sitona lineatus	1			
Stenolophus mixtus	1			
Stenolophus teutonus	5		26,732	26,732
Stenus fulvicornis	1			
Stenus guttula	7		7	7
Stenus pallipes	1			
Stenus providus	2		20	20
Stenus pusillus	2	12		12
Tachyporus nitidulus	3		3	3
Tachys bistriatus	7			
Telmatophilus typhae	7		9	9
Tetartopeus angustatus*	4			
Thinobius brevipennis*	3			
Tipula lateralis[#]	0	24		24
Tipula oleracea[#]	0	955		955
Xantholinus longiventris	1			

Hand search counts are shown in black and pitfall trap counts in blue text. There were no specimens classified as 'big' for the hand search sample.

[≠]*Altica chamaenerii not found on the UK species list – potential metabarcoding misidentification*

[#]*not a beetle – potential prey species?*

[§]*Potential taxonomic misidentification*

[*]*species not found on BOLD database but genus present*

Table S5.3 Highcliffe beetles site 3 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Acupalpus dubius	1			
Acupalpus flavicollis	2			
Acupalpus parvulus	4		14	14
Aleochara bipustulata	1			
Altica aenescens[≠]	0		7	7
Altica chamaenerii[≠]	0		5	5
Altica longicollis	0		67	67
Altica lythri	3		3,343	3,343
Anotylus tetracarinatus	6		8	8
Bembidion articulatum	6			
Bembidion illigeri	24		79; 271	350
Bembidion lunulatum	1		4	4
Bradycellus harpalinus	1		81	81
Brassicogethes (Meligethes) aeneus	8		12; 3	15
Carpelimus corticinus	1		299	299
Carpelimus erichsoni	2			
Carpelimus similis	1			
Cercyon haemorrhoidalis	1			
Cercyon ustulatus	1			
Chaetocnema arida	1			
Chlaenius vestitus	14	12,315; 4		12,319
Coccidula scutellata	1		1,170	1,170
Ctesias serra	0	3	5	8
Curimopsis setigera[*]	1			
Demetrias atricapillus	1			
Dryops luridus	1			
Dyschirius aeneus	11			
Dyschirius politus	1			
Elodes pseudo minuta	0		175; 17,682	17,857
Enicmus transversus	1		8	8
Erichsonius cinerascens	1		12	12
Gronops lunatus	1			
Gyrohypnus fracticornis	2	5	11,734	11,739

Helophorus aequalis	1			
Helophorus brevipalpis[§]	0		564	564
Helophorus minutus	2			
Heterocerus fusculus	38		158; 509	667
Hydroglyphus geminus	1			
Isotomurus palustris[#]	0		267	267
Laccobius atratus[*]	1			
Laccobius bipunctatus	0		3	3
Laccobius sinuatus	2			
Laccobius striatulus	1			
Leiodes lunicollis[*]	1			
Limnichus pygmaeus	1			
Lithobius forficatus[#]	0		36	36
Longitarsus parvulus	1			
Nebria breviollis	0	3		3
Nebria salina	2	27,058		27,058
Otiorhynchus ligneus	1			
Paradromius linearis	1		32	32
Paranchus albipes	2	3,793		3,793
Parocyusa (Tetralaucopora) longitarsis	5		38	38
Platystethus alutaceus	2			
Sitona lineatus	2		13	13
Stenolophus teutonus	8		38,223; 1,097	39,320
Stenus cicindeloides	1		3	3
Stenus providus	1			
Tachyporus nitidulus	1		9	9
Tachys bistriatus	6	4		4
Thinobius brevipennis[*]	1			
Tipula lateralis[#]	0	3		3

Hand search counts are shown in black and pitfall trap counts in blue text. No reads were generated for pitfall trap 'big' sample.

[≠]*Altica aenescens and Altica chamaenerii not found on the UK species list – potential metabarcoding misidentification*

[§]*Potential taxonomic misidentification*

[#]*not a beetle – potential prey species?*

[*]*species not found on BOLD database but genus present*

Table S5.4 Highcliffe beetles site 4 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Acupalpus dubius	1			
Agonum muelleri	1			
Agriotes lineatus	1	20		20
Agriotes proximus[≠]	0	33		33
Alaus melanops[≠]	0		5	5
Aleochara bipustulata	4	5		5
Altica palustris	2			
Anotylus tetracarinatus	13		6	6
Apion rubiginosum	1			
Armadillidium vulgare[#]	0	20	4	24
Autalia rivularis	1			
Bembidion deletum	3		22,390; 174	22,564
Bembidion illigeri	4	7	15,517	15,524
Bembidion stephensii	2		8	8
Bembidion tetracolum	1			
Bledius atricapillus[*]	2			
Bledius opacus	2			
Bledius spectabilis	1			
Bradysia scabricornis[#]	0	3		3
Brassicogethes (Meligethes) aeneus	0	172	12	184
Carpelimus corticinus	1			
Carpelimus erichsoni	2			
Cassida hemisphaerica[*]	2			
Cathormiocerus socius[±]	1			
Chaetarthria seminulum	22	4	21	25
Chaetocnema hortensis	7			
Chlaenius vestitus	50	28	3	31
Cicindela campestris	9	2,462		2,462
Corticarina curta	2			
Corynoptera trepida[#]	0	26	13	39
Cryptops hortensis[#]	0	13		13
Ctesias serra	0	7,403	249	7,652
Drusilla canaliculata	12		106	106

Dryocoetes villosus	0	52		52
Dryops ernesti	14		27	27
Dryops luridus	9		152	152
Dyschirius aeneus	3			
Dyschirius politus	2			
Enicmus transversus	2			
Euscelis confinis[#]	0	5		5
Gabrius coxalus	2		134	135
Georissus crenulatus	79	6	59	65
Glyptotendipes pallens[#]	0	5		5
Harpalus latus	1	8		8
Helophorus minutus	1			
Heterocerus fusculus	1			
Hippuriphila modeeri	5			
Ischnopterapion virens	1			
Isotomurus palustris[#]	0	16	3; 3	22
Kissister minimus[±]	1			
Limnichus pygmaeus	3			
Longitarsus parvulus	1			
Longitarsus pratensis	1			
Mecinus circulates	1			
Meligethes carinulatus[*]	1			
Meligethes lugubris[*]	3			
Molophilus obscurus[#]	0		2,292	2,292
Myrmica scabrinodis[#]	0		33	33
Nebria salina	6	12,837; 3		12,840
Neliocarus faber[±]	2			
Oomorphus concolor[±]	1			
Opatrum sabulosum	1			
Orchestia gammarellus[#]	0	93	41	134
Otiorhynchus ovatus	1		30	30
Oxypoda brevicornis	1			
Paederus littoralis	1			
Paradromius linearis	1			
Paranchus albipes	15	33,819	6	33,825
Phaedon tumidulus	1			
Phalangium opilio[#]	0	6		6
Philorhizus melanocephalus	1		10	10
Platynothrus peltifer[#]	0	6		6
Porcellio scaber[#]	0	41	72	113

Protapion assimile	1			
Psilothrix viridicoeruleus	2			
Pterostichus nigrita	1			
Quedius semiobscurus	1			
Rhyzobius litura	3		25	25
Rhizoglyphus robini[#]	0	3		3
Trachyphloeus (Romualdius) angustisetulus	1			
Scatella paludum[#]	0		5	5
Silpha tristis	18	1,799		1,799
Sitona lineatus	2		29	29
Sitona waterhousei	3			
Sminthurinus elegans[#]	0	6		6
Stenichnus poweri[*]	1			
Stenus guttula	29		4,757	4,757
Stenolophus teutonus	0	3		3
Trichosirocalus troglodytes	1		8	8
Xantholinus longiventris	4		72	72

Hand search counts are shown in black and pitfall trap counts in blue text.

[≠]*Agriotes proximus and Alaus melanops not found on the UK species list – potential metabarcoding misidentification*

[#]*not a beetle – potential prey species?*

[*]*species not found on BOLD database but genus present*

[±]*species nor genus found on BOLD database*

Table S5.5 Total Highcliffe beetles community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Acupalpus dubius	2			
Acupalpus flavicollis	7			
Acupalpus meridianus	1			
Acupalpus parvulus	5		14	14
Agelastica alni	1			
Agonum muelleri	1			
Agonum viduum	1			
Agriotes lineatus	1	20		20
Agriotes proximus[≠]	0	33		33
Alaus melanops[≠]	0		5	5
Aleochara bipustulata	13	5	34	39
Altica aenescens[≠]	0		7	7
Altica chamaenerii[≠]	0		8	8
Altica longicollis	0		79	79
Altica lythri	8		4,967	4,967
Altica palustris	2			
Amara aenea	1			
Amischa analis	1			
Anacaena limbata	1			
Anisodactylus binotatus	0	192		192
Anotylus nitidulus	2			
Anotylus rugosus	4		256	256
Anotylus tetracarinatus	39		18	18
Aphthona euphorbiae	1			
Apion rubiginosum	1			
Armadillidium vulgare[#]	0	20	4	24
Autalia rivularis	1			
Barypeithes pellucidus	1		26	26
Bembidion articulatum	13		4	4
Bembidion cruciatum	12		12,024	12,024
Bembidion deletum	35		59,165	59,165
Bembidion dentellum	1			
Bembidion illigeri	73	11	19,351	19,362
Bembidion lunulatum	5		8	8
Bembidion quadrimaculatum	4		10; 24	34
Bembidion stephensii	4		6,124	6,124
Bembidion tetracolum	1			
Bledius atricapillus[*]	2			
Bledius opacus	2			
Bledius spectabilis	1			
Bradycellus harpalinus	1		81	81
Bradysia scabricornis[#]	0	3		3

Brassicogethes (Meligethes) aeneus	17	176	182	358
Byrrhus pilula	1	2,461	9	2,470
Carpelimus corticinus	11		299	299
Carpelimus erichsoni	18			
Carpelimus incongruous[*]	1			
Carpelimus rivularis	1			
Carpelimus similis	1			
Cassida hemisphaerica[*]	2			
Cathormiocerus socius[±]	1			
Cercyon haemorrhoidalis	1			
Cercyon ustulatus	2			
Chaetarthria seminulum	22	4	21	25
Chaetocnema arida	1			
Chaetocnema hortensis	15			
Cheilotrichia cinerascens[#]	0		6	6
Chlaenius vestitus	87	12,399	8	12,407
Cicindela campestris	12	14,564		14,564
Coccidula rufa	1		7	7
Coccidula scutellata	2		1,170	1,170
Coelostoma orbiculare	4		64	64
Corticarina curta	2			
Corynoptera trepida[#]	0	26	22	48
Crepidodera fulvicornis	2		22	22
Cryptops hortensis[#]	0	13		13
Ctesias serra	0	7,411	997	8,408
Curimopsis setigera[*]	2			
Demetrias atricapillus	1			
Dinaraea angustula	3			
Drusilla canaliculata	25	20	224	244
Dryocoetes villosus	0	52		52
Dryops ernesti	14		27	27
Dryops luridus	14		152	152
Dyschirius aeneus	23		4	4
Dyschirius politus	6			
Elodes pseudo minuta	0		175; 17,682	17,857
Enicmus transversus	3		8	8
Entomobrya lanuginose[#]	0		67	67
Erichsonius cinerascens	5		28	28
Euaesthetus laeviusculus	1			
Euscelis confinis[#]	0	5		5
Gabrius coxalus	2		134	135
Georissus crenulatus	80	6	59	65
Glyptotendipes pallens[#]	0	5		5
Gnypeta carbonaria	1			
Gronops lunatus	1			
Grypus equiseti	2		23	23
Gyrinus substriatus	1			

Gyrohypnus fracticornis	2	5	11,734	11,739
Harpalus affinis	1			
Harpalus latus	1	8		8
Helophorus aequalis	5		4	4
Helophorus brevipalpis	0		4,699	4,699
Helophorus grandis	1			
Helophorus minutus	7			
Heterocerus fusculus	113		1,625	1,625
Hippuriphila modeeri	5			
Hydroglyphus geminus	1			
Hydroporus planus	1			
Hygrotus confluens	1			
Ilybius fuliginosus	1			
Ischnopterapion loti	1		826	826
Ischnopterapion virens	1			
Isotomurus palustris[#]	0	16	173	189
Isotomurus plumosus[#]	0		40	40
Kissister minimus[±]	1			
Laccobius atratus[*]	1			
Laccobius bipunctatus	0		3	3
Laccobius minutus	3			
Laccobius sinuatus	5			
Laccobius striatulus	5			
Lasius niger[#]	0		36	36
Leiodes lunicollis[*]	3			
Leistus fulvibarbis	2	6,629		6,629
Leistus spinibarbus	1			
Limnichus pygmaeus	7			
Lithobius forficatus[#]	0		36	36
Lobrathium multipunctum[*]	1			
Longitarsus dorsalis	4		112	112
Longitarsus parvulus	6		313	313
Longitarsus pratensis	1			
Mecinus circulates	1			
Mecinus pascuorum	1			
Meligethes carinulatus[*]	1			
Meligethes lugubris[*]	3			
Mocyta fungi agg.	1			
Molophilus obscurus[#]	0		2,292	2,292
Myrmica scabrinodis[#]	0		33	33
Nebria brevicollis	43	52,860	20	52,880
Nebria salina	19	40,546	20	40,566
Neliocarus faber[±]	2			
Notaris scirpi	1			
Noterus clavicornis	1			
Notiophilus substriatus	8			
Oomorphus concolor[±]	1			
Opatrum sabulosum	1			

Orchestia gammarellus[#]	0	93	41	134
Otiorhynchus ligneous	1			
Otiorhynchus ovatus	1		30	30
Oulema melanopus	1		6,176	6,176
Oxypoda brevicornis	1			
Pachnida nigella	4			
Paederus littoralis	1			
Paradromius linearis	2		32	32
Paranchus albipes	25	37,924	130	38,054
Parocyusa (Tetralaucopora) longitarsis	13		55	55
Phaedon tumidulus	1			
Phalangium opilio[#]	0	6	15	21
Philhygra palustris	0		5	5
Philonthus quisquiliarius	1			
Philorhizus melanocephalus	1		10	10
Plateumaris sericea	1			
Platynothrus peltifer[#]	0	6		6
Platystethus alutaceus	4			
Porcellio scaber[#]	0	41	72	113
Prinerigone vagans[#]	0	7		7
Propylea quatuordecimpunctata	1		849	549
Protapion assimile	1			
Psilothrix viridicoeruleus	6			
Pterostichus madidus	2	159		159
Pterostichus nigrita	20	2,367	39	2,406
Pterostichus rhaeticus	0	6		6
Quedius schatzmayri	1			
Quedius semiobscurus	1			
Rhinoncus pericarpius	2			
Rhizoglyphus robini[#]	0	3		3
Rhyzobius litura	3		25	25
Trachyphloeus (Romualdius) angustisetulus	1			
Scatella paludum[#]	0	3	5	8
Silpha tristis	18	1,799		1,799
Sitona cylindricollis	1			
Sitona Lepidus	1			
Sitona lineatus	9		406	406
Sitona waterhousei	3			
Sminthurinus elegans[#]	0	6		6
Stenichnus poweri[*]	1			
Stenolophus mixtus	1			
Stenolophus teutonus	15	3	68,479	68,482
Stenus cicindeloides	1		3	3
Stenus fulvicornis	1			
Stenus guttula	36		4,764	4,764

Stenus pallipes	1			
Stenus providus	3		20	20
Stenus pusillus	2	12		12
Tachyporus nitidulus	4		12	12
Tachys bistriatus	15	4		4
Telmatophilus typhae	7		9	9
Tetartopeus angustatus [*]	4			
Thinobius brevipennis [*]	4			
Tipula lateralis [#]	0	27		27
Tipula oleracea [#]	0	955		955
Trichosirocalus troglodytes	1		8	8
Xantholinus longiventris	5		72	72

[≠] *Altica aenescens, Altica chamaenerii, Agriotes proximus and Alaus melanops not found on the UK species list – potential metabarcoding misidentification*

[#]*not a beetle – potential prey species?*

[*]*species not found on BOLD database but genus present*

[±]*species nor genus found on BOLD database*

APPENDIX 6. Highcliffe Spiders community composition

Table S6.1 Highcliffe spiders unit 1 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Araeoncus crassiceps	1			
Arctosa leopardus	0		3,224	3,224
Bathyphantes gracilis	1		28	28
Brassicogethes (Meligethes) aeneus[#]	0	37		37
Corynoptera trepida[#]	0		5	5
Ctesias setta[#]	0	153		153
Diplocephalus cristatus	1		8	8
Erigone atra	1			
Hahnia nava	1			
Nalassus laevioctostriatus[#]	0	4		4
Oedothorax apicatus	0		9	9
Pardosa hortensis	1		16,487	16,487
Pardosa prativaga	0		23	23
Pardosa proxima	1		13	13
Phrurolithus festivus	1			
Pirata piraticus	1	3		3
Prinerigone vagans	1		4	4
Scatella paludum[#]	0	9		9
Strophosoma capitatum[#]	0	6		6
Trochosa ruricola	1	1,941		1,941
Xysticus audax	0		29	29
Xysticus cristatus	1		823	823

[#]*Not a spider – potential prey species*

Table 6.2 Highcliffe spiders unit 2 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Antistea elegans	2		111	111
Araeoncus crassiceps	2			
Araeoncus humilis	1			
Arctosa leopardus	2	6,114	382	6,496
Argenna subnigra*	1			
Brassicogethes (Meligethes) aeneus#	0	15	4	19
Ctesias serra#	0	84		84
Corynoptera trepida#	0	12		12
Diplostyla concolor	1		20	20
Drassodes cupreus	1			
Drassodes lapidosus	0	592		592
Erigone atra	2		152	152
Erigone dentipalpis	1		108	108
Euophrys frontalis	2		47	47
Hahnia nava	2		6	6
Isotomurus palustris#	0	87	7	94
Lasius niger	0		4	4
Melanotus villosus#	0	8		8
Micaria pulicaria	2		69	69
Nalassus laevioctostriatus	0		4	4
Odiellus spinosus	1			
Pachygnatha clercki	1	569		569
Pardosa hortensis	2	297	9,345	9,642
Pardosa nigriceps	1	130	377	507
Pardosa palustris	1	5,196		5,196
Pardosa pullata	1	5	748	753
Phaeocedus braccatus	1			
Phrurolithus festivus	2		104	104
Pirata latitans*	2			
Pirata piraticus	2	4		4
Stenolophus teutonus#	0	3		3
Tetrix ceperoi#	0		10	10
Trochosa robusta	1			
Trochosa ruricola	1	682		682
Xysticus audax	0		65	65
Xysticus cristatus	2		4,637	4,637
Xysticus kochi	2	23,937	14,588	38,525
Zelotes latreillei	2	755		755

#*not a spider – potential prey species?*

**species not found on BOLD database but genus present*

Table S6.3 Highcliffe spiders unit 3 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample[±]	Read Count – 'Small' sub-sample	Total Read Count
Alaus melanops[#]	0		6	6
Antistea elegans	0		311	311
Araeoncus crassiceps	1			
Bolla atahuallpai[#]	0		18	18
Dicranomyia halterella[#]	0		397	397
Erigone dentipalpis	1		18,367	18,367
Hahnia nava	1			
Heterocerus fusculus[#]	0		5	5
Hymenoptera	7			
Isotomurus palustris[#]	0		7	7
Micaria pulicaria	1		875	875
Pardosa hortensis	1		25,158	25,158
Scatella paludum[#]	0		4	4

[±]*there were no specimens classified as 'big' for this sample*

[#]*not a spider – potential prey species?*

Table 6.4 Highcliffe spiders unit 4 community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample±	Read Count – 'Small' sub-sample	Total Read Count
Alopecosa pulverulenta	1		12	12
Arctosa perita	1		501	501
Drassodes cupreus	1		156	156
Heliophanus flavipes	1			
Homalenotus quadridentatus*	1			
Micaria pulicaria	1			
Ozyptila simplex	1		11	11
Pachygnatha degeeri	1		48	48
Pardosa pullata	1		1,273	1,273
Pardosa nigriceps	0		74	74
Pardosa riparia	0		31	31
Thanatus striatus	1			
Tibellus oblongus	1		21,733	21,733
Trochosa terricola	1			
Xysticus cristatus	1		180	180
Xysticus kochi	1		24,516	24,516

**species nor genus found on BOLD database*

±*there were no specimens classified as 'big' for this sample*

Table S6.5 Total Highcliffe spiders community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Alaus melanops[#]	0		6	6
Alopecosa pulverulenta	1		12	12
Antistea elegans	2		422	422
Araeoncus crassiceps	4			
Araeoncus humilis	1			
Arctosa leopardus	2	6,114	3,606	9,720
Arctosa perita	1		501	501
Argenna subnigra*	1			
Bathyphantes gracilis	1		28	28
Bolla atahuallpai[#]	0		18	18
Brassicogethes (Meligethes) aeneus[#]	0	52	4	56
Corynoptera trepida[#]	0	12	5	17
Ctesias setta[#]	0	237		237
Dicranomyia halterella[#]	0		397	397
Diplocephalus cristatus	1		8	8
Diplostyla concolor	1		20	20
Drassodes cupreus	2		156	156
Drassodes lapidosus	0	592		592
Erigone atra	3		152	152
Erigone dentipalpis	2		18,475	18,475
Euophrys frontalis	2		47	47
Hahnia nava	4		6	6
Heliophanus flavipes	1			
Heterocerus fusculus[#]	0		5	5
Homalenotus quadridentatus*	1			
Hymenoptera[#]	7			
Isotomurus palustris[#]	0	87	7	94
Lasius niger	0		4	4
Melanotus villosus[#]	0	8		8
Micaria pulicaria	4		944	944
Nalassus laevioctostriatus[#]	0	4	4	4
Odiellus spinosus	1			
Oedothorax apicatus	0		9	9

Ozyptila simplex	1		11	11
Pachygnatha clercki	1	569		569
Pachygnatha degeeri	1		48	48
Pardosa hortensis	4	297	50,990	50,990
Pardosa nigriceps	1	130	451	581
Pardosa palustris	1	5,196		5,196
Pardosa prativaga	0		23	23
Pardosa proxima	1		13	13
Pardosa pullata	2	5	2,021	2,021
Pardosa riparia	0		31	31
Phaeocedus braccatus	1			
Phrurolithus festivus	3		104	104
Pirata latitans*	2			
Pirata piraticus	3	7		7
Prinerigone vagans	1		4	4
Scatella paludum[#]	0	9	4	13
Stenolophus teutonus[#]	0	3		3
Strophosoma capitatum[#]	0	6		6
Tetrix ceperoi[#]	0		10	10
Thanatus striatus	1			
Tibellus oblongus	1		21,733	21,733
Trochosa robusta	1			
Trochosa ruricola	2	2,623		2,623
Trochosa terricola	1			
Xysticus audax	0		94	94
Xysticus cristatus	8	83684	5,640	5,640
Xysticus kochi	1	59,747	63,623	123,370
Zelotes latreillei	2	755		755

[#]*not a spider – potential prey species?*

**species not found on BOLD database but genus present*

APPENDIX 7. Highcliffe flies community composition

Table S7.1 Highcliffe flies community composition

Species	Number of individuals identified taxonomically	Read Count – 'Big' sub-sample	Read Count – 'Small' sub-sample	Total Read Count
Clinocera stagnalis	2			
Dicranomyia chorea	1		92	92
Dicranomyia goritiensis*	1			
Dicranomyia modesta	2		121	121
Dicranophragma nemorale*	1			
Eloeophila maculata	1			
Eloeophila submarmorata	1		2,979	2,979
Erioptera fusculenta*	1			
Geomyza tripunctata	1		269	269
Gonomyia sp. Female	1			
Hydrellia maura	0		50	50
Ilisia maculata*	1			
Melieria omissa	1			
Metriocnemus fusipes	0		4	4
Molophilus obscurus	1		4,816	4,816
Parydra littoralis*	1			
Phylidorea ferruginea	1		15	15
Pseudolimnophila lucorum	0	75	44	119
Pseudolimnophila sepium	1			
Ptychoptera contaminata	1			
Scathophaga stercoraria	1	484		484
Scatella paludum	0		195	195
Suillia imberbis*	1			
Sylvicola punctatus	2			
Sylvicola stackelbergii	0		102	102
Symplecta stictica	1		53	53
Terellia ruficauda	0		195	195
Tetrix ceperoi	2	2,583		2,583
Tipula lateralis	1	53		53
Tipula oleracea	1	1,015		1,015

**species not found on BOLD database but genus present*

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

Natural England publications are available as accessible pdfs from www.gov.uk/natural-england.

Should an alternative format of this publication be required, please contact our enquiries line for more information: 0300 060 3900 or email enquiries@naturalengland.org.uk.

ISBN 978-1-78354-894-1

Catalogue code: NECR388

NATURAL ENGLAND