Nested metabarcode tagging: a robust tool for studying species interactions in ecology and evolution

James JN Kitson[1], Christoph Hahn[1], Richard J Sands[2,3], Nigel A Straw[2], Darren M Evans[1], David H Lunt[1]

1. Evolutionary and Environmental Genomics Group, School of Biological, Biomedical and Environmental Biology, University of Hull, Hull, UK
2. Centre for Ecosystems, Society and Biosecurity, Forest Research, Alice Holt Lodge, Farnham, Surrey, UK
3. Centre for Biological Sciences, The University of Southampton, Highfield Campus, Southampton, UK.

Abstract

Here we present a next generation sequencing approach for use in ecological studies which allows for individual level sample tracking through the use of hierarchically organised molecular identification tags. We demonstrate its utility on a sample data set examining levels of parasitism in a recently arrived invasive species in the UK and discuss further possibilities for our approach in ecology and evolution.

Introduction

Massively parallel ‘next generation’ sequencing (NGS) has made an enormous impact on many diverse areas of biology such as evolution (Heliconius Genome Consortium, 2012; Wagner et al.,

2013) and immunology (Spurgin et al., 2011). The quantity of sequence data produced by NGS surveys whole genomes at depth, and the field of population genomics has added much to our ability to characterise complex ecological and evolutionary systems (e.g. Toju et al., 2014). The nature of NGS however does not lend itself naturally to certain types of experimental design. This is a common situation in ecology and evolution where sequence data is required from a large number of individuals for accurate population genetic metrics, to achieve sampling across a broad geographic range, or to produce large scale *individual-level* information to describe an ecological community and its interactions. Pool-seq type approaches (Schlötterer et al., 2014) can represent population diversity but lose the individual level metadata, whereas genome reduction approaches such as RAD-seq (Davey et al., 2010) can increase the number of individuals processed but still do not easily scale to large sample sizes. For many studies however genome-wide sequence data is not required and taxonomic identifications can be reliably recovered from a single locus.

A common approach in ecology and evolution is 'community metabarcoding' where a bulk DNA sample from one environment is PCR amplified for a standard barcode locus, sequenced, and taxa comprising the community identified bioinformatically (e.g. Taberlet et al., 2012; Yu et al., 2012). Information however is at the level of the ecological sample rather than the individual, which can restrict its utility. We suggest that NGS has not penetrated ecology and evolution as pervasively as it would otherwise have done because there has been no widely applied method to metabarcode while preserving individual organism level metadata.

Here, we describe an experimental demonstration of a simple 'nested metabarcoding' approach generating large amounts of NGS data linked unambiguously to the source sample (individual). This protocol is straightforward to scale to thousands of individuals in a single experiment using an inexpensive design ideally suited to population level metabarcoding in ecology and evolution. We also discuss how the ability to link metadata to individuals also opens many new avenues for research including the ability to create larger more highly resolved ecological networks for habitat management and restoration (Evans et al., 2016 - in final review).

The use of unique MID tags (8-mer oligonucleotide sequences) added to the PCR primers is a well-tested approach to multiplex samples in a single NGS experiment (Binladen et al., 2007). Large numbers of unique MID-labelled primers can be expensive however, and complex to organise in a laboratory environment, making it unusual to have more than 96-384 primer combinations in a single

52 experiment. This bias towards small numbers of tags per run leads to relatively few samples each
53 being very deeply sequenced, whether this is required or not, and it is likely that the cost per sample
54 may not be particularly favourable compared to Sanger sequencing. If the number of samples could
55 be greatly increased, however, this would allow cost-effective population level experiments.

56 We demonstrate the utility of nested metabarcoding to quantify and identify parasitism rates in the
57 invasive lepidopteran *Thaumetopoea processionea*, the Oak Processionary Moth (hereafter OPM).
58 We amplify part of the COI barcode fragment for 919 individuals in a single sequencing run and
59 identify parasitoid sequences from caterpillar DNA extractions, describing in detail for the first time
60 parasitism rates for one population of this important invasive species. Our approach is a hierarchical
61 (or "nested") tagging design using MID-tags introduced both with unique PCR primer combinations
62 and also with library-specific sequence tags. This approach makes the tag number multiplicative
63 rather than additive and we show that we can track populations of individuals without a significant
64 change in laboratory protocol complexity. There is a large literature of different varieties of Illumina
65 tagging and library construction and while this nested tagging approach of employing both PCR and
66 library tags has been reported in part before (e.g. Daigle, Simen & Pochart, 2001; Binladen et al.,
67 2007; Tang et al., 2015), our modifications allow us to incorporate a very large number of samples in
68 a single sequencing run and still pull apart individual level information afterwards. With this, we can
69 link sequences to individual level metadata (e.g. locations and ecological measurements) allowing us
70 to apply molecular tools to previously difficult to address ecological questions that require robustly
71 quantified data. This will open up fields such as dietary and parasitism studies which were traditionally
72 observational in nature to molecular investigation and allows us to more easily investigate these
73 phenomena at population or community scales by producing large, phylogenetically referenced
74 ecological networks. This is something that standard metabarcoding and the newer PCR-free
75 metabarcoding (Tang et al., 2015) cannot currently do within reasonable financial and time
76 constraints.

77

Methods

The nested metabarcoding approach.

We employ a modification to the standard Illumina 16S bacterial metabarcoding protocol (Illumina, 2011). In the original protocol two rounds of PCR were used to; (PCR1) isolate and amplify the gene region of interest and then (PCR2) add a set of molecular identification tags (MID tags) and the Illumina MiSeq adapter sequences. Here we do the same but add an additional set of MIDs in PCR1 to further increase the resolution of sample identification. Each MID tag is composed of a unique 8-nucleotide sequence allowing them to be bioinformatically linked back to the individual sample. We include MIDs in both the forward and reverse primers with either eight unique forward tags and twelve unique reverse tags, or *vice versa*, giving 96 unique combinations of tags that can be arranged on a plate (See Fig1 for general primer design). A plate of PCRs with these tagged primers is carried out with each PCR well being given a unique combination of tags. The PCR products are then pooled into a separate pre-library for each plate of samples (PCR1, Fig2). The pre-library is then used as a template for a second round of PCR which adds the adapters necessary for Illumina sequencing. This reaction also adds two additional MID tags that uniquely identify the plate (PCR2, Fig2). These tagged pre-libraries can then be purified, pooled and sequenced on a single Illumina MiSeq run.

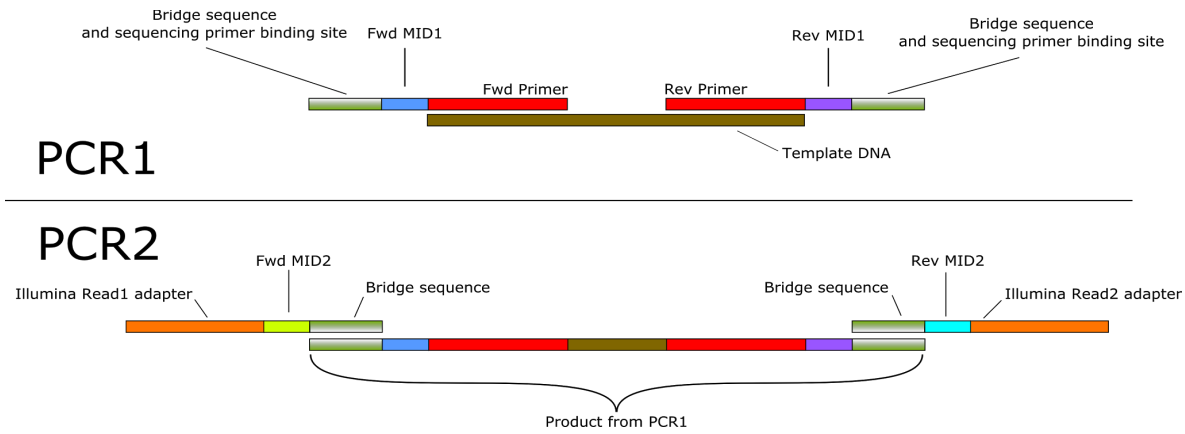


Fig1. Simplified primer structure for a nested metabarcoding approach

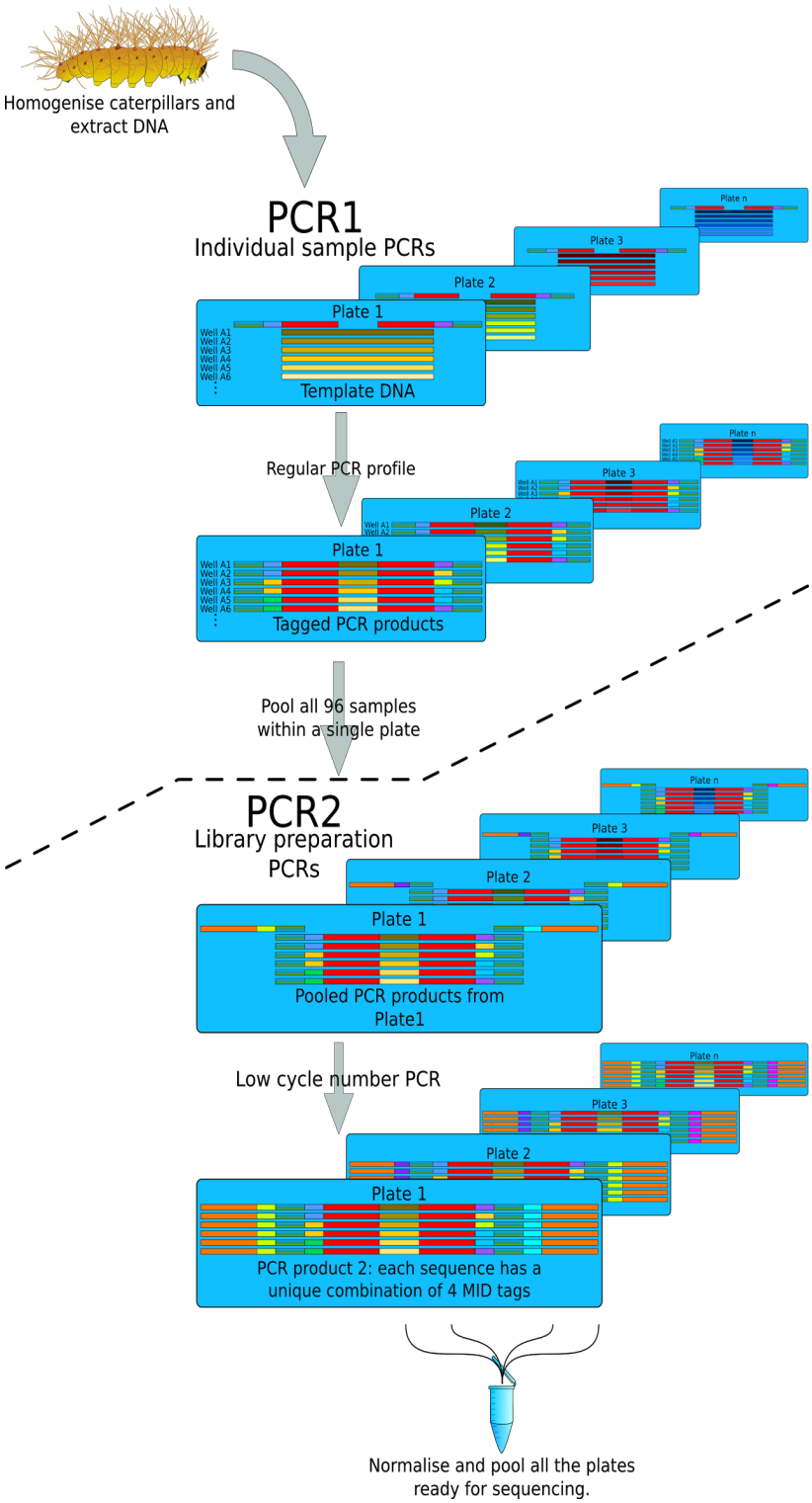


Fig2. Preparation of PCR amplicon libraries for Illumina MiSeq using a nested metabarcoding approach. Colour choices have the same meaning as in Fig1 shades of colours represent the same target sequences in different individuals.

103 *Sampling and laboratory protocols.*

104 For this study 919 OPM caterpillars and pupae were extracted from 25 nests collected in Richmond
105 Park, London, UK in July 2014 (full collection data is available in Table S1). Caterpillars were placed
106 in deep well plates and individually perforated using autoclaved toothpicks. Caterpillars were then
107 digested overnight at 37^{o}C in 670 µl of digestion buffer (20mM EDTA, 120mM NaCl, 50mM Tris and
108 1% SDS) with 30 µl of 10 mg/ml Proteinase K solution. Ten microliters of the digestion supernatant
109 was then used as the starting material for a 70 µl HotSHOT DNA extraction (Truett et al., 2000) which
110 was then diluted 1/100 for PCR amplification.

111 A 310 bp fragment of the Cytochrome C Oxidase subunit I barcode region (*coxI)* was
112 amplified using primers modified from LeRay (2013) to include standard Illumina MIDs and bridge
113 sequences (see Fig1 and Table S2). PCRs were carried out over 45 cycles (95^{o}C for 15s, 51^{o}C for
114 15s and 72^{o}C for 30s) in 20 µl reactions using MyFi Mix (Bioline), 1 µl of template DNA and each
115 primer (final concentration - 0.5 µM). Extra cycles were required as long primers are known to cause a
116 lag in PCR amplification (Schnell, Bohmann & Gilbert, 2015). PCRs were checked on a gel to gauge
117 success rates and 10 µl of each product from a plate was pooled together (without quantification) to
118 produce each pre-library, resulting in ten separate pre-libraries. Two aliquots of each pre-library were
119 gel purified to remove excess primers using QIAquick gel extraction kit (Qiagen). Purified pre-libraries
120 were quantified using a nanodrop ND-1000 (Thermo Scientific) and pooled ready for the library
121 preparation PCR and Illumina MiSeq V3 (2 x 300bp) sequencing (Macrogen, South Korea). Each
122 plate contained 92 OPM samples, two negative samples (18MΩ H_2O), and two positive samples. The
123 first positive contained a mixture of extracted template DNA from *Astatotilapia calliptera* (a cichlid
124 fish), *Comaster audax* (a crinoid) and *Triops cancriformis* (a tadpole shrimp) and was amplified at the
125 same time as the OPM samples (hereafter denoted DNA positive). The positive samples were chosen
126 due to their low probability of occurring in UK oak trees. All samples were sequenced (including
127 positives and negatives) even when no band was present as PCR products may still exist below gel
128 detectable levels. The second positive contained a mixture of PCR products from each of these
129 species that had been independently amplified using primers with the correct combination of tags and
130 combined before being added directly to the pre-library during pooling (hereafter denoted PCR
131 positive). The PCR positive was quantified using a nanodrop ND-1000 and a volume was calculated
132 that meant we were adding 1/95th the total DNA of each pre-library as PCR positive.

133

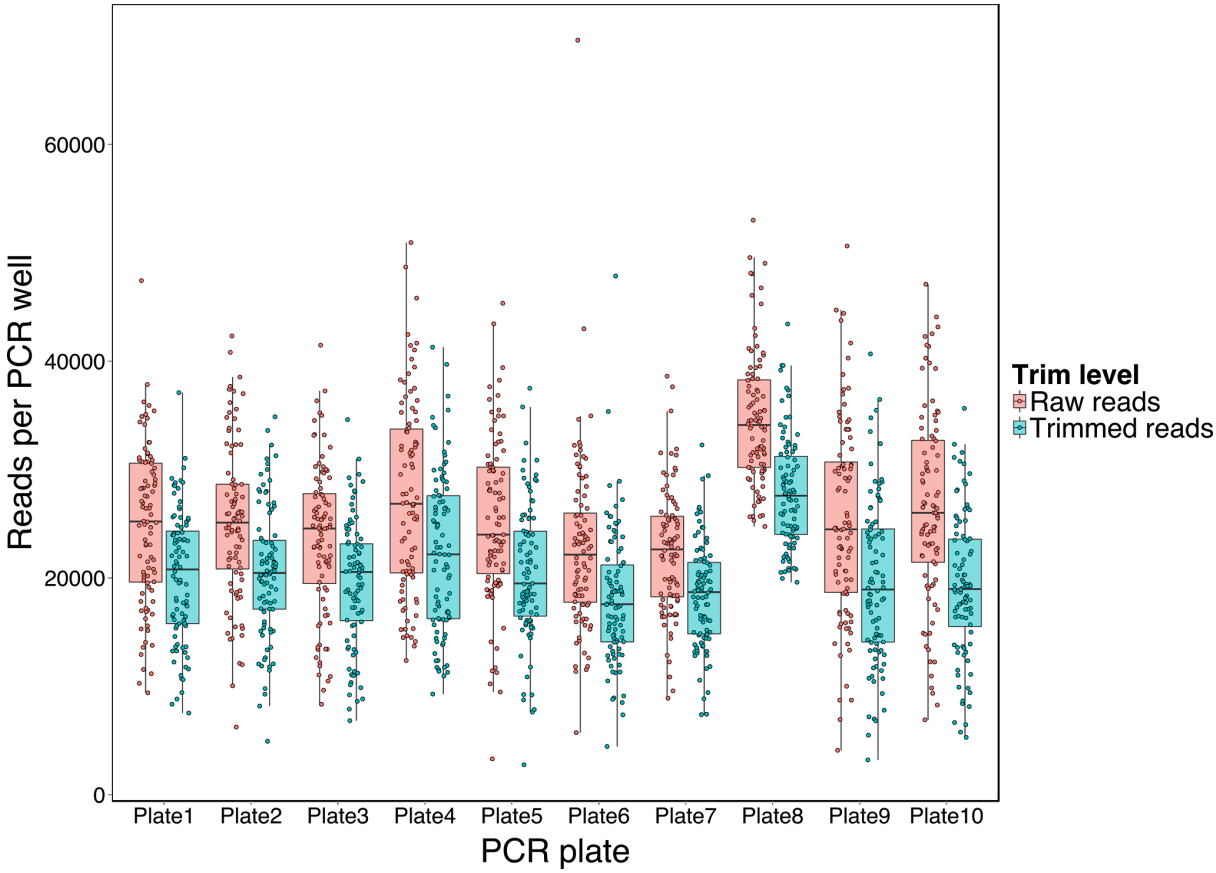
134 *Bioinformatic processing of Illumina MiSeq output.*

135 Processing of Illumina data from raw sequences to taxonomic assignment was performed using a
136 custom pipeline for reproducible analysis of metabarcoding data metaBEAT v0.8
137 (https://github.com/HullUni-bioinformatics/metaBEAT).

138 Individual steps performed as part of the pipeline are as follows: In brief, reads were demultiplexed
139 using the *process_shortreads* script from the Stacks software suite (Catchen et al. 2013).
140 Trimmomatic 0.32 (Bolger et al. 2014) was subsequently used for quality trimming and PCR-primer
141 clipping of the raw reads in two steps: (1) reads were end-trimmed to phred Q30 using a sliding
142 window approach (5bp window size) and (2) PCR-primers were clipped off the reverse complemented
143 sequences. Reads shorter than 100bp after quality trimming/primer clipping were discarded. Paired-
144 end sequences were subsequently merged (minimum overlap 10bp) using FLASH 1.2.11 (Magoc &
145 Salzberg 2011). Successfully merged reads were length filtered to retain only amplicons of the
146 expected length (313bp +- 10%). The remaining high-quality sequences were clustered using vsearch
147 v.1.1 (https://github.com/torognes/vsearch) across a range of clustering similarity thresholds (0.9-1.0).
148 Clustering results were further filtered based on the number of reads assigned to each cluster
149 (minimum cluster coverage) in order to minimize of cross-contamination effects between wells. Single
150 representative sequences from each cluster were subjected to a BLAST search (Zhang et al. 2000)
151 against a custom reference database, which was compiled from all available *coxI* sequences for
152 *Thaumetopoea* spp. as well as representative sequences for the positive controls (*A. calliptera, C.*
153 *audax, T. cancriformis*), downloaded from Genbank. The database was further extended by Sanger
154 sequences produced using established protocols (Folmer et al., 1994) for arthropods retrieved during
155 OPM nest sampling, including the known specialist OPM parasitoid *Carcelia iliaca* (Genbank
156 accession KT345964) only recently discovered in the UK (Sands et al., 2015). The custom reference
157 database compiled for this study is available at Github **<repository to be confirmed>**. Taxonomic
158 assignment of clusters was performed using a lowest common ancestor (LCA) approach similar to the
159 strategy used by MEGAN (Huson et al. 2007), such that for each query we identify the taxa receiving
160 the top 10% (bit-score) BLAST hits and subsequently determine the lowest taxonomic level shared by
161 all taxa in the list.

Results

Overall we had an apparent PCR success rate of 99.5% (i.e. 99.5% of sample wells produced a visible band on a gel). In total we produced 25,313,722 sequences from a single MiSeq v3 Illumina run and retained 20,096,646 after quality trimming. For the 919 moth samples, read depth per well ranged from 3,312-69,624 reads before quality trimming (mean = 25,793, sd = 8,327) and from 2,764-47,856 reads after quality trimming (mean = 20,639, sd = 6,755) (Fig3). With the exception of plate 8, all PCR plates had similar read depths and variation per well. We suspect that an error in loading concentration for plate 8 is the cause of a generally higher read count compared to other plates but as our samples were sequenced using a commercial company, all library normalisation steps were performed away from our lab and we do not have the information to verify this. Figure 4 indicates that positive wells were overrepresented in our sequencing suggesting that our DNA quantification with the Nanodrop was not accurate enough to adequately normalise the positives relative to each pre-library.



177

178 Fig3. Boxplots of read depth per PCR well for each plate (positives and negatives excluded) with

179 actual read depth for each PCR well overlaid as scatter plots.

180

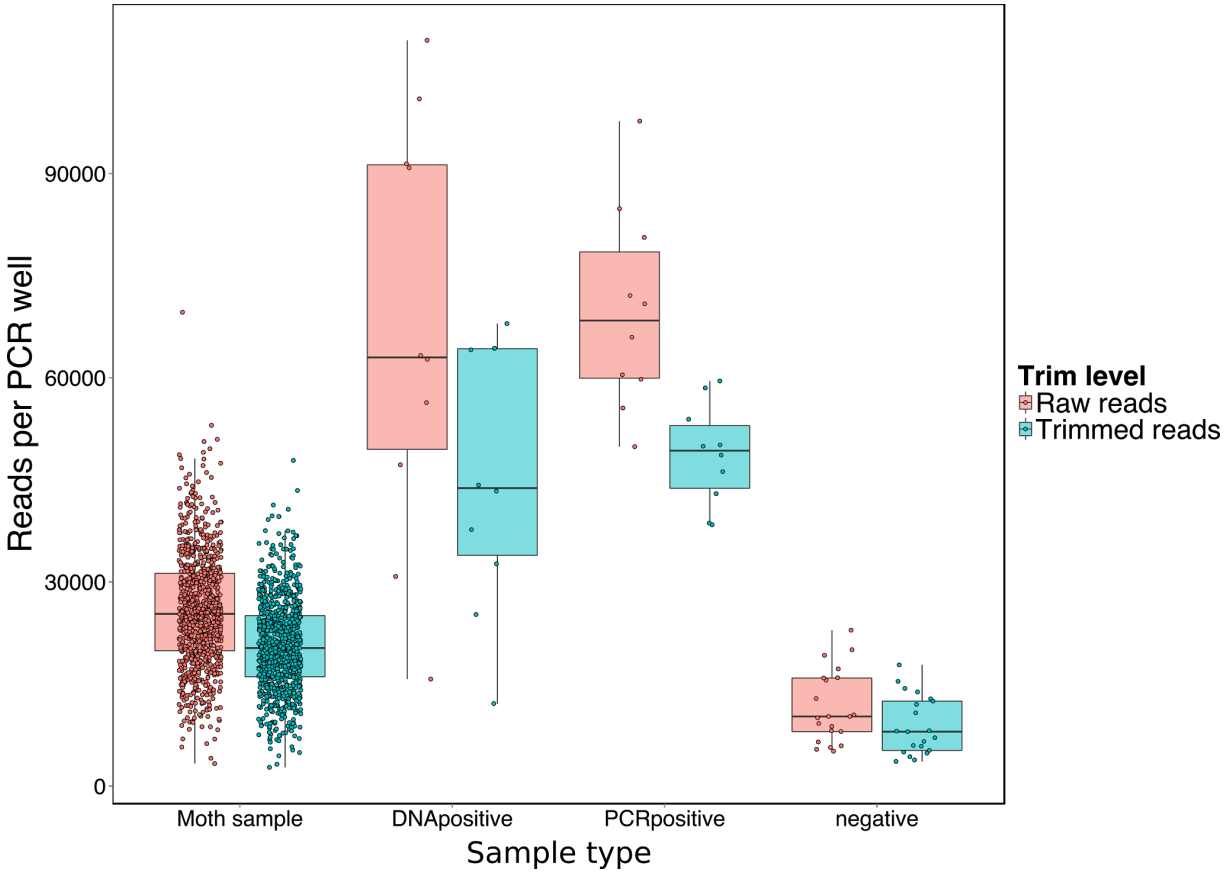


Fig4. Boxplots of read depth per PCR well for each type of PCR well with actual read depth for each PCR overlaid as scatter plots.

Analysis of clustering parameters revealed that the parameters chosen for sequence similarity and minimum cluster size have a strong effect on the number of putative OTUs defined in each well (Fig5). As we have single caterpillars in each PCR well we chose a set of stringent parameters that resulted in a relatively low number of OTUs identified per well. Clustering with a similarity of 95% and minimum cluster size of 200 reads results in a mean of 1.9 OTUs per well and a standard deviation of 0.9 OTUs per well.

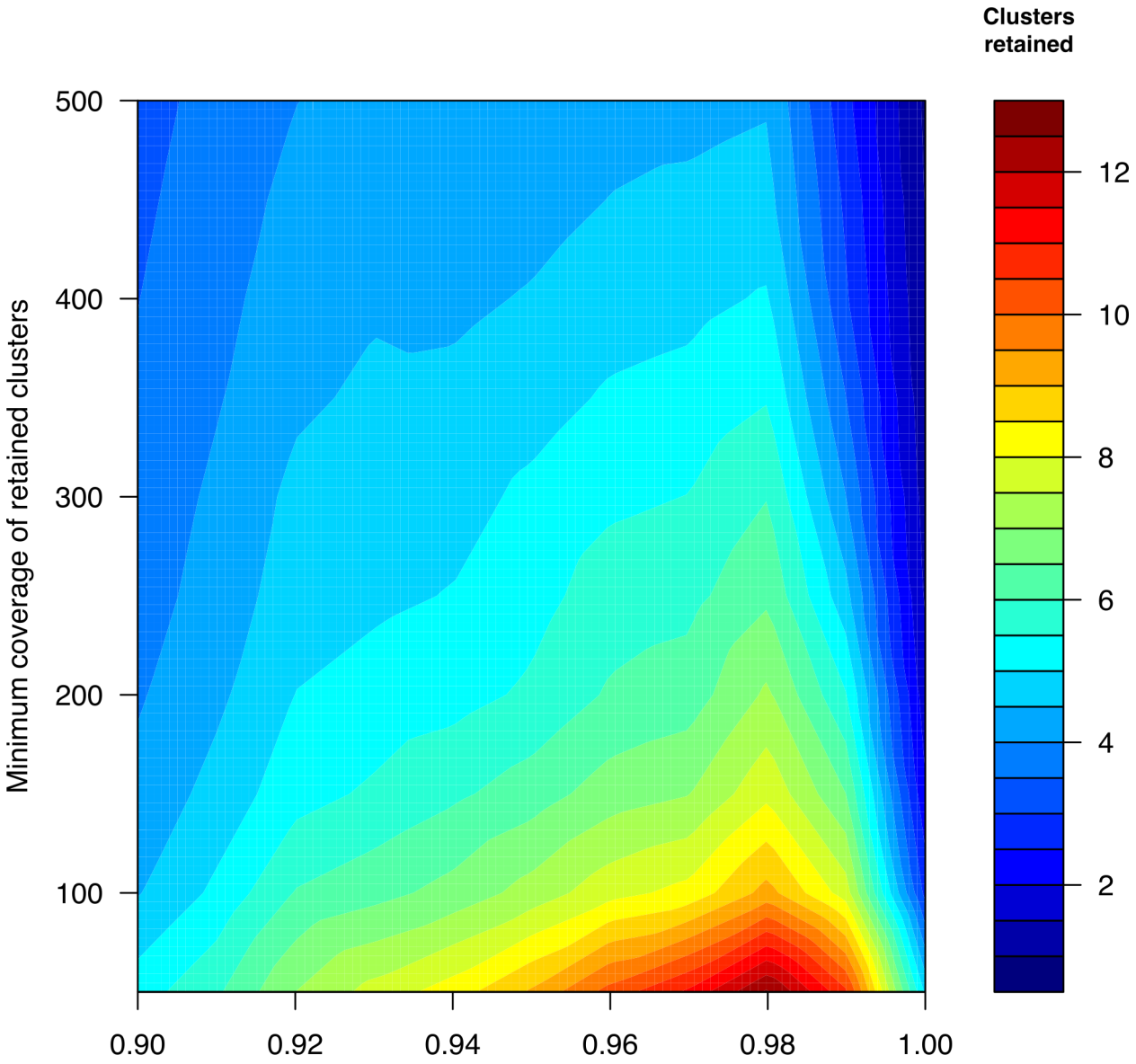


Fig5. Visualisation of parameter space for clustering stringency. The mean number of putative OTUs (clusters) retained per well after trimming and clustering is determined by both the sequence similarity threshold for clustering and the minimum cluster size retained. Understanding how these two values interact allows the researcher to determine sensible values for both before moving on to taxonomic assignment of OTUs.

Sequences were assigned by the taxonomy identification pipelines predominantly to either OPM or its known parasitoid fly *Carcelia iliaca* (Fig 6). Our data indicated that 29.8% of OPM caterpillars sampled from Richmond park, London, were parasitised by *C. iliaca*.

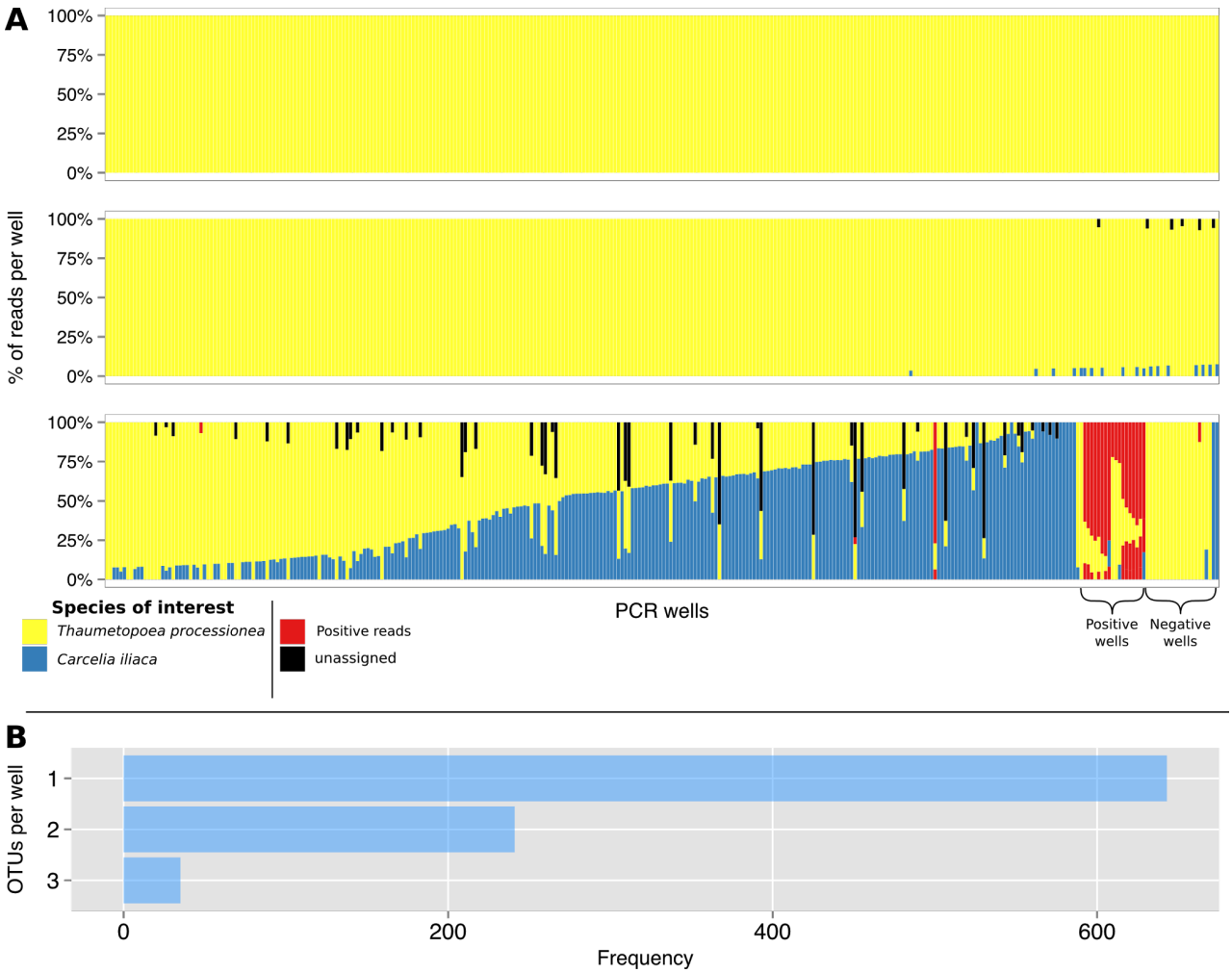


Fig6. (A) Percentages of reads and taxonomic identifications for 919 individuals. (B) the frequency of different numbers of OTUs per well.

Discussion

Nested metabarcoding

Here we tested the ability of a NGS nested metabarcoding design to produce individual level data for a large number of samples in a single sequencing run. We found approximately 26,000x coverage for each well before sequence filtering, allowing us to adopt a high stringency for sequence quality. The depth of coverage found in our experiment allows us to distinguish multiple unique sequences in each well, representing the host, parasitoids, and (potentially) any other species interacting with the moths. Thus, our approach leads more easily to a much more complete understanding of the ecological interactions than standard Sanger barcoding approaches (cf. Wirta et al., 2014; Derocles et al., 2015).

214 Although applied here to parasitised individuals, we anticipate that environmental or community
215 sequencing approaches could be applied in exactly the same way to a range of study systems.
216 *Thaumetopoea processionea* in our samples was parasitised with a single parasitoid species already
217 known from the literature (*Carcelia iliaca*: Tachinidae: Diptera -Sobczyk, 2014). Other parasitoids
218 reported for this moth were not included in our reference database and may be present at a low
219 frequency in the unassigned category. Their absence in our data set could also be due to our samples
220 being almost exclusively late instar caterpillars and many of the recorded parasitoids are known egg
221 and pupal parasitoids (Sobczyk, 2014). It is also possible that the arrival of a parasitoid exhibits a lag
222 period after that of its host (e.g. Stone et al., 2012) so much of the parasitoid community associated
223 with OPM in its native range may simply not be present in the UK. More thorough sampling of OPM
224 life stages and a more complete reference database would resolve this and is the topic of ongoing
225 work in our laboratory.

226 The rare presence of positive sequences in sample and negative wells as well as positives indicates
227 that contamination between wells can be a problem when preparing samples with a standard protocol.
228 However, there is still sufficient read depth in each well to control for this through stringent quality
229 filtering and produce accurate barcoding results from simple mixed templates. Unfortunately read
230 depths for sequences found in negative wells cannot be used as a reliable cutoff to filter background
231 contamination as any template DNA that contaminates a negative PCR well has no competition for
232 reagents (as compared to a sample well) and so will be disproportionately amplified. Future runs
233 using this method should use much more stringent lab protocols such as oil sealed reactions and the
234 use of individual lids rather than plastic films.

235 The relative costs of NGS and Sanger sequencing varies with scale of the experiment, with
236 commercial UK prices for Sanger reads approximately 1/300th the cost of an Illumina MiSeq at time of
237 writing. For small numbers of individuals and a single barcode locus Sanger may be much more cost
238 effective. As the quantity of data required increases however NGS has the potential to be
239 considerably cheaper, as the costs of a single NGS run are largely fixed, irrespective of how many
240 individuals are included. Although our experiment would not have even been feasible with Sanger due
241 to the sequence complexity of the sample, an experiment to barcode ~1000 individuals we estimate
242 that the costs are approximately 1/5th that of the equivalent Sanger experiment.

243 Ecological sciences are appreciating more than ever the power of incorporating ecological networks

244 rather than simple species lists into monitoring approaches. Such experiments, though producing a
245 new level of ecological information, require individual level rather than community data, and so have
246 previously been little assisted by NGS.

247 Conventional metabarcoding produces vast numbers of reads which are excellent for error correction
248 and detection of rare sequences but most modern statistical analyses for detecting subtle ecological
249 effects require more than presence/absence data. Some authors have attempted to relate read depth
250 to biomass or numbers of individuals both for PCR based metabarcoding (e.g. Elbrecht & Leese,
251 2015) and PCR free metabarcoding (e.g. Tang et al., 2015). Attempting to measure sample sizes or
252 biomass from read depth presents a number of challenges. Firstly PCR based approaches can be
253 biased by variation in amplification efficiency across different taxa (for example, variation in primer
254 binding affinities across different taxa or base composition variation affecting enzyme efficiency). PCR
255 free approaches to metabarcoding attempt to circumvent this by removing the PCR step and all the
256 associated biases completely (e.g. Tang et al., 2015). In theory, read depth should then correlate with
257 copy number for a given locus but in reality we have little knowledge of how sequenceable DNA
258 availability is affected by extraction method and more importantly, how read depth then correlates with
259 biomass or numbers of individuals. PCR free metabarcoding is further constrained as much of the
260 read depth which could be used for sequencing additional specimens is used for sequencing
261 additional areas of genome that are not necessary for identification. Our approach allows us to use
262 presence/absence data across a large number of individual specimens to produce quantitative data
263 analysable with standard statistical tests at the same time as reducing over sequencing of any single
264 individual.

265

266 *Future approaches employing nested metabarcoding*

267 Our approach has been demonstrated in a single context of determining parasitism. The nested
268 barcoding approach however has a diverse range of applications to which it could be applied. One of
269 the most immediate applications is the description of community data in the context of a single
270 individual. Here we show that several species are present within a single caterpillar, and other sets of
271 primers could more broadly sample the parasites and mutualists of individuals. This approach could
272 be used to determine a range of interspecies interactions (e.g. parasitism, predation, root fungal

communities) and build ecological networks. Networks present a much more complete measure of an ecosystem than the presence/absence data usually produced by metabarcoding studies and has great potential for ecosystem management and restoration (Evans et al., 2016- in final review). The nested barcoding approach is currently being used in our laboratory to survey the pollen community carried by individual pollinators allowing key functional ecological information to be gathered across a large sample size of individuals.

Recently, there has been a great interest in understanding how gut flora vary between individuals in different habitats, with different diets, or between related species (e.g. Sharon et al., 2010; Brucker & Bordenstein, 2013; Ceja-Navarro et al., 2015). It is likely that nested metabarcoding has a deep enough coverage per individual to allow characterisation of even rich bacterial communities. This approach could also be used to process bacterial communities in environmental or medical samples, soil mesofauna, bulk insect samples, or any other complex community while still keeping the number of individual samples high to help replication and detailed spatial or temporal sampling. Should the read number be insufficient for a given experiment the same samples could be loaded onto a sequencer with higher throughput (e.g. Illumina HiSeq rather than MiSeq) to address this issue as long as the paired-end nature of the sequences can be maintained.

Whole genome sequencing for phylogenomics produces a very rich dataset for systematics that has resolved many previously intractable problems. As resources are often limited, there can be a tradeoff however between the number of loci and the number of species sampled in a phylogenetic design. Nested metabarcoding approaches could perhaps be employed to multiplex 5-10 loci per individual, each with the same MID, and the scale that to 1000-2000 individuals in a single MiSeq run. Densely sampling taxonomic space in this way may prove valuable in some experiments.

This same approach would clearly be valuable for intraspecific studies too where 5-10 nuclear loci could represent complex population genetic data much better than single cytoplasmic loci. Nested approaches such as described here would also phase SNPs and indels within a locus which Sanger approaches do not. This is advantageous in many population genomics datasets.

Conclusions

Here we have demonstrated the utility of nested barcoding to an exemplar dataset in ecology and evolution, the characterisation of parasite-host community data in an invasive species. The hierarchical tagging approach in NGS we describe will allow a large diversity of advances in ecology and evolution, which will be of increasing importance as we attempt to quantify functional changes in ecological networks with climate change, intensifying agriculture, and species loss.

Reproducibility statement

To ensure reproducibility of all our analyses we have deposited Jupyter notebooks with all commands in Github **<repository to be confirmed>**. Raw sequence data has been submitted to the SRA with accession number **<will be added upon acceptance by NCBI>** and all trimmed and clustered versions of the data are also included in the manuscript repository along with primary scripts and any additional results files. The metaBEAT pipeline, and other analyses, were run in a Docker container (https://hub.docker.com/r/chrishah/metabeat/; v0.8 was used for the current study) in order to make our entire analysis environment available for replication if required.

Supp Material

1. Table S1: Collection locations for samples used as part of this study.
2. Table S2: Primers used as part of this study.

Acknowledgements

We would like to thank Amir Szitenberg for helpful discussion on bioinformatic pipelines and Gillian Jonusas (Royal Parks) for access and logistics in Richmond Park.

References

Binladen J., Gilbert MTP., Bollback JP., Panitz F., Bendixen C., Nielsen R., Willerslev E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PloS one* 2:e197.

Brucker RM., Bordenstein SR. 2013. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus Nasonia. *Science* 341:667–669.

Ceja-Navarro JA., Vega FE., Karaoz U., Hao Z., Jenkins S., Lim HC., Kosina P., Infante F., Northen TR., Brodie EL. 2015. Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. *Nature communications* 6:7618.

Daigle D., Simen BB., Pochart P. 2001. High-Throughput Sequencing of PCR Products Tagged with Universal Primers Using 454 Life Sciences Systems. In: *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.,.

Davey JW., Davey JL., Blaxter ML., Blaxter MW. 2010. RADSeq: next-generation population genetics. *Briefings in functional genomics* 9:416–423.

Derocles SAP., Evans DM., Nichols PC., Evans SA., Lunt DH. 2015. Determining plant - leaf miner - parasitoid interactions: a DNA barcoding approach. *PloS one* 10:e0117872.

Elbrecht V., Leese F. 2015. Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass--Sequence Relationships with an Innovative Metabarcoding Protocol. *PloS one* 10:e0130324.

Evans DM., Kitson JJN., Lunt DH., Straw NA., Pocock MJO. 2016. Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems.

Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* 3:294–299.

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487:94–98.

350 Illumina. 2011. Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq
351 System. *Illumina technical note*.
352 Leray M., Yang JY., Meyer CP., Mills SC., Agudelo N., Ranwez V., Boehm JT., Machida RJ.
353 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI
354 region for metabarcoding metazoan diversity: application for characterizing coral reef
355 fish gut contents. *Frontiers in zoology* 10:34.
356 Sands RJ., Jonusas G., Straw NA., Kitson JJN., Raper CM. 2015. *Carcelia Iliaca* (Diptera:
357 Tachinidae), a specific parasitoid of the Oak Processionary Moth (Lepidoptera:
358 Thaumetopoeidae), new to Great Britain. *British Journal of Entomology and Natural*
359 *History* 28.
360 Schlötterer C., Tobler R., Kofler R., Nolte V. 2014. Sequencing pools of individuals - mining
361 genome-wide polymorphism data without big funding. *Nature reviews. Genetics* 15:749–
362 763.
363 Schnell IB., Bohmann K., Gilbert MTP. 2015. Tag jumps illuminated - reducing sequence-to-
364 sample misidentifications in metabarcoding studies. *Molecular ecology resources*
365 15:1289–1303.
366 Sharon G., Segal D., Ringo JM., Hefetz A., Zilber-Rosenberg I., Rosenberg E. 2010.
367 Commensal bacteria play a role in mating preference of Drosophila melanogaster.
368 *Proceedings of the National Academy of Sciences of the United States of America*
369 107:20051–20056.
370 Sobczyk T. 2014. *Der Eichenprozessionsspinner in Deutschland*. Bundesamt für
371 Naturschutz.
372 Spurgin LG., van Oosterhout C., Illera JC., Bridgett S., Gharbi K., Emerson BC., Richardson
373 DS. 2011. Gene conversion rapidly generates major histocompatibility complex diversity
374 in recently founded bird populations. *Molecular ecology* 20:5213–5225.
375 Stone GN., Lohse K., Nicholls JA., Fuentes-Utrilla P., Sinclair F., Schönrogge K., Csóka G.,
376 Melika G., Nieves-Aldrey J-L., Pujade-Villar J., Tavakoli M., Askew RR., Hickerson MJ.
377 2012. Reconstructing Community Assembly in Time and Space Reveals Enemy Escape

in a Western Palearctic Insect Community. *Current biology: CB* 22:532–537.

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

Tang M., Hardman CJ., Ji Y., Meng G., Liu S., Tan M., Yang S., Moss ED., Wang J., Yang C., Bruce C., Nevard T., Potts SG., Zhou X., Yu DW. 2015. High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in ecology and evolution / British Ecological Society* 6:1034–1043.

Toju H., Guimarães PR., Olesen JM., Thompson JN. 2014. Assembly of complex plant-fungus networks. *Nature communications* 5:5273.

Truett GE., Heeger P., Mynatt RL., Truett AA., Walker JA., Warman ML. 2000. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *BioTechniques* 29:52, 54.

Wagner CE., Keller I., Wittwer S., Selz OM., Mwaiko S., Greuter L., Sivasundar A., Seehausen O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular ecology* 22:787–798.

Wirta HK., Hebert PDN., Kaartinen R., Prosser SW., Várkonyi G., Roslin T. 2014. Complementary molecular information changes our perception of food web structure. *Proceedings of the National Academy of Sciences of the United States of America* 111:1885–1890.

Yu DW., Ji Y., Emerson BC., Wang X., Ye C., Yang C., Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in ecology and evolution / British Ecological Society* 3:613–623.