1 NGS metabarcoding proves successful for quantitative assessment of
2 symbiont abundance: the case of feather mites on birds

3
4 J. Diaz-Real^{1,2}, D. Serrano³, A. Piriz⁴ and R. Jovani^{1*}
5

6 ¹Department of Evolutionary Ecology. Estación Biológica de Doñana (CSIC), Avda. Americo
7 Vesputio s/n, ES-41092 Sevilla, Spain.

8 ²Present address: Department of Ecology and Animal Biology. Universidade de Vigo.
9 Campus As Lagoas Marconsende, ES-36310 Vigo, Pontevedra, Spain.

10 ³Department of Conservation Biology. Estación Biológica de Doñana (CSIC), Avda. Americo
11 Vesputio s/n, ES-41092 Sevilla, Spain.

12 ⁴Molecular Ecology Lab. Estación Biológica de Doñana (CSIC), Avda. Americo Vesputio
13 s/n, ES-41092 Sevilla, Spain.

14
15 *Corresponding author:

16 Roger Jovani

17 Department of Evolutionary Ecology. Estación Biológica de Doñana (CSIC), Avda. Americo
18 Vesputio s/n, ES-41092 Sevilla, Spain.

19 E-mail: jovani@ebd.csic.es. Fax: (+34) 954 621 125.

Abstract

Understanding the ecological function of species and the structure of communities is crucial in the study of ecological interactions among species. For this purpose, not only the occurrence of particular species but also their abundance in ecological communities is required. However, abundance quantification of species through morphological characters is often difficult or time/money consuming when dealing with elusive or small taxa. Here we tested the use of next-generation sequencing (NGS) for abundance estimation of two species of feather mites (*Proctophyllodes stylifer* and *Pteronyssoides parinus*) under five different proportions (16:1, 16:4, 16:16, 16:64, and 16 vs 256 mites) against a mock community composed by *Proctophyllodes clavatus* and *Proctophyllodes sylviae*. In all mixtures, we retrieved sequence reads from all species. We found a strong linear relationship between 454 reads and the real proportion of individuals in the mixture for both focal species. The slope for *Pr. stylifer* was close to one (0.904), and the intercept close to zero (-0.007), thus showing an almost perfect correspondence between real and estimated proportions. The slope for *Pt. parinus* was 0.351 and the intercept 0.307, showing that while the estimated proportion increased linearly relative to real proportions of individuals in the samples, proportions were overestimated at low real proportions and underestimated at larger ones. Additionally, pyrosequencing replicates from each DNA extraction were highly repeatable ($R=0.920$ and 0.972 , respectively), showing that the quantification method is highly consistent given a DNA extract. Our study suggests that NGS is a promising tool for abundance estimation of feather mites' communities in birds.

Keywords

COI, high-throughput sequencing, mtDNA, NGS quantitative test, pyrosequencing.

Introduction

The study of ecological interactions among species requires the understanding of the ecological function of the organisms and the structure of communities. This demands knowing not only the occurrence of species in an area and the intensity of their interactions, but also the quantification of species' abundances (Huber et al. 2007). This is particularly challenging when studying elusive species, degraded material (e.g. diet samples) or when dealing with species difficult to identify by morphology. The latter scenario is common in the study of symbiotic interactions where differentiation of taxa of minute symbionts by morphological characters is often complicated. Species-level identification is the most challenging task (Carew et al. 2013), producing more errors than higher taxonomical level identification (Jones 2008), and commonly only some sex and/or age classes (e.g. adult males) allow morphological identification to the species level. This, jointly with the typical large numbers of symbionts (hundreds or thousands in a single host), makes the quantification of the abundance of each individual symbiont at the species level highly challenging when not simply impossible.

As a result, molecular techniques based on DNA sequencing have been developed as an alternative to morphological identification, because their potential to obtain fast, easy and accurate identification to the species level. Hebert *et al.* (2003) promoted the use of cytochrome oxidase I (COI) as a molecular barcode for identifying animal species, which have been shown to be successful for most invertebrates (with few exceptions, e.g. Meier et al. 2006; McFadden et al. 2011). Furthermore mtDNA is easy to amplify due to the high number of copies present in cells (Galtier et al. 2009).

Development of Next-Generation Sequencing (NGS) technologies has allowed the characterization of communities through the assessment of richness indices of species through DNA barcodes. This is commonly known as “metabarcoding” which refers to identification of multiple species appearing in a bulk sample or degraded material (Taberlet et al. 2012). Several studies have shown the utility of NGS for diversity assessment in: microbial flora of humans (Ahn et al. 2011; Marzorati et al. 2013), and of comestibles (Nam et al. 2012), diet studies (Soininen et al. 2009; Deagle

et al. 2009) and natural environmental samples (Yu et al. 2012; Carew et al. 2013; Ji et al. 2013, Gómez-Rodríguez et al. 2015). It is generally assumed that the number of reads obtained from NGS for a given species or taxon positively correlates with its relative abundance in the analysed sample. However, several recent approaches in diverse taxa reveal a more complex scenario. While some controlled tests showed promising results (Hajibabaei et al. 2011; Piloni et al. 2012; Shokralla et al. 2012; Carew et al. 2013; Weber and Pawlowski 2013), other studies concluded that the number of reads of a given taxon was not correlated with its relative abundance (Porazinska et al. 2009; Amend et al. 2010; Egge et al. 2013). This has been proposed to be the consequence of multiple biological and methodological issues (Deagle et al. 2014) such as biases during DNA extraction and PCR amplification (Carew et al. 2013; Kermarrec et al. 2013), assignation of reads to Multiplex Identifiers (Berry et al. 2011), different number of copies of the sequenced gene in different unicellular (Weber and Pawlowski 2013) or multicellular species (e.g. species of different body size, Gómez-Rodríguez et al. 2015), and differential primer affinity for DNA template (Hajibabaei et al. 2011).

In this study we experimentally tested the utility of NGS for quantitative assessment of the relative abundance of feather mites (Acariformes: Astigmata: Analgoidea) that live on the surface of wing feathers (Gaud and Atyeo 1996). Although feather mites are the commonest ectosymbionts on birds, their ecology is seldom studied (Dabert and Mironov 1999; Proctor 2003). Their tiny size most < 0.5 mm), high number of individuals per host, absence of keys for most genera, low number of expert taxonomists and the fact that for many genera only adult males have clear species-specific morphology make NGS a promising tool for feather mite species identification and relative quantification. Here, we tested whether the number of reads in NGS corresponding to *Proctophyllodes styliifer* (Proctophyllodidae, Buckholz 1869) or *Pteronyssoides parinus* (Pteronyssidae, Koch 1940) correlates with the proportion of individuals of those species in a mock mixture of *Proctophyllodes clavatus* (Fritsch 1961) and *Proctophyllodes sylviae* (Gaud 1957). With this purpose we carried out a controlled test of ten mock communities with five different proportions of a close (same genus) and a distant (different genus) species (Table 1). To look for consistence in pyrosequencing estimates of proportions, we also made five different replicated amplifications of each of the ten DNA templates.

97 **Material and methods**

98 **Samples collection**

99 Feather mites were sampled from birds captured in two localities in South Spain: Roblehondo (La
 100 Iruela), in Sierras de Cazorla, Segura y las Villas Natural Park (Jaén), and Manecorro (El Rocío), in
 101 Doñana National Park (Huelva) during August and September 2013. Birds were captured using mist
 102 nests, kept individually in cloth bags, immediately banded, inspected for feather mites and released.
 103 Four species of feather mites belonging to two genera were collected from the flight feathers
 104 (primaries, secondaries and tertiaries) of four different bird host species: *Pr. stylifer* and *Pt. parinus*
 105 from blue tits *Cyanistes caeruleus*; *Pr. sylviae* from blackcaps *Sylvia atricapilla* and Sardinian
 106 warblers *S. melanocephala*; and *Pr. clavatus* from garden warblers *S. borin*. Given the high specificity
 107 of these feather mite species and the low number of mite species found in these birds, feather mite
 108 samples could be identified to the species level under the stereomicroscope and stored separately in
 109 98% ethanol until extraction. Later matching of sequences with our DNA barcoding library (Doña et
 110 al. 2015) confirmed our initial identifications under the stereomicroscope.

111 **Experimental design**

112 Sets of 16 individuals of *Pr. stylifer* or 16 individuals of *Pt. parinus* were confronted with a varying
 113 number of individuals from the mix *Pr. sylviae* / *Pr. clavatus* (“pool” hereafter; Table 1). Only male
 114 adult mites were used in this study. In this way, we were confronting the two focal species with
 115 different proportions of congeneric (*Proctophyllodes*) or non-congeneric species (*Proctophyllodes* vs
 116 *Pteronyssoides*). Once DNA was extracted from each of the ten combinations shown in Table 1, five
 117 replicates of each sample were amplified by PCR and pyrosequenced (see below). Thus, we could
 118 estimate the correlation between the proportion of individuals introduced in the vials with the
 119 proportion of the sequences retrieved, and also the robustness of the pyrosequencing procedure.

120 **DNA extraction**

Genomic DNA was extracted using HotSHOT (Truett et al. 2000). Total volume of extraction was variable, depending on the number of feather mites being extracted in the same experimental sample. Feather mites in experimental samples with fewer individuals (17-32 mites) were extracted in a final volume of 50 µl and the rest in 100 µl (samples with 80-272 mites). After extraction, exoskeletons were separated from the extraction volume and stored in 98% ethanol.

DNA amplification and sequencing

To build the library of specific amplicons, we used the universal tailed amplification design (Guidelines from amplicon experimental design, ROCHE November 2012). It is based on two rounds of PCR; the first round amplifies the target sequences and the second PCR attaches the multiplex identifiers (MID) and the 454 adaptors to each amplicon. This allowed to barcode and prepare the amplicons for pyrosequencing. In the first PCR, a segment of approximately 650-bp of the COI region was amplified with degenerated primers bcdF05 and bcdR04 (Dabert et al. 2008). PCRs were carried out in 20 µl reaction volumes containing 1x (NH₄)₂SO₄ reaction buffer (Bioline), 2.5 mM MgCl₂, BSA 1x, 0.25 mM dNTPs, 1 µM of each primer, 1.25 U BIOTAQ™ (Bioline) and 5 µl of DNA template. The PCR followed a touchdown PCR profile: initial denaturation for 3 min at 95°C; 20 cycles of denaturing for 1 min at 95°C, annealing for 30 sec starting at 55°C and decreasing by 0.5°C/cycle and elongation for 1 min at 72°C; followed by 20 cycles of denaturing for 1 min at 95°C, annealing for 30 sec at 45°C and elongation for 1 min at 72°C; then, a final step for 5 min at 72°C. Five PCR replicates of each extract were done, and thus we carried out 50 PCR reactions in total.

PCR products were visualized by electrophoresis in a 2% agarose gel. Visible bands of the COI fragment were cleaned with homemade SPRI beads (Rohland and Reich 2012) and amplified again in the second round. This second PCR was carried out in 20 µl reaction volumes containing 2 µl of 1x (NH₄)₂SO₄ reaction buffer (Bioline), 2.5 mM MgCl₂, 0.25 mM dNTPs, 1 µM of each primer, 1.25 U BIOTAQ™ (Bioline) and 1 µl of PCR product from first PCR, following the profile: initial denaturation for 2 min at 94° followed by 35 cycles of denaturing for 30 sec at 92 °C, annealing for 30 sec at 56°C and elongation for 45 sec at 72°C; then, a final step for 5 min at 72°C.

The amplification products were electrophoresed in a 2.5% agarose gel and quantitatively assessed using the Quantity One software (Bio-Rad) that estimates band intensities. All PCR products were pooled into equimolar concentrations, purified using SPRI beads and quantified fluorimetrically with Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). The pool was amplified in an emulsion PCR using the GS Junior Titanium emPCR kit (Lib A) with the emPCR amplification protocols for long fragments, and to increase the number of reads we followed the one-way reads amplicon sequencing design (ROCHE Technical Bulletin: Amplicon Sequencing with Various emPCR Amplification Conditions, February 2011; Guidelines from Amplicon Experimental Design, ROCHE November 2012). The pool was pyrosequenced on a 454 GS Junior System using the Sequencing Method Manual GS Junior Titanium Series following manufacturer's instructions (Roche) at Estación Biológica de Doñana (EBD-CSIC) in Seville, Spain.

Bioinformatics analysis

Only high quality fragments of reads between 530 and 850 bp length were used for this experiment. Reads were separated by MID and the Novo Assemble tool of Geneious (Biomatters) was used for collapsing reads within each MID with maximum 3% of mismatches per read. All contigs generated were aligned by MUSCLE with default settings and with known sequences of the species analyzed (Genebank Accession Numbers: *Pr. stylifer*: KP193704 – KP193716; *Pt. parinus*: KP193755 - KP193757; *Pr. clavatus*: KP193516 - KP193523; *Pr. sylviae*: KP193717 - KP193725, Doña et al. 2015). The alignments were visually inspected to identify and count individually the contigs corresponding to each species. In this step, chimeras were identified and omitted. The alignment with all the final sequences used in this experiment is available in Dryad (doi:XXXXXX).

Statistical analyses

Due to the lack of independence of data belonging to the same extraction process, we used Linear Mixed Models (LMMs) to obtain the intercept and slope of the relationship between the proportion of individuals estimated by NGS (dependent variable) and the real proportion of individuals in each extraction for *Pr. stylifer* and *Pt. parinus* (independent variable). The extraction sample was set as the

random effect, and the real proportion of individuals of the focal species in each experimental sample was set as fixed effect.

Results

A total of 84,015 reads were obtained. 49,217 were fragments between 530 and 850 bp of length (corresponding to the COI region). After splitting sequences by MID we obtained a total number of 48,627 reads. 584 reads were not assigned to a known MID or linker, and 6 reads were assigned to a non-present MID. After purging all chimeras and low quality sequences we obtained 36,752 reads corresponding to all 50 MIDs. Median coverage read per MID was 763 (range 18 – 2,545). Only nine of the 50 MIDs had coverage lower than 500 reads per MID of which only four had fewer than 100 reads.

In all 50 mock mixtures, sequences from the focal species (either *Pr. styliifer* or *Pt. parinus*) and from species of the pool were retrieved, even when the species pool was composed of only one or four individuals. Likewise, focal species were always detected even when occurring in the lowest proportions (16 vs 256). There was, however, a likely technical problem with the extraction of DNA from the combination of 16 *Pt. parinus* vs one individual from the pool (Fig. 1A). In any case, the problem occurred before pyrosequencing as shown by the consistent result from the five pyrosequencing replicates of this extraction, leading to a strong underestimation of the focal species. We excluded these five samples from further analyses (but see Discussion).

Results were strikingly repeatable between replicates from same DNA extractions for both focal species ($R_{Pt.parinus}=0.920$, $p<0.001$; $R_{Pr.styliifer}=0.972$, $p<0.001$; Fig. 1). We found a positive linear correlation between the proportion of pyrosequencing reads and the proportion of individuals in the mixture for both focal feather mite species (*Pr.styliifer*: $F=418.03$, $p<0.0001$; *Pt.parinus*: $F=195.71$, $p<0.0001$). The slope for *Pr. styliifer* was close to one (0.904) and the intercept close to zero (-0.007). The estimated proportion of *Pt. parinus* increased at increasing real proportions of individuals in the samples, with a flatter slope of 0.351 and a higher intercept of 0.307.

199

200 **Discussion**

201 Our results show that NGS provides a promising tool for quantifying the relative abundance of feather
202 mites from birds. First, we found high consistency among the proportions of individuals of species
203 estimated by NGS from the same DNA extract. This is highly promising because it means that NGS
204 introduced little noise into the data. However, we encountered a potential extraction failure in one of
205 the samples (rightmost points in Fig. 1A), suggesting that the DNA extraction step could bias the
206 abundance estimation by NGS (Kermarrec et al. 2013).

207 Second, for the two focal species studied here, *Pr. stylifer* and *Pt. parinus*, NGS was very
208 informative in relative terms. That is, the higher the real proportion of a mite species, the higher was
209 the estimated proportion by pyrosequencing. However, the slope of the relationship differed between
210 species. This tells us that, without a previous pilot study, it is not possible to be sure of calculating
211 with precision the real abundance of each species in a sample from the relative frequencies obtained
212 through NGS. Moreover, note that the behaviour of pyrosequencing could potentially change
213 depending on the species composition of the sample (as shown in Fig. 1). The relationship between
214 real and estimated proportions of *Pt. parinus* (Fig. 1) is difficult to explain: when the proportion of *Pt.*
215 *parinus* was above 50% it was underestimated by pyrosequencing, but it was overestimated below
216 50%. In other words, the most abundant taxa (either *Pt. parinus* or *Pr. clavatus* / *Pr. sylviae*) was the
217 one being underestimated. Whatever the cause, it seems clear that only in highly controlled scenarios
218 with a known low number of species and conducting pilot studies as the one reported here it could be
219 achieved a reliable estimate of real absolute abundances in a sample. However, pyrosequencing has
220 proved to be a promising tool to compare the abundance of species between samples.

221 Third, pyrosequencing allowed us to detect individuals present in very low proportions (<6%),
222 supporting previous findings for other groups of organisms (Pochon et al. 2013; Zhan et al. 2013).
223 This is crucial for the detection of elusive species of feather mites expected to be in scarce proportions.

Recent literature performing similar studies in other organisms from bacteria to invertebrates have found contradictory results on the power of NGS to calculate the relative abundance of different taxa from a sample (see Introduction). A recent study also questioned the COI-based metabarcoding application, mainly because the primer-binding region is not always highly conserved (Deagle et al. 2014). However, as suggested by (Deagle et al. 2014), and as we have found here, COI is an adequate option as a metabarcoding marker in studies where the number of taxa is small and a reference database is available.

Overall, although our results are promising, we recommend detailed pilot studies like the one reported here before deciding to apply this method in community studies (see also (Amend et al. 2010)), especially when a high accuracy of estimation of the relative abundance of species is needed. However, our study also shows that even when the relationship between the number of reads from NGS and the relative proportion of species is not perfect, the estimated abundance may be corrected by using regression parameters from pilot studies. Overall, we show that at least for analgoid feather mites, metabarcoding through NGS of COI is a highly promising tool to study the evolutionary ecology of the intimate association between birds and feather mites.

Acknowledgements

Funding was provided by the Spanish Ministries of Science and Innovation (Ramón y Cajal research contract RYC-2009-03967 to RJ, research project CGL2011-24466 to RJ). We would like to thank the lab technicians of the Molecular Ecology Laboratory of Estación Biológica de Doñana for support during lab-work. We also thank Sergey V. Mironov for his collaboration on our ongoing feather mites barcoding library project; Mirka Dabert for advice on the extraction protocol; José Luis Arroyo for the support during field work; and Jorge Doña, Iñigo Salvanueva, Miguel Alcaide, Elena Marmesat and two anonymous reviewers for generous advices.

Compliance with Ethical Standards

Disclosure of potential conflicts of interest

This study was funded by the Spanish Ministries of Science and Innovation (Ramón y Cajal research contract RYC-2009-03967 to RJ, research project CGL2011-24466 to RJ).

The authors declare that they have no conflict of interest

Research involving human participants and/or animals

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. The study was conducted under the authorization of Junta de Andalucía (permission provided by Consejería de Medio Ambiente), Generalitat de Catalunya (permit number 43430497) and Espacio Natural de Doñana and Dirección General del Medio Natural (permit 2013/27). No endangered species were involved in this study. All birds were studied with non-invasive methods and released at sampling locality few minutes after capture. The protocols were approved by the appropriate ethics committees: the Dirección General del Medio Natural, and the Subcomité de Bioética (CSIC).

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

Ahn J, Yang L, Paster BJ, Ganly I, Morris L, Pei Z, Hayes RB (2011) Oral microbiome profiles: 16S rRNA pyrosequencing and microarray assay comparison. PLoS One 6:e22788. doi: 10.1371/journal.pone.0022788

Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with
pyrosequencing: does read abundance count? *Mol Ecol* 19:5555–5565. doi: 10.1111/j.1365-
294X.2010.04898.x

Berry D, Ben Mahfoudh K, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon
pyrosequencing bias amplification. *Appl Environ Microbiol* 77:7846–7849. doi:
10.1128/AEM.05220-11

Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA (2013) Environmental monitoring using next
generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Front Zool*
10:45. doi: 10.1186/1742-9994-10-45

Dabert J, Ehrnsberger R, Dabert M (2008) *Glaucalgae tytonis* sp. n. (Analgoidea, Xolalgidae) from the
barn owl *Tyto alba* (Strigiformes, Tytonidae): compiling morphology with DNA barcode data for
taxon descriptions in mites (Acari). *Zootaxa* 1719:41–52.

Dabert J, Mironov S V. (1999) Origin and evolution of feather mites (Astigmata). *Exp Appl Acarol*
23:437–454.

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the
cytochrome c oxidase subunit I marker: not a perfect match. *Biol Lett* 10:20140562.

Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing
prey DNA in faeces. *Mol Ecol* 18:2022–2038. doi: 10.1111/j.1365-294X.2009.04158.x

Doña J, Diaz-Real J, Mironov S, Bazaga P, Serrano D, Jovani R (2015) DNA barcoding and mini-
barcoding as a powerful tool for feather mite studies. *Mol Ecol Resour.* doi: 10.1111/1755-
0998.12384

Egge E, Bittner L, Andersen T, Audic S, de Vargas C, Edvardsen B (2013) 454 Pyrosequencing to
describe microbial eukaryotic community composition, diversity and relative abundance: a test
for marine haptophytes. *PLoS One* 8:e74371. doi: 10.1371/journal.pone.0074371

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular
diversity: a reappraisal. *Mol Ecol* 18:4541–4550. doi: 10.1111/j.1365-294X.2009.04380.x

Gaud J, Atyeo WT (1996) Feather mites of the world (Acarina, Astigmata). The supraespecific taxa.
Part I. *Ann Mus R l’Afrique Cent Sci Zool* 277:3–193.

Gómez-Rodríguez C, Crampton-Platt A, Timmermans MJTN, Baselga A, Vogler AP (2015)
Validating the power of mitochondrial metagenomics for community ecology and phylogenetics
of complex assemblages. *Meth Ecol Evol.* doi: 10.1111/2041-210X.12376

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-
generation sequencing approach for biomonitoring applications using river benthos. *PLoS One*
6:e17497. doi: 10.1371/journal.pone.0017497

308 Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA
309 barcodes. *Proc Biol Sci* 270:313–321. doi: 10.1098/rspb.2002.2218

310 Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML (2007)
311 Microbial population structures in the deep marine biosphere. *Science* 318:97–100. doi:
312 10.1126/science.1146689

313 Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, Kitching R, Dolman PM, Woodcock
314 P, Edwards FA, et al (2013) Reliable, verifiable and efficient monitoring of biodiversity via
315 metabarcoding. *Ecol Lett* 16:1245–57. doi: 10.1111/ele.12162

316 Jones FC (2008) Taxonomic sufficiency: The influence of taxonomic resolution on freshwater
317 bioassessments using benthic macroinvertebrates. *Environ Rev* 16:45–69. doi: 10.1139/A07-010

318 Kermarrec L, Franc A, Rimet F, Chaumeil P, Humbert JF, Bouchez A (2013) Next-generation
319 sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater
320 diatoms. *Mol Ecol Resour* 13:607–619. doi: 10.1111/1755-0998.12105

321 Marzorati M, Maignien L, Verhelst A, Luta G, Sinnott R, Kerckhof FM, Boon N, Van de Wiele T,
322 Possemiers S (2013) Barcoded pyrosequencing analysis of the microbial community in a
323 simulator of the human gastrointestinal tract showed a colon region-specific microbiota
324 modulation for two plant-derived polysaccharide blends. *Antonie Van Leeuwenhoek* 103:409–
325 420. doi: 10.1007/s10482-012-9821-0

326 McFadden CS, Benayahu Y, Pante E, Thoma JN, Nevarez PA, France SC (2011) Limitations of
327 mitochondrial gene barcoding in Octocorallia. *Mol Ecol Resour* 11:19–31. doi: 10.1111/j.1755-
328 0998.2010.02875.x

329 Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of
330 high intraspecific variability and low identification success. *Syst Biol* 55:715–728. doi:
331 10.1080/10635150600969864

332 Nam Y-D, Park S, Lim S-I (2012) Microbial composition of the Korean traditional food “kochujang”
333 analyzed by a massive sequencing technique. *J Food Sci* 77:M250–M256. doi: 10.1111/j.1750-
334 3841.2012.02656.x

335 Piloni G, Granitsiotis MS, Engel M, Lueders T (2012) Testing the limits of 454 pyrotag sequencing:
336 reproducibility, quantitative assessment and comparison to T-RFLP fingerprinting of aquifer
337 microbes. *PLoS One* 7:e40467. doi: 10.1371/journal.pone.0040467

338 Pochon X, Bott NJ, Smith KF, Wood SA (2013) Evaluating detection limits of next-generation
339 sequencing for the surveillance and monitoring of international marine pests. *PLoS One*
340 8:e73935. doi: 10.1371/journal.pone.0073935

341 Porazinska DL, Giblin-Davis RM, Faller L, Farmerie W, Kanzaki N, Morris K, Powers TO, Tucker
342 AE, Sung W, Thomas WK (2009) Evaluating high-throughput sequencing as a method for

343 metagenomic analysis of nematode diversity. *Mol Ecol Resour* 9:1439–1450. doi:
344 10.1111/j.1755-0998.2009.02611.x

345 Proctor HC (2003) Feather mites (Acari: Astigmata): ecology, behavior, and evolution. *Annu Rev*
346 *Entomol* 48:185–209. doi: 10.1146/annurev.ento.48.091801.112725

347 Rohland N, Reich D (2012) Cost-effective , high-throughput DNA sequencing libraries for
348 multiplexed target capture. *Genome Res* 22:939–946. doi: 10.1101/gr.128124.111.1

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

351 Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebo JH,
352 Ims R a, Yoccoz NG, Taberlet P (2009) Analysing diet of small herbivores: the efficiency of
353 DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition
354 of complex plant mixtures. *Front Zool* 6:16. doi: 10.1186/1742-9994-6-16

355 Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation
356 biodiversity assessment using DNA metabarcoding. *Mol Ecol* 21:2045–50. doi: 10.1111/j.1365-
357 294X.2012.05470.x

358 Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, Warman ML (2000) Preparation of PCR-
359 Quality Mouse Genomic DNA with Hot Sodium Hydroxide and Tris (HotSHOT). *Biotechniques*
360 29:52–54.

361 Weber AA-T, Pawlowski J (2013) Can abundance of protists be inferred from sequence data: a case
362 study of foraminifera. *PLoS One* 8:e56739. doi: 10.1371/journal.pone.0056739

363 Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: metabarcoding
364 of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol Evol* 3:613–
365 623. doi: 10.1111/j.2041-210X.2012.00198.x

366 Zhan A, Hulák M, Sylvester F, Huang X, Adebayo AA., Abbott CL, Adamowicz SJ, Heath DD,
367 Cristescu ME, MacIsaac HJ (2013) High sensitivity of 454 pyrosequencing for detection of rare
368 species in aquatic communities. *Methods Ecol Evol* 4:558–565. doi: 10.1111/2041-210X.12037

369

370

371 **Table 1.** Experimental design showing the number of individual mites in each experimental sample.

372

Experimental sample	Feather mite species		
	<i>Proctophyllodes</i>	<i>Proctophyllodes</i>	<i>Pteronyssoides</i>
	<i>sylviae/clavatus</i>	<i>stylifer</i>	<i>parinus</i>
1	1	0	16
2	4	0	16
3	16	0	16
4	64	0	16
5	256	0	16
6	1	16	0
7	4	16	0
8	16	16	0
9	64	16	0
10	256	16	0

Figure captions

Figure 1. Proportion of reads from NGS vs. real proportion of individuals for *Pteronyssoides parinus* and *Proctophyllodes stylifer* in each experimental sample. The dashed lines represent a line with zero intercept and unit slope, and the continuous lines correspond to the real linear tendency of the points. Crosses correspond to reads of an experimental sample with a problem during DNA extraction (see Results and Discussion).

A

B

381

382