1. Title: New tools for diet analyses: nanopore sequencing of metagenomic DNA from stomach contents to quantify diet in an invasive population of rats

2. Authors: William Pearman, Adam N. H. Smith, Georgia Breckell, James Dale, Nikki E. Freed, Olin K. Silander

3. Addresses of all authors: Institute of Natural and Mathematical Sciences, Massey University, Auckland 0745, New Zealand

4. Keywords: Nanopore sequencing, diet analysis, rat, invasive

5. Communicating authors: Olin K. Silander, Institute of Natural and Mathematical Sciences, Massey University, Auckland 0745, New Zealand, olinsilander@gmail.com, +64 9 213 6618; Nikki E. Freed, Institute of Natural and Mathematical Sciences, Massey University, Auckland 0745, New Zealand, freednikki@gmail.com, +64 9 213 6639

6. Running title: Quantifying rat diets by direct nanopore sequencing

Abstract

Introduced species of mammals in New Zealand have had catastrophic effects on populations of diverse native species. Quantifying the diets of these omnivorous and predatory species is critical for understanding which native species are most impacted, and to prioritize which mammal species and locations should be targeted with control programmes. A variety of methods have been applied to quantify diet components in animals, including visual inspection of gut contents (Daniel 1973; Pierce and Boyle 1991), stable isotope analysis (Major et al. 2007; Carreon-Martinez and Heath 2010), and time-lapse video (Brown and Brown 1997; Dunlap and Pawlik 1996). Increasingly, DNA-based metabarcoding methods are being used (King et al. 2008; Soininen et al. 2009). These metabarcoding methods require a PCR step using primers that bind to highly conserved genomic regions (e.g. mitochondrial COI) to amplify specific regions for sequencing. This step introduces significant bias, primarily due to the lack of a universal primer set (King et al. 2008). Here we show that direct metagenomic sequencing using the Oxford Nanopore Minion allows rapid quantification of rat diets. Using a sample of rats collected from within 100km of Auckland, NZ, we show that these rats consume a wide variety of plant, invertebrate, vertebrate, and fungal taxa, with substantial differences in diet content between locales. We then show that, based on diet content alone, it is possible to pinpoint the sampling location of an individual rat within tens of kilometres. We expect that the rapidly increasing accuracy and throughput of nanopore-based sequencing, as well as increases in the species diversity of genomic databases, will soon allow rapid and unbiased assessments of animal diets in field settings.

Introduction

The relatively recent introduction of mammalian species to New Zealand has had significant negative effects on many of the native animal populations, including insects (Gibbs 1998), reptiles (Towns, Daugherty, and Cree 2001), molluscs (Stringer et al. 2003), and birds (Diamond and Veitch 1981; Dowding and Murphy 2001). These mammalian predators include possums, rats, stoats, and hedgehogs, as well as domestic pets such as cats (Gillies and Clout 2003). Currently, an ambitious plan is being put into place that aims for the eradication of all mammalian predators by 2050 (http://www.doc.govt.nz/predator-free-2050; Russell et al. 2015). A useful step toward this goal would be to prioritise the management of predators and geographical areas in which native species are experiencing the highest levels of predation. To do so requires establishing the diet content of these predators.

Unbiased and sensitive assessment of diet content is extremely difficult to achieve due to the limited accuracy of available methods. Identification of prey items using visual examination of stomach contents is strongly affected by which items are most easily degraded (for example, soft-bodied species). Stable isotope analysis is imprecise, yielding only broad information on diet such as relative consumption of protein and plant matter, as well as information on whether prey items are terrestrial or marine in origin (Hobson 1987; Basha et al. 2016). Time-lapse video (Brown et al. 2008) requires identification of the specific prey item, often difficult or impossible for small prey items or in low-light conditions.

Perhaps the most widely applied current molecular method is DNA metabarcoding. This approach relies on the PCR amplification and DNA sequencing of conserved regions from nuclear, mitochondrial, and/or plastid genomes (King et al. 2008). With adequate primer selection, this method can detect a wide range of species, and does not require specific expertise necessary for other methods (for example identifying degraded prey items).

However, DNA metabarcoding is not free from bias. PCR primers must be specifically tailored to particular sets of taxa or species (Jarman et al. 2002). Although more "universal" PCR primer pairs have been developed (for example targeting all bilaterians or even all eukaryotes; (Jarman, Deagle, and Gales 2004), all primer sets exhibit bias towards certain taxa. Tedersoo et al. (Tedersoo et al. 2015) found five-fold differences in fungal operational taxonomic units (OTU) estimates when using different sets of fungal-specific PCR primer pairs. Leray et al. (Leray et al. 2013) found that published universal primer pairs (i.e. those that do not target specific taxa) were capable of amplifying only between 57% and 91% of tested metazoan species, with as few as 33% of species in some phyla being amplified at all (e.g. cnidarians). Deagle et al. argued that in general, COI regions are simply not sufficiently conserved, and thus should not be used for metabarcoding studies at all (Deagle et al. 2014). Finally, Pawluczyk et al. showed that different loci from the same species exhibit up to 2,000-fold differences in qPCR-estimated DNA quantity within samples (Pawluczyk et al. 2015). It has even been shown that the polymerase itself can bias diversity metrics when using metabarcoding methods (Pereira et al. 2018). For these reasons, a less biased molecular-based method is desirable.

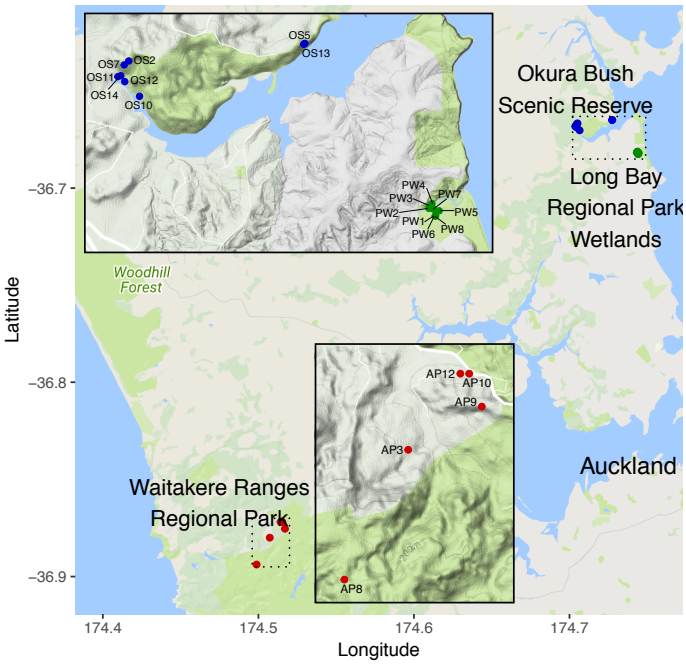
Metagenomic approaches, in which all DNA in the sample is directly sequenced, offer a promising alternative. Such methods have been used for diet analysis in only very few and recent instances (Srivathsan et al. 2015; Paula et al. 2015; Srivathsan et al. 2016). Although metagenomic approaches decrease the bias arising from PCR amplification of specific DNA regions, additional biases can arise, as the presence or absence of species and genera can only be inferred for those species or genera present in genomic databases. Although this is similarly true for metabarcoding approaches, metabarcode databases are considerably more comprehensive in terms of species representation as compared to genomic databases. Importantly, genomic sequence databases are rapidly increasing in species diversity, as are the methods to query these large databases (Wood and Salzberg 2014; Kim et al. 2016).

Here, we quantify rat diet composition using a metagenomic approach based on nanopore sequencing (Oxford Nanopore Technologies). The data suggest that rat diets around Auckland, New Zealand, comprise a wide range of plant, animal, and fungal families. We show using multivariate analyses that rat diets are distinguishable by habitat and location. This work establishes long-read metagenomic methods as a straightforward approach for diet quantification.

Materials and Methods

Study Areas

We trapped rats from three locations near Auckland, New Zealand. Each location comprised a different type of habitat: undisturbed inland native forest (Waitakere Regional Parklands, WP); native bush surrounding an estuary (Okura Bush Walkway, OB); and restored coastal wetland (Long Bay Regional Park, LB) (**Fig. 1**). Traps in OB and LB were baited with peanut butter, apple, and cinnamon wax pellets; or bacon fat and flax pellets. Traps in WP were baited with chicken eggs, rabbit meat, or cinnamon scented poison pellets. From 16 November to 16 December 2016, traps were surveyed by established volunteer conservation groups at each site every 48 hours. A total of 36 rats were collected from these locations. The majority of rats collected (34/36) were determined to be male by visual inspection, and identified as *Rattus norvegicus*. These 34 rats were selected for further analysis.



111 **Fig. 1. Location of rat sampling sites** in the greater Auckland area in the North Island
112 of New Zealand. Each point indicates a trap where one rat was captured, with the colour
113 of the points indicating the three broad locations: the native estuarine bush habitat of
114 Okura Bush (OB), the restored wetland of Long Bay (LB), and the native forest of
115 Waitakere Park (WP). The two insets show the three locations in higher resolution with
116 topographical details. Green indicates park areas. Precise geographical coordinates
117 were only available for five out of eight rats in WP.

118

119 DNA Isolation

120 Within 48 hours of trapping, rats were stored at either -20°C or -80°C until dissection.
121 We removed intact stomachs from each animal, rinsed them with 70% ethanol, and
122 removed the contents. After snap freezing in liquid nitrogen, we homogenised the
123 stomach contents using a sterile mini blender to ensure sampling was representative of
124 the entire stomach.

We purified DNA from 10-20 mg of homogenised stomach contents using the Promega Wizard Genomic DNA Purification Kit, with the following modifications to the Animal Tissue protocol: after protein precipitation, we transferred the supernatant to a new tube and centrifuged a second time to minimise protein carryover. The DNA pellet was washed twice with ethanol. These modifications were performed to improved DNA purity. We rehydrated precipitated DNA by incubating overnight in molecular biology grade water at 4°C and stored the DNA at -20°C. DNA quantity, purity, and quality was ascertained by nanodrop and agarose gel. The DNA samples were ranked according quantity and purity (based on A260/A280 and secondarily, A230/A280 ratios). The eight highest quality DNA samples from each location were selected for DNA sequencing.

DNA Sequencing

Sequencing was performed on two different dates (24 January 2017 and 17 March 2017) using a MinION Mk1B device and R9.4 chemistry. For each sequencing run, DNA from each rat was barcoded using the 1D Native Barcoding Kit (Barcode expansion kit EXP-NBD103 with sequencing kit SQK-LSK108) following the manufacturer's instructions. Twelve samples were pooled and run on each flow cell, for a total of 24 individual rats. The flow cells had 1373 active pores (January) and 1439 active pores (March). Sequencing was performed using local base calling in MinKnow v1.3.25 (January) or MinKnow v1.5.5 (March), but both runs were re-basecalled after data collection using Albacore 2.2.7 with the following command: *read_fast5_basecaller.py -i ./fast5_input -t 48 -s basecalled_output -f FLO-MIN106 -k SQK-LSK108 --barcoding -r -o fastq,fast5 -q 0 --disable_filtering.*

Sequence classification

All sequences were BLASTed (blastn v2.6.0+) against a locally compiled database consisting of the combined NCBI other_genomic and nt databases (downloaded on 13th June 2018 from NCBI). Default blastn parameters were used (gapopen 5, gapextend 2), and only hits with an e-value of 0.01 or less were retained. Due to the predominance of

short indels present in nanopore sequence data, we used the datasets basecalled in January and March of 2017 to test whether changing these default penalties affected the results (gapopen 1, gapextend 1). We found that these adjusted parameters did not qualitatively change our results.

We assigned sequence reads to specific taxon levels using MEGAN6 (v.6.11.7 June 2018) (Huson et al. 2016). We only used reads with BLAST hits having an e-value of $1x10^{-20}$ or lower (corresponding to a bit score of 115 or higher) and an alignment length of 100 bp or more. To assign reads to taxon levels, we considered all hits having bit scores within 20% of the bit score of the best hit (MEGAN parameter Top Percent).

Statistical analyses

Multivariate analyses were done using the software PRIMER v7 (Clarke and Gorley 2015). The data used in the multivariate analyses were in the form of a sample- (i.e. individual rat) by-family matrix of read counts. All bacteria, rodent, and primate families were removed. The majority of rodent hits were to rat and mouse, resulting from the rats' own DNA (see below). The majority of the primate hits were to human sequences, which probably resulted from sample contamination.

The read counts were converted to proportions per individual rat, by dividing by the total count for each rat, to account for the fact that the number of reads varied substantially among rats (Clarke et al. 2006). The proportions were then square-root transformed so that subsequent analyses were informed by the full range of taxa, rather than just the most abundant families (Clarke and Green 1988). We then calculated a matrix of Bray-Curtis dissimilarities, which quantified the difference in the gut DNA of each pair of rats based on the square-root transformed proportions of read counts across families (Clarke et al. 2006).

We used unconstrained ordination (specifically, non-metric multidimensional scaling (nMDS)) applied to the dissimilarity matrix to examine the overall patterns in the diet composition among rats. To assess the degree to which the diet compositions of rats

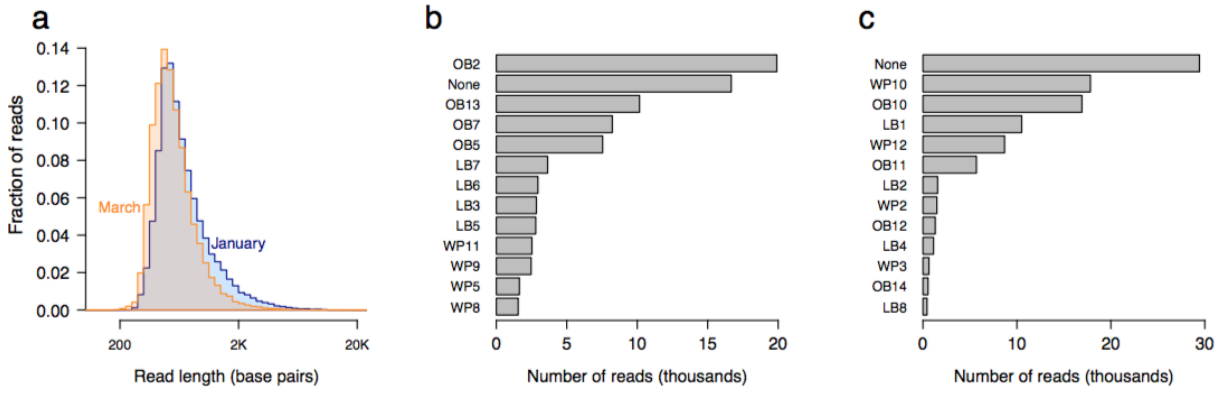
were distinguishable among the three locations, we applied canonical analysis of principal coordinates (CAP) (Anderson and Willis 2003) to the dissimilarity matrix. CAP is a constrained ordination which aims to find axes through multivariate data that best separates *a priori* groups of samples (in this case, the groups are the locations from which the rats were sampled); it is akin to linear discriminant analysis but CAP can be used with any resemblance matrix. The out-of-sample classification success was evaluated using a leave-one-out cross-validation procedure (Anderson and Willis 2003).

We used Similarity Percentage (SIMPER) (Clarke 1993)) to characterise taxa that distinguished between the locations. This allowed us to identify the families with the greatest percentage contributions to (1) the Bray-Curtis similarities of diets within each location and (2) the Bray-Curtis dissimilarities between each pair of locations.

Results

DNA sequencing and assignment of reads to taxa

After DNA isolation and sequencing, we obtained a total of 82,977 reads from the January sequencing run and 96,150 reads from the March run. Median read lengths were 606 bp and 527 bp for the January and March datasets, respectively (**Fig. 2A**). These lengths are considerably shorter than other nanopore sequencing results from both our and others work (Jain et al. 2016). This is most likely due to degradation of the DNA during digestion in the stomach as well as fragmentation during DNA isolation (Deagle, Eveson, and Jarman 2006) and sequencing library preparation. The median quality scores per read ranged from 7-12 (0.80 - 0.94 accuracy) for both runs (**Fig. S1**). The number of reads per barcoded rat sample varied by 10-fold for January and up to 40-fold in March(**Fig. 2B** and **2C**). This is primarily due to the highly variable quality of DNA in each sample. However, read length and quality were similar for all samples (**Fig. S1**).



204

205 **Fig. 2. Results of nanopore metagenomic sequencing of rat stomach contents. (a)**
206 **Read length distribution for January and March nanopore runs.** Read lengths
207 varied between 300 and 3,000 bp. **(b) and (c) Barcode distributions for January and**
208 **March runs, respectively.** We multiplexed the samples on the flow cells, using 12
209 barcodes per flow cell. The distribution of read numbers across barcodes was quite
210 uneven, varying by up to 40-fold in some cases. 20% (January) and 30% (March) of all
211 reads could not be assigned to a barcode ("None"). The inability to assign these reads
212 to a barcode is due primarily to their lower quality.

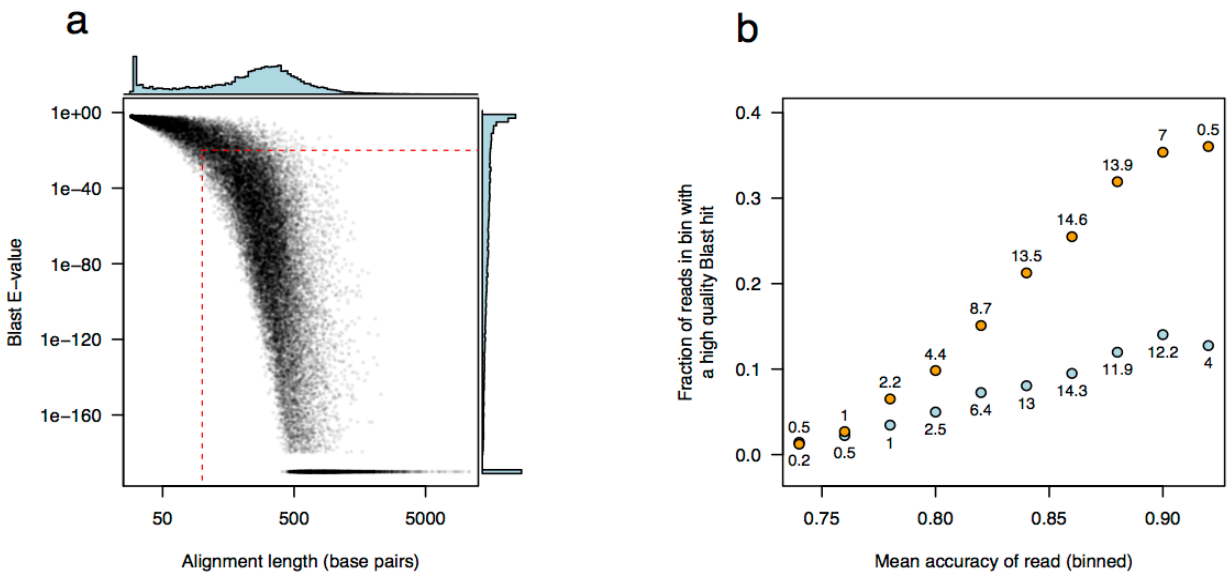
213

214 To quantify diet contents we first BLASTed all sequences against a combined database
215 of the NCBI nt database (the partially non-redundant nucleotide sequences from all
216 traditional divisions of GenBank excluding genome survey sequence, EST, high-
217 throughput genome, and whole genome shotgun
218 (ftp://ftp.ncbi.nlm.nih.gov/blast/db/README)) and the NCBI other_genomic database
219 (RefSeq chromosome records for non-human organisms
220 (ftp://ftp.ncbi.nlm.nih.gov/blast/db/README)). We used BLAST as it is generally viewed
221 as the gold standard method in metagenomic analyses (McIntyre et al. 2017). Of the
222 133,022 barcoded reads, 30,535 (23%) hit a sequence in the combined nt and
223 other_genomic database at an e-value cut-off of 0.01.

As an initial assessment of the quality of these hits, we examined the alignment lengths and e-values. We found a bimodal distribution of alignment lengths and a highly skewed distribution of e-values (**Fig. 3A**). We hypothesized that many of the short alignments with high e-values were false positives. We thus first filtered this hit set, only retaining BLAST hits with e-values less than 1e-20. Similar quality filters have been imposed previously (Srivathsan et al. 2015). A total of 22,331 hits passed this filter. Mean read quality had substantial effects on the rates of BLAST success, with almost 40% of high accuracy read having hits in the March dataset, as compared to 1% of low accuracy hits (**Fig. 3B**).

To specifically assign each sequence read to a taxon, we analysed the BLAST results in MEGAN6 (Huson et al. 2016). The algorithm employed in MEGAN6 assigns reads to a most recent common ancestor (MRCA) taxon level. For example, if a read has BLAST hits to five species, three of which have bit scores within 20% of the best hit, the read will be assigned to the genus, family, order, or higher taxon level that is the MRCA of those best-hit three species (Huson et al. 2007). If a read matches one species far better than to any other, by definition, the MRCA is that species.

Of the reads from the rat stomachs with BLAST hits, 31% were assigned by MEGAN as being bacterial. 55% of these were *Lactobacillus spp*. These results match previous studies on rat stomach microbiomes, which have found lactobacilli to be the dominant taxa (Maurice et al. 2015; Brownlee and Moss 1961; Li et al. 2017; Horáková, Zierdt, and Beaven 1971). Plant-associated *Pseudomonas* and *Lactococcus* taxa were also common, at 7% and 6%, respectively.



247 **Fig. 3. BLAST hits of metagenomic reads. (a) Biplot showing the e-value and**
248 **alignment length of the top BLAST hit for each read.** We observed a bimodal
249 distribution of alignment lengths and a skewed distribution of e-values. The y-axis is
250 plotted on a log scale, with zero e-values suppressed by adding a small number (1e-
251 190) to each e-value. The horizontal red dotted line indicates the e-value cut-off we
252 implemented and the vertical red dotted line indicates the length cut-off (e-value < 1e-20
253 and alignment length of 100, respectively) to decrease false positive hits. **(b) The**
254 **fraction of reads with high quality BLAST hits (e-value < 1e-20) increases as a**
255 **function of read accuracy.** We binned the data according to mean read accuracy (bin
256 width of 0.02) and calculated the fraction of reads within each bin that had a high quality
257 BLAST hit for the January and March runs separately (blue and orange points,
258 respectively). The number of reads in each bin is indicated above or below each point
259 (in thousands). There is a clear positive correlation between mean accuracy and the
260 likelihood of a high-quality BLAST hit, reaching almost 40% for very high quality reads
261 (accuracy greater than 92.5%).

262

MEGAN assigned reads to a wide range of eukaryotic taxa. To conservatively infer taxon presence, we first reclassified MEGAN species-level assignments to the level of genus. However, after this, many clear false positive assignments remained (e.g. hippo and naked mole rat). These matches were generally short and of low identity. To reduce such false positive taxon inferences, we used information from reads assigned to the genera *Rattus* (rat) and *Mus* (mouse). We inferred that the reads assigned to *Rattus* (2,696 reads in total) were true positive genus-level assignments and that the reads assigned to *Mus* (2,798 reads in total) were false positive genus-level assignments (and not true positive *Mus*-derived reads). Although rats are known to prey on mice (Bridgman et al. 2013), if this had occurred, we would expect that (1) the ratio of mouse to rat reads would be higher in the subset of rats that had predated mice; (2) in those same rats, the percent identity of the reads assigned to *Mus* would be higher than in rats that had not predated mice. However, we found that the ratio of mouse to rat reads was similar for all rats, and there was no evidence of higher percent identities for *Mus* reads from rats that had higher ratios.

Notably, the mean percent identity values of the best BLAST hits for these two groups of reads differed substantially, with *Rattus* reads having a median identity of 86.4%, and *Mus* 81.0% (**Fig. 4A**). The mean percent identity for *Rattus* reads corresponds very well to that expected given the mean quality scores of the reads (assuming the true sequence of the read is 100% identical to *Rattus*, 86.4% identity corresponds to a mean quality score of 8.7).

There was also a clear difference in the alignment lengths: the median ratio of alignment length to read length was 0.57 for *Rattus* and 0.52 for *Mus* (**Fig. 4B**). Importantly, the majority of diet items have percent identities that overlap with the *Rattus* reads, and alignment length to read length ratios that exceed the *Rattus* reads, suggesting that many diet taxa assignments are correct down to the level of genus. However, to minimise the rates of false positive taxon assignments of diet items, we implemented cut-offs based on the characteristics of the *Mus*- and *Rattus*-assigned reads. For genus-level assignment, we required at least 82.5% identity and an

alignment length to read length ratio of at least 0.55. These cut-offs exclude 88% of the reads falsely assigned to *Mus*, instead assigning them correctly to one taxon level higher, the Family *Muridae*.

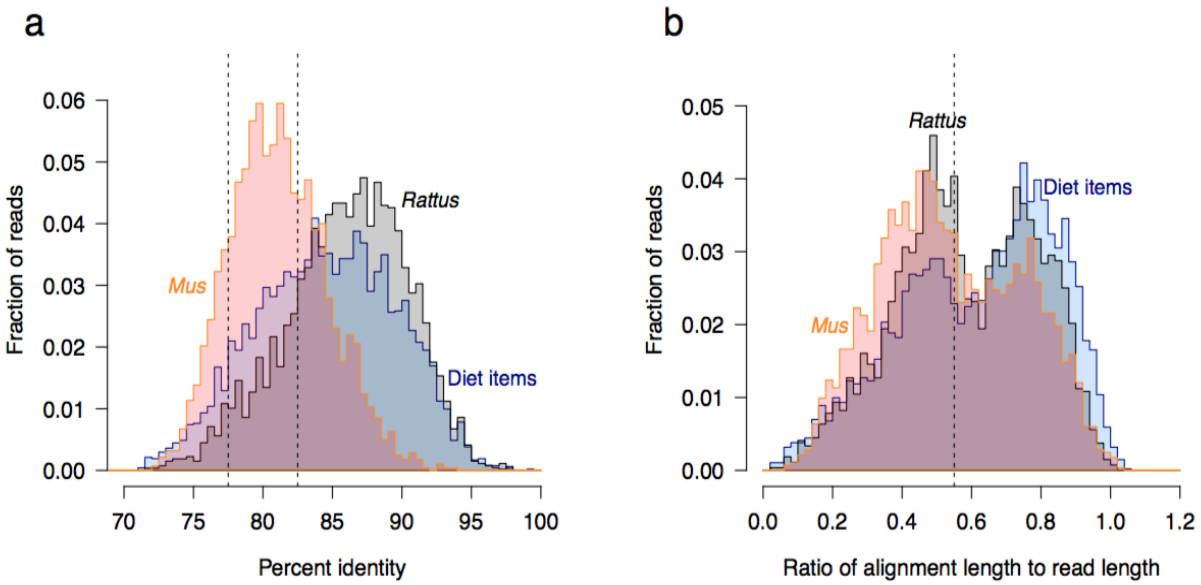


Fig. 4. Distributions of percent identity and length for alignments of reads matching *Rattus* (rat), *Mus* (mouse), and diet items. (a) Percent identity for alignments of rat (*Rattus*) and diet items is much higher than for mouse (*Mus*). Histograms are shown for the percent identity of the alignment of the top BLAST hit with the read. *Mus* matches show a clear shift to the left (lower percent identity) as compared to *Rattus* and diet items. Although different genera, *Mus* and *Rattus* are in the same family (*Muridae*). The dotted lines indicate the cut-offs that we implemented for inferring reads as belonging to a specific genus (above 82.5% identity) or family (above 77.5% identity). **(b) Ratios of alignment lengths to read lengths of rat (*Rattus*) and diet items are higher than for mouse (*Mus*).** This plot is analogous to that in (a). The dotted line indicates the cut-off that we implemented for inferring reads as belonging to a specific genus (above 0.55).

For family-level assignments, we required 77.5% identity, an alignment length to read length ratio of at least 0.1, and a total alignment length of at least 150 bp. Using higher cut-offs for the ratio of alignment length to read length excluded a large number of likely true positive taxa for which only short mtDNA or rDNA database sequences were present in the databases. For all other read-to-taxon assignments, we placed the read at the level of Order, or used the taxon level assigned by MEGAN. Using these cut-offs, 16% of all reads were classified at the Genus level; 71% were classified at the Family-level or below; 89% were classified at the Order-level or below; and 98% were classified at the Phylum-level or below.

Even though we imposed these relatively stringent filters, we observed that a small number of likely false positive taxa remained. Most were single reads with short alignments: *Poeciliidae* (177 bp); *Salmonidae* (172 bp); *Cyprinodontiformes* (140 bp and 177 bp); and *Octopodidae* (151 bp). The exception to this were three reads from two rats matching *Buthidae* (scorpions), which had alignment lengths of 762 bp, 664 bp, and 298 bp. It is unlikely these are true positives, and instead that these rats predated harvestmen, a closely related sister taxon within *Arachnida*. Despite the presence of these false positive taxa, to ensure we could identify the majority of taxa at high resolution (i.e. family-level), we did not further increase the stringency of our filters.

Identification of stomach contents

Within each rat, a wide variety of plant, animal, and fungal families and orders were discernible, ranging from two to 25 orders per rat (mean 8.7) (**Fig. 5**). Plants were the primary diet item, with the largest fraction of rats consuming four predominant orders: *Poales* (grasses), *Fabales* (legumes), *Arecales* (palms), and *Araucariales* (podocarps). The dominance of plant matter (fruits and seeds) in rat diets has been established previously (Sweetapple and Nugent 2007; Riofrío-Lazo and Páez-Rosas 2015). Animal taxa made up a smaller component of each rat's diet, with *Insecta* dominating: *Hymenoptera*, *Coleoptera*, *Lepidoptera* (moths and butterflies), *Blattodea* (cockroaches), *Diptera* (flies), and *Phasmatodea* (stick insects). In addition

Stylommatophora (slugs and snails) were present in substantial numbers (**Fig. 6A** and **6B**). Fungi were only a small component, although several families and orders were present: *Sclerotiniaceae* (plant pathogens), *Saccharomycetaceae* (budding yeasts), *Mucorales* (pin molds), *Russulales* (brittlegills and milk-caps), and *Chytotheriales* (black yeasts). Finally, for many rats, a substantial proportion of the stomach contents were parasitic worms (primarily *Spirurida* (nematodes) and *Hymenolepididae* (tapeworms)).

Close examination of this data suggested that specific families and orders were overrepresented in the diets of rats from particular locations. For example, six out of eight rats from the native estuarine bush habitat (OB) consumed *Arecaceae*, while only one in the restored wetland area (LB) did. All three rats that consumed *Phaseanidae* were from the native estuarine habitat (OB). All five rats that consumed *Solanales* were from the restored wetland area. These patterns suggested that it might be possible to use diet components alone to pinpoint the habitat from which each rat was sampled.

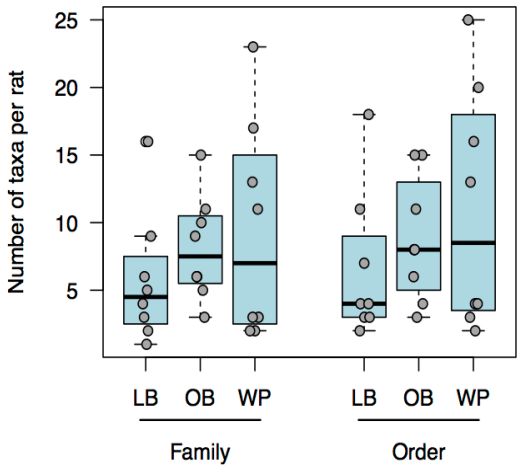


Fig. 5. Numbers of taxa in individual rats. Each boxplot indicates the range of families (left boxes) or orders (right boxes) consumed by each rat in each location (OB: Okura Bush; LB: Long Bay Park; WP: Waitakere Park). The numbers for individual rats (eight per location) are plotted in grey.

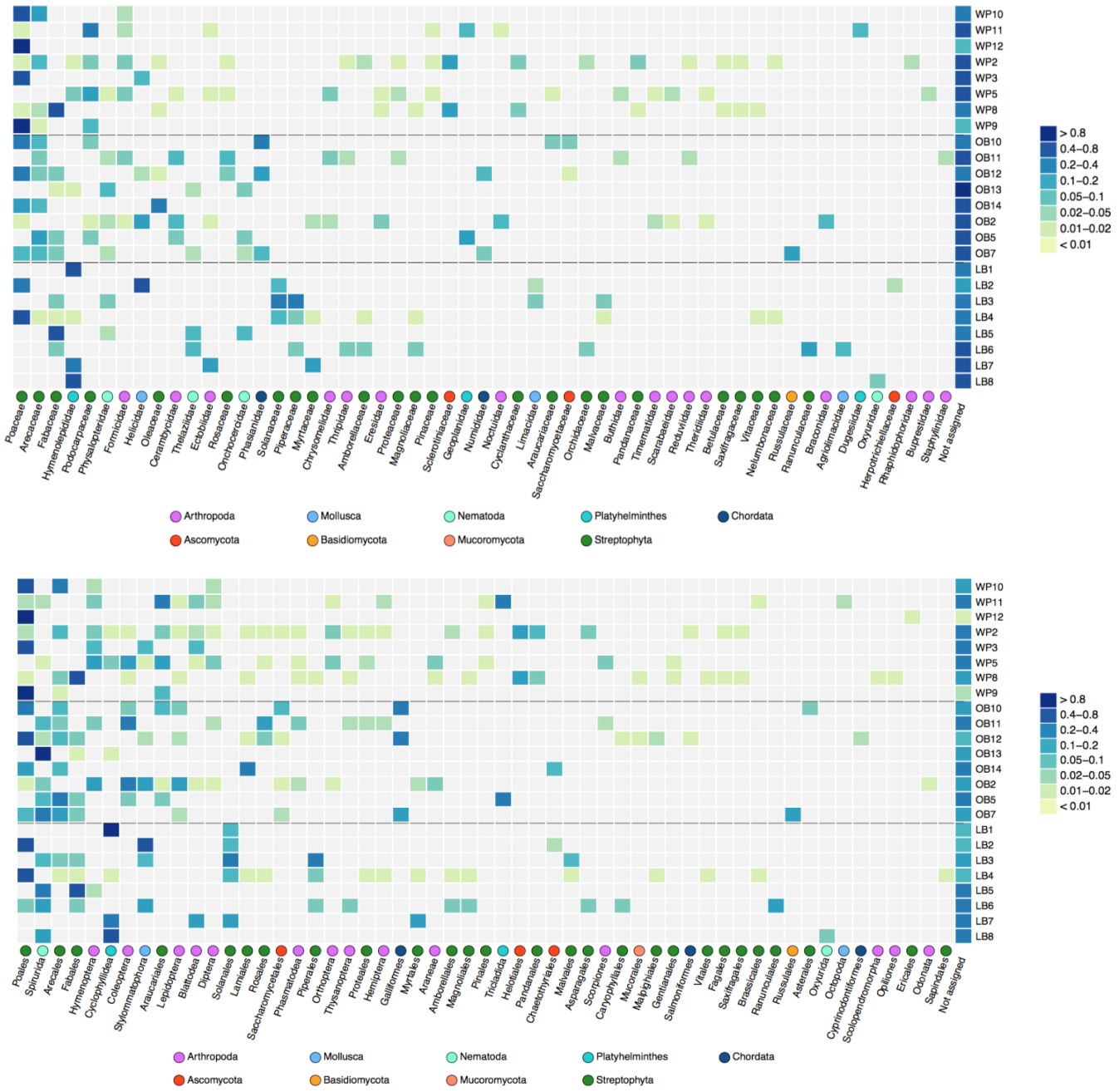


Fig. 6. Proportions of taxa in the diets of individual rats. Reads assigned to taxa at the family and order level are shown in the top and bottom panels, respectively. Each row corresponds to a single rat, with the proportions of reads for that rat assigned to each family or order indicated in shades of blue and yellow. Reads that were not assigned to a specific family or order are on the right end of the figure. The families and orders have been sorted so that the most common diet components appear on the left. Only the 55 most common families are shown. Note that the colour gradations presented on the scale are not linear.

363 nMDS and CAP analysis by location

364 In order to determine if diet composition of the rats differed consistently between
365 locations, we first performed an unconstrained analysis using nMDS on taxa assigned
366 at the family level. Using family rather than order or genus provides a balance between
367 how precisely we identify the taxon of diet item (genus, family, order), and whether we
368 assign a taxon at all. While family-level assignments are less precise than genus-level,
369 only 16% of all reads were classified at the genus level, while 71% were classified at the
370 family level.

371 The family-level unconstrained ordination (nMDS; **Fig. 7a**) showed no obvious grouping
372 of rats with respect to the locations, indicating that locations did not correspond to the
373 predominant axes of variation among the diets. However, a constrained ordination
374 analysis (CAP; **Fig. 7b**) identified axes of variation that distinguished the diets of rats
375 from different locations. The CAP axes correctly classified the locations of 19 out of 24
376 (79%) rats using a leave-one-out procedure.

377 The families that best characterised the three locations (i.e., had the greatest within-
378 location SIMPER scores) varied among locations. WP had average Bray-Curtis within-
379 location similarity of 13%; mostly attributable to *Hymenolepidae* (accounting for 51% of
380 the within-group similarity), *Solanaceae* (11%), and *Fabaceae* (11%). The average
381 similarity for OB was 21%, with the greatest contributing taxa being *Arecaceae* (33%),
382 *Poaceae* (23%), *Fabaceae* (9%), and *Phasianidae* (8%). The average similarity for WP
383 was 24%, with the greatest contributing taxa being *Poaceae* (72%) (**Table S4**). The
384 families with the largest correlations with the first two principal coordinates and thus
385 responsible for the separation between groups) were *Arecaceae*, *Podocarpaceae*,
386 *Phasianidae*, and *Pinaceae* (**Fig. 7c**).

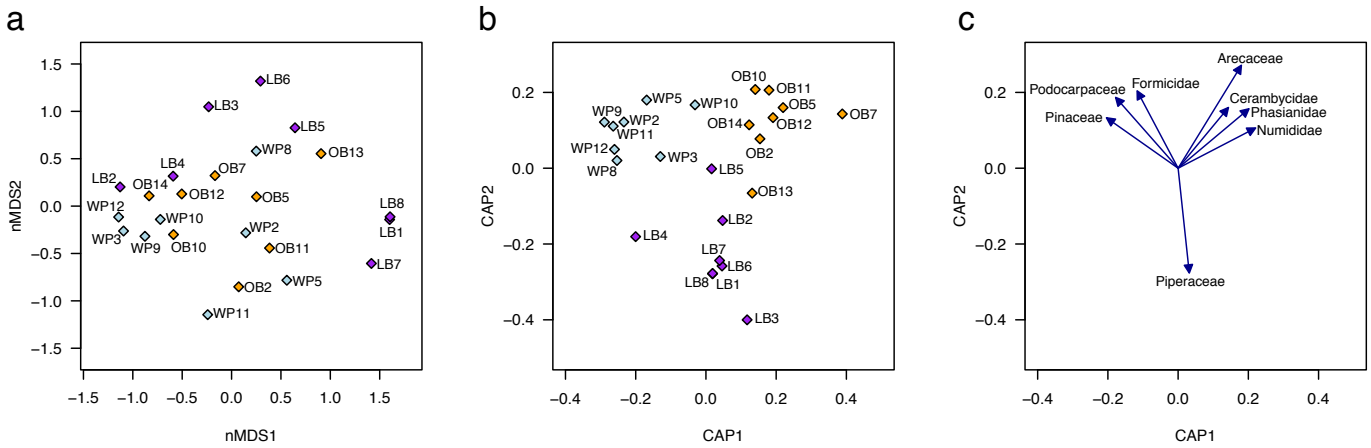


Fig. 7. Unconstrained nMDS (a) and constrained CAP (b) ordinations of the diets of rats from three locations. Both ordinations were based on Bray-Curtis dissimilarities of square root transformed proportions of reads attributed to each family. The locations were a native estuarine bush (OB, orange); a restored marine wetland (LB, purple); and a native forest (WP, light blue). The CAP ordination is repeated in panel **(c) with the rats omitted to show the Pearson correlations between families and the two cap axes.** The eight families with the strongest correlations are shown, indicating the taxa most strongly associated with each location.

Discussion

Accuracy and sensitivity

Here we have shown that using a simple long-read metagenomic DNA sequencing approach allows rapid classification of rat diet components. We expect that this technique for dietary analysis could be applied to a variety of animals and sample types, including samples that require less invasive collection (e.g. fecal matter). The sensitivity of this approach will likely improve as the accuracy and yield of Oxford Nanopore sequencing increases, and as the diversity of taxa in genomic sequence databases increases. Several aspects of the data support this.

First, we note that we did not find BLAST hits for the majority of reads. This is partially due the relatively low accuracy of the Oxford Nanopore sequencing platform at the time we preformed this analysis (approximately 87%). However, the fraction of reads yielding hits in the database increased substantially for higher quality reads, approaching 40% for very high-quality reads (**Fig. 3b)**. Other factors also likely reduced the numbers of BLAST hits, such as the paucity of genome sequence data for many taxa. This is convincingly illustrated by comparing across taxa the fraction of genomic hits to mitochondrial or rDNA sequence hits: although genome data exists for only a few taxa, mtDNA and rDNA sequence data are present in the database for the vast majority of animal and plant genera.

For animals with sequenced genomes in the database, we found that only a minority of reads mapped to mitochondrial or rDNA sequences (e.g. 32% of *Rattus* hits and 22% of *Anoplophora* hits). For plants with sequenced genomes, the fraction of mtDNA or rDNA-matching reads was even lower: between 4% (*Aegilops*) and 6% (*Triticum* and *Solanum*). For genera with little or no genomic sequence, the vast majority of matches were solely to mtDNA, rDNA, or microsatellite loci: 90% of *Phoenix* (date palm) hits; all *Helix* (snail) and *Rhaphidophora* **(**cave weta) hits. All *Artioposthia* (New Zealand flatworm) hits were to rDNA. These results indicate that for genera with no genomic sequence data, we have underestimated the actual number of sequences from that taxon by approximately three- to twenty-fold (for animals and plants, respectively).

As the species sampling of genomic databases increases (Lewin et al. 2018), the taxon-level precision of this method will improve. Given the current rate of genomic sequencing, with careful sampling, the vast majority of multicellular plant and animal families (and even genera) will likely have at least one type species with a sequenced genome within the next decade. Continued advancement in sequence database search algorithms as compared to current methods (Wood and Salzberg 2014; Kim et al. 2016; Nasko et al. 2018) should considerably decrease the computational workload necessary to find matching sequences.

To decrease biases in genomic databases, some previous studies have performed metagenomic classification using mitogenome data alone (Srivathsan et al. 2016; Paula et al. 2016). Using such methods, they found between 0.004% and 0.008% of all metagenomic reads matched mitogenomes from diet taxa. Limiting database searches to mitogenomes partially ameliorates biases in terms of taxon field in terms of taxon representation (i.e. most taxa will have similar levels of genomic representation in the databases). However, it considerably decreases diet resolution given that for some taxa, only a small percentage of sequence reads derive from the mitochondria as opposed to the nuclear genome.

It is important to note that our interest in diet also includes resolving relative biomass and relative numbers of each prey species, neither of which necessarily correlate well with the amount of DNA (either mitochondrial or nuclear) purified from a sample. Even a simple correction for the fraction of reads matching mitochondrial versus nuclear genomes is difficult, as different plant and animal tissues differ considerably in the relative amounts of mitochondrial versus nuclear DNA (e.g. leaf versus fruit).

Methodological advantages

We found that rats consumed many soft-bodied species (e.g. mushrooms, flat worms, slugs, and lepidopterans) that would be difficult to identify using visual inspection of stomach contents. Capturing such a wide variety of taxa would be difficult to quantify using metabarcoding methods, as there are no universal 18S or COI universal primers capable of amplifying sequences in all these taxa. While it might be possible to use primer sets targeted at different phyla or orders, quantitatively comparing diet components across these using sequences amplified with different primer sets is extremely difficult due to differences in primer binding and PCR efficiency.

The nanopore MinION-based sequencing method used in this simple metagenomic approach has several advantages. Compared to other high throughput sequencing technologies (e.g. Illumina, IonTorrent, or PacBio), there is no initial capital investment

460 required to use the platform. On a per-sample basis, data generation is inexpensive
461 (approximately $150 USD per barcoded sample, and approximately half this price if
462 reagents are purchased in bulk). Library preparation and sequencing can be extremely
463 rapid, going from DNA sample to sequence in less than two hours (Zaaijer et al. 2017).
464 Furthermore, the sequencing platform itself is highly portable. As the cost of nanopore-
465 based sequencing continues to decrease (both per sample and per base pair), it should
466 become possible to use molecular methods for routine ecological monitoring of species
467 presence or absence in field settings, without significant investment in infrastructure
468 (Kamenova et al. 2017). Finally, we suggest that our approach of standardising the read
469 counts by sample, followed by an optional transformation such as square root and
470 dissimilarity-based multivariate ordination, offers a useful analytical pipeline for
471 analysing metagenomic diet-composition data.

472 Conclusion

473 Here we have shown that a rapid long-read metagenomic approach is able to accurately
474 characterise diet taxa at the family-level and distinguish between the diets of rats
475 according to the locations from which they were sourced, with almost 80% out-of-
476 sample classification success. This approach also identifies the taxa responsible for
477 such patterns. This information may be used to guide conservation efforts toward
478 specific areas and habitats in which native species are most at risk from this highly
479 destructive introduced predator.

480

481 ## Acknowledgements

482 This work was supported by a Massey University Research Fund to NF, a Marsden
483 Fund Grant (15-MAU-136) to JD and Marsden Fund Grant (MAU1703) to OS. Thanks to
484 Friends of Okura Bush, Mary Stewart from Auckland Council, and Gillian Wadams and
485 the volunteers at the Waitakere Ranges for collecting rat samples. Sample collection
486 was performed under (Auckland Council Permit to Undertake Research WS1064).

487

488 ## Data Accessibility

489 Sequence data are available in ENA (accession number PRJEB27647)

490 ## Author Contributions

491 WP, JD, NF, and OS conceived the project. WP performed the stomach dissections.
492 WP and NF optimised the genomic DNA isolation and library preparation. NF performed
493 the nanopore sequencing. GB and OS processed and performed quality control on the
494 sequencing data. WP and OS performed the sequence classification. WP, AS, NF, and
495 OS analysed the data. WP, NF, AS, and OS wrote the paper, with input from all authors.

496

Supplemental Tables

Table S1. Read numbers and total base pairs for each barcode in the January sequencing run.

Rat	Total reads	Total Mbp	Mean length
OB2	19.907	14.62	734
WP11	10,164	9.63	947
WP5	8237	6.78	823
LB7	7548	7.04	933
OB13	3644	3.63	995
WP9	2954	2.4	814
OB5	2850	2.06	721
WP8	2801	2.32	827
LB6	2531	1.6	632
OB7	2473	1.87	756
LB5	1641	1.16	705
LB3	1554	0.99	636
None	16673	13.01	781
Total	82,977	67.1	

Table S2. Read numbers and total base pairs for each barcode in the March sequencing run.

Barcode	Total	Total Mbp	Mean
LB1	17,820	9.21	517
LB8	16,923	13.13	776
WP2	10,511	7.00	666
LB4	8684	4.92	567
OB11	5689	3.40	598
WP10	1563	0.99	633
OB12	1479	0.89	604
WP12	1309	0.78	596
LB2	1127	0.76	676
WP3	637	0.73	1141
OB14	541	0.37	683
OB10	435	0.24	555
None	29,432	21.33	725
Total	96,150	63.75	

504 **Table S3.** SIMPER analysis of family contributions to group similarities. Abbreviations:
505 Av.Abund: average abundance; Av.Si: average similarity; Sim/SD: similarity over
506 standard deviation; Percent.contr: percent contribution.

507

Family	Av.Abun	Av.Si	Sim/SD	Percent.cont	Group
Hymenolepididae	3.37	6.87	0.34	51.2	LB
Solanaceae	1.57	1.48	0.34	11.1	LB
Fabaceae	1.74	1.41	0.44	10.5	LB
Arecaceae	2.86	7.11	1	33.4	OB
Poaceae	2.87	4.82	0.55	22.7	OB
Fabaceae	1.17	1.98	0.51	9.3	OB
Phasianidae	1.79	1.67	0.34	7.9	OB
Poaceae	5.08	17.61	0.62	72.1	WP

508

509 **Table S4.** SIMPER analysis of family contributions to group dissimilarities. The families
510 are sorted by percent contribution to group dissimilarity. Abbreviations: Abund.Group1
511 and Abund.Group2: abundance of each family in groups 1 and 2 (see Group 1 and 2
512 columns). Av.Diss: average dissimilarity; Diss/SD: Dissimilarity over standard deviation;
513 Percent.contr: percent contribution.

Family	Abund.Group1	Abund.Group2	Av.Diss	Diss/SD	Percent.contr	Group1	Group2
Poaceae	1.95	5.08	15.15	1.04	16.74	LB	WP
Poaceae	2.87	5.08	11.29	1.26	13.78	OB	WP
Hymenolepididae	3.37	0.48	10.8	0.73	11.93	LB	WP
Hymenolepididae	3.37	0.29	9.37	0.79	10.32	LB	OB
Poaceae	1.95	2.87	8.37	1.1	9.22	LB	OB
Arecaceae	0.05	2.86	6.99	1.41	7.7	LB	OB
Arecaceae	2.86	1.31	5.92	1.29	7.23	OB	WP
Fabaceae	1.74	1.05	6.14	0.67	6.78	LB	WP
Podocarpaceae	0	2.38	5.34	0.83	5.9	LB	WP
Podocarpaceae	0.71	2.38	4.82	0.99	5.88	OB	WP
Fabaceae	1.74	1.17	4.87	0.81	5.37	LB	OB
Fabaceae	1.17	1.05	4.31	0.84	5.26	OB	WP
Phasianidae	1.79	0	4.03	0.7	4.92	OB	WP
Phasianidae	0	1.79	4.28	0.72	4.71	LB	OB
Physalopteridae	1.63	0.18	3.79	0.59	4.63	OB	WP
Physalopteridae	0.46	1.63	4.15	0.65	4.57	LB	OB
Solanaceae	1.57	0	3.95	0.65	4.37	LB	WP
Solanaceae	1.57	0	3.56	0.67	3.92	LB	OB
Helicidae	0.87	0.47	3.46	0.52	3.82	LB	WP
Oleaceae	1.06	0.25	3.09	0.44	3.77	OB	WP
Arecaceae	0.05	1.31	3.4	0.65	3.75	LB	WP
Piperaceae	1.42	0	3.36	0.65	3.71	LB	WP
Helicidae	0.87	0.82	3.33	0.58	3.66	LB	OB
Cerambycidae	1.41	0.18	2.99	0.77	3.65	OB	WP
Cerambycidae	0	1.41	3.07	0.74	3.38	LB	OB
Oleaceae	0	1.06	3.06	0.41	3.37	LB	OB
Piperaceae	1.42	0	3.04	0.67	3.35	LB	OB
Onchocercidae	0.39	1.04	3.02	0.77	3.33	LB	OB
Geoplanidae	0.8	0.48	2.64	0.49	3.22	OB	WP
Thelaziidae	0.9	0.64	2.85	0.77	3.14	LB	OB
Helicidae	0.82	0.47	2.54	0.61	3.1	OB	WP
Onchocercidae	1.04	0	2.49	0.66	3.04	OB	WP
Formicidae	0.44	1.17	2.39	1.04	2.92	OB	WP

Sclerotiniaceae	0	1.28	2.63	0.57	2.91	LB	WP
Sclerotiniaceae	0	1.28	2.3	0.57	2.81	OB	WP
Formicidae	0	1.17	2.46	0.94	2.72	LB	WP
Rosaceae	0.92	0.3	2.13	0.58	2.61	OB	WP
Ectobiidae	0.63	0.34	2.24	0.5	2.48	LB	WP
Thelaziidae	0.9	0	2.13	0.56	2.36	LB	WP
Geoplanidae	0	0.8	2.04	0.37	2.25	LB	OB
Rosaceae	0	0.92	1.99	0.5	2.19	LB	OB
Numididae	0.78	0	1.68	0.56	2.05	OB	WP
Myrtaceae	0.66	0	1.8	0.38	1.99	LB	WP
Thelaziidae	0.64	0	1.53	0.51	1.86	OB	WP
Ranunculaceae	0.66	0	1.36	0.37	1.5	LB	WP

Datafile S1. Table of read BLAST hits and assigned MEGAN taxa with reads reclassified at the family or order level by filtering on read length to alignment length ratio and percent identity.

Datafile S2. Table of read BLAST hits and assigned MEGAN taxa with no filters applied.

Supplemental Figures

Fig S1. Biplots of read lengths and qualities for each barcode in the January and March runs.

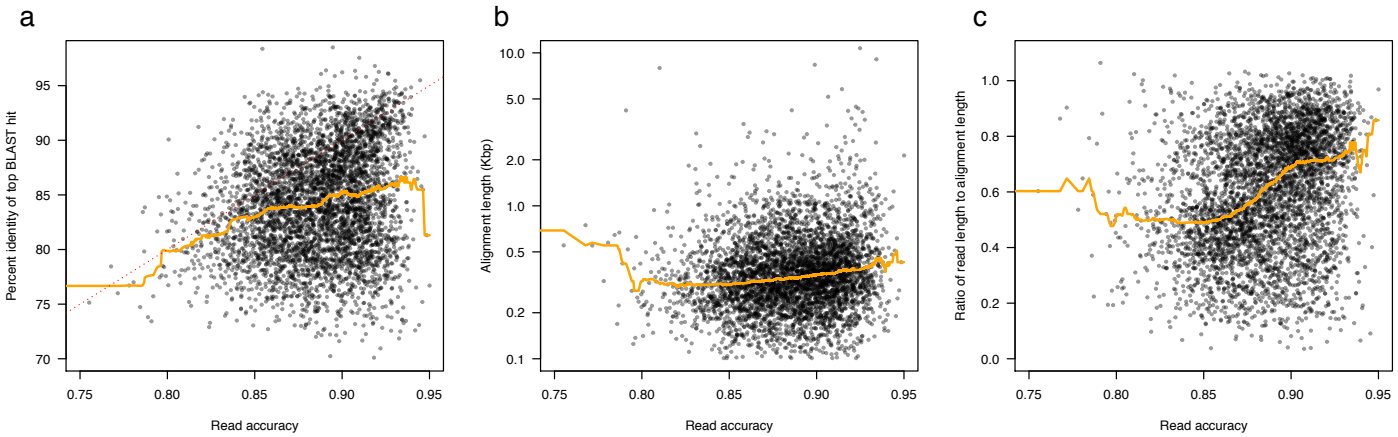


Fig S2. Correlation of read accuracy with alignment characteristics. **(a)** Read accuracy is positively correlated with the percent identity of the top BLAST hit. Points show a subsample of reads; orange line indicates a running median; red dotted line is the y=x line, which is expected if accuracy corresponds exactly to percent identity. **(b)** Read accuracy and alignment length show no relationship. **(c)** Read accuracy and the ratio of read length to alignment length are positively correlated.

References

1. Anderson, Marti J., and Trevor J. Willis. 2003. “Canonical Analysis Of Principal Coordinates: A Useful Method Of Constrained Ordination For Ecology.” *Ecology* 84 (2): 511–25.

2. Basha, Walaa A., Andrew T. Chamberlain, Moushira E. Zaki, Wafaa A. Kandeel, and Nagui H. Fares. 2016. “Diet Reconstruction through Stable Isotope Analysis of Ancient Mummified Soft Tissues from Kulubnarti (Sudanese Nubia).” *Journal of Archaeological Science: Reports* 5 (February): 71–79.

3. Bridgman, L. J., J. Innes, Craig Gillies, Neil Fitzgerald, and Carolyn M. King. 2013. “Do Ship Rats Display Predatory Behaviour towards House Mice?” *Animal Behaviour* 86 (2)): 257–68.

4. Brown, K. P., H. Moller, J. Innes, and P. Jansen. 2008. “Identifying Predators at Nests of Small Birds in a New Zealand Forest.” *The Ibis* 140 (2): 274–79.

5. Brownlee, A., and W. Moss. 1961. “The Influence of Diet on Lactobacilli in the Stomach of the Rat.” *The Journal of Pathology* 82 (2): 513–16.

6. Carreon-Martinez, L., and D. D. Heath. 2010. “Revolution in Food Web Analysis

and Trophic Ecology: Diet Analysis by DNA and Stable Isotope Analysis." *Molecular Ecology* 19 (1): 25–27.

7. Clarke, K. R. 1993. "Non-Parametric Multivariate Analyses of Changes in Community Structure." *Austral Ecology* 18 (1): 117–43.

8. Clarke, K. R., and R. N. Gorley. 2015. "PRIMER v7: User Manual/Tutorial." PRIMER-E, Plymouth.

9. Clarke, K. R., and R. H. Green. 1988. "Statistical Design and Analysis for a 'Biological Effects' Study." *Marine Ecology Progress Series* 46: 213–26.

10. Clarke, K. Robert, K. Robert Clarke, Paul J. Somerfield, and M. Gee Chapman. 2006. "On Resemblance Measures for Ecological Studies, Including Taxonomic Dissimilarities and a Zero-Adjusted Bray–Curtis Coefficient for Denuded Assemblages." *Journal of Experimental Marine Biology and Ecology* 330 (1): 55–80.

11. Daniel, M. J. 1973. "Seasonal Diet Of The Ship Rat (Rattus Rattus) In Lowland Forest In New Zealand." *Proceedings of the New Zealand Ecological Society* 20: 21–30.

12. Deagle, Bruce E., J. Paige Eveson, and Simon N. Jarman. 2006. "Quantification of Damage in DNA Recovered from Highly Degraded Samples--a Case Study on DNA in Faeces." *Frontiers in Zoology* 3 (August): 11.

13. Deagle, Bruce E., Simon N. Jarman, Eric Coissac, François Pompanon, and Pierre Taberlet. 2014. "DNA Metabarcoding and the Cytochrome c Oxidase Subunit I Marker: Not a Perfect Match." *Biology Letters* 10 (9). https://doi.org/10.1098/rsbl.2014.0562.

14. Diamond, J. M., and C. R. Veitch. 1981. "Extinctions and Introductions in the New Zealand Avifauna: Cause and Effect?" *Science* 211 (4481): 499–501.

15. Dowding, John E., and Elaine C. Murphy. 2001. "The Impact of Predation by Introduced Mammals on Endemic Shorebirds in New Zealand: A Conservation Perspective." *Biological Conservation* 99 (1): 47–64.

16. Dunlap, M., and J. R. Pawlik. 1996. "Video-Monitored Predation by Caribbean Reef Fishes on an Array of Mangrove and Reef Sponges." *Marine Biology* 126 (1): 117–23.

17. Gibbs, George W. 1998. "Why Are Some Weta (Orthoptera: Stenopelmatidae) Vulnerable yet Others Are Common?" *Journal of Insect Conservation* 2 (3-4):

161–66.

18. Gillies, Craig, and Mick Clout. 2003. “The Prey of Domestic Cats (Felis Catus) in Two Suburbs of Auckland City, New Zealand.” *Journal of Zoology* 259 (3): 309–15.

19. Hobson, Keith A. 1987. “Use of Stable-Carbon Isotope Analysis to Estimate Marine and Terrestrial Protein Content in Gull Diets.” *Canadian Journal of Zoology* 65 (5): 1210–13.

20. Horáková, Z., C. H. Zierdt, and M. A. Beaven. 1971. “Identification of Lactobacillus as the Source of Bacterial Histidine Decarboxylase in Rat Stomach.” *European Journal of Pharmacology* 16 (1): 67–77.

21. Huson, Daniel H., Alexander F. Auch, Ji Qi, and Stephan C. Schuster. 2007. “MEGAN Analysis of Metagenomic Data.” *Genome Research* 17 (3): 377–86.

22. Huson, Daniel H., Sina Beier, Isabell Flade, Anna Górska, Mohamed El-Hadidi, Suparna Mitra, Hans-Joachim Ruscheweyh, and Rewati Tappu. 2016. “MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data.” *PLoS Computational Biology* 12 (6): e1004957.

23. Jain, Miten, Hugh E. Olsen, Benedict Paten, and Mark Akeson. 2016. “The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community.” *Genome Biology* 17 (1): 239.

24. Jarman, S. N., B. E. Deagle, and N. J. Gales. 2004. “Group-Specific Polymerase Chain Reaction for DNA-Based Analysis of Species Diversity and Identity in Dietary Samples.” *Molecular Ecology* 13 (5): 1313–22.

25. Jarman, S. N., N. J. Gales, M. Tierney, P. C. Gill, and N. G. Elliott. 2002. “A DNA-Based Method for Identification of Krill Species and Its Application to Analysing the Diet of Marine Vertebrate Predators.” *Molecular Ecology* 11 (12): 2679–90.

26. Kamenova, S., T. J. Bartley, D. A. Bohan, J. R. Boutain, R. I. Colautti, I. Domaizon, C. Fontaine, et al. 2017. “Chapter Three - Invasions Toolkit: Current Methods for Tracking the Spread and Impact of Invasive Species.” In *Advances in Ecological Research*, edited by David A. Bohan, Alex J. Dumbrell, and François Massol, 56:85–182. Academic Press.

27. Kim, Daehwan, Li Song, Florian P. Breitwieser, and Steven L. Salzberg. 2016. “Centrifuge: Rapid and Sensitive Classification of Metagenomic Sequences.” *Genome Research* 26 (12): 1721–29.

28. King, R. A., D. S. Read, M. Traugott, and W. O. C. Symondson. 2008. "Molecular Analysis of Predation: A Review of Best Practice for DNA-Based Approaches." *Molecular Ecology* 17 (4): 947–63.

29. Leray, Matthieu, Joy Y. Yang, Christopher P. Meyer, Suzanne C. Mills, Natalia Agudelo, Vincent Ranwez, Joel T. Boehm, and Ryuji J. Machida. 2013. "A New Versatile Primer Set Targeting a Short Fragment of the Mitochondrial COI Region for Metabarcoding Metazoan Diversity: Application for Characterizing Coral Reef Fish Gut Contents." *Frontiers in Zoology* 10 (June): 34.

30. Lewin, Harris A., Gene E. Robinson, W. John Kress, William J. Baker, Jonathan Coddington, Keith A. Crandall, Richard Durbin, et al. 2018. "Earth BioGenome Project: Sequencing Life for the Future of Life." *Proceedings of the National Academy of Sciences of the United States of America* 115 (17): 4325–33.

31. Li, Dongyao, Haiqin Chen, Bingyong Mao, Qin Yang, Jianxin Zhao, Zhennan Gu, Hao Zhang, Yong Q. Chen, and Wei Chen. 2017. "Microbial Biogeography and Core Microbiota of the Rat Digestive Tract." *Scientific Reports* 8 (April): 45840.

32. Major, H. L., I. L. Jones, M. R. Charette, and A. W. Diamond. 2007. "Variations in the Diet of Introduced Norway Rats (*Rattus norvegicus*) Inferred Using Stable Isotope Analysis." *Journal of Zoology* 271 (4): 463–68.

33. Maurice, Corinne F., Sarah C. L. Knowles, Joshua Ladau, Katherine S. Pollard, Andy Fenton, Amy B. Pedersen, and Peter J. Turnbaugh. 2015. "Marked Seasonal Variation in the Wild Mouse Gut Microbiota." *The ISME Journal* 9 (11): 2423–34.

34. McIntyre, Alexa B. R., Rachid Ounit, Ebrahim Afshinnekoo, Robert J. Prill, Elizabeth Hénaff, Noah Alexander, Samuel S. Minot, et al. 2017. "Comprehensive Benchmarking and Ensemble Approaches for Metagenomic Classifiers." *Genome Biology* 18 (1): 182.

35. Nasko, Daniel J., Sergey Koren, Adam M. Phillippy, and Todd J. Treangen. 2018. "RefSeq Database Growth Influences the Accuracy of K-Mer-Based Species Identification," 1–21.

36. Paula, Débora P., Benjamin Linard, David A. Andow, Edison R. Sujii, Carmen S. S. Pires, and Alfried P. Vogler. 2015. "Detection and Decay Rates of Prey and Prey Symbionts in the Gut of a Predator through Metagenomics." *Molecular Ecology Resources* 15 (4): 880–92.

37. Paula, Débora P., Benjamin Linard, Alex Crampton-Platt, Amrita Srivathsan,

685 Martijn J. T. N. Timmermans, Edison R. Sujii, Carmen S. S. Pires, Lucas M.
686 Souza, David A. Andow, and Alfried P. Vogler. 2016. “Uncovering Trophic
687 Interactions in Arthropod Predators through DNA Shotgun-Sequencing of Gut
688 Contents.” *PloS One* 11 (9): e0161841.
689
690 38. Pawluczyk, Marta, Julia Weiss, Matthew G. Links, Mikel Egaña Aranguren, Mark
691 D. Wilkinson, and Marcos Egea-Cortines. 2015. “Quantitative Evaluation of Bias
692 in PCR Amplification and next-Generation Sequencing Derived from
693 Metabarcoding Samples.” *Analytical and Bioanalytical Chemistry* 407 (7): 1841–
694 48.
695
696 39. Kerry P. Brown. 1997. “Predation at Nests of Two New Zealand Endemic
697 Passerines; Implications for Bird Community Restoration.” *Pacific Conservation*
698 *Biology: A Journal Devoted to Conservation and Land Management in the Pacific*
699 *Region* 3 (2): 91.
700
701 40. Pereira, Rui P. A., Jorg Peplies, Ingrid Brettar, and Manfred G. Hoefle. 2018.
702 “Impact of DNA Polymerase Choice on Assessment of Bacterial Communities by
703 a Legionella Genus-Specific next-Generation Sequencing Approach.” *bioRxiv*.
704 https://doi.org/10.1101/247445.
705
706 41. Pierce, Graham J., and Peter R. Boyle. 1991. “A Review of Methods for Diet
707 Analysis in Piscivorous Marine Mammals.” *Oceanography and Marine Biology:*
708 *An Annual Review* 29: 409–86.
709
710 42. Riofrío-Lazo, Marjorie, and Diego Páez-Rosas. 2015. “Feeding Habits of
711 Introduced Black Rats, *Rattus rattus*, in Nesting Colonies of Galapagos Petrel on
712 San Cristóbal Island, Galapagos.” *PloS One* 10 (5): e0127901.
713
714 43. Russell, James C., John G. Innes, Philip H. Brown, and Andrea E. Byrom. 2015.
715 “Predator-Free New Zealand: Conservation Country.” *Bioscience* 65 (5): 520–25.
716
717 44. Soininen, Eeva M., Alice Valentini, Eric Coissac, Christian Miquel, Ludovic Gielly,
718 Christian Brochmann, Anne K. Brysting, et al. 2009. “Analysing Diet of Small
719 Herbivores: The Efficiency of DNA Barcoding Coupled with High-Throughput
720 Pyrosequencing for Deciphering the Composition of Complex Plant Mixtures.”
721 *Frontiers in Zoology* 6 (August): 16.
722
723 45. Srivathsan, Amrita, Andie Ang, Alfried P. Vogler, and Rudolf Meier. 2016. “Fecal
724 Metagenomics for the Simultaneous Assessment of Diet, Parasites, and
725 Population Genetics of an Understudied Primate.” *Frontiers in Zoology* 13 (April):
726 17.
727
728 46. Srivathsan, Amrita, John C. M. Sha, Alfried P. Vogler, and Rudolf Meier. 2015.

"Comparing the Effectiveness of Metagenomics and Metabarcoding for Diet Analysis of a Leaf-Feeding Monkey (Pygathrix Nemaeus)." *Molecular Ecology Resources* 15 (2): 250–61.

47. Stringer, Ian A. N., Suzanne M. Bassett, Megan J. McLean, Jay McCartney, and G. Richard Parrish. 2003. "Biology and Conservation of the Rare New Zealand Land Snail Paryphanta Busbyi Watti (Mollusca, Pulmonata)." *Invertebrate Biology: A Quarterly Journal of the American Microscopical Society and the Division of Invertebrate Zoology/ASZ* 122 (3): 241–51.

48. Sweetapple, Peter J., and Graham Nugent. 2007. "Ship Rat Demography and Diet Following Possum Control in a Mixed Podocarp—hardwood Forest." *New Zealand Journal of Ecology* 31 (2): 186–201.

49. Tedersoo, Leho, Sten Anslan, Mohammad Bahram, Sergei Põlme, Taavi Riit, Ingrid Liiv, Urmas Kõljalg, et al. 2015. "Shotgun Metagenomes and Multiple Primer Pair-Barcode Combinations of Amplicons Reveal Biases in Metabarcoding Analyses of Fungi." *MycoKeys* 10: 1.

50. Towns, David R., Charles H. Daugherty, and Alison Cree. 2001. "Raising the Prospects for a Forgotten Fauna: A Review of 10 Years of Conservation Effort for New Zealand Reptiles." *Biological Conservation* 99 (1): 3–16.

51. Wood, Derrick E., and Steven L. Salzberg. 2014. "Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments." *Genome Biology* 15 (3): R46.

52. Zaaijer, Sophie, Assaf Gordon, Daniel Speyer, Robert Piccone, Simon Cornelis Groen, and Yaniv Erlich. 2017. "Rapid Re-Identification of Human Samples Using Portable DNA Sequencing." *eLife* 6 (November). https://doi.org/10.7554/eLife.27798.