Evaluating metabarcoding to analyse diet composition of species foraging in anthropogenic landscapes using Ion Torrent and Illumina sequencing

Marie-Amélie Forin-Wiart[1,2.3] (MAFW), Maxime Galan[4] (MG), Sylvain Piry[4] (SP), Jean-François Cosson[4] (JFC), Claire Larose[5] (CL) and Marie-Lazarine Poulle[1,2] (MLP)

[1] Université de Reims Champagne-Ardenne (URCA), SFR Cap-Santé, EA3800-PROTAL, UFR de Médecine, 1 rue du Maréchal Juin, 51098 Reims Cedex, France

[2] URCA, CERFE, 5 rue de la Héronnière, 08240 Boult-aux-Bois, France

[3] Université de Strasbourg, CNRS, Département Ecologie, Physiologie et Ethologie, IPHC UMR 7178, F-67087 Strasbourg, France.

[4] CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Université de Montpellier, 755 avenue du Campus Agropolis, CS 300 16, 34988 Montferrier-sur-Lez cedex, France

[5] DIANA PetFood Division, Z.A. du Gohélis, 56250, Elven, France

Corresponding author: Marie-Amélie Forin-Wiart, Université de Reims Champagne-Ardenne, CERFE, 5 rue de la Héronnière, 08240 Boult-aux-Bois, France. email: ma.forinw@hotmail.fr

Running title: HTS platform sensitivity in diet metabarcoding

Abstract

Determining the foraging niche of wild species is a challenging task, recently overcome through DNA metabarcoding. This method is successfully used to identify ingested food items of faecal samples. However, several limitations can affect the output of this method. Here, two benchtop high-throughput sequencing (HTS) platforms were evaluated, the Ion PGM (Life Technologies) and the MiSeq (Illumina). We assessed their ability to identify a large taxonomic array of food items for species having a diet possibly containing human-processed products and prey items. We sequenced faecal DNA samples obtained from domestic cats *Felis silvestris catus*, generalist carnivores presenting numerous specific DNAs in their diet. We assessed platform sensitivity and data reproducibility using both a cat fed on a controlled mixed diet along with a calibration study and free-ranging individuals sampled from a rural area. All the samples were analysed in independent PCR duplicates on each platform. The results of the calibration study showed that the output from MiSeq was more sensitive and reproducible than that from Ion PGM. Detection variability was linked to DNA degradation: 82% of the expected occurrences were detected for intact preys, versus 54% for pet food and 31% for out-of-date raw fish and meat. The Ion PGM detected significantly less human-linked food, birds, field voles, murids and shrews in the field-collected samples than the MiSeq platform. Pooling the replicates from both platforms and filtering HTS data allowed identifying at least one food item in 87.4% of the 326 field-collected samples. A total of 29 prey taxa (90% at species level) and 33 human-related food taxa (76% at genus level) were identified. The findings indicate that using HTS platforms such as MiSeq, which provide reads of sufficiently high quantity and quality, with sufficient numbers of technical replicates, is a robust and non-invasive approach to determine diet composition in animals foraging on a wide range of items in anthropogenic landscapes.

Key-words: cyt*b*, HTS data filtering, MiSeq sequencing, non-invasive approach, sequencer bias, variability in DNA degradation

Introduction

With increasing urbanisation and loss of natural habitat, several species have adapted to human-dominated landscapes by shifting their dietary niche to include human-processed food items [1]. For instance, scavengers such as gulls *Larus spp.*, ravens *Corvus spp.*, Eurasian badgers *Meles meles*, opossums *Didelphis virginiana*, racoons *Procyon lotor*, or wild boar *Sus scrofa*, as well as opportunistic generalist carnivores such as coyotes *Canis latrans* [2], or foxes *Vulpes spp.* [3,4], presently have dietary compositions that differ substantially from their natural diet. Several large carnivores such as bears, wolves and hyenas have also altered their diet to include significant parts of human-processed products (see Bateman and colleagues [5] for a review). Understanding to what extent species in close contact with humans may have changed their diet to include human-processed items is critical both from a conservation perspective to determine the likelihood of human-wildlife conflicts, and from an evolutionary perspective to determine species plasticity in foraging niches in a context of global change. Thus, accurately identifying a large diversity of food items in the diet composition of studied species, including those from human-linked food sources is necessary for ecologists.

Foraging ecology studies generally rely on the preservation of hard remains found in the faeces, pellets, or stomach contents of studied species to determine their diet, using visual recognition of morphologically features by macro- or microscopic methods [6]. Whereas useful, these techniques generally present a poor resolution in terms of taxa identification or preclude the identification of food items that do not leave behind hard remains [7,8]. In contrast, DNA metabarcoding has been demonstrated as an accurate method for diet analysis using faecal samples as a source of DNA [9]. As a result, this technique has become a key application in high-throughput sequencing (HTS) technologies in ecological research [10]. Its aim is the

analysis of biological samples (of which there is no prior knowledge) that contain a mixture of DNA using PCR amplification of a gene marker such as universal DNA barcodes.

Metabarcoding in HTS technologies allows DNA to be read from multiple templates for hundreds of samples in the same sequencing run. A subunit of a mitochondrial gene – for example, a 136-bp mini-barcode on the cytochrome *b* gene (cyt*b*) – is enough to distinguish between closely related rodent species even in degraded or ancient DNA samples [11]. By comparing obtained sequences to a standard reference library of taxonomically identified organisms (e.g. Bold or GenBank), the taxa or species present in a sample can be identified with high confidence. In this way, metabarcoding of faecal samples allows ingested food items to be identified by characterizing the barcode sequences in the samples [12]. This technique has been successfully used to analyse the diet of species of birds [13], mammals [11,14,15], reptiles [16,17] and fish [18].

However promising, metabarcoding for dietary analyses is not free of limitations or bias (see Table 1 in Galan and colleagues [19] for a review). Contamination from alien DNA during fieldwork or cross-contamination during lab procedures can occur, resulting in false positive results [20]. Furthermore, the library preparation protocol, the targeted DNA region, the primer choice and the sequencing platform can all affect the identification of the composition of environmental samples [21]. The degree of DNA degradation in an analysed sample also influences the sensitivity of PCR-based methods [22]. In addition, the sequencing of PCR products from enriched samples that contain predator DNA could reduce the potential detection of the prey sequences [15]. As a result of these potential biases, alone or in combination, the relative abundance of some species in faecal samples can be under- or overestimated in a DNA analysis, while other species may not be amplified and so remain absent from the bioinformatics output [23]. Although several studies have described most of

these challenges, to our knowledge the specific use of HTS platforms to analyse samples from a dietary perspective has never been evaluated.

The main objective of this work was to assess the ability of benchtop HTS platforms in identifying a large taxonomic array for species having a diet possibly containing human-processed products. To maximise the variability in DNA samples obtained from the diet and ensure the presence of human-linked food items, we focused on a high generalist carnivore species, the domestic cat *Felis silvestris catus*. Cats appear as a particularly interesting model to test for metabarcoding sensitivity since in addition to human-processed products, free-ranging individuals are known to feed on a diversity of prey items including rodent species, birds, reptiles, amphibians and insects [24,25]. Specifically, the sensitivity of the two most popular sequencers available for metabarcoding analyses, the Life Technologies Ion PGM (Personal Genome Machine) platform and the Illumina MiSeq platform, in identifying food items from faecal DNA having undergone different degrees of degradation (fresh prey items versus processed food versus out-of-date raw fish and meat) was investigated. This was achieved by collecting faecal samples from (1) a cat fed on a mixed diet of known composition in an initial calibration study, and (2) free-ranging individuals sampled from a rural area. Since all samples were analysed in independent PCR duplicates on each HTS platform, this allowed us to compare the reproducibility of diet results both within the platform and among platforms.

Materials and methods

Faecal sample collection

Calibration study. A feeding trial was performed with a housebound cat fed a known diet during 22 consecutive days in May 2012 and 27 consecutive days in November 2012 (see Table S1 for details). The daily diet of the housebound cat was constant in quantity but

differed in composition from day to day. Daily rations included one or two of the following categories of food item: (1) processed wet or dry pet food (commercial meat-based cat food, Eco+ product line, SCAMARK, France); and (2a) prey items or (2b) out-of-date raw fish or meat. The daily rations of individual prey species, processed pet food, and out-of-date raw fish and meat were weighed and distributed evenly in three to four meals. The faecal samples were then collected from a litter tray that was cleaned out daily, and these were put in labelled plastic bags and stored at -20 °C. See Table S1 for detailed information related to daily ration contents.

Field study. The faecal samples from free-ranging domestic cats were collected in a rural area located in northeastern France (Ardennes region) as part of a joint study (Forin-Wiart et al. unpublished data). Cats were sampled from a monitored free-ranging domestic cat population composed of 162 individuals, 40% of which were free-ranging outdoor cats fed *ad libitum* by their owners, and 60% of which were farm cats fed from time to time with milk, dry pet food or leftovers (see Forin-Wiart and colleagues [26] for the census and cat feeding status definitions). Over 100 fresh faecal samples were collected in June–July 2011, November–December 2011 and March 2012 along an 11-km survey transect delineated in villages and isolated farms and a 15-km transect in two areas surrounding the villages. Additional faecal samples were also collected from the litter trays of free-ranging individuals. Putative cat faecal samples were identified in the field by their shape and size [27] prior to molecular identification based on cat-specific quantitative PCR (qPCR, see below). The 359 putative felid faecal samples were labelled with their location (using a GPS) and stored at -20 °C in plastic bags.

DNA extraction

All the DNA extractions were performed at the SAS SPYGEN (France) laboratory in a room dedicated to degraded or rare DNA. Total DNA was extracted from approximately 500 mg of faecal material taken from all parts of the scat, as predator and prey DNA are not evenly distributed. The QIAamp DNA Stool Mini Kit was used following the manufacturer's instructions (Qiagen). Ten extraction blanks (containing no faecal sample) were included as a negative control to test for contamination. DNA extraction success was determined by real-time PCR (qPCR) amplification with a cat species-specific primer [28] by SAS SPYGEN. This step was also used to validate the origin of field-collected faecal samples.

DNA metabarcoding

A 136-bp mini-barcode of the cytochrome *b* gene was targeted, as it is effective in identifying a large range of vertebrates, including ungulates, rodents, birds and fish, from degraded samples [11,29,30].

Sequencing was successively performed using the Ion Personal Genome Machine (Ion PGM; Life Technologies) platform and the MiSeq (Illumina) platform to analyse the 400 DNA extractions obtained from the putative cat faecal samples (41 samples from the calibration study and 359 from the field study). Each DNA extraction was sequenced four times: twice independently on each HTS platform. Due to the specificities of each platform, the DNA library preparation differed slightly (see below).

The Ion PGM libraries were performed twice for each DNA extraction by amplifying the 136-bp fragment of cytochrome *b* described in Galan and colleagues [11]. We used the fusion primers L15411F (5'-CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG NNNNNNN GAY AAA RTY CCV TTY CAY C-3') and H15546R-PGM (5'CCT CTC TAT GGG CAG TCG GTG AT NNNNNNN AAR TAY CAY TCD GGY TTR AT-3') modified in 5' by the addition of individual-specific 7-bp MIDs (Multiplex IDentifiers NNNNNNN) and adaptors

required for the emPCR and the Ion PGM sequencing. PCRs were performed following the procedure detailed in Galan and colleagues [11]. After PCR pooling and size selection by gel extraction, the amplicon libraries were sequenced by the company Genes Diffusion (Douai, France) on an Ion PGM system using an Ion 316 Chip Kit version 2 (Ion Torrent, Life Technologies).

The MiSeq libraries were performed using a two-step PCR protocol (see Illumina Application Note Part #15044223) combined with the dual-index paired-end sequencing approach described in Kozich and colleagues [31] due to the significant length of the Illumina adaptors (see details in Galan and colleagues [32]). During PCR1, the 136-bp fragment of cyt*b* was amplified twice for each DNA extraction with the primers modified in 5' by the addition of a partial overhang Illumina adapter: L15411F-MiSeq 5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG (none or C or TC or ATG) GAY A AA RTY CCV TTY CAY CC-3' and H15546R-MiSeq 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G (none or C or TG or GCT) AAR TAY CAY TCD GGY TTR AT-3'. The alternative bases between the partial adaptors and the target-specific primers corresponded to a 0- to 3-bp 'heterogeneity spacer' designed to mitigate the issues caused by low sequence diversity amplicons in Illumina sequencing [33]. These four versions of each primer were mixed before PCR1, which was performed in the same conditions as the Ion PGM PCR libraries, i.e. following the procedure detailed in Galan and colleagues [11].

PCR1 was then used as the DNA template for PCR2, consisting of a limited-cycle amplification step to add multiplexing indices and Illumina sequencing adapters P5 and P7 at both ends of each DNA fragment. The primers P5 (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC AC NNNNNNNN TCG TCG GCA GCG TC-3') and P7 (5'-CAA GCA GAA GAC GGC ATA CGA GAT NNNNNNNN GTC TCG TGG GCT CGG-3') were synthesized with the different index sequences (NNNNNNNN) described in Kozich and

colleagues [31]. PCR2 was carried out in a 10-µL reaction volume using 5 µL of PCR Multiplex Kit Master Mix (Qiagen) and 0.7 µM of each primer. One microliter of the PCR1 product was added to each well. PCR2 was started with an initial denaturation step of 95 °C for 15 min, followed by 5 cycles of denaturation at 94 °C for 40 s, annealing at 45 °C for 45 s and extension at 72 °C for 60 s, followed by a final extension step at 72 °C for 10 min. After PCR pooling and size selection by gel extraction, the amplicon library was sequenced at the local Plateau GPTR-GENOTYPAGE facility (Montpellier, France) on a MiSeq instrument by paired-end sequencing using a MiSeq 500 Cycle Kit version 2 (Illumina).

Data processing

Bioinformatics and taxonomic identification. The paired-end reads 1 and 2 produced by MiSeq were assembled with the 'make.contigs' tool available in the mothur program [34]. SESAME barcode software (SEquence Sorter & AMplicon Explorer [35]) was used for both Ion PGM and MiSeq data to sort the sequences and identify the haplotypes at a taxonomic level (genus or species) using GenBank and three different similarity thresholds for BLAST+FASTA assignment (95%, 98% and 100%).

Data filtering for false positive and negative occurrences. In line with Esling and colleagues [36], a mistagging threshold was calculated to prevent laboratory contamination or sample cross-contamination. This threshold corresponds to the number of reads for a particular haplotype below which a positive occurrence could be due to contamination. It is essential to filter out putative false positive results, especially for food items originating from highly represented item sequences in the sequencing run. The mistagging threshold was calculated using the maximal mistagging frequency in a sequencing run, *i.e.* the higher percentage of sequences assigned to a wrong sample and determined by the known mixed-diet study (e.g. the number of sequences of a rodent species assigned to a specific sample when this species was not

expected in the targeted sample). The mistagging threshold corresponds to the total number of reads in a sequencing run for a particular haplotype, multiplied by the maximal mistagging frequency (e.g. if an Ion PGM run produced 34,000 reads of an *Arvicola terrestris* haplotype and if the maximal mistagging frequency = 0.05%, then the mistagging threshold = 17 reads for this *Arvicola terrestris* haplotype).

Another possible source of bias is that DNA from a food item of low biomass or from a highly degraded sample can be stochastically amplified during the PCR process, resulting in a proportion of false negative results [37]. To overcome this problem of stochastic amplification, a specific food item had to occur in at least 50% of the PCR replicates performed for each DNA extraction of a given sample. Following these stages of data filtering, a table of occurrences of observed food items was obtained per sequencing run in order to compare the performance of the two platforms (see below). In a final step, to determine the diet represented by each faecal sample, the food occurrence data from both HTS platforms was pooled to enhance food item detection.

Data analysis

Model selection was based on the Information Theoretic approach and was performed with the Akaike Information Criterion (AIC) following the model selection approach described by Zuur and colleagues [38]. All statistical analyses were conducted with the statistical software R 3.3 (R Development Core Team 2016 [3939]), and a value of alpha = 0.05 was used for all tests.

Calibration study. The ability of the two HTS platforms to detect consumed items from a diet of no processed food (prey items only) or of different degrees of processed food (wet and dry pet food, raw fish or meat) was assessed through descriptive statistics. For each HTS platform and each food item category (prey, pet food or raw fish and meat), the number of faeces containing haplotype sequences informative enough to assign a food item (observed

occurrence) was compared to the minimum expected occurrence in order to identify the sensitivity and bias of each sequencing strategy. The minimum expected occurrence was determined by considering that the DNA associated with a given food item ingested by the housebound cat was expected to be detected in the scat emitted at least 12 h after the ingestion of the food item, according to the transit time previously assessed by Forin-Wiart and colleagues [26]. The accordance of food item detection (i.e. its occurrence in one, both or none of the replicates from a sequencing run) for each HTS platform and each successfully amplified scat was also estimated. Additionally, for mixed meals (i.e. including two food item categories), the proportion of each food item category by mass ingested by the housebound cat relative to the total mass of the food ingested was compared to the proportion of each food item category identified by haplotype information in the sequence reads relative to the total number of reads of a specific faecal sample. This measure allowed the investigation of whether the proportion of a given food item category in the diet of cats was reflected in the HTS sequence counts. These proportions were compared using Fisher's exact tests for count data.

Field study. Small- and medium-sized rodent, insectivore, wild bird and reptile DNA in field-collected samples were placed in the 'Prey items' category and its sub-categories (see Table S2 for details). In contrast, DNA from domestic ungulates (cow, goat, sheep, pig) and farmed birds (chicken, duck, guinea fowl, Perdicinae *spp.*, turkey), as well as other mammals (e.g. rabbit *Oryctolagus cuniculus*), farmed fish and game species (*Capreolus capreolus*, *Cervus elaphus* and *Sus scrofa*) were placed in the 'Human-linked food' category and its sub-categories (see Table S2 for details). This latter category was assumed to originate from pet food, leftovers or hunting scraps. These food item categories were defined to fit those previously obtained during diet studies.

To identify the sensitivity of each sequencing protocol, the probability of detecting a food item in field-collected samples, i.e. presence/absence (Table S3), was modelled as a function of the HTS platform (fPlatform), the food item category (fFoodItem, reptile occurrence was excluded for modelling purposes) and the interaction between fPlatform and fFoodItem, using a Generalized Linear Model (GLM) with a binomial link function. The fit of the full model (lme4 package [40]) and the reduced model was compared using a deviance test (G^2) as an alternative to an individual parameter-based approach by considering fSample as a random effect. Multiple comparisons with interaction terms were then adjusted using Tukey's post-hoc analyses (lsmeans package [41]).

Finally, the diet obtained from field-collected faecal samples was assessed through descriptive statistics.

Results

Sequencing results

The Ion PGM and MiSeq platforms provided 4,042,017 and 4,042,606 raw sequences passing filter respectively. For the Ion PGM reads, the percentage of bases with a quality score of 30 (Q30: incorrect base call probability of 1 in 1000) or higher was 61.6%; for the MiSeq reads, the percentage of bases with a quality score of 30 or higher was 96.9%. For the Ion PGM run, a high proportion of truncated sequences without reverse tags was observed (2,490,973 reads < 126bp excluding tags and primers) due to a significant decrease in quality at the end of the sequences. After demultiplexing, 688,530 (using Ion PGM) and 2,509,592 (using MiSeq) sequences of cyt*b* were assigned to the faecal samples and were identified to the closest taxonomic rank (genus or species) using the consensus of the three similarity thresholds (95%, 98% and 100%).

In the calibration study, a high proportion of cat haplotypes were obtained: 41,129 reads (80%) for Ion PGM and 235,673 reads (98%) for MiSeq at the 95% similarity threshold. There were also a high proportion of unidentified sequences: 5,069 reads (12%) for Ion PGM and 1,434 reads (0.6%) for MiSeq at the 95% similarity threshold. In the study of field-collected data, sequences from predators (cat, but also Canids and Mustelids, see below) represented 60% of the total read count using Ion PGM, and 88% of the total read count using MiSeq. For both platforms, a high proportion of cat haplotypes (301,329 reads [51%] for Ion PGM and 1,690,359 reads [83%] for MiSeq) and unidentified sequences (70,050 reads [12%] for Ion PGM and 125,086 reads [6%] for MiSeq, with a ≤ 95% identity using GenBank sequences) were obtained. As found in a previous study by Deagle and colleagues [42], some human DNA sequences were also present, accounting for 0.06% and 0.6% of the sequences generated by Ion PGM and MiSeq respectively.

DNA metabarcoding on known mixed-diet samples

In the MiSeq run, one out of 1934 sequences of a *Myodes glareolus* haplotype (n = 1, resulting in a frequency of 0.0517) was detected during the last 21 days of the calibration study despite the fact the housebound cat was fed only with pet food and raw fish or meat (see Table S1 for details), so this constituted a mistagging event. Accordingly, the maximal mistagging frequency in a sequencing run was set at 0.05%.

The DNA of prey items was mostly recovered in duplicates from the MiSeq (7 out of 8 replicated occurrences) and the Ion PGM (3 out of 4 replicated occurrences) runs, i.e. 100% of prey species expected in a given scat were detected (Table 1 and Table S1). MiSeq was the only platform to detect the Norway rat, the house mouse and a bird species (Table 1). DNA belonging to pet food contents was preferentially recovered with the MiSeq platform (85% of observed occurrences) as compared to the Ion PGM platform (45%) and was largely

recovered by duplicates with the MiSeq platform (71%) (Table 1). The DNA from raw fish and meat was scarcely detected by either method and was never detected in Ion PGM PCR replicates (Table 1 and Table S1).

Both platforms were better at detecting DNA from prey items (82% of the expected occurrences) than DNA from pet food (54%), or from raw fish and meat (31%) (Table 1). Independently of the food item considered, the percentage of faecal samples in which the expected DNA of a food item was detected (see Table S1) was higher for the MiSeq run (25/27, 93%) than for the Ion PGM run (14/27, 52%) (Table 1). Combining the results of the HTS runs from both platforms, the sequencing of food item DNA in the faeces collected from the housebound cat succeeded in 66% of the samples (27 out of 41 samples).

Of the six expected prey species, the wood mouse *Apodemus sp.* was not detected, while at least one of the two HTS platforms sequenced four other rodent species (water vole, field vole, Norway rat and house mouse) and a bird species (Table 1). DNA from the Norway rat and field vole, each expected in one scat, as with the other food items, was recovered in two consecutive faeces. DNA from the house mouse was poorly detected (i.e. only four informative reads from the MiSeq run), and the wood mouse was not detected at all. A delay of 12 to 24 hours between prey ingestion and the occurrence of their remains in the scat was identified for each prey species eaten and detected by both methods.

The proportion of sequences (Ion PGM + MiSeq) associated with the different food item categories did not match the proportion of the food item consumed ($p < 0.001$; Fig. 1a), even when the platforms were considered independently (Fig. 1b and 1c). Prey items were considerably overrepresented: they accounted for 66% of the sequence reads assigned to food items, whereas they represented only 5% of the biomass ingested by the cat. Within the prey item category, rodent and bird biomass (90.4% and 9.6% respectively) were representative of

their read proportions (95.7% and 4.3% respectively, $p > 0.05$). Conversely, human-linked food items (pet food and raw fish or meat) were considerably underrepresented: pet food accounted for 33% and raw fish or meat 0.5% of the sequence reads, whereas these respectively represented 83% and 12% of the biomass ingested.

DNA metabarcoding on field-collected samples

Comparing predator identification using cat-specific PCR amplification and DNA metabarcoding, 2.4% of faecal samples (10/415) identified as cat faeces using cat-specific PCR amplification were found to be emitted by another predator species using DNA metabarcoding; 5% of faecal samples (20/415) not identified as cat faeces using cat-specific PCR amplification were found to be emitted by a cat using DNA metabarcoding. Of the 359 putative felid faecal samples collected, 90.8% (326/359) were confirmed as originating from domestic cats using sequencing. The others were determined to be the dejections of Canids (4.5% from the domestic dog *Canis lupus familiaris* and 3.9% from the red fox,) or Mustelids (0.6% from the Eurasian badger *Meles meles* and 0.3% from *Martes sp.*). All a priori DNA metabarcoding identifications of predators were confirmed unambiguously by both HTS platforms and by the four PCR replicates.

Of the 159 occurrences of prey items detected with Ion PGM, 65% (n = 103) were identified in duplicates compared to 72% (130/180) for MiSeq (Table 2 and Table S2). Of the 115 occurrences of human-linked food items detected with Ion PGM, 39% (n = 45) were identified in duplicates compared to 48% (90/189) for MiSeq (Table 2 and Table S2). The two platforms jointly confirmed 72.2% (143/198) of the occurrences of prey items, and 61% (115/189) of the occurrences of human-linked food (pet food and scraps) (Table 2 and Table S2). In terms of the cat faecal samples, Ion PGM detected at least one food item in 73.6% (240/326) of the investigated samples, and MiSeq detected at least one food item in 83.1%

(271/326) of cases. Combining the results from both HTS platforms, 87.4% (285/326) of the investigated samples contained at least one identified food item.

A likelihood ratio test applied to the model selected for food item detection indicated that the model without random effects was significantly better than the one with random effects ($L = 870.09$, $df = 1$, $p > 0.05$). The top-ranked model included fPlatform and fFoodItem and the interaction between these two variables ($R^2 = 0.17$, Table S4). Ion PGM detected significantly less human-linked food (-2.01 ± 0.31 SE on the log odds ratio scale, $p < 0.001$), birds (-4.98 ± 1.10 SE, $p < 0.001$), field voles (-1.50 ± 0.72 SE, $p < 0.05$), murids (-3.30 ± 1.10 SE, $p < 0.01$) and shrews (-2.60 ± 1.24 SE, $p < 0.05$) than MiSeq.

A total of 521 different occurrences of food items were identified and classified as prey items or human-linked food. A total of 29 different prey taxa were identified, of which 25 were identified without ambiguity up to the species level (9 rodents, 3 insectivores, 12 birds and 1 reptile), representing 90% of the prey items (287/319, Table S2). One reptile item was identified as being the slow worm *Anguis fragilis*. Out of 202 different faecal samples, 478 occurrences related to human-linked food items were detected. Game species were also detected in addition to the expected occurrences of pet food, food leftovers and hunting remains. A total of 33 different taxa related to human-linked food were identified, of which 17 were identified without ambiguity up to the species level (30%, 155/478) and 23 up to the genus level (76%, 364/478, Table S2).

Discussion

Effect of food item categories and their DNA integrity

The use of the short cyt*b* mini-barcode provided a good sensitivity of food item detection, particularly for prey items. Galal-Khallaf and colleagues [18] highlighted the advantage of

amplifying a short mini-barcode when employing a DNA metabarcoding approach on degraded DNA in highly processed food, based on 454 Next Generation Sequencing (NGS) to detect fish species in commercial fish feed samples. However, in our study non-reproducible occurrences between replicates were observed using Ion PGM and MiSeq runs independently (86% for Ion PGM and 56% for MiSeq runs during the calibration study, and 48% for Ion PGM and 40% for MiSeq runs during the field study). These findings emphasize the need to generate sufficient numbers of technical replicates (at least four were required in the present study) to detect rare or degraded DNA, in order to limit the presence of false negatives as well as to remove inconsistent results likely to be false positives [37].

Results from the calibration study revealed that DNA related to processed food is rarely recovered in duplicates and is less detected than the DNA of prey items by both HTS platforms. This result might be linked to a DNA integrity gradient between prey items, pet food and raw fish and meat. Prey items were frozen immediately after death during the calibration study, and in the field study they were probably the result of hunting rather than scavenging behaviour [25], ensuring maximum DNA integrity before ingestion. In addition, in this food category, hard parts (e.g. bones versus flesh or other soft parts) limit DNA degradation during digestion, and so partly solve problems of variable digestibility [7]. In contrast, the stages of food processing (e.g. cooking, dehydration, heat sterilization, etc.) involved in producing pet food induce a higher rate of DNA degradation than for prey items, leading to damaged and degraded DNA [22,43]. However, food processing also suppresses the activity of micro-organisms such as bacteria, so the integrity of the DNA remains the same regardless of the duration of food storage before ingestion. In the case of the category of raw fish and meat, which is not processed, DNA integrity depends on the length of storage and the activity of micro-organisms. In this study, the use of out-of-date raw fish and meat stored in a supermarket refrigerator for several days/weeks might explain the low success of DNA

detection related to this food category. The link between food category and DNA integrity could be confirmed through a calibration study including the evaluation of DNA integrity and a DNA metabarcoding experiment on various food item categories before and after ingestion.

The effect of the HTS platform

In addition to variability in the DNA detection of food items, variations in the ability of the different HTS platforms to detect food items were identified. Overall, during the calibration study, MiSeq was more effective at detecting expected food items and duplicating this (93% of expected items were detected, of which 71% were duplicate-confirmed) than Ion PGM (52% of expected items were detected, of which 29% were duplicate-confirmed). A similar pattern was observed in the analysis of the field-collected faecal samples. Furthermore, MiSeq was also able to detect more human-linked food, as well as species of bird, field vole, murids and shrews than Ion PGM in field collected-samples. This outcome could be most easily explained by the difference in the number of sequence reads assigned to the samples between the two HTS platforms (i.e. low sequencing depth for the Ion PGM platform), linked with the better sequencing quality of the MiSeq platform.

Another strong sequencing bias was noticed in results from the Ion PGM platform, particularly for species haplotypes that included short homopolymer repeats. In our case, a loss of quality at the end of the cyt*b* mini-barcode, and consequently to the presence of several short homopolymer stretches in tandem (e.g. the last 22 bases of the mini-barcode from *Felis silvestris catus*: 5'-CCC TTT AAA TAC CCC TCC CCA T-3') explained a high proportion of truncated sequences without reverse tags. This bias has been previously observed in other public datasets generated by this platform and suggests that it is a pervasive problem [44]. This may also partly explain the bias found between biomass and number of reads by leading to a

lower representation of homopolymer-rich haplotype sequences in the DNA metabarcoding studies using Ion Torrent technology (e.g. Deagle and colleagues [45]).

Common biases in diet analysis using DNA metabarcoding

The proportion of reads from prey items or human-linked food was not related to the ingested biomass of these food categories in the housebound cat diet. On a mass-to-mass basis, prey item DNA was better detected than DNA associated with human-linked food. This result was not related to a taxonomic bias introduced by primer-template mismatches, but rather related to food item properties. This discrepancy is likely to be associated with the variable digestibility of different food items and/or DNA quality (see above). In line with previous studies [45-47], these results reinforce the preference for a qualitative approach that involves creating a list of food items consumed and their frequency of occurrence [12,42] rather than relating reads to ingested biomass, particularly in studies on species with mixed diets. This issue has been highlighted in several recent DNA-based dietary studies investigating the COI barcode region [48] or a more conserved mitochondrial region such as the mitochondrial 16S rRNA gene [49]. Deagle and colleagues [45] also noted a lack of relationship between biomass and the proportion of reads when comparing the proportion of three fish species detected in seal scat to the proportion of these species in a constant diet. An alternative would be to use conserved markers such as those targeting the mitochondrial 12S rRNA or the nuclear 18S rRNA [50], but these provide limited taxonomic resolution for vertebrates compared to markers targeting mitochondrial cytochrome *b*. One way to increase pet food detection in dietary studies would be to target not only vertebrate and invertebrate components of the diet but also cereals contained in pet food. DNA metabarcoding that targets different genes, including those of plants, has been successfully used on an omnivorous carnivore, the brown bear *Ursus arctos* [14]. DNA identified in faeces could also reflect secondary predation cases, e.g. insects consumed by shrews or seeds ingested by rodents [11,51].

Another bias related to an approach based on amplifying a cyt*b* 136-bp mini-barcode without a blocking oligonucleotide on the targeted predator species is that it can lead to an over-representation of the sequences assigned to the predator in the PCR product. As a result, some samples can exclusively yield predator sequences by inhibiting the amplification of rare DNA reads (e.g. prey items). This phenomenon is probably exacerbated because predator DNA has higher integrity than DNA originating from human-linked food. This hypothesis is consistent with the high proportion of cat haplotypes in the present study (80% for Ion PGM and 98% for MiSeq when faecal samples were rapidly stored, and 51% for Ion PGM and 83% for MiSeq in more hazardous sample conservation). Of the field-collected faecal samples, 18% produced only cat sequences. Similarly, in a study by Shehzad and colleagues [15], 21% of faecal samples were found to be entirely composed of leopard cat *Prionailurus bengalensis* DNA when investigating leopard cat diet using metabarcoding without a blocking oligonucleotide on the targeted predator species. The high proportion of predator sequences in the PCR product reduces the sensitivity of HTS platforms. This phenomenon contributes to explaining the relatively low sequencing success rate obtained with Ion PGM, for which the total number of reads was lower than with MiSeq.

In line with D'Amore and colleagues [21], these results indicate that studies using DNA metabarcoding for conservation management purposes should evaluate their outputs in comparison with a calibration study or at least with biological expectations. It should be noted that our calibration study also indicated underestimation (e.g. small murids) and overestimation (e.g. DNA from large species [Norway rat] or numerously ingested prey species [field vole] found in more than one sample) of food items in cat diets; this was linked to molecular bias (e.g. sequencing bias for particular taxa due to homopolymer stretches or primer mismatches) or targeted species biology (i.e. pyloric retention in Felids). Several previous studies also support this pattern [26,52,53].

This study also confirmed that species assignment is subject to errors related to species sequences from the cyt*b* gene submitted to GenBank (e.g. closely related rodents/birds that are not distributed within the study range), and thus potentially poorly allocated. DNA metabarcoding is mainly used to investigate faeces with no previous knowledge about the diet of the targeted species; a reference database containing voucher specimens from the study area would be useful to validate the detection of prey species.

DNA metabarcoding improves results in dietary studies

When dealing with conservation or epidemiological concerns, misidentified faecal samples from non-targeted species within the analysed sample can dramatically affect study conclusions. In the present study, 9% of the field-collected faecal samples were misidentified using morphological criterion. Hence, identifying the faeces origin with species-specific primer pairs and quantitative PCR [15,54] is one alternative. However, the results presented here and elsewhere [42] have shown that a positive detection in real-time PCR may reflect the contamination of samples with a small amount of DNA from targeted species. This can be illustrated through the human contamination usually detected in sequencing runs. In contrast, a positive species-specific PCR can partly reflect cross-contamination of samples or field contamination. Combining several replicates of PCR amplification with and without a blocking oligonucleotide would increase food item detection sensitivity per sample, particularly for small samples or when targeting the DNA of rare prey species. However, the number of primer mismatches and the concentration of blocking oligonucleotide can affect the detection of target species [48].

DNA metabarcoding allowed the identification of 90% of prey items up to species level. It led to both a high detection rate and the high-resolution identification of closely related prey species (e.g. *M. arvalis* versus *M. agrestis*, *A. flavicollis* versus *A. sylvaticus*, Table S2). Using

a DNA sequence database and bird-specific primer pairs to analyse faecal samples, Zarzoso-Lacoste and colleagues [55] highlighted similar efficiency in using molecular ecology for the identification of bird species in cat and rat diets. Ordinarily, this level of precision is achieved only in cat diet studies that rely on prey animals brought back to the home (e.g. 92% identified to at least class level [56]), not with conventional methods analysing scat and stomach contents (e.g. from 25% [57] to 59% [58]).

The results of this study demonstrate that DNA metabarcoding has the great advantage of allowing the detection of occurrences of food items that leave undetectable remains in the field (such as human-related food items) in a non-invasive sampling approach. Even unprecedented, this method gives insights on the potential origin of ingested food items (e.g., pet food or scavenged human-processed food versus scavenged carrion), with a taxonomic identification of 76% of vertebrate cyt*b* variants to the genus level. This makes this method useful for evaluating the exploitation of human-linked foods in urban wildlife populations, above all in endangered species, for which it is preferable to use non-invasive survey approaches. Nevertheless, the results of the calibration study make it clear that the rate of false negatives is partly linked to DNA integrity (and so to food item categories). This finding suggests that even if DNA metabarcoding allows more efficient detection of specific food items, certain items in the diet may remain highly underestimated (e.g. 46% for pet food and 70% for out-of-date raw fish and meat), which could have a significant impact on conservation strategies. It should be noted, however, that sequence proportions matched biomass within a food category (e.g. rodents and birds).

To conclude, the present study joins several metabarcoding studies considering the biases and limitations of this approach, particularly in terms of using benchtop HTS platforms. In line with Piñol and colleagues [48], our findings indicate that universal primers bias the detection of degraded DNA in a heterogeneous mixture. This can drastically affect the identification of the

ingested taxa for species having a diet possibly containing human-linked food items, and thus the inferences drawn about this diet regarding the impacts of human activities on wildlife populations and food-web dynamics in anthropogenic landscapes. Nevertheless, generating sufficient numbers of technical replicates and using a sequencer such as MiSeq to produce a high quantity and quality of reads appears to be a robust approach for investigating the exploitation of a large range of food items for populations relying on a mixed-diet through non-invasive sampling. Indeed, compared to conventional approaches based on morphological analysis, DNA metabarcoding allows the identification of a large number of food items up to species level and a wide range of taxa.

Acknowledgements

We are grateful to the people in the study area who allowed us to prospect for and collect samples, to the cat owners who collected samples for the study, to Matthieu Bastien, Ivan Puga Gonzales, Nancy Rebout and Claire Saraux for their statistical advice, and to Vincent Viblanc for its constructive comments. We also sincerely thank Spécialités Pet Food (SPF) for funding the DNA extraction and the two HTS runs.

Data accessibility

Should the manuscript be accepted, the data supporting the results (raw sequence reads (fastq format), raw table of abundance and filtered table of occurrence including taxonomic affiliations) will be published along with the manuscript in a Dryad Digital Repository.

Author contributions

MAFW and MLP contributed to the study design and sample collection. MG conducted the molecular analysis. MAFW, MG and SP conducted the bioinformatics analysis. MAFW conducted the data analysis, performed the statistical analysis and prepared the tables and

graphs. MAFW, MLP and MG led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

References

1 McCleery, R. A., Moorman, C. E. & Peterson, M. N. *Urban wildlife conservation: theory and practice*. (Springer, 2014).

2 Murray, M. *et al.* Greater consumption of protein□poor anthropogenic food by urban relative to rural coyotes increases diet breadth and potential for human-wildlife conflict. *Ecography* **38**, 1235-1242 (2015).

3 Robardet, E. *et al.* Infection of foxes by *Echinococcocus multilocularis* in urban and suburban areas of Nancy, France: influence of feeding habits and environment. *Parasite* **15**, 77-85 (2008).

4 Newsome, S. D., Ralls, K., Job, C. V. H., Fogel, M. L. & Cypher, B. L. Stable isotopes evaluate exploitation of anthropogenic foods by the endangered San Joaquin kit fox (*Vulpes macrotis mutica*). *J. Mammal.* **91**, 1313-1321, doi:10.1644/09-mamm-a-362.1 (2010).

5 Bateman, P. W., Fleming, P. A. & Le Comber, S. Big city life: carnivores in urban environments. *J. Zool.* **287**, 1-23, doi:10.1111/j.1469-7998.2011.00887.x (2012).

6 Pompanon, F. *et al.* Who is eating what: Diet assessment using next generation sequencing. *Mol. Ecol.* **21**, 1931-1950 (2012).

7 Yoccoz, N. G. The future of environmental DNA in ecology. *Mol. Ecol.* **21**, 2031-2038 (2012).

8 Egeter, B., Bishop, P. J. & Robertson, B. C. Detecting frogs as prey in the diets of introduced mammals: a comparison between morphological and DNA-based diet analyses. *Molecular Ecology Resources* **15**, 306-316, doi:10.1111/1755-0998.12309 (2015).

9 Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045-2050 (2012).

10 Shokralla, S., Spall, J. L., Gibson, J. F. & Hajibabaei, M. Next□generation sequencing technologies for environmental DNA research. *Mol. Ecol.* **21**, 1794-1805 (2012).

11 Galan, M., Pagès, M. & Cosson, J.-F. Next-generation sequencing for rodent barcoding: species identification from fresh, degraded and environmental samples. *PLoS ONE* **7**, e48374, doi:10.1371/journal.pone.0048374 (2012).

12 Valentini, A., Pompanon, F. & Taberlet, P. DNA barcoding for ecologists. *Trends Ecol. Evol.* **24**, 110-117, doi:10.1016/j.tree.2008.09.011 (2009).

13 Coghlan, M. L. *et al.* Metabarcoding avian diets at airports: Implications for birdstrike hazard management planning. *Investigative Genetics* **4** (2013).

14 De Barba, M. *et al.* DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources* **14**, 306-323 (2014).

15 Shehzad, W. *et al.* Carnivore diet analysis based on next-generation sequencing: Application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Mol. Ecol.* **21**, 1951-1965 (2012).

16 Brown, D. S., Ebenezer, K. L. & Symondson, W. O. C. Molecular analysis of the diets of snakes: Changes in prey exploitation during development of the rare smooth snake *Coronella austriaca*. *Mol. Ecol.* **23**, 3734-3743 (2014).

17 Brown, D. S., Jarman, S. N. & Symondson, W. O. C. Pyrosequencing of prey DNA in reptile faeces: Analysis of earthworm consumption by slow worms. *Molecular Ecology Resources* **12**, 259-266 (2012).

18 Galal-Khallaf, A., Osman, A. G., Carleos, C. E., Garcia-Vazquez, E. & Borrell, Y. J. A case study for assessing fish traceability in Egyptian aquafeed formulations using pyrosequencing and metabarcoding. *Fisheries Research* **174**, 143-150 (2016).

19 Galan, M. *et al.* 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife: the importance of cleaning post-sequencing data before estimating positivity, prevalence and co-infection. *bioRxiv*, 039826 (2016).

20 Ficetola, G. F. *et al.* Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* **15**, 543-556 (2015).

21 D'Amore, R. *et al.* A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics* **17**, 55 (2016).

22 Sharma, N., Thind, S. & Sharma, D. Effect of meat processing on genomic DNA quality and specific gene amplification. *Journal of Applied Animal Research* **28**, 69-72 (2005).

23 Coissac, E., Riaz, T. & Puillandre, N. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Mol. Ecol.* **21**, 1834-1847, doi:10.1111/j.1365-294X.2012.05550.x (2012).

24 Fitzgerald, B. M. & Turner, D. C. in *The domestic cats: The Biology of its Behaviour* (eds D.C. Turner & P. Bateson) 151-175 (Cambridge Univ. Press, 2000).

25 Spotte, S. in *Free-ranging cats: Behavior, ecology, management* (ed Wiley Blackwell) Ch. 9, 181-213 (John Wiley & Sons, 2014).

26 Forin-Wiart, M.-A., Gotteland, C., Gilot-Fromont, E. & Poulle, M.-L. Assessing the homogeneity of individual scat detection probability using the bait-marking method on a monitored free-ranging carnivore population. *European Journal of Wildlife Research* **60**, 665-672, doi:10.1007/s10344-014-0833-0 (2014).

27 Corbett, L. K. *Feeding ecology and seasonal organization of wildcats (Felis silvestris) and domestic cats (Felis catus) in Scotland.*, PhD thesis, University of Aberdeen, Scotland, (1979).

28 Lopez, J. V., Cevario, S. & O'Brien, S. J. Complete Nucleotide Sequences of the Domestic Cat (Felis catus) Mitochondrial Genome and a Transposed mtDNA Tandem Repeat (Numt) in the Nuclear Genome. *Genomics* **33**, 229-246, doi:http://dx.doi.org/10.1006/geno.1996.0188 (1996).

29 Kappel, K., Haase, I., Käppel, C., Sotelo, C. G. & Schröder, U. Species identification in mixed tuna samples with next-generation sequencing targeting two short cytochrome b gene fragments. *Food Chem.* **234**, 212-219 (2017).

30 Teletchea, F., Bernillon, J., Duffraisse, M., Laudet, V. & Hänni, C. Molecular identification of vertebrate species by oligonucleotide microarray in food and forensic samples. *J. Appl. Ecol.* **45**, 967-975 (2008).

31 Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112-5120 (2013).

32 Galan, M. *et al.* Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. *Molecular Ecology Resources*, 1-16, doi:10.1111/1755-0998.12749 (2017).

33 Fadrosh, D. *et al.* An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**, 6 (2014).

34 Schloss, P. D. *et al.* Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial

Communities. *Appl. Environ. Microbiol.* **75**, 7537-7541, doi:10.1128/aem.01541-09 (2009).

35 Piry, S., Guivier, E., Realini, A. & Martin, J. F. | SE| S| AM| E| Barcode: NGS-oriented software for amplicon characterization–application to species and environmental barcoding. *Molecular Ecology Resources* **12**, 1151-1157 (2012).

36 Esling, P., Lejzerowicz, F. & Pawlowski, J. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res.* **43**, 2513-2524 (2015).

37 Ficetola, G. F., Taberlet, P. & Coissac, E. How to limit false positives in environmental DNA and metabarcoding? *Molecular ecology resources* **16**, 604-607 (2016).

38 Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A. & Smith, G. M. *Mixed effects models and extensions in ecology with R.* (Springer, 2009).

39 R: A language and environment for statistical computing. (Vienna, Austria, 2016).

40 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* (2014).

41 Lenth, R. V. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software* **69**, 1-33, doi:10.18637/jss.v069.i01 (2016).

42 Deagle, B. E., Kirkwood, R. & Jarman, S. N. 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 (2009).

43 De Battisti, C. *et al.* Pyrosequencing as a tool for rapid fish species identification and commercial fraud detection. *J. Agric. Food Chem.* **62**, 198-205 (2013).

44 Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology* **30**, 434-439 (2012).

45 Deagle, B. E., Thomas, A. C., Shaffer, A. K., Trites, A. W. & Jarman, S. N. Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: Which counts count? *Molecular Ecology Resources* **13**, 620-633 (2013).

46 Egge, E. *et al.* 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PloS one* **8**, e74371 (2013).

47 Pochon, X., Bott, N. J., Smith, K. F. & Wood, S. A. Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS One* **8**, e73935 (2013).

48 Piñol, J., Mir, G., Gomez-Polo, P. & Agustí, N. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources* **15**, 819-830, doi:10.1111/1755-0998.12355 (2015).

49 Deagle, B. E., Chiaradia, A., McInnes, J. & Jarman, S. N. Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conserv. Genet.* **11**, 2039-2048 (2010).

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

51 Sheppard, S. K. *et al.* Detection of secondary predation by PCR analyses of the gut contents of invertebrate generalist predators. *Mol. Ecol.* **14**, 4461-4468 (2005).

52 Krauze-Gryz, D., Gryz, J. & Goszczyński, J. Predation by domestic cats in rural areas of central Poland: an assessment based on two methods. *J. Zool.* **288**, 260-266, doi:10.1111/j.1469-7998.2012.00950.x (2012).

53 Pires, M. M., Widmer, C. E., Silva, C. & Setz, E. Z. F. Differential detectability of rodents and birds in scats of ocelots, *Leopardus pardalis* (Mammalia: Felidae). *Zoologia (Curitiba)* **28**, 280-283 (2011).

54 Knapp, J., Umhang, G., Poulle, M.-L. & Millon, L. Development of a real-time PCR for a sensitive one-step coprodiagnosis allowing both the identification of carnivore feces and the detection of Toxocara spp. and Echinococcus multilocularis. *Applied and environmental microbiology* **82**, 2950-2958 (2016).

55 Zarzoso-Lacoste, D. *et al*. Improving morphological diet studies with molecular ecology: An application for invasive mammal predation on island birds. *Biol. Conserv.* **193**, 134-142, doi:10.1016/j.biocon.2015.11.018 (2016).

56 Kauhala, K., Talvitie, K. & Vuorisalo, T. Free-ranging house cats in urban and rural areas in the north: useful rodent killers or harmful bird predators? *Folia Zoologica* **64** (2015).

57 Germain, E., Ruette, S. & Poulle, M.-L. Likeness between the food habits of European wildcats, domestic cats and their hybrids in France. *Mammalian Biology* **74**, 412-417 (2009).

58 Biró, Z., Lanszki, J., Szemethy, L., Heltai, M. & Randi, E. Feeding habits of feral domestic cats (*Felis catus*), wild cats (*Felis silvestris*) and their hybrids: trophic niche overlap among cat groups in Hungary. *Journal of Zoology (London)* **266**, 187-196 (2005).

Figures and tables

Table 1 The minimum expected occurrences and the actual observed occurrences of food items in HTS runs of 41 faecal samples collected from a housebound cat. The observed occurrences are shown according to the HTS platform used, the PCR replicates and the pooled occurrences from both platforms.

Food category	**Food items**	**Minimum expected occurrence**	**Observed occurrence in the Ion PGM run**			**Observed occurrence in the MiSeq run**			**Observed occurrences (Ion PGM & MiSeq)**
			in one replicate	in duplicates	Total	in one replicate	in duplicates	Total	
Prey items	Water vole *Arvicola terrestris*	3	0	2	2	0	2	2	2
	Field vole *Microtus agrestis*	1	1	1	2	0	1	1	2
	Norway rat *Rattus norvegicus*	1	0	0	0	1	1	2	2
	House mouse *Mus musculus*	2	0	0	0	0	1	1	1
	Wood mouse *Apodemus sp.*	2	0	0	0	0	0	0	0
	Chestnut-eared finch *Taeniopygia guttata*	2	0	0	0	0	2	2	2
	Total prey items	11	1	3	4	1	7	8	9
Pet food*		37	8	1	9	8	10	18	20
Out-of-date raw fish or meat**		13	1	0	1	1	2	4	4
TOTAL		**61**	**10**	**4**	**14**	**10**	**20**	**30**	**33**

***** Pet food taxa recovered (Suidae, *Sus sp.*, *Bos sp.*, *Ovis sp.*, Caprie, *Gallus sp.*, *Anser sp.*, *Meleagris gallopavo*, *Numida meleagris*, *Salmo salar*).

** Out-of-date raw meat taxa recovered (*Sus scrofa*, *Ovis aries*, *Cairina moschata*).

Table 2 Observed occurrences of food items in 285 field-collected faecal samples from free-ranging cats. The observed occurrences are shown according to the HTS platform used and the PCR replicates. The final observed occurrence corresponds to the pooled results from the Ion PGM and MiSeq runs. See S1 Table for more details about food sub-categories.

Food category	**Food sub-category**	**Observed occurrence in the Ion PGM run**			**Observed occurrence in the MiSeq run**			**Observed occurrence**
		in one replicate	in duplicates	Total	in one replicate	in duplicates	Total	
Prey items	Small/medium-sized rodents	88	133	221	119	111	230	271
	Insectivores	1	3	4	7	2	9	10
	Wild birds	3	3	6	22	13	35	37
	Reptiles	0	0	0	1	0	1	1
	Total prey items	92	139	231	149	126	275	319
Human-linked food	Game species	4	1	5	8	6	14	15
	Anthropogenic food	117	53	169	289	151	439	187
	Total human-linked food	121	54	174	297	157	453	202
	TOTAL	**213**	**193**	**405**	**446**	**283**	**728**	**521**
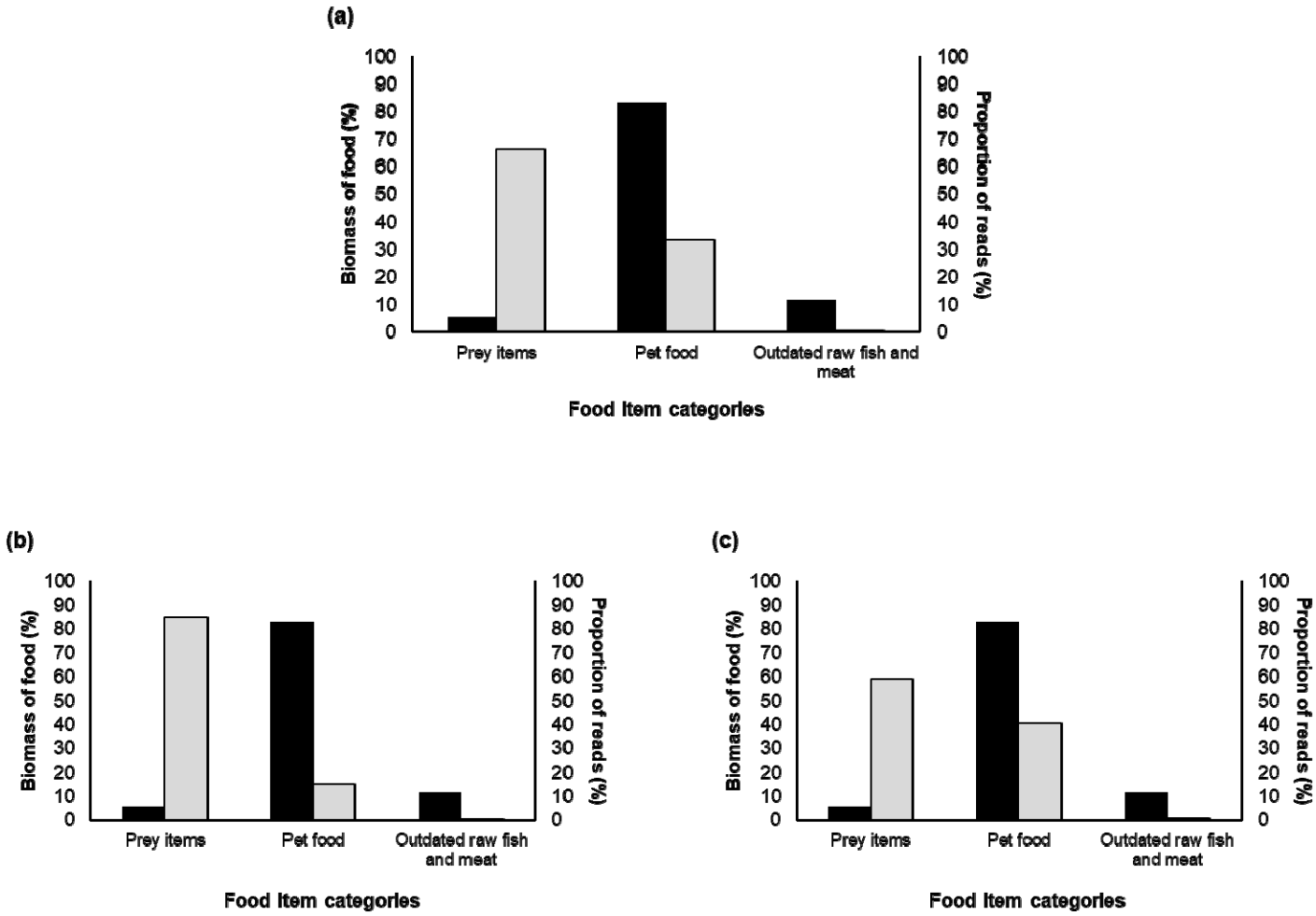
Figure 1 Comparison between the proportion of biomass of the three food categories fed to the housebound cat (black) versus the overall proportion of sequence reads assigned to that food category (light grey) from 28 scat samples for (a) pooled replicates (Ion PGM + MiSeq runs), (b) Ion PGM replicates only, and (c) MiSeq replicates only.