

## INVITED REVIEW

# Who is eating what: diet assessment using next generation sequencing

FRANCOIS POMPANON,<sup>\*1</sup> BRUCE E. DEAGLE,<sup>†‡1</sup> WILLIAM O. C. SYMONDSON,<sup>§1</sup> DAVID S. BROWN,<sup>§</sup> SIMON N. JARMAN<sup>†</sup> and PIERRE TABERLET<sup>\*</sup>

*\*Université Grenoble 1, Laboratoire d'Ecologie Alpine, CNRS, UMR 5553, 2233 Rue de la Piscine, 38041 Grenoble Cedex 9, France, †Australian Antarctic Division, Channel Highway, Kingston, Tas. 7050, Australia, ‡Department of Biology, University of Victoria, Victoria, BC V8W 3N5, Canada, §Cardiff School of Biosciences, Cardiff University, Biomedical Sciences Building, Museum Avenue, Cardiff CF10 3AX, UK*

## Abstract

The analysis of food webs and their dynamics facilitates understanding of the mechanistic processes behind community ecology and ecosystem functions. Having accurate techniques for determining dietary ranges and components is critical for this endeavour. While visual analyses and early molecular approaches are highly labour intensive and often lack resolution, recent DNA-based approaches potentially provide more accurate methods for dietary studies. A suite of approaches have been used based on the identification of consumed species by characterization of DNA present in gut or faecal samples. In one approach, a standardized DNA region (DNA barcode) is PCR amplified, amplicons are sequenced and then compared to a reference database for identification. Initially, this involved sequencing clones from PCR products, and studies were limited in scale because of the costs and effort required. The recent development of next generation sequencing (NGS) has made this approach much more powerful, by allowing the direct characterization of dozens of samples with several thousand sequences per PCR product, and has the potential to reveal many consumed species simultaneously (DNA metabarcoding). Continual improvement of NGS technologies, on-going decreases in costs and current massive expansion of reference databases make this approach promising. Here we review the power and pitfalls of NGS diet methods. We present the critical factors to take into account when choosing or designing a suitable barcode. Then, we consider both technical and analytical aspects of NGS diet studies. Finally, we discuss the validation of data accuracy including the viability of producing quantitative data.

*Keywords:* DNA barcoding, DNA metabarcoding, faeces, food webs, herbivory, predation

*Received 7 September 2011; revision received 4 November 2011; accepted 9 November 2011*

## Introduction

The diversity of life on Earth and the abundance of each species are in large part a product of the evolution of strategies for survival. A major part in this struggle for existence is played by predator–prey, host–parasite and herbivore–plant interactions, between which multi-dimensional webs of interactions have arisen. The complexity of these food webs can be considerable even at

the simplest level of connectance, while analysis of fully quantitative webs involving generalist species has so far been beyond our reach. Perturbation in one part of a food web has ramifications throughout the web and change outcomes in ways that become more difficult to predict as complexity increases (Pimm 2002). However, such predictions are essential if we are to analyse, and determine the mechanistic processes behind, major areas of community ecology and hence better understand ecosystem function (e.g. Estes *et al.* 2011). Food web dynamics are important in areas of ecology as diverse as conservation biology, agroecology, penetration of ecosystems by alien species, the effects of

Correspondence: François Pompanon, Fax: +33 4 76 51 42 79;

E-mail: francois.pompanon@ujf-grenoble.fr

<sup>1</sup>These authors contributed equally.

biodiversity on ecosystem function, and energy flows within and between ecosystems.

Understanding food webs ultimately requires reconstruction and modelling of the overall population interactions of the species involved. An important part of this process is analysis of precisely what has been eaten in the field, information that is difficult to obtain for generalist predators and herbivores. Thus, we need accurate techniques for determining the identities and proportions of different food items in a species diet. The application of DNA-based techniques has improved our ability to determine diets, and the range of DNA-based approaches used has been extensively considered elsewhere (Symondson 2002; Sheppard & Harwood 2005; Sunderland *et al.* 2005; Gariépy *et al.* 2007; King *et al.* 2008a). Our aim here is not to cover the diversity of methods in detail but to review recent developments based on next generation sequencing (NGS) technologies that have provided us with more precise information on the dietary ranges of predators and herbivores. We also present the critical steps necessary for setting up a NGS diet assessment, examine the utility of NGS as a means of improving the speed, accuracy and quantitative potential that this technology makes theoretically possible and also discuss its limitations.

### From behavioural observations to NGS diet assessment

#### *Visual observations*

Initially, diets were determined mainly by direct observation of feeding, or by microscopic examination of gut contents or faeces. While these approaches have provided much useful information in some cases, direct observation precludes working on small invertebrates, most nocturnal species, anything beneath the soil, under water, hidden or elusive. Microscopic examination also has major limitations. It is labour intensive and depends heavily upon the skills of the person identifying species from masticated, semi-digested fragments of plants and animals (Holechek *et al.* 1982; Moreby 1988; Ingerson-Mahar 2002; Sunderland *et al.* 2005). Moreover, it precludes studying the diet of fluid feeders (most invertebrates) and identifying foods that leave no hard remains or simply lack diagnostic taxonomic features.

#### *Molecular approaches*

To overcome these difficulties, ecologists have innovatively adapted a variety of technologies. A suite of techniques based upon prey-specific antibodies have been used to analyse gut contents with some success

(reviewed by Boreham & Ohiagu 1978; Symondson 2002). In most cases, these tests were developed as markers for specific target prey species and were best utilized for screening a range of predators consuming a small number of target species.

However, to understand food web dynamics, the whole dietary breadth needs to be measured. Attempts to measure consumption of the whole array of prey in the guts of predators were initially made using a protein electrophoretic approach (e.g. Walrant & Loreau 1995). This worked reasonably well for stenophagous predators (Murray *et al.* 1989; Solomon *et al.* 1996) but not for highly generalist predators for which an individual may have several prey species in the gut simultaneously resulting in an uninterpretable superimposed protein banding pattern. Similar problems are encountered when interpreting plant alkane fingerprints to determine herbivore diets (Dove & Mayes 1996). Plant material in the diet has also been analysed using near infrared reflectance spectroscopy (NIRS) (Kaneko & Lawler 2006; Rothman *et al.* 2009). Calibration of the technique for different plant material and consumers is critical and obtaining constituent identities for complex mixtures of plants difficult, but this approach is best suited for estimating nutritional components such as total nitrogen or starch (Foley *et al.* 1998).

Studies using stable isotopes can tell us a lot about energy flow through (e.g. Ponsard & Arditi 2000) and between ecosystems (e.g. Gratton *et al.* 2008), or simply dietary shifts (e.g. Ostrom *et al.* 1997). Such work provides information on long-term diets, rather than simply components within a recent meal (reviewed by Sunderland *et al.* 2005). However, isotopic enrichment is subject to many variables (Vanderklift & Ponsard 2003; Lecomte *et al.* 2011), and the isotopic signatures of different food items are often not great enough, or consistent enough, for clear resolution of trophic links (Traugott *et al.* 2007). This approach that has the strong advantage of being integrative can be replaced or complemented by DNA-based methods when a precise identification of the taxa consumed in the last meals is needed.

An early DNA approach to look at dietary diversity was DNA profiling through amplification of the gut contents using general or group-specific primers, followed by temperature or denaturing gradient gel electrophoresis (TGGE or DGGE) (Deagle *et al.* 2005; Harper *et al.* 2006). Problems with cryptic bands and haplotypic variation (Lessa & Applebaum 1993) make interpretation difficult. However, where diversity in the diet per se is the object, rather than component identity, this approach has merit, as it has in the study of microbial community diversity (e.g. Muyzer *et al.* 1993; Felske *et al.* 1998). DGGE/TGGE had also been used as

a separation tool, followed by sequencing (Martin *et al.* 2006; Tollit *et al.* 2009), but the technique could show significant amplification bias, and more robust approaches have been developed.

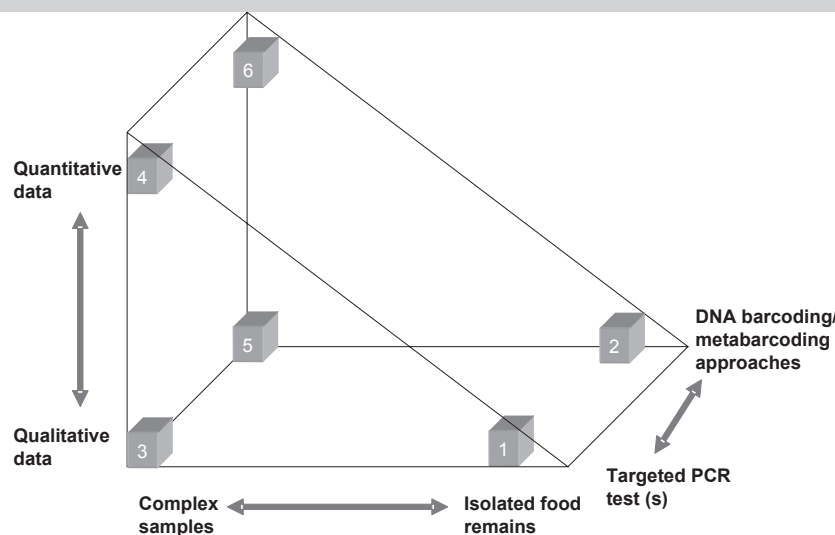
#### Recent DNA-based approaches

The last 12 years have witnessed the development of two main strands in the field of DNA-based dietary analysis, both of which use PCR to amplify DNA present in dietary samples to obtain enough material for subsequent analyses. One of these is a targeted approach using specific PCR primers designed to

examine predation on one or a few specific prey or prey groups (see Box 1; reviewed by King *et al.* 2008a). It was first used by Asahida *et al.* (1997) to examine predation by sand shrimps (*Crangon affinis*) on stone flounder (*Kareius bicoloratus*). This approach has proved to be useful for studying predator-prey systems as divergent as whale sharks feeding on larval crabs (Meekan *et al.* 2009) down to Collembola feeding on nematodes (Read *et al.* 2006). Efforts have also been made to combine many taxon-specific fluorescence-labelled primers to look at the range of prey consumed through multiplex PCR followed by fragment analysis (Harper *et al.* 2005; King *et al.* 2010a,b).

#### Box 1 Place of next generation sequencing analyses among DNA-based methods for dietary studies

The potential for DNA to be used as a dietary biomarker was first highlighted 20 years ago (Höss *et al.* 1992) and since then a range of DNA-based approaches have been applied to identify food remains in stomach contents and faeces. These studies can be categorized based on: (i) type of sample analysed (ii) whether diagnostic PCR or DNA-barcoding (i.e. sequencing) is used for identification and (iii) whether quantitative data is sought from individual samples [Fig. 1; following Darling & Blum (2007)].



**Fig. 1** Classification of DNA-based diet studies. In categories 1 and 2 DNA is extracted from individual food remains followed by either: species identification using diagnostic PCR tests [e.g. species-level PCR identification of salmon bones from seal faeces (Kvitrud *et al.* 2005)]; or direct sequencing of PCR amplicons [e.g. COI barcoding of insect fragments from bat faeces (Clare *et al.* 2009)]. In categories 3 and 4 DNA is extracted from unsorted dietary samples followed by either: presence/absence PCR tests [e.g. DNA isolated from whole spiders and tested with species-specific PCR for their Collembola prey (Agustí *et al.* 2003)]; or quantitative PCR [e.g. quantitative feeding estimates of copepods eating algae (Troedsson *et al.* 2009)]. Finally, in the metabarcoding approach employed in categories 5 and 6 DNA is extracted from unsorted dietary samples followed by cloning and sequencing of PCR amplicons [e.g. identification of plants eaten by primates by cloning of chloroplast DNA from their faeces (Bradley *et al.* 2007)]. In dietary DNA metabarcoding the distinction between quantitative and non-quantitative data is somewhat arbitrary.

High-throughput sequencing has substantially enhanced diet studies using the DNA barcoding approach (Fig. 1; categories 5 and 6) by removing the need to clone PCR products, and by drastically reducing the unit cost of sequencing. This has transformed dietary DNA barcoding from being an approach limited in scale, due to costs and technical difficulty, into an approach that provides a wealth of data from each sample and is cost-effective for large diet studies. One of the consequences of having large dataset with counts of sequences from food items is the potentially misleading ability to infer a quantitative picture of diet (see main text for further discussion).

However, even with multiplexing only the limited range of prey for which the primers were designed can be detected.

The second strand has been based upon the amplification of DNA using general or group-specific primers, followed by the cloning and sequencing of amplicons to identify individual taxa (see Fig. 1 and Box 1). This dietary barcoding approach has been used to analyse the diets of extinct mammals from coprolites (Poinar *et al.* 1998, 2001; Hofreiter *et al.* 2000, 2003), the stomach contents of an ancient human (Rollo *et al.* 2002), and in several contemporary ecological studies that analysed predation and herbivory. They include diet analysis of passerine birds (Sutherland 2000), whales (Jarman *et al.* 2004), marine amphipods and bivalves (Blankenship & Yayanos 2005), penguins (Deagle *et al.* 2007), dolphins (Dunshea 2009), bats (Zeale *et al.* 2011), primates (Bradley *et al.* 2007) and krill (Passmore *et al.* 2006). The cloning step can be skipped in studies where DNA can be extracted from particulate remains in faeces and directly sequenced (Clare *et al.* 2009, 2011). The cloning approach can detect any dietary components by using suitably general primers, without any need to predict target prey species. However, this is a labour-intensive approach, requiring sequencing of many clones. Thus, it is not well suited to the mass screening that is generally required to obtain a comprehensive analysis of animal diets.

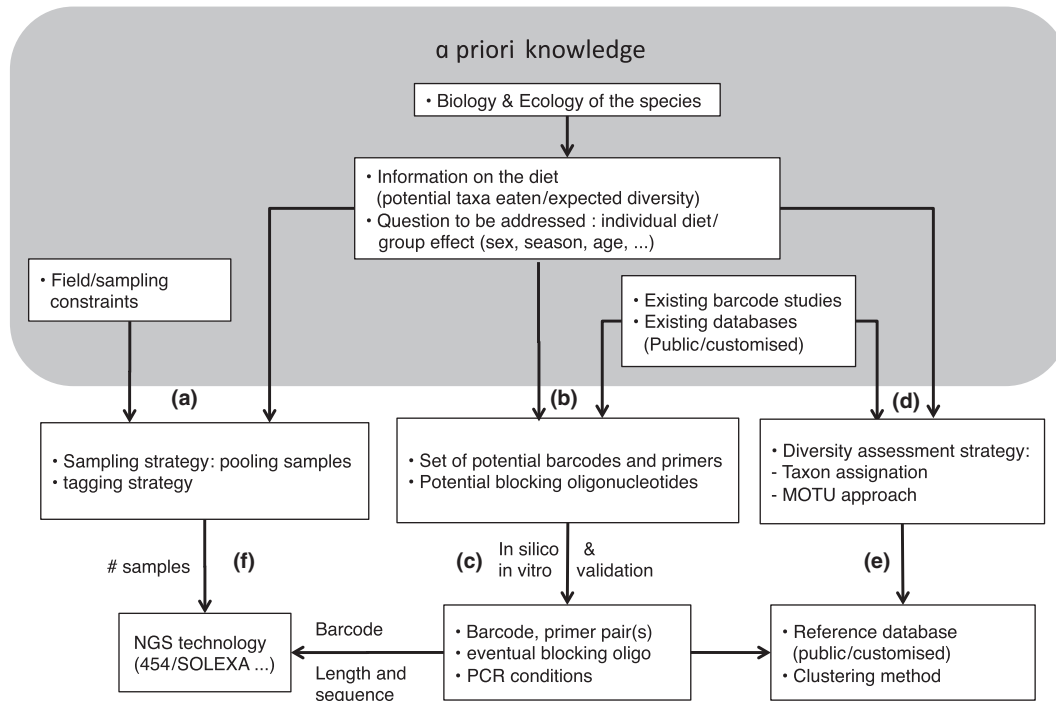
The information arising from clone sequencing can now be obtained much more easily with the recent explosion in the development of NGS technologies (e.g. Shendure & Ji 2008). As the technology has spread and costs reduced, ecologists have been quick to exploit this and turn it into a powerful new tool for ecological studies including dietary analysis (Valentini *et al.* 2009b). At the same time, there has been a massive expansion of sequence databases through the DNA barcoding of animals and plants (Hebert *et al.* 2003; Chase *et al.* 2007; Taberlet *et al.* 2007), without which accurate interpretation of the thousands of sequences generated by NGS would be impossible (Valentini *et al.* 2009a,b). Thus, the potential of NGS for characterizing simultaneously many species from an environmental sample through sequencing of DNA barcodes (DNA metabarcoding approach) is clearly enormous. The first papers on NGS diet assessment dealt with predation by Australian fur seals (Deagle *et al.* 2009), little penguins (Deagle *et al.* 2010; Murray *et al.* 2011), bats (Bohmann *et al.* 2011) and slow worms (Brown 2011), and herbivory by a broad range of taxa including birds, molluscs, insects, and mammals such as bears, marmots or voles (Soininen *et al.* 2009; Valentini *et al.* 2009a), as well as domestic and wild ruminants (Pegard *et al.* 2009; Kowalczyk *et al.* 2011; Rayé *et al.* 2011).

## What barcode to choose?

### *Coverage and resolution*

Because of the constraints linked to the simultaneous characterization of broad ranges of taxa from environmental samples, the DNA regions defined as standard by the Consortium for the Barcoding of Life (CBoL, <http://www.barcodeoflife.org/>) are not always the best suited, leading to the choice of other DNA barcodes (*sensu lato*, Valentini *et al.* 2009b) and primers to amplify them for reliable NGS diet analyses. This choice should be determined by the question being addressed and knowledge of the ecology and biology of the study species (see Fig. 2). The range of likely consumed species that can have their DNA amplified by the barcode primers (i.e. the taxonomic coverage) is the primary consideration, so studies of herbivorous animals, for example, require barcoding primers that amplify DNA from a wide range of plants (e.g. Valentini *et al.* 2009a). The taxonomic resolution (i.e. resolution capacity) of the barcode region also needs to be considered as the saturation of the marker should be avoided, and some barcode regions will only identify taxa above the species level. Looking for highly conserved primers to increase the taxonomic coverage may favour less variable DNA regions with, concomitantly, a low-resolution capacity, which may not be sufficient for answering all dietary questions. For example, primer sets with binding sites that are invariant among eukaryotes (e.g. LSU rDNA) will often amplify less variable regions than those found in mitochondria or chloroplasts, although some have good enough resolution for answering many questions (e.g. Sonnenberg *et al.* 2007). Conversely, designing short broad coverage barcodes in variable regions such as COI used for animal barcoding may be difficult, because potential priming sites would not be conserved enough over a broad range of taxa. This is illustrated by the Uni-minibar barcode targeting COI (Meusnier *et al.* 2008) evaluated for vertebrates. About 60% of the taxa potentially amplified differ at each priming site by 4–8 mismatches, probably leading to a failure or a low rate of amplification (Ficetola *et al.* 2010).

It may be useful to employ a DNA barcode with a broad taxonomic coverage but low resolution, combined with other DNA barcodes that resolve some of the higher taxonomic units to species level, a strategy sometimes called hierarchical barcoding (Moszczyńska *et al.* 2009). For example, in a study of seal diet, Deagle *et al.* (2009) added to the information provided by a low-resolution primer set (which identified bilaterian animals to family), by using group-specific primer sets that resolved key groups (i.e. bony fish and squid) to the



**Fig. 2** Setting up a next generation sequencing (NGS) diet assessment. The flow chart starts by emphasizing the importance of information on the biology and ecology of the species for designing the study. Letters (a) to (f) point out the key steps of the design: (a) field constraints, as well as decisions regarding whether or not to produce individual dietary data according to the question addressed, determines the pooling and tagging strategies that might be used to sort sequence reads. (b) The potential barcode and primers are defined according to (i) the species to identify (i.e. the range of potentially eaten species), (ii) the species to prevent from amplification (eventual blocking oligonucleotides) and (iii) a priori knowledge of the performance of existing barcodes. (c) *In silico* and then *in vitro* validation of potential barcodes will allow selection of the final set of primers and oligonucleotides. (d) The strategy used for diversity assessment (taxon assignment vs. clustering for designing anonymous MOTUs) depends on the existence of reference data sets and on the questions addressed, that may not always require the identification of food items (e.g. study of shifts in diets based on shared MOTUs). (e) According to the strategy chosen, a clustering method may be chosen and/or a reference database may be set up based on reference sequences already existing and/or produced for the study (public vs. customized databases). (f) The number of samples studied as well as the barcode length and sequence (more or less error prone according to the sequencing method) contribute to the choice of the NGS technology.

species level. This provided a general overview of seal diet while allowing species level identification of the most important prey.

### Amplification efficiency

The amplification efficiency of a DNA barcoding primer set is also important in ensuring that most dietary samples analysed will produce sequences. Even when primers work on target DNA templates in isolation, amplification can be highly skewed within a multitemplate PCR. For example, in a study on penguin diet, Deagle *et al.* (2007) found that krill was the prevalent prey item (based on stomach contents and species-specific PCR tests carried out on faecal DNA); however, amplicons generated from faeces using primers that were relatively conserved among prey species recovered only fish and squid sequences. In this case, a minor primer mismatch in the krill sequence led to

extensive under-representation of krill as squid and fish were preferentially amplified. Sequence variation in primer-binding sites such as this can strongly bias the pool of sequences generated in a PCR. In some situations, this variation in the primer-binding site may be used to help reduce predator DNA amplification (Deagle *et al.* 2007), but this may lead to a bias in representation of food species proportions.

DNA degradation in dietary samples limits the length of fragments that can be successfully amplified by PCR (Symondson 2002; Deagle *et al.* 2006; Troedsson *et al.* 2009). For this reason, the length of barcoding regions used for dietary analysis is generally in the 100–250 bp range, which inevitably reduces taxonomic resolution. It is possible in some situations to use barcodes as long as 585 bp (Juen & Traugott 2005), 650 bp (the COI marker defined by Folmer *et al.* 1994; Blankenship & Yayanos 2005) or even more when DNA is extracted from intact prey remains (Clare *et al.* 2009, 2011). However, in

many cases, such a long marker will reduce the success of the amplification so that sequences will not be recovered from many samples. Minimal variation in the barcode size among food species is also required to prevent preferential long fragment dropout during PCR (Pompanon *et al.* 2005) and biases because of differences in DNA copy number for long vs. short fragments in degraded samples (Deagle *et al.* 2006). Rayé *et al.* (2011) suggest that the short *trnL* fragment size of *Helianthemum nummularium* relative to other plant species could lead to its preferential amplification and thus explain its high proportion in their data set. DNA degradation also means that markers that occur in multiple copies per cell, such as mitochondrial and chloroplastic DNA regions or regions of nuclear ribosomal RNA genes, are usually used. While this improves the sensitivity of the PCR assays for detecting food species DNA, it does mean that there may be more variation in DNA:biomass ratios among species as copy numbers of these regions may vary extensively between food species.

#### Reference databases

The quality and coverage of databases of DNA barcodes available for NGS diet studies is often a major factor in choosing a barcoding primer set (Fig. 2). Where there is not a good collection of potential food DNA available, making a customized database is possible (e.g. Deagle *et al.* 2009; Rayé *et al.* 2011), but this may not be common for predation. The most extensive database is undoubtedly that for the IBOL (<http://www.barcodeoflife.org/>) but for animals, this has largely been generated for a mitochondrial region that is too large to amplify reliably from diet samples in most cases and has a variable primer-binding site among species. Primers that target a shorter portion of this region have been developed for identifying samples with degraded DNA (Meusnier *et al.* 2008) but may have restricted range of application. However, the region has been used in some dietary work targeting particular groups of prey (e.g. Zeale *et al.* 2011). The choice of plant barcoding regions used by CBoL, *matK* and *rbcL*, is unfortunately also different from those regions that have been most successfully used in DNA diet work, such as the *trnL* (UAA) intron (Valentini *et al.* 2009a). Despite all of the necessary compromises, a range of useful DNA metabarcoding primer sets is available for NGS diet assessment (Table 1).

#### Preventing amplification of non-target species

A problem encountered when designing primers with a broad taxonomic coverage is the possible amplifica-

tion of undesirable species. For example, broad coverage primers that amplify all eukaryotes have been used in DNA diet assessment (Martin *et al.* 2006; Riemann *et al.* 2010), but this strategy will produce amplicons from the predator and from gut parasites and symbionts. The amplification of DNA from the predator can be particularly problematic because it is often more abundant than the prey templates (e.g. Deagle *et al.* 2006) and can prevent their detection. Several solutions have been devised to deal with this difficulty. One set of approaches is based on the targeted removal of undesired dominant DNA templates. Digestion of predator DNA by endonuclease activity prior to (Dunshea 2009) and following (Blankenship & Yayanos 2005) PCR with broad coverage primers has had limited success. A promising alternative, termed 'suicide polymerase endonuclease restriction' (SuPER) (Green & Minz 2005), involves the design of a target-specific PCR primer and identification of a restriction site within the amplicon. During a joint PCR and endonuclease digest, only double-stranded target DNA is cut, leaving single stranded rare DNA templates intact. A requirement of this method, however, is finding enzymes capable of working at a high enough temperature to maintain PCR primer stringency. Another strategy is to block the amplification of predator DNA (reviewed by Vestheim *et al.* 2011). This can be accomplished using an artificially synthesized DNA analogue, such as peptide nucleic acid (PNA, Ørum *et al.* 1993) or locked nucleic acid (LNA, Dominguez & Kolodney 2005) designed to attach somewhere along the target predator DNA fragment and prevent DNA polymerase from extending the broad coverage PCR primers along the entire length of the fragment. The use of PNA in such 'PCR clamping' (Ørum *et al.* 1993) has been applied successfully to dietary analysis (Chow *et al.* 2011).

A cost-effective alternative, based on the same principle, is the use of an oligonucleotide modified at the 3' end to prevent polymerase extension, with the 3' hydroxyl group replaced by either a phosphate group (Carlson *et al.* 2003), a reversed (3' to 5') nucleotide (Corless *et al.* 2006), or a spacer-C3-CPG (Vestheim & Jarman 2008). The spacer-C3-CPG modification has been successfully applied in dietary studies (Vestheim & Jarman 2008; Deagle *et al.* 2009). Vestheim & Jarman (2008) found that a 10-fold excess of spacer-C3-CPG-blocking oligonucleotide compared with PCR primers was sufficient to reduce dominant predator amplicons from 100% to as low as 2.2% and even further at higher concentrations. The approach has been found to be efficient when designed to overlap with the primer-binding site, thus competitively preventing PCR primers from annealing [von Wintzingerode *et al.* 1997; Peano *et al.*

**Table 1** Main DNA barcoding primer sets for next generation sequencing (NGS) diet studies. The table references the main primer sets already used in DNA-based diet studies that are suitable for NGS diet assessment. The table is not exhaustive, and priority was given to primer sets with the broadest taxonomic coverage according to their ecological interest. However, more group-specific primers can be useful, such as those developed by Corse *et al.* (2010) for studying the diet of freshwater organisms

DNA region	Primers (5'-3')	Taxonomic range*	Amplicon length <sup>††</sup> (bp)	Size of public database (no. of accessions and species in EMBL) <sup>†§</sup>	Taxonomic resolution <sup>††</sup> (% family/genus/species identified)	Reference primers	Reference diet study
Plant <i>trnL</i>	<i>g</i> : GGGCAATCCTGAGCCAA <i>h</i> : CCATTGAGTCTTGCACCTATC	Green plants	12–134	53324/32890	65/34/18	Taberlet <i>et al.</i> (2007)	Valentini <i>et al.</i> (2009a), Soiminen <i>et al.</i> (2009), Rayé <i>et al.</i> (2011), Kowalczyk <i>et al.</i> (2011), Staudacher <i>et al.</i> (2011)
<i>rbcL</i>	<i>ZlaF</i> : ATGTCACCACCAACAGAGACTAAAGC <i>19bR</i> : CTTCTCAGGTTGGAATCCAG	Land plants	110	1726/1276	73/42/25	Hofreiter <i>et al.</i> (2000)	Hofreiter <i>et al.</i> (2000), Kim <i>et al.</i> (2011), Staudacher <i>et al.</i> (2011)
<i>trnH-psbA</i>	<i>psbAF</i> : GTTATGCATGAACGTAATGCTC <i>trnHR-2</i> : CGCCATGGTGGATTCACAAT	Land Plant	185–887	1059/656	98/97/78	Sang <i>et al.</i> (1997), Kress <i>et al.</i> (2005)	Stech <i>et al.</i> (2010)
Animals <i>12S</i>	<i>12SV5F</i> : TAGAACAGGCTCCTCTAG <i>12SV5R</i> : TTAGATACCCCACTATGC	Vertebrates	76–110	15974/7715	94/88/73	Riaz <i>et al.</i> (2011)	Riaz <i>et al.</i> (2011), Shehzad <i>et al.</i> (2012)
<i>16S</i>	<i>Chord_16S_F</i> : CGAGAAGACCCTRTGGAGCT 2 possible reverse primers <i>Chord_16S_R</i> : GGATTGCCCTGTTATCCCT <i>Chord_16 s_R_Short</i> : CCTNGGTGCCCCCAAC	Chordata	59–301 13–221	41834/16299 47276/16766	93/93/84 89/87/74	Deagle <i>et al.</i> (2009) based on Dumshea <i>et al.</i> (2008)	Deagle <i>et al.</i> (2009)
<i>COI</i>	<i>ZBI-Ar#F1c</i> : AGATATTGGAAACWTTATATTTTATTTTGG <i>ZBI-Ar#R2c</i> : WACTAAATCAATWCCAAATCCTCC	Insecta + Arachnida	157	3659/1697	93/95/83	Zeale <i>et al.</i> (2011)	Zeale <i>et al.</i> (2011), Bohmann <i>et al.</i> (2011)
<i>12S</i>	<i>185F</i> : TGTGTACTGCCGTCGTAAGCA <i>14233R</i> : AAGAGCGACGGCGGATGTGT	Lumbricidae (earthworms)	164–214	95/69	100/100/100	Harper <i>et al.</i> (2005)	King <i>et al.</i> (2010b), Eitzinger & Traugott (2011), Brown <i>et al.</i> (2012)
<i>16S</i>	<i>LR-J-12887</i> : CCGGTCTGAACTCAGATCACGT <i>LR-N-13398</i> : CGCCTGTTATCAAAAACAT	Most invertebrates and vertebrates	396–640	6246/3416	96/99/99	Simon <i>et al.</i> (1994)	Kasper <i>et al.</i> (2004)

Table 1 Continued

DNA region	Primers (5'-3')	Taxonomic range*	Amplicon length <sup>†‡</sup> (bp)	Size of public database (no. of accessions and species in EMBL) <sup>§</sup>	Taxonomic resolution <sup>†¶</sup> (% family/genus/species identified)	Reference primers	Reference diet study
<i>CytB</i>	CB3: GAGGAGCAACTGTAATTACTAA CB4: AAAAGAAARTATCATTCAGGTTGAAT 2 possible forward primers Col3F: GGACGATYTRTRGTTTCGT Col4F: GCTACAGCCTGAACAATWVG Col5R: TCTTGGCAAAATGCTTCCAGTA	Insecta + Arachnida  Collembola	343-364  89-230 134-136	4335/1117  143/79 152/87	89/95/96  61/70/65 23/25/32	Pons (2006)  Kuusk & Agustí (2008)	Pons (2006)  Kuusk & Agustí (2008) Chapman et al. (2010), Eitzinger & Traugott (2011) Brown (2011)
<i>COI</i>	LCOI490: GGTCAACAATAATCATAAAGATATTGG CI-N-1777: ACTTATATTGTTTATACGAGGAA LEP-FI: AITCAACCAATCATAAAGATAT LEP-RI: TAAACTTCTGGATGTCCAAAAA	Most animals  Lepidoptera + other animals	229-235  656-709	727/381  1850/979	96/98/98  98/99/99	Brown (2011)  Hebert et al. (2004) modified from Folmer et al. (1994)	Clare et al. (2009), Clare et al. (2011)
Eukaryotes <i>Nuclear LSU rDNA</i>	nucLSUDfi: CGTCTTGAACAACACGGACCAAG nucLSUDri: GCATAGTTCACCATCTTTCCGGG	Eukaryotes (data for plants and animals respectively)	122-180 129-559	3450/2875 18347/12915	54/64/74 72/87/93	Modified from Sonnenberg et al. (2007)	Vestheim & Jarman (2008)
<i>Nuclear LSU rDNA</i>	Short28SF: GTGTAACAACCTCACCTGCCG Short28SR: GCTACTACCACCAAGATCTG	Eukaryotes (data for plants and animals respectively)	132-169 140-332	860/633 5436/4523	63/82/87 70/81/91	Vestheim & Jarman (2008)	Vestheim & Jarman (2008)

\*Taxonomic range is given according to cited publications and gives an indication of the range taxa for which primers have been designed. However, within this range, some group of species may not be amplified and/or the range of taxa amplified may be broader.

<sup>†</sup>Parameters have been estimated by analysing the results of an *in silico* PCR as described in Ficetola et al. (2010). *In silico* PCR was performed on the release 107 of the EMBL public database (May 2011) and retained 10-1000 bp amplicons from accessions containing the priming sites with up to three mismatches per primer (except on the last three bases of the primer 3' end). *In silico* PCRs were restricted to Viridiplantae (taxid 33090) or Eumetazoa (taxid 6072) without *Homo sapiens* (taxid 9606) for plant and animal barcode primers, respectively. Primers with a broader taxonomic coverage (i.e. amplifying LSU rDNA) were tested on both groups.

<sup>‡</sup>Amplicon length is given without primers (range given for 99.9% of the sequences). Amplicons longer than 150-200 bp may not be adequate for degraded templates.

<sup>§</sup>The size of reference data sets is estimated by the number of accessions and species measured using EcoPCR and ecoTaxSpecificity, respectively (Ficetola et al. 2010). It gives an idea of the availability of reference sequence in public databases but does not prejudice either their quality or the relevance of building a customized database.

<sup>¶</sup>Taxonomic resolution is estimated as the percentage of taxa unambiguously identified for a given taxonomic level, using ecoTaxSpecificity (Ficetola et al. 2010). The estimation obtained from the whole EMBL data set is indicative and reflects the level of identification of the sequences from the public database. Actually, it strongly varies according to the reference data set and thus to the potential food species. The estimations of taxonomic resolution are not strictly comparable because they are not calculated on the same data set. Moreover, these values are under-estimated, because of errors in the database, such as wrong taxonomic assignment. Thus, we strongly recommend testing the performance of the primer pairs in the context of the study using dedicated bioinformatic tools (e.g. Ficetola et al. 2010; see text).



2005; Vestheim & Jarman 2008; Shehzad *et al.* 2012]. As it may be challenging to find target-specific differences directly flanking the PCR primer sites, the use of a long dual priming oligonucleotide (DPO) (Chun *et al.* 2007) may be used to effectively extend the region flanking the primer site in which a suitable blocking oligonucleotide site can be found (Vestheim & Jarman 2008). A DPO containing two separate binding regions with distinct annealing properties connected by a polydeoxyinosine linker and when modified with a C3 spacer can be as effective as a short blocking oligonucleotide (Chun *et al.* 2007; Vestheim & Jarman 2008). Thus, when blocking of non-target amplification is required, the barcode region chosen is ideally sufficiently variable in a region near the conserved PCR primer-binding sites for designing a blocking oligonucleotide that targets predator but not prey DNA.

Therefore, while an ideal DNA barcode primer set should consistently amplify short and unique DNA fragment already recorded in a well-curated database with equal efficiency from all food species, all primer sets developed so far compromise on some aspects of this situation. Finding the best compromise, which has long been empiric, may now go through comparative evaluation of several barcodes *in silico* (Fig. 2). New bioinformatic tools have been developed for estimating the resolution capacity of DNA barcodes (proportion of taxa unambiguously identified) and the taxonomic coverage of their associated primers (proportion of species amplified in a given range of taxa) from a set of reference sequences such as a complete mitochondrial genomes for vertebrates (Ficetola *et al.* 2010) or as internal transcribed spacers (ITS) for fungi (Bellemain *et al.* 2010). Based on this approach, designing the best-suited barcodes and associated primers for a given range of food species is now possible (Riaz *et al.* 2011).

### Technical aspects

King *et al.* (2008a) provide a detailed review of technical issues related to molecular diagnostics, from sampling in the field through to DNA extraction, primer design, PCR and laboratory procedures. Although this review concentrated upon predation studies, very similar issues are applicable to measurement of herbivory. Our intention here, however, is to focus on experimental design, sampling, preparation and pooling of samples (see Fig. 2), tagging of primers, and quantitative issues particularly relevant to NGS.

### Experimental design

Sampling design in NGS diet studies must take several factors into consideration if the expectation/hope is that

quantitative data on consumption of each dietary component (sequence counts for each taxon) could be built into food webs, providing population-level information directly, rather than through the sums of individual trophic links that are represented in conventional webs (see 'Quantitative aspects' below). Owing to sampling effects, the more individual consumers included in each pooled sample, the better the proportions of different sequences generated will reflect the diet of the population as a whole. Similarly, at the single sample level, Deagle *et al.* (2005) showed that DNA subsampled from different parts of the same sea lion faecal mass generated different dietary profiles, reflecting successive meals. Blending of the whole faecal mass ensures that the full dietary range pertaining to that mass is properly represented within subsamples taken for DNA extraction (Deagle *et al.* 2005; Matejusova *et al.* 2008). However, additional handling during blending might increase the contamination risk. Alternatively, subsamples can be taken from different parts of the faecal mass, pooled, homogenized and resampled (Deagle *et al.* 2009; Kowalczyk *et al.* 2011). Conversely, where predators produce many small faecal pellets, several faeces may be blended together and resampled (e.g. Deagle *et al.* 2010). A potential disadvantage of this approach is that individual diet samples may vary in food DNA content (because of age, size, etc.), resulting in some contributing disproportionately to the pooled DNA template. Taking multiple subsamples from each blend, whether derived from a single or several faeces, increases the chances of detecting all prey. In the same way, we recommend performing multiple amplifications per subsample (Taberlet *et al.* 2012) to cope with possible PCR drift.

Once a suitable barcoding primer set has been chosen and samples collected, the production of sequence data involves PCR amplification and carrying out NGS. Even when the diversity of DNA in a dietary sample corresponds to dozens of species, individual samples can be characterized by from a few hundred to a few thousands sequences. This means that, even with reduced-throughput NGS platforms, amplicons from several hundred samples can be pooled and sequenced together in a single run. Ideally, uniquely tagged primers are used in the PCR to amplify DNA from each sample. These tags then act as identifiers to recover data from each sample post-sequencing using a bioinformatics approach (see the study by Binladen *et al.* 2007; Meyer *et al.* 2007; Coissac 2012). Tags can be used to sort sequences by individual consumer (Pegard *et al.* 2009; Soininen *et al.* 2009; Valentini *et al.* 2009a; Kowalczyk *et al.* 2011) or to separate 'treatments', thus providing an average view of diet for each uniquely tagged group. For example, tags may be used to create sequence libraries for a consumer

species feeding in different habitats or locations (Deagle *et al.* 2009; Brown 2011), or at different times of year, to compare adults with juveniles or to investigate differences between the diets of each sex separately. Once PCR is completed, the construction of amplicon pools is in itself an important step and usually involves normalization of DNA concentrations of the individual amplicons before mixing (see the study by Harris *et al.* 2010 for possible methods).

#### *Exploratory NGS followed by targeted analysis*

Many ecological studies require data on the diets of individuals rather than populations, and statistical comparisons are only possible where there is sufficient replication. When large numbers of individuals are analysed separately, data may be broken down to examine the effects of multiple variables upon predation or herbivory (e.g. consumer age, sex, the time of year, temperature, spatial factors, etc.). Currently it is feasible, but not considered practical (or economic) to use separate tags for potentially hundreds of individual predators. However, a two-stage process is possible (Brown 2011). Many of the prey consumed by a generalist predator, for example, may be too rarely consumed to be of any significance in terms of energetics, food web dynamics or predator behaviour. Thus, identifying the main food sources might be adequate for modelling the primary dynamics within a food web, and NGS is good at identifying the major species in the diet, those that are primarily responsible for sustaining populations of consumers. However, identifying rare prey species may also be meaningful, for example when studying the impact of predation on threatened species. Species-specific primers can subsequently be designed that target these critical species, followed by mass screening of individuals, using multiplexing and fragment analysis to make the task more efficient (Harper *et al.* 2005; King *et al.* 2010b) or real-time PCR where quantification is advantageous and/or sensitivity improved (Zhang *et al.* 2007; Lundgren *et al.* 2009; Schmidt *et al.* 2009; Weber & Lundgren 2009). However, as the price of NGS is decreasing, and as the number of sequence reads per experiment is continuously increasing, such a two-step approach might not be justified in the near future.

#### *NGS platforms*

The first DNA-based diet studies applying NGS have used the pyrosequencing technology of Roche/454 Life Sciences (Margulies *et al.* 2005). This was the first commercially available platforms and has an advantage

over current Illumina/Solexa and AB SOLiD/Agencourt technologies because of its longer read length (Hudson 2008). However, technologies with shorter read length may be adequate for studying the shortest barcodes. For example, the Illumina HiSeq paired-end reads length (i.e. 100 bp) is compatible with the *trnL* approach (Valentini *et al.* 2009a). Another advantage of the 454 technology is ironically its lower sequencing capacity. Indeed many of the newest sequencing platforms are aimed at genome resequencing, or *de novo* genome assembly, and their capacity is higher than required for most directed amplicon sequencing projects, leading to an unnecessary increase in both the time needed for analysis and data storage capacity. However, their sequencing power (e.g. the 50 Mb produced in a HiSeq lane) remains compatible with analysing multiplexes of barcodes for hundreds or even thousands of samples. Thus, the choice of the NGS technology to use may depend on several parameters such as the barcode length, the number of barcodes used and the sample size (see Fig. 2). More complete description of the features and biases of the different NGS platforms can be found elsewhere (Glenn 2011).

A number of new instruments that cater to 'small-scale' amplicon sequencing projects are starting to be produced. Roche's GS Junior System uses pyrosequencing technology to produce about 70 000 amplicon sequences of about 450 bp in length (<http://my454.com>) for a consumable cost of 1300 €. Life Technologies' Ion Torrent Personal Genome Machine will initially generate 100 000 reads per run with a length of 100 bp at a consumable cost of 500 € per run (Perkel 2011). The Illumina MiSeq sequencing system (<http://www.illumina.com/systems/miseq.ilmn>) can produce more than 3 400 000 paired-end reads of 150 bp for a consumable cost of 600 €. Currently, these types of systems are the best adapted for NGS diet assessment because they have sufficient capacity to analyse large sample sizes combined with lower per-run costs and reduced computing requirements. However, given the frequent upgrades and evolution of the technologies, these specifications will become quickly obsolete. Therefore, even laboratories planning to routinely conduct NGS diet analyses may consider outsourcing high-throughput sequencing to companies that can follow the latest technological developments and propose the best adapted service currently available.

#### **Data analysis**

As for other environmental sequencing experiments, the benefit of using NGS resides in the ability to analyse multiplexed PCR products obtained using tagged

primers with a broad taxonomic coverage. The output consists of DNA sequences (i.e. reads) and corresponding quality values (for each nucleotide of each sequence read). Primary analysis consists of (i) discarding sequences with errors, (ii) sorting the remaining sequences according to their tag, (iii) clustering them and/or assigning them to a taxon. As for all studies generating DNA sequences, there is a need for public access to the data produced in NGS environmental studies. A solution is to deposit unique sequences in DRYAD (<http://datadryad.org/>).

### *Dealing with errors*

Erroneous reads may affect the analysis in two ways. First, an error occurring on the tag may lead to assignment of a sequence to the wrong sample group. We can overcome this by using an appropriate minimum number of differences among tags to maintain the ability to correctly assign a sequence to a sample even when errors occur on the tag sequence (Coissac 2012). Second, errors affecting the barcode sequence between primers may lead to taxon misidentification. These errors may occur during the sequencing process. For example, 454 technology is known for its low reliability during homopolymer extension (Huse *et al.* 2007). Such read errors may be identified through low-quality values. However, analysing PCR products from a unique sequence template shows that most of errors result from degradation of template DNA or nucleotide misincorporation during PCR occurring before the sequencing step and are not related to a low-quality value. This is why previous studies have been conservative, discarding rare sequences, as well as more frequent sequences that were close to a very common sequence (Valentini *et al.* 2009a). Indeed, the strategies set up for removing errors are based on the assumption that the correct sequence is the most common one, that is the error is rare and occurs late during the replication process. However, nucleotide misincorporation may sometimes occur during an early step in the PCR, leading to erroneous sequences predominating. A conservative approach would therefore be to discard a frequent sequence that occurs in a single sample when aiming to describe the diet at a population or group level or to replicate PCRs for each sample (i.e. multitube approach) when aiming at describing precisely differences among individuals. Another common problem leading to erroneous sequences is the occurrence of PCR-generated chimeras (Qiu *et al.* 2001) that is promoted when amplifying degraded DNA (Pääbo *et al.* 1990). Such chimeric sequences can be managed during the taxon assignment step by assigning the sequence to a higher taxonomic rank that includes both taxa composing the chimera, or

to unassign it if the taxa are too distant from anything in the database (Soininen *et al.* 2009).

To date, dealing with errors has mainly involved discarding sequences using more or less arbitrarily chosen thresholds. Several studies have discarded sequences occurring <4 times (Valentini *et al.* 2009a; Rayé *et al.* 2011), but there is no general rule, and one may use a threshold relative to the number of sequences per sample. A drawback of removing less frequent sequences is that rare food items could be missed. Protocols used to date may reliably estimate the diet at a population/group level, but we do not yet know how we can interpret between-individual differences for taxa represented by low frequency sequences. However, new methods are being developed that remove the noise from sequence data sets by examining chimeras, by building clusters of sequences relating erroneous sequences to the correct one (e.g. Quince *et al.* 2011), and we can predict that more explicit error models and the development of new related methods will decrease the impact of read errors on diet assessment.

### *Taxon assignment*

Another critical step in the analysis is the sequence assignment to a taxon. This is possible through comparison of each sequence to a reference database (e.g. ecoTag, Pegard *et al.* 2009; Taxonator, Jones *et al.* 2011). The reference database may be a subset of public databases and/or a set of sequences specifically produced for the study. This allows faster processing, by avoiding inappropriate comparisons and cleaning the reference data set from errors. An efficient way of selecting from a public database (e.g. GenBank, EMBL), the sequences corresponding to the barcode of interest, is to perform an *in silico* PCR using the barcode primers (Pegard *et al.* 2009; Soininen *et al.* 2009). *In silico* PCR programs such as ecoPCR may allow imperfect matches between each barcode primer and its binding site to mimic *in vitro* PCR (Ficetola *et al.* 2010). It is well known that public databases contain sequencing errors and a few wrong taxonomic assignments (Harris 2003). As a consequence, it is important to remove errors when using public data, because the quality (as well as the comprehensiveness) of the reference database determines the reliability and accuracy of taxon identification. When possible, the best solution is to build a customized reference database gathering all potential food items and not only those deduced from an a priori knowledge of the diet. Examples of this include the study by Soininen *et al.* (2009), who studied subarctic voles using a database of 842 widespread and/or ecologically important arctic plant species, and the study by Rayé *et al.* (2011), who based their study of the diet of chamois from the

Bauges Massif on a reference database of 475 plant species collected in this massif. Using a customized database has two major advantages. First, it allows the diet to be analysed with reference to what is actually available and ecologically meaningful. Second, it allows a more accurate taxonomic assignment. Indeed, a species sharing the same barcode sequence with related species living in different ecosystems or biogeographical regions may be identified at a higher taxonomic level (i.e. genus, family) when using a database built at the worldwide level. However, public databases may be useful for identifying sequences not assigned when using the customized database (Soininen *et al.* 2009).

Although unrecognized sequences are often simply a result of incomplete barcoding databases, they can also be a source of new discoveries. For example, species may be discovered in the diets of predators or herbivores that are not known to exist in that area, or which may be new to science. This also applies to cloning approaches, but the far greater exploratory power intrinsic to massive parallel sequencing will greatly accelerate the discovery process. Molecular phylogenetics frequently reveals extensive cryptic speciation in sympatry among the relatively few organisms that have been analysed in any detail (e.g. Williams *et al.* 2006; Finston *et al.* 2007; Heethoff *et al.* 2007; King *et al.* 2008b). Predation on cryptic species or lineages can be studied using specific primers (King *et al.* 2010b). NGS will provide archives of trophic links that can be mined for proof of trophic interactions in future, when new species designations have been defined and (where possible) sequences matched to morphologies.

The use of public databases is also useful when building a comprehensive local database is out of reach such as in tropical forests (Gonzalez *et al.* 2009). This may lead to a low-resolution taxa assignment (e.g. at the family level) that may be sufficient for estimating the diversity of the diet or the overlap between trophic niches of different species. An anonymous survey (i.e. without taxonomic information) may also be informative. The diversity of a diet can be assessed for example using Ecological Clades that are monophyletic groups containing species sharing precise ecological features (e.g. Corse *et al.* 2010). Molecular operational taxonomic Units (MOTUs) that represents groups defined on a common barcode sequence are also currently used for assessing ecological diversity when taxon assignment is not possible (see for example, the studies by Floyd *et al.* 2002; Blaxter *et al.* 2005; Bohmann *et al.* 2011), and dedicated program exist that MOTUs (e.g. jMOTU, Jones *et al.* 2011). The clustering approach that is used to build MOTUs is of critical importance and will have to take into account PCR and sequencing errors.

## Validation of data accuracy

While it is clear that NGS has the potential to be enormously useful in dietary studies (Valentini *et al.* 2009b), unforeseen biases can influence the recovery of sequences as in all DNA-based studies (Pompanon *et al.* 2005). Therefore, it is important to follow the general advice on best practice for DNA-based diet studies (King *et al.* 2008a) and whenever possible validate steps of the process. We will outline a few points that are particularly relevant for NGS studies.

### Validity of sequence data

From sample collection, through to DNA extraction and PCR, it is necessary to be extremely careful to avoid cross-contamination. The low amount of target DNA in many dietary samples, combined with extreme sensitivity of PCR and the ability to recover thousands of sequences per sample, means that even minor contamination will be represented in the final data set. This is clearly illustrated by the recovery of contaminating human DNA sequences (0.5–2.4% of sequences) in amplicon sequencing studies using vertebrate barcoding markers (Binladen *et al.* 2007; Deagle *et al.* 2009). In these studies, the human sequences could simply be discounted, but if contamination had been target DNA it would have been difficult (or impossible) to detect and could produce misleading results. Amplicons from previous experiments are the most potent source of target DNA contamination, necessitating physical separation of laboratory space and equipment used in pre- and post-PCR amplification steps (King *et al.* 2008b). It is also important to carry out DNA extractions from tissue, and primer testing experiments using target DNA, in isolation from dietary sample processing (see the studies by Taberlet *et al.* 1999; Cooper & Poinar 2000 for general discussion of laboratory contamination issues). Post-PCR contamination (e.g. between lanes in NGS systems) can be detected more easily than pre-PCR because reads can be assigned to the right sample by looking at the primer and Tag sequences. As additional amplicon sequencing studies using NGS are published, it may become apparent that a low level of contamination is inevitable and all sequences below a certain threshold might best be excluded from final analysis. To minimize representation of contaminants, Valentini *et al.* (2009a) recommended no more than 35 cycles during PCRs. Regardless, minimizing the potential for contamination of dietary samples by exogenous DNA is crucial to obtain accurate diet data. Contamination originates in many forms and some are ecological. Species unintentionally consumed can be detected in the case of secondary predation where a predator eats

another predator that contains prey in its guts (Sheppard *et al.* 2005). Similarly, vertebrate grazers may unintentionally consume large numbers of invertebrates, and marine species may accidentally consume planktonic components. According to the goal of the study, this information can be considered as contamination (e.g. when studying food preference) or not (when looking at the actual intake).

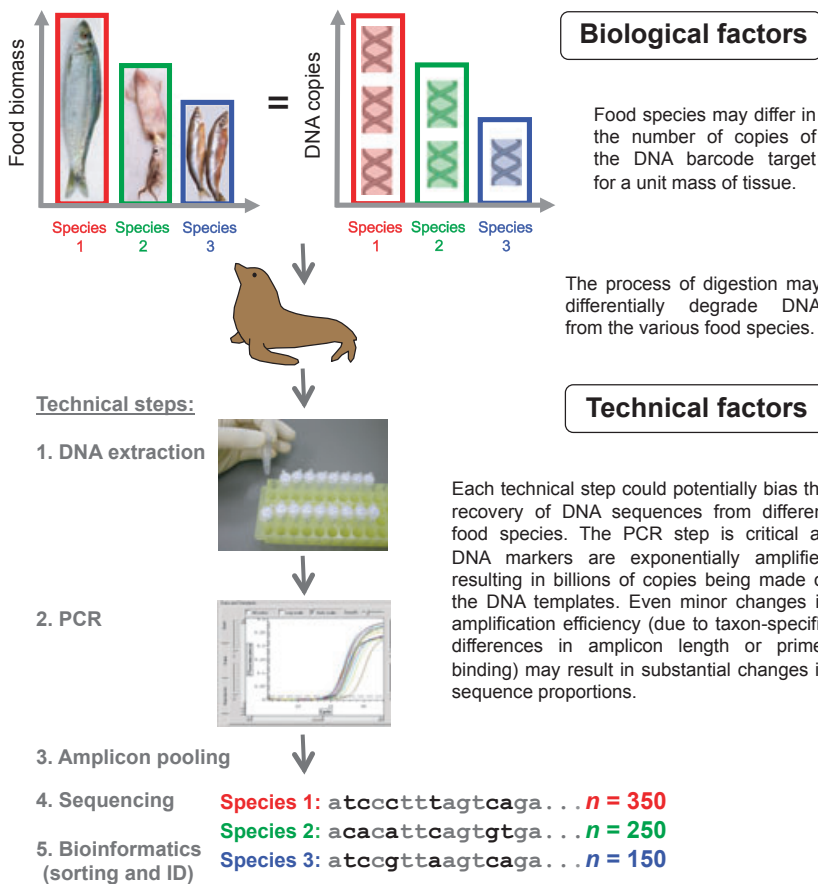
Beyond contamination, there are several other technical issues that could influence the diversity of sequences in the final data set. Most critically is the ability of the barcoding primers to reliably amplify the intended targets, and this should be taken into account when designing barcode primers, as discussed earlier. However, different sequencing platforms may also preferentially sequence certain amplicons (Dohm *et al.* 2008) or fail to completely sequence particular amplicons (Deagle *et al.* 2009), and this may also lead to some sequences from some targets being unrepresented in the final data set.

#### *Quantitative aspects*

The intent of many dietary studies is not simply to analyse diversity, but to obtain quantitative data on the relative amounts of different foods consumed by a species. For this to be possible using NGS, the proportional biomass of foods eaten needs to be reflected in the proportions of the recovered DNA sequences. It makes intuitive sense that if a large biomass of a particular food item was consumed, this would be mirrored in the amount of DNA in the dietary sample and ultimately in the recovered DNA sequences. However, results from many different fields of research have shown that obtaining quantitative data in amplicon sequencing studies is fraught with problems (e.g. von Wintzingerode *et al.* 1997; Polz & Cavanaugh 1998; Acinas *et al.* 2005; Porazinska *et al.* 2009; Amend *et al.* 2010). The most basic requirement is that the quantitative signature must be retained during all technical steps (see Fig. 3). A clear potential for bias exists during PCR when target DNA is exponentially amplified; even a 2% difference in amplification efficiency between two initially equal targets can lead to a 30% divergence in DNA copy number over 35 cycles. Biases may also appear during DNA extraction (e.g. Martin-Laurent *et al.* 2001), DNA pooling (e.g. Harris *et al.* 2010), sequencing (e.g. Porazinska *et al.* 2010) and during bioinformatic sorting (Amend *et al.* 2010). Beyond these technical issues, a number of biological features may obscure the quantitative signal (see Fig. 3). These include variation in tissue cell density (and therefore DNA per gram of tissue eaten), inter- and intraspecific variation in gene copy number (particularly relevant to

markers for mtDNA, chloroplast DNA and nuclear ribosomal gene clusters) (Prokopowich *et al.* 2003), differential survival of DNA during digestion (e.g. Deagle & Tollit 2007), and differences in the state of digestion (particularly in stomach contents e.g. Troedsson *et al.* 2009). Finally, it is unclear what the effect of averaging data from many samples will have on the quantitative estimates. On the one hand, stochastic variation will be minimized, but systemic biases will likely persist. Pooling a standardized amount of PCR product before NGS will result in each sample having equal weight in the final dietary data set, and this will affect the quantitative data (e.g. a dissected stomach containing a single small prey item would count for as much as a stomach full of many items).

Given this substantial list of obstacles, how can we know whether NGS data accurately reflect diets, either quantitatively or even qualitatively? One possibility is to use alternative diet analysis methods in parallel to help validate the DNA-based data. Soininen *et al.* (2009) used microhistological identification of plant fragments in vole stomachs to complement their DNA-based approach. This did provide some assurance that the NGS DNA results were reasonable, but was hindered by the high degree of uncertainty in the microhistological identification (Soininen *et al.* 2009). This is likely to be a common problem because DNA-based methods are most usefully applied in situations where other methods of diet analysis are problematic. Deagle *et al.* (2009) also compared their prey sequence data from fur seal faeces with traditional prey hard-part identification and found a good agreement between data sets, increasing confidence in both methods of diet analysis for fur seals. The sequence depth afforded by NGS also allowed Deagle *et al.* (2009) to sequence multiple mtDNA barcoding markers targeting the same prey species to cross-validate the precision of their data. This is a useful approach to help uncover major marker-specific, or primer-specific, biases (see also Murray *et al.* 2011). Validation studies based on captive feeding trials with animals fed a known diet will probably provide the best test of the methodology in the long run. Many studies have examined DNA-based approaches for studying diet with captive animals (e.g. Hoogendoorn & Heimpel 2001; Deagle *et al.* 2005; Foltan *et al.* 2005; Weber & Lundgren 2009) and a few have specifically focused on Quantitative aspects (Deagle & Tollit 2007; Bowles *et al.* 2011). One study has investigated the accuracy of quantitative diet data from high-throughput amplicon sequencing (Deagle *et al.* 2010). In this study on captive penguins, the authors found differences in digestibility of the fish prey and suggested that estimates of diet composition were possible but should be given wide confidence intervals. With these



**Fig. 3** Summary of the biological and technical factors that may affect the proportions of DNA sequences recovered from samples in a next generation sequencing diet study. Given the number of potential biases even a well-designed dietary barcoding study is likely to only provide semi-quantitative data on the diet of a species. This is a limitation shared with all methods of diet analysis.

limitations in mind, where possible, it would be prudent to design NGS dietary studies that are comparative and not as dependent on absolute quantification (e.g. measuring spatial or temporal variation in diet).

## Conclusion

Our analysis of the potential and pitfalls of the NGS approach to dietary analyses, detailed in this review, was designed to help researchers to choose what technique(s) to use before embarking upon a new study. All approaches do not bring the same information and other methods might be used instead of, or together with, NGS methods. DNA-based methods are especially efficient for determining the taxa eaten, but when an estimation of nutritional components is required (e.g. total nitrogen for herbivores), the use of NIRS may be required (e.g. Kaneko & Lawler 2006; Rothman *et al.* 2009) even if the calibration step is critical. Also, the use of stable isotopes gives integrative information that DNA-based approaches cannot provide. It reflects the diet over longer periods of time and gives insight into how what was ingested is used. A limitation of this approach is the necessity of obtaining prior knowledge of the isotopic signals for different prey types (e.g.

Moore & Semmens 2008). A combination of DNA-based and stable isotopes analyses has already been proven to be effective (Hardy *et al.* 2010). NIRS (Foley *et al.* 1998) and micro-histological studies may also be of interest when information on the part of the organism consumed is required. Other DNA-based techniques may also be preferred to NGS sequencing when quantitative data are expected, such as real-time PCR, which may also be used for validating the semi-quantitative aspect of the NGS approach in a given situation (see 'Quantitative aspects' section). Classical sequencing approaches, or prey-specific primers, can also complement an NGS approach that provides insufficiently resolution for a given group, by allowing discriminating among potential eaten species within this group [e.g. distinction between wild mouflon and domestic sheep eaten by snow leopards, W. Shehzad, T. M. MacCarthy, F. Pompanon, L. Purejav, E. Coissac, T. Riaz, P. Taberlet unpublished]. If predation or herbivory on a limited number of specific target species is required, then a multiplex-PCR approach is probably a quicker, simpler and less expensive approach (Harper *et al.* 2005; King *et al.* 2010a,b; Traugott *et al.* 2012). However, NGS diet assessment takes advantage of the previous progress in DNA-based methods for diet analysis (King *et al.*

2008a) and will benefit from the continued development of NGS technologies for characterizing biodiversity in environmental samples (Valentini *et al.* 2009b). The major advantage of NGS-based techniques is that they are ideal for providing precise taxonomic identification of food items within highly diverse diets, whatever the type of sample and especially non-invasive and degraded ones, with low effort compared to the large sample sizes that can be analysed. NGS diet assessment has already been applied to study herbivores and carnivores from highly diverse taxa such as insects, spiders, birds, molluscs, mammals and reptiles (Deagle *et al.* 2009, 2010; Valentini *et al.* 2009a; Bohmann *et al.* 2011; Brown 2011; Murray *et al.* 2011) and can be adapted to analyse the diet of most organisms. Valentini *et al.* (2009a) showed that NGS diet assessment is especially well suited to studying extremely diverse herbivore diets. The lack of resolution of some barcodes in particular cases can be overcome by the simultaneous use of several barcodes (Deagle *et al.* 2009; Valentini *et al.* 2009a; Rayé *et al.* 2011).

Given that a species can be identified in a complex substrate using broad coverage primers when its DNA represents a low proportion of the target DNA (Pegard *et al.* 2009), even uncommon food species can be documented. In cases where samples are dominated by non-target DNA that would be co-amplified with food species (e.g. DNA from predator or parasites), the use of blocking oligonucleotides provides a practical route to preferentially amplify the DNA of interest (Vestheim & Jarman 2008). NGS-based methods may also uncover a higher dietary diversity than traditional methods by detecting DNA from prey that leave no hard-parts in the faeces. For example, occasional predation of rays and sharks by the Australian fur seal revealed in a DNA-based study was previously missed in traditional faecal analysis because of the digestion of their cartilaginous skeletons (Deagle *et al.* 2009). Similarly, Brown *et al.* (in press) were able to use NGS to analyse predation by lizards on different earthworm species, none of which can be distinguished by microscopic examination of faeces. However, the power of NGS diet assessment is at the same time limited by problems such as amplification of contaminants and PCR errors. Thus, if the sequencing of the PCR products is outsourced to a company or to a sequencing centre, prior template preparation should be carried out in laboratory space dedicated to minimizing contamination and the data analysis should take into account potential for sequence errors. Even with such constraints, all published studies recognize that methods based on NGS reduce the effort necessary for carrying out diet assessment. This is true in comparison with hard-parts analysis of faeces (e.g. Deagle *et al.* 2009; Soininen *et al.* 2009) and also with other

DNA-based techniques. Thus, the time and money saved can potentially be reallocated advantageously to developing a more sophisticated and extensive sampling strategy and increasing the accuracy of the study. While semi-quantitative interpretation of dietary barcoding data is probably the best that can be hoped for, given the difficulties with alternate approaches of diet determination, DNA-based methods may be one of the most accurate approaches available for increasing our understanding of trophic relationships in a diverse range of food webs.

### Acknowledgement

We would like to thank Alice Valentini and Ludovic Gielly for fruitful discussions.

### References

- Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology*, **71**, 8966–8969.
- Agustí N, Shayler SP, Harwood JD *et al.* (2003) Collembola as alternative prey sustaining spiders in arable ecosystems: prey detection within predators using molecular markers. *Molecular Ecology*, **12**, 3467–3475.
- Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular Ecology*, **19**, 5555–5565.
- Asahida T, Yamashita Y, Kobayashi T (1997) Identification of consumed stone flounder, *Kareius bicoloratus* (Basilewsky), from the stomach contents of sand shrimp, *Crangon affinis* (De Haan) using mitochondrial DNA analysis. *Journal of Experimental Marine Biology and Ecology*, **217**, 153–163.
- Bellemain E, Carlsen T, Brochmann C *et al.* (2010) ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. *BMC Microbiology*, **10**, e189.
- Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One*, **2**, e197.
- Blankenship LE, Yayanos AA (2005) Universal primers and PCR of gut contents to study marine invertebrate diets. *Molecular Ecology*, **14**, 891–899.
- Blaxter M, Mann J, Chapman T *et al.* (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1935–1943.
- Bohmann K, Monadjem A, Lehmkuhl Noer C *et al.* (2011) Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. *PLoS One*, **6**, e21441.
- Boreham PFL, Ohiagu CE (1978) The use of serology in evaluating invertebrate prey-predator relationships: a review. *Bulletin of Entomological Research*, **68**, 171–194.

- Bowles E, Schulte PM, Tollit DJ, Deagle BE, Trites AW (2011) Proportion of prey consumed can be determined from faecal DNA using real-time PCR. *Molecular Ecology Resources*, **11**, 530–540.
- Bradley BJ, Stiller M, Doran-Sheehy DM *et al.* (2007) Plant DNA sequences from feces: potential means for assessing diets of wild primates. *American Journal of Primatology*, **69**, 699–705.
- Brown DS (2011) *Molecular Analysis of the Diet of British Reptiles*. PhD Thesis, Cardiff University, UK.
- Brown DS, Jarman SN, Symondson WOC (2012) Pyrosequencing of prey DNA in reptile faeces. *Molecular Ecology Resources*, in press.
- Carlson CM, Dupuy A, Fritz S, Roberg-Perez K, Fletcher CF, Largaespada DA (2003) Transposon mutagenesis of the mouse germline. *Genetics*, **165**, 243–256.
- Chapman EG, Romero SA, Harwood JD (2010) Maximising collection and minimising risk: does vacuum suction sampling increase the likelihood for misinterpretation of food web connections. *Molecular Ecology Resources*, **10**, 1023–1033.
- Chase MW, Cowan RS, Hollingsworth PM *et al.* (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon*, **56**, 295–299.
- Chow S, Suzuki S, Matsunaga T, Lavery S, Jeffs A, Takeyama H (2011) Investigation on natural diets of larval marine animals using peptide nucleic acid-directed polymerase chain reaction clamping. *Marine Biotechnology*, **13**, 305–313.
- Chun JY, Kim KJ, Hwang IT *et al.* (2007) Dual priming oligonucleotide system for the multiplex detection of respiratory viruses and SNP genotyping of CYP2C19 gene. *Nucleic Acids Research*, **35**, e40.
- Clare EL, Fraser EE, Braid HE, Fenton MB, Hebert PDN (2009) Species on the menu of a generalist predator, the eastern red bat (*Lasiurus borealis*): using a molecular approach to detect arthropod prey. *Molecular Ecology*, **18**, 2532–2542.
- Clare EL, Barber BR, Sweeney BW, Hebert PDN, Fenton MB (2011) Eating local: influences of habitat on the diet of little brown bats (*Myotis lucifugus*). *Molecular Ecology*, **20**, 1772–1780.
- Coissac E (2012) oligoTag: a program for designing sets of tag for Next Generation Sequencing of multiplexed samples. In: *Data Production and Analysis in Population Genomics* (eds Pompanon F, Bonin A), in press. Humana Press, New York.
- Cooper A, Poinar HN (2000) Ancient DNA: do it right or not at all. *Science*, **289**, 1139.
- Corless CL, Harrell P, Lacouture M *et al.* (2006) Allele-specific polymerase chain reaction for the imatinib-resistant KIT D816V and D816F mutations in mastocytosis and acute myelogenous leukaemia. *Journal of Molecular Diagnostics*, **8**, 604–612.
- Corse E, Costedoat C, Chappaz R *et al.* (2010) A PCR-based method for diet analysis in freshwater organisms using 18S rDNA barcoding on faeces. *Molecular Ecology Resources*, **10**, 96–108.
- Darling JA, Blum MJ (2007) DNA-based methods for monitoring invasive species: a review and prospectus. *Biological Invasions*, **9**, 751–765.
- Deagle BE, Tollit DJ (2007) Quantitative analysis of prey DNA in pinniped faeces: potential to estimate diet composition? *Conservation Genetics*, **8**, 743–747.
- Deagle BE, Tollit DJ, Jarman SN *et al.* (2005) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Molecular Ecology*, **14**, 1831–1842.
- Deagle BE, Eveson JP, Jarman SN (2006) Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces. *Frontiers in Zoology*, **3**, 11.
- Deagle BE, Gales NJ, Evans K *et al.* (2007) Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (*Eudyptes chrysolophus*). *PLoS One*, **2**, e831.
- Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, **18**, 2022–2038.
- Deagle BE, Chiaradia A, McInnes J, Jarman SN (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conservation Genetics*, **11**, 2039–2048.
- Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, **36**, e105.
- Dominguez PL, Kolodney MS (2005) Wild-type blocking polymerase chain reaction for detection of single nucleotide minority mutations from clinical specimens. *Oncogene*, **24**, 6830–6834.
- Dove H, Mayes RW (1996) Plant wax components: a new approach to estimating intake and diet composition in herbivores. *Journal of Nutrition*, **126**, 13–26.
- Dunshea G, Barros NB, Wells RS *et al.* (2008) Pseudogenes and DNA-based diet analyses: a cautionary tale from a relatively well sampled predator-prey system. *Bulletin of Entomological Research*, **98**, 239–248.
- Dunshea G (2009) DNA-based diet analysis for any predator. *PLoS One*, **4**, e5252.
- Eitzinger B, Traugott M (2011) Which prey sustains cold-adapted invertebrate generalist predators in arable land? Examining prey choices by molecular gut-content analysis. *Journal of Applied Ecology*, **48**, 591–599.
- Estes JA, Terborgh J, Brashares JS *et al.* (2011) Trophic downgrading of planet earth. *Science*, **333**, 301–306.
- Felske A, Akkermans ADL, de Vos WD (1998) Quantification of 16S rRNAs in complex bacterial communities by multiple competitive reverse transcription – PCR in temperature gradient gel electrophoresis fingerprints. *Applied and Environmental Microbiology*, **64**, 4581–4587.
- Ficetola F, Coissac E, Zundel S *et al.* (2010) *In silico* comparison of primers for DNA barcoding. *BMC genomics*, **11**, e434.
- Finston TL, Johnson MS, Humphreys WF, Eberhard SM, Halse SA (2007) Cryptic speciation in two widespread subterranean amphipod genera reflects historical drainage patterns in an ancient landscape. *Molecular Ecology*, **16**, 355–365.
- Floyd R, Abebe E, Papert A *et al.* (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology*, **11**, 839–850.
- Foley WJ, McIlwee A, Lawler I *et al.* (1998) Ecological applications of near infrared reflectance spectroscopy—a tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance. *Oecologia*, **116**, 293–305.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.



- Foltan P, Sheppard S, Konvicka M, Symondson WOC (2005) The significance of facultative scavenging in generalist predator nutrition: detecting decayed prey in the guts of predators using PCR. *Molecular Ecology*, **14**, 4147–4158.
- Gariépy TD, Kuhlmann U, Gillott C, Erlandson M (2007) Parasitoids, predators and PCR: the use of diagnostic molecular markers in biological control of arthropods. *Journal of Applied Entomology*, **131**, 225–240.
- Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11**, 759–769.
- Gonzalez MA, Baraloto C, Engel J *et al.* (2009) Identification of Amazonian trees with DNA barcodes. *PLoS One*, **4**, e7483.
- Gratton C, Donaldson J, Vander Zanden MJ (2008) Ecosystem linkages between lakes and the surrounding terrestrial landscape in northeast Iceland. *Ecosystems*, **11**, 764–774.
- Green SJ, Minz D (2005) Suicide polymerase endonuclease restriction, a novel technique for enhancing PCR amplification of minor DNA templates. *Applied and Environmental Microbiology*, **71**, 4721–4727.
- Hardy CM, Krull ES, Hartley DM, Oliver RL (2010) Carbon source accounting for fish using combined DNA and stable isotope analyses in a regulated lowland river weir pool. *Molecular Ecology*, **19**, 197–212.
- Harper GL, King RA, Dodd CS *et al.* (2005) Rapid screening of invertebrate predators for multiple prey DNA targets. *Molecular Ecology*, **14**, 819–827.
- Harper GL, Sheppard SK, Harwood JD *et al.* (2006) Evaluation of temperature gradient gel electrophoresis for the analysis of prey DNA within the guts of invertebrate predators. *Bulletin of Entomological Research*, **96**, 295–304.
- Harris DJ (2003) Can you bank on GenBank? *Trends in Ecology & Evolution*, **18**, 317–319.
- Harris JK, Sahl JW, Castoe TA *et al.* (2010) Comparison of normalization methods for construction of large, multiplex amplicon pools for next-generation sequencing. *Applied and Environmental Microbiology*, **76**, 3863–3868.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B*, **270**, 313–321.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the Neotropical skipper butterfly *Astrartes fuligator*. *Proceedings of the National Academy of Sciences, USA*, **101**, 14812–14817.
- Heethoff M, Domes K, Laumann M *et al.* (2007) High genetic divergences indicate ancient separation of parthenogenetic lineages of the oribatid mite *Platynothrus peltifer* (Acari, Oribatida). *Journal of Evolutionary Biology*, **20**, 392–402.
- Hofreiter M, Poinar HN, Spaulding WG *et al.* (2000) A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology*, **9**, 1975–1984.
- Hofreiter M, Betancourt JL, Sbriller AP, Markgraf V, McDonald HG (2003) Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Cura, Neuquen Province, southwest Argentina. *Quaternary Research*, **59**, 364–378.
- Holechek JL, Vavra M, Pieper RD (1982) Botanical composition determination of range diets: a review. *Journal of Range Management*, **35**, 309–315.
- Hoogendoorn M, Heimpel GE (2001) PCR-based gut content analysis of insect predators: using ribosomal ITS-1 fragments from prey to estimate predation frequency. *Molecular Ecology*, **10**, 2059–2067.
- Höss M, Kohn M, Pääbo S, Knauer F, Schroder W (1992) Excrement analysis by PCR. *Nature*, **359**, 199.
- Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources*, **8**, 3–17.
- Huse SM, Huber JA, Morrison HG *et al.* (2007) Accuracy and quality of massively-parallel DNA pyrosequencing. *Genome Biology*, **8**, R143.
- Ingerson-Mahar J (2002) Relating diet and morphology in adult carabid beetles. In: *The Agroecology of Carabid Beetles* (ed. Holland J), pp. 111–136. Intercept, Andover, UK.
- Jarman SN, Deagle BE, Gales NJ (2004) Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. *Molecular Ecology*, **13**, 1313–1322.
- Jones M, Ghoorah A, Blaxter M (2011) jMOTU and taxonator: turning DNA barcode sequences into annotated operational taxonomic units. *PLoS One*, **6**, e19259.
- Juen A, Traugott M (2005) Detecting predation and scavenging by DNA gut-content analysis: a case study using a soil insect predator-prey system. *Oecologia*, **142**, 344–352.
- Kaneko H, Lawler IR (2006) Can near infrared spectroscopy be used to improve assessment of marine mammal diets via fecal analysis? *Marine Mammal Science*, **22**, 261–275.
- Kasper ML, Reeson AF, Cooper SJB, Perry KD, Austin AD (2004) Assessment of prey overlap between a native (*Polistes humilis*) and an introduced (*Vespula germanica*) social wasp using morphology and phylogenetic analyses of 16s rDNA. *Molecular Ecology*, **13**, 2037–2048.
- Kim BJ, Lee NS, Lee SD (2011) Feeding diets of the Korean water deer (*Hydropotes inermis argyropus*) based on a 202 bp rbcL sequence analysis. *Conservation Genetics*, **12**, 851–856.
- King RA, Read DS, Traugott M, Symondson WOC (2008a) Molecular analysis of predation: a review of best practice for DNA-based approaches. *Molecular Ecology*, **17**, 947–963.
- King RA, Tibble AL, Symondson WOC (2008b) Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Molecular Ecology*, **17**, 4684–4698.
- King RA, Moreno-Ripoll R, Agustí N *et al.* (2010a) Multiplex reactions for the molecular detection of predation on pest and non-pest invertebrates in agroecosystems. *Molecular Ecology Resources*, **11**, 270–373.
- King RA, Vaughan IP, Bell JR, Bohan DA, Symondson WOC (2010b) Prey choice by carabid beetles feeding on an earthworm community analysed using species- and lineage-specific PCR primers. *Molecular Ecology*, **19**, 1721–1732.
- Kowalczyk R, Taberlet P, Coissac E *et al.* (2011) Influence of management practices on large herbivore diet—Case of European bison in Białowieża Primeval Forest (Poland). *Forest Ecology and Management*, **261**, 821–828.
- Kress WJ, Wurdack KJ, Zimmer EA *et al.* (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences, USA*, **102**, 8369–8374.
- Kuusk AK, Agustí N (2008) Group-specific primers for DNA-based detection of springtails (Hexapoda: Collembola) within predator gut contents. *Molecular Ecology Resources*, **8**, 678–681.

- Kuusk AK, Ekblom B (2010) Lycosid spiders and alternative food: feeding behaviour and implications for biological control. *Biological Control*, **55**, 20–26.
- Kvitrud MA, Riemer SD, Brown RF, Bellinger MR, Banks MA (2005) Pacific harbor seals (*Phoca vitulina*) and salmon: genetics presents hard numbers for elucidating predator-prey dynamics. *Marine Biology*, **147**, 1459–1466.
- Lecomte N, Ahlström Ø, Ehrlich D *et al.* (2011) Intrapopulation variability shaping isotope discrimination and turnover: experimental evidence in Arctic foxes. *PLoS One*, **6**, e21357.
- Lessa EP, Applebaum G (1993) Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology*, **2**, 119–129.
- Lundgren JG, Ellsbury ME, Prischmann DA (2009) Analysis of the predator community of a subterranean herbivorous insect based on polymerase chain reaction. *Ecological Applications*, **19**, 2157–2166.
- Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.
- Martin DL, Ross RM, Quetin LB, Murray AE (2006) Molecular approach (PCR-DGGE) to diet analysis in young Antarctic krill *Euphausia superba*. *Marine Ecology Progress Series*, **319**, 155–165.
- Martin-Laurent F, Philippot L, Hallet S *et al.* (2001) DNA extraction from soils: old bias for new microbial diversity analysis methods. *Applied and Environmental Microbiology*, **67**, 2354–2359.
- Matejusova I, Doig F, Middlemas SJ *et al.* (2008) Using quantitative real-time PCR to detect salmon prey in scats of grey *Helichoerus grypus* and harbour *Phoca vitulina* seals in Scotland – an experimental field study. *Journal of Applied Ecology*, **45**, 632–640.
- Meehan MG, Jarman SN, McLean C, Schultz MB (2009) DNA evidence of whale sharks (*Rhincodon typus*) feeding on red crab (*Gecarcoidea natalis*) larvae at Christmas Island, Australia. *Marine and Freshwater Research*, **60**, 607–609.
- Meusnier I, Singer GAC, Landry JF *et al.* (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.
- Meyer M, Stenzel U, Myles S, Pruber K, Hofreiter M (2007) Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Research*, **35**, e97.
- Moore JW, Semmens BX (2008) Incorporating uncertainty and prior information into stable isotope mixing models. *Ecology Letters*, **11**, 470–480.
- Moreby SJ (1988) An aid to the identification of arthropod fragments in the faeces of gamebird chicks (galliformes). *Ibis*, **130**, 519–526.
- Moszczyńska A, Locke S, McLaughlin JD, Marcogliese D, Crease T (2009) Development of primers for the mitochondrial cytochrome c oxidase I gene in digenetic trematodes (Platyhelminthes) illustrates the challenge of barcoding parasitic helminths. *Molecular Ecology Resources*, **9**, 75–82.
- Murray RA, Solomon MG, Fitzgerald JD (1989) The use of electrophoresis for determining patterns of predation in arthropods. In: *Electrophoretic Studies on Agricultural Pests* (eds Loxdale HD, den Hollander J), pp. 467–483. Clarendon Press, Oxford.
- Murray DC, Bunce M, Cannell BL *et al.* (2011) DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches. *PLoS One*, **6**, e25776. doi: 10.1371/journal.pone.0025776.
- Muyzer G, De Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, **59**, 695–700.
- Ørum H, Nielsen PE, Egholm M, Berg RH, Buchardt O, Stanley C (1993) Single-base pair mutation analysis by PNA directed PCR clamping. *Nucleic Acids Research*, **21**, 5332–5336.
- Ostrom PH, Colunga-Garcia M, Gage SH (1997) Establishing pathways of energy flow for insect predators using stable isotope ratios: field and laboratory evidence. *Oecologia*, **109**, 108–113.
- Pääbo S, Irwin DM, Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic amplification. *Journal of Biological Chemistry*, **265**, 4718–4721.
- Passmore AJ, Jarman SN, Swadling KM *et al.* (2006) DNA as a dietary biomarker in Antarctic krill, *Euphausia superba*. *Marine Biotechnology*, **8**, 686–696.
- Peano C, Lesignoli F, Gulli M *et al.* (2005) Development of a peptide nucleic acid polymerase chain reaction clamping assay for semiquantitative evaluation of genetically modified organism content in food. *Analytical Biochemistry*, **344**, 174–182.
- Pegard A, Miquel C, Valentini A *et al.* (2009) Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from feces. *Journal of Agricultural and Food Chemistry*, **57**, 5700–5706.
- Perkel J (2011) Making contact with sequencing's fourth generation. *BioTechniques*, **50**, 93–95.
- Pimm SL (2002) *Food Webs*. University of Chicago Press, Chicago, Illinois.
- Poinar HN, Hofreiter M, Spaulding WG *et al.* (1998) Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science*, **281**, 402–406.
- Poinar HN, Kuch M, Sobolik KD *et al.* (2001) A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences, USA*, **98**, 4317–4322.
- Polz MF, Cavanaugh CM (1998) Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, **64**, 3724–3730.
- Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.
- Pons J (2006) DNA-based identification of preys from non-destructive, total DNA extractions of predators using arthropod universal primers. *Molecular Ecology Notes*, **6**, 623–626.
- Ponsard S, Arditi R (2000) What can stable isotopes ( $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}$ ) tell about the food web of soil macro-invertebrates? *Ecology*, **81**, 852–864.
- Porazinska DL, Giblin-Davis RM, Faller L *et al.* (2009) Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources*, **9**, 1439–1450.

- Porazinska DL, Sung W, Giblin-Davis RM, Thomas WK (2010) Reproducibility of read numbers in high-throughput sequencing analysis of nematode community composition and structure. *Molecular Ecology Resources*, **10**, 666–676.
- Prokopowich CD, Gregory TR, Crease TJ (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, **46**, 48–50.
- Qiu XY, Wu LY, Huang HS *et al.* (2001) Evaluation of PCR-generated chimeras: mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology*, **67**, 880–887.
- Quince C, Lanzen A, Davenport RJ *et al.* (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**, e38.
- Rayé G, Miquel C, Coissac E *et al.* (2011) New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: chamois diet in autumn as a case study. *Ecological Research*, **26**, 265–276.
- Read DS, Sheppard SK, Bruford MW, Glen DM, Symondson WOC (2006) Molecular detection of predation by soil microarthropods on nematodes. *Molecular Ecology*, **15**, 1963–1972.
- Riaz T, Shehzad W, Viari A *et al.* (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, e145. doi: 10.1093/nar/gkr732.
- Riemann L, Alfredsson H, Hansen MM *et al.* (2010) Qualitative assessment of the diet of European eel larvae in the Sargasso Sea resolved by DNA barcoding. *Biology Letters*, **6**, 819–822.
- Rollo F, Ubaldi M, Ermini L, Marota I (2002) Ötzi's last meals: DNA analysis of the intestinal content of the Neolithic glacier mummy from the Alps. *Proceedings of the National Academy of Sciences, USA*, **99**, 12594–12599.
- Rothman JM, Chapman CA, Hansen JL, Cherney DJR, Pell A (2009) Rapid assessment of the nutritional value of foods eaten by mountain gorillas: applying near-infrared reflectance spectroscopy to primatology. *International Journal of Primatology*, **30**, 729–742.
- Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution and biogeography of Paeonia (Paeoniaceae). *American Journal of Botany*, **84**, 1120–1136.
- Schmidt JEU, Almeida JRM, Rosati C, Arpaia S (2009) Identification of trophic interactions between *Macrolophus caliginosus* (Heteroptera: Miridae) and *Myzus persicae* (Homoptera: Aphididae) using real time PCR. *BioControl*, **54**, 383–391.
- Shehzad W, Riaz T, Nawaz MA *et al.* (2012) A universal approach for carnivore diet analysis based next generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*. In press.
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1035–1045.
- Sheppard SK, Harwood JD (2005) Advances in molecular ecology: tracking trophic links through predator-prey food-webs. *Functional Ecology*, **19**, 751–762.
- Sheppard SK, Bell JR, Sunderland KD, Fenlon J, Skirvin DJ, Symondson WOC (2005) Detection of secondary predation by PCR analyses of the gut contents of invertebrate generalist predators. *Molecular Ecology*, **14**, 4461–4468.
- Simon C, Frati F, Beckenbach A *et al.* (1994) Evolution, weighing, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.
- Soininen EM, Valentini A, Coissac E *et al.* (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Frontiers in Zoology*, **6**, 16.
- Solomon MG, Fitzgerald JD, Murray RA (1996) Electrophoretic approaches to predator-prey interactions. In: *The Ecology of Agricultural Pests: Biochemical Approaches* (eds Symondson WOC, Liddell JE), pp. 457–468. Chapman & Hall, London.
- Sonnenberg R, Arne W, Nolte AW, Tautz D (2007) An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Frontiers in Zoology*, **4**, 6.
- Staudacher K, Wallinger C, Schallhart N, Traugott M (2011) Detecting ingested plant DNA in soil-living insect larvae. *Soil Biology & Biochemistry*, **43**, 346–350.
- Stech M, Kolvoort E, Loonen MJJE *et al.* (2010) Bryophyte DNA sequences from faeces of an arctic herbivore, barnacle goose (*Branta leucopsis*). *Molecular Ecology Resources*, **11**, 404–408.
- Sunderland KD, Powell W, Symondson WOC (2005) Populations and communities. In: *Insects as Natural Enemies: A Practical Perspective* (ed. Jervis MA), pp. 299–434. Springer, Berlin.
- Sutherland RM (2000) *Molecular Analysis of Avian Diets*. PhD Thesis, University of Oxford, UK.
- Symondson WOC (2002) Molecular identification of prey in predator diets. *Molecular Ecology*, **11**, 627–641.
- Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution*, **14**, 323–327.
- Taberlet P, Coissac E, Pompanon F *et al.* (2007) Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, **35**, e14.
- Taberlet P, Prud'homme S, Campione E *et al.* (2012) Extraction of extracellular DNA from large amount of soil for metabarcoding studies. *Molecular Ecology*, in press.
- Tollit DJ, Schulze AD, Trites AW *et al.* (2009) Development and application of DNA techniques for validating and improving pinniped diet estimates. *Ecological Applications*, **19**, 889–905.
- Traugott M, Pazmandi C, Kaufmann R, Juen A (2007) Evaluating <sup>15</sup>N/<sup>14</sup>N and <sup>13</sup>C/<sup>12</sup>C isotope ratio analysis to investigate trophic relationships of elaterid larvae (Coleoptera: Elateridae). *Soil Biology & Biochemistry*, **39**, 1023–1030.
- Traugott M, Bell JR, Raso L, Sint D, Symondson WOC (2012) Generalist predators disrupt parasitoid aphid control by direct and coincidental intraguild predation. *Bulletin of Entomological Research*, in press. doi: 10.1017/S0007485311000551.
- Troedsson C, Simonelli P, Nagele V, Nejstgaard JC, Frischer ME (2009) Quantification of copepod gut content by differential length amplification quantitative PCR (dla-qPCR). *Marine Biology*, **156**, 253–259.
- Valentini A, Miquel C, Nawaz MA *et al.* (2009a) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. *Molecular Ecology Resources*, **9**, 51–60.
- Valentini A, Pompanon F, Taberlet P (2009b) DNA barcoding for ecologists. *Trends in Ecology and Evolution*, **24**, 110–117.

- Vanderklift MA, Ponsard S (2003) Sources of variation in consumer diet  $\delta^{15}\text{N}$  enrichment: a meta-analysis. *Oecologia*, **136**, 169–182.
- Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples—a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, **5**, 12.
- Vestheim H, Deagle BE, Jarman SN (2011) Application of blocking oligonucleotides to improve signal-to-noise ratio in a PCR. *Methods in Molecular Biology*, **687**, 265–274.
- Walrant A, Loreau M (1995) Comparison of iso-enzyme electrophoresis and gut content examination for determining the natural diets of the groundbeetle species *Abax ater* (Coleoptera: Carabidae). *Entomologia Generalis*, **19**, 253–259.
- Weber DC, Lundgren JG (2009) Detection of predation using qPCR: effect of prey quantity, elapsed time, chaser diet, and sample preservation on detectable quantity of prey DNA. *Journal of Insect Science*, **9**, e41.
- Williams HE, Ormerod SJ, Bruford MW (2006) Molecular systematics and phylogeography of the cryptic species complex *Baetis rhodani* (Ephemeroptera, Baetidae). *Molecular Phylogenetics and Evolution*, **40**, 370–382.
- von Wintzingerode F, Gobel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, **21**, 213–229.
- Zeale MRK, Butlin RK, Barker GLA, Lees DC, Jones G (2011) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, **11**, 236–244.
- Zhang GF, Lü ZC, Wan FH, Lövei GL (2007) Real-time PCR quantification of *Bemisia tabaci* (Homoptera: Aleyrodidae) B-biotype remains in predator guts. *Molecular Ecology Notes*, **7**, 947–954.
- 
- F.P. studies adaptation through population genomics approaches and develops molecular markers for describing biodiversity. B.D. works on the application of DNA-based identification methods in dietary studies and also on stickleback evolutionary genomics. W.O.C.S. leads a research group developing molecular techniques for studying predator–prey interactions. D.B. studies the diet of British reptiles using DNA-based methods. S.J.’s research applies DNA-based techniques and bioinformatics to better understand key aspects of marine animal populations. P.T. is interested in conservation genetics and develops molecular methods for studying non-invasive samples, including the description of biodiversity with environmental DNA.
-