

Invited Review

SPECIAL ISSUE: Interactions between White-Tailed Deer and Invasive Plants in North American Forests

Reconstructing a herbivore's diet using a novel *rbcL* DNA mini-barcode for plants

David L. Erickson¹, Elizabeth Reed¹, Padmini Ramachandran¹, Norman A. Bourg², William J. McShea^{3*} and Andrea Ottesen¹

¹ Molecular Methods and Subtyping Branch, Division of Microbiology, Centre for Food Safety and Applied Nutrition (CFSAN), FDA, 5001 Campus Drive, College Park, MD 20740, USA

² U.S. Geological Survey, National Research Program - Eastern Branch, 12201 Sunrise Valley Drive, MS 430, Reston, VA 20192, USA

³ Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, VA 22630, USA

Received: 26 August 2016; **Editorial decision:** 28 March 2017; **Accepted:** 13 April 2017; **Published:** 21 April 2017

Guest Editor: David Gorchov

Citation: Erickson DL, Reed E, Ramachandran P, Bourg NA, McShea WJ, Ottesen A. 2017. Reconstructing a herbivore's diet using a novel *rbcL* DNA mini-barcode for plants. *AoB PLANTS* 9: plx015; doi:10.1093/aobpla/plx015

Abstract. Next Generation Sequencing and the application of metagenomic analyses can be used to answer questions about animal diet choice and study the consequences of selective foraging by herbivores. The quantification of herbivore diet choice with respect to native versus exotic plant species is particularly relevant given concerns of invasive species establishment and their effects on ecosystems. While increased abundance of white-tailed deer (*Odocoileus virginianus*) appears to correlate with increased incidence of invasive plant species, data supporting a causal link is scarce. We used a metabarcoding approach (PCR amplicons of the plant *rbcL* gene) to survey the diet of white-tailed deer (fecal samples), from a forested site in Warren County, Virginia with a comprehensive plant species inventory and corresponding reference collection of plant barcode and chloroplast sequences. We sampled fecal pellet piles and extracted DNA from 12 individual deer in October 2014. These samples were compared to a reference DNA library of plant species collected within the study area. For 72% of the amplicons, we were able to assign taxonomy at the species level, which provides for the first time—sufficient taxonomic resolution to quantify the relative frequency at which native and exotic plant species are being consumed by white-tailed deer. For each of the 12 individual deer we collected three subsamples from the same fecal sample, resulting in sequencing 36 total samples. Using Qiime, we quantified the plant DNA found in all 36 samples, and found that variance within samples was less than variance between samples ($F = 1.73$, $P = 0.004$), indicating additional subsamples may not be necessary. Species level diversity ranged from 60 to 93 OTUs per individual and nearly 70% of all plant sequences recovered were from native plant species. The number of species detected did reduce significantly (range 4–12) when we excluded species whose OTU composed <1% of each sample's total. When compared to the abundance of native and non-natives plants inventoried in the local community, our results support the observation that white-tailed deer have strong foraging preferences, but these preferences were not consistent for species in either class. Deer forage behaviour may favour some exotic species, but not all.

Keywords: Deer diet; exotic plants; fecal DNA; forest; metagenomics; North America; oaks; white-tailed deer.

* Corresponding author's e-mail address: mcsheaw@si.edu

Published by Oxford University Press on behalf of the Annals of Botany Company 2017. This work is written by US Government employees and is in the public domain in the US.

Introduction

Trophic interactions among species are thought to be a critical regulatory mechanism affecting species fitness and community composition (Paine 1966; Shears and Babcock 2003). These interactions can take many forms, the most familiar of which include predator-prey (Letnic *et al.* 2009), host-pathogen (Hajishengallis *et al.* 2012) and plant-herbivore (Creed 2000). These interactions appear to be part of a broader cascade of interaction effects, all with significant impacts on individual species and the communities they inhabit (Cumplido *et al.* 2016). While invertebrate herbivory has been widely studied and implicated in structuring plant communities (Janzen 1970; Maron and Crone 2006; McShea *et al.* 2007), large herbivores can also shift plant community composition in both forests (Frerker *et al.* 2014) and grasslands (Augustine and McNaughton 1998), although their role is less well understood. Improved quantification of ungulate diet will greatly contribute to our understanding of how foraging behaviour sustains or degrades natural communities, and importantly, how the removal of species from communities through extinction and extirpation may have a cascading effect on natural communities.

Increasingly, molecular genetic tools to quantify trophic interactions have been explored from a technical perspective (Hibert *et al.* 2013; Ficetola *et al.* 2016). These studies follow the pathway of metagenomic analyses first pioneered in environmental and human microbiome studies of bacteria (Lane *et al.* 1985; Hugenholtz *et al.* 1998), using PCR amplicon data from DNA samples containing a mixture of species, and large reference databases for those genes, notably 16S in microbial communities (Baker and Banfield 2003; Sogin *et al.* 2006; Turnbaugh *et al.* 2009). In animal studies the CO1 gene, the official DNA barcode in animals, has been utilized (Deagle *et al.* 2009; Leray *et al.* 2013), while in herbivores, different markers have been used (Pompanon *et al.* 2012; García-Robledo *et al.* 2013). Several program packages have been developed to help streamline the process of taxonomic identification of the sequences in mixed samples (such as fecal samples containing plant DNA), notably Qiime (Caporaso *et al.* 2010). Most studies in eukaryotes have focused on accurately measuring diversity (Leray and Knowlton 2015; Goldberg *et al.* 2015) and have begun to quantify interactions between species (Murray *et al.* 2011). There have been comparatively fewer studies that directly engage in hypothesis testing (Leray *et al.* 2015; Kartzinel *et al.* 2015; Soininen *et al.* 2009), compared to bacterial metagenomics (Blaser 2014; Thomsen and Willerslev 2015). But numerous reviews have highlighted the utility of these methods in

ecology (Valentini *et al.* 2009; Bourlat *et al.* 2013; Clare 2014), and increasing numbers of reports implementing metagenomic methods are being published.

Studies that use metagenomics to diagnose interactions between different trophic levels in natural communities are rare (e.g. Kartzinel *et al.* 2015; Murray *et al.* 2011; Soininen *et al.* 2009), particularly compared to the widespread use of metagenomics to quantify organismal diversity in an array of environments (Taberlet *et al.* 2012). Nonetheless, using metagenomics in trophic interactions has been well reviewed (Valentini *et al.* 2009; Bourlat *et al.* 2013; Clare 2014). Most metagenomic studies of inter-trophic interactions have focused on predator diets, (e.g. Deagle *et al.* 2009; 2010; Bourlat *et al.* 2013; Leray *et al.* 2013, 2015), with few focusing on herbivore diet (Soininen *et al.* 2009; Kartzinel *et al.* 2015). Yet the power that metagenomics offers in revealing the strength of trophic interactions suggests that this will become an increasingly common tool, particularly for quantifying animal diet when the animals' feeding behaviour is not directly observable. Deducing herbivore diet is relatively simple, as the DNA of the herbivore itself will not confound the bioinformatic analysis of the plants they consume. This is in contrast to carnivore diets where the DNA of the carnivore may be PCR amplified along with the diet of its prey (Leray *et al.* 2013). Yet metagenomic studies investigating trophic interactions among herbivores and their plant community have also been rare, especially relative to the role that herbivory plays in shaping natural plant communities (Côté *et al.* 2004, 2014; McShea 2012; Holm *et al.* 2013; Frerker *et al.* 2014).

Historically, observations of herbivore diet have largely taken three forms: direct observation (Holechek *et al.* 1982), exclusion experiments (Bowers 1993; Parker *et al.* 2006) and micro-histological analysis of animal scats (Alipayo *et al.* 1992; Kowalczyk *et al.* 2011). All of these methods have various limitations, but the use of genetics to deduce animal diet from feces was challenging because the traditional sequencing technology could only deduce one species at a time with accuracy, thereby being limited to use only with highly specialized herbivores (García-Robledo *et al.* 2013). However, for generalist herbivores with broad diets, the recent advances in massively parallel Next-Generation Sequencing methods enables us to simultaneously sequence DNA from all plant species in a mixture (Soininen *et al.* 2009).

In the forests of North America, exotic insects and plants have caused significant ecosystem alteration and a body of theory has developed around the mechanisms by which some exotic species become invasive (Sanders *et al.* 2003). There are multiple hypotheses for invasive species success that rely on preferential consumption of

native plant species. For example, the enemy-release hypothesis (Keane and Crawley, 2002) proposes that native species have the full complement of natural enemies (herbivores), but invasive species do not; likewise the novel defense hypothesis postulates exotic plants offer novel, and therefore more effective, plant defenses against herbivores (Enge *et al.* 2012; Macel *et al.* 2014). Deer are known to feed selectively on forest plants (Rooney 2001), but direct observation of deer behaviour is difficult and recent experiments have instead used a common garden approach to determine preference (Averill *et al.* 2016). But such experiments have limited scope due to the small number of plants that can be compared relative to the natural diet breadth of most herbivores. Effective means to deduce the diet of deer in wild populations would be critical to understanding their impact on invasive plants.

White-tailed deer (*Odocoileus virginianus*) are a widespread and often important herbivore in eastern deciduous forests of North America (McShea 2012). Throughout most of eastern North America, present densities of deer exceed historical levels (McCabe and McCabe 1997). These high densities have been implicated in changing the composition of herbaceous and woody plant communities in many forest ecosystems (Vavra *et al.* 2007), shifting the chemical composition of soils and successional pathways (Pastor and Naiman 1992), and facilitating the invasion of exotic plant species (Shen *et al.* 2016). However, although the abundance of deer has corresponded with increased rates of invasive plant abundance, the mechanisms behind this relationship remain uncertain (Averill *et al.* 2016). Using metagenomics to help determine the forage species and their relative frequency in wild populations is of major economic and conservation consequence.

The availability of reference sequence data for plants may be one limitation that prevents the more widespread application of metagenomics to the study of herbivore diets. The Smithsonian Conservation Biology Institute (SCBI) is the site of one of the few natural habitats where most plants within a study area have been sequenced to obtain a genetic fingerprint through DNA barcoding (Valentini 2009; Erickson *et al.* 2014). In this study, we use a metagenomic approach of applying a PCR-based mini-barcode marker in conjunction with the analysis package Qiime (Caporaso *et al.* 2010) to address several questions affecting the diagnosis of animal diets in natural environments, namely: (i) what is the efficacy of using DNA recovered from fecal samples to produce PCR products that can distinguish plant species?; (ii) Do repeat samples of the same fecal pile indicate caution is needed when sampling fecal DNA for diet information?; (iii) Can a single locus *rbcl* barcode distinguish species in

a reference DNA library and be used to accurately assign fecal DNA to tree, shrub and herb species? and (iv) can observational and relative abundance data from the number of operational taxonomic units (OTUs) and barcoding be used to infer herbivore diet selection for or against invasive plants?

Methods

Study organisms and study site

This study focused on the diet of white-tailed deer (*Odocoileus virginianus*) at the SCBI in Front Royal, VA, adjacent to the northern end of Shenandoah National Park. Within SCBI our sampling focused around a 26 ha Large Forest Dynamics Plot (LFDP) because the vascular plant flora was exhaustively sampled and genotyped at the accepted DNA barcode loci (CBOL Plant Working Group 2009). Those DNA barcode reference plant sequences were deposited in GenBank, and are available under accessions: KP642769–KP644136. Vouchers for each plant included in the reference set are stored at the National Museum of Natural History's herbarium. We list plant species as native, introduced and invasive. The terminology of exotic plants is problematic (Richardson *et al.* 2000; Valéry *et al.* 2008; Pereyra 2016). We differentiate between two types of exotic plants; introduced species that persist within native plant communities, and invasive species that are undergoing rapid spread in range and habitat associations, and tend to rapidly increase in numbers once established. We referenced the USDA to identify introduced species (USDA, NRCS 2017) and the state of Virginia to identify those introduced species that are considered invasive in our region (VA Dept. Cons. and Rec. 2015). When plants were only identified to genus and the genus contained a mixture introduced and native species we designated the genus according to the most abundant species. For example, the genus *Agrotis*, *Festuca* and *Poa* were considered introduced and genus *Acer* was considered native.

Sampling of the vascular plant flora was accomplished by collecting leaf material and voucher specimens for all woody vascular plant species >1 cm DBH, vines, and herbs present on the LFDP to produce a complete DNA barcode library for this representative mature secondary forest. This was achieved with a three-pronged survey approach, encompassing a subset of 330 one m² woody seedling plots in place on the LFDP for systematic sampling, intensive collection within a 4-ha long-term deer enclosure located within the LFDP, and walking searches for unrecorded species conducted across the LFDP. Whenever possible, at least three individuals of a given species were sampled for leaf tissue, collected and

subsequently genotyped at the four sequence regions of *rbcl*, *matK*, *trnH-psbA* and ITS (Kress et al. 2009). The DNA barcodes from these collections were combined with the previously generated barcodes from 66 species of woody plants, vines and herbs collected previously on the LFDP, as well as the 62 species of woody trees and shrubs sampled during the initial plot census (Bourg et al. 2013, Erickson et al. 2014).

The deer diet, as estimated from the fecal samples, was compared to the relative abundance of plants within the study area. In June/July 2009, we sampled 160 plots within the study area to estimate species richness and relative abundance for the site. We selected 40 random points within the mature forest and arrayed four 1 × 1 m plots around each selected point. At each plot we identified all woody and herbaceous stems (up to 2 m height) to species and recorded abundances for all species (number of individuals for woody and forb species, and percent cover for graminoids and ferns). For this paper, we used the presence of species in each 1 × 1 m plot to calculate the proportion of plots containing the species and thereby the relative abundance of each species detected.

Fecal sampling and DNA isolation

Twelve deer fecal samples were collected in the field on October 10–15, 2014 and collections were within forested areas within 200 m of the LFDP. Samples were only collected from obviously ‘fresh’ fecal piles, as evidenced by no degradation of pellet form and all pellets moist on the surface. The pellets were handled with sterile gloves and ~10 g were placed in a marked ziplock bag and homogenized within the bag. The steps by which these samples were processed is outlined in Fig. 1. Upon returning to lab, three samples, ~0.1 g each of fecal material, were then subsampled from each bag and placed into 2 mL polystyrene tubes with screw caps and O-ring seals to prevent leakage and contamination, resulting in 36 total samples. Each collected sample was mixed with 800 µL of Zymo Xpedition™ Lysis/Stabilization Solution, then thoroughly mixed by vortexing with the Lysis buffer on a laboratory mixer to ensure complete suspension of the buffer and sample. Following mixing, samples were frozen and stored at –20 °C in Front Royal, Virginia until DNA isolation in the laboratory.

DNA isolation was conducted in the Laboratories for Analytical Biology at the Smithsonian Institution (LAB-SI). Frozen samples were transferred to LAB-SI and kept at –20 °C. For DNA extraction, the samples were thawed and then processed utilizing the Zymo ZR Fecal DNA MiniPrep kit to recover and purify DNA. After thawing, the 36 samples were immediately centrifuged at maximum speed (14 000 rpm) in a mini-centrifuge for 5 min, after

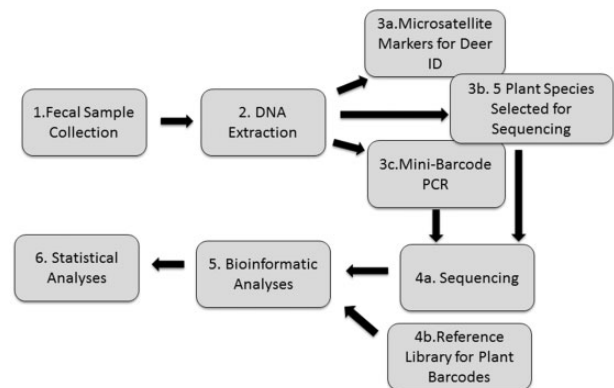


Figure 1. Flow chart of procedures for collecting, extracting and quantifying DNA used to identify deer forage species and deer identification for 12 samples. (1) Samples were collected within a 1 km² at Smithsonian’s Conservation Biology in Front Royal, VA in October 2014. (2) The DNA extraction was conducted using Zymo Fecal DNA Mini-prep Kit. (3a) Known microsatellite markers were used verify each of the 12 samples were from separate individuals. (3b) In parallel to the PCR based mini-barcode approach, WGS data was generated and analysed using five whole chloroplast sequences in CLC Bio. (3c) PCR was conducted using primers designed for mini-Barcode *rbcl* region. (4a) Sequencing of amplicons was performed on an Illumina MiSeq V2. (4b) Voucher DNA Barcode samples from plant species collected in region were obtained from GenBank. (5) Data quality control was performed in CLC Bio Genomics Workbench (ver. 7.4) while bioinformatics analysis, where sequences were assigned taxonomic identity, were performed using Qiime. (6) Primary statistics contrasting animal diet included PERMANOVA and ANOSIM which utilized OTU tables from Qiime analysis.

which 400 µL of supernatant was used for downstream DNA isolation. Use of the kit involved separating solids from the lysis buffer, binding of the DNA to silica in a spin column and elution in 100 µL of 10 mM Tris-1 mM EDTA (pH 8.0) buffer. The eluted DNA was quantified to ensure its capture, and then stored at –20 °C until used in PCR. We employed a negative control, composed of sterile water, that was processed in parallel to the 12 fecal samples. The supernatant for the negative control was then used in both the PCR and WGS sequencing (see below).

To ensure that each of fecal samples were derived from different individuals, we used two microsatellites developed for genetic fingerprinting of white-tailed deer (Anderson et al. 2002), using markers denoted as N and Q. PCR amplification using the DNA recovered from fecal samples was conducted for all 36 DNA samples following Anderson et al. (2002) and the different size alleles for each animal were resolved using an Agilent TapeStation with high sensitivity tapes.

PCR

A mini-barcode approach was chosen for PCR amplification of plant DNA recovered from the 36 fecal samples.

The mini-barcodes were derived from the standard *rbcl* plant DNA barcode region using the accepted reverse primer from standard DNA barcodes and a forward primer built following multiple sequence alignment of over 500 *rbcl* sequences derived from 30 separate orders (including the Polypodiales, Pinales and the remainder from the most speciose clades of Magnoliophyta), and identification of a sub-region with no mismatches in the critical first five bases of the 3' end for the set of 500 sequences used in our design. While this analysis did not explicitly quantify potential primer bias, which is possible, sequence alignment and *in-silico* testing using complete chloroplast genomes for species from Asteraceae, Fagales, Malpighiales and Poaceae suggested limited primer mismatch. The new forward locus specific primer sequence we used was: 5'-CTTACCAGYCTTGATCGTTACAAA GG-3', and the reverse primer sequence which is the reverse primer for the official *rbcl* DNA barcode region is 5'-GTAAATCAAGTCCACRCG-3' (Kress and Erickson 2007), which is the accepted and widely implemented standard primer used in *rbcl* plant barcoding (CBOL Plant Working Group 2009). The amplicon size from this primer pair is a constant 379 base pairs across our alignment of the 500 *rbcl* sequences. Primers were synthesized as fusion primers with the Illumina Nextera adaptors at the 5' end, following the 16S Metagenomic Protocol, as Forward fusion primer: 5'-TCGTGCGCA GCGTCAGATGTGTATAAGAGACAG-[locus specific sequence], and Reverse fusion primer: 5'-GTCTCGTGGGC TCGGAGATGTGTATAAGAGACAG-[locus specific sequence].

Extracted DNA was used for conducting PCR to amplify *rbcl* minibarcode. All DNA samples were subjected to identical PCR conditions. The reaction cocktail consisted of 12.5 μ L of 2 \times EmeraldAmp MAX PCR Master Mix, 0.1 μ L each of 100 μ M forward and reverse primer, 1 μ L of DNA and 11.3 μ L of ultrapure water. Thermocycling was conducted in a ABI 2720 with the program: 1 cycle of 95 $^{\circ}$ C 4 min; 35 cycles of: 94 $^{\circ}$ C for 20 s, 55 $^{\circ}$ C for 30 s, 72 $^{\circ}$ C for 1 min; 1 final extension of 72 $^{\circ}$ C for 5 min; and a hold at 9 $^{\circ}$ C until the reaction was removed. Each DNA sample had its PCR reaction repeated once, and the two reactions were pooled in equal concentration as measured using a Qubit fluorometer prior to PCR cleanup. The negative control was used in this protocol to test for how laboratory processing affected results.

The Next Generation Sequencing library for the 36 fecal pellet PCR samples followed the Illumina 16S protocol (see Illumina 2013). Amplicons from the *rbcl* PCR were verified through gel electrophoresis on an Agilent TapeStation, then purified with Ampure beads (0.8 \times volume of beads to PCR volume) and washed with 80% Ethanol twice. DNA was eluted from beads by mixing

52.5 μ L of 10 mM Tris (pH 8.5) buffer with the beads, and recovering 50 μ L of clean DNA solution following separation of the beads with a magnet. Clean up was performed in a 96 well plate using a magnet plate from Ambion. The cleaned PCR product was then used in a second PCR, which added a unique combination of Nextera XT indexes (Illumina catalog number FC-131-1002) for each of the 36 samples. For each index PCR amplification we used: 5 μ L of cleaned PCR product, 5 μ L each of the Nextera XT index primers (S5XX and N7XX, respectively), 25 μ L of 2 \times EmeraldAmp MAX PCR Master Mix and 10 μ L of H₂O. Each reaction conducted on an ABI 2720 using the program: one cycle of 95 $^{\circ}$ C 3 min; 8 cycles of: 95 $^{\circ}$ C 30 s, 55 $^{\circ}$ C 30 s, 72 $^{\circ}$ C 30 s; 1 final extension of: 72 $^{\circ}$ C 5 min; followed by a hold at 9 $^{\circ}$ C until the reaction was removed. The index PCR was cleaned using a 1.12 \times volume of Ampure beads (56 μ L of Ampure beads mixed with 50 μ L PCR), and washed twice with 80% Ethanol while on a magnet stand. Clean index PCR was eluted from beads using 27.5 μ L of 10 mM Tris (pH 8.5) buffer, with 25 μ L of solution recovered.

Each cleaned index PCR was quantified using fluorometric assay with a Qubit 3.0 system, using 3 μ L of DNA in each measurement. Products were pooled in equal molar concentrations to a final 40 mM solution containing all 36 samples.

Evaluation of *rbcl*mini-barcode locus

To evaluate the ability of the *rbcl*-mini-barcode locus used to correctly identify plant species from specimens collected at SCBI, we tested different fractions of the standard *rbcl* barcode locus relative to the mini-barcode. This involved alignment of all *rbcl* reference sequences for SCBI ($n = 234$), including at least two individuals per species. That alignment was then trimmed to different lengths: 100, 200, 300 and 400 bp. We used a sliding window that was used to extract 100 bp segments from along the length of the reference DNA barcode that is 540 bp in length. The sliding window was in 20 bp increments, such that there were 23 different 100 bp segments, 18 different 300 bp segments, eight different 400 bp segments and three different 500 bp segments. Each of these segments, which represented a sampling of the aligned reference database, was used in the SpeciesIdentifier module in TaxonDNA v1.8 (Meier et al. 2006) to quantify rates of identification to different taxonomic levels. For each sequence segment tested, we recorded the rate of correct identification to species, genus and family level. The average value for identification to each of the three taxonomic levels was then plotted (Fig. 2). The values for species level identification for the mini-barcode and for the full length DNA barcode were

determined using TaxonDNA and plotted with the results for the sliding window of difference size selections.

Sequencing and post-sequence processing

The 40 mM sample of pooled, indexed PCR was diluted to 4 pM and sequenced using an Illumina MiSeq with a 500 bp V2 kit with 251 cycles. Sequences produced on the MiSeq had adaptor and index sequences trimmed on the instrument by bcl2fastq Conversion Software (v1.8.4; <http://support.illumina.com>), with each of the 36 fecal DNA samples resulting in separate forward and reverse sequences. For each of the 36 fecal DNA samples, the software package CLC Genomics Workbench (ver. 7.4) was used to trim sequences, remove locus specific PCR primer sequences and merge the forward and reverse strands. All sequences were quality filtered and trimmed before merging into contigs. The software was used to remove all sequences containing ambiguous bases, and trimmed all sequences at both ends using quality scores to a 0.03 limit (equivalent to a Q value of 30), with further detail on the trimming algorithm found here (http://resources.qiagenbioinformatics.com/manuals/clcgenomics-workbench/801/index.php?manual=Quality_trimming.html). Only sequences that could be merged were used in subsequent analyses, and we used the following parameters within CLC Bio Genomics Workbench (ver. 7.4): mismatch cost = 2, Gap cost = 4, Max unaligned = 0 Minimum score = 0; again for more description of the implementation of the parameters see: http://resources.qiagenbioinformatics.com/manuals/clcgenomics-workbench/801/index.php?manual=Merge_overlapping_pairs.html. For each of the 36 samples, we used CLC Bio Genomics Workbench (ver. 7.4) to randomly select 100 000 merged sequences for use in downstream analysis. The 100 000 randomly selected sequences were then exported in FASTA format for downstream use in assignment of plant identity.

Metagenomic analysis of fecal plant DNA

We define an OTU as a DNA sequence cluster that corresponds to a taxonomic group, either Family, Genus or Species (Blaxter et al. 2005). Whether an OTU can be assigned to a specific taxonomic level depends on the resolution of the mini-barcode sequence. We assigned species identifications to the OTU extracted from the fecal DNA using Qiime (Caporaso et al. 2010) in combination with the reference sequences for the LFDP to identify the species of plants consumed by the white-tailed deer. We followed the 454 SOP (see: <http://qiime.org/tutorials/tutorial.html>), which involved removal of chimera and OTU picking, assignment of taxonomy for each OTU based on the reference database and output

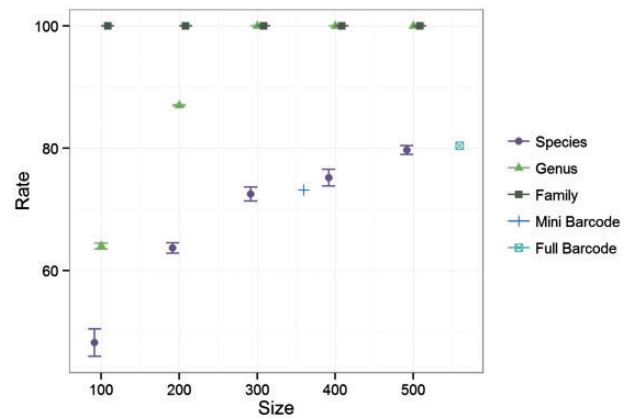


Figure 2. The rate of correct identification for increasingly larger portions of the *rbcl*-mini-barcode as applied to the reference sequence data from the SCBI plant community. The rate of correct identification for a given size range of *rbcl* is plotted on Y-axis, size of the tested size range of *rbcl* is plotted on X-axis. A 20 bp sliding window was applied to the 540 bp *rbcl* reference sequences and used to test how well increasingly larger fragments of *rbcl* distinguish at different taxonomic levels (family, genus, species). The points represent the average and standard deviation of the rate of identification for each size range tested. Also plotted are the rates of species identification for the *rbcl*-mini-Barcode used in this study, and for the full length *rbcl* DNA barcode that is recognized by CBOL.

of a BIOM table that contained all taxonomic and abundance data. The BIOM table was converted using `assign_taxonomy.py`, to list the taxonomy of identified OTU and the number of sequences that were in each OTU. We implemented the closed reference OTU algorithm, focusing only on assignment to the plant species found in our reference library for SCBI through the “`pick_closed_reference_otus.py`” option in Qiime. We used `uclust` as the cluster algorithm for generating OTU at a 97% similarity threshold. The reference sequences were only for the *rbcl* gene, and were trimmed to the same constant 379 bp corresponding to our mini-barcode and pared such that only a single sequence representing each species was in the database. The second step after OTU clustering was selecting a single representative sequence from each OTU to be used in assignment of taxonomy to each OTU. For assignment of the representative sequences to sequences in the reference database, we also used `uclust` at a 97% identity threshold. Following assignment of taxonomy to each OTU, we then output a JSON Biom table containing the taxonomic assignments of each OTU observed along with counts for the number of sequences found in each OTU summarized for each of the 36 samples. The Biom table was subsequently converted to a tab delimited table with individual counts of the plants observed in each sample. We summarized values for each of the 12 individual

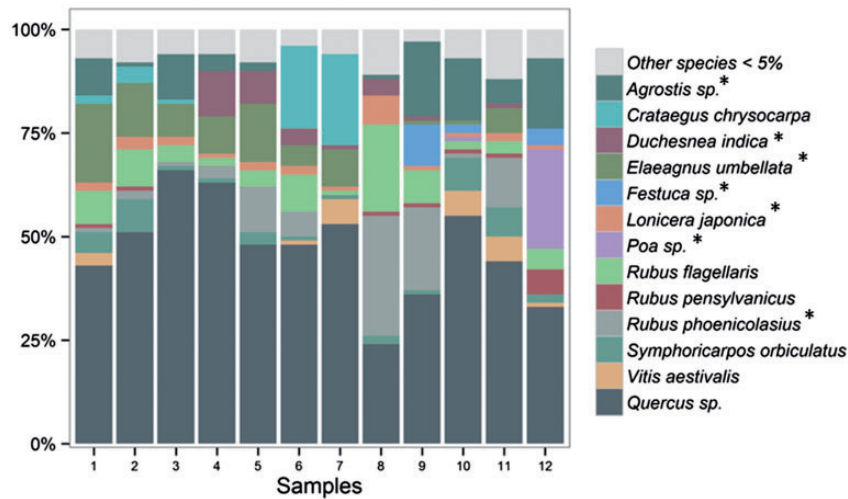


Figure 3. Stacked bar-graph showing the frequency of different plant species found in each of the 12 deer individuals. Only the 15 most common species, which comprise more than 95 % of all diversity are graphed. For each of the 12 individual deer, the three sub-replicates were collapsed into a single value by averaging counts of the plants observed. Exotic species are identified with an asterisk.

deer, see **Supporting Information—Table S1**, and this table was used to construct **Fig. 3**, a stacked bar-plot of the plant species observed for each of the 12 individual deer where the three replicates per deer were averaged. In **Supporting Information—Table S1**, each of the 12 samples were further summarized by including only species that were present at >1% representation within that sample. Figures were made in either excel (**Fig. 4**) or the R package using ggplot.

After construction of the Biom table, we then aligned the reference *rbCL* sequences to make a phylogenetic tree that could be used to estimate the relative genetic alpha and beta diversity among samples. Alignment of reference sequences was conducted using the alignment program MUSCLE (Edgar 2004). The aligned sequences were then applied to build a phylogeny of the sequences with FastTree (Price et al. 2009). This molecular phylogeny can be used to describe the diversity of plants eaten in each sample, such that we can quantify the diversity of all plants found within and between fecal samples (see Erickson et al. 2014 for discussion of use of *rbCL* barcode phylogeny for quantifying community phylogenetic diversity). While a molecular phylogeny for a wide diversity of angiosperms using only a mini-barcode is imperfect, the *rbCL* locus is one of the most informative loci in reconstructing phylogenetic relationships among plant groups (Erickson et al. 2014). Beta diversity, defined as genetic distance among the 36 samples, was calculated in Qiime using the “Beta_diversity.py” algorithm, which used the Biom table and the phylogenetic tree of reference samples to estimate a matrix of the phylogenetic distance among of each of the 36 samples using weighted UniFrac distance estimation (Lozupone and Knight 2005).

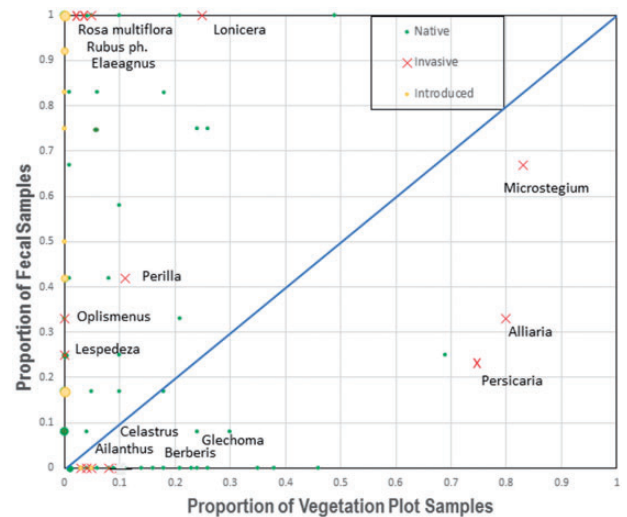


Figure 4. Relative occurrence of plant species in 12 fecal samples and in 1601 × 1 m vegetation plots in the same region. The dashed line indicates the equal detection of plants in both samples. Points above the line indicate higher occurrence in the fecal samples than the plant community and points below the line indicate lower occurrence in the samples than would be predicted from the plant community samples. Those points representing invasive exotic species are identified to genus. The figure is derived from values presented in **Supporting Information—Table S2**.

Whole genome sequencing

Because a number of published studies have suggested that PCR bias can lead to skewed estimates of community composition in microbes (e.g. Acinas et al. 2005), we submitted the one replicate from each of the 12 deer in addition to a negative control, for whole genomic shotgun (WGS) sequencing, using the TruSeq kit from

Illumina. For each of the 12 samples, 300 ng of DNA was sheared using Covaris shearing matrices and the size selected to include a range of 300–600 bp fragments using Ampure beads. We used a 0.8× volume of Ampure beads to remove larger fragments, and a second selection using 1.8× volume of Ampure beads to remove smaller fragments. Libraries were prepared using the TruSeq kit and the resulting DNA was sequenced using a MiSeq and a V2 500 bp kit. Analysis of the shotgun data followed the same quality filtering as above using CLC Bio Genomics Workbench (ver. 7.4), but reads were not merged and the forward and reverse sequences were kept separate. We then identified five complete chloroplast genomes that corresponded to species (only genus in the case of *Quercus*) observed in the PCR based assays (see above) which included (species and GenBank accession: *Quercus aliena* NC_026790.1, *Lonicera japonica* NC_026839.1, *Elaeagnus macrophylla* NC_028066.1, *Festuca arundinacea* NC_011713.2 and *Vitis aestivalis* NC_029454.1). We then used CLC Bio Genomics Workbench (ver. 7.4) to map the reads from each of the 12 WGS sequence sets back to each of the 5 complete chloroplasts using reference mapping algorithm. This allowed us to quantify if sequence data from WGS contained any information from the chloroplast that would allow us to identify the diet of the animals—with the expectation that sequence data from WGS would not suffer any of the problems that arise from PCR bias when it comes to identifying how much of each species the animals consume.

Native versus exotic species consumption

We analysed the diet selection of deer in three ways. First, we compared the relative abundance of plant species observed in the diet to the relative plant abundance as detected in an understory plant survey of the study area. For comparison between environmental samples and fecal DNA we combined species from the same genus when there was uncertainty in species identification at the seedling stage (e.g. *Lonicera*, *Acer*, *Rubus*).

Second, we constructed a bipartite graph using the OTU table from Qiime, and implemented the bipartite package in R (Dormann et al. 2009). To improve legibility of the graph, only the 25 most abundant plant species observed in deer diet were included in the bipartite analysis (these 25 species accounted for more than 95 % of all observations). For the 12 deer, each of the three replicates were averaged, so the graph contained 12 deer mapped to 25 plant species (Fig. 5). The bipartite graph visually represents the symmetry, or lack thereof, in diet choice for each of the 12 individual deer. We also used the null.t.test option within the Bipartite software

package to compare the symmetry of connections within our network to that of a randomly distributed network with the same number of observations using $n = 25$.

A third analysis was to construct a nonmetric MDS graph, using the multidimensional scaling function in the R “stats” package. The nMDS graph provides a visual representation of the similarity in diet of the 12 deer individuals, and shows how the three replicates per individual may group together. The distance table produced through the Beta_diversity.py analysis was used as input for the R package. We further included a random sample in the nMDS graph. The random sample was generated by partitioning the 100 000 OTU observations evenly among the 139 different plants that were observed in the diet of the 12 deer. Thus for the random sample, each of the plants in the matrix was assigned 719 counts, as if an individual deer ate every available plant at an equal rate. To evaluate whether sample origin significantly influenced total community composition, we employed non-parametric PERMANOVA, analysis of similarity (ANOSIM), and distance based redundancy analysis modules available in Qiime (using the Vegan R package), with 1000 permutations for each test. In this way, we contrasted within versus between variance in diet composition for the 12 deer and their sub-replicates.

Results

DNA isolation and PCR and sequencing

Use of the ZR Fecal DNA MiniPrep™ provided high DNA yield and successful PCR amplification with the *rbcl*-mini-barcode. The DNA concentration of samples ranged from 30 to 60 ng/μL and PCR using the *rbcl*-mini-barcode was successful for all 36 samples using Illumina fusion tagged primers. The number of sequences obtained for each sample ranged from 105 271 to 582 045, from which 100 000 sequences per sample were sampled for use in OTU assessment following Yoccoz et al. (2012).

The two microsatellite markers, developed by Anderson et al. (2002), were applied to each sample and no two of the 12 deer shared the exact same genotype, and thus each of the 12 individuals were unique.

DNA mini-barcode performance

The mini-barcode based on the *rbcl* gene was highly effective, both in PCR success, and in resolving species identity in the LFDP [see **Supporting Information—Table S1**]. The rates of correct identification for species, genera and families (Fig. 2) demonstrates that although the mini-barcode marker was nearly 30 % smaller than the full length, official DNA barcode for *rbcl*, it captured 90 % of the identification power of the full length DNA

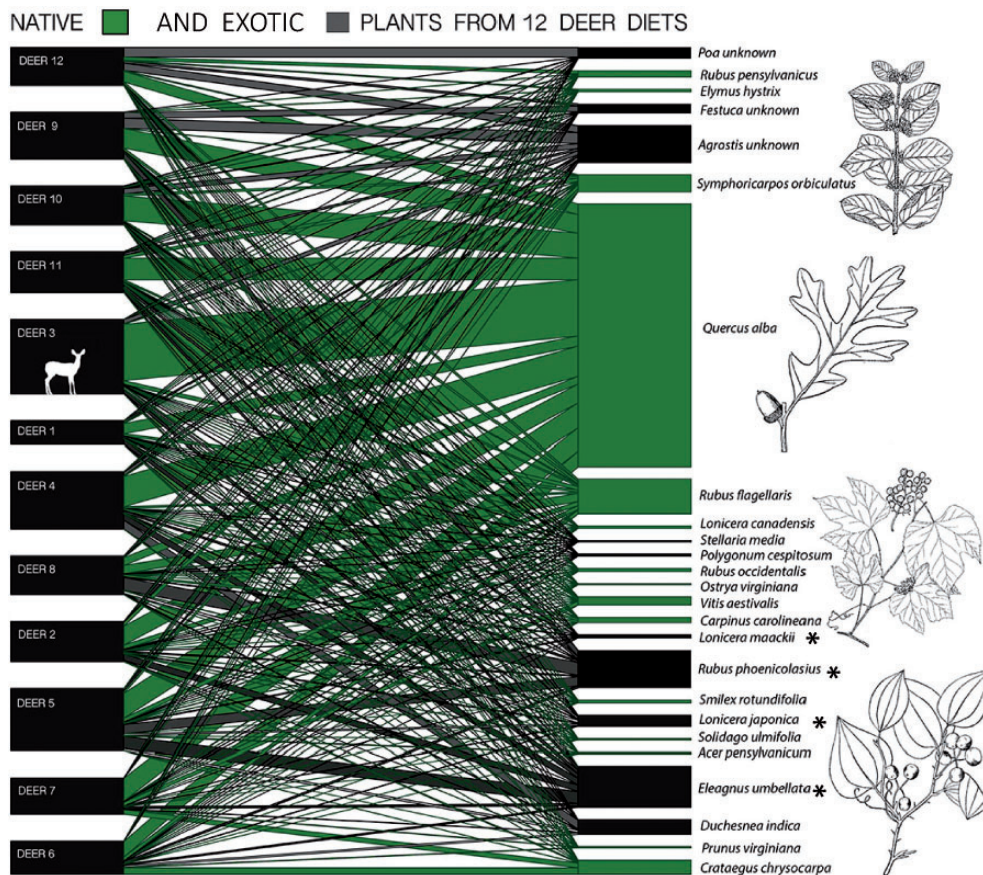


Figure 5. A bipartite graph showing the foraging network for 12 individuals of deer. Only the 25 most common plant samples (which capture more than 95 % of the observations) are mapped. Native and exotic species are indicated by green and black colours, respectively. The graph is quantitative with the width of the bar connecting deer to plants corresponding to the relative proportion of that plant found in each deer's diet. Invasive exotic species are identified with an asterisk.

barcode. The full length DNA barcode was able to distinguish among species 82 % of the time in the LFDP, whereas the mini-barcode correctly assigned identity 72 % of the time on the same set of species. The identification rate is a function of the database used, and we recognize that the 82 % identification rate is for species in our SCBI reference library and does not reflect identification rates for all plants. The 72 % identification rate reflects that 26 % of the species in the reference database were identified to the level of genus or family, but does not mean they will be arbitrarily assigned to another taxon. In total, all species could be identified to species ($n = 173$), Genus ($n = 46$) or Family ($n = 15$), and the 15 species that could only be identified to Family were all from the Poaceae. We note that reducing the threshold for identification did not affect rates of identification and the rate of correct identification at a 1 % and 2 % threshold were the same as a 3 % threshold. Figure 2 displays the rate of correct identification for the mini-barcode when applied to the SCBI community and shows how smaller to larger size samples of the

rbcl barcode provide increasing rates of correct identification for different taxonomic levels. The smallest size range of 100 bp provided accurate identifications to family, and at 300 bp to genus. The rate of identification for Species saturates at $\sim 80\%$ using the full-length DNA barcode; thus the observed rate of 72 % identification, at a 3 % sequence divergence for the mini-barcode, captures most of the diagnostic information in the gene. Our use of a 3 % threshold for discriminating among species also means that sequencing error is unlikely to result in incorrect taxonomic assignments. We note that the amplicon size of 379 is larger than some other mini-barcode markers, notably the P6-loop employed by Soinen et al. (2009), and Kartzinel et al. (2015). Just as more work to quantify PCR bias is needed, so too is work to establish whether there is loss of signal from more degraded diet items. This will be relevant if, for example, there are changes in diet preference over time, such that older, more degraded DNA reflect legitimate diet choice as animals alter their foraging preferences.

OTU diversity and structure

The deer diets had between 61 and 91 distinct OTU (mean = 71), although the abundance was highly skewed to a few plant species (Fig. 3). When we exclude species detected in <1% of the OTU, our species drops to 4–12 distinct OTUs per sample (mean = 8.3). The most abundant species found in their fecal samples were *Quercus* sp. (we could not distinguish among the DNA samples from the different oak species and consider them as a single OTU), *Elaeagnus umbellata*, *Rubus phoenicolasius*, *Agrostis* sp., *Rubus flagellaris*, *Symphoricarpon orbiculatus*, *Rubus pensilvanicus*, *Duchesnea indica*, *Festuca* sp., *Poa* sp. and *Crataegus chrysoarpa*. *Quercus* sp. alone accounted for 46% of all observations when summed across all samples, with no other species occurring at more than seven percent. The 12 most abundantly observed species accounted for 95% of all reads [Fig. 3 and see Supporting Information—Table S1]. Although there was a high degree of asymmetry in diet, there was a very high degree of selectivity to the full spectrum of plants available in their environment [see Supporting Information—Fig. S1]. In total, three plant families, accounted for a large majority of the diet, Fagaceae (46%), Poaceae (12%) and Rosaceae (29%), even though the abundance of other families (e.g., Asteraceae) is high in our census. The negative control produced a weak PCR amplicon which was sequenced with other samples, but no plant sequences were observed and nearly all sequences were assigned to bacterial reference samples.

Native versus exotic forage

The forest-wide inventory of plant species detected 335 vascular plant species within SCBI. The 1 × 1 m plot survey detected 103 species within the vicinity of the fecal collections [see Supporting Information—Table S2]. Of these species, 29 were detected in both the fecal samples and the 1 × 1 m plots [see Supporting Information—Table S2]. A comparison of plant relative abundance in fecal and plot samples found little convergence [Fig. 4 and see Supporting Information—Table S2]. Most species were either more abundant in fecal or plot samples with a significant mean difference between the paired samples (\bar{X} difference in samples = 0.40 ± 0.028 SE; $P < 0.001$; $n = 149$). For those plant species found in at least 10% of the plot surveys or fecal samples, there were 13 species that were at least twice as abundant in the plot surveys than the fecal samples and 92 species of plants that were at least twice as abundant in the fecal samples as the plot surveys. Most of the latter were species commonly found in old fields.

Native plants were closer to equivalence in the two samples than exotic plants (\bar{X} difference in abundance = 0.37 and 0.54, respectively; $df = 147$; $t = -2.32$, $P = 0.01$), but there was no pattern with respect to exotic species. Some were >25% more abundant in the feces than the plot samples (i.e. *Lonicera* sp., *Elaeagnus* sp., *Rubus phoenicolasius*, *Rosa multiflora*) and some were absent or rare from fecal samples but more common in plots (>0.25; i.e. *Alliaria petiolata*, *Lespedeza* sp., *Oplismenus hirtellus*, *Berberis*, *Periscaria perfoliata*). We distinguish between invasive and introduced exotic species [see Supporting Information—Table S2]. Counts of the 25 most abundant native versus exotic species from the species table showed that over 65% of all sequences assigned to OTU were assigned to native species (Fig. 5). However, when *Quercus* sp. were excluded, half of the remaining sequence data were assigned to non-native species, with 30% of OTUs being invasive exotic species and 20% introduced exotic species [see Supporting Information—Fig. S1].

Bipartite network

A quantitative network mapping of deer to the 25 most common plant species showed a high proportional occurrence of *Quercus* in the deer diet, but also demonstrated a very wide diet breadth (Fig. 5), including species from 12 different plant families. The distribution of species found in their diet was highly asymmetrical, when tested with null.t.test in the bipartite R package ($t = 21.07$, $P < 0.001$).

Nonmetric MDS

The nonmetric multidimensional scaling graph (Fig. 6) demonstrates the clustering of deer along the second MDS axis relative to the randomly distributed sample, reflecting that the diet of deer is highly non-random. All three methods used to quantify variance within and between samples, (PERMANOVA pseudo $F = 1.73$; $P = 0.004$; ANOSIM $r^2 = 0.27$, $P = 0.002$; and DBRDA pseudo $F = 1.66$, $P = 0.006$) found that the variance between samples was higher than variance among sub-samples. We likewise detected significant associations between community composition (using the Bray-Curtis distance metric) and sample origin, meaning sub-samples from the same fecal pile were more similar to each other than they were to samples derived from other fecal piles. As such, these results reflect that there is greater variance among individual deer samples than there is within sub-replicates, and that a single sample from each fecal pile was sufficient to describe the diversity of diet for that animal.

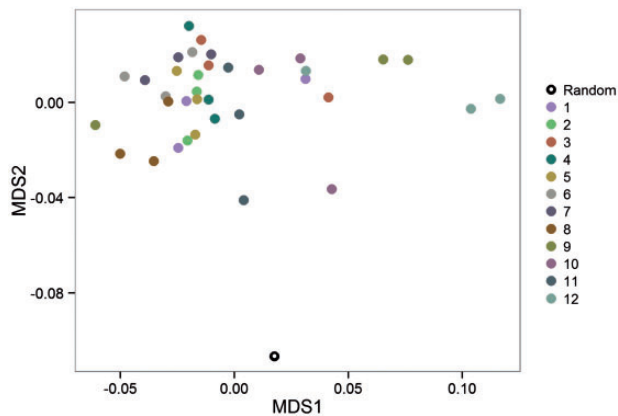


Figure 6. A nonmetric multidimensional scaling (nMDS) graph of the weighted UniFrac distances among the 36 deer samples is plotted. Each of the 36 points is labelled and a reference sample corresponding to completely random diet (see text) from the LFDP community is plotted for reference. The three sub-samples from each of the 12 individual deer are the same colour.

Whole genome sequencing

For all of 60 combinations of WGS reads (12 samples \times five chloroplast genomes) using 100 000 sequence reads per sample, we found that none of the reads mapped to the chloroplast genomes. The negative control run in parallel for the WGS library failed to produce a sequence library that could be sequenced, and was not included in the sequencing run. In contrast to the WGS reads from the fecal samples, we have observed in other analyses that WGS reads from plant tissues map to chloroplast references at a rate of 0.5–2% (unpublished data) using a reference mapping strategy in CLC Bio Genomics Workbench as we conducted here. Thus the failure to identify reads was likely attributable to plant DNA being very rare, and the fecal DNA being dominated by host and microbial DNA.

Discussion

This is a metagenomic study of forage selection by white-tailed deer on exotic and native plant species. Our observations are not conclusive, but do indicate selective feeding by deer (Fletcher et al. 2001; Keane and Crawley 2002; Frerker et al. 2014; Averill et al. 2016), and are particularly relevant given that a number of widespread exotic species were relatively rare in the samples examined. Our results correspond closely to those found by Averill et al. (2016, 2017) of some exotic plants, such as *Alliaria petiolata* and *Microstegium vimineum*, being under represented in deer fecal samples, and other exotic plants such as *Lonicera japonica* and *Rosa multiflora*, being over represented in the same fecal samples. We

could discern no common trait on which exotic plant species were not consumed by deer.

WGS versus PCR

We tested both a PCR amplicon approach and a whole-genome-shotgun (WGS) approach to determine if PCR approaches exhibit bias in the OTU that can be observed. A number of studies have suggested that PCR can often lead to altered estimates of the number and abundance of OTU in communities that are very diverse (Larsen et al. 2014; Wood and Salzberg 2014). Methods that use WGS to infer microbial diversity have been able to circumvent the potential bias observed in PCR (e.g. Rusch et al. 2007). In addition, WGS data from fecal samples has been used successfully to quantify the plant diet of a primate herbivore, *Pygathrix nemaeus* (Srivathsan et al. 2015) using very deep sequencing (over an order of magnitude greater than used in our study). However, our results found that a PCR amplicon strategy revealed a wide diversity of plant species, whereas the WGS approach was unable to recover and assign sequence data to whole genome reference for species that were commonly observed in the PCR amplicon analysis. When we then mapped WGS reads from each of the 12 deer individuals to the five complete chloroplast genomes using both MIRA and CLC Bio Genomics Workbench we found no sequences could be assigned to the plant chloroplasts, and that sequences were overwhelmingly microbial. These results strongly contrast with Srivathsan et al. (2016), and may be attributable to our focal herbivore species being a ruminant versus a primate, in addition to differences in the number of WGS sequences used. We note that Covaris shearing of the fecal DNAs may have further degraded the plant sequences which were likely small to begin with, making them too small to be recovered in the library processing steps after shearing. Alternatively, we observed a wide number of plant species using PCR amplicon methods [Fig. 3 and see Supporting Information—Table S1]. The *rbcL* mini-barcode recovered sequence data throughout the flowering plant clade, from ferns to Asterids and Rosids, suggesting little bias in recovery using that marker. The gene did have limits in assigning species designation to the grass family (Poaceae), a majority of the OTUs that could not be assigned to species were from this family.

Marker selection and identification rate

We employed a *rbcL* based mini-barcode PCR marker derived from a subset of the larger *rbcL* DNA barcode locus (CBOL Plant Working Group 2009). This mini-barcode captured a majority of the identification power (nearly

90%) of the complete DNA barcode (Fig. 2), but with 70% the size of the amplicon. Smaller amplicon size is known to improve recovery of increasingly degraded plant samples (Valentini *et al.* 2009), such as what may be found in scat used to deduce herbivore diet. The rate of identification within the SCBI plot for the mini-barcode was 72%, which is very similar to the overall rate of identification reported for the combined two locus plant DNA barcode advocated by CBOL (76% resolution) (CBOL Plant Working Group 2009). The rate of identification is also similar to the P6-loop advocated by Taberlet and colleagues (Valentini *et al.* 2009; Pompanon *et al.* 2012). However the much larger available database of *rbcl* sequence data due to its assignment as the official DNA barcode locus, and its historical use in plant molecular systematics means that many more species have been sequenced at this locus, and may be more likely uniquely identified using *rbcl*. We also observed that amplification of the *rbcl* mini-barcode was not affected by addition of a phusion primer adaptor, facilitating NGS library preparation. Lastly, although the P6 marker has been successfully used where larger PCR amplicons fail (e.g. Kartzinel *et al.* 2015), the marker is highly A-T rich and contains many short mononucleotide repeats that may affect sequencing error rates and hence correct taxonomic assignment. We believe that use of the *rbcl* mini-barcode reported here will best serve future studies that wish to diagnose plant-herbivore trophic interactions. We did not perform explicit *in-silico* testing of primer bias with our *rbcl* mini-barcode, and we note that such analysis is appropriate for subsequent analyses. Published reports that describe primer bias have focused on loss of relatively rare and disjunct species, and as such would not likely affect our conclusions. In addition, primer matching to common invasive species that were not common in our diet analysis, such as *Alliaria petiolata*, *Vitis rotundifolia* and *Rosa multiflora*, showed that the primers matched them perfectly and thus their prevalence in our diet observations was likely not due to PCR amplification bias.

There were some problematic assignments from the DNA process. These fall into two groups; the first are plants commonly associated with old fields that were found in fecal samples but not detected in our forest plot samples. This group includes several of the grasses, both native and exotic. The second instance is for eastern hemlock (*Tsuga canadensis*), which is a native species detected in two fecal samples but is not common in the areas and the closest known individuals would be around residential houses within a 2 km of the collection sites. The field sampling was conducted at the height of the acorn fall, and a previous study at the site recorded

deer moving up to 3 km to consume acorns within this stand (McShea and Schwede 1993).

Sampling in the field

The foraging behaviour of the animals and the relatively small volume of sample taken from each fecal pellet group prompted questions that it would be difficult to fully characterize the diet breadth of an animal from a single small subsample. To test this concern, we subdivided each fecal sample into three sub-replicates, and then sequenced each independently. This produced 36 samples from our original 12 samples. We further sought to minimize PCR bias through repeating the PCR twice for each of the 36 samples and then pooling each of the PCR replicates in equal concentration. We then examined within and between sample variance for the different samples. The direct analysis of the variance within and between samples pointed toward greater variance among individuals compared to variance within the three replicates taken from each individual's fecal sample. As new species were still being detected in the third subsample of several fecal samples we cannot closely estimate the number of subsamples needed to exhaustively characterize diet, but the higher variance between samples than within indicates we are approaching the maximum needed. The MDS projection, which characterizes the diversity of the entire set of plants found in each sample, similarly pointed to the clustering of the deer diets relative to a random sample, suggesting that while deer diets are diverse, they are not random and thus a sampling strategy using a single sub-sample per fecal is sufficient. A larger unknown is the length of time plant tissues reside within the gut; we speculate that the very large diet breadth observed reflects a significant time (>1 days), as well as reflecting the sensitivity of the methods.

We discount the role that pollen may have played in increasing the observed number of OTU as collections were made in the fall when relatively few species are in flower. However, we acknowledge that, as with any type of environmental sequencing, that our samples may include DNA from pollen, soil or other natural sources. Our observations suggest a much higher number of plants in the diet compared to reports from microhistological studies (Kowalczyk *et al.* 2011). For a highly mobile generalist herbivore, the sensitivity of the metagenomic methods likely reveal legitimate patterns of widespread selective browsing.

Low frequency OTU

We observed a very high number of OTU in our study, although their distribution was highly asymmetric. The 15

most common species accounted for 95% of all OTU counts, and the remaining observations recorded relatively few counts, and their observation did not significantly affect our quantification of native versus exotic plant forage choice. However, the issue of low frequency OTU is important, and reflects more generally how we identified true versus false observations. A number of studies have explored this issue (de Barba *et al.* 2014, Ficetola *et al.* 2016), and in cases where there is not a high level of asymmetry, their sampling methodology may ensure more stringent OTU assignment. Those studies correctly note that caution should be taken where single read counts are used as data and closely related species are considered in the reference database. In these cases mutations introduced through PCR or sequence error cannot be distinguished from legitimate low frequency OTU observations or even background environmental contamination. For our purposes, the high number of low frequency OTU likely reflect a combination of all of these processes. When we exclude species whose OTU units are <1% of total units for each sample, we do greatly reduce the number of species detected. However, we note that most species we identified at SCBI exceeded the 3% threshold for sequence differences; and that negative controls used in both the PCR and WGS assays failed to contain plant sequence data, suggesting that laboratory contamination was not a factor.

Quantifying trophic interactions

The critical role of herbivory in structuring natural communities is widely accepted (Janzen 1970; Huntly 1991; Danell *et al.* 2006), although herbivory is implicated in both promoting the diversity of plant communities (Janzen 1970; Crawley 1983; Bowers 1993; Parks *et al.* 2005), and reducing diversity through promoting the spread of invasive species due to herbivore preference of native species (Keane and Crawley 2002; Jogesh *et al.* 2008). Experimental trials of animals in captivity or exclusion experiments have found evidence for both increasing and decreasing diversity in natural communities (Augustine and McNaughton 1998; Côté *et al.* 2014; Kalisz *et al.* 2014); which suggests that herbivory may indeed be inconsistent in its effects.

Our observation suggests that white-tailed deer are selective feeders (Figs. 3–5). The effect of selecting native species in higher proportion than they exist in the field was not observed for all native plants, and the diet choice of the deer was clearly focused on a few species, but reflected preference for native species. The widespread success of invasive plants in the local floristic community, particularly *Alliaria petiolata*, may well be

promoted in part due to the selective browsing of native species by white-tailed deer.

Along with competitive advantages for light and water, and altered disturbance regimes, selective browsing of herbivores is a primary agent for successful invasions of exotic species (Ricciardi *et al.* 2013). We report evidence for selective browsing by deer with 105 common species being detected in feces at least 25% more or less than found in plot surveys. The list of these selected species includes several ecologically important woody species, such as *Quercus*, *Ulmus*, *Acer*, *Juglans* and *Viburnum* (McShea *et al.* 2007). Given that deer are highly generalist feeders, they are likely less affected by changes in community composition, but our increasing ability to measure and quantify diet from non-destructive methods such as through fecal DNA, will allow us to better quantify how changes in plant community structure may affect network dynamics in natural communities.

Conclusions and Future Challenges

In conclusion, we observed a wide diversity of plant species in the diet of 12 individuals of white-tailed deer collected at the SCBI long term research site [Fig. 3 and see Supporting Information—Table S1]. Fecal pellets had total genomic DNA extracted using the Zymo fecal DNA kit, with the resulting DNA PCR amplified using a *rbcl*-mini-barcode marker set, first described in this paper. The *rbcl* mini-barcode marker exhibited a high rate of species level identification, and could resolve all sequences from plants found at the site to at least genus, if not species (Fig. 2). The availability of an exhaustive reference library of *rbcl* sequences compiled for the SCBI site allowed us to use reference based assignment of recovered *rbcl* sequences, and demonstrated a wide range of species in the diet, suggesting the universality of the markers. The ability to readily recover plant DNA data from fecal pellets of a ruminant suggests we will be able to directly quantify plant diets and address hypotheses regarding the influence of ungulate herbivores on the structure of natural plant communities.

The use of metagenomics to assess diet of ungulates has great promise as it allows for species-specific identification of forage plants in natural communities. The techniques use by the research community will depend on several factors. Our ability to identify 72% of OTUs to species may be sufficient to determine native or exotic species consumed by deer, but may not satisfy requirements for studies of grazing species where the bulk of the diet is rapidly evolving grass species; in that case, use of a second locus barcode may be necessary. The

reagent costs are not high (~\$20/sample), but the expertise and access to sequencing equipment will play a larger role in determining the final cost. Its adoption also relies on an established reference library, which we had for this study site, and the rapid collection of fecal samples prior to degradation which was facilitated by the field fecal DNA kits used. In addition, there still remain some major technical questions before broadening the use of this technique. We have assumed that the amount of OTUs found in the feces reflects the proportion of the plant in the diet. Studies using coarse samples and visual inspection of prepared slides have not found this assumption to be valid. However, the sensitivity of the metagenomics method infers that we are not heavily reliant on examining only large undigested plant segments. This technique should be far more reliable at detecting plant consumption but those tests remain to be conducted. We are not sure of the amount of sampling needed to characterize the single fecal pile of an individual or the daily diet. We collected three samples for each fecal pile and these samples showed remarkable consistency but, as could be expected, were not in agreement for rare OTUs. A larger question was the forage period a single fecal pile represents; we found a high diversity of plant species in single fecal piles but are not sure if these samples represent an hour, day, week or month of foraging. With a complex digestive track and regurgitation of plant materials to increase digestion we do not know what time period we are sampling for this ungulate. A controlled experiment with captive deer placed on specific diets would answer these questions.

Sources of Funding

David L Erickson and Padmini Ramachandran were funded as an ORISE Fellow by the Office of Regulatory Science (ORS), Centre for Food Safety and Applied Nutrition (CFSAN), U.S. Food and Drug Administration. Funding for DNA sequencing was provided by the Division of Microbiology, ORS, CFSAN, U.S. Food and Drug Administration. Funding for DNA Sequencing was provided through support to the Division of Microbiology at the Food and Drug Administration. The plant DNA Barcode Library and the plant inventory were completed through the Smithsonian Institution's Global Genomics Initiative via a Smithsonian DNA Barcoding Network grant.

Contributions by the Authors

The experiment was designed by David Erickson and William McShea. The field data was collected by William

McShea and Norman Bourg. DNA sequencing and interpretation was completed by David Erickson, Padmini Ramachandran and Andrea Ottesen. Data Analysis included William McShea, David Erickson and Elizabeth Reed. The manuscript was prepared by William McShea, David Erickson, Norman Bourg and Andrea Ottesen. DNA sequence data are available from DL Erickson on request; **Supporting Information—Table S1** is a summary of these data.

Acknowledgements

Additional field and data assistance was provided by Xiaoli Shen, Chad Stewart, Olivia Cosby and Brandon Ross.

Conflict of Interest Statement

All authors had no conflict of interests. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Supporting Information

The following additional information is available in the online version of this article—

Figure S1. The percent of total OTUs for the 25 most common plant species for each of 36 fecal samples (3 samples from 12 fecal piles). Sample ID indicates the fecal pile and the subsample (e.g. Sample 1.1, 1.2, 1.3 are 3 sub-samples of the same fecal pile). These 25 species represent >95% of the plant species detected in the samples. Exotic species are identified with an asterisk.

Table S1. List of species identified based on comparison of DNA samples extracted from fecal pellets from 12 deer when compared to reference DNA barcode created for site. For each of the 12 deer, 3 sub-samples were prepared from each fecal pile and the sum of OTUs for each species listed in table. For those deer where the species was not found in each sub-sample we have indicated if two (**) or one (*) sub-samples contained the plant species. Some OTUs could not be classified to species level and only the genus level is indicated. Species are listed in their taxonomic order. The bottom of each column sums the total number of species detected and the number of species if values less than 1% of the total OTUs are excluded.

Table S2. List of species detected and their relative abundance in either 1601 × 1 m plots within the forest around the Large Tree Plot or within the 12 fecal samples obtained in the same general location. When species level identification was not possible for either the fecal

sample or the individual plant we have listed the genus. Plants are listed based on their relative abundance in the plot survey. We have categorized plants as native, introduced or invasive. The data in this table are used to create Fig. 4 in Paper.

Literature Cited

- 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**:8966–9.
- Alipayo D, Valdez R, Holechek JL, Cardenas M. 1992. Evaluation of microhistological analysis for determining ruminant diet botanical composition. *Journal of Range Management* **1**: 148–152.
- Anderson JD, Honeycutt RL, Gonzales RA, Gee KL, Skow LC, Gallagher RL, Honeycutt DA, DeYoung RW. 2002. Development of microsatellite DNA markers for the automated genetic characterization of white-tailed deer populations. *The Journal of Wildlife Management* **66**:67–74.
- Augustine DJ, McNaughton SJ. 1998. Ungulate effects on the functional species composition of plant communities: herbivore selectivity and plant tolerance. *The Journal of Wildlife Management* **62**:1165–1183.
- Averill KM, Mortensen DA, Smithwick EA, Post E. 2016. Deer feeding selectivity for invasive plants. *Biological Invasions* **18**:1247–1263.
- Averill KM, Mortensen DM, Smithwick EAH, Susan Kalisz S, McShea WJ, Bourg NA, Parker JD, Royo AA, Abrams MD, Anurag A, Agrawal AA, Apsley DK, Blossey B, Caraher KL, Johnson SE, Masson R, Nuzzo VA, Snitzer J. 2017. Native deer browsing increases the degree of plant invasion. *AoBPlants*.
- Baker BJ, Banfield JF. 2003. Microbial communities in acid mine drainage. *FEMS Microbiology Ecology* **44**:139–152.
- Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E. 2005. Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**: 1935–1943.
- Blaser MJ. 2014. The microbiome revolution. *The Journal of Clinical Investigation* **124**:4162–4165.
- Bourlat SJ, Borja A, Gilbert J, Taylor MI, Davies N, Weisberg SB, Griffith JF, Lettieri T, Field D, Benzie J, Glöckner FO. 2013. Genomics in marine monitoring: new opportunities for assessing marine health status. *Marine Pollution Bulletin* **74**:19–31.
- Bourg NA, McShea WJ, Thompson JR, McGarvey JC, Shen X. 2013. Initial census, woody seedling, seed rain and stand structure data for the SCBI SIGEO Large Forest Dynamics Plot (*Ecological Archives* E094-195). *Ecology* **94**:2111.
- Bowers MA. 1993. Influence of herbivorous mammals on an old-field plant community: years 1–4 after disturbance. *Oikos* **1**: 129–141.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**:335–336.
- CBOL Plant Working Group, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences USA* **106**:12794–12797.
- Clare EL. 2014. Molecular detection of trophic interactions: emerging trends, distinct advantages, significant considerations and conservation applications. *Evolutionary Applications* **7**: 1144–1157.
- Côté SD, Beguin J, de Bellefeuille S, Champagne E, Thiffault N, Tremblay JP. 2014. Structuring effects of deer in boreal forest ecosystems. *Advances in Ecology*. doi:10.1155/2014/917834.
- Côté SD, Rooney TP, Tremblay JP, Dussault C, Waller DM. 2004. Ecological impacts of deer overabundance. *Annual Review of Ecology, Evolution, and Systematics* **31**:113–147.
- Crawley MJ. 1983. *Herbivory, the dynamics of animal-plant interactions*. Oxford: Blackwell Scientific Publications.
- Creed RP. 2000. Is there a new keystone species in North American lakes and rivers? *Oikos* **91**:405–408.
- Cumplido JH, Glauser G, Benrey B. 2016. Cascading effects of early-season herbivory on late-season herbivores and their parasitoids. *Ecology* **97**:1283–1297.
- Danell K, Bergström R, Duncan P, Pastor J. 2006. *Large herbivore ecology, ecosystem dynamics and conservation*. Cambridge: Cambridge University Press.
- 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.
- de Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P. 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources* **14**:306–332.
- Dormann CF, Fruend J, Bluethgen N, Gruber B. 2009. Indices, graphs and null models: analyzing bipartite ecological networks. *The Open Ecology Journal* **2**:7–24.
- Edgar R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.
- Enge S, Nylund GM, Harder T, Pavia H. 2012. An exotic chemical weapon explains low herbivore damage in an invasive alga. *Ecology* **93**:2736–2745.
- Erickson DL, Jones FA, Swenson NG, Pei N, Bourg NA, Chen W, Davies SJ, Ge XJ, Hao Z, Howe RW, Huang CL. 2014. Comparative evolutionary diversity and phylogenetic structure across multiple forest dynamics plots: a mega-phylogeny approach. *Frontiers in Genetics* **5**:358.
- Ficetola GF, Taberlet P, Coissac E. 2016. How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources* **16**:604–607.
- Fletcher JD, Shipley LA, McShea WJ, Shumway DL. 2001. Wildlife herbivory and rare plants: the effects of white-tailed deer, rodents, and insects on growth and survival of Turk's cap lily. *Biological Conservation* **101**:229–238.
- Frerker K, Sabo A, Waller D. 2014. Long-term regional shifts in plant community composition are largely explained by local deer impact experiments. *PLoS One* **9**:e115843.
- García-Robledo C, Erickson DL, Staines CL, Erwin TL, Kress WJ. 2013. Tropical plant-herbivore networks: reconstructing species interactions using DNA barcodes. *PLoS One* **8**:e52967.

- Goldberg Caren S, Strickler KM, Pilliod DS. 2015. Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation* **183**:1–3.
- Hajishengallis G, Darveau RP, Curtis MA. 2012. The keystone-pathogen hypothesis. *Nature Reviews Microbiology* **10**:717–725.
- Hibert F, Taberlet P, Chave J, Scotti-Saintagne C, Sabatier D, Richard-Hansen C. 2013. Unveiling the diet of elusive rainforest herbivores in Next Generation Sequencing era? The tapir as a case study. *PLoS One* **8**:e60799.
- Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology* **180**:4765–4774.
- Holeček JL, Vavra M, Pieper RD. 1982. Botanical composition determination of range herbivore diets: a review. *Journal of Range Management* **1**:309–315.
- Holm JA, Thompson JR, McShea WJ, Bourg NA. 2013. Interactive effects of chronic deer browsing and canopy gap disturbance on forest successional dynamics. *Ecosphere* **4**: 144.
- Huntly N. 1991. Herbivores and the dynamics of communities and ecosystems. *Annual Review of Ecology and Systematics* **22**: 477–503.
- Illumina Corporation. 2013. 16S Metagenomic Sequencing Library Preparation. https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf.
- Janzen DH. 1970. Herbivores and the number of tree species in tropical forests. *American Naturalist* **104**:501–528.
- Jogesh T, Carpenter D, Cappuccino N. 2008. Herbivory on invasive exotic plants and their non-invasive relatives. *Biological Invasions* **10**:797–804.
- Kalisz S, Spigler RB, Horvitz CC. 2014. In a long-term experimental demography study, excluding ungulates reversed invader's explosive population growth rate and restored natives. *Proceedings of the National Academy of Sciences USA* **111**: 4501–4506.
- Kartzinel TR, Chen PA, Coverdale TC, Erickson DL, Kress WJ, Kuzmina ML, Rubenstein DI, Wang W, Pringle RM. 2015. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National Academy of Sciences USA* **112**:8019–8024.
- Keane RM, Crawley MJ. 2002. Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology and Evolution* **17**:164–170.
- Kowalczyk R, Taberlet P, Coissac E, Valentini A, Miquel C, Kamiński T, Wójcik JM. 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, Erickson DL. 2007. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. *PLoS One* **2**:e508.
- Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjurjo O, Bermingham E. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Science of the USA* **106**: 18621–18626.
- Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences USA* **82**:6955–6959.
- Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Pontén T, Aarestrup FM, Ussery DW, Lund O. 2014. Benchmarking of methods for genomic taxonomy. *Journal of Clinical Microbiology* **52**:1529–1539.
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* **10**:1.
- Leray M, Knowlton N. 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences* **112**: 2076–2081.
- Leray M, Meyer CP, Mills SC. 2015. Metabarcoding dietary analysis of coral dwelling predatory fish demonstrates the minor contribution of coral mutualists to their highly partitioned, generalist diet. *PeerJ*: epub.
- Letnic M, Koch F, Gordon C, Crowther MS, Dickman CR. 2009. Keystone effects of an alien top-predator stem extinctions of native mammals. *Proceedings of the Royal Society of London B: Biological Sciences* **276**:3249–3256.
- Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* **71**:8228–8235.
- Macel M, Vos RC, Jansen JJ, Putten WH, Dam NM. 2014. Novel chemistry of invasive plants: exotic species have more unique metabolomic profiles than native congeners. *Ecology and Evolution* **4**: 2777–2786.
- Maron JL, Crone E. 2006. Herbivory: effects on plant abundance, distribution and population growth. *Proceedings of the Royal Society of London B: Biological Sciences* **273**:2575–2584.
- McShea WJ. 2012. Ecology and management of white-tailed deer in a changing world. *Annals of the New York Academy of Sciences* **1249**:45–56.
- McShea WJ, Healy WM, Devers P, Fearer T, Koch FH, Stauffer D, Waldon J. 2007. Forestry matters: decline of oaks will impact wildlife in hardwood forests. *The Journal of Wildlife Management* **71**:1717–1728.
- McShea WJ, Schwede G. 1993. Variable acorn crops: responses of white-tailed deer and other mast consumers. *Journal of Mammalogy* **74**:999–1006.
- McCabe TR, McCabe RE. 1997. Recounting whitetails past. In: McShea WJ, Underwood HB, Rappole JH, eds. *The science of overabundance: Deer ecology and population management*. Washington, DC: Smithsonian Press, 11–26.
- Meier R, Shiyang K, Vaidya G, Ng PK. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* **55**:715–728.
- Murray DC, Bunce M, Cannell BL, Oliver R, Houston J, White NE, Barrero RA, Bellgard MI, Haile J. 2011. DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches. *PLoS One* **6**:e25776.
- Paine RT. 1966. Food web complexity and species diversity. *American Naturalist* **1**:65–75.
- Parker JD, Burkepile DE, Hay ME. 2006. Opposing effects of native and exotic herbivores on plant invasions. *Science* **311**:1459–1461.
- Parks CG, Radosevich SR, Endress BA, Naylor BJ, Anzinger D, Rew LJ, Maxwell BD, Dwire KA. 2005. Natural and land-use history of the Northwest mountain ecoregions (USA) in relation to patterns of

- plant invasions. *Perspectives in Plant Ecology, Evolution and Systematics* **7**:137–158.
- Pastor J, Naiman RJ. 1992. Selective foraging and ecosystem processes in boreal forests. *American Naturalist* **139**:690–705.
- Pereyra PJ. 2016. Revisiting the use of the invasive species concept: an empirical approach. *Austral Ecology* **41**:519–528.
- Pompanon F, Deagle BE, Symondson WO, Brown DS, Jarman SN, Taberlet P. 2012. Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology* **21**:1931–1950.
- Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution* **26**:1641–1650.
- Richardson DM, Pyšek P, Rejmánek M, Barbour MG, Panetta FD, West CJ. 2000. Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions* **6**:93–107.
- Ricciardi A, Hoopes MF, Marchetti MP, Lockwood JL. 2013. Progress toward understanding the ecological impacts of nonnative species. *Ecological Monographs* **83**:263–282.
- Rooney TP. 2001. Deer impacts on forest ecosystems: a North American perspective. *Forestry* **74**:201–208.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K. 2007. The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biology* **5**(3):e77.
- Sanders NJ, Gotelli NJ, Heller NE, Gordon DM. 2003. Community disassembly by an invasive species. *Proceedings of the National Academy of Sciences USA* **100**:2474–2477.
- Shears NT, Babcock RC. 2003. Continuing trophic cascade effects after 25 years of no-take marine reserve protection. *Marine Ecology: Progress Series* **246**:1–6.
- Shen X, Bourg NA, McShea WJ, Turner BL. 2016. Long-term effects of white-tailed deer exclusion on the invasion of exotic plants: a case study in a Mid-Atlantic temperate forest. *PLoS One* **11**:e0151825.
- Srivathsan A, Sha J, Vogler AP, Meier R. 2015. Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). *Molecular Ecology Resources* **15**:250–261.
- Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences USA* **103**:12115–12120.
- Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebo JH, Ims RA, Yoccoz NG, Taberlet P. 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.
- Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. 2012. Environmental DNA. *Molecular Ecology* **21**:1789–1793.
- Thomsen PF, Willerslev E. 2015. Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* **183**:4–18.
- Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. 2009. A core gut microbiome in obese and lean twins. *Nature* **457**:480–484.
- United States Department of Agriculture, Natural Resources Conservation Service. *Plants Database*. Retrieved from <https://plants.usda.gov/java/> (6 Feb 2017).
- Valentini A, Pompanon F, Taberlet P. 2009. DNA barcoding for ecologists. *Trends in Ecology and Evolution* **24**:110–117.
- Valéry L, Hervé F, Lefeuvre J, Simberloff D. 2008. In search of a real definition of the biological invasion phenomenon itself. *Biological Invasions* **10**:1345–1351.
- Vavra M, Parks CG, Wisdom MJ. 2007. Biodiversity, exotic plant species, and herbivory: the good, the bad, and the ungulate. *Forest Ecology and Management* **246**:66–72.
- Virginia Department of Conservation and Recreation. 2015. *Virginia Invasive Plant Species List*. Retrieved from <http://www.dcr.virginia.gov/natural-heritage/invspdflist> (6 Feb 2017).
- Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* **15**:R46.
- Yoccoz NG, Bråthen KA, Gielly L, Haile J, Edwards ME, Goslar T, Von Stedingk H, Brysting AK, Coissac E, Pompanon F, Sønstebo JH. 2012. DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology* **21**:3647–3655.