High-throughput sequencing-based waterbird diet analysis: application to wintering herbivorous geese

Yuzhan Yang¹, Aibin Zhan², Lei Cao², Fanjuan Meng², Wenbin Xu³

¹School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, China
²Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China
³Anhui Shengjin Lake National Nature Reserve Administration, Chizhou, Anhui, China

Corresponding Author:
Lei Cao²
18 Shuangqing Road, Haidian District, Beijing, China, 100085

Email address: caolei@ustc.edu.cn
Abstract

Food availability and diet selection are important factors influencing the abundance and
distribution of wild waterbirds. In order to better understand changes in waterbird populations, it
is essential to figure out what they feed on. However, analyzing diet could be difficult and
inefficient using traditional methods, such as microhistologic observation. Here, we addressed
this gap of knowledge by investigating diet of greater white-fronted goose *Anser albifrons* and
bean goose *Anser fabalis*, which are obligate herbivores wintering in China, mostly in Middle
and Lower Yangtze River Floodplain. We firstly prepared a local plant reference library by
selecting an optimal marker gene (P6 loop of chloroplast trnL intron) and amplifying the most
common plants that these geese would consume. Then, utilizing DNA metabarcoding, we
discovered 15 food items in total from feces of these birds. Of the 15 unique dietary sequences,
10 could be identified at species-level. As for greater white-fronted goose, 73% of sequences
belonged to Poaceae spp., and 26% belonged to Carex spp. In contrast, almost all sequences of
bean goose belonged to Carex spp. (99%). Using the same samples, microhistology provided
consistent food composition with metabarcoding results for greater white-fronted goose, while 13%
of Poaceae was recovered for bean goose. In addition, two other taxa were discovered only
through microhistologic analysis. Although most of the identified taxa matched relatively well
between the two methods, DNA metabarcoding gave taxonomically more detailed information.

Discrepancies were likely due to biased PCR amplification in metabarcoding, low discriminating
power of current marker genes for monocots, and biases in microhistologic analysis. The diet
differences between two geese species might indicate deeper ecological significance beyond the
scope of this study. We concluded that DNA metabarcoding provided new perspectives for
studies of herbivorous waterbird diets and inter-specific interactions, as well as new possibilities
to investigate interactions between herbivores and plants. In addition, microhistologic analysis
should be used together with metabarcoding methods to integrate these information.
**Introduction**

Wetlands are one of the most important ecosystems in nature, and they harbor a variety of ecosystem services such as protection against floods, water purification, climate regulation and recreational opportunities (Brander, Flora & Vermaat, 2006). Waterbirds are typically wetland-dependent animals upon which they could get abundant food and suitable habitats (Ma et al., 2010). Waterbird abundance and distribution could reflect the status of wetland structure and functions, making them important bio-indicators for wetland health (Fox et al., 2011). Among all factors affecting waterbird community dynamics, food availability is frequently considered to play one of the most important roles (Wang et al., 2013). However, recently suitable food resources have tended to decrease or even disappear due to deterioration and loss of natural wetlands (Fox et al., 2011). As a result, waterbirds are forced to discard previous habitats and sometimes even feed in agricultural lands (Zhang et al., 2011). In addition, migratory waterbirds may aid the dispersal of aquatic plants or invertebrates by carrying and transporting them between water bodies at various spatial scales (Reynolds, Miranda & Cumming, 2015).

Consequently, long-time monitoring and systematic studies of waterbird diets are essential to understand population dynamics of waterbirds, as well as to establish effective management programs for them (Wang et al., 2012).

Traditional methods for waterbird diet analysis, were direct observation in the field (Swennen & Yu, 2005) or microhistologic analysis of remnants in feces and/or gut contents (James & Burney, 1997; Fox et al., 2007). While these approaches have been proved useful in some cases, they are relatively labor-intensive and greatly skill-dependent (Fox et al., 2007;
Samelius & Alisauskas, 1999; Symondson, 2002). Applications of other methods for analyzing gut contents or feces were also restricted due to inherent limitations, as reviewed by Pompanon et al. (Pompanon et al., 2012). Recently, metabarcoding methods, based on high-throughput sequencing, have provided new perspectives for diet analysis and biodiversity assessment (Taberlet et al., 2007; Creer et al., 2010). These methods provide higher taxonomic resolution and enormous sequence output simultaneously from large-scale environmental samples, such as soil, water and feces (Shokralla, Spall & Gibson, 2012; Bohmann et al., 2014). Owing to these advantages, metabarcoding has been widely employed in diet analysis of herbivores (Taberlet et al., 2012; Ando et al., 2013; Hibert et al., 2013), carnivores (Deagle, Kirkwood & Jarman, 2009; Shehzad et al., 2012) and omnivores (De Barba et al., 2014). But pitfalls of metabarcoding should not be ignored when choosing suitable techniques for new studies. For instance, many researches have shown that it is difficult to obtain quantitative data using metabarcoding (Sun et al., 2015). This drawback might result from both technical issues of this method and relevant biological features of samples (Pompanon et al., 2012).

One paramount prerequisite of metabarcoding methods is to select robust genetic markers and corresponding primers (Zhan et al., 2014; Zhan & Maclsaac, 2015). For diet study of herbivores, at least eight chloroplast genes and two nuclear genes are used as potential markers for land plants (Hollingsworth, Graham & Little, 2001). Although mitochondrial cytochrome c oxidase I (COI) is extensively recommended as a standard barcode for animals, its relatively low rate of evolution in botanical genomes precludes it being an optimum for plants (Wolfe, Li & Sharp, 1987; Fazekas et al., 2008). The internal transcribed spacer (ITS) is excluded due to...
divergence discrepancies of individuals and low reproducibility (Álvarez & Wendel, 2003). A variety of combinations and comparisons have been performed for the eight candidate genes, however, none proved equally powerful for all cases (Fazekas et al., 2008). Consequently, it is more effective to choose barcodes for a circumscribed set of species occurring in a regional community (Kress et al., 2009). Another equally important aspect of metabarcoding application is the construction of reference libraries which assist taxonomic assignment (Rayé et al., 2011; Xu et al., 2015). It is difficult to accurately interpret sequence reads without a reliable reference library (Elliott & Jonathan Davies, 2014).

Diet analysis is a central issue in waterbird research, both for deciphering waterfowl population dynamics and interpreting inter- or intra-specific interactions of cohabitating species (Zhao et al., 2015). For instance, more than 60% of bean goose Anser fabalis and almost 40% of greater white-fronted goose Anser albifrons populations along the East Asian – Australian Flyway Route winter at the Shengjin Lake National Nature Reserve (Zhao et al., 2015). Previous studies based on microhistologic observation illustrated that the dominant composition of their diets were monocotyledons, such as Carex spp. (Zhao et al., 2012), Poaceae (Zhang et al., 2011), and a relatively small proportion of non-monocots (referred to as dicotyledons in study of “Zhao, Cao & Fox, 2013”). However, few food items could be identified to species-level, mainly owing to variable tissue structures within plants, similar morphology between relative species, and a high level of degradation after digestion (Zhang et al., 2011; Zhao et al., 2012; Zhao, Cao & Fox, 2013). Ambiguous identification has hindered understanding of waterbird population dynamics and potential to establish effective conservation plans for them.
In this study, we aimed to improve this situation using a metabarcoding method to analyze diets of these species (see flowchart in Fig. 1). By examining the efficiency of eight candidate genes (*rbcL, rpoC1, rpoB, matK, trnH-psbA, trnL* (UAA), *atpF-atpH*, and *psbK-psbI*), we selected robust genes and corresponding primers for reference library construction and high-throughput sequencing. Subsequently, we used the metabarcoding method to investigate diet composition of these two species based on *feces* collected from Shengjin Lake. Finally, we discussed and compared results from microhistology and DNA metabarcoding using the same samples to assess the utility and efficiency of these two methods.

**Materials and Methods**

**Ethics Statement**

Our research work did not involve capture or any direct manipulation or disturbances of animals. We collected samples of plants and *feces* for molecular analyses. We got access to the reserve under the permission of Shengjin Lake National Nature Reserve Administration (Chizhou, Anhui, China), which is responsible for the management of the protected area and wildlife. We were forbidden to capture or disturb geese in the field.

**Study Area**

Shengjin Lake (116°55´ - 117°15´ E, 30°15´ -30°30´ N) was established as National Nature Reserve in 1997, aiming to protect diverse waterbirds including geese, cranes and storks. The water level fluctuates greatly in this lake, with maximal water level of 17 m during summer (flood season) but only 10 m during winter (dry season). Due to this fluctuation, receding waters expose two large *Carex* spp. meadows and provide suitable habitats for waterbirds. This makes
Shengjin Lake one of the most important wintering sites for migratory waterbirds (Zhao et al., 2015). Greater white-fronted goose and bean goose are the dominant herbivores wintering (from October to April) in this area, accounting for 40% and 60% of populations along the East Asian – Australian Flyway Route, respectively (Zhao et al., 2015).

Field Sampling

The most common plant species that these two geese may consume were collected in May 2014 and January 2015, especially species belonging to Carex and Poaceae. Fresh and intact leaves were carefully picked, tin-packaged in the field and stored at -80 °C in the laboratory before further treatment. Morphological identification was carried out with the assistance of two botanists (Profs Zhenyu Li and Shuren Zhang from Institute of Botany, Chinese Academy of Sciences). In total, 87 specimens were collected, belonging to 25 families, 53 genera and 70 species (Table S1).

All feces were collected at the reserve (Fig. 2) in January 2015. Based on previous studies and the latest waterbird survey, sites with big flocks of geese (i.e. more than 200 individuals) were chosen (Zhang et al., 2011). As soon as geese finished feeding and feces were defecated, fresh droppings were picked and stored in dry ice. Droppings of bean goose were generally thicker than those of smaller greater white-fronted goose, to the degree that these could be reliably distinguished in the field (Zhao et al., 2015). Disposal gloves were changed for each sample to avoid cross contamination. To avoid repeated sampling and make sure samples were from different individuals, each sample was collected with a separation of more than two meters. In total, 21 feces were collected, including 11 for greater white-fronted goose and 10 for bean.
goose. All samples were transported to laboratory in dry ice and then stored at -80 °C until further analysis.

**Selection of Molecular Markers and Corresponding Primers**

In this part, we aimed to select gene markers with adequate discriminating power for our study. We included eight chloroplast genes - rbcL, rpoC1, rpoB, matK, trnH-psbA, trnL (UAA), atpF-atpH, and psbK-psbI for estimation. Although Shengjin Lake included an array of plant species, we focused mainly on the most likely food resources (Xue et al., 2008; Zhao et al., 2015) that geese would consume for candidate gene tests. These covered eleven genera and the family Poaceae (Table S2, Table 2). For tests of all candidate genes, we recovered sequences of representative species in the selected groups from GenBank.

We calculated inter-specific divergence within every genus or family based on the Kimura 2-parameter model (K2P) using MEGA version 6 (Tamura et al., 2013). We also constructed molecular trees based on UPGMA using MEGA and characterized the resolution of species by calculating the percentage of species recovered as monophyletic based on phylogenetic trees (Rf?). Secondly, primers selected out of eight of candidate genes were used to amplify all 87 specimens and to check their amplification efficiency and universality. Thirdly, we calculated inter-specific divergence based on sequences that we obtained from last step. Generally, a robust barcode gene is obtained when the minimal inter-specific distance exceeds the maximal intra-specific distance (e.g. existence of barcoding gaps). For reference database building, we calculated the rate of discrimination for the species in each family (Rf) by dividing the number of unique sequences per family by the number of
species resolved as monophyletic clades in each family. Finally, to allow the recognition of sequences after high-throughput sequencing, both of the forward and reverse primers of the selected marker gene were tagged specifically for each sample with 8nt nucleotide codes at the 5’ end (Parameswaran et al., 2007).

**DNA Extraction, Amplification and Sequencing**

Two hundred milligrams of leaf was used to extract the total DNA from each plant sample using a modified CTAB protocol (Cota-Sanchez, Remarchuk & Ubayasena, 2006). DNA extraction of feces was carried out using the same protocol with minor modification in incubation time (elongate to 12 h). Each fecal sample was crushed thoroughly and divided into four quarters. All quarters of DNA extracts were then pooled together. DNA extraction was carried out in a clean room used particularly for this study. For each batch of DNA extraction, negative controls (i.e. extraction without feces) were included to monitor possible contamination.

For plant DNA extracts, PCR amplifications were carried out in a volume of 25µl with ~100 ng total DNA as template, 1U of Taq Polymerase (Takara, Dalian, Liaoning Prov., China), 1× PCR buffer, 2 mM of Mg²⁺, 0.25 mM of dNTPs, 0.1 µM of forward primer and 0.1 µM of reverse primer. After 4 min at 94 °C, the PCR cycles were as follows: 35 cycles of 30 s at 94 °C, 30 s at 56 °C and 45 s at 72 °C, and the final extension was 10 min at 72 °C. We applied the same PCR conditions for all primers. All the successful PCR products were sequenced with Genewiz (Suzhou, Jiangsu Prov., China).

For fecal DNA extracts, PCR mixtures (25µl) were prepared in six replicates for each sample to reduce biased amplification. Each replicate was subjected to the same amplification conditions and the PCR cycles were as follows: 35 cycles of 30 s at 94 °C, 30 s at 56 °C, 45 s at 72 °C, and the final elongation was 10 min at 72 °C.
procedure used for plant extracts. Each set of six replicates was pooled and purified using the Sangon PCR product purification kit (Sangon Biotech, Shanghai, China). Quantification was carried out to ensure equilibrium of contribution of each sample using the NanoDrop ND-2000 UV-Vis Spectrophotometer (NanoDrop Technologies, Delaware, United States of America).

High-throughput sequencing was performed using Illumina MiSeq platform following manufacturer’s instructions by BGI (Shenzhen, Guangdong Prov., China). Reads of high-throughput sequencing could be found at NCBI’s Sequence Read Archive (Accession number: SRP070470).

Data Analysis for Estimating Diet Composition

After high-throughput sequencing, pair-ended reads were merged with using the UPARSE pipeline (http://drive5.com/uparse, Edgar, 2010). Reads were then split into independent files according to unique tags using RDP pipeline (http://rdp.cme.msu.edu/). We removed sequences i) that didn’t perfectly match tags and primer sequences; ii) that contained ambiguous nucleotide (N’s). Tags and primers were then trimmed using the RDP pipeline. Further quality filtering based on the UPARSE pipeline discarded sequences with i) quality score less than 30 (<Q30) and ii) shorter than 100 bp and longer than 200 bp. Unique sequences were clustered to operational taxonomy units (OTUs) at the similarity threshold of 98% (Edgar, 2013). All OTUs were assigned to unique taxonomy with local blast 2.2.30+ (Altschul et al., 1990). We detected a plant within the reference library for each sequence with the threshold of length coverage > 98%, identity > 98% and e-value < 1.0 e^{-50}. If a query sequence matched two or more taxa, it was assigned to a higher taxonomic level which included all taxa.
Microhistology analysis

We used the method described by Zhang et al. (2011) to perform microhistologic examination of fecal samples. Each sample was first washed with pure water and filtered with a 25-µm filter. Subsequently, the suspension was examined under a light microscope at 10× magnification for quantification statistics and at 40× magnification for species identification. We compared photos of visible fragments with an epidermis database of plants from Shengjin Lake to identify food items (Source?...Zhang? Fox?).

Results

Selection of Genes and Corresponding Primers and Reference Library Constructing

A total of 3,296 representative sequences were recovered from GenBank, ranging from 0 to 345 sequences per gene, per genus (Table S2). Among the candidate genes, trnL, trnH-psbA, matK and rbcL showed largest inter-specific divergence in seven, three, one and one taxonomic groups, respectively. These four genes also displayed relatively high resolution of species (Table S2). For example, with matK gene, 77% of Carex could be identified to species-level. However, our results indicated that none of these eight genes could simultaneously differentiate all 12 genera or families to species-level (Table S2). Considering the inter-specific divergence and resolution of species, we chose the most commonly used chloroplast genes rbcL, matK, trnH-psbA and trnL for further tests.

Primers for these four genes (Table 1) were used to amplify the plants that we collected in the field. The selected primers for trnL and rbcL successfully amplified 100% and 91% of all species, respectively, while primers for trnH-psbA and matK amplified only 71% and 43%.
respectively. Therefore, we chose trnL and rbcL to test their discriminating power in our target plants.

We calculated the inter-specific divergence within genera and families with at least two species to compare their discriminating power. Maximal, minimal and mean inter-specific distances were calculated for seven dominant genera and six dominant families (Table 2). Neither gene could differentiate species of *Vallisneria* Linn. (mean=0.000±0.000%) or *Artemisia* Linn. (mean=0.000±0.000%). But trnL showed a larger divergence range for the other six genera and five families. Hence, we chose trnL as the barcoding gene for reference library constructing and high-throughput sequencing for our study. The discriminating power of trnL was strong for most species (Table 3). However, some species could only be identified at genus-level or family-level with trnL. For instance, five species of *Potamogeton* shared the same sequences and this made them to be identified at genus-level. Species could be identified easily to genus and family, except for three grasses (Poaceae) *Beckmannia syzigachne*, *Phalaris arundinacea*, and *Polypogon fugax*, which shared identical sequences.

**Data Processing for Estimating Diet Composition**

In total, 0.21 and 0.18 million reads were generated for greater white-fronted goose (GWFG) and bean goose (BG), respectively (Table 4). The number of recovered OTUs ranged from 8 to 123 for GWFG and BG samples. We used local BLAST to compare these sequences with the Shengjin Lake reference database. Finally, with DNA metabarocoding, 12 items were discovered in the feces of GWFG, including one at family-level, three at genus-level and eight at species-level (Table 5). Four items were discovered in the feces of BG, including one at...
genus-level and three at species-level. **In total,** this method identified 15 taxa in feces of these geese.

However, the sequence percentage of each food item varied greatly (Table 5). For GWFG, the majority of sequences (96.36%) were composed of only five items: Poaceae spp. (47.98%), Poa annua (21.86%), Carex heterolepis (17.51%), Carex spp. (9.01%), and Alopecurus aequalis (3.21%). For BG, almost all the sequences belonged to Carex heterolepis (99.49%). Other items only occupied a relatively small proportion of sequences. In addition, the presence of each item per sample was also unequal (Table S3). For example, in GWFG, Carex heterolepis, Carex spp., Poa annua and Potentilla supina were present in almost all the samples, while Stellaria media, Asteraceae sp. and Lapsana apogonoldes occurred in only about one third of samples,

When microhistologic examination were performed using the same samples, eight items were found in the feces of greater white-fronted goose, including one at family-level, four at genus-level and three at species-level (Table 5). Dominant items were Poaceae spp. (45.68%), Alopecurus Linn. (30.93%) and Carex heterolepis (16.39%). Seven items were found in the feces of bean goose, including four at genus-level and three at species-level (Table 5). Dominant items were Carex heterolepis (62.85%), Asteraceae sp. (14.55%), and Alopecurus Linn. (13.18%).

**Discussion**

**Marker Selection and Reference Library Constructing for Diet Analysis**

With greatly reduced cost, extremely high throughput and **information content, metabarcoding** has revolutionized the exploration and quantification of dietary analysis from noninvasive samples containing degraded DNA (Fonseca et al., 2010; Shokralla et al., 2014). Despite
enormous potential to boost data acquisition, successful application of this technology relies
greatly on the power and efficiency of genetic markers and corresponding primers (Bik et al.,
2012; Zhan et al., 2014). In order to select the most appropriate marker gene for our study, we
compared the performance of eight commonly used chloroplast genes, rbcL, rpoB, rpoC1, matK,
trnL, trnH-psbA, atpF-atpH, and psbK-psbl and their corresponding primers. Although a higher
level of discriminating power was shown in several studies, atpF-atpH, psbK-psbI, rpoB and
rpoC1 were not as commonly used as other barcoding genes (Hollingsworth, Graham & Little,
2001). As one of the most rapidly evolving coding genes of plastid genomes, matK was
considered as the closest plant analogue to the animal barcode COI (Hilu & Liang, 1997).
However, matK was difficult to amplify using available primer sets, with only 43% of successful
amplification in this study. In spite of the higher species discrimination success of trnH-psbA
than rbcL+matK in some groups, the presence of duplicated loci, microinversions and premature
termination of reads by mononucleotide repeats lead to considerable proportion (30% in this
study) of low-quality sequences and over-estimation of genetic difference when using trnH-psbA
(Graham et al., 2000; Whitlock Hale & Groff, 2010). In contrast, the barcode region of rbcL is
easy to amplify, sequence, and align in most plants and was recommended as the standard
barcode for land plants (Chase et al., 2007). The relatively modest discriminating power
(compared to trnL) precludes its application for our study aiming to recover high resolution of
food items. Consequently, trnL was selected out of eight candidate markers, with 100%
amplification success, more than 90% of high quality sequences, and relatively large
inter-specific divergence.
One of the biggest obstacles in biodiversity assessment and dietary analysis is the lack of a comprehensive reference library, without which it is impossible to accurately interpret and assign sequences generated from high-throughput sequencing (Valentini, Pompanon & Taberlet 2009; Barco et al., 2015). In this study, we constructed a local reference library by amplifying the most common species (70 morpho-species in total) during the wintering period with the trnL gene. Although not all of them could be identified at species-level with trnL due to relatively low inter-specific divergence, many species could be separated with distinctive sequences. Previous studies have recommended group-specific barcodes to differentiate closely related plants at the species level (Li et al., 2015). For instance, matK has been proved to be more efficient for the discrimination of Carex spp. (Starr, Naczi & Chouinard, 2009). However, the primer set of matK failed to amplify species of Carex spp. in our study, suggesting the universality of selected primer pairs should be tested in each study (Zhan et al., 2014).

**Applications of Metabarcoding for Geese Diet Analysis**

A variety of recent studies have demonstrated the great potential of metabarcoding for dietary analysis, mainly owing to the high throughput, high discriminating power, and the ability to process large-scale samples simultaneously (Creer et al., 2010; Taberlet et al., 2012; Shehzad et al., 2012). In this study, we applied this method to recover diets of herbivorous geese and provided standard protocols for dietary analysis of these two ecologically important waterbirds. Our results further proved the more objective, less experience-dependent and more time-efficient character of DNA metabarcoding. However, not all the species in the reference library could be identified at species-level, owing to low inter-specific divergence. We suggest that multiple
group-specific markers to be incorporated in the future, as in De Barba et al. (2014). Two species, *Carex thunbergii* and *Fabaceae* sp., were only discovered via microhistologic analysis rather than metabarcoding. This failure might reflect the biased fragment amplification of current technology, of which dominant templates could act as inhibitors of less dominant species (Piñol et al., 2015). However, three species of *Poaceae* were only discovered using metabarcoding. In total, more taxa and higher resolution were attained using metabarcoding. But microhistology still proved a powerful supplementary. Previous studies using metabarcoding usually detected dozens of food items, even as many as more than one hundred species. For instance, 18 taxa prey were identified for leopard cat (Shehzad et al., 2012); 44 plant taxa were recovered in feces of red-headed wood pigeon (Ando et al., 2013); while more than 100 taxa were found in diet studies of brown bear (De Barba et al., 2014). The relatively narrow diet spectrum of herbivorous geese may lead to misunderstanding that this result of our study is merely an artefact due to small sampling effort. However, this result is credible since these two geese species only feed on *Carex* meadow, where the dominant vegetation is *Carex* spp., with other species such as *Poaceae* and dicots (Zhao et al., 2015). Even though other wetland plants exist, they usually composed only a small proportion of the geese diets.

Quantification of food composition is another key concern in dietary analysis. Although the relative percentage of sequences were not truly a quantitative estimate of diet, taxa of the majority sequences in this study were in accord with microhistologic observations, which was considered an efficient way to provide quantitative results (Wang et al., 2013, Discrepancies). This might come from the semi-quantitative nature of metabarcoding methods (Sun et al., 2015).
is likely derived from PCR amplification, which always entails biases caused by universal
primer-template mismatches, annealing temperature or number of PCR cycles (Zhan et al., 2014;
Piñol et al., 2015). Other methods such as shot-gun sequencing or metagenomic sequencing
could be incorporated in the future to give information on abundance of food items (Srivathsan et
al., 2015).

**Implications for Waterbird Conservation and Wetland Management**

For long-distance migratory waterbirds, such as the wild geese in this study, their abundance and
distribution are greatly influenced by diet availability and habitat use (Wang et al., 2013). For
example, waterbirds may be restricted at (forced to leave) certain areas due to favoring (loss) of
particular food (Wang et al., 2013), while the recovery of such food may contribute to return of
bird populations (Noordhuis et al., 2002). Results of both metabarcoding and microhistologic
analysis in this study revealed that Carex and Poaceae were dominant food components which is
in accordance with previous studies. The increasing number of these two geese wintering at the
Shengjin Lake may be attributed to the expansion of Carex meadow, which offers access to
abundant food resources (Zhao et al., 2015). Considering the long-distance migratory character of
these birds, it is important to maintain energy balances and good body conditions in wintering
areas because this might further influence their departure dates and reproductive success after
arriving at breeding areas (Prop, Black & Shimmings, 2003). Based on this, it is important for
wetland managers to maintain suitable habitats and food resources for sustainable conservation of
waterbirds, which highlights the significance of diet information. Our study also indicated that
overlap and dissimilarity existed between the diets of these two geese. As we all know, animals
foraging in the same habitats may compete for limited food resources (Madsen & Mortensen, 1987). This discrepancy of food composition may arise from the avoidance of inter-specific competition (Zhao et al., 2015). However, with the increase of these two species in Shengjin Lake, further research is needed to investigate the mechanisms of food resource partitioning and spatial distribution.

Shengjin Lake is one of the most important wintering sites for tens of thousands of migratory watebirds, while annual life cycles of these birds depend on the whole migratory route, including breeding sites, stop-over sites and wintering sites (Kear, 2006). Thus, a molecular reference library covering all the potential food items along the whole migratory route will be useful both for understanding of wetland connections and waterbird conservation. Besides, the ability of DNA metabarcoding to process lots of samples simultaneously enables rapid analyses and makes this method helpful for waterbird studies.

Acknowledgements

We are very grateful to the stuff of the Shengjin Lake National Nature Reserve for their excellent assistance during the field work. Great thanks to Zhujun Wang and An An for feces collection in the field. We thank Song Yang for collecting plants in the Shengjin Lake Reserve. We also thank Profs. Zhenyu Li and Shuren Zhang for plant identification. Special thanks to Drs. Meijuan Zhao, Xin Wang, Fanjuan Meng and Peihao Cong for preparing the epidermis database and guiding microhistologic analysis.
References


Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods*


Madsen J, Mortensen CE. 1987. Habitat exploitation and interspecific competition of moulting geese in East...


