

DATA NOTE

Open Access



Using next generation sequencing of alpine plants to improve fecal metabarcoding diet analysis for Dall's sheep

Kelly E. Williams^{1,4*} , Damian M. Menning², Eric J. Wald³, Sandra L. Talbot², Kumi L. Rattenbury³ and Laura R. Prugh¹

Abstract

Objectives: Dall's sheep (*Ovis dalli dalli*) are important herbivores in the mountainous ecosystems of northwestern North America, and recent declines in some populations have sparked concern. Our aim was to improve capabilities for fecal metabarcoding diet analysis of Dall's sheep and other herbivores by contributing new sequence data for arctic and alpine plants. This expanded reference library will provide critical reference sequence data that will facilitate metabarcoding diet analysis of Dall's sheep and thus improve understanding of plant-animal interactions in a region undergoing rapid climate change.

Data description: We provide sequences for the chloroplast *rbcl* gene of 16 arctic-alpine vascular plant species that are known to comprise the diet of Dall's sheep. These sequences contribute to a growing reference library that can be used in diet studies of arctic herbivores.

Keywords: Alpine, Arctic, Boreal, Dall's sheep, Diet, Fecal, Metabarcoding, Chloroplast, Plant

Objective

Dall's sheep (*Ovis dalli dalli*) are endemic to alpine ecosystems of northwestern North America, and their populations have been declining in recent decades [1–4]. Climate change may be altering alpine plant communities and contributing to these declines. Dall's sheep have a generalist plant diet; they were observed eating 110 different plant species in the Yukon Territory, Canada through traditional observational methods [5]. However, the diet of Dall's sheep remains relatively poorly characterized and represents a gap in understanding how climate change is affecting plant-animal interactions in alpine ecosystems.

The level of taxonomic resolution of items consumed in a diet study greatly affects ecological analysis [6]. DNA based tools can infer diet composition with higher resolution and reduces cost, time, and effort compared to observational, morphological, and microhistological methods [7, 8]. Specifically, DNA metabarcoding uses universal primers for multispecies identification to mass-amplify DNA barcodes using PCR that are then read using next generation sequencing and assigned to the appropriate taxon [9]. DNA barcoding includes a reference database of potential diet components, providing the capability to identify diet items to a desirable taxonomic resolution, ensuring that all components will be detected and assigned [10]. Next generation sequencing of DNA from fecal samples has been successfully used to characterize diets of a variety of species, including ungulates [11, 12]. However, metabarcoding has not yet been used to assess the diet of Dall's sheep. Lack of sequence data for some arctic/alpine plants known to be grazed

*Correspondence: kel.elizabeth.williams@gmail.com

¹ School of Environmental and Forest Sciences, University of Washington, Seattle, USA

Full list of author information is available at the end of the article



Table 1 Overview of data files for arctic plant *rbcL* sequencing

Label	Name of data file/data set	File types	Data repository and identifier
Data file 1	<i>Elymus borealis</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW538513 [19]
Data file 2	<i>Gentiana propinqua</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW538515 [20]
Data file 3	<i>Juncus mertensianus</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548523 [21]
Data file 4	<i>Luzula arctica</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548524 [22]
Data file 5	<i>Ranunculus kamchaticus</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548525 [23]
Data file 6	<i>Oxytropis scammaniana</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548526 [24]
Data file 7	<i>Packera ogorukensis</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548527 [25]
Data file 8	<i>Penstemon gormanii</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548528 [26]
Data file 9	<i>Saxifraga caespitosa</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548529 [27]
Data file 10	<i>Silene tayloriae</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548530 [28]
Data file 11	<i>Smelowskia integrifolia</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548531 [29]
Data file 12	<i>Stellaria alaskana</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548532 [30]
Data file 13	<i>Taraxacum lyratum</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW548533 [31]
Data file 14	<i>Anemone lithophila</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW526257 [32]
Data file 15	<i>Carex pyrenaica</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW538514 [33]
Data file 16	<i>Elymus latiglumis</i> rbcL contig	*.gb	https://identifiers.org/ncbi/insdc:MW537582 [34]

upon by Dall's sheep currently limits the development and application of metabarcoding for alpine herbivore diet studies.

To improve capabilities for diet analysis of Dall's sheep and other arctic herbivores, we used a python script [13] to identify gaps in archived nucleotide sequence data for species known to comprise the diet of Dall's Sheep, then obtained specimens of 16 species of arctic/alpine vascular plants for which sequence information was missing or underrepresented in publicly archived databases. We then sequenced the *rbcL* gene of the plant chloroplast genome, which is one of the most commonly used barcoding regions for plants [9, 14].

Data description

Plant specimens were obtained from herbarium specimens collected from the various arctic or alpine sites across mainland Alaska (Additional file 1). Plant tissue was extracted at the U. S. Geological Survey Alaska Science Center, employing a CTAB-PVP protocol modified from Stewart and Via [15] as reported by Muñiz-Salazar et al. [16]. Extracts were quantified and shipped to the School of Environmental and Forest Sciences Genetics Lab at the University of Washington for PCR amplification and NexteraXT library preparation for sequencing. The *rbcL* gene region of each specimen was amplified via a two-step PCR protocol [17] with a primary amplification with tailed primers (*rbcLaf*+adaptor, *rbcLr506*+adaptor) followed by a second round of amplification to anneal NexteraXT indices. Amplicons were quantified using a Qubit 4 Fluorometer (ThermoFisher) and diluted with dH₂O to the recommended

starting concentration for library preparation, 0.2 ng/μL (Illumina). Tagmentation, library amplification, and clean-up steps were completed according to the NexteraXT library preparation protocol (Illumina) with a variation of using New England Biolabs AMPure XP beads for cleanup instead of Agentcourt AMPure beads. The libraries were normalized and pooled prior to sequencing on an Illumina Miseq platform. Samples were paired-end sequenced in a 2 × 300 bp format.

Illumina sequence reads were processed in Geneious Prime 2020.2.4. Forward and reverse read files (fastq) were paired upon import, then quality trimmed with BBDuk trimmer (minimum quality 20, minimum overlap 20, minimum length 20). Sequences were normalized, then aligned and assembled using the de novo assembly tool (Geneious Prime). Assembled contigs were uploaded and annotated using BankIt, then submitted to GenBank [18] (Table 1).

Limitations

The following are limitations for these data files:

1. We sequenced one DNA extraction from each plant species.
2. The sequencing project was funded through a grant to train new users on Illumina Nextera sequencing.

Abbreviations

rbcL: Large subunit of ribulose 1, 5 bisphosphate carboxylase/oxygenase (RUBISCO or RuBPCase); CTAB-PVP: DNA extraction method using

cetyltrimethylammonium bromide as a detergent-based extraction buffer and polyvinylpyrrolidone, which is added to remove phenolic compounds from plant DNA extracts [15, 16]; PCR: Polymerase chain reaction; NexteraXT: NexteraXT DNA library preparation kit enables sequencing of small genomes, PCR amplicons, and plasmids (Illumina); Miseq: Illumina Miseq Next Generation Sequencer is an integrated instrument that performs clonal amplification, genomic DNA sequencing, and data analysis with base calling, alignment, variant calling, and reporting in a single run (Illumina).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13104-021-05590-z>.

Additional file 1. Table of information about the plant specimens used for rbcl sequencing.

Acknowledgements

The Illumina team at the University of Washington provided valuable training on Illumina sequencing technology. Plant specimens were obtained from the U.S. Geological Survey and the University of Alaska, Anchorage herbarium. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Authors' contributions

EW, KR, DM, and ST identified need and with KW and LP contributed to study design. ST obtained plant specimens and performed DNA extraction at the U.S. Geological Survey Alaska Science Center, Alaska. DM scanned sequence data archived in GenBank to identify data gaps for candidate plant species. KR and DM chose the final list for analysis based on these data gaps, available information on Dall's sheep diets, and expected plants in habitat where populations have declined (e.g., Brooks Range). KW performed laboratory work for library preparation and sequencing and assembled sequences in Geneious Prime. KW and LP wrote the manuscript, and DM, EW, and ST edited the manuscript. All authors read and approved the final manuscript.

Funding

Illumina and NASA's Arctic and Boreal Vulnerability Experiment program (Grant NNX15AU21A to LRP). The United States Geological Survey and National Park Service provided funding in terms of salary and field and laboratory processes.

Availability of data and materials

Please see Table 1 and references [19–34] for details and links to the data. Please see Additional file 1 for a table of information about the plant specimens used for sequencing.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

There are no competing interests.

Author details

¹School of Environmental and Forest Sciences, University of Washington, Seattle, USA. ²U.S. Geological Survey, Alaska Science Center, Anchorage, USA. ³Arctic Network Inventory & Monitoring Division, National Park Service, Fairbanks, USA. ⁴Present Address: Department of Evolutionary Anthropology, Duke University, Durham, USA.

Received: 11 February 2021 Accepted: 28 April 2021

Published online: 07 May 2021

References

- Alaska Department of Fish and Game (ADFG). Division of Wildlife Conservation, Wildlife Management Report (Juneau, AK: Alaska Department of Fish and Game) Trends in Alaska sheep populations, hunting, and harvests. 2014.
- Koizumi CL, Carey J, Branigan M, Callaghan, K. Status of Dall's sheep (*Ovis Dall'si Dall'si*) in the Northern Richardson Mountains. Yukon Fish and Wildlife Branch Report. 2011. TRC-11–01. Whitehorse, Yukon, Canada.
- Rattenbury KL, Schmidt JH. Declining sheep populations in Alaska's Arctic Parks. *Alaska Park Science* 2017; 16: 67–69. <https://www.nps.gov/articles/aps-16-1-15.htm>.
- Rattenbury KL, Schmidt JH, Swanson DK, Borg BL, Mangipane BA, Soutanes PJ. Delayed spring onset drives declines in abundance and recruitment in a mountain ungulate. *Ecosphere*. 2018;9(11):e02513. <https://doi.org/10.1002/ecs2.2513>.
- Hoefs M, Cowan I. Ecological investigation of a population of Dall's sheep (*Ovis Dall'si Dall'si* Nelson). *Syesis*. 1979;12:1–81.
- Smith MA, Eveleigh ES, McCann KS, Merilo MT, McCarthy PC, Van Rooyen KI. Barcoding a quantified food web: cryptis, concepts, ecology and hypotheses. *PLoS ONE*. 2011;6:e14424.
- Alberdi A, Aizpurua O, Bohmann K, Gopalakrishnan S, Lynggaard C, Nielsen M, et al. Promises and pitfalls of using high-throughput sequencing for diet analysis. *Mol Ecol Resour*. 2019;19:327–48. <https://doi.org/10.1111/1755-0998.12960>.
- Andriollo T, Gillet F, Michaux JR, Reudi M. The menu varies with metabarcoding practices: a case study with bat *Plecotus auritus* 2019. *PLoS ONE*. 2019;14(7):e0219135.
- Pompanon F, Deagle BE, Symondson WOC, Brown DS, Jarman SN. Who is eating what: diet assessment using next generation sequencing. *Mol Ecol*. 2012;21:1931–50. <https://doi.org/10.1111/j.1365-294X.2011.05403.x>.
- Nielsen JM, Clare EL, Hayden B, Brett MT, Kratina P. Diet tracing in ecology: Method comparison and selection. *Methods Ecol Evol*. 2018;9:278–91.
- Kartzinel TR, Chen PA, Coverdale TC, Erickson DL, Kress WJ, Kuzmina ML, et al. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proc Natl Acad Sci USA*. 2015;112:8019–24. <https://doi.org/10.1073/pnas.1503283112>.
- McShea WJ, Sukmasuang R, Erickson DL, Herrmann V, Ngoprasert D, Bhumpakphan N, et al. Metabarcoding reveals diet diversity in an ungulate community in Thailand. *Biotropica*. 2019;51:923–37. <https://doi.org/10.1111/btp.12720>.
- Menning DM, Talbot S. Python scripts for bioinformatics, 2017. U.S. Geological Survey data release. 2018. <https://doi.org/https://doi.org/10.5066/F74F1N24>.
- Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, et al. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol Ecol Resour*. 2009;9:439–57. <https://doi.org/10.1111/j.1755-0998.2008.02439.x>.
- Stewart CN, Via LE. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques*. 1993;14(5):748–50 (PMID: 8512694).
- Muñiz-Salazar RM, Talbot SL, Sage GK, Ward DH, Cabello-Pasini A. Population genetic structure of annual and perennial populations of *Zostera marina* L. along the Pacific coast of Baja California and the Gulf of California. *Mol Ecol*. 2005;14:711–22.
- de Vere N, Jones LE, Gilmore T, Moscrop J, Lowe A, Smith D, et al. Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Sci Rep*. 2017;7:42838. <https://doi.org/10.1038/srep42838>.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Res*. 2013;41(Database issue):D36–42. <https://doi.org/10.1093/nar/gks1195> (PMID: 23193287).
- Williams K, et al. *Elymus borealis* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW538513>. 2021.
- Williams K, et al. *Gentiana propinqua* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW538515>. 2021.
- Williams K, et al. *Juncus mertensianus* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548523>. 2021.
- Williams K, et al. *Luzula arctica* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548524>. 2021.

23. Williams K, et al. *Ranunculus kamchaticus* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548525>. 2021.
24. Williams K, et al. *Oxytropis scammaniana* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548526>. 2021.
25. Williams K, et al. *Packera ogotorukensis* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548527>. 2021.
26. Williams K, et al. *Penstemon gormanii* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548528>. 2021.
27. Williams K, et al. *Saxifraga caespitosa* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548529>. 2021.
28. Williams K, et al. *Silene tayloriae* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548530>. 2021.
29. Williams K, et al. *Smelowskia integrifolia* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548531>. 2021.
30. Williams K, et al. *Stellaria alaskana* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548532>. 2021.
31. Williams K, et al. *Taraxacum lyratum* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW548533> (2021).
32. Williams K, et al. *Anemone lithophilia* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW526257>. 2021.
33. Williams K, et al. *Carex pyrenaica* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW538514>. 2021.
34. Williams K, et al. *Elymus latiglumis* rbcl partial GenBank <https://identifiers.org/ncbi/insdc:MW537582>. 2021.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

